

Assessing the causal role of epigenetic clocks in the development of multiple cancers: a Mendelian randomization study

Fernanda Morales-Berstein^{1,2}, Daniel L McCartney³, Ake T Lu⁴, Konstantinos K Tsilidis^{5,6}, Emmanouil Bouras⁶, Philip Haycock^{1,2}, Kimberley Burrows^{1,2}, Amanda I Phipps^{7,8}, Daniel D Buchanan⁹, Iona Cheng¹⁰, the PRACTICAL consortium*, Richard M Martin^{1,2,11}, George Davey Smith^{1,2}, Caroline L Relton^{1,2}, Steve Horvath^{4,12}, Riccardo E Marioni³, Tom G Richardson^{1,2,13}, Rebecca C Richmond^{1,2}

¹MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

²Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

³Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, Scotland, UK

⁴Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

⁵Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK

⁶Department of Hygiene and Epidemiology, School of Medicine, University of Ioannina, Ioannina, Greece

⁷Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;

⁸Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA

⁹Department of Clinical Pathology, Melbourne Medical School, University of Melbourne, Parkville, Victoria, Australia

¹⁰Cancer Prevention Institute of California, Fremont, CA, USA

¹¹National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol, Bristol, UK

¹²Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, CA, USA

¹³Novo Nordisk Research Centre, Headington, Oxford, OX3 7FZ, UK

Correspondence to: F Morales-Berstein – fernanda.moralesberstein@bristol.ac.uk

*Members from the PRACTICAL consortium are provided in the Supplementary material.

1 **ABSTRACT**

2

3 **Background:** Epigenetic clocks have been associated with cancer risk in several
4 observational studies. Nevertheless, it is unclear whether they play a causal role in cancer
5 risk or if they act as a non-causal biomarker.

6

7 **Methods:** We conducted a two-sample Mendelian randomization (MR) study to examine
8 the genetically predicted effects of epigenetic age acceleration as measured by HannumAge
9 (9 single-nucleotide polymorphisms (SNPs)), Horvath Intrinsic Age (24 SNPs), PhenoAge (11
10 SNPs) and GrimAge (4 SNPs) on multiple cancers (i.e., breast, prostate, colorectal, ovarian
11 and lung cancer). We obtained genome-wide association data for biological ageing from a
12 meta-analysis (N=34,710), and for cancer from the UK Biobank (N cases=2,671–13,879; N
13 controls=173,493–372,016), FinnGen (N cases=719–8,401; N controls=74,685–174,006) and
14 several international cancer genetic consortia (N cases=11,348–122,977; N
15 controls=15,861–105,974). Main analyses were performed using multiplicative random
16 effects inverse variance weighted (IVW) MR. Individual study estimates were pooled using
17 fixed effect meta-analysis. Sensitivity analyses included MR-Egger, weighted median,
18 weighted mode and Causal Analysis using Summary Effect Estimates (CAUSE) methods,
19 which are robust to some of the assumptions of the IVW approach.

20

21 **Results:** Meta-analysed IVW MR findings suggested that higher GrimAge acceleration
22 increased the risk of colorectal cancer (OR=1.12 per year increase in GrimAge acceleration,
23 95%CI 1.04–1.20, $p=0.002$). The direction of the genetically predicted effects was consistent
24 across main and sensitivity MR analyses. Among subtypes, the genetically predicted effect

25 of GrimAge acceleration was greater for colon cancer (IVW OR=1.15, 95%CI 1.09–1.21,
26 $p=0.006$), than rectal cancer (IVW OR=1.05, 95%CI 0.97–1.13, $p=0.24$). We also found
27 evidence that higher GrimAge acceleration decreased the risk of prostate cancer (pooled
28 IVW OR=0.93 per year increase in GrimAge acceleration, 95%CI 0.87–0.99, $p=0.02$). This was
29 supported by MR sensitivity analyses, but did not replicate in MR analyses using data on
30 parental history of prostate cancer in UK Biobank (IVW OR=1.00, 95%CI 0.96–1.04, $p=1.00$).

31 Results were less consistent for associations between other epigenetic clocks and cancers.

32

33 **Conclusions:** GrimAge acceleration may increase the risk of colorectal cancer. Additionally,
34 there is more limited evidence that it may be protective against prostate cancer. Findings
35 for other clocks and cancers were inconsistent. Further work is required to investigate the
36 potential mechanisms underlying the results.

37

38 **Funding:** FMB was supported by a Wellcome Trust PhD studentship in Molecular, Genetic
39 and Lifecourse Epidemiology (218495/Z/19/Z). KKT was supported by a Cancer Research UK
40 (C18281/A29019) programme grant (the Integrative Cancer Epidemiology Programme) and
41 by the Hellenic Republic’s Operational Programme “Competitiveness, Entrepreneurship &
42 Innovation” (OPΣ 5047228). PH was supported by Cancer Research UK (C18281/A29019).

43 RMM was supported by the NIHR Biomedical Research Centre at University Hospitals Bristol
44 and Weston NHS Foundation Trust and the University of Bristol and by a Cancer Research
45 UK (C18281/A29019) programme grant (the Integrative Cancer Epidemiology Programme).

46 The views expressed are those of the author(s) and not necessarily those of the NIHR or the
47 Department of Health and Social Care. GDS and CLR were supported by the Medical

48 Research Council (MC_UU_00011/1 and MC_UU_00011/5) and by a Cancer Research UK

49 (C18281/A29019) programme grant (the Integrative Cancer Epidemiology Programme). REM
50 was supported by an Alzheimer's Society project grant (AS-PG-19b-010) and NIH grant (U01
51 AG-18-018, PI: Steve Horvath). RCR is a de Pass Vice Chancellor's Research Fellow at the
52 University of Bristol.

53 INTRODUCTION

54

55 DNA methylation (DNAm) is an epigenetic biomarker that can be used as an estimator of
56 chronological age. Biological age, as predicted by DNAm patterns at specific cytosine-
57 phosphate-guanine (CpG) sites, may differ from chronological age on an individual basis.

58 Observational evidence suggests that epigenetic age acceleration (i.e., when an individual's
59 biological age is greater than their chronological age) may be associated with an increased
60 risk of mortality and age-related diseases, including cancer¹.

61

62 Epigenetic clocks are heritable indicators of biological ageing derived from DNAm data. Each

63 clock is based on DNAm levels measured at a different set of CpG sites, which capture

64 distinctive features of epigenetic ageing². 'First-generation' epigenetic clocks, such as

65 HannumAge³ and Intrinsic HorvathAge⁴, have been derived from DNAm levels at CpG sites

66 found to be strongly associated with chronological age. HannumAge is trained on 71 age-

67 related CpGs found in blood³, while Intrinsic HorvathAge is based on 353 age-related CpGs

68 found in several human tissues and cell types, and is further adjusted for blood cell counts⁴.

69 More recently, 'second-generation' epigenetic clocks, such as, PhenoAge⁵ and GrimAge⁶,

70 have been developed to predict age-related morbidity and mortality. PhenoAge

71 incorporates data from 513 age-related CpGs associated with mortality and nine clinical

72 biomarkers (i.e., albumin, creatinine, serum glucose, C-reactive protein, lymphocyte

73 percentage, mean corpuscular volume, red cell distribution width, alkaline phosphatase and

74 leukocyte count)⁵, and GrimAge includes data from 1,030 age-related CpGs associated with

75 smoking pack-years and seven plasma proteins (i.e., cystatin C, leptin, tissue inhibitor

76 metalloproteinases 1, adrenomedullin, beta-2-microglobulin, growth differentiation factor

77 15, and plasminogen activation inhibitor 1 (PAI-1))⁶. Due to differences in their composition,
78 HannumAge and Intrinsic HorvathAge are better predictors of chronological age^{3 4}, while
79 PhenoAge and GrimAge stand out for their ability to predict health and lifespan⁵⁻⁷.

80

81 Several studies suggest that HannumAge, Intrinsic HorvathAge, PhenoAge and GrimAge
82 acceleration are associated with cancer risk^{5 8-13}. In contrast, others indicate that evidence in
83 support of this claim is weak or non-existent¹⁴⁻¹⁷. This lack of consensus could be explained
84 by biases that often affect observational research, such as reverse causation (e.g., cancer
85 influencing the epigenome and not the other way around) and residual confounding (e.g.,
86 unmeasured, or imprecisely measured confounders of the association between epigenetic
87 age acceleration and cancer)¹⁸.

88

89 The strength of the associations between epigenetic age acceleration and different cancers
90 has also been found to vary across epigenetic clocks. For instance, positive associations
91 between epigenetic age acceleration and colorectal cancer seem to be much stronger when
92 biological age is estimated using second-generation clocks (i.e., PhenoAge and GrimAge)¹⁰
93 rather than first-generation clocks (i.e., HannumAge and Intrinsic HorvathAge)^{14 16}. Lack of
94 consensus across epigenetic clocks could be explained by differences in their algorithms
95 (which may reflect different mechanisms of biological ageing), as well as heterogeneity in
96 study designs¹. Furthermore, even if there were a consensus, it would still be unclear
97 whether age-related DNA methylation plays a causal role in cancer risk or if it merely acts as
98 a non-causal prognostic biomarker.

99

100 Mendelian randomization (MR), a method that uses genetic variants as instrumental
101 variables to infer causality between a modifiable exposure and an outcome, is less likely to
102 be affected by residual confounding and reverse causation than traditional observational
103 methods¹⁹. A recent genome-wide association study (GWAS) meta-analysis has revealed 137
104 genetic loci associated with epigenetic age acceleration (as measured by six epigenetic
105 biomarkers) that may be used within an MR framework²⁰.

106

107 McCartney et al.²⁰ used IVW MR, MR-Egger, weighted median and weighted mode methods
108 to explore the genetically predicted effects of HannumAge, Intrinsic HorvathAge, PhenoAge
109 and GrimAge acceleration on breast, ovarian and lung cancer. Here we extend this analysis
110 to include colorectal and prostate cancer (two of the most common cancers worldwide²¹)
111 and use additional methods and datasets to verify the robustness of our findings.

112

113 The aim of this two-sample MR study was to examine the genetically predicted effects of
114 epigenetic age acceleration (as measured by HannumAge³, Horvath Intrinsic Age⁴, PhenoAge⁵
115 and GrimAge⁶) on multiple cancers (i.e., breast, prostate, colorectal, ovarian and lung cancer)
116 using summary genetic association data from (1) McCartney et al. (N=34,710)²⁰, (2) the UK
117 Biobank (N cases=2,671–13,879; N controls=173,493–372,016), (3) FinnGen (N cases=719–
118 8,401; N controls=74,685–174,006) and (4) several international cancer genetic consortia (N
119 cases=11,348–122,977; N controls=15,861–105,974).

120

121 **METHODS**

122

123 **Reporting guidelines**

124

125 This study has been reported according to the STROBE-MR guidelines²² (**Supplementary**
126 **material**).

127

128 **Genetic instruments for epigenetic age acceleration**

129

130 We obtained summary genetic association estimates for epigenetic age acceleration
131 measures of HannumAge³, Intrinsic HorvathAge⁴, PhenoAge⁵ and GrimAge⁶ from a recent
132 GWAS meta-analysis of biological ageing²⁰, which included 34,710 participants of European
133 ancestry. Across the 28 European ancestry studies considered in the analysis, 57.3% of
134 participants were female. A detailed description of the methods that were used can be
135 found in the publication²⁰. In short, outlier samples with clock methylation estimates of +/-
136 5 s.d. from the mean were excluded from further analysis. SNPs were genotyped and
137 imputed independently for each cohort included in the meta-analysis and genotypes were
138 imputed using either the HRC or the 1000 Genomes Project Phase 3 reference panels.

139

140 We selected GWAS-significant SNPs ($p < 5 \times 10^{-8}$) for each epigenetic age acceleration
141 measure and performed linkage disequilibrium (LD) clumping ($r^2 < 0.001$) to remove
142 correlated variants using the European reference panel from the 1000 Genomes Project
143 Phase 3 v5.

144

145 We identified 9 independent SNPs for HannumAge, 24 for Intrinsic HorvathAge, 11 for
146 PhenoAge and 4 for GrimAge (**Supplementary Table 1**). The genetic instruments for
147 HannumAge, Intrinsic HorvathAge, PhenoAge and GrimAge acceleration explained 1.48%,
148 4.41%, 1.86% and 0.47% of the trait variance, respectively. All the selected SNPs had F-
149 statistics greater than 10 (HannumAge median 38 and range 31–99, Intrinsic HorvathAge
150 median 47 and range 31–240, PhenoAge median 45 and range 32–89, GrimAge median 36
151 and range 31–45).

152

153 **Genetic association data sources for cancer outcomes**

154

155 We obtained summary-level genetic association data for cancer outcomes from the UK
156 Biobank and several international cancer genetic consortia (**Table 1**). Data were obtained
157 from UK Biobank, FinnGen, the Breast Cancer Association Consortium (BCAC), the Ovarian
158 Cancer Association Consortium (OCAC), the Consortium of Investigators of Modifiers of
159 BRCA1/2 (CIMBA), the Prostate Cancer Association Group to Investigate Cancer Associated
160 Alterations in the Genome (PRACTICAL), the International Lung Cancer Consortium (ILCCO)
161 and the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO). Further
162 details of the studies and the data obtained are described in the **Supplementary Methods**.

163

164 We extracted genetic association data for the selected SNPs from each cancer GWAS (for
165 breast, prostate, colorectal, ovarian and lung cancers). LD proxies ($r^2 > 0.8$) were used when
166 the SNPs of interest were missing from the cancer GWAS dataset. The proxies were located
167 using the MR-Base platform, which calculates LD using the European subset of individuals
168 from the 1000 Genomes Project reference panel as above²³. The “LDlinkR” R package

169 version 1.1.2 was used to find proxies for cancer data that were not included in the MR-Base
170 platform. The exposure and outcome datasets were then harmonised to ensure the genetic
171 associations reflect the same effect allele. Palindromic SNPs with minor allele frequencies
172 (MAF) <0.3 were aligned, while those with MAF \geq 0.3 or mismatching strands were excluded.

