

SUPPLEMENT

Table of Contents

Supplemental Figures

| | |
|--|-----|
| Figure S1. Methylation by RPMM class..... | iii |
| Figure S2. Cross-validation of RPMM classes generated using Infinium methylation array data from the primary and validation study sets..... | iv |

Supplemental Tables

| | |
|---|------|
| Table S1. Description and comparison of study populations..... | v |
| Table S2. Cross-validation of RPMM CpG classes..... | vi |
| Table S3. Distribution of CpG loci by bioinformatically-derived class..... | vii |
| Table S4. Association of aging/exposures with CpG island shore methylation.... | viii |
| Table S5. Marginal model results for the association of methylation with sequence features and age, adjusted for gender..... | ix |
| Table S6. Marginal model results for the association of methylation with sequence features gender and hair dye use, adjusted for age..... | x |
| Table S7. Marginal model results for the association of methylation with sequence features and ever-smoking, adjusted for age and gender..... | xi |
| Table S8. Marginal model results for the association of methylation with sequence features and pack-years of smoking, adjusted for age and gender..... | xii |
| Table S9. Marginal model results for the association of methylation with sequence features and consumption of ≤ 6.5 alcoholic drinks per week (compared with non-drinkers), adjusted for age and gender..... | xiii |
| Table S10. Marginal model results for the association of methylation with sequence features and consumption of > 6.5 alcoholic drinks per week (compared with non-drinkers), adjusted for age and gender..... | xiv |
| Table S11. Marginal model results for the association of methylation with sequence features and arsenic exposure, adjusted for age and gender..... | xv |

| | |
|--|-------|
| Table S12. Marginal model results for the association of methylation with sequence features and selenium exposure, adjusted for age and gender..... | xvi |
| Table S13. Marginal model results for the association of methylation with sequence features and tanning lamp use, adjusted for age and gender..... | xvii |
| Table S14. Marginal model results for the association of methylation with sequence features and lifetime painful sunburns, adjusted for age and gender..... | xviii |
| Supplemental Analysis | |
| Analysis S1. Performance assessment of recursively partitioned mixture model (RPMM) versus metric hierarchical clustering for classification of high-density DNA methylation data..... | xix |

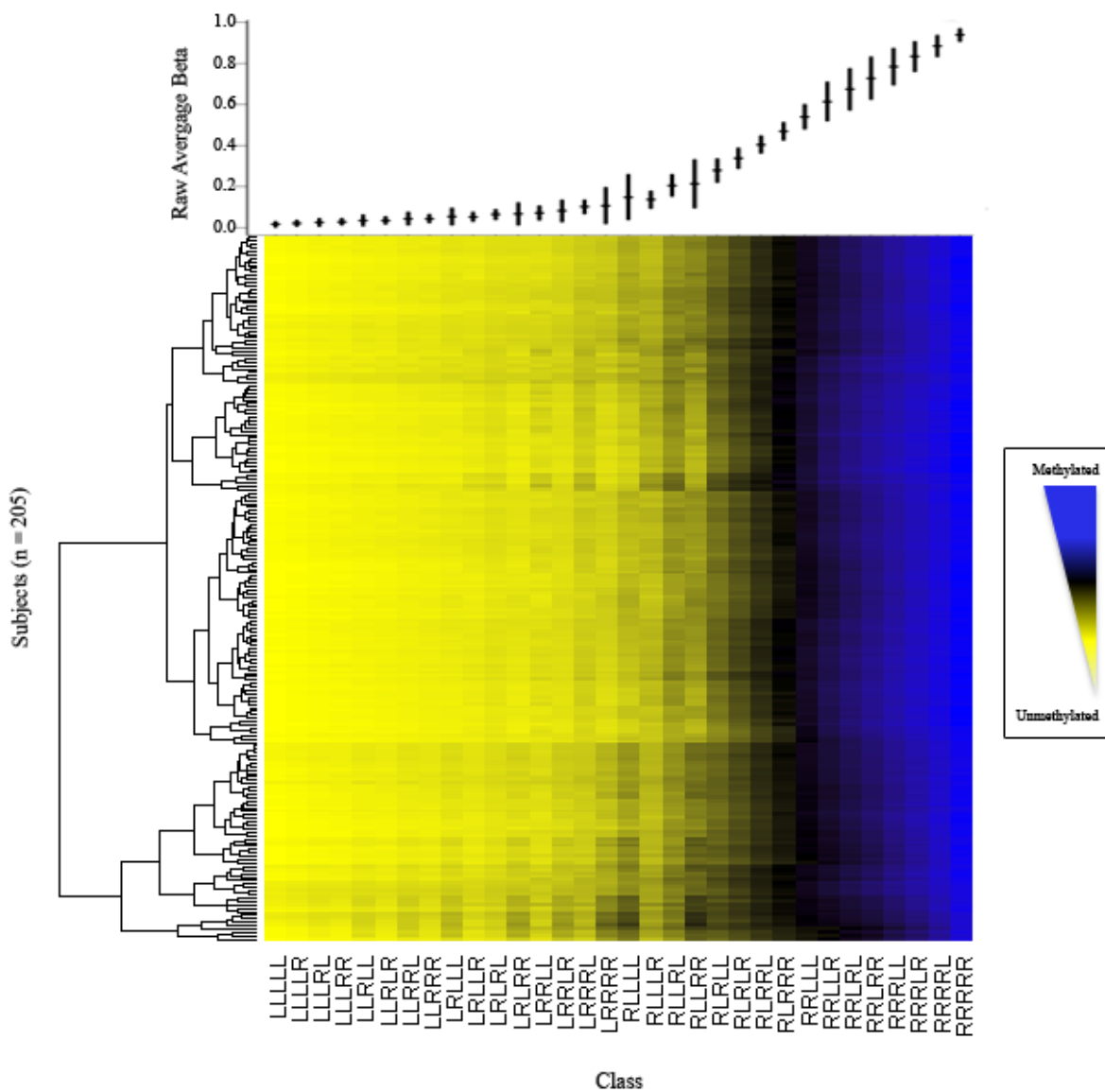


Figure S1. Methylation by RPMM class. The heatmap shows average methylation (beta) by class, with the 32 RPMM-based methylation classes shown in columns and the subjects represented by rows. Raw average beta values and 95% confidence intervals are displayed per class at the top of the figure.

