Supplementary Materials

This document contains supplementary materials for the manuscript:

Epigenome-wide analysis identifies DNA methylation mediators of treatment-related cardiometabolic risk in survivors of childhood cancer

Tiffany Eulalio[1], Yoonji Kim[1], Xiaoxi Meng[1], Noel-Marie Plonski[1], Kyla Shelton[1], Heather Mulder[2], Emily Plyler[2], John Easton[2], Xiang Chen[2], Jinghui Zhang[2], Emily Walker[3], Geoffrey Neale[3], Min Ni[4], John T. Lucas Jr.[5], Nilanjan Chatterjee[6], Ziqiao Wang[6], Deokumar Srivastava[7], Bonnie Ky[8], Stephanie B. Dixon[1,4], Kirsten K. Ness[1], Melissa M. Hudson[1,4], Gregory T. Armstrong[1], Zhaoming Wang[1,2,9]*

[1]Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN, USA.

[2]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA.

[3]Hartwell Center, St. Jude Children's Research Hospital, Memphis, TN, USA.

[4]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA.

[5]Department of Radiation Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA.

[6]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.

[7]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA.

[8]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[9]Health Informatics Institute, Morsani College of Medicine & Bellini College of Artificial Intelligence, Cybersecurity and Computing, University of South Florida, Tampa, Florida 33612

Corresponding author:

Zhaoming Wang, PhD

*Present affiliation: Health Informatics Institute, Morsani College of Medicine & Bellini College of Artificial Intelligence, Cybersecurity and Computing, University of South Florida, Tampa, Florida 33612, USA.

Phone: 813-396-0209, Email: zhaomingwang@usf.edu

Contents:

Supplementary Methods

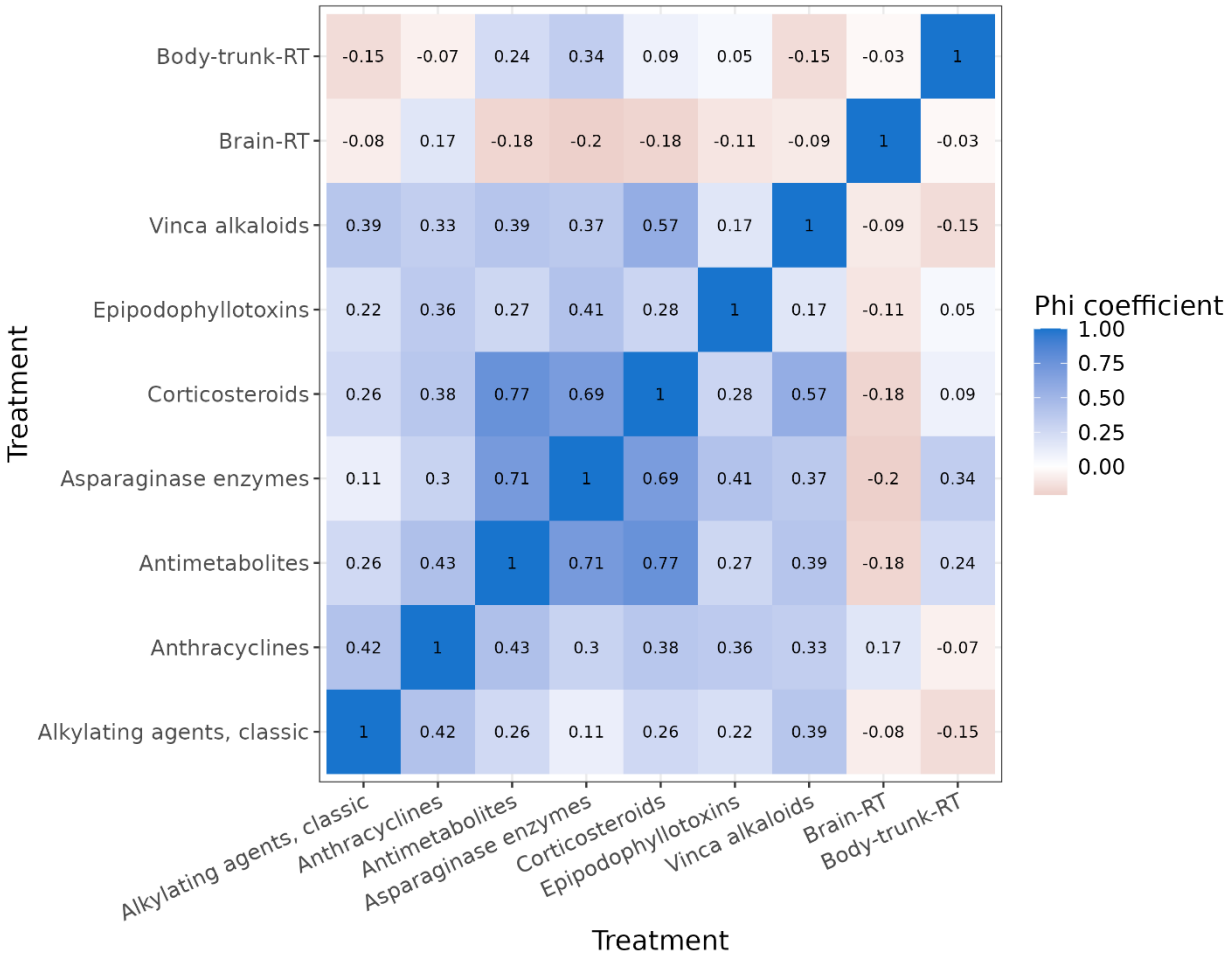
Methods of clinical assessment and severity grading for cardiometabolic outcomes

Cardiometabolic Outcome	Clinical Assessment	CTCAE Grading Rubric
Glucose Metabolism[$]	Fasting glucose and insulin, Hemoglobin A1c	1: Asymptomatic, clinical or diagnostic observations only, pharmacologic intervention not indicated or initiated (e.g., dietary modification); 2: Symptomatic, oral agent indicated or initiated; 3: Severe symptoms, insulin indicated or initiated; 4: Life threatening consequences, urgent intervention indicated or initiated; 5: Death
Body Mass Index (Obesity or Overweight)	Height measured using a wall-mounted stadiometer, weight measured on electric scale (weight in kilograms/height in square meters)	For age >=20-years, 1: Not applicable; 2: BMI 25-29.9 kg/m2; 3: BMI 30-39.9 kg/m2; 4: BMI >=40 kg/m2; 5: Not applicable For age 2-<20-years, 1: Not applicable; 2: BMI >85th percentile <95th percentile; 3: BMI >95th percentile; 4: Not applicable; 5: Not applicable
Blood Pressure (Hypertension)	Resting blood pressure (average of three measurements)	1: Prehypertension (systolic BP 120-139 mm Hg or diastolic BP 80-89 mm Hg) from resting BP in HPL; 2: Stage 1 hypertension (systolic BP 140-159 mm Hg or diastolic BP 90-99 mm Hg); medical intervention indicated or initiated, recurrent or persistent (>=24 hrs), symptomatic increase by >20 mm Hg (diastolic) or to >140/90 mm Hg if previously WNL, monotherapy indicated or initiated, for Pediatric: recurrent or persistent (>=24 hrs) BP >ULN, monotherapy indicated or initiated; 3: Stage 2 hypertension (systolic BP >=160 mm Hg or diastolic BP >=100 mm Hg); medical intervention indicated; more than one drug or more intensive therapy than previously used indicated or initiated, for Pediatric: Same as adult; 4: Life-threatening consequences (e.g., malignant hypertension, transient or permanent neurologic deficit, hypertensive crisis); urgent intervention indicated; 5: Death attributed to hypertension