173

174 **Power calculations**

175

176 Statistical power was calculated using an online calculator for MR available at:

177 <https://shiny.cnsgenomics.com/mRnd/>. Calculations were performed separately for each

178 clock-cancer combination. They were based on a type 1 error rate of 0.05, the proportion of

179 phenotypic variance explained by genetic variants (R^2) for each measure of epigenetic age

180 acceleration, and the total number of cases and controls included in the meta-analysis for

181 each cancer. Across combinations of the four epigenetic clock acceleration and five cancer

182 measures, we had 80% power to detect ORs as small as 1.04–1.39 (**Supplementary Table 2**).

183

184 **Statistical analysis**

185

186 We estimated the genetically predicted effects of epigenetic age acceleration (as measured

187 by HannumAge³, Horvath Intrinsic Age⁴, PhenoAge⁵ and GrimAge⁶) on multiple cancers (i.e.,

188 breast, prostate, colorectal, ovarian and lung cancer) using a two-sample MR framework

189 (**Figure 1**).

190

191 *Main analyses*

192

193 Main analyses were performed using multiplicative random effects inverse variance
194 weighted (IVW) MR, a method that combines the genetically predicted effect of epigenetic
195 age acceleration on cancer across genetic variants²⁴. We used fixed effect meta-analysis to
196 pool results across studies (i.e., UK Biobank, FinnGen and international consortia). For
197 colorectal cancer, we only pooled FinnGen and GECCO estimates, since UK Biobank
198 participants were already included in GECCO. I^2 statistics and their corresponding
199 confidence intervals were used to estimate heterogeneity across study estimates²⁵. A
200 Benjamini-Hochberg false discovery rate (FDR) <5% was used to correct the pooled main
201 IVW results for multiple testing²⁶.

202

203 *Sensitivity analyses*

204

205 MR assumes genetic instruments for epigenetic age acceleration are 1) associated with
206 epigenetic age acceleration (relevance assumption), 2) independent of confounders
207 (independence assumption) and 3) only associated with cancer through their effect on
208 epigenetic age acceleration (exclusion restriction assumption)^{27 28}.

209

210 As a sensitivity analysis and to test for potential violations of the relevance assumption, we
211 calculated F-statistics and the R^2 for each measure of epigenetic age acceleration²⁹. Other
212 sensitivity analyses included MR-Egger³⁰, weighted median³¹ and weighted mode³²
213 methods, which are robust to some of the assumptions of the IVW approach (described in
214 the ***Supplementary Methods***). These results were also pooled across studies, as explained
215 above. Consistency across different MR methods would suggest that it is less likely that the
216 independence and exclusion restriction assumptions are violated.

217

218 Where associations between genetically predicted epigenetic age acceleration and cancer
219 were identified, we additionally performed single-SNP two-sample MR analysis to assess
220 whether the effects were likely to be driven by a single SNP. We also used Causal Analysis
221 using Summary Effect Estimates (CAUSE)³³, a method that uses genome-wide summary
222 statistics to disentangle causality (i.e., SNPs are associated with cancer through their effect
223 on epigenetic age acceleration) from correlated horizontal pleiotropy (i.e., SNPs are
224 associated with epigenetic age acceleration and cancer through a shared heritable factor),
225 while taking into account uncorrelated horizontal pleiotropy (i.e., SNPs are associated with
226 epigenetic age acceleration through separate mechanisms). It uses Bayesian modelling to
227 assess whether the sharing model (i.e., model that fixes the causal effect at zero) fits the
228 data at least as well as the causal model (i.e., model that allows a causal effect different
229 from zero). Additionally, Cochran's Q statistics were used to quantify global heterogeneity
230 across SNP-specific MR estimates³⁴ and MR-Egger intercept tests were performed to detect
231 uncorrelated horizontal pleiotropy³⁰.

232

233 *Secondary analyses*

234

235 As a secondary analysis, we conducted two-sample MR of epigenetic age acceleration and
236 cancer subtypes (i.e., breast cancer: ER+, ER-, triple negative, luminal B/HER2-negative-like,
237 HER2-enriched-like, luminal A-like, luminal B-like, BRCA1 and BRCA2; ovarian cancer: high
238 grade serous, low grade serous, invasive mucinous, clear cell, endometrioid, BRCA1 and
239 BRCA2; prostate cancer: advanced, advanced [vs non-advanced], early onset, high risk [vs
240 low risk], and high risk [vs low and intermediate risk]; lung cancer: adenocarcinoma and

241 squamous cell; colorectal cancer: colon-specific, proximal colon-specific, distal colon-
242 specific, rectal-specific, male and female) (***Supplementary Methods***).

243

244 We also performed two-sample MR analyses of epigenetic age acceleration and parental
245 history of cancer in the UK Biobank for breast, prostate, lung and bowel cancer
246 (***Supplementary Methods***). Data on parental history of ovarian cancer were not available in
247 UK Biobank. Family history data correlate with combined hospital record and questionnaire
248 data and it has been suggested that they provide better power to detect GWAS-significant
249 associations for some phenotypes in the UK Biobank³⁵. Therefore, we expected these results
250 to be consistent with those obtained in the main analyses. Although these analyses were
251 conducted as a form of replication, their effect estimates are not directly comparable to
252 those obtained in the main analyses due to cases in the GWAS-by-proxy of parental
253 endpoints being defined as either or both parents reportedly having a type of cancer.

254

255 MR results were reported as the odds ratio (OR) of site-specific cancer per one year increase
256 in genetically predicted epigenetic age acceleration.

257

258 LD Score regression³⁶ was used to identify genome-wide genetic correlations between
259 epigenetic age acceleration and cancer. Genetic correlations were estimated using full
260 GWAS summary statistics for the epigenetic clocks and cancer, as well as the 1000 Genomes
261 Project European LD reference panel. Traits with mean heritability chi-square values <1.02
262 were excluded from the analyses.

263

264 All MR analyses were performed using R software version 4.0.2. Two sample MR analyses
265 were conducted using the “TwoSampleMR” package version 0.5.5. Meta-analyses of IVW
266 results were performed using the “meta” package version 4.18. CAUSE analyses were
267 conducted using the “cause” package version 1.2.0. Forest plots were created using the
268 “ggforestplot” package version 0.1.0. LD Scores were computed using the “ldsc” command
269 line tool version 1.0.1. The code used in this study is available at:
270 https://github.com/fernandam93/epiclocks_cancer.

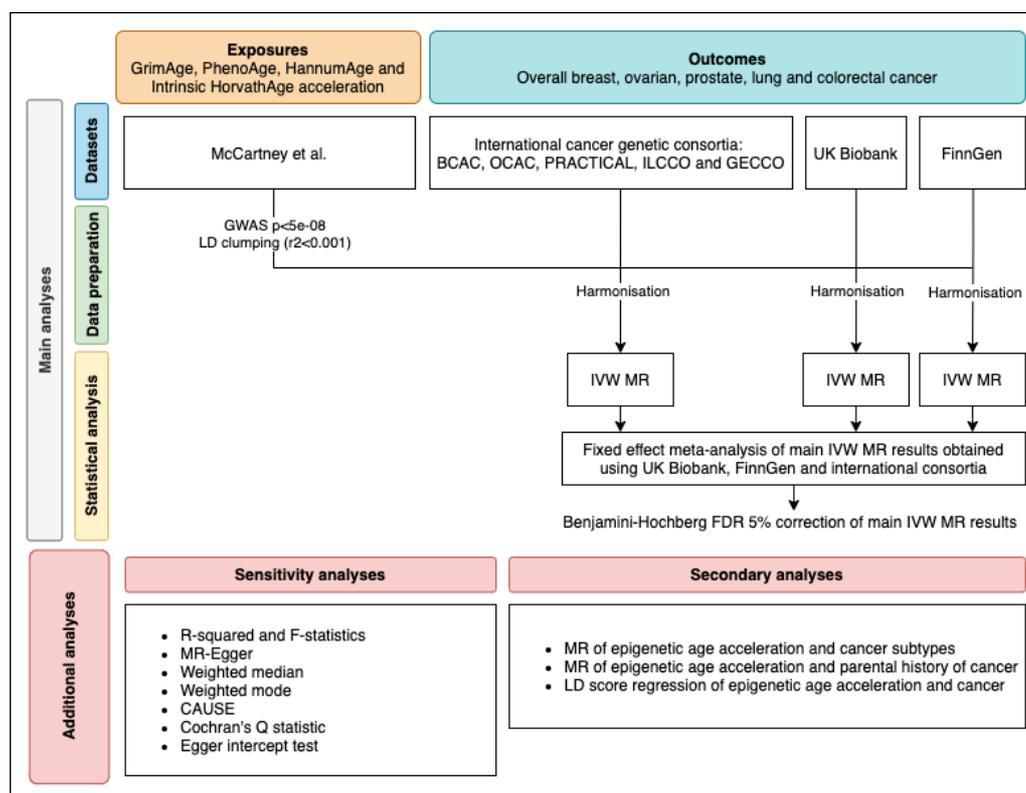
271 **Table 1.** Numbers of overall cancer cases and controls by data source.

Cancer type	Source	N cases (%) *	N controls
Breast	BCAC	122,977 (53.7%)	105,974
	UK Biobank	13,879 (6.5%)	198,523
	FinnGen	8,401 (7.8%)	99,321
Ovarian	OCAC	25,509 (38.4%)	40,941
	UK Biobank	1,218 (0.6%)	198,523
	FinnGen	719 (0.7%)	99,321
Prostate	PRACTICAL	79,148 (56.4%)	61,106
	UK Biobank	9,132 (5.0%)	173,493
	FinnGen	6,311 (7.8%)	74,685
Lung	ILCCO	11,348 (41.7%)	15,861
	UK Biobank	2,671(0.7%)	372,016
	FinnGen	1,681 (1.0%)	173,933
Colorectal	GECCO	58,131 (46.3%)	67,347
	UK Biobank	5,657 (1.5%)	372,016
	FinnGen	3,022 (1.7%)	174,006

272 Abbreviations: BCAC, Breast Cancer Association Consortium; OCAC, Ovarian Cancer Association Consortium; PRACTICAL, Prostate
 273 Cancer Association Group to Investigate Cancer Associated Alterations in the Genome; ILCCO, International Lung Cancer Consortium;
 274 GECCO, Genetics and Epidemiology of Colorectal Cancer Consortium.

275 *Percentage (%) of cases within each study source was calculated using the following formula: $100 * N \text{ cases} / (N \text{ cases} + N \text{ controls})$

276



277 **Figure 1.** Flowchart summarising study methods.
 278

279 Abbreviations: BCAC, Breast Cancer Association Consortium; OCAC, Ovarian Cancer Association Consortium; PRACTICAL, Prostate
 280 Cancer Association Group to Investigate Cancer Associated Alterations in the Genome; ILCCO, International Lung Cancer Consortium; GECCO,
 281 Genetics and Epidemiology of Colorectal Cancer Consortium; LD, linkage disequilibrium; IVW, inverse variance weighted; MR, Mendelian
 282 randomization; FDR, false discovery rate; GWAS, genome-wide association study; CAUSE, Causal Analysis Using Summary Effect estimates.

283 **RESULTS**

284

285 *Breast cancer*

286

287 We did not find strong evidence of causality between epigenetic age acceleration and breast
288 cancer (GrimAge IVW OR=0.98, 95%CI 0.95–1.00, $p=0.08$; PhenoAge IVW OR=0.99, 95%CI
289 0.98–1.01, $p=0.23$; HannumAge IVW OR=0.99, 95%CI 0.97–1.02, $p=0.63$; and Intrinsic
290 HorvathAge IVW OR=0.99, 95%CI 0.98–1.00, $p=0.13$) (**Figure 2, Supplementary Figures 1**
291 **and 5–7, Supplementary Tables 3–6 and 9–10**).

292

293 *Ovarian cancer*

294

295 There was also limited evidence of causality between epigenetic age acceleration and
296 ovarian cancer (GrimAge IVW OR=0.99, 95%CI 0.93–1.06, $p=0.78$; PhenoAge IVW OR=0.98,
297 95%CI 0.96–1.01, $p=0.24$; HannumAge IVW OR=1.00, 95%CI 0.96–1.04, $p=0.95$; and Intrinsic
298 HorvathAge IVW OR=1.00, 95%CI 0.97–1.02, $p=0.89$) (**Figure 2, Supplementary Figures 1**
299 **and 5–7, Supplementary Tables 3–6 and 9–10**).

300

301 *Prostate cancer*

302

303 Meta-analysed IVW MR findings suggested that genetically predicted GrimAge acceleration
304 decreased the risk of prostate cancer (OR=0.93 per year increase in GrimAge acceleration,
305 95%CI 0.87–0.99, $p=0.02$) (**Figure 2, Supplementary Tables 3–6**). These results withstood
306 multiple testing correction (FDR $p=0.04$), there was little evidence of heterogeneity across

307 study estimates ($I^2=0\%$, 95%CI 0–41%, $p=0.84$) and the direction of the genetically predicted
308 effect was consistent across main and sensitivity MR analyses (i.e., MR-Egger, weighted
309 median and weighted mode) (**Figure 3, Supplementary Tables 3–6**).

310

311 We further explored the genetically predicted effect of GrimAge on prostate cancer using
312 PRACTICAL data only. Single-SNP analysis revealed that the effect was not driven by a single
313 SNP (**Supplementary Table 7**), although there was some evidence of heterogeneity across
314 SNPs (Cochran's Q 10.63, $p=0.01$) (**Supplementary Table 8**). Additionally, results suggested
315 no evidence of detectable uncorrelated horizontal pleiotropy (MR-Egger intercept = -0.03,
316 95%CI -0.34–0.28, $p=0.87$) (**Supplementary Table 8**). However, we were unable to rule out
317 correlated pleiotropy (CAUSE OR=0.98, 95%CI 0.94–1.02, $p=0.64$; shared $q=3\%$, 95%CI=0–
318 25%) (**Supplementary Figure 2**).