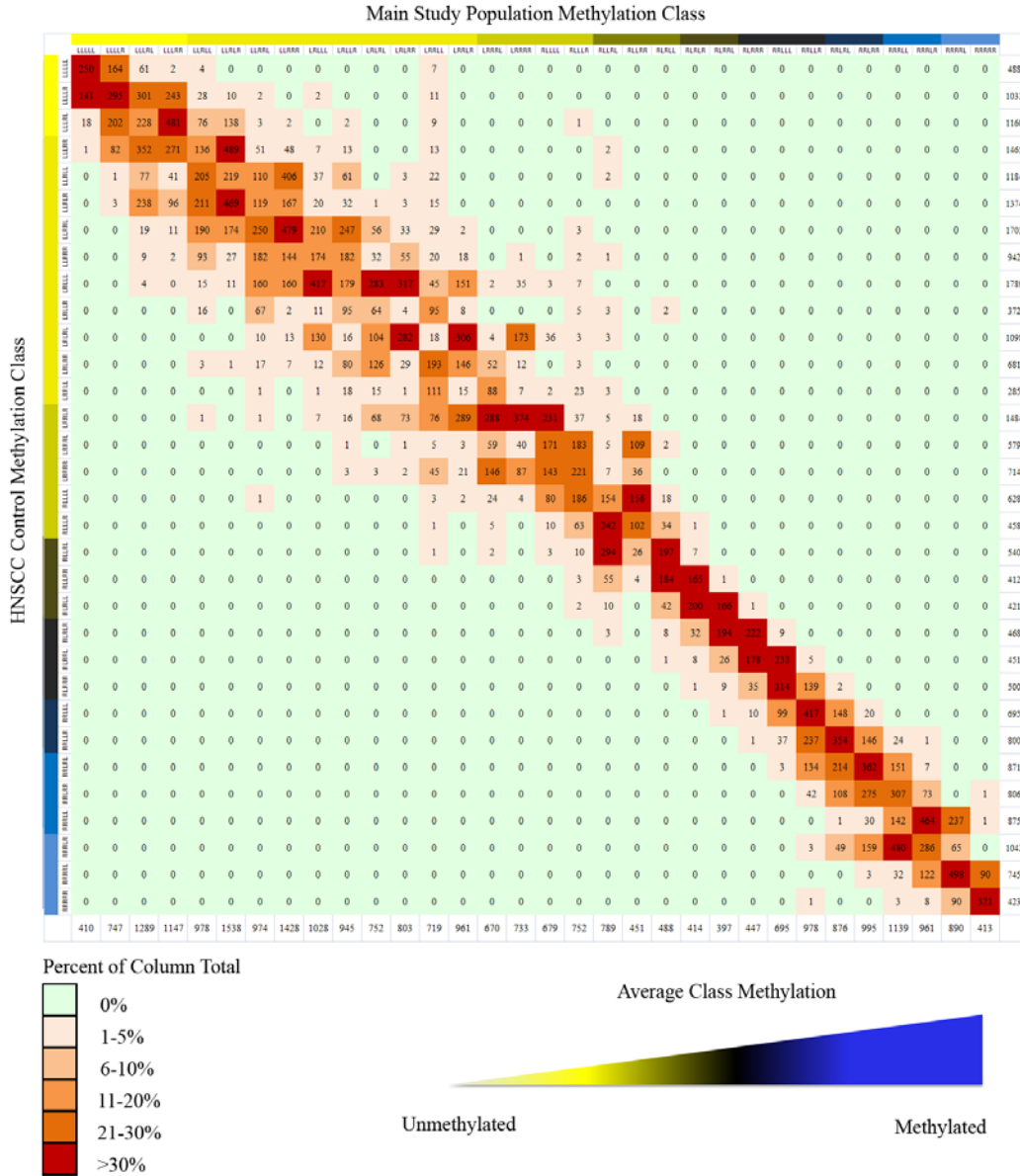


Figure S2. Cross-validation of RPMM classes generated using Infinium methylation array data from the primary and validation study sets. Each box shows the number of overlapping CpGs by class, presented as a percent of the column totals (indicated by the key on the bottom right). The columns represent the classes generated for the 205 primary study subjects, and the rows represent the classes for the 92 validation subjects. The total number of CpGs per class for the primary study subjects (column totals) are shown across the bottom margin and the total number of CpGs per class for the validation study subjects (row totals) are provided in the right margin.

Table S1. Description and comparison of study populations. Describes and compares personal characteristics and exposures of the primary study subjects and those used for the cross-validations (validation subjects).

| Exposure | Primary Study Subjects (n = 205) | Validation Subjects (n = 92) | P-Value |
|--|---|-------------------------------------|----------------------|
| Age | | | |
| Median years (range) | 64 (28-74) | 59 (32-86) | 0.69 ^a |
| Gender, n (%) | | | |
| Female | 65 (31.7%) | 28 (30.4%) | 0.89 ^b |
| Male | 140 (62.3%) | 64 (69.6%) | |
| Race/ethnicity | | | |
| Caucasian (non-Hispanic) | 205 (100%) | 85 (92.4%) | < 0.001 ^b |
| Other | 0 (0%) | 7 (7.6%) | |
| Cigarette smoking, n (%) | | | |
| Never | 58 (28.3%) | 32 (34.8%) | 0.28 ^b |
| Ever | 147 (71.7%) | 60 (65.2%) | |
| Median pack-years (range) | 26 (0.1-176.0) | 24.5 (0.5-85.0) | 0.66 ^a |
| Alcohol consumption, n (%) | | | |
| Non-drinker | 58 (28.6%) | 9 (10.0%) | 0.001 ^b |
| ≤ 6.5 drinks per week | 74 (36.5%) | 42 (46.7%) | |
| > 6.5 drinks per week | 71 (35.0%) | 39 (43.3%) | |
| Hair dye use, n (%) | | | |
| Never | 145 (71.1%) | --- | --- |
| Ever | 59 (28.9%) | --- | --- |
| Arsenic (toenail) | | | |
| Median µg/g (range) | 0.09 (0.02-0.81) | --- | --- |
| Selenium (toenail) | | | |
| Median µg/g (range) | 0.85 (0.57-2.84) | --- | --- |
| Ultraviolet radiation exposure | | | |
| Tanning lamp use, n (%) | | | |
| Never | 137 (86.7%) | --- | --- |
| Ever | 21 (13.3%) | --- | --- |
| Median lifetime painful sunburns (range) | 1 (0-222) | --- | --- |

^a Wilcoxon rank-sum test

^b Fisher's exact test

^c Restricted to ever-smokers

Table S2. Cross-validation of RPMM CpG classes. Infinium methylation array data was available for blood DNA from a second set of 92 healthy subjects (validation subjects), which was employed for cross-validation of the RPMM CpG classes derived from the 205 primary study subjects in the main analysis. The omnibus test for the relationship between class methylation and exposure was performed using the CpG classes generated from each methylation array dataset, each applied to each study population.

| Exposure | Omnibus P-Value (Spearman) | | | | Omnibus P-Value (Multiple Regression) | | | |
|----------------------------------|----------------------------|---------------------|------------------|---------------------|---------------------------------------|---------------------|------------------|---------------------|
| | Primary RPMM | | Validation RPMM | | Primary RPMM | | Validation RPMM | |
| | Primary Subjects | Validation Subjects | Primary Subjects | Validation Subjects | Primary Subjects | Validation Subjects | Primary Subjects | Validation Subjects |
| Age (years) ^a | < 0.001 | 0.03 | < 0.001 | 0.06 | < 0.001 | 0.04 | < 0.001 | 0.05 |
| Female gender ^b | 0.04 | 0.05 | 0.04 | 0.07 | 0.23 | 0.02 | 0.04 | 0.04 |
| Cigarette smoking | | | | | | | | |
| Ever-smoker ^c | 0.06 | 0.69 | 0.20 | 0.64 | 0.29 | 0.59 | 0.37 | 0.62 |
| Pack-years ^{c,d} | 0.006 | 0.77 | 0.01 | 0.83 | 0.41 | 0.74 | 0.35 | 0.63 |
| Alcohol consumption ^e | | | | | | | | |
| Non-drinker (reference) | --- | --- | --- | --- | --- | --- | --- | --- |
| ≤ 6.5 drinks/week | 0.29 | 0.10 | 0.22 | 0.08 | 0.49 | 0.24 | 0.36 | 0.22 |
| > 6.5 drinks/week | 0.34 | 0.06 | 0.26 | 0.04 | 0.47 | 0.19 | 0.44 | 0.19 |

^a The multiple regression model was adjusted for gender

^b The multiple regression model was adjusted for age

^c The multiple regression model was adjusted for age and gender

^d Restricted to ever-smokers

^e Median = 6.5 alcoholic drinks per week among drinkers

Table S3. Distribution of CpG loci by bioinformatically-derived class.