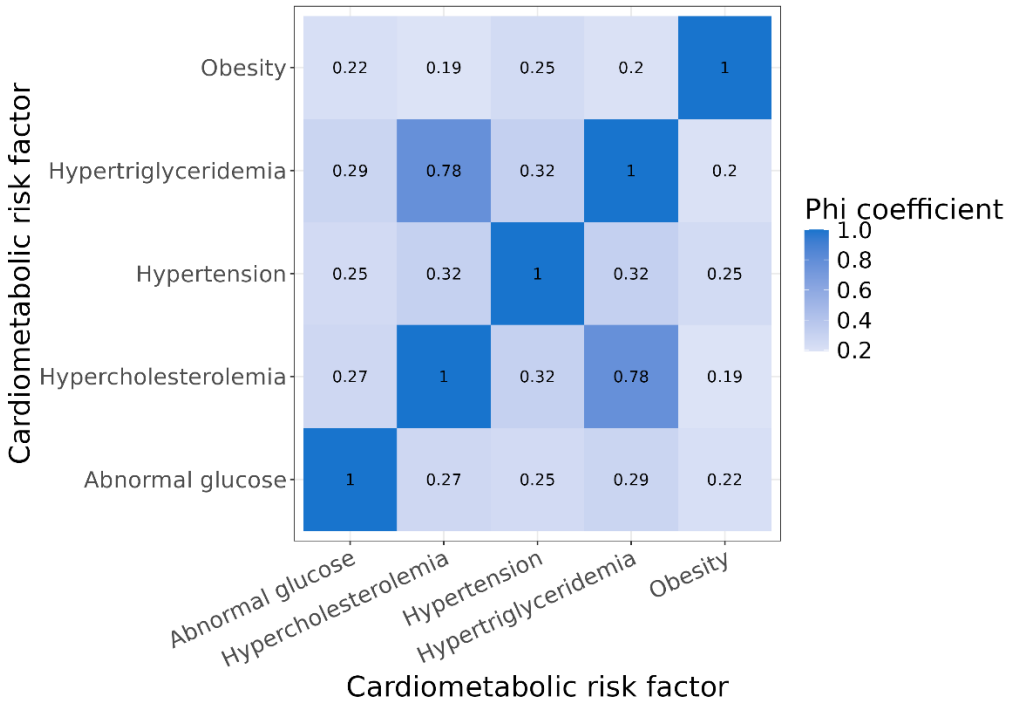
Dyslipidemia (Hypertriglyceridemia)	Fasting lipid panel including triglycerides	1: 150 mg/dL-300 mg/dL; 2: >300 mg/dL-500 mg/dL or treatment with one lipid lowering agent; 3: >500 mg/dL-1000 mg/dL or treatment with >=2 lipid lowering agents; 4: >1000 mg/dL; life-threatening consequences
Dyslipidemia (Hypercholesterolemia)	Fasting lipid panel including total cholesterol, LDL, HDL	1: >200 mg/dL-300 mg/dL; 2: >300-400 mg/dL or treatment with one lipid lowering agent; 3: >400-500 mg/dL or treatment with >=2 lipid lowering agent; 4: >500 mg/dL

Abbreviations: BG, blood glucose; BP, blood pressure; CABG, coronary artery bypass graft; CTCAE, Common Terminology Criteria for Adverse Events; EF, ejection fraction; HbA1c, glycated hemoglobin test; HDL, high-density lipoprotein; HPL, human placental lactogen; Kg/m2, LDL, low-density lipoprotein; Mg/dl, milligrams per deciliter; mm, millimeter; OGTT, oral glucose tolerance test; ULN, upper limit of normal; WNL, within normal limits. [$]Inclusion criteria are the following: 1) Impaired fasting glucose: fasting blood glucose (BG) 100-125 mg/dL in isolation; 2) Pre-diabetes: fasting BG 100-125 mg/dL AND Hemoglobin A1C 5.7-6.4% OR OGTT 140-199 mg/dL; 3) Diabetes mellitus: fasting BG ≥126 mg/dL on 2 separate tests OR random glucose ≥200 mg/dL or HbA1c ≥6.5%