319

320 Among subtypes, the protective effect of GrimAge acceleration appeared strongest in
321 relation to advanced (IVW OR=0.90, 95%CI 0.82–1.00, $p=0.04$) and early onset prostate
322 cancer (IVW OR=0.76, 95%CI 0.62–0.92, $p=0.005$) (**Figure 4, Supplementary Table 9**).

323

324 However, the prostate cancer results must be interpreted with caution, as findings did not
325 replicate when using parental history of prostate cancer as the outcome (IVW OR=1.00 per
326 year increase in GrimAge acceleration, 95%CI 0.96–1.04, $p=1.00$) (**Figure 5, Supplementary**
327 **Table 10**). Furthermore, LD Score regression results did not provide strong, consistent
328 evidence of a genetic correlation between GrimAge and prostate cancer (PRACTICAL $r_g=-$
329 0.13, $p=0.12$; UK Biobank $r_g=0.03$, $p=0.76$; FinnGen $r_g=-0.02$, $p=0.89$) (**Supplementary Figure**
330 **3, Supplementary Table 11**)

331

332 We did not find consistent evidence of causality between other measures of epigenetic age
333 acceleration and prostate cancer (PhenoAge IVW OR=1.01, 95%CI 0.99–1.03, $p=0.31$;
334 HannumAge IVW OR=0.98, 95%CI 0.95–1.02, $p=0.39$; and Intrinsic HorvathAge IVW
335 OR=0.99, 95%CI 0.98–1.01, $p=0.42$) (**Figure 2, Supplementary Figures 1 and 5–7,**
336 **Supplementary Tables 3–6 and 9–10**).

337

338 *Lung cancer*

339

340 Meta-analysed IVW MR findings suggested that genetically predicted Intrinsic HorvathAge
341 acceleration decreased the risk of lung cancer (OR=0.97 per year increase in Intrinsic
342 HorvathAge acceleration, 95%CI 0.93–1.00, $p=0.03$) (**Figure 2, Supplementary Tables 3–6**).
343 However, these results did not survive multiple testing correction (FDR $p=0.16$) and were
344 not strongly supported by sensitivity analyses (**Supplementary Figure 1, Supplementary**
345 **Tables 3–6**).

346

347 We did not find evidence of causality between other measures of epigenetic age
348 acceleration and lung cancer (GrimAge IVW OR=1.00, 95%CI 0.91–1.09, $p=0.96$; PhenoAge
349 IVW OR=0.97, 95%CI 0.94–1.00, $p=0.06$; and HannumAge IVW OR=0.99, 95%CI 0.95–1.04,
350 $p=0.82$) (**Figure 2, Supplementary Figures 1 and 5–7, Supplementary Tables 3–6 and 9–10**).

351

352 *Colorectal cancer*

353

354 Meta-analysed IVW MR findings suggested that genetically predicted GrimAge acceleration
355 increased the risk of colorectal cancer (OR=1.12 per year increase in GrimAge acceleration,
356 95%CI 1.04–1.20, $p=0.002$) (**Figure 2, Supplementary Tables 3, 5 and 6**). These results
357 survived multiple testing correction (FDR $p=0.009$) and there was little evidence of
358 heterogeneity across FinnGen and GECCO estimates ($I^2=0\%$, 95%CI “NA”, $p=0.61$).
359 Additionally, the direction of the genetically predicted effect was consistent across main and
360 sensitivity MR analyses (i.e., MR-Egger, weighted median and weighted mode) (**Figure 3,**
361 **Supplementary Tables 3, 5 and 6**) and results replicated when using UK Biobank data alone
362 (IVW OR=1.15, 95%CI 1.04–1.28, $p=0.007$) (**Figure 2, Supplementary Table 4**).

363

364 We further explored the genetically predicted effect of GrimAge on colorectal cancer using
365 GECCO data only. Single-SNP analysis revealed that the effect was not driven by a single SNP
366 (**Supplementary Table 7**). Moreover, there was no detectable evidence of uncorrelated
367 horizontal pleiotropy (MR-Egger intercept = -0.13, 95%CI -0.38–0.12, $p=0.41$), but there was
368 some evidence of heterogeneity across individual SNP estimates (Cochran’s Q 8.90, $p=0.03$)
369 (**Supplementary Table 8**) and no evidence against bias due to correlated pleiotropy (CAUSE
370 OR=1.00, 95%CI 0.96–1.04, $p=1.00$; shared $q=4\%$, 95%CI=0–24%) (**Supplementary Figure 4**).

371

372 Among subtypes, we found strong evidence for a causal relationship between GrimAge
373 acceleration and colon cancer (IVW OR=1.15, 95%CI 1.09–1.21, $p=0.006$). In contrast, we did
374 not find such evidence for rectal cancer (IVW OR=1.05, 95%CI 0.97–1.13, $p=0.24$). After
375 further stratification, the magnitude of the genetically predicted effect of GrimAge
376 acceleration on colon cancer was the same for distal (IVW OR=1.16, 95%CI 1.03–1.29,
377 $p=0.01$) and proximal colon cancer (IVW OR=1.16, 95%CI 0.97–1.40, $p=0.11$). Also, sex-

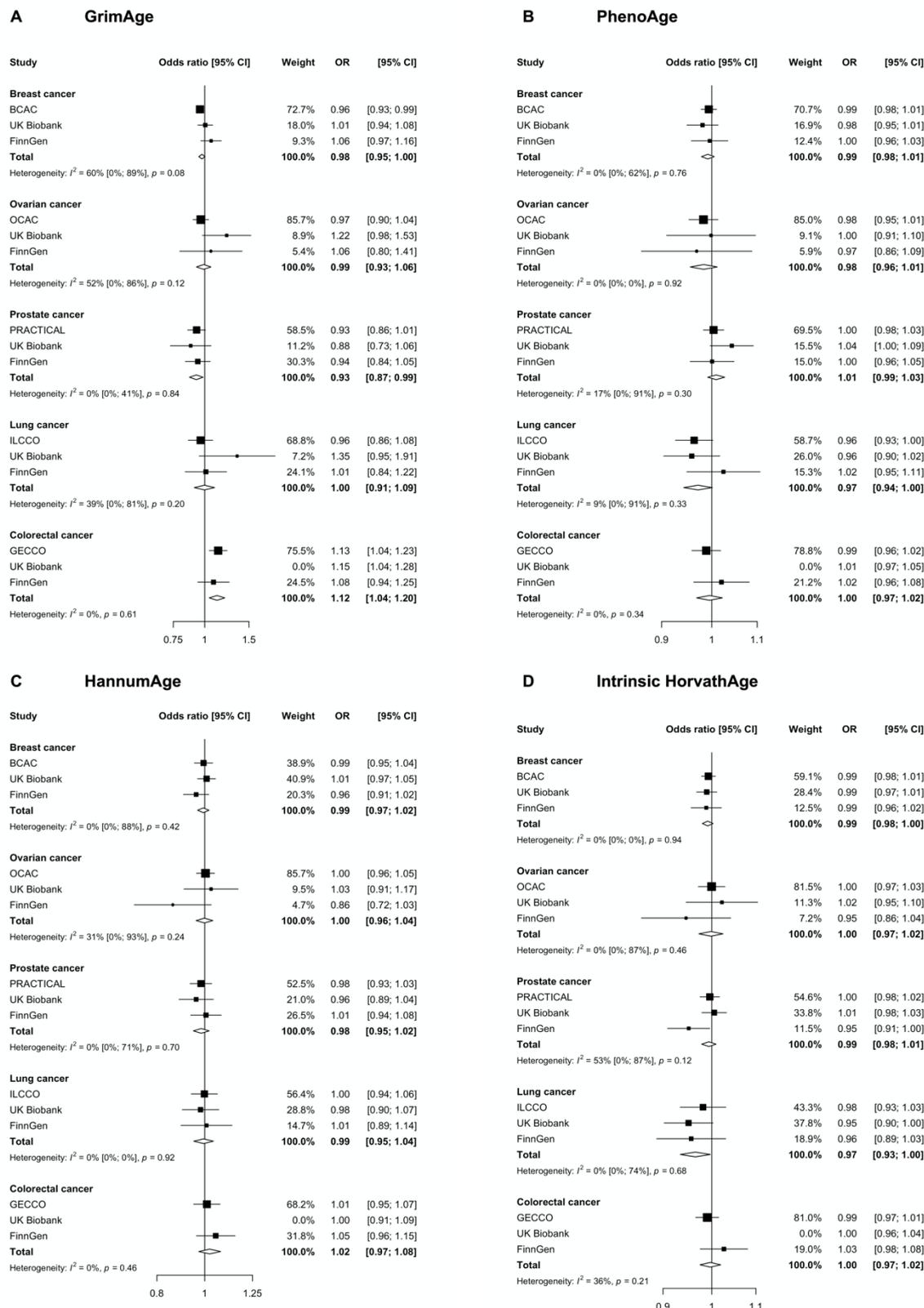
378 stratified results suggest that GrimAge acceleration may influence colorectal cancer in both
379 males (IVW OR=1.12, 95%CI 1.00–1.25, $p=0.05$) and females (IVW OR=1.14, 95%CI 1.04–
380 1.26, $p=0.008$) (**Figure 4, Supplementary Table 9**).

381

382 These findings were further supported by evidence of a positive association between
383 GrimAge acceleration and parental history of colorectal cancer (OR=1.06, 95%CI 1.00–1.12,
384 $p=0.03$) (**Figure 5, Supplementary Table 10**). Additionally, LD Score regression coefficients
385 for GrimAge acceleration and colorectal cancer were also in the expected direction (GECCO
386 $rg=0.28$, $p<0.001$; UK Biobank $rg=0.15$, $p=0.21$; FinnGen $rg=0.27$, $p=0.29$) (**Supplementary**
387 **Figure 3, Supplementary Table 11**).

388

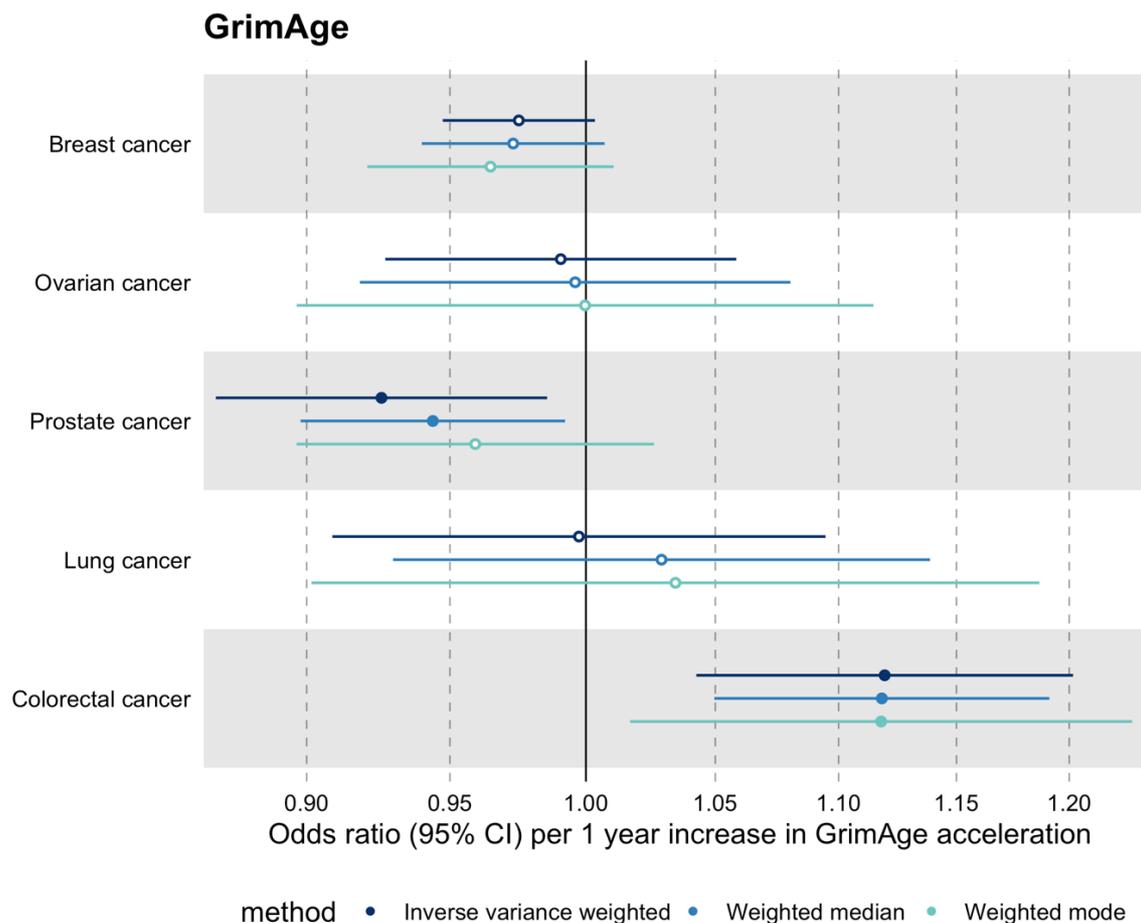
389 We did not find consistent evidence of causality between other measures of epigenetic age
390 acceleration and colorectal cancer (PhenoAge IVW OR=1.00, 95%CI 0.97–1.02, $p=0.73$;
391 HannumAge IVW OR=1.02, 95%CI 0.97–1.08, $p=0.37$; and Intrinsic HorvathAge IVW
392 OR=1.00, 95%CI 0.97–1.02, $p=0.79$) (**Figure 2, Supplementary Figures 1 and 5–7,**
393 **Supplementary Tables 3–6 and 9–10**).



394
395
396
397
398
399
400
401

Figure 2. Fixed effect meta-analysis of inverse-variance weighted Mendelian randomization estimates for genetically predicted effects of epigenetic age acceleration on multiple cancers. Odds ratios and 95% confidence intervals are reported per 1 year increase in (A) GrimAge acceleration, (B) PhenoAge acceleration, (C) HannumAge acceleration and (D) Intrinsic HorvathAge acceleration. GrimAge, PhenoAge, HannumAge and Intrinsic HorvathAge acceleration were instrumented by 4, 11, 9 and 24 genetic variants, respectively. All meta-analysis estimates were calculated using data from UK Biobank, FinnGen and international consortia, except for colorectal cancer estimates, which exclude UK Biobank data to avoid double counting.