| Class Number | Genomic Attributes | n _{loci} | Percent Total |
|--------------|--------------------------|-------------------|---------------|
| 1 | CGI PcG LINE2 TFBS | 7 | < 0.1% |
| 2 | CGI TFBS | 12958 | 48.90% |
| 3 | CGI PcG TFBS | 2692 | 10.20% |
| 4 | CGI PcG | 330 | 1.20% |
| 5 | CGI MIR TFBS | 253 | 1.00% |
| 6 | CGI | 2347 | 8.90% |
| 7 | LINE2 MIR TFBS | 1 | < 0.1% |
| 8 | CGI LINE2 TFBS | 130 | 0.50% |
| 9 | CGI LINE2 | 33 | 0.10% |
| 10 | CGI PcG MIR | 4 | < 0.1% |
| 11 | CGI PcG MIR TFBS | 33 | 0.10% |
| 12 | CGI MIR | 71 | 0.30% |
| 13 | CGI LINE1 TFBS | 35 | 0.10% |
| 14 | PcG LINE1 | 2 | < 0.1% |
| 15 | PcG LINE1 TFBS | 1 | < 0.1% |
| 16 | PcG TFBS | 315 | 1.20% |
| 17 | PcG MIR | 8 | < 0.1% |
| 18 | CGI LINE1 | 24 | 0.10% |
| 19 | PcG LINE2 TFBS | 6 | < 0.1% |
| 20 | PcG | 64 | 0.20% |
| 21 | CGI ALU TFBS | 173 | 0.70% |
| 22 | CGI PcG ALU TFBS | 21 | 0.10% |
| 23 | TFBS | 4436 | 16.70% |
| 24 | PcG MIR TFBS | 13 | < 0.1% |
| 25 | CGI PcG ALU | 10 | < 0.1% |
| 26 | CGI PcG LINE1 TFBS | 1 | < 0.1% |
| 27 | MIR TFBS | 184 | 0.70% |
| 28 | None | 1638 | 6.20% |
| 29 | CGI PcG LINE2 | 4 | < 0.1% |
| 30 | LINE1 TFBS | 48 | 0.20% |
| 31 | LINE2 TFBS | 103 | 0.40% |
| 32 | LINE1 | 74 | 0.30% |
| 33 | MIR | 99 | 0.40% |
| 34 | CGI ALU | 121 | 0.50% |
| 35 | LINE2 | 88 | 0.30% |
| 36 | ALU TFBS | 79 | 0.30% |
| 37 | PcG LINE2 | 3 | < 0.1% |
| 38 | ALU | 68 | 0.30% |
| 39 | CGI PcG LINE1 | 5 | < 0.1% |
| 40 | PcG ALU | 3 | < 0.1% |
| 41 | PcG ALU TFBS | 1 | < 0.1% |
| Total | | 26486 | 100% |

Abbreviations: CGI = CpG island; TFBS = transcription factor binding sites; PcG = polycomb group protein target gene; MIR = mammalian wide-interspersed repeat

Note: Classes are denoted by the applicable attributes separated by a “|” (e.g. a class of CpGs located in CGI and *LINE-1* element would be symbolized as CGI|LINE1)

Supplemental Table S4. Association of aging/exposures with CpG island shore methylation.

CpG island shores were defined as sequences within 2kb distance of CpG islands, with the north shore in the 5' direction (4264 CpGs) and south shore in the 3' direction (3739 CpGs) of the island using the annotation of the Illumina 450K methylation array, which includes annotation for 94.4% of the CpG loci found on the HumanMethylation27k BeadArray. The association of each exposure with average methylation of the shores was assessed using multiple linear regression models. P-values for each respective adjusted estimate are presented in the table below.

| Exposure | p-value | |
|--|--------------------|--------------------|
| | North Shore | South Shore |
| Age ^a | 0.32 | 0.27 |
| Female gender ^b | 0.09 | 0.08 |
| Smoking | | |
| Ever-smoking ^c | 0.24 | 0.25 |
| Pack-years ^{c,d} | 0.92 | 0.95 |
| Alcohol consumption | | |
| Non-drinker | <i>Reference</i> | <i>Reference</i> |
| ≤ 6.5 drinks per week ^c | 0.26 | 0.29 |
| > 6.5 drinks per week ^c | 0.07 | 0.07 |
| Hair dye ever-use ^c | 0.67 | 0.78 |
| Arsenic ^c | 0.55 | 0.52 |
| Selenium ^c | 0.50 | 0.50 |
| Ultraviolet radiation exposure | | |
| Tanning lamp ever-use ^c | 0.28 | 0.30 |
| Lifetime painful sunburns ^c | 0.50 | 0.48 |

^a Adjusted for gender

^b Adjusted for age and hair dye use

^c Adjusted for age and gender

^d Restricted to ever-smokers

Table S5. Marginal model results for the association of methylation with sequence features and age, adjusted for gender.

| | Estimate | Mean | SE | Z | P-value |
|--------------|-----------------|-------------|-----------|----------|----------------|
| (Intercept) | 0.90284 | 0.90287 | 0.00249 | 363.04 | < 0.00001 |
| CGI | -0.51457 | -0.51458 | 0.00320 | -161.02 | < 0.00001 |
| PcG | -0.00010 | -0.00007 | 0.00069 | -0.10 | 0.91916 |
| TFBS | -0.06879 | -0.06879 | 0.00053 | -128.78 | < 0.00001 |
| LINE1 | 0.15308 | 0.15310 | 0.00145 | 105.52 | < 0.00001 |
| LINE2 | 0.09217 | 0.09219 | 0.00053 | 174.12 | < 0.00001 |
| Alu | 0.41501 | 0.41503 | 0.00299 | 138.76 | < 0.00001 |
| MIR | 0.07012 | 0.07014 | 0.00047 | 150.74 | < 0.00001 |
| Age | -0.00607 | -0.00604 | 0.00195 | -3.09 | 0.00197 |
| Age*CGI | 0.00593 | 0.00591 | 0.00252 | 2.35 | 0.01893 |
| Age*PcG | 0.00181 | 0.00176 | 0.00049 | 3.61 | 0.00031 |
| Age*TFBS | 0.00035 | 0.00035 | 0.00040 | 0.86 | 0.39036 |
| Age*LINE1 | -0.00111 | -0.00108 | 0.00104 | -1.04 | 0.30067 |
| Age*LINE2 | -0.00143 | -0.00143 | 0.00040 | -3.61 | 0.00031 |
| Age*Alu | -0.00279 | -0.00283 | 0.00230 | -1.23 | 0.21903 |
| Age*MIR | -0.00172 | -0.00173 | 0.00037 | -4.65 | < 0.00001 |
| Female | -0.00630 | -0.00633 | 0.00478 | -1.32 | 0.18524 |
| Female*CGI | 0.01427 | 0.01415 | 0.00671 | 2.11 | 0.03497 |
| Female*PcG | -0.00204 | -0.00215 | 0.00127 | -1.70 | 0.08979 |
| Female*TFBS | 0.00225 | 0.00225 | 0.00103 | 2.19 | 0.02853 |
| Female*LINE1 | -0.00556 | -0.00557 | 0.00275 | -2.02 | 0.04314 |
| Female*LINE2 | -0.00105 | -0.00109 | 0.00104 | -1.05 | 0.29445 |
| Female*Alu | -0.01418 | -0.01415 | 0.00626 | -2.26 | 0.02372 |
| Female*MIR | -0.00152 | -0.00152 | 0.00096 | -1.59 | 0.11243 |

Abbreviations: CGI = CpG islands; TFBS = transcription factor binding sites; PcG = polycomb group protein target gene; LINE = long interspersed repeat elements; MIR = mammalian wide-interspersed repeat

Note: exposure/sequence feature interactions are shown as exposure*feature (i.e. the effect of age on methylation CpGs located in CGI is denoted by Age*CGI)

Table S6. Marginal model results for the association of methylation with sequence features gender and hair dye use, adjusted for age.