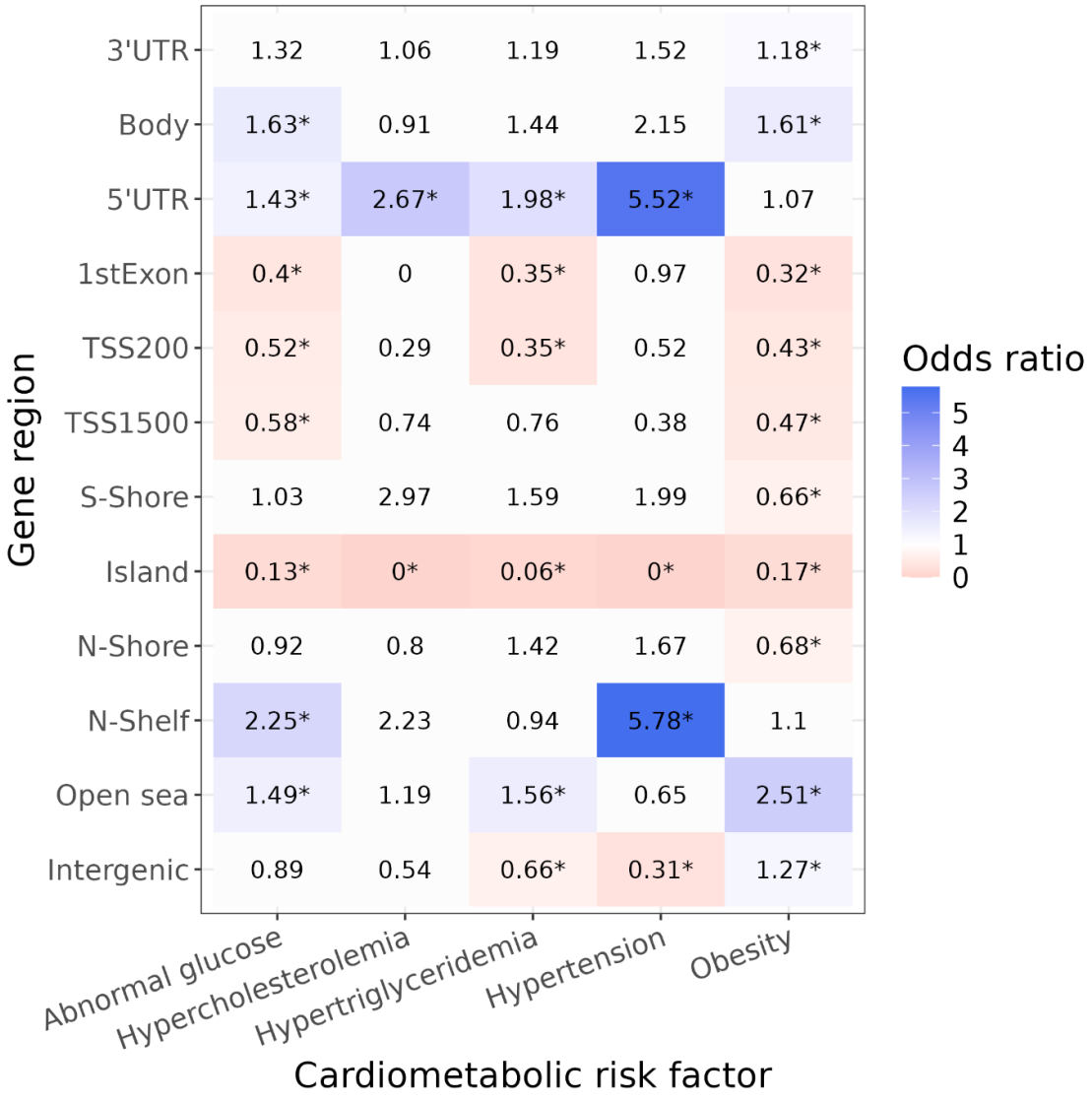
Supplementary Figures



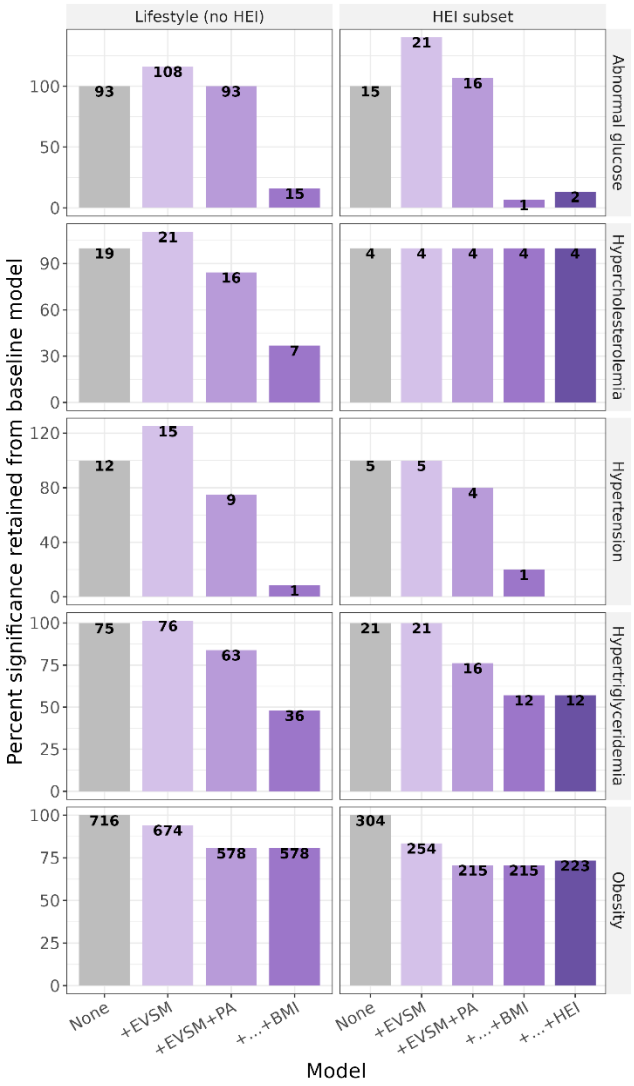
Supplementary Figure 1 Pairwise correlation heatmap for nine cancer treatments. A heatmap depicts the Phi coefficient for each of the pairwise comparisons among nine cancer treatments (N = 2,938 survivors). Blue indicates positive correlations, while red denotes negative correlations. Source data are provided as a Source Data file. **Abbreviations**: RT, radiotherapy.



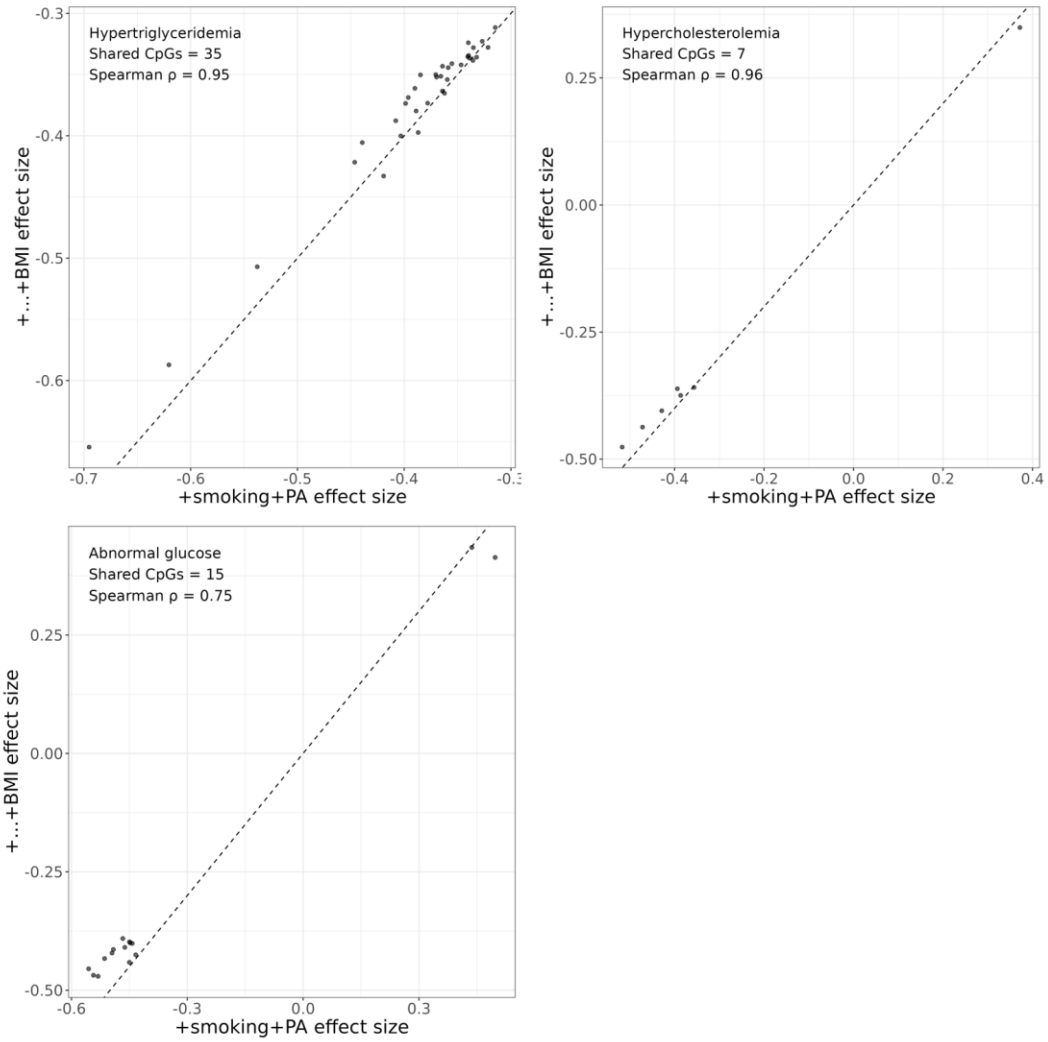
Supplementary Figure 2 Pairwise correlations among five prevalent case cardiometabolic risk factors. A heatmap displays the Phi coefficient for each pairwise comparisons of cardiometabolic risk factors (N = 2,938 survivors), with darker shades of blue indicating stronger positive correlations. Source data are provided as a Source Data file.



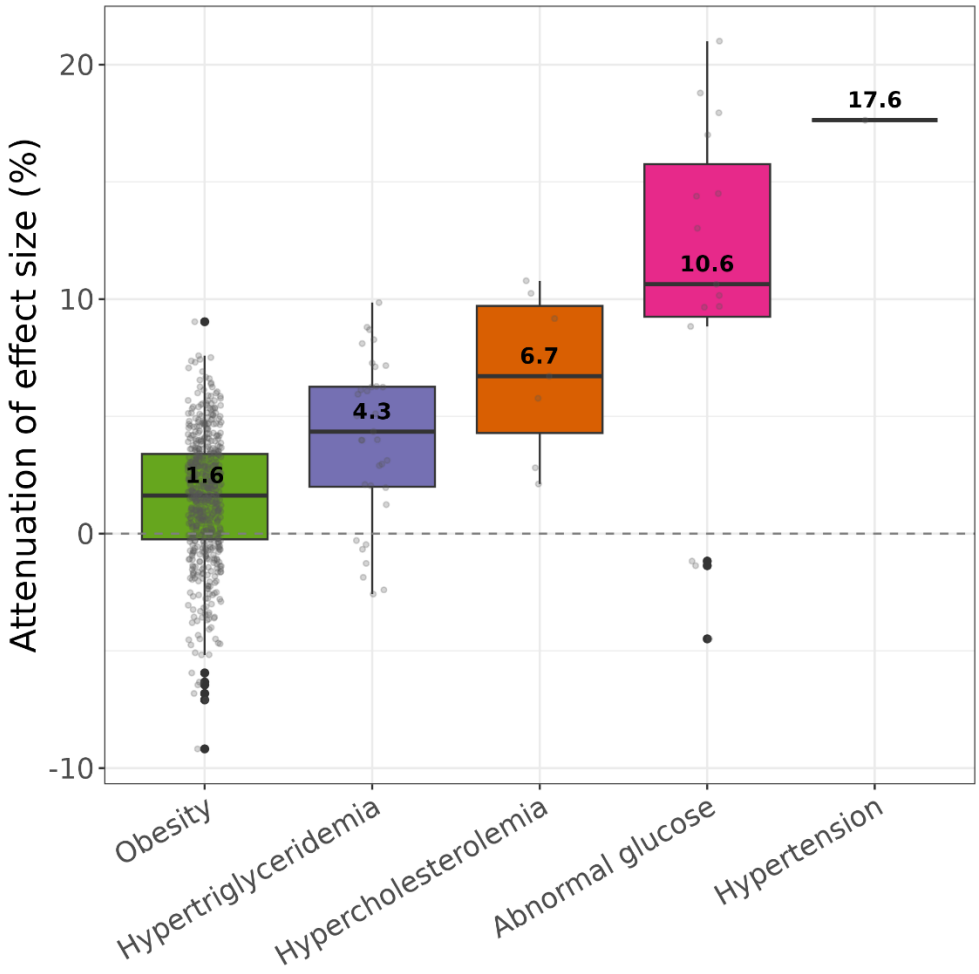
Supplementary Figure 3 Over- and under-representation of EWAS hits within CpG islands and gene functional regions. A heatmap illustrates the representation of CMRF-associated CpGs across 14 gene regions (y-axis) and five CMRFs (x-axis; N = 2,938 survivors). Odds ratios were calculated using two-sided Fisher's exact tests, with significant associations (Benjamini-Hochberg adjusted $p < 0.05$) marked by an asterisk (*). Blue indicates overrepresentation of CpGs in a given region, while red denotes underrepresentation. Source data are provided as a Source Data file. **Abbreviations:** Body, gene body; CMRF, cardiometabolic risk factor; Island, CpG island; N-Shore, north shore of CpG island; S-Shore, south shore of CpG island; TSS, transcription start site; UTR, untranslated region.