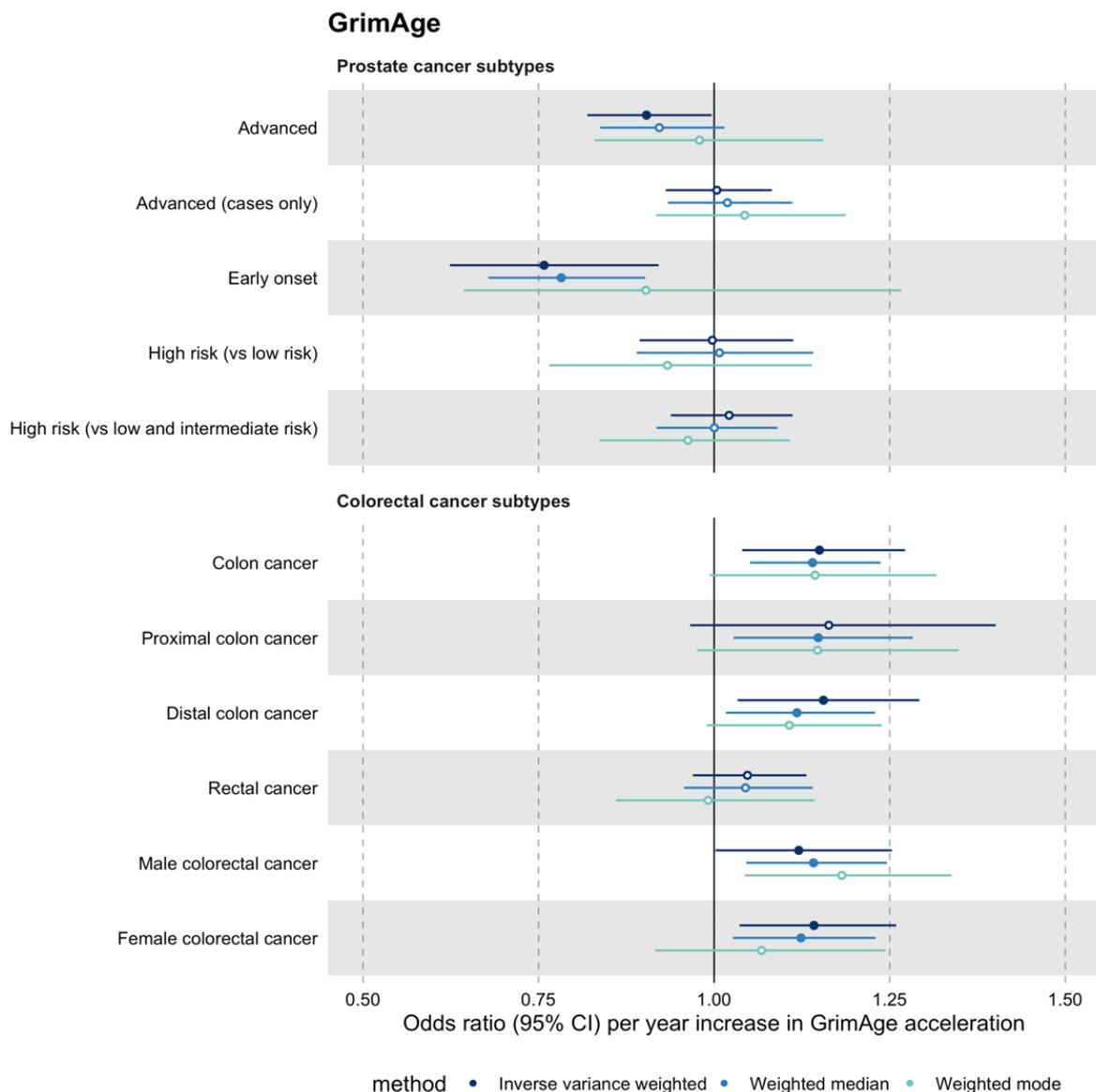
402



403
404
405
406
407
408
409

Figure 3. Fixed effect meta-analysis of Mendelian randomization estimates for genetically predicted effects of GrimAge acceleration on multiple cancers.

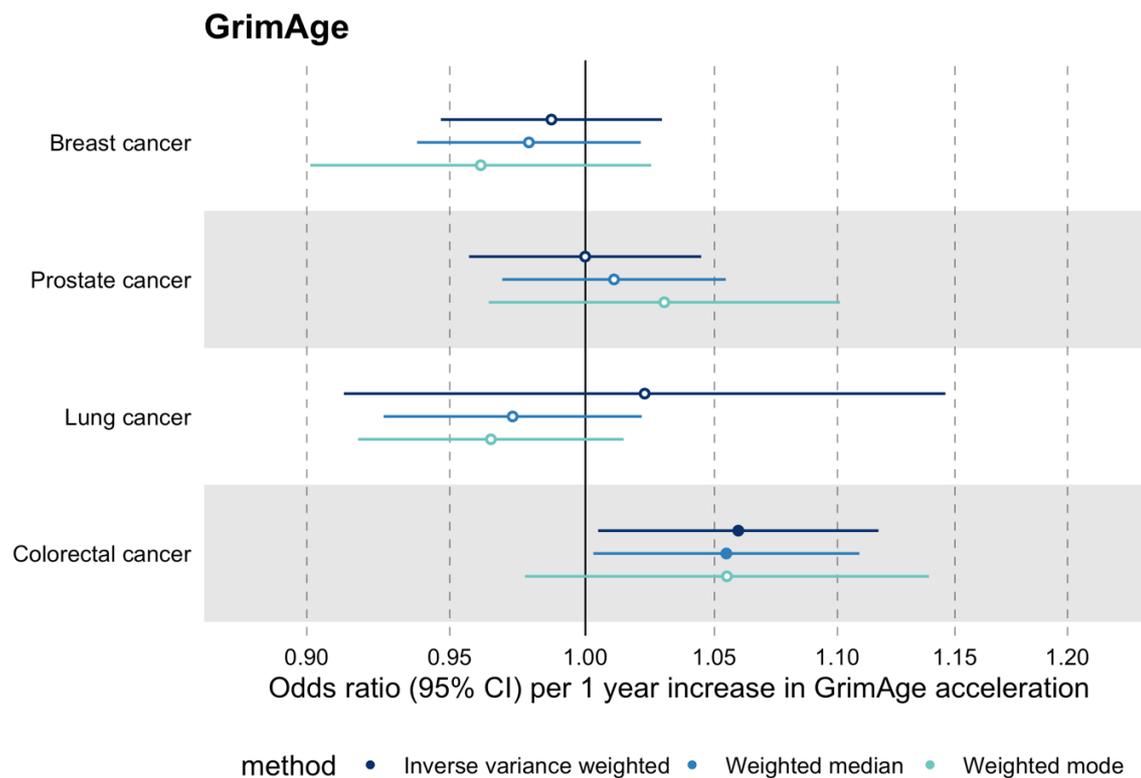
Odds ratios and 95% confidence intervals are reported per 1 year increase in GrimAge acceleration. GrimAge acceleration was instrumented by four genetic variants. Results were obtained using inverse variance weighted MR (dark blue), weighted median (sky blue) and weighted mode (turquoise) methods. All meta-analysis estimates were calculated using data from UK Biobank, FinnGen and international consortia, except for colorectal cancer estimates, which exclude UK Biobank data to avoid double counting.



410
411
412
413
414
415
416

Figure 4. Mendelian randomization estimates for genetically predicted effects of GrimAge acceleration on prostate and colorectal cancer subtypes.

Odds ratios and 95% confidence intervals are reported per 1 year increase in GrimAge acceleration. GrimAge acceleration was instrumented by four genetic variants. Results were obtained using inverse variance weighted MR (dark blue), weighted median (sky blue) and weighted mode (turquoise) methods. Data sources: PRACTICAL and GECCO.



417
418
419
420
421
422

Figure 5. Mendelian randomization estimates for genetically predicted effects of GrimAge acceleration on parental history of multiple cancers.

Odds ratios and 95% confidence intervals are reported per 1 year increase in GrimAge acceleration. GrimAge acceleration was instrumented by four genetic variants. Results were obtained using inverse variance weighted MR (dark blue), weighted median (sky blue) and weighted mode (turquoise) methods. Data source: UK Biobank.

423 **DISCUSSION**

424

425 In this comprehensive two-sample MR study of epigenetic age acceleration and multiple
426 cancers, we found evidence to suggest that genetically predicted GrimAge acceleration may
427 increase the risk of colorectal cancer in both males and females. Among subtypes, effects
428 appeared to be stronger in relation to colon than rectal cancer. Our MR results also
429 suggested that genetically predicted GrimAge acceleration may decrease the risk of prostate
430 cancer. The protective effect appeared strongest for advanced and early onset prostate
431 cancer. Nevertheless, we cannot rule out the possibility of bias, as the prostate cancer
432 results did not replicate in the MR analysis using data regarding parental history of cancer,
433 and LD Score regression analyses did not indicate a genetic correlation between the traits.
434 We also identified a protective effect of Intrinsic HorvathAge acceleration on lung cancer,
435 but this did not pass multiple testing correction and was substantially attenuated in
436 sensitivity analyses. Finally, we found no consistent evidence for other measures of
437 epigenetic age acceleration and cancers.

438

439 Our estimates for the association between GrimAge and colorectal cancer were consistent
440 with those reported in Dugue et al.¹⁰, an observational nested case-control study in the
441 Melbourne Collaborative Cohort Study (RR=1.04 per year increase in GrimAge acceleration,
442 95%CI 1.01–1.07, $p=0.02$). However, our findings contrast with those highlighted in Hillary et
443 al.¹⁵. The latter authors observed no evidence of an association between GrimAge
444 acceleration and colorectal cancer after correcting for multiple testing. Nevertheless, it is
445 possible that their analyses were underpowered, as their sample only included 63 colorectal

446 cancer cases (0.66%). More importantly, the direction of the reported estimate is consistent
447 with our findings and those presented in Dugue et al.¹⁰.

448

449 As in our study, Dugue et al.¹⁰ provided evidence of a weak inverse association between
450 GrimAge acceleration and prostate cancer (RR=0.86 per 5-year increase in GrimAge
451 acceleration, 95%CI 0.74–1.01, $p=0.07$). Nevertheless, associations in both studies seem to
452 have been driven by different prostate cancer subtypes. Stratified results in Dugue et al.¹⁰
453 suggested that the protective effect of GrimAge acceleration is driven by non-aggressive
454 prostate cancer, while we only found evidence of a protective effect for early onset prostate
455 cancer and advanced prostate cancer (which both tend to be more aggressive³⁷).

456

457 The weak inverse association observed in Dugue et al.¹⁰ could have emerged due to
458 detection bias. Health-conscious individuals are more likely to use health services and
459 undergo prostate cancer screening^{38 39}. Hence, their likelihood of being diagnosed with less
460 severe cancers may be greater, despite them potentially also having lower GrimAge
461 acceleration. This could lead to spurious conclusions, such as that people with higher
462 GrimAge acceleration are less likely to develop non-aggressive prostate cancer. The inverse
463 effect in our study could be explained by selection bias, as cases of early onset prostate
464 cancer (who possibly have higher GrimAge acceleration) are likely to be missed if studies
465 only recruit middle- to late-aged participants. Likewise, individuals with longer lifespans
466 (and possibly lower GrimAge acceleration) are more likely to develop advanced prostate
467 cancer, simply because they have not died of other illnesses (survival bias).

468

469 Observational evidence for the association between other measures of epigenetic ageing
470 and cancer is inconclusive (the pre-existing evidence has been summarised in
471 **Supplementary Table 12**). For instance, epigenetic clock acceleration has been positively
472 associated with breast^{8 11 12} and lung cancer^{5 9 10} in some studies. However, Durso et al.¹⁶,
473 Hillary et al.¹⁵ and Dugue et al.¹⁴ did not find strong evidence to support this. In some cases,
474 observational evidence is stronger for some clocks than it is for others. For example, for
475 colorectal cancer, evidence of a positive association was much stronger for second-
476 generation clocks¹⁰ than for first-generation clocks^{14 16}. For prostate cancer, evidence of an
477 association was only found for GrimAge^{10 14}, as in our study. To date, the association
478 between epigenetic age acceleration and ovarian cancer has not been explored
479 observationally. Although our findings were less susceptible to biases that often influence
480 observational research, they still provide no compelling evidence of a causal association
481 between several measures of epigenetic clock acceleration and cancer.

482

483 This MR study had several strengths. For instance, we pooled results from multiple sources
484 using fixed effect meta-analysis to improve the precision of the estimates presented in
485 McCartney et al²⁰. We also conducted extra sensitivity analyses, such as single-SNP and
486 CAUSE analyses, to assess the validity of the MR assumptions. Moreover, we performed
487 subtype-specific MR analyses and sought to replicate our results using UK Biobank GWAS
488 data on parental history of cancer and LD Score regression. Additionally, our findings
489 contribute to the identification of modifiable targets for future interventions aimed at
490 reversing epigenetic ageing for the prevention of cancer. Compared to clinical trials, MR
491 provides a cheaper, quicker, and ethical way of assessing the long-term impact of

492 interventions on epigenetic ageing. This is especially relevant while attempts to develop
493 interventions which reverse epigenetic ageing are still in early stages⁴⁰⁻⁴³.