| | Estimate | Mean | SE | Z | P-value |
|--------------------|-----------------|-------------|-----------|----------|----------------|
| (Intercept) | 0.90357 | 0.90349 | 0.00247 | 365.69 | < 0.00001 |
| CGI | -0.51571 | -0.51561 | 0.00319 | -161.54 | < 0.00001 |
| PcG | -0.00034 | -0.00035 | 0.00072 | -0.49 | 0.62303 |
| TFBS | -0.06895 | -0.06893 | 0.00054 | -127.90 | < 0.00001 |
| LINE1 | 0.15367 | 0.15367 | 0.00145 | 106.19 | < 0.00001 |
| LINE2 | 0.09230 | 0.09228 | 0.00052 | 178.49 | < 0.00001 |
| Alu | 0.41586 | 0.41578 | 0.00298 | 139.60 | < 0.00001 |
| MIR | 0.07015 | 0.07014 | 0.00048 | 145.81 | < 0.00001 |
| Age | -0.00612 | -0.00607 | 0.00189 | -3.22 | 0.00129 |
| Age*CGI | 0.00584 | 0.00571 | 0.00241 | 2.36 | 0.01804 |
| Age*PcG | 0.00168 | 0.00170 | 0.00047 | 3.61 | 0.00031 |
| Age*TFBS | 0.00034 | 0.00031 | 0.00038 | 0.82 | 0.41467 |
| Age*LINE1 | -0.00114 | -0.00109 | 0.00101 | -1.08 | 0.28052 |
| Age*LINE2 | -0.00144 | -0.00142 | 0.00040 | -3.55 | 0.00038 |
| Age*Alu | -0.00276 | -0.00262 | 0.00221 | -1.19 | 0.23476 |
| Age*MIR | -0.00173 | -0.00172 | 0.00036 | -4.83 | < 0.00001 |
| Female | -0.00107 | -0.00158 | 0.00532 | -0.30 | 0.76633 |
| Female*CGI | 0.00781 | 0.00864 | 0.00757 | 1.14 | 0.25365 |
| Female*PcG | -0.00237 | -0.00235 | 0.00131 | -1.80 | 0.07245 |
| Female*TFBS | 0.00130 | 0.00140 | 0.00117 | 1.20 | 0.23012 |
| Female*LINE1 | -0.00149 | -0.00188 | 0.00298 | -0.63 | 0.52900 |
| Female*LINE2 | -0.00017 | -0.00026 | 0.00117 | -0.22 | 0.82372 |
| Female*Alu | -0.00901 | -0.00979 | 0.00688 | -1.42 | 0.15462 |
| Female*MIR | -0.00124 | -0.00138 | 0.00119 | -1.16 | 0.24528 |
| Hair dye use | -0.00884 | -0.00873 | 0.00645 | -1.35 | 0.17548 |
| Hair dye use*CGI | 0.01120 | 0.01095 | 0.00817 | 1.34 | 0.18025 |
| Hair dye use*PcG | 0.00080 | 0.00079 | 0.00125 | 0.63 | 0.52733 |
| Hair dye use*TFBS | 0.00164 | 0.00162 | 0.00130 | 1.25 | 0.21199 |
| Hair dye use*LINE1 | -0.00689 | -0.00675 | 0.00323 | -2.09 | 0.03648 |
| Hair dye use*LINE2 | -0.00151 | -0.00148 | 0.00130 | -1.14 | 0.25354 |
| Hair dye use*Alu | -0.00889 | -0.00867 | 0.00730 | -1.19 | 0.23493 |
| Hair dye use*MIR | -0.00045 | -0.00040 | 0.00122 | -0.33 | 0.74329 |

Abbreviations: CGI = CpG islands; TFBS = transcription factor binding sites; PcG = polycomb group protein target gene; LINE = long interspersed repeat elements; MIR = mammalian wide-interspersed repeat

Note: exposure/sequence feature interactions are shown as exposure*feature (i.e. the effect of age on methylation CpGs located in CGI is denoted by Age*CGI)

Table S7. Marginal model results for the association of methylation with sequence features and ever-smoking, adjusted for age and gender.

| | Estimate | Mean | SE | Z | P-value |
|--------------------|-----------------|-------------|-----------|----------|----------------|
| (Intercept) | 0.90628 | 0.90601 | 0.00416 | 217.96 | < 0.00001 |
| CGI | -0.51559 | -0.51533 | 0.00543 | -94.96 | < 0.00001 |
| PcG | 0.00023 | 0.00026 | 0.00118 | 0.22 | 0.82589 |
| TFBS | -0.06886 | -0.06884 | 0.00090 | -76.87 | < 0.00001 |
| LINE1 | 0.15416 | 0.15402 | 0.00248 | 62.19 | < 0.00001 |
| LINE2 | 0.09288 | 0.09283 | 0.00085 | 109.16 | < 0.00001 |
| Alu | 0.41522 | 0.41508 | 0.00509 | 81.6 | < 0.00001 |
| MIR | 0.07040 | 0.07037 | 0.00083 | 84.63 | < 0.00001 |
| Age | -0.00603 | -0.00609 | 0.00196 | -3.11 | 0.00187 |
| Age*CGI | 0.00592 | 0.00592 | 0.00261 | 2.27 | 0.02327 |
| Age*PcG | 0.00181 | 0.00181 | 0.00045 | 4.06 | 0.00005 |
| Age*TFBS | 0.00034 | 0.00034 | 0.00040 | 0.85 | 0.39493 |
| Age*LINE1 | -0.00109 | -0.00106 | 0.00108 | -0.99 | 0.32429 |
| Age*LINE2 | -0.00142 | -0.00144 | 0.00042 | -3.42 | 0.00062 |
| Age*Alu | -0.00279 | -0.00275 | 0.00238 | -1.16 | 0.24758 |
| Age*MIR | -0.00172 | -0.00172 | 0.00038 | -4.55 | 0.00001 |
| Female | -0.00696 | -0.00697 | 0.00490 | -1.42 | 0.15441 |
| Female*CGI | 0.01446 | 0.01440 | 0.00680 | 2.12 | 0.03424 |
| Female*PcG | -0.00210 | -0.00214 | 0.00128 | -1.66 | 0.09630 |
| Female*TFBS | 0.00227 | 0.00227 | 0.00105 | 2.17 | 0.03025 |
| Female*LINE1 | -0.00577 | -0.00575 | 0.00278 | -2.06 | 0.03908 |
| Female*LINE2 | -0.00119 | -0.00119 | 0.00104 | -1.15 | 0.25047 |
| Female*Alu | -0.01422 | -0.01419 | 0.00631 | -2.25 | 0.02460 |
| Female*MIR | -0.00157 | -0.00155 | 0.00094 | -1.65 | 0.09860 |
| Ever Smoking | -0.00451 | -0.00435 | 0.00461 | -0.94 | 0.34526 |
| Ever-smoking*CGI | 0.00133 | 0.00117 | 0.00611 | 0.19 | 0.84831 |
| Ever-smoking*PcG | -0.00044 | -0.00047 | 0.00125 | -0.38 | 0.70629 |
| Ever-smoking*TFBS | 0.00009 | 0.00009 | 0.00100 | 0.09 | 0.92786 |
| Ever-smoking*LINE1 | -0.00141 | -0.00130 | 0.00269 | -0.48 | 0.62875 |
| Ever-smoking*LINE2 | -0.00094 | -0.00090 | 0.00098 | -0.92 | 0.35523 |
| Ever-smoking*Alu | -0.00028 | -0.00020 | 0.00574 | -0.03 | 0.97270 |
| Ever-smoking*MIR | -0.00036 | -0.00034 | 0.00092 | -0.37 | 0.71199 |

Abbreviations: CGI = CpG islands; TFBS = transcription factor binding sites; PcG = polycomb group protein target gene; LINE = long interspersed repeat elements; MIR = mammalian wide-interspersed repeat

Note: exposure/sequence feature interactions are shown as exposure*feature (i.e. the effect of age on methylation CpGs located in CGI is denoted by Age*CGI)

Table S8. Marginal model results for the association of methylation with sequence features and pack-years of smoking, adjusted for age and gender.