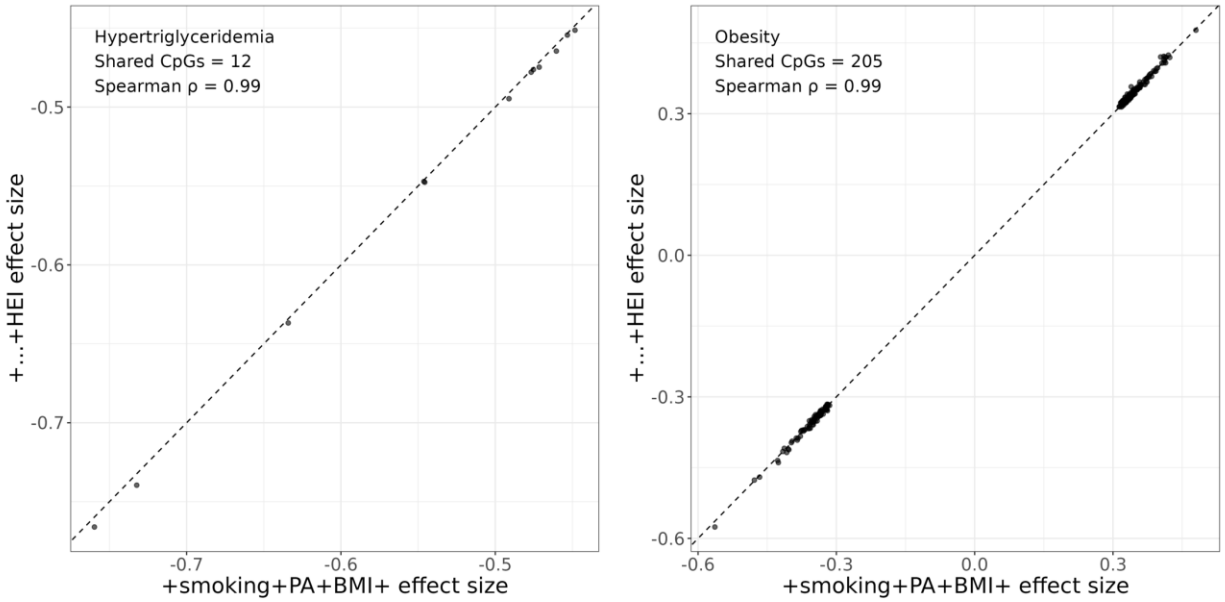
Supplementary Figure 4 Retention of EWAS-significant CpGs after lifestyle adjustment. Bars show the percentage of baseline (no-lifestyle) genome-wide significant CpG-CMRF signals (two-sided $p < 9 \times 10^{-8}$) retained at each adjustment step in two cohorts: the Lifestyle cohort (N = 2,037; smoking, physical activity, and BMI available) and the HEI subset (N = 1,382; all four lifestyle variables). Numbers on bars are the retained counts. Models were fit sequentially: None; +EVSM (ever smoked); +EVSM+PA (physical activity); +EVSM+PA+BMI; and, in the HEI subset only, +...+HEI (Healthy Eating Index-2010). BMI was not included as a covariate when the outcome was obesity. Source data are provided as a Source Data file. **Abbreviations**: EVSM, ever smoked; PA, physical activity; BMI, body mass index; HEI, Healthy Eating Index.



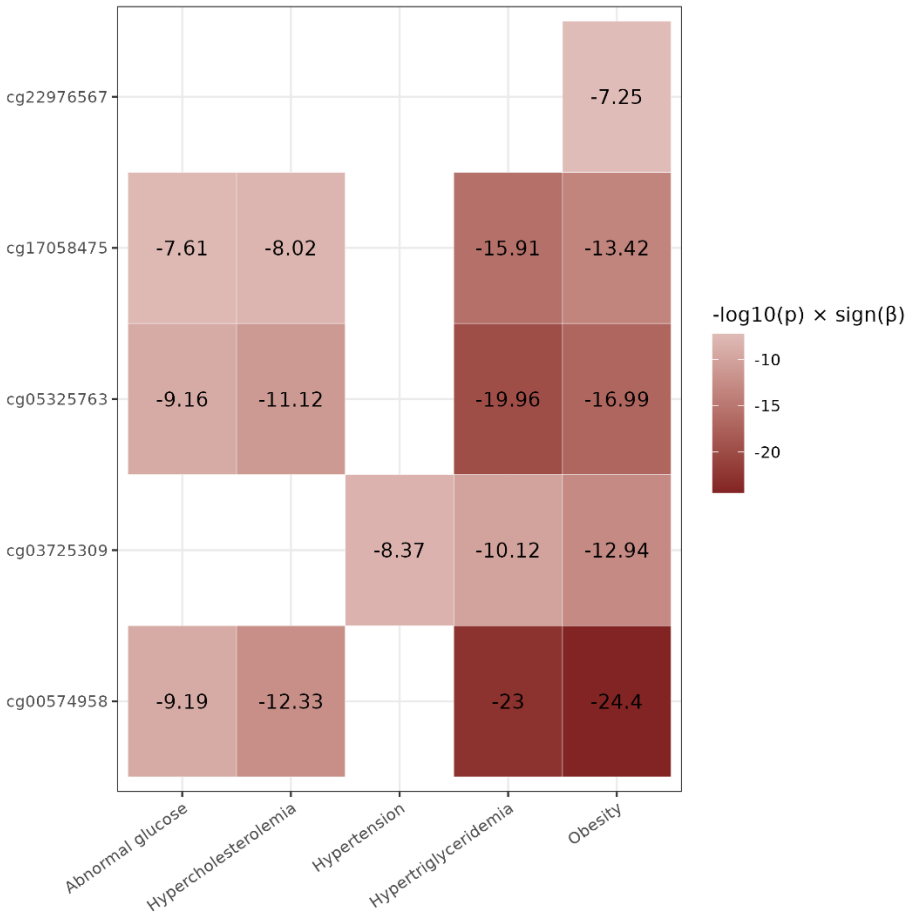
Supplementary Figure 5 Concordance of EWAS effect sizes after adjustment for BMI. Scatter plots show CpG-level EWAS effect estimates from models adjusting for smoking and physical activity (x-axis) versus models additionally adjusted for BMI (y-axis; N = 2,037 survivors). Each point represents a CpG that was genome-wide significant (two-sided $p < 9 \times 10^{-8}$) in both models for the indicated CMRF. The dashed line indicates $y = x$ (no change). Concordance was assessed using Spearman's rank correlation (two-sided), with ρ shown in each panel. Obesity is not shown because BMI was not included as a covariate when obesity was the outcome; hypertension is not shown due to only one shared CpG. Source data are provided as a Source Data file. **Abbreviations**: BMI, body mass index; PA, physical activity.