494

495 The findings from this study should be interpreted in light of its limitations. We only
496 identified four genetic instruments for GrimAge acceleration, which explained 0.47% of the
497 variance in the trait. This could lead to two issues: low statistical power and horizontal
498 pleiotropy. First, our GrimAge analyses were underpowered to detect ORs <1.20 for
499 colorectal cancer and >0.84 for prostate cancer. Therefore, it is possible that our findings do
500 not reflect a true effect (we identified ORs=1.12 and 0.93 for colorectal and prostate cancer,
501 respectively). Similarly, our study was underpowered to detect genetically predicted effects
502 of GrimAge acceleration on cancer subtypes and cancers with smaller sample sizes (i.e.,
503 ovarian and lung cancer). Some of our sensitivity analyses, such as the MR-Egger intercept
504 test used to detect uncorrelated horizontal pleiotropy, also had low power, resulting in
505 imprecise estimates. The weighted mode method may also be misleading in this context, as
506 its use is limited in the presence of very few SNPs. Although these limitations potentially
507 undermine the validity of our results, it is reassuring that point estimates for the genetically
508 predicted effect of GrimAge acceleration on prostate and colorectal cancer were consistent
509 across MR methods and study populations. However, since CAUSE analyses did not provide
510 evidence against confounding by correlated horizontal pleiotropy, it is possible that the
511 genetically predicted effects identified are attributed to correlated pleiotropy (whereby
512 SNPs are associated with epigenetic age acceleration and cancer through a shared heritable
513 factor) rather than a causal effect of GrimAge on cancer risk.

514

515 One could argue that because the results for GrimAge acceleration were inconsistent with
516 those obtained for other measures of epigenetic age acceleration, chance and horizontal
517 pleiotropy are more likely explanations for our findings. However, inconsistencies across
518 epigenetic ageing measures do not necessarily invalidate our results. They may simply
519 reflect differences in how clocks were trained (i.e., they were trained on different outcomes,
520 tissues, and populations). Different clocks may capture information on distinct underlying
521 biological ageing mechanisms². For example, GrimAge was trained on mortality and smoking
522 (factors which are closely related to cancer risk), which may explain why it outperforms
523 other measures of epigenetic ageing in predicting time-to-cancer⁶. Although little is known
524 about the underlying mechanisms, GrimAge may plausibly influence cancer risk through its
525 effect on hormonal, inflammatory and metabolic processes⁴⁴⁻⁴⁶. Furthermore, Wang et al.⁴⁷
526 suggests that the disruption of epigenomic maintenance systems, reflected by either
527 epigenetic clock acceleration or deceleration, may be involved in the process of
528 carcinogenesis.

529

530 Although promising in terms of consistency and biological plausibility, further research is
531 required to confirm our findings. For example, multivariable MR^{48 49} could be used to
532 disentangle the causal effects of GrimAge acceleration on cancer from shared heritable
533 factors such as and blood cell composition. Additionally, our analyses could be replicated
534 using other large independent cancer datasets to increase power. It would also be useful to
535 replicate our analyses once a larger GWAS of epigenetic ageing with more genetic
536 instruments for GrimAge acceleration is available. This would allow for a more rigorous
537 assessment of horizontal pleiotropy and may be used to assess clustering of genetic variants
538 to reveal distinct biological mechanisms underlying the effects⁵⁰.

539

540 The selection of “super controls” (e.g., in UK Biobank, FinnGen and GECCO), with no other
541 cancers, related lesions (i.e., benign, in situ, uncertain or unspecified behaviour neoplasms)
542 or reported family history of cancer, could have inflated cancer GWAS effect sizes (and our
543 MR estimates), because “super controls” are healthier than the general population and are
544 less likely to be genetically predisposed to develop cancer.

545

546 Another limitation is that we did not have access to individual level data. Therefore, we
547 were unable to stratify the analyses by potential effect modifiers, such as sex, smoking and
548 menopausal status. Moreover, we did not have sex-specific instruments for sex-specific
549 cancers. However, it is unlikely that the genetic architecture of epigenetic clock acceleration
550 differs across sexes, as DNAm levels at individual clock CpGs are highly correlated between
551 males and females^{51 52}.

552

553 Finally, to reduce bias due to population stratification, this study was conducted using data
554 from participants of European ancestry only. The GWAS data used for the analyses had been
555 adjusted for the top genetic principal components for the same reason. Assortative mating
556 is unlikely to be a problem in the context of this study because we would not expect people
557 to select partners based on their epigenetic age acceleration. Despite this, confounding due
558 to population stratification and assortative mating cannot be ruled out completely, as it is
559 not possible to test the second MR assumption (i.e., independence assumption).

560 Furthermore, more research is required to see if our results could translate to other
561 ancestries.

562

563 In conclusion, our findings suggest that genetically predicted GrimAge acceleration may
564 increase the risk of colorectal cancer. There is more limited evidence that it may decrease
565 the risk of prostate cancer. Findings were less consistent for other epigenetic clocks and
566 cancers. Further work is required to investigate the potential mechanisms underlying the
567 genetically predicted effects identified in this study.

568

569 **Acknowledgements**

570

571 We thank Richard Wilkinson for proofreading several versions of the manuscript. We would
572 also like to acknowledge the participants and investigators of the FinnGen and UK Biobank
573 studies. GWAS data on parental history of cancer were generated using the UK Biobank
574 Resource under application number 15825. Finally, we would like to thank the BCAC, OCAC,
575 PRACTICAL, ILCCO, GECCO and CIMBA consortiums for their contributions.

576

577 **Competing interests**

578

579 ATL and SH declare that UC Regents filed the patent “DNA METHYLATION BASED
580 BIOMARKERS FOR LIFE EXPECTANCY AND MORBIDITY” (in pending status), and that the
581 Epigenetic Clock Development Foundation and Foxo Labs hold licenses. SH receives
582 consulting fees from the Epigenetic Clock Development Foundation and royalties for patents
583 involving epigenetic clocks. REM has received a speaker fee from Illumina and is an advisor
584 to the Epigenetic Clock Development Foundation. TGR is employed part time by Novo
585 Nordisk outside of this work. The other authors declare that they have no competing
586 interests.

587 **REFERENCES**

588

- 589 1. Fransquet PD, Wrigglesworth J, Woods RL, et al. The epigenetic clock as a predictor of
590 disease and mortality risk: a systematic review and meta-analysis. *Clin Epigenetics*
591 2019;11(1):62. doi: 10.1186/s13148-019-0656-7 [published Online First: 2019/04/13]
- 592 2. Liu Z, Leung D, Thrush K, et al. Underlying features of epigenetic aging clocks in vivo and in
593 vitro. *Aging Cell* 2020;19(10):e13229. doi: 10.1111/accel.13229 [published Online
594 First: 2020/09/16]
- 595 3. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal
596 quantitative views of human aging rates. *Mol Cell* 2013;49(2):359-67. doi:
597 10.1016/j.molcel.2012.10.016 [published Online First: 2012/11/28]
- 598 4. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*
599 2013;14:R115.
- 600 5. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and
601 healthspan. *AGING* 2018;10(4)
- 602 6. Lu AT, Quach A, Wilson JG, et al. DNA methylation GrimAge strongly predicts lifespan and
603 healthspan. *AGING* 2019;11(2):303-27.
- 604 7. McCrory C, Fiorito G, Hernandez B, et al. GrimAge Outperforms Other Epigenetic Clocks in
605 the Prediction of Age-Related Clinical Phenotypes and All-Cause Mortality. *J Gerontol*
606 *A Biol Sci Med Sci* 2021;76(5):741-49. doi: 10.1093/gerona/glaa286 [published Online
607 First: 2020/11/20]
- 608 8. Ambatipudi S, Horvath S, Perrier F, et al. DNA methylome analysis identifies accelerated
609 epigenetic ageing associated with postmenopausal breast cancer susceptibility. *Eur J*
610 *Cancer* 2017;75:299-307. doi: 10.1016/j.ejca.2017.01.014 [published Online First:
611 2017/03/05]
- 612 9. Levine ME, Hosgood HD, Chen B, et al. DNA methylation age of blood predicts future
613 onset of lung cancer in the women's health initiative. *AGING* 2015;7(9):690-700.
- 614 10. Dugue PA, Bassett JK, Wong EM, et al. Biological Aging Measures Based on Blood DNA
615 Methylation and Risk of Cancer: A Prospective Study. *JNCI Cancer Spectr*
616 2021;5(1):pkaa109. doi: 10.1093/jncics/pkaa109 [published Online First:
617 2021/01/15]
- 618 11. Kresovich JK, Xu Z, O'Brien KM, et al. Methylation-Based Biological Age and Breast
619 Cancer Risk. *J Natl Cancer Inst* 2019;111(10):1051-58. doi: 10.1093/jnci/djz020
620 [published Online First: 2019/02/23]
- 621 12. Kresovich JK, Xu Z, O'Brien KM, et al. Epigenetic mortality predictors and incidence of
622 breast cancer. *AGING* 2019;11(24):11975-87.
- 623 13. Zheng Y, Joyce BT, Colicino E, et al. Blood Epigenetic Age may Predict Cancer Incidence
624 and Mortality. *EBioMedicine* 2016;5:68-73. doi: 10.1016/j.ebiom.2016.02.008
625 [published Online First: 2016/04/15]
- 626 14. Dugue PA, Bassett JK, Joo JE, et al. DNA methylation-based biological aging and cancer
627 risk and survival: Pooled analysis of seven prospective studies. *Int J Cancer*
628 2018;142(8):1611-19. doi: 10.1002/ijc.31189 [published Online First: 2017/12/03]
- 629 15. Hillary RF, Stevenson AJ, McCartney DL, et al. Epigenetic measures of ageing predict the
630 prevalence and incidence of leading causes of death and disease burden. *Clin*
631 *Epigenetics* 2020;12(1):115. doi: 10.1186/s13148-020-00905-6 [published Online
632 First: 2020/08/02]

- 633 16. Durso DF, Bacalini MG, Sala C, et al. Acceleration of leukocytes' epigenetic age as an
634 early tumor- and sex-specific marker of breast and colorectal cancer. *Oncotarget*
635 2017;8(14):23237-45.
- 636 17. Wang C, Ni W, Yao Y, et al. DNA methylation-based biomarkers of age acceleration and
637 all-cause death, myocardial infarction, stroke, and cancer in two cohorts: The NAS,
638 and KORA F4. *EBioMedicine* 2021;63:103151. doi: 10.1016/j.ebiom.2020.103151
639 [published Online First: 2020/12/07]
- 640 18. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for
641 establishing the causal role of epigenetic processes in pathways to disease. *Int J*
642 *Epidemiol* 2012;41(1):161-76. doi: 10.1093/ije/dyr233 [published Online First:
643 2012/03/17]
- 644 19. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute
645 to understanding environmental determinants of disease? *Int J Epidemiol*
646 2003;32(1):1-22. doi: 10.1093/ije/dyg070 [published Online First: 2003/04/12]
- 647 20. McCartney DL, Min JL, Richmond RC, et al. Genome-wide association studies identify 137
648 genetic loci for DNA methylation biomarkers of aging. *Genome Biol* 2021;22(1):194.
649 doi: 10.1186/s13059-021-02398-9 [published Online First: 2021/07/01]
- 650 21. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of
651 Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*
652 2021;71(3):209-49. doi: 10.3322/caac.21660 [published Online First: 2021/02/05]
- 653 22. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of
654 observational studies in epidemiology using mendelian randomisation (STROBE-MR):
655 explanation and elaboration. *Bmj* 2021 doi: 10.1136/bmj.n2233
- 656 23. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal
657 inference across the human phenome. *Elife* 2018;7 doi: 10.7554/eLife.34408
658 [published Online First: 2018/05/31]
- 659 24. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with
660 multiple genetic variants using summarized data. *Genet Epidemiol* 2013;37(7):658-
661 65. doi: 10.1002/gepi.21758 [published Online First: 2013/10/12]
- 662 25. von Hippel PT. The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC*
663 *Med Res Methodol* 2015;15:35. doi: 10.1186/s12874-015-0024-z [published Online
664 First: 2015/04/17]
- 665 26. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful
666 Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B*
667 *(Methodological)* 1995;57(1):289-300.
- 668 27. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to
669 causal inference. *Statistical Methods in Medical Research* 2007;16:309-30. doi:
670 10.1177/0962280206077743
- 671 28. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a
672 guide, glossary, and checklist for clinicians. *BMJ* 2018;362:k601. doi:
673 10.1136/bmj.k601 [published Online First: 2018/07/14]
- 674 29. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies
675 with weak instruments. *Stat Med* 2011;30(11):1312-23. doi: 10.1002/sim.4197
676 [published Online First: 2011/03/25]
- 677 30. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments:
678 effect estimation and bias detection through Egger regression. *Int J Epidemiol*
679 2015;44(2):512-25. doi: 10.1093/ije/dyv080 [published Online First: 2015/06/08]