| | Estimate | Mean | SE | Z | P-value |
|------------------|-----------------|-------------|-----------|----------|----------------|
| (Intercept) | 0.90164 | 0.90198 | 0.00466 | 193.42 | < 0.00001 |
| CGI | -0.51425 | -0.51442 | 0.00541 | -95.08 | < 0.00001 |
| PcG | -0.00014 | -0.00013 | 0.00115 | -0.11 | 0.91222 |
| TFBS | -0.06886 | -0.06889 | 0.00091 | -75.33 | < 0.00001 |
| LINE1 | 0.15097 | 0.15104 | 0.00239 | 63.15 | < 0.00001 |
| LINE2 | 0.09189 | 0.09194 | 0.00091 | 101.58 | < 0.00001 |
| Alu | 0.41409 | 0.41423 | 0.00484 | 85.64 | < 0.00001 |
| MIR | 0.07010 | 0.07014 | 0.00082 | 85.98 | < 0.00001 |
| Age | -0.00675 | -0.00662 | 0.00235 | -2.81 | 0.00490 |
| Age*CGI | 0.00593 | 0.00574 | 0.00336 | 1.70 | 0.08824 |
| Age*PcG | 0.00185 | 0.00188 | 0.00063 | 2.96 | 0.00308 |
| Age*TFBS | 0.00040 | 0.00036 | 0.00053 | 0.67 | 0.50053 |
| Age*LINE1 | -0.00157 | -0.00148 | 0.00138 | -1.07 | 0.28505 |
| Age*LINE2 | -0.00147 | -0.00145 | 0.00050 | -2.92 | 0.00348 |
| Age*Alu | -0.00304 | -0.00285 | 0.00312 | -0.91 | 0.36189 |
| Age*MIR | -0.00163 | -0.00162 | 0.00050 | -3.26 | 0.00111 |
| Female | -0.01102 | -0.01163 | 0.00633 | -1.84 | 0.06638 |
| Female*CGI | 0.02046 | 0.02118 | 0.00915 | 2.31 | 0.02068 |
| Female*PcG | -0.00142 | -0.00140 | 0.00169 | -0.83 | 0.40566 |
| Female*TFBS | 0.00301 | 0.00312 | 0.00139 | 2.24 | 0.02490 |
| Female*LINE1 | -0.00814 | -0.00846 | 0.00365 | -2.32 | 0.02027 |
| Female*LINE2 | -0.00270 | -0.00278 | 0.00133 | -2.09 | 0.03625 |
| Female*Alu | -0.01912 | -0.01974 | 0.00834 | -2.37 | 0.01797 |
| Female*MIR | -0.00264 | -0.00272 | 0.00124 | -2.19 | 0.02828 |
| Pack-years | 0.00005 | 0.00004 | 0.00012 | 0.34 | 0.73733 |
| Pack-years*CGI | -0.00006 | -0.00006 | 0.00014 | -0.40 | 0.68630 |
| Pack-years*PcG | -0.00001 | -0.00001 | 0.00003 | -0.35 | 0.72560 |
| Pack-years*TFBS | -0.00001 | -0.00001 | 0.00002 | -0.23 | 0.81421 |
| Pack-years*LINE1 | 0.00008 | 0.00008 | 0.00006 | 1.38 | 0.16644 |
| Pack-years*LINE2 | 0.00001 | 0.00001 | 0.00002 | 0.44 | 0.65675 |
| Pack-years*Alu | 0.00008 | 0.00008 | 0.00012 | 0.63 | 0.52898 |
| Pack-years*MIR | 0.00001 | 0.00001 | 0.00002 | 0.32 | 0.74532 |

Abbreviations: CGI = CpG islands; TFBS = transcription factor binding sites; PcG = polycomb group protein target gene; LINE = long interspersed repeat elements; MIR = mammalian wide-interspersed repeat

Note: exposure/sequence feature interactions are shown as exposure*feature (i.e. the effect of age on methylation CpGs located in CGI is denoted by Age*CGI)

Table S9. Marginal model results for the association of methylation with sequence features and consumption of ≤ 6.5 alcoholic drinks per week (compared with non-drinkers), adjusted for age and gender.

| | Estimate | Mean | SE | Z | P-value |
|-------------------------|----------|----------|---------|--------|-----------|
| (Intercept) | 0.89858 | 0.89857 | 0.00430 | 208.91 | < 0.00001 |
| CGI | -0.51314 | -0.51326 | 0.00516 | -99.41 | < 0.00001 |
| PcG | 0.00015 | 0.00017 | 0.00119 | 0.14 | 0.88849 |
| TFBS | -0.06857 | -0.06859 | 0.00082 | -83.98 | < 0.00001 |
| LINE1 | 0.15164 | 0.15168 | 0.00229 | 66.29 | < 0.00001 |
| LINE2 | 0.09247 | 0.09248 | 0.00085 | 108.50 | < 0.00001 |
| Alu | 0.41520 | 0.41527 | 0.00461 | 90.11 | < 0.00001 |
| MIR | 0.06934 | 0.06933 | 0.00083 | 83.31 | < 0.00001 |
| Age | -0.00520 | -0.00527 | 0.00246 | -2.14 | 0.03249 |
| Age*CGI | 0.00472 | 0.00475 | 0.00320 | 1.48 | 0.13809 |
| Age*PcG | 0.00212 | 0.00212 | 0.00063 | 3.35 | 0.00080 |
| Age*TFBS | 0.00024 | 0.00025 | 0.00053 | 0.47 | 0.63869 |
| Age*LINE1 | -0.00032 | -0.00032 | 0.00136 | -0.24 | 0.81117 |
| Age*LINE2 | -0.00113 | -0.00115 | 0.00052 | -2.19 | 0.02829 |
| Age*Alu | -0.00207 | -0.00208 | 0.00310 | -0.67 | 0.50132 |
| Age*MIR | -0.00149 | -0.00150 | 0.00048 | -3.16 | 0.00157 |
| Female | -0.00719 | -0.00720 | 0.00600 | -1.20 | 0.23035 |
| Female*CGI | 0.01614 | 0.01640 | 0.00773 | 2.12 | 0.03393 |
| Female*PcG | -0.00280 | -0.00277 | 0.00153 | -1.81 | 0.07096 |
| Female*TFBS | 0.00266 | 0.00270 | 0.00121 | 2.23 | 0.02592 |
| Female*LINE1 | -0.00595 | -0.00605 | 0.00313 | -1.93 | 0.05325 |
| Female*LINE2 | -0.00112 | -0.00111 | 0.00125 | -0.89 | 0.37229 |
| Female*Alu | -0.01683 | -0.01707 | 0.00713 | -2.40 | 0.01661 |
| Female*MIR | -0.00127 | -0.00126 | 0.00115 | -1.10 | 0.27344 |
| ≤ 6.5 drinks/week | 0.00433 | 0.00441 | 0.00548 | 0.81 | 0.42072 |
| < 6.5 drinks/week*CGI | -0.00139 | -0.00137 | 0.00716 | -0.19 | 0.84804 |
| < 6.5 drinks/week*PcG | 0.00032 | 0.00026 | 0.00135 | 0.19 | 0.84881 |
| < 6.5 drinks/week*TFBS | -0.00019 | -0.00017 | 0.00118 | -0.15 | 0.88248 |
| < 6.5 drinks/week*LINE1 | 0.00080 | 0.00084 | 0.00309 | 0.27 | 0.78511 |
| < 6.5 drinks/week*LINE2 | -0.00062 | -0.00065 | 0.00115 | -0.56 | 0.57437 |
| < 6.5 drinks/week*Alu | -0.00088 | -0.00081 | 0.00680 | -0.12 | 0.90568 |
| < 6.5 drinks/week*MIR | 0.00072 | 0.00072 | 0.00106 | 0.68 | 0.49501 |