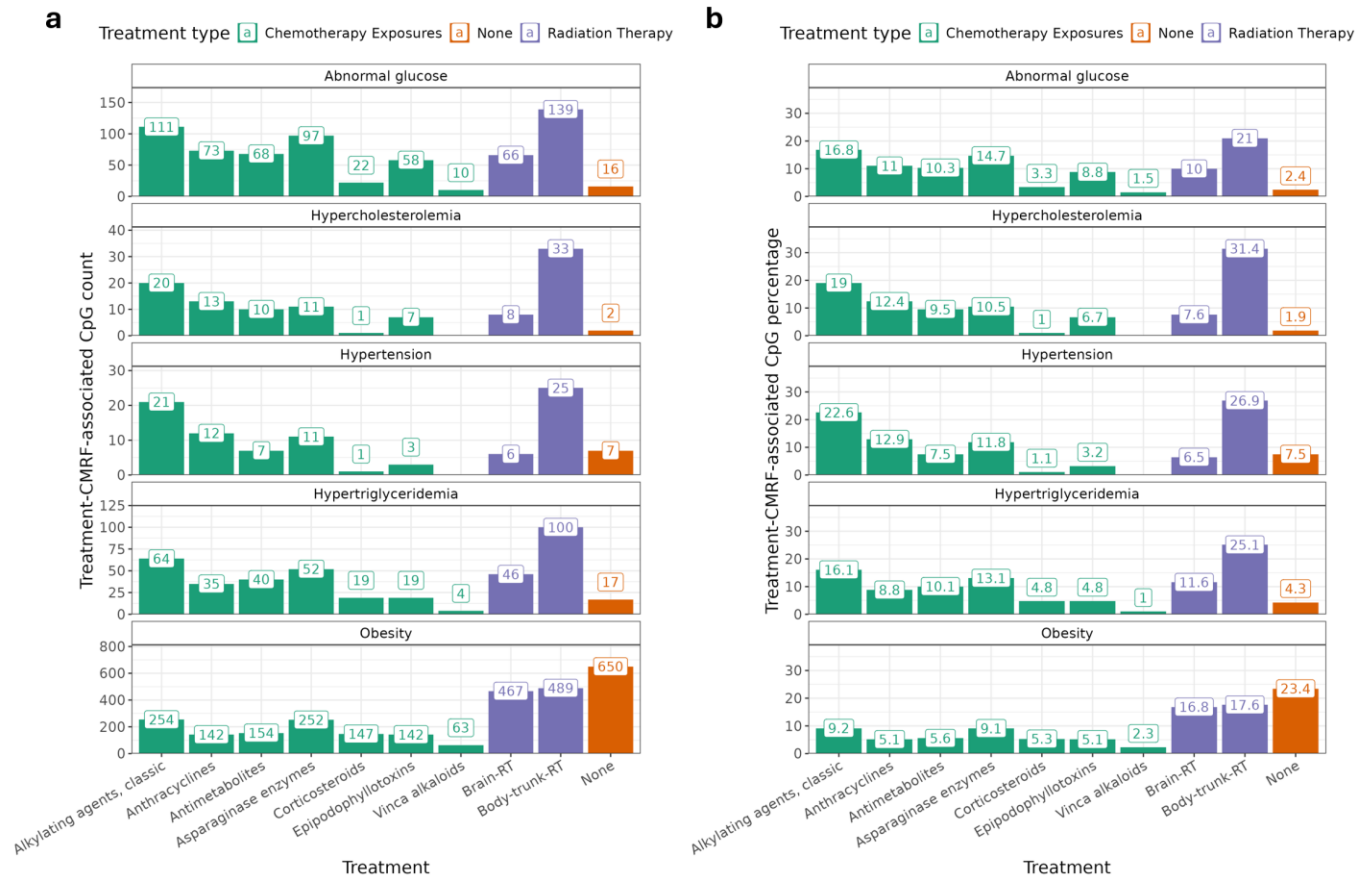
Supplementary Figure 6 Attenuation of EWAS effect sizes after lifestyle adjustment. Boxplots show the percent change in absolute EWAS effect size when moving from the base model (no lifestyle) to the fully lifestyle-adjusted model (ever smoked + physical activity + BMI; BMI omitted when obesity was the outcome; N = 2,037 survivors). Attenuation was computed as 100 x [1 - |lifestyle β| / |base β|]; positive values indicate shrinkage of the association after adjustment, negative values indicate strengthening. Each point represents an individual CpG (from the base model discovery set). Center lines denote the median, boxes represent the interquartile range (IQR), and whiskers extend to 1.5x IQR. Numbers on boxes indicate medians. The dashed line marks 0% (no change). Source data are provided as a Source Data file.



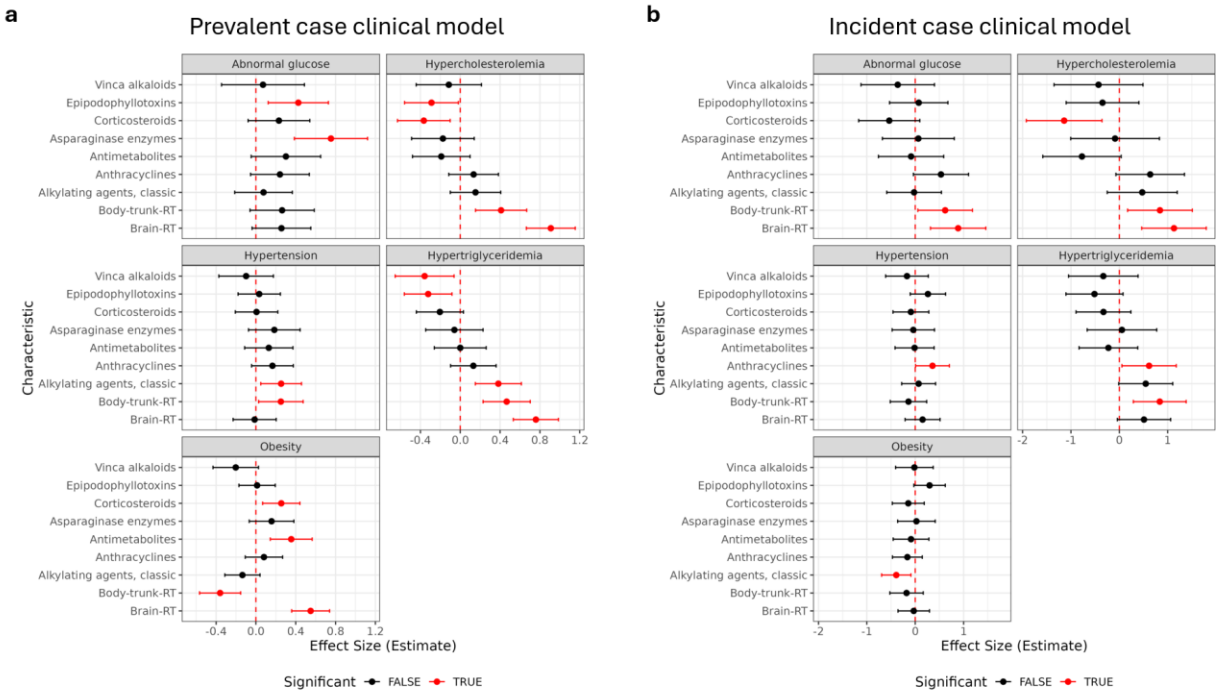
Supplementary Figure 7 Effect size concordance after HEI adjustment. Scatter plots comparing EWAS effect sizes for CpGs significant in both models within the HEI subset: smoking+PA+BMI (x-axis) versus smoking+PA+BMI+HEI (y-axis; N = 1,382). Each point represents a CpG. The dashed line *denotes y = x*. Concordance was quantified using Spearman's rank correlation (two-sided), with ρ shown in each panel, indicating minimal additional impact of HEI adjustment beyond smoking, PA, and BMI. Source data are provided as a Source Data file. **Abbreviations**: BMI, body mass index; HEI, Healthy Eating Index; PA, physical activity.