- 680 31. Bowden J, Davey Smith G, Haycock PC, et al. Consistent Estimation in Mendelian
681 Randomization with Some Invalid Instruments Using a Weighted Median Estimator.
682 *Genet Epidemiol* 2016;40(4):304-14. doi: 10.1002/gepi.21965 [published Online First:
683 2016/04/12]
- 684 32. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian
685 randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*
686 2017;46(6):1985-98. doi: 10.1093/ije/dyx102 [published Online First: 2017/10/19]
- 687 33. Morrison J, Knoblauch N, Marcus JH, et al. Mendelian randomization accounting for
688 correlated and uncorrelated pleiotropic effects using genome-wide summary
689 statistics. *Nat Genet* 2020;52(7):740-47. doi: 10.1038/s41588-020-0631-4 [published
690 Online First: 2020/05/27]
- 691 34. Bowden J, Del Greco MF, Minelli C, et al. Improving the accuracy of two-sample
692 summary-data Mendelian randomization: moving beyond the NOME assumption. *Int*
693 *J Epidemiol* 2019;48(3):728-42. doi: 10.1093/ije/dyy258 [published Online First:
694 2018/12/19]
- 695 35. DeBoever C, Tanigawa Y, Aguirre M, et al. Assessing Digital Phenotyping to Enhance
696 Genetic Studies of Human Diseases. *Am J Hum Genet* 2020;106(5):611-22. doi:
697 10.1016/j.ajhg.2020.03.007 [published Online First: 2020/04/11]
- 698 36. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across
699 human diseases and traits. *Nat Genet* 2015;47(11):1236-41. doi: 10.1038/ng.3406
700 [published Online First: 2015/09/29]
- 701 37. Salinas CA, Tsodikov A, Ishak-Howard M, et al. Prostate cancer in young men: an
702 important clinical entity. *Nat Rev Urol* 2014;11(6):317-23. doi:
703 10.1038/nrurol.2014.91 [published Online First: 2014/05/14]
- 704 38. Li Y, Doukas DJ. Health motivation and emotional vigilance in genetic testing for prostate
705 cancer risk. *Clin Genet* 2004;66(6):512-6. doi: 10.1111/j.1399-0004.2004.00351.x
706 [published Online First: 2004/11/04]
- 707 39. Rai T, Clements A, Bukach C, et al. What influences men's decision to have a prostate-
708 specific antigen test? A qualitative study. *Fam Pract* 2007;24(4):365-71. doi:
709 10.1093/fampra/cmm033 [published Online First: 2007/07/24]
- 710 40. Fahy GM, Brooke RT, Watson JP, et al. Reversal of epigenetic aging and
711 immunosenescent trends in humans. *Aging Cell* 2019;18(6):e13028. doi:
712 10.1111/accel.13028 [published Online First: 2019/09/10]
- 713 41. Fitzgerald KN, Hodges R, Hanes D, et al. Potential reversal of epigenetic age using a diet
714 and lifestyle intervention: a pilot randomized clinical trial. *AGING* 2021;13(7):9419-
715 32. doi: 10.18632/aging.202913
- 716 42. Gensous N, Garagnani P, Santoro A, et al. One-year Mediterranean diet promotes
717 epigenetic rejuvenation with country- and sex-specific effects: a pilot study from the
718 NU-AGE project. *Geroscience* 2020;42(2):687-701. doi: 10.1007/s11357-019-00149-0
719 [published Online First: 2020/01/26]
- 720 43. Chen L, Dong Y, Bhagatwala J, et al. Effects of Vitamin D3 Supplementation on Epigenetic
721 Aging in Overweight and Obese African Americans With Suboptimal Vitamin D
722 Status: A Randomized Clinical Trial. *J Gerontol A Biol Sci Med Sci* 2019;74(1):91-98.
723 doi: 10.1093/gerona/gly223 [published Online First: 2018/09/27]
- 724 44. Yu M, Hazelton WD, Luebeck GE, et al. Epigenetic Aging: More Than Just a Clock When It
725 Comes to Cancer. *Cancer Res* 2020;80(3):367-74. doi: 10.1158/0008-5472.CAN-19-
726 0924 [published Online First: 2019/11/07]

- 727 45. Bottazzi B, Riboli E, Mantovani A. Aging, inflammation and cancer. *Semin Immunol*
728 2018;40:74-82. doi: 10.1016/j.smim.2018.10.011 [published Online First:
729 2018/11/10]
- 730 46. Lau CE, Robinson O. DNA methylation age as a biomarker for cancer. *Int J Cancer*
731 2021;148(11):2652-63. doi: 10.1002/ijc.33451 [published Online First: 2021/01/05]
- 732 47. Wang T, Maden SK, Luebeck GE, et al. Dysfunctional epigenetic aging of the normal
733 colon and colorectal cancer risk. *Clin Epigenetics* 2020;12(1):5. doi: 10.1186/s13148-
734 019-0801-3 [published Online First: 2020/01/05]
- 735 48. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic
736 genetic variants to estimate causal effects. *Am J Epidemiol* 2015;181(4):251-60. doi:
737 10.1093/aje/kwu283 [published Online First: 2015/01/30]
- 738 49. Sanderson E, Davey Smith G, Windmeijer F, et al. An examination of multivariable
739 Mendelian randomization in the single-sample and two-sample summary data
740 settings. *Int J Epidemiol* 2019;48(3):713-27. doi: 10.1093/ije/dyy262 [published
741 Online First: 2018/12/12]
- 742 50. Foley CN, Mason AM, Kirk PDW, et al. MR-Clust: clustering of genetic variants in
743 Mendelian randomization with similar causal estimates. *Bioinformatics*
744 2021;37(4):531-41. doi: 10.1093/bioinformatics/btaa778 [published Online First:
745 2020/09/12]
- 746 51. Grodstein F, Lemos B, Yu L, et al. Characteristics of Epigenetic Clocks Across Blood and
747 Brain Tissue in Older Women and Men. *Front Neurosci* 2020;14:555307. doi:
748 10.3389/fnins.2020.555307 [published Online First: 2021/01/26]
- 749 52. Tajuddin SM, Hernandez DG, Chen BH, et al. Novel age-associated DNA methylation
750 changes and epigenetic age acceleration in middle-aged African Americans and
751 whites. *Clin Epigenetics* 2019;11(1):119. doi: 10.1186/s13148-019-0722-1 [published
752 Online First: 2019/08/21]
753

754 **SUPPLEMENTARY MATERIAL**

755

756 **Supplement 1. STROBE-MR checklist of recommended items to address in reports of**
757 ***Mendelian randomization studies***²²

758

	Item Nº	Recommendation	Comments/page Nº
Title and abstract			
	1	Indicate Mendelian randomization (MR) as the study's design in the title and/or the abstract if that is a main purpose of the study.	'Mendelian randomization' in title and abstract
Introduction			
Background	2	Explain the scientific background and rationale for the reported study. Is a potential causal relationship between exposure and outcome plausible? Justify why MR is a helpful method to address the study question.	Pages 5-7
Objectives	3	State specific objectives clearly, including pre-specified causal hypotheses (if any). State that MR is a method that, under specific assumptions, intends to estimate causal effects.	Page 7
Methods			
Study design and data sources	4	Present key elements of study design early in the article. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following: (a) Setting: Describe the study design and the underlying population, if possible. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, when available. (b) Participants: Give the eligibility criteria, and the sources and methods of selection of participants. Report the sample size, and whether any power or sample size calculations were carried out prior to the main analysis. (c) Explain how the analysed sample size was arrived at. (d) Describe measurement, quality and selection of genetic variants. (e) For each exposure, outcome and other relevant variables, describe methods of assessment and diagnostic criteria for diseases. (f) Provide details of ethics committee approval and participant informed consent, if relevant.	Pages 8-10, Supplementary methods, Table 1, Figure 1
Assumptions	5	Explicitly state the three core IV assumptions for the main analysis (relevance, independence, and exclusion restriction) as well as assumptions for any additional or sensitivity analysis.	Page 11, Supplementary methods
Statistical methods: main analysis	6	Describe statistical methods and statistics used. (a) Describe how quantitative variables were handled in the analyses (i.e., scale, units, model). (b) Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected. (c) Describe the MR estimator (e.g. two-stage least squares, Wald ratio) and related statistics. Detail the included covariates and, in case of two-sample MR,	Page 11, Supplementary methods

		whether the same covariate set was used for adjustment in the two samples. (d) Explain how missing data were addressed. (e) If applicable, say how multiple testing was addressed.	
Assessment of assumptions	7	Describe any methods or prior knowledge used to assess the assumptions or justify their validity.	Pages 11-13, Supplementary methods
Sensitivity analyses	8	Describe any sensitivity analyses or additional analyses performed (e.g. comparison of effect estimates from different approaches, independent replication, bias analytic techniques, validation of instruments, simulations).	Pages 11-13, Supplementary methods
Software and pre-registration	9	(a) Name statistical software and package(s), including version and settings used. (b) State whether the study protocol and details were pre-registered (as well as when and where).	Pages 13-14, Supplementary methods
Results			
Descriptive data	10	(a) Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow diagram. (b) Report summary statistics for phenotypic exposure(s), outcome(s) and other relevant variables (e.g. means, SDs, proportions). (c) If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies. (d) For two-sample MR: i. Provide justification of the similarity of the genetic variant-exposure associations between the exposure and outcome samples. ii. Provide information on the number of individuals who overlap between the exposure and outcome studies.	Not applicable, as we used summary statistics from previously published genome-wide association studies. We cite these accordingly. Furthermore, there is no sample overlap between the exposure and outcome studies.
Main results	11	(a) Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale. (b) Report MR estimates of the relationship between exposure and outcome, and the measures of uncertainty from the MR analysis on an interpretable scale, such as odds ratio or relative risk per SD difference. (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time-period. (d) Consider plots to visualize results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure).	Pages 16-20, Figure 2, Supplementary figures and tables
Assessment of validity	12	(a) Report the assessment of the validity of the assumptions. (b) Report any additional statistics (e.g., assessments of heterogeneity across genetic variants, such as I^2 , Q statistic or E-value).	Pages 16-20 and Figure 3, Supplementary figures and tables
Sensitivity and additional analyses	13	(a) Report any sensitivity analyses to assess the robustness of the main results to violations of the assumptions. (b) Report results from other sensitivity analyses or additional analyses.	Pages 16-20 and Figures 3-5, Supplementary figures and tables

		(c) Report any assessment of direction of causal relationship (e.g., bidirectional MR). (d) When relevant, report and compare with estimates from non-MR analyses. (e) Consider any additional plots to visualize results (e.g., leave-one-out analyses).	
Discussion			
Key results	14	Summarize key results with reference to study objectives.	Page 25
Limitations	15	Discuss limitations of the study, taking into account the validity of the IV assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias, and any efforts to address them.	Pages 25-31
Interpretation	16	(a) Meaning: Give a cautious overall interpretation of results in the context of their limitations and in comparison with other studies. (b) Mechanisms: Discuss underlying biological mechanisms that could drive a potential causal relationship between the investigated exposure and the outcome, and whether the gene-environment equivalence assumption is reasonable. Use causal language carefully, clarifying that IV estimates may provide causal effects only under certain assumptions. (c) Clinical relevance: Discuss whether the results have clinical or public policy relevance, and to what extent they inform effect sizes of possible interventions.	Pages 25-31
Generalisability	17	Discuss the generalizability of the study results (a) to other populations (i.e. external validity), (b) across other exposure periods/timings, and (c) across other levels of exposure.	Page 30
Other information			
Funding	18	Describe sources of funding and the role of funders in the present study and, if applicable, sources of funding for the databases and original study or studies on which the present study is based.	Abstract
Data and data sharing	19	Provide the data used to perform all analyses or report where and how the data can be accessed, and reference these sources in the article. Provide the statistical code needed to reproduce the results in the article, or report whether the code is publicly accessible and if so, where.	Supplementary methods
Conflicts of interest	20	All authors should declare all potential conflicts of interest.	Page 31

759
760
761
762

Adapted from: BMJ 2021;375:n2233 <http://dx.doi.org/10.1136/bmj.n2233>

763 **Supplement 2. PRACTICAL consortia authorship and acknowledgements**

764

765 **PIs from the PRACTICAL (<http://practical.icr.ac.uk/>) consortia:**

766

767 Rosalind A. Eeles^{1,2}, Christopher A. Haiman³, Zsofia Kote-Jarai¹, Fredrick R. Schumacher^{4,5},
768 Sara Benlloch^{6,1}, Ali Amin Al Olama^{6,7}, Kenneth R. Muir⁸, Sonja I. Berndt⁹, David V. Conti³,
769 Fredrik Wiklund¹⁰, Stephen Chanock⁹, Ying Wang¹¹, Catherine M. Tangen¹², Jyotsna
770 Batra^{13,14}, Judith A. Clements^{13,14}, APCB BioResource (Australian Prostate Cancer
771 BioResource)^{15,14}, Henrik Grönberg¹⁰, Nora Pashayan^{16,17}, Johanna Schleutker^{18,19}, Demetrius
772 Albanes⁹, Stephanie Weinstein⁹, Alicja Wolk²⁰, Catharine M. L. West²¹, Lorelei A. Mucci²²,
773 Géraldine Cancel-Tassin^{23,24}, Stella Koutros⁹, Karina Dalsgaard Sørensen^{25,26}, Eli Marie
774 Grindedal²⁷, David E. Neal^{28,29,30}, Freddie C. Hamdy^{31,32}, Jenny L. Donovan³³, Ruth C. Travis³⁴,
775 Robert J. Hamilton^{35,36}, Sue Ann Ingles³⁷, Barry S. Rosenstein³⁸, Yong-Jie Lu³⁹, Graham G.
776 Giles^{40,41,42}, Robert J. MacInnis^{40,41}, Adam S. Kibel⁴³, Ana Vega^{44,45,46}, Manolis
777 Kogevinas^{47,48,49,50}, Kathryn L. Penney⁵¹, Jong Y. Park⁵², Janet L. Stanford^{53,54}, Cezary
778 Cybulski⁵⁵, Børge G. Nordestgaard^{56,57}, Sune F. Nielsen^{56,57}, Hermann Brenner^{58,59,60},
779 Christiane Maier⁶¹, Jeri Kim⁶², Esther M. John⁶³, Manuel R. Teixeira^{64,65,66}, Susan L.
780 Neuhausen⁶⁷, Kim De Ruyck⁶⁸, Azad Razack⁶⁹, Lisa F. Newcomb^{53,70}, Davor Lessel⁷¹, Radka
781 Kaneva⁷², Nawaid Usmani^{73,74}, Frank Claessens⁷⁵, Paul A. Townsend^{76,77}, Jose Esteban
782 Castela⁷⁸, Monique J. Roobol⁷⁹, Florence Menegaux⁸⁰, Kay-Tee Khaw⁸¹, Lisa Cannon-
783 Albright^{82,83}, Hardev Pandha⁷⁷, Stephen N. Thibodeau⁸⁴, David J. Hunter⁸⁵, Peter Kraft⁸⁶,
784 William J. Blot^{87,88}, Elio Riboli⁸⁹

785

786 ¹The Institute of Cancer Research, London, SM2 5NG, UK

787 ²Royal Marsden NHS Foundation Trust, London, SW3 6JJ, UK

788 ³Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of
789 Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los
790 Angeles, CA 90015, USA

791 ⁴Department of Population and Quantitative Health Sciences, Case Western Reserve
792 University, Cleveland, OH 44106-7219, USA

793 ⁵Seidman Cancer Center, University Hospitals, Cleveland, OH 44106, USA.