Abbreviations: CGI = CpG islands; TFBS = transcription factor binding sites; PcG = polycomb group protein target gene; LINE = long interspersed repeat elements; MIR = mammalian wide-interspersed repeat

Note: exposure/sequence feature interactions are shown as exposure*feature (i.e. the effect of age on methylation CpGs located in CGI is denoted by Age*CGI)

Table S10. Marginal model results for the association of methylation with sequence features and consumption of > 6.5 alcoholic drinks per week (compared with non-drinkers), adjusted for age and gender.

| | Estimate | Mean | SE | Z | P-value |
|-------------------------|-----------------|-------------|-----------|----------|----------------|
| (Intercept) | 0.89690 | 0.89678 | 0.00438 | 204.70 | < 0.00001 |
| CGI | -0.51171 | -0.51171 | 0.00538 | -95.03 | < 0.00001 |
| PcG | -0.00043 | -0.00048 | 0.00132 | -0.36 | 0.71930 |
| TFBS | -0.06826 | -0.06825 | 0.00085 | -80.29 | < 0.00001 |
| LINE1 | 0.15135 | 0.15130 | 0.00247 | 61.29 | < 0.00001 |
| LINE2 | 0.09241 | 0.09239 | 0.00090 | 102.39 | < 0.00001 |
| Alu | 0.41317 | 0.41316 | 0.00490 | 84.36 | < 0.00001 |
| MIR | 0.06931 | 0.06930 | 0.00084 | 82.51 | < 0.00001 |
| Age | -0.00645 | -0.00652 | 0.00270 | -2.42 | 0.01557 |
| Age*CGI | 0.00710 | 0.00724 | 0.00331 | 2.19 | 0.02875 |
| Age*PcG | 0.00228 | 0.00228 | 0.00065 | 3.50 | 0.00046 |
| Age*TFBS | 0.00043 | 0.00044 | 0.00050 | 0.88 | 0.37681 |
| Age*LINE1 | -0.00192 | -0.00197 | 0.00145 | -1.36 | 0.17502 |
| Age*LINE2 | -0.00160 | -0.00164 | 0.00057 | -2.88 | 0.00401 |
| Age*Alu | -0.00408 | -0.00418 | 0.00288 | -1.45 | 0.14644 |
| Age*MIR | -0.00182 | -0.00186 | 0.00050 | -3.74 | 0.00019 |
| Female | -0.00239 | -0.00214 | 0.00562 | -0.38 | 0.70378 |
| Female*CGI | 0.01106 | 0.01091 | 0.00820 | 1.33 | 0.18319 |
| Female*PcG | -0.00157 | -0.00143 | 0.00174 | -0.82 | 0.41339 |
| Female*TFBS | 0.00179 | 0.00174 | 0.00120 | 1.44 | 0.14930 |
| Female*LINE1 | -0.00411 | -0.00403 | 0.00339 | -1.19 | 0.23439 |
| Female*LINE2 | -0.00066 | -0.00068 | 0.00132 | -0.52 | 0.60635 |
| Female*Alu | -0.01064 | -0.01049 | 0.00756 | -1.39 | 0.16523 |
| Female*MIR | -0.00097 | -0.00099 | 0.00118 | -0.84 | 0.39910 |
| > 6.5 drinks/week | 0.00851 | 0.00827 | 0.00585 | 1.41 | 0.15790 |
| > 6.5 drinks/week*CGI | -0.00355 | -0.00320 | 0.00770 | -0.42 | 0.67791 |
| > 6.5 drinks/week*PcG | 0.00001 | -0.00007 | 0.00151 | -0.05 | 0.96376 |
| > 6.5 drinks/week*TFBS | -0.00075 | -0.00068 | 0.00120 | -0.57 | 0.57146 |
| > 6.5 drinks/week*LINE1 | 0.00308 | 0.00291 | 0.00330 | 0.88 | 0.37744 |
| > 6.5 drinks/week*LINE2 | -0.00029 | -0.00031 | 0.00115 | -0.27 | 0.79056 |
| > 6.5 drinks/week*Alu | 0.00238 | 0.00203 | 0.00686 | 0.30 | 0.76740 |
| > 6.5 drinks/week*MIR | 0.00109 | 0.00106 | 0.00110 | 0.97 | 0.33187 |

Abbreviations: CGI = CpG islands; TFBS = transcription factor binding sites; PcG = polycomb group protein target gene; LINE = long interspersed repeat elements; MIR = mammalian wide-interspersed repeat

Note: exposure/sequence feature interactions are shown as exposure*feature (i.e. the effect of age on methylation CpGs located in CGI is denoted by Age*CGI)

Table S11. Marginal model results for the association of methylation with sequence features and arsenic exposure, adjusted for age and gender.

| | Estimate | Mean | SE | Z | P-value |
|---------------|-----------------|-------------|-----------|----------|----------------|
| (Intercept) | 0.90894 | 0.90945 | 0.00371 | 245.34 | < 0.00001 |
| CGI | -0.52034 | -0.52084 | 0.00440 | -118.32 | < 0.00001 |
| PcG | -0.00041 | -0.00040 | 0.00096 | -0.41 | 0.67980 |
| TFBS | -0.06966 | -0.06974 | 0.00073 | -96.08 | < 0.00001 |
| LINE1 | 0.15549 | 0.15562 | 0.00192 | 81.08 | < 0.00001 |
| LINE2 | 0.09293 | 0.09307 | 0.00078 | 119.88 | < 0.00001 |
| Alu | 0.41997 | 0.42042 | 0.00391 | 107.54 | < 0.00001 |
| MIR | 0.07090 | 0.07096 | 0.00067 | 106.65 | < 0.00001 |
| Age | -0.00786 | -0.00786 | 0.00201 | -3.91 | 0.00009 |
| Age*CGI | 0.00770 | 0.00769 | 0.00270 | 2.85 | 0.00440 |
| Age*PcG | 0.00179 | 0.00175 | 0.00053 | 3.33 | 0.00087 |
| Age*TFBS | 0.00063 | 0.00062 | 0.00043 | 1.45 | 0.14831 |
| Age*LINE1 | -0.00188 | -0.00187 | 0.00110 | -1.69 | 0.09089 |
| Age*LINE2 | -0.00172 | -0.00175 | 0.00041 | -4.24 | 0.00002 |
| Age*Alu | -0.00428 | -0.00429 | 0.00247 | -1.74 | 0.08238 |
| Age*MIR | -0.00196 | -0.00196 | 0.00038 | -5.12 | < 0.00001 |
| Female | -0.00588 | -0.00578 | 0.00495 | -1.17 | 0.24323 |
| Female*CGI | 0.01409 | 0.01373 | 0.00693 | 1.98 | 0.04743 |
| Female*PcG | -0.00217 | -0.00210 | 0.00130 | -1.61 | 0.10773 |
| Female*TFBS | 0.00222 | 0.00215 | 0.00107 | 2.00 | 0.04525 |
| Female*LINE1 | -0.00541 | -0.00526 | 0.00282 | -1.86 | 0.06239 |
| Female*LINE2 | -0.00099 | -0.00102 | 0.00106 | -0.96 | 0.33642 |
| Female*Alu | -0.01418 | -0.01380 | 0.00637 | -2.17 | 0.03039 |
| Female*MIR | -0.00153 | -0.00151 | 0.00095 | -1.59 | 0.11221 |
| Arsenic | -0.05237 | -0.05711 | 0.03095 | -1.85 | 0.06501 |
| Arsenic*CGI | 0.04864 | 0.05318 | 0.03283 | 1.62 | 0.10523 |
| Arsenic*PcG | 0.00343 | 0.00264 | 0.00666 | 0.40 | 0.69192 |
| Arsenic*TFBS | 0.00731 | 0.00820 | 0.00541 | 1.52 | 0.12937 |
| Arsenic*LINE1 | -0.02036 | -0.02180 | 0.01286 | -1.70 | 0.09001 |
| Arsenic*LINE2 | -0.00629 | -0.00745 | 0.00608 | -1.23 | 0.22048 |
| Arsenic*Alu | -0.04162 | -0.04635 | 0.02770 | -1.67 | 0.09428 |
| Arsenic*MIR | -0.00645 | -0.00680 | 0.00450 | -1.51 | 0.13121 |