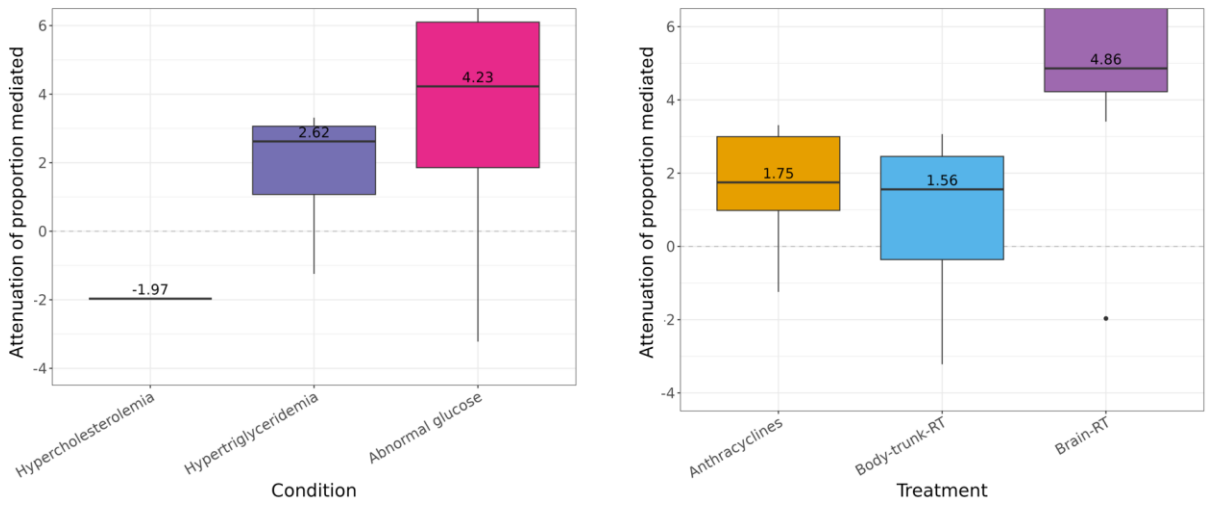
Supplementary Figure 8 Multi-phenotype ("hub") CpGs across five CMRFs in the lifestyle-adjusted EWAS. Heatmap displays the strength of signed association for five CpGs (rows) associated with multiple CMRFs (columns) in the final lifestyle-adjusted EWAS (N = 2,037 survivors). Tile show -log10(p) x sign(β) from logistic regression models adjusted for smoking and physical activity, with additional adjustment for BMI where applicable (BMI not included for obesity). Darker shading indicates stronger evidence of association. Blank cells indicate CpG-CMRF pairs that did not reach genome-wide significance (two-sided $p < 9 \times 10^{-8}$). Source data are provided as a Source Data file.



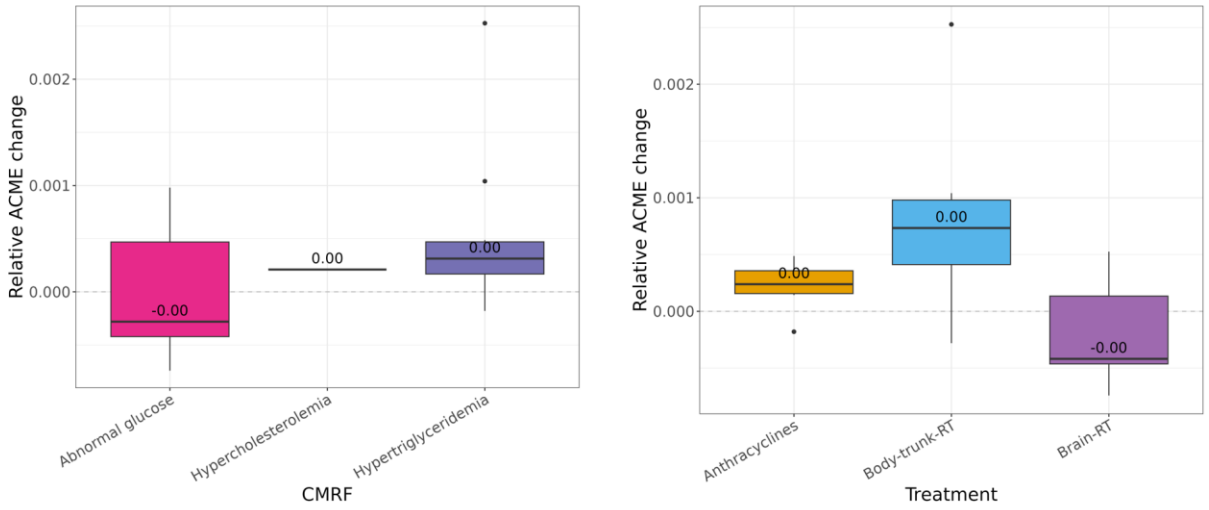
Supplementary Figure 9 Counts and percentages of CMRF- and treatment-associated CpGs. a) Bar chart showing the number of CpGs significantly associated with each CMRF (facets) and cancer treatments (x-axis). **b)** Corresponding percentages. CpGs were identified from EWAS using logistic regression with two-sided testing with genome-wide significance defined as $p < 9 \times 10^{-8}$ (N = 2,938 survivors). Colors denote chemotherapy (green), radiotherapy (purple), or no treatment association (orange). **Abbreviations:** CMRF, cardiometabolic risk factor; RT, radiotherapy.



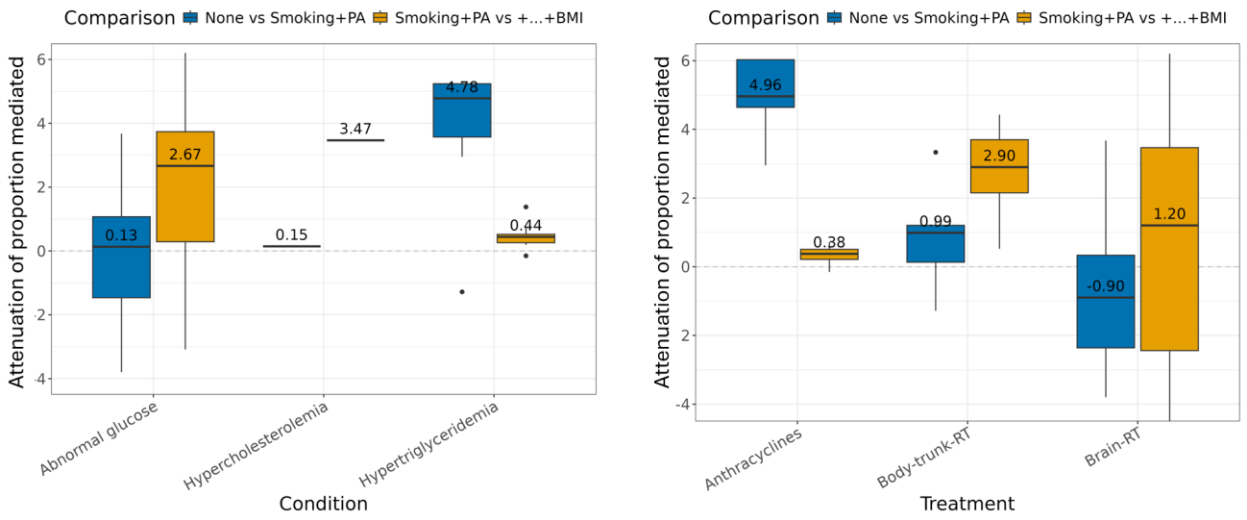
Supplementary Figure 10 CMRF-treatment associations. Forest plots display effect sizes for associations between each treatment exposure and CMRF from linear regression models for **a)** prevalent cases (N = 2,938 survivors) and **b)** incident cases (N = 2,802 survivors). Points represent estimated regression coefficients, with horizontal whiskers showing 95% confidence intervals (Estimate ± 1.96 x SE). Statistical significance was assessed using two-sided Wald tests; associations meeting the significance threshold ($p < 0.05$) are shown in red. Source data are provided as a Source Data file. **Abbreviations:** CMRF, cardiometabolic risk factor; RT, radiotherapy; SE, standard error.