794 ⁶Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care,
795 University of Cambridge, Strangeways Research Laboratory, Cambridge CB1 8RN, UK

796 ⁷University of Cambridge, Department of Clinical Neurosciences, Stroke Research Group, R3,
797 Box 83, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK

798 ⁸Division of Population Health, Health Services Research and Primary Care, University of
799 Manchester, Oxford Road, Manchester, M13 9PL, UK

800 ⁹Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda,
801 Maryland, 20892, USA

802 ¹⁰Department of Medical Epidemiology and Biostatistics, Karolinska Institute, SE-171 77
803 Stockholm, Sweden

804 ¹¹Department of Population Science, American Cancer Society, 250 Williams Street, Atlanta,
805 GA 30303, USA

806 ¹²SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA 98109,
807 USA

808 ¹³Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical
809 Innovation and School of Biomedical Sciences, Queensland University of Technology,

810 Brisbane QLD 4059, Australia
811 ¹⁴Translational Research Institute, Brisbane, Queensland 4102, Australia
812 ¹⁵Australian Prostate Cancer Research Centre-Qld, Queensland University of Technology,
813 Brisbane; Prostate Cancer Research Program, Monash University, Melbourne; Dame Roma
814 Mitchell Cancer Centre, University of Adelaide, Adelaide; Chris O'Brien Lifehouse and
815 ¹⁶Department of Applied Health Research, University College London, London, WC1E 7HB,
816 UK
817 ¹⁷Centre for Cancer Genetic Epidemiology, Department of Oncology, University of
818 Cambridge, Strangeways Laboratory, Worts Causeway, Cambridge, CB1 8RN, UK
819 ¹⁸Institute of Biomedicine, University of Turku, Finland
820 ¹⁹Department of Medical Genetics, Genomics, Laboratory Division, Turku University
821 Hospital, PO Box 52, 20521 Turku, Finland
822 ²⁰Department of Surgical Sciences, Uppsala University, 75185 Uppsala, Sweden
823 ²¹Division of Cancer Sciences, University of Manchester, Manchester Academic Health
824 Science Centre, Radiotherapy Related Research, The Christie Hospital NHS Foundation Trust,
825 Manchester, M13 9PL UK
826 ²²Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA
827 02115, USA
828 ²³CeRePP, Tenon Hospital, F-75020 Paris, France.
829 ²⁴Sorbonne Universite, GRC n°5 , AP-HP, Tenon Hospital, 4 rue de la Chine, F-75020 Paris,
830 France
831 ²⁵Department of Molecular Medicine, Aarhus University Hospital, Palle Juul-Jensen
832 Boulevard 99, 8200 Aarhus N, Denmark
833 ²⁶Department of Clinical Medicine, Aarhus University, DK-8200 Aarhus N
834 ²⁷Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway
835 ²⁸Nuffield Department of Surgical Sciences, University of Oxford, Room 6603, Level 6, John
836 Radcliffe Hospital, Headley Way, Headington, Oxford, OX3 9DU, UK
837 ²⁹University of Cambridge, Department of Oncology, Box 279, Addenbrooke's Hospital, Hills
838 Road, Cambridge CB2 0QQ, UK
839 ³⁰Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, CB2
840 ORE, UK
841 ³¹Nuffield Department of Surgical Sciences, University of Oxford, Oxford, OX1 2JD, UK
842 ³²Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK
843 ³³Population Health Sciences, Bristol Medical School, University of Bristol, BS8 2PS, UK
844 ³⁴Cancer Epidemiology Unit, Nuffield Department of Population Health, University of
845 Oxford, Oxford, OX3 7LF, UK
846 ³⁵Dept. of Surgical Oncology, Princess Margaret Cancer Centre, Toronto ON M5G 2M9,
847 Canada
848 ³⁶Dept. of Surgery (Urology), University of Toronto, Canada
849 ³⁷Department of Preventive Medicine, Keck School of Medicine, University of Southern
850 California/Norris Comprehensive Cancer Center, Los Angeles, CA 90015, USA
851 ³⁸Department of Radiation Oncology and Department of Genetics and Genomic Sciences,
852 Box 1236, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York,
853 NY 10029, USA
854 ³⁹Centre for Cancer Biomarker and Biotherapeutics, Barts Cancer Institute, Queen Mary
855 University of London, John Vane Science Centre, Charterhouse Square, London, EC1M 6BQ,
856 UK

857 ⁴⁰Cancer Epidemiology Division, Cancer Council Victoria, 615 St Kilda Road, Melbourne, VIC
858 3004, Australia
859 ⁴¹Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global
860 Health, The University of Melbourne, Grattan Street, Parkville, VIC 3010, Australia
861 ⁴²Precision Medicine, School of Clinical Sciences at Monash Health, Monash University,
862 Clayton, Victoria 3168, Australia
863 ⁴³Division of Urologic Surgery, Brigham and Womens Hospital, 75 Francis Street, Boston, MA
864 02115, USA
865 ⁴⁴Fundación Pública Galega Medicina Xenómica, Santiago de Compostela, 15706, Spain.
866 ⁴⁵Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago De Compostela,
867 15706, Spain.
868 ⁴⁶Centro de Investigación en Red de Enfermedades Raras (CIBERER), Spain
869 ⁴⁷ISGlobal, Barcelona, Spain
870 ⁴⁸IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
871 ⁴⁹Universitat Pompeu Fabra (UPF), Barcelona, Spain
872 ⁵⁰CIBER Epidemiología y Salud Pública (CIBERESP), 28029 Madrid, Spain
873 ⁵¹Channing Division of Network Medicine, Department of Medicine, Brigham and Women's
874 Hospital/Harvard Medical School, Boston, MA 02115, USA
875 ⁵²Department of Cancer Epidemiology, Moffitt Cancer Center, 12902 Magnolia Drive,
876 Tampa, FL 33612, USA
877 ⁵³Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle,
878 Washington, 98109-1024, USA
879 ⁵⁴Department of Epidemiology, School of Public Health, University of Washington, Seattle,
880 Washington 98195, USA
881 ⁵⁵International Hereditary Cancer Center, Department of Genetics and Pathology,
882 Pomeranian Medical University, 70-115 Szczecin, Poland
883 ⁵⁶Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen,
884 Denmark
885 ⁵⁷Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University
886 Hospital, Herlev, 2200 Copenhagen, Denmark
887 ⁵⁸Division of Clinical Epidemiology and Aging Research, German Cancer Research Center
888 (DKFZ), D-69120, Heidelberg, Germany
889 ⁵⁹German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), D-69120
890 Heidelberg, Germany
891 ⁶⁰Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National
892 Center for Tumor Diseases (NCT), Im Neuenheimer Feld 460, 69120 Heidelberg, Germany
893 ⁶¹Humangenetik Tuebingen, Paul-Ehrlich-Str 23, D-72076 Tuebingen, Germany
894 ⁶²The University of Texas M. D. Anderson Cancer Center, Department of Genitourinary
895 Medical Oncology, 1515 Holcombe Blvd., Houston, TX 77030, USA
896 ⁶³Departments of Epidemiology & Population Health and of Medicine, Division of Oncology,
897 Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94304 USA
898 ⁶⁴Department of Genetics, Portuguese Oncology Institute of Porto (IPO-Porto), 4200-072
899 Porto, Portugal
900 ⁶⁵Biomedical Sciences Institute (ICBAS), University of Porto, 4050-313 Porto, Portugal
901 ⁶⁶Cancer Genetics Group, IPO-Porto Research Center (CI-IPOP), Portuguese Oncology
902 Institute of Porto (IPO-Porto), 4200-072 Porto, Portugal
903 ⁶⁷Department of Population Sciences, Beckman Research Institute of the City of Hope, 1500

- 904 East Duarte Road, Duarte, CA 91010, 626-256-HOPE (4673)
905 ⁶⁸Ghent University, Faculty of Medicine and Health Sciences, Basic Medical Sciences,
906 Proeftuinstraat 86, B-9000 Gent
907 ⁶⁹Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur,
908 Malaysia
909 ⁷⁰Department of Urology, University of Washington, 1959 NE Pacific Street, Box 356510,
910 Seattle, WA 98195, USA
911 ⁷¹Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, D-20246
912 Hamburg, Germany
913 ⁷²Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical
914 University of Sofia, Sofia, 2 Zdrave Str., 1431 Sofia, Bulgaria
915 ⁷³Department of Oncology, Cross Cancer Institute, University of Alberta, 11560 University
916 Avenue, Edmonton, Alberta, Canada T6G 1Z2
917 ⁷⁴Division of Radiation Oncology, Cross Cancer Institute, 11560 University Avenue,
918 Edmonton, Alberta, Canada T6G 1Z2
919 ⁷⁵Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, KU
920 Leuven, BE-3000, Belgium
921 ⁷⁶Division of Cancer Sciences, Manchester Cancer Research Centre, Faculty of Biology,
922 Medicine and Health, Manchester Academic Health Science Centre, NIHR Manchester
923 Biomedical Research Centre, Health Innovation Manchester, Univeristy of Manchester, M13
924 9WL
925 ⁷⁷The University of Surrey, Guildford, Surrey, GU2 7XH, UK
926 ⁷⁸Genetic Oncology Unit, CHUVI Hospital, Complejo Hospitalario Universitario de Vigo,
927 Instituto de Investigación Biomédica Galicia Sur (IISGS), 36204, Vigo (Pontevedra), Spain
928 ⁷⁹Department of Urology, Erasmus University Medical Center, 3015 CE Rotterdam, The
929 Netherlands
930 ⁸⁰"Exposome and Heredity", CESP (UMR 1018), Faculté de Médecine, Université Paris-Saclay,
931 Inserm, Gustave Roussy, Villejuif
932 ⁸¹Clinical Gerontology Unit, University of Cambridge, Cambridge, CB2 2QQ, UK
933 ⁸²Division of Epidemiology, Department of Internal Medicine, University of Utah School of
934 Medicine, Salt Lake City, Utah 84132, USA
935 ⁸³George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, Utah
936 84148, USA
937 ⁸⁴Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905,
938 USA
939 ⁸⁵Nuffield Department of Population Health, University of Oxford, United Kingdom
940 ⁸⁶Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology,
941 Harvard School of Public Health, Boston, MA, USA
942 ⁸⁷Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center,
943 2525 West End Avenue, Suite 800, Nashville, TN 37232 USA.
944 ⁸⁸International Epidemiology Institute, Rockville, MD 20850, USA
945 ⁸⁹Department of Epidemiology and Biostatistics, School of Public Health, Imperial College
946 London, SW7 2AZ, UK
947

948 **CRUK and PRACTICAL consortium Funding Acknowledgements:**

949 This work was supported by the Canadian Institutes of Health Research, European
950 Commission's Seventh Framework Programme grant agreement n° 223175 (HEALTH-F2-

951 2009-223175), Cancer Research UK Grants C5047/A7357, C1287/A10118, C1287/A16563,
952 C5047/A3354, C5047/A10692, C16913/A6135, and The National Institute of Health (NIH)
953 Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01 (the GAME-ON
954 initiative).

955

956 We would also like to thank the following for funding support: The Institute of Cancer
957 Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate
958 Research Campaign UK (now PCUK), The Orchid Cancer Appeal, Rosetrees Trust, The
959 National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK.
960 We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The
961 Institute of Cancer Research, The Royal Marsden NHS Foundation Trust, and Manchester
962 NIHR Biomedical Research Centre. The Prostate Cancer Program of Cancer Council Victoria
963 also acknowledge grant support from The National Health and Medical Research Council,
964 Australia (126402, 209057, 251533, , 396414, 450104, 504700, 504702, 504715, 623204,
965 940394, 614296,), VicHealth, Cancer Council Victoria, The Prostate Cancer Foundation of
966 Australia, The Whitten Foundation, PricewaterhouseCoopers, and Tattersall's. EAO, DMK,
967 and EMK acknowledge the Intramural Program of the National Human Genome Research
968 Institute for their support.

969

970 Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) [U19
971 CA 148537 for ELucidating Loci Involved in Prostate cancer SuscEptibility (ELLIPSE) project
972 and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract
973 number HHSN268201200008I]. Additional analytic support was provided by NIH NCI U01
974 CA188392 (PI: Schumacher).

975

976 Research reported in this publication also received support from the National Cancer
977 Institute of the National Institutes of Health under Award Numbers U10 CA37429 (CD
978 Blanke), and UM1 CA182883 (CM Tangen/IM Thompson). The content is solely the
979 responsibility of the authors and does not necessarily represent the official views of the
980 National Institutes of Health.

981

982 Funding for the iCOGS infrastructure came from: the European Community's Seventh
983 Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175)
984 (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174,
985 C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National
986 Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19
987 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence
988 (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team
989 in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer
990 Research Foundation, and the Ovarian Cancer Research Fund.

SUPPLEMENTARY METHODS

Cancer datasets

UK Biobank

The UK Biobank is a large cohort study including around 500,000 individuals aged 40 to 69 years at the time of recruitment (2006–2010). The cohort has been described in detail in previous publications^{1,2}. In short, all participants provided written informed consent, after which baseline data were collected using sociodemographic, lifestyle and health-related questionnaires, physical and cognitive assessments, and biological samples. Participants' data were linked to their health records for longitudinal follow-up. The study obtained ethical approval from the National Information Governance Board for Health and Social Care and the North-West Multicenter Research Ethics Committee (Ref: 11/NW/0382).

Cancer cases (diagnosed prior or after enrolment) were obtained from the UK Cancer Registry (updated to April 2019). They were then coded according to the ninth and tenth editions of the International Classification of Diseases (ICD-9 and ICD-10, respectively) as follows: breast (ICD-9: 174; ICD-10: C50), ovarian (ICD-9: 183; ICD-10: C56), prostate (ICD-9: 185; ICD-10: C61), lung (ICD-9: 162; ICD-10: C34) and colorectal cancer (ICD-9: 153; ICD-10: C18-C20). Controls excluded individuals with any type of cancer (self-reported and/or recorded in cancer registry), as well as those with benign, in situ, uncertain or unspecified behaviour neoplasms (ICD-9: 210-239; ICD-10: D00-D49).