Abbreviations: CGI = CpG islands; TFBS = transcription factor binding sites; PcG = polycomb group protein target gene; LINE = long interspersed repeat elements; MIR = mammalian wide-interspersed repeat

Note: exposure/sequence feature interactions are shown as exposure*feature (i.e. the effect of age on methylation CpGs located in CGI is denoted by Age*CGI); Arsenic was measured from toenail clippings ($\mu\text{g/g}$)

Table S12. Marginal model results for the association of methylation with sequence features and selenium exposure, adjusted for age and gender.

| | Estimate | Mean | SE | Z | P-value |
|----------------|----------|----------|---------|--------|-----------|
| (Intercept) | 0.89250 | 0.89274 | 0.0123 | 72.57 | < 0.00001 |
| CGI | -0.50833 | -0.50633 | 0.01557 | -32.51 | < 0.00001 |
| PcG | 0.00039 | 0.00003 | 0.00307 | 0.01 | 0.99141 |
| TFBS | -0.06741 | -0.06709 | 0.00248 | -27.03 | < 0.00001 |
| LINE1 | 0.15079 | 0.15034 | 0.00617 | 24.35 | < 0.00001 |
| LINE2 | 0.09160 | 0.09098 | 0.00253 | 35.98 | < 0.00001 |
| Alu | 0.40981 | 0.40731 | 0.01467 | 27.77 | < 0.00001 |
| MIR | 0.06958 | 0.06912 | 0.00232 | 29.79 | < 0.00001 |
| Age | -0.00670 | -0.00669 | 0.00186 | -3.59 | 0.00033 |
| Age*CGI | 0.00660 | 0.00660 | 0.00236 | 2.80 | 0.00515 |
| Age*PcG | 0.00175 | 0.00175 | 0.00048 | 3.65 | 0.00026 |
| Age*TFBS | 0.00046 | 0.00046 | 0.00037 | 1.24 | 0.21539 |
| Age*LINE1 | -0.00142 | -0.00139 | 0.00100 | -1.40 | 0.16262 |
| Age*LINE2 | -0.00158 | -0.00156 | 0.00039 | -4.05 | 0.00005 |
| Age*Alu | -0.00333 | -0.00330 | 0.00215 | -1.53 | 0.12572 |
| Age*MIR | -0.00182 | -0.00182 | 0.00037 | -4.96 | < 0.00001 |
| Female | -0.00675 | -0.00667 | 0.00496 | -1.34 | 0.17891 |
| Female*CGI | 0.01507 | 0.01520 | 0.00697 | 2.18 | 0.02929 |
| Female*PcG | -0.00224 | -0.00230 | 0.00137 | -1.67 | 0.09404 |
| Female*TFBS | 0.00236 | 0.00239 | 0.00108 | 2.20 | 0.02770 |
| Female*LINE1 | -0.00583 | -0.00589 | 0.00279 | -2.11 | 0.03471 |
| Female*LINE2 | -0.00112 | -0.00110 | 0.00111 | -0.99 | 0.32050 |
| Female*Alu | -0.01507 | -0.01523 | 0.00653 | -2.33 | 0.01959 |
| Female*MIR | -0.00164 | -0.00165 | 0.00098 | -1.69 | 0.09168 |
| Selenium | 0.01242 | 0.01199 | 0.01344 | 0.89 | 0.37229 |
| Selenium*CGI | -0.00775 | -0.01005 | 0.01670 | -0.60 | 0.54745 |
| Selenium*PcG | -0.00050 | -0.00006 | 0.00333 | -0.02 | 0.98536 |
| Selenium*TFBS | -0.00168 | -0.00205 | 0.00264 | -0.78 | 0.43751 |
| Selenium*LINE1 | 0.00287 | 0.00339 | 0.00651 | 0.52 | 0.60248 |
| Selenium*LINE2 | 0.00075 | 0.00145 | 0.00282 | 0.51 | 0.60796 |
| Selenium*Alu | 0.00651 | 0.00940 | 0.01577 | 0.60 | 0.55117 |
| Selenium*MIR | 0.00071 | 0.00125 | 0.00258 | 0.48 | 0.62885 |

Abbreviations: CGI = CpG islands; TFBS = transcription factor binding sites; PcG = polycomb group protein target gene; LINE = long interspersed repeat elements; MIR = mammalian wide-interspersed repeat

Note: exposure/sequence feature interactions are shown as exposure*feature (i.e. the effect of age on methylation CpGs located in CGI is denoted by Age*CGI); Selenium was measured from toenail clippings ($\mu\text{g/g}$)

Table S13. Marginal model results for the association of methylation with sequence features and tanning lamp use, adjusted for age and gender.

| | Estimate | Mean | SE | Z | P-value |
|------------------------|-----------------|-------------|-----------|----------|----------------|
| (Intercept) | 0.89701 | 0.89673 | 0.00290 | 309.32 | < 0.00001 |
| CGI | -0.50902 | -0.50870 | 0.00407 | -125.08 | < 0.00001 |
| PcG | -0.00173 | -0.00175 | 0.00067 | -2.62 | 0.00890 |
| TFBS | -0.06758 | -0.06753 | 0.00066 | -102.92 | < 0.00001 |
| LINE1 | 0.15084 | 0.15069 | 0.00172 | 87.48 | < 0.00001 |
| LINE2 | 0.09166 | 0.09160 | 0.00067 | 136.92 | < 0.00001 |
| Alu | 0.40991 | 0.40965 | 0.00377 | 108.53 | < 0.00001 |
| MIR | 0.06928 | 0.06922 | 0.00058 | 118.61 | < 0.00001 |
| Age | -0.00756 | -0.00754 | 0.00224 | -3.37 | 0.00077 |
| Age*CGI | 0.00769 | 0.00762 | 0.00299 | 2.55 | 0.01078 |
| Age*PcG | 0.00196 | 0.00195 | 0.00050 | 3.87 | 0.00011 |
| Age*TFBS | 0.00059 | 0.00059 | 0.00045 | 1.30 | 0.19508 |
| Age*LINE1 | -0.00198 | -0.00198 | 0.00119 | -1.67 | 0.09537 |
| Age*LINE2 | -0.00165 | -0.00163 | 0.00046 | -3.56 | 0.00037 |
| Age*Alu | -0.00423 | -0.00418 | 0.00266 | -1.57 | 0.11663 |
| Age*MIR | -0.00194 | -0.00195 | 0.00044 | -4.46 | < 0.00001 |
| Female | -0.00304 | -0.00280 | 0.00532 | -0.53 | 0.59912 |
| Female*CGI | 0.01131 | 0.01125 | 0.00751 | 1.50 | 0.13418 |
| Female*PcG | -0.00151 | -0.00153 | 0.00132 | -1.15 | 0.24848 |
| Female*TFBS | 0.00162 | 0.00163 | 0.00112 | 1.45 | 0.14758 |
| Female*LINE1 | -0.00415 | -0.00412 | 0.00295 | -1.39 | 0.16304 |
| Female*LINE2 | -0.00080 | -0.00075 | 0.00117 | -0.64 | 0.51933 |
| Female*Alu | -0.01190 | -0.01195 | 0.00686 | -1.74 | 0.08154 |
| Female*MIR | -0.00081 | -0.00077 | 0.00106 | -0.73 | 0.46410 |
| Tanning lamp use | -0.00293 | -0.00272 | 0.00733 | -0.37 | 0.71056 |
| Tanning lamp use*CGI | 0.00862 | 0.00810 | 0.01095 | 0.74 | 0.45959 |
| Tanning lamp use*PcG | 0.00394 | 0.00398 | 0.00197 | 2.02 | 0.04356 |
| Tanning lamp use*TFBS | 0.00091 | 0.00082 | 0.00147 | 0.56 | 0.57530 |
| Tanning lamp use*LINE1 | -0.00375 | -0.00356 | 0.00380 | -0.94 | 0.34843 |
| Tanning lamp use*LINE2 | -0.00143 | -0.00137 | 0.00161 | -0.85 | 0.39490 |
| Tanning lamp use*Alu | -0.00715 | -0.00663 | 0.00974 | -0.68 | 0.49635 |
| Tanning lamp use*MIR | -0.00177 | -0.00173 | 0.00148 | -1.17 | 0.24075 |