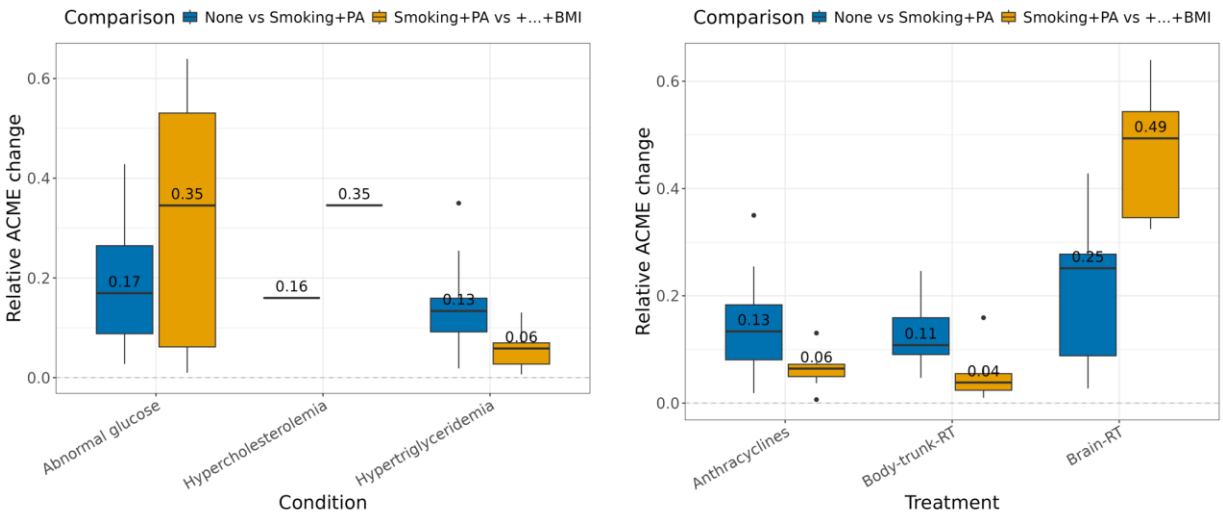
Supplementary Figure 11 Change in mediation proportion using incident outcomes. Boxplots show the change in the proportion of treatment effects on CMRFs mediated by DNAm when restricting analyses to incident outcomes that developed strictly after DNAm blood draw (N = 2,802 survivors). Results are presented separately for conditions and treatment exposures, calculated as 100 x (primary proportion mediated – incident proportion mediated). Positive values indicate greater mediation under the primary analysis compared with the incident-only analysis. Center lines denote the median, boxes represent the interquartile range (IQR) and whiskers extend 1.5x IQR. Source data are provided as a Source Data file. **Abbreviations:** CMRF, cardiometabolic risk factor; DNAm, DNA methylation; RT, radiotherapy.



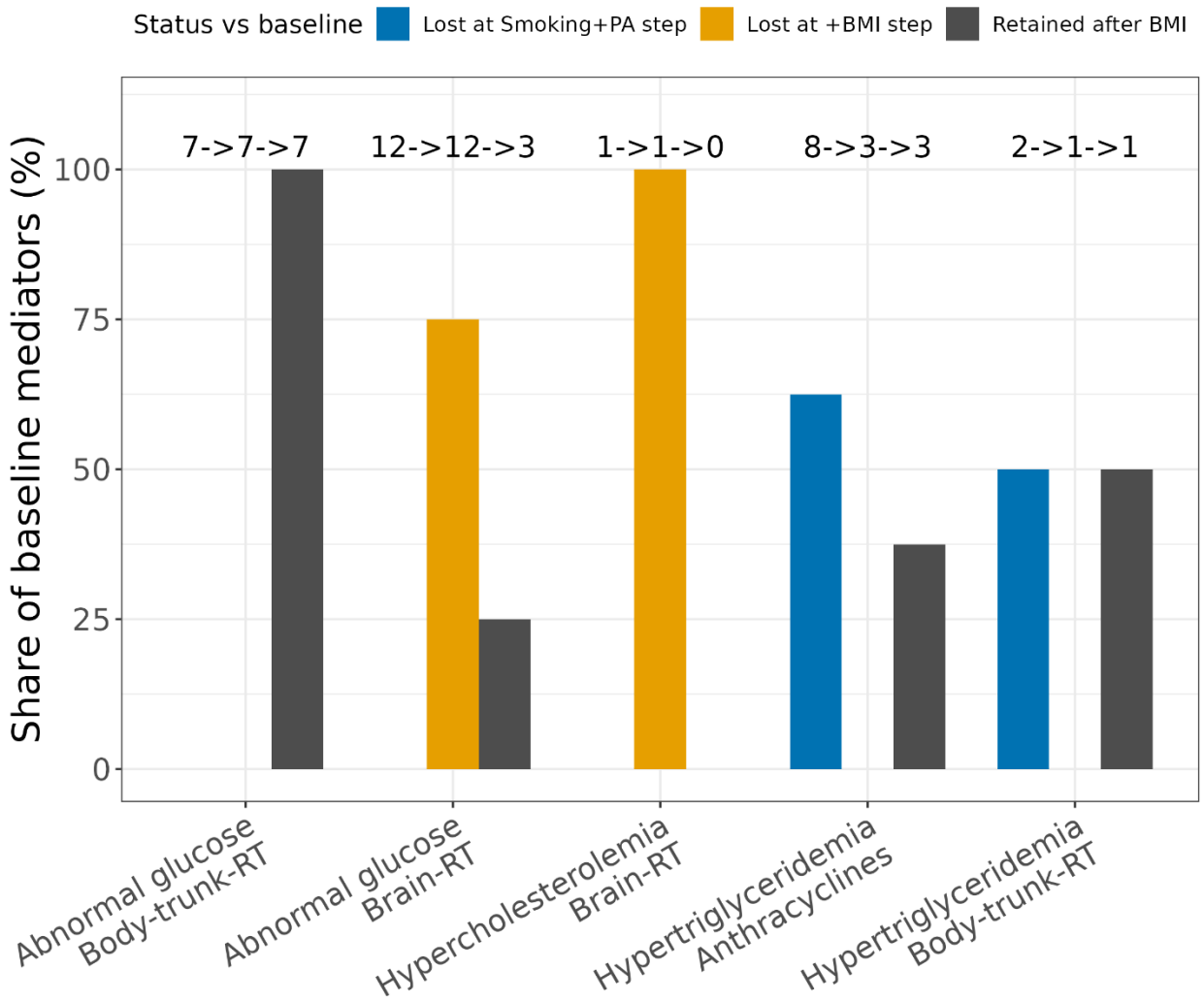
Supplementary Figure 12 Relative change in mediation effects using incident outcomes. Boxplots show the relative change in the ACME when restricting analyses to incident outcomes that developed strictly after the DNAm blood draw (N = 2,802 survivors), calculated as (primary ACME – incident ACME)/primary ACME. Results are shown separately for conditions and treatment exposures. Center lines denote the median, boxes represent the interquartile range (IQR) and whiskers extend 1.5x IQR. Numbers on boxes show median differences. Source data are provided as a Source Data file. **Abbreviations:** ACME, average causal mediation effect; CMRF, cardiometabolic risk factor, RT, radiotherapy.