Abbreviations: CGI = CpG islands; TFBS = transcription factor binding sites; PcG = polycomb group protein target gene; LINE = long interspersed repeat elements; MIR = mammalian wide-interspersed repeat

Note: exposure/sequence feature interactions are shown as exposure*feature (i.e. the effect of age on methylation CpGs located in CGI is denoted by Age*CGI)

Table S14. Marginal model results for the association of methylation with sequence features and lifetime painful sunburns, adjusted for age and gender.

| | Estimate | Mean | SE | Z | P-value |
|------------------------|-----------------|-------------|-----------|----------|----------------|
| (Intercept) | 0.89782 | 0.89778 | 0.00298 | 301.55 | < 0.00001 |
| CGI | -0.50879 | -0.50889 | 0.00404 | -125.86 | < 0.00001 |
| PcG | -0.00134 | -0.00139 | 0.00070 | -1.98 | 0.04817 |
| TFBS | -0.06763 | -0.06763 | 0.00065 | -103.63 | < 0.00001 |
| LINE1 | 0.15073 | 0.15072 | 0.00176 | 85.78 | < 0.00001 |
| LINE2 | 0.09170 | 0.09174 | 0.00068 | 135.00 | < 0.00001 |
| Alu | 0.40981 | 0.40988 | 0.00374 | 109.63 | < 0.00001 |
| MIR | 0.06920 | 0.06920 | 0.00059 | 117.89 | < 0.00001 |
| Age | -0.00756 | -0.00755 | 0.00217 | -3.48 | 0.00051 |
| Age*CGI | 0.00726 | 0.00722 | 0.00294 | 2.46 | 0.01394 |
| Age*PcG | 0.00166 | 0.00165 | 0.00048 | 3.47 | 0.00052 |
| Age*TFBS | 0.00056 | 0.00056 | 0.00045 | 1.25 | 0.21091 |
| Age*LINE1 | -0.00179 | -0.00179 | 0.00121 | -1.48 | 0.13950 |
| Age*LINE2 | -0.00157 | -0.00154 | 0.00047 | -3.32 | 0.00091 |
| Age*Alu | -0.00390 | -0.00388 | 0.00270 | -1.43 | 0.15144 |
| Age*MIR | -0.00181 | -0.00180 | 0.00043 | -4.17 | 0.00003 |
| Female | -0.00264 | -0.00271 | 0.00550 | -0.49 | 0.62150 |
| Female*CGI | 0.01157 | 0.01202 | 0.00747 | 1.61 | 0.10774 |
| Female*PcG | -0.00106 | -0.00114 | 0.00135 | -0.84 | 0.39904 |
| Female*TFBS | 0.00159 | 0.00168 | 0.00113 | 1.48 | 0.13813 |
| Female*LINE1 | -0.00429 | -0.00445 | 0.00300 | -1.48 | 0.13774 |
| Female*LINE2 | -0.00087 | -0.00090 | 0.00125 | -0.72 | 0.47103 |
| Female*Alu | -0.01206 | -0.01252 | 0.00688 | -1.82 | 0.06859 |
| Female*MIR | -0.00100 | -0.00103 | 0.00110 | -0.93 | 0.34996 |
| Painful sunburns | -0.00014 | -0.00012 | 0.00011 | -1.06 | 0.29058 |
| Painful sunburns*CGI | 0.00010 | 0.00010 | 0.00016 | 0.60 | 0.54845 |
| Painful sunburns*PcG | 0.00000 | 0.00000 | 0.00002 | 0.20 | 0.84108 |
| Painful sunburns*TFBS | 0.00002 | 0.00002 | 0.00003 | 0.64 | 0.52306 |
| Painful sunburns*LINE1 | -0.00004 | -0.00003 | 0.00007 | -0.49 | 0.62077 |
| Painful sunburns*LINE2 | -0.00002 | -0.00002 | 0.00003 | -0.83 | 0.40763 |
| Painful sunburns*Alu | -0.00009 | -0.00009 | 0.00015 | -0.60 | 0.54686 |
| Painful sunburns*MIR | -0.00001 | -0.00001 | 0.00003 | -0.30 | 0.76289 |

Abbreviations: CGI = CpG islands; TFBS = transcription factor binding sites; PcG = polycomb group protein target gene; LINE = long interspersed repeat elements; MIR = mammalian wide-interspersed repeat

Note: exposure/sequence feature interactions are shown as exposure*feature (i.e. the effect of age on methylation CpGs located in CGI is denoted by Age*CGI)

Supplemental Analysis S1. Performance assessment of recursively partitioned mixture model (RPMM) versus metric hierarchical clustering for classification of high-density DNA methylation data.

For the present manuscript, we sought to cluster 26,486 autosomal CpG loci based on methylation data from Illumina Infinium HumanMethylation27k BeadArray across 205 primary study samples. Although metric (nonparametric) hierarchical clustering is a well-characterized method, it does not scale to tens of thousands of cases; consequently, we used a Gaussian variant of Recursively Partitioned Mixture Model (RPMM), a hierarchical mixture-model algorithm described by Houseman et al. (*BMC Bioinformatics* 2008;9:365) and implemented in the R library RPMM (<http://cran.r-project.org/web/packages/RPMM/index.html>). Standardized average betas (i.e. z-scores computed using CpG-specific means and standard deviations) for 26,486 CpGs were clustered via RPMM according to their relative variation of 205 beta values, as described below. Thus, the CpGs were hierarchically clustered based on their pattern of relative variation. The resulting hierarchy of classes was pruned to 5 binary levels, resulting in $2^5=32$ classes of CpGs. Note that we compared the consistency of RPMM clustering to that of metric clustering (using Euclidean distance with Ward's linkage) by pairwise analysis of 100 resampling experiments, each of which was analyzed using both RPMM and Ward's method. We sampled 1000 probe sets at a time for each experiment, and on a pairwise basis between sampling runs, used the adjusted Rand index (Rand WM. *JASA* 1971;66(336):846–50) to compare the consistency of the clustering of CpGs that were sampled in both runs (i.e. the intersection of the sampled CpGs between two experiments). Note that a separate Rand index was computed for RPMM and for Ward's method, for each pair of experiments. Thus the mean adjusted Rand index for each of the two methods was computed by averaging 4950 unique pairs of experiments; standard errors were computed from the approximate sampling distribution obtained by bootstrapping the 100 individual experiments and averaging the resulting pairwise comparisons. Mean adjusted Rand index for RPMM was 0.560 (se = 0.006) while for metric hierarchical clustering it was 0.442 (se = 0.004). Additionally, to assess performance with fewer hierarchical classes, the comparison was repeated pruning at 2 binary levels (4 classes), yielding a mean adjusted Rand Index of 0.881 (se = 0.006) for RPMM versus 0.574 (se = 0.009) for metric hierarchical clustering. Thus, RPMM also appeared to provide more consistent clustering than the more common nonparametric approach.