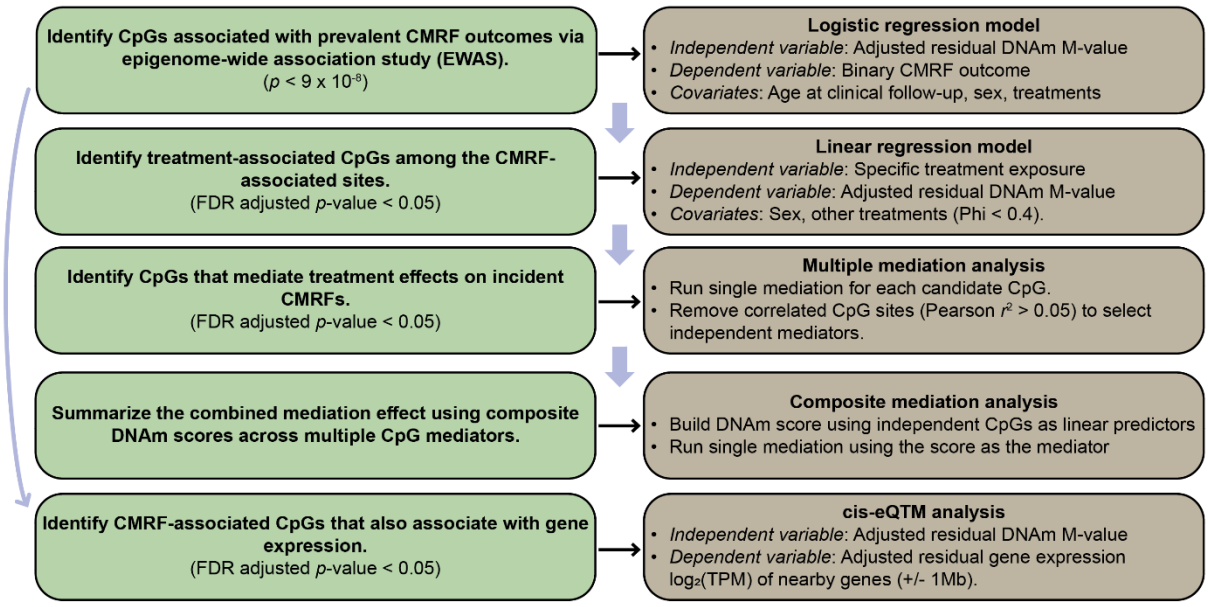
Supplementary Figure 13 Change in mediation proportion after lifestyle adjustment. Boxplots show changes in the proportion of treatment effects mediated by DNAm (PM; percentage points) across screened CpG mediators when adding lifestyle covariates to mediation models (N = 2,037 survivors). Two nested model comparisons are shown: None vs Smoking+PA (blue; difference in model with no lifestyle covariates minus model adjusting for smoking and physical activity) and Smoking+PA vs +...+BMI (gold; difference in model adjusting for smoking and physical activity minus model adjusting for smoking, physical activity, and BMI). Differences were calculated as the proportion mediated in the less-adjusted model minus the more-adjusted model; positive values indicate attenuation of mediation after adding covariates. Left, results grouped by CMRF; right, grouped by treatment. Horizontal dashed line marks zero; points denote outliers; numbers on boxes give median difference in PM. Center lines denote the median, boxes represent the interquartile range (IQR) and whiskers extend 1.5x IQR. Numbers on boxes show median differences. Source data are provided as a Source Data file. **Abbreviations:** BMI, body mass index; PA, physical activity; RT, radiotherapy.



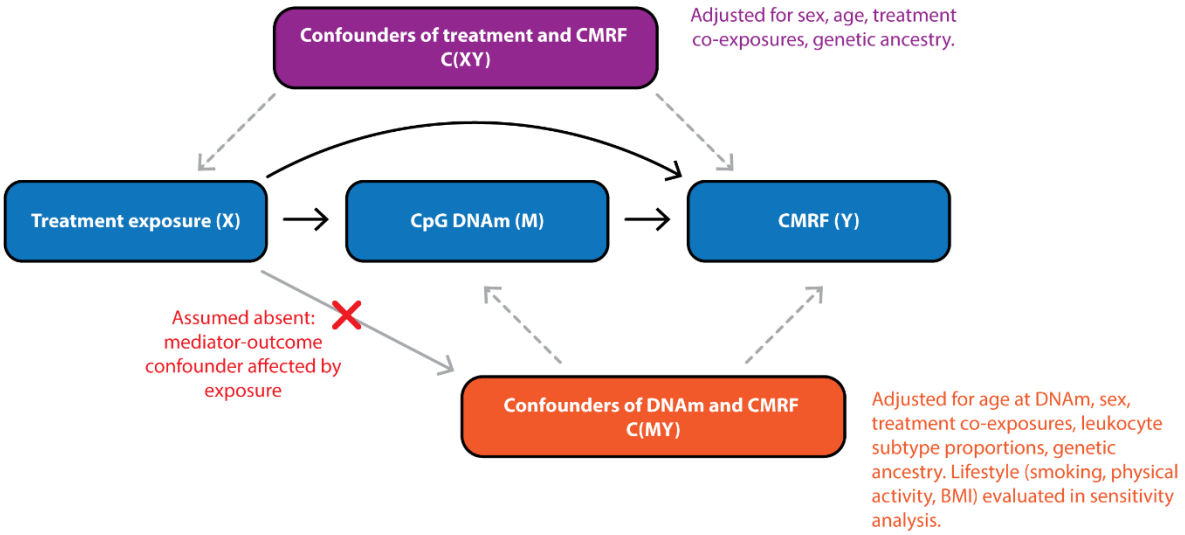
Supplementary Figure 14 Relative change in mediation effects across lifestyle-adjusted models. Boxplots summarize the relative change in the average causal mediation effect (ACME) when moving between nested mediation models (N = 2,037 survivors): None vs Smoking+PA (blue; relative difference in model with no lifestyle covariates minus model adjusting for smoking and physical activity) and Smoking+PA vs +...+BMI (gold; relative difference in model adjusting for smoking and physical activity minus model adjusting for smoking, physical activity, and BMI). Relative difference is computed as |ACME of model 1 – ACME of model 2|/|ACME model 1|. Left, results grouped by CMRF; right, grouped by treatment. Horizontal dashed line marks zero; points denote outliers; numbers on boxes indicate median relative change. Center lines denote the median, boxes represent the interquartile range (IQR) and whiskers extend 1.5x IQR. Numbers on boxes show median differences. Source data are provided as a Source Data file. **Abbreviations:** BMI, body mass index; PA, physical activity; RT, radiotherapy.



Supplementary Figure 15 Stepwise loss and retention of CpG mediators after lifestyle adjustment. For each treatment-CMRF pair, CpG mediators identified in the baseline no-lifestyle model (M1) were classified as lost after adding smoking and physical activity (M3, blue), lost only after subsequently adding BMI (M4, gold), or retained after BMI adjustment (dark grey; N = 2,037 survivors). Bar heights represent the proportion of baseline mediators (M1 = 100%). Labels above bars indicate the number of mediators at each step (M1->M3->M4). Source data are provided as a Source Data file. **Abbreviations:** BMI, body mass index; CMRF, cardiometabolic risk factor, PA, physical activity; RT, radiotherapy.



Supplementary Figure 16 Statistical analysis workflow. A flowchart outlining the sequential steps and models used in this study, including CMRF EWAS, candidate treatment associations analysis, mediation analyses, composite mediation analysis, and eQTM analysis. **Abbreviations:** CMRF, cardiometabolic risk factors; DNAm, DNA methylation; eQTM, expression quantitative trait methylation; FDR, false discovery rate; Mb, megabase; TPM, transcripts per million.



Supplementary Figure 17 Directed acyclic graph for treatment (X), CpG DNAm (M), and CMRF (Y) in the mediation analysis. Solid arrows denote causal relations (X to M to Y and X to Y). Dashed arrows indicate confounding paths adjusted in the models C(XY) (sex, age at most recent follow-up, treatment co-exposures, genetic ancestry) and C(MY) (age at DNAm, sex, treatment co-exposures, leukocyte subtype proportions, genetic ancestry); lifestyle factors were evaluated in sensitivity analyses. The path X to C(MY) is assumed absent (no mediator-outcome confounder expected by exposure).