Sample-level quality control (QC) involved removing any individuals who had non-white British genetic ancestry, sex chromosome aneuploidies, who withdrew consent from the UK Biobank study and who were closely related to other participants. Variant-level QC consisted in imputing SNPs using the Haplotype Reference Consortium (HRC) and restricting SNPs to a minor allele frequency (MAF) >0.1%, a genotyping rate > 0.015 and a Hardy-Weinberg Equilibrium (HWE) $p > 1 \times 10^{-4}$. LD pruning was performed to an r^2 cutoff of 0.1 using PLINK v2³. In order to reduce false positive signals, SNPs were removed when MAF was below our expectations (we would expect at least 25 minor alleles in cases), as recommended in <http://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas>.

The GWAS analysis in the UK Biobank consisted of 13,879 cases and 198,523 controls for breast cancer, 1,218 cases and 198,523 controls for ovarian cancer, 9,132 cases and 173,493 controls for prostate cancer, 2,671 cases and 372,016 controls for lung cancer and 5,657 cases and 372,016 controls for colorectal cancer. It was performed using BOLT-LMM v2.3.5⁴, adjusting for sex and genotyping chip. BOLT-LMM uses a linear mixed model to account for population stratification and cryptic relatedness in the UK Biobank. Lung cancer associations were estimated twice, once adjusting for genotyping chip and once without. For sex-specific cancers, analyses were limited to individuals of the pertinent sex (only females were used for breast and ovarian cancers, whereas only males were used for prostate cancer). Beta coefficients and their corresponding standard errors were finally transformed to log odds ratios (ORs)⁵.

We also performed a GWAS analysis of parental history of cancer reported by UK Biobank participants (i.e., breast, prostate, lung and bowel cancer) using BOLT-LMM software v2.3.5⁴. Age and sex were included as covariates in the model as before. For sex-specific cancers, analyses were restricted to individuals of the relevant sex (i.e., maternal history only for breast cancer and paternal history only for prostate cancer). We obtained 35,356 breast cancer cases and 206,992 controls, in addition to 31,527 prostate cancer cases and 160,579 controls. For other cancers, we combined maternal and paternal history of cancer, thus obtaining a total of 51,073 lung cancer cases and 404,606 controls, as well as 45,213 bowel cancer cases and 412,429 controls. GWAS of these outcomes have previously provided strong concordance with those based on hospital records⁶. They have also provided consistent results in MR⁷.

FinnGen

The FinnGen R5 release includes data on 218,792 individuals of Finnish ancestry, obtained from Finnish biobanks and digital health registry records⁸. Complete study details are available elsewhere (<https://www.finngen.fi/en>). In brief, samples were excluded for the following reasons: ambiguous gender, genotype missingness >5%, heterozygosity ± 4 s.d. and non-Finnish ancestry. SNPs were genotyped using Illumina and Affymetrix arrays. Variants were excluded for the following reasons: missingness >2%, HWE $p < 1 \times 10^{-6}$ and minor allele count <3. Genotypes were imputed using the Finnish SISu v3 reference panel. The GWAS analysis was conducted using SAIGE v0.36.3.2, a mixed model logistic regression R/C++ package. Sex, age, genotyping batch and the first 10 genetically derived principal components were included as covariates in the analysis. We used FinnGen R5 release data on breast (8,401 cases and 99,321 controls), ovarian (719 cases and 99,321 controls), prostate (6,311 cases and 74,685 controls), lung (1,681 cases and 173,933 controls) and colorectal cancer (3,022 cases and 174,006 controls). We used the “EXALLC” cancer variables, which excluded other cancers from controls.

Breast Cancer Association Consortium

The GWAS summary data for breast cancer were obtained from a Breast Cancer Association Consortium (BCAC) meta-analysis performed by Michailidou et al.⁹. This included 122,977 cases and 105,974 controls (69,501 cases of ER+ and 21,468 of ER- breast cancer). All studies that contributed to this meta-analysis have been fully detailed in previous publications⁹⁻¹¹. In sum, samples were excluded if they had a low call rate (<95%), abnormally high or low heterozygosity (4.89 s.d. from the mean), <80% European ancestry, probable duplicates and/or close relatives within and across studies. Genetic variants were genotyped using the Illumina OncoArray and iCOGS arrays and genotypes were imputed using the 1000 Genomes Project Phase 3 reference panel.

We also obtained summary data for breast cancer subtypes from a BCAC GWAS meta-analysis by Zhang et al.¹². The study comprised data on luminal A-like (7,325 cases), luminal B-like (1,682 cases), luminal B/HER2-negative-like (1,779 cases), HER2-enriched-like (718 cases) and triple-negative (2,006 cases) invasive breast cancer subtypes and 20,815 controls. The details of the study can be found in the publication. In brief, the analyses excluded cases of carcinoma in situ, cases missing data on tumour characteristics and cases for which there

were no controls available in their respective countries. Participants were also excluded if age at diagnosis/enrolment was missing. Genotypes were obtained using OncoArray and iCOGS arrays. Imputation was performed using the 1000 Genomes Project Phase 3 reference panel. OncoArray and iCOGS datasets were analysed separately and pooled using fixed-effect meta-analysis.

Ovarian Cancer Association Consortium

We used ovarian cancer genetic summary statistics from an Ovarian Cancer Association Consortium (OCAC) study by Phelan et al.¹³. This comprised 25,509 cases and 40,941 controls. Subtypes included high grade serous (13,037 cases), low grade serous (1,012 cases), invasive mucinous (1,417 cases), clear cell (1,366 cases) and endometrioid (2,810 cases) ovarian cancers. This study combined genotype data from OCAC and Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA) genotyping projects. These have been fully described in the publication. In short, samples with >27% non-European ancestry were excluded, as were those with a genotyping call rate <95%, excessively low or high heterozygosity. Non-females and duplicates were also removed. SNPs were genotyped using several Illumina arrays (OncoArray, iSelect iCOGS, 550k, HumanOmni 2.5M, 610 Quad and 317k). Imputations were performed separately for each genotyping project using the 1000 Genomes Project v3 reference panel.

Consortium of Investigators of Modifiers of BRCA1/2

We also used CIMBA GWAS data for breast and ovarian cancers in BRCA1 and BRCA2 mutation carriers^{13 14}. The genotyping and imputation procedures that were used have been described elsewhere^{13 14}. In brief, samples were excluded if they were non-female, had discordant genotypes in known sample duplicates, had >19% non-European ancestry, a genotyping call rate <95% or extremely low or high heterozygosity ($p < 1 \times 10^{-6}$). SNPs were genotyped using Illumina's Oncoarray and iSelect Collaborative Oncological Gene-Environment Study (iCOGS) arrays. Imputation was performed using the 1000 Genomes Project Phase 3 reference panel.

Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome

Prostate cancer GWAS summary data were acquired from a Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) study by Schumacher et al.¹⁵. This included 79,148 cases and 61,106 controls. It also comprised data on prostate cancer subtypes: 15,167 advanced cases vs. 58,308 healthy controls; 14,160 advanced cases vs. 62,421 non-advanced controls; 6,988 early-onset cases (age at diagnosis ≤ 55 years) vs. 44,256 healthy controls; 15,561 high aggressive cases vs. 9,739 low aggressive controls; and 20,658 high aggressive cases vs. 38,093 low/intermediate aggressive controls.

Prostate cancer aggressiveness was defined as follows:

- Low aggressive: tumor stage $\leq T1$ **AND** Gleason score ≤ 6 **AND** prostate-specific antigen (PSA) < 10 ng/mL.
- Intermediate aggressive: tumor stage T2 **OR** Gleason score = 7 **OR** PSA 10–20 ng/mL.

- High aggressive: tumor stage T3/T4, N1 or M1 **OR** Gleason score ≥ 8 **OR** PSA >20 ng/mL.
- Advanced: metastatic disease **OR** Gleason score ≥ 8 **OR** PSA >100 ng/mL **OR** death due to prostate cancer.

Study details are available in the publication. In brief, individuals were excluded if they presented a call rate <95%, extreme heterozygosity (>4.9 s.d. from the mean), if they were duplicates or if they were related to other participants. Only men of European ancestry (>80%) were included in the GWAS. Studies were genotyped using Illumina (OncoArray, Human 610, 60k, Infinium HumanHap 550, iSELECT, iCOGS and Human Omni 2.5) and Affymetrix GeneChip (500k and 5.0k) genotyping arrays and SNPs were imputed to the 1000 Genomes Project Phase 3 reference panel.

International Lung Cancer Consortium

For lung cancer, we used genetic summary data obtained from an International Lung Cancer Consortium (ILCCO) GWAS meta-analysis of 11,348 cases and 15,861 controls by Wang et al.¹⁶. We also used lung cancer subtype data including 3,275 squamous cell lung carcinoma cases and 15,038 controls, as well as 3,442 lung adenocarcinoma cases and 14,894 controls. Individual studies included in the meta-analysis have been explained in prior publications¹⁷⁻²⁰. In summary, sample QC consisted in excluding any individuals of non-European ancestry, with low call rates (<90%), or abnormally high or low heterozygosity ($p < 1 \times 10^{-4}$). Duplicates and closely related individuals were also removed. Genotyping was performed using Illumina HumanHap 317k, 317k+240S, 370Duo, 550k, 610k or 1M arrays. SNPs were imputed from the 1000 Genomes Project Phase 1 v3 reference panel.

Genetics and Epidemiology of Colorectal Cancer Consortium

Colorectal cancer GWAS summary statistics were retrieved from a Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) GWAS meta-analysis by Huyghe et al.²¹. This comprised 58,131 cases (31,288 male and 26,843 female) and 67,347 controls (34,527 male and 32,820 female). Cases were defined as patients with colorectal cancer or advanced adenoma.

Data on colorectal cancer subtypes were obtained from another GECCO publication by Huyghe et al.²². This included 48,214 cases and 64,159 controls (32,002 colon, 15,706 proximal colon, 14,376 distal colon and 16,212 rectal cancer cases).

Colorectal cancer subtypes were defined as follows:

- Proximal colon cancer: any primary tumour starting in the cecum, ascending colon, hepatic flexure, or transverse colon (ICD-9: 153.4, 153.6, 153.0, or 153.1, respectively).
- Distal colon cancer: any primary tumour starting in the splenic flexure, descending colon, or sigmoid colon (ICD-9 codes: 153.7, 153.2, or 153.3, respectively)
- Colon cancer: proximal and distal colon cancer cases, in addition to colon cancer cases with unspecified site.

- Rectal cancer: any primary tumour starting in the rectum or rectosigmoid junction (ICD-9 codes: 154.1, or 154.0, respectively)

Controls excluded individuals with known history of cancer or reported family history of colorectal cancer. QC procedures have been explained in the publications^{21,22}. In brief, the studies excluded samples with evidence of DNA contamination, high missing genotype rates, unintentional duplicate pairs and sex discrepancies. Closely related individuals and those of non-European ancestry were also excluded. Genotyping was conducted using Illumina (300k, Oncoarray, 1M, 550k, 610k, OmniExpress, OmniExpressExome, 300/240S and custom iSelect) and Affymetrix (Axiom and 500k) arrays. Imputation was performed to the HRC reference panel.

Sensitivity analyses

MR-Egger assumes that the association between SNPs and epigenetic age acceleration is not correlated with SNPs that affect cancer via pleiotropic pathways (Instrument Strength Independent of Direct Effect—InSIDE assumption)²³. The weighted median method assumes that at least half of the SNPs in the analysis are valid instruments. The weighted mode approach presupposes that the most frequent association estimate is not affected by pleiotropy, meaning it must correspond to the true causal effect (Zero Modal Pleiotropy Assumption—ZEMPA)²⁴.

Data availability

Summary statistics for epigenetic age acceleration measures of HannumAge, Intrinsic HorvathAge, PhenoAge and GrimAge were downloaded from: <https://datashare.ed.ac.uk/handle/10283/3645>. Summary statistics for international cancer genetic consortiums were obtained from their respective data repositories. Colorectal cancer data were obtained following the submission of a written request to the GECCO committee, which may be contacted by email at kafdem@fredhutch.org/upeters@fredhutch.org. Breast, ovarian, prostate and lung cancer data were accessed via MR-Base (<http://app.mrbase.org/>), which holds complete GWAS summary data from BCAC, OCAC, PRACTICAL and ILCCO. Breast cancer subtype data were obtained from BCAC and can be downloaded from: <http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/oncoarray-and-combined-summary-result/gwas-summary-associations-breast-cancer-risk-2020/>. Data on breast and ovarian cancer in BRCA1 and BRCA2 carriers were obtained from CIMBA and can be downloaded from: <http://cimba.ccge.medschl.cam.ac.uk/oncoarray-complete-summary-results/>. Prostate cancer subtype data are not publicly available through MR-Base but can be accessed upon request. These data are managed by the PRACTICAL committee, which may be contacted by email at practical@icr.ac.uk. FinnGen data is publicly available and can be accessed here: https://www.finngen.fi/en/access_results. UK Biobank data can be accessed through the MR-Base platform. Parental history of cancer data were obtained from the UK Biobank study under application #15825 and can be accessed via an approved application to the UK Biobank (<https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access>).

Code availability

The GWAS analysis for cancers in UK Biobank was performed using BOLT-LMM v2.3.5 (<http://data.broadinstitute.org/alkesgroup/BOLT-LMM/>). All MR analyses and visualisations were conducted using R software v4.0.2 (<https://www.r-project.org/>). For cancer datasets that were not obtained from the MR-Base platform, LD proxies were identified using the “LDlinkR” v1.1.2 R package (<https://github.com/CBIIT/LDlinkR>). Two-sample MR analyses were conducted using the “TwoSampleMR” v0.5.6 R package (<https://github.com/MRCIEU/TwoSampleMR>). Meta-analyses were performed using the “meta” v4.18 R package (<https://github.com/guido-s/meta/>). MR-CAUSE analyses were conducted using the “cause” v1.2.0 R package (<https://github.com/jean997/cause>). Plots were created using the “ggforestplot” v0.1.0 R package (<https://github.com/nightingalehealth/ggforestplot>). LD scores were computed using the “ldsc” v1.0.1 command line tool (<https://github.com/bulik/ldsc>). The code used in this study is available at: https://github.com/fernandam93/epiclocks_cancer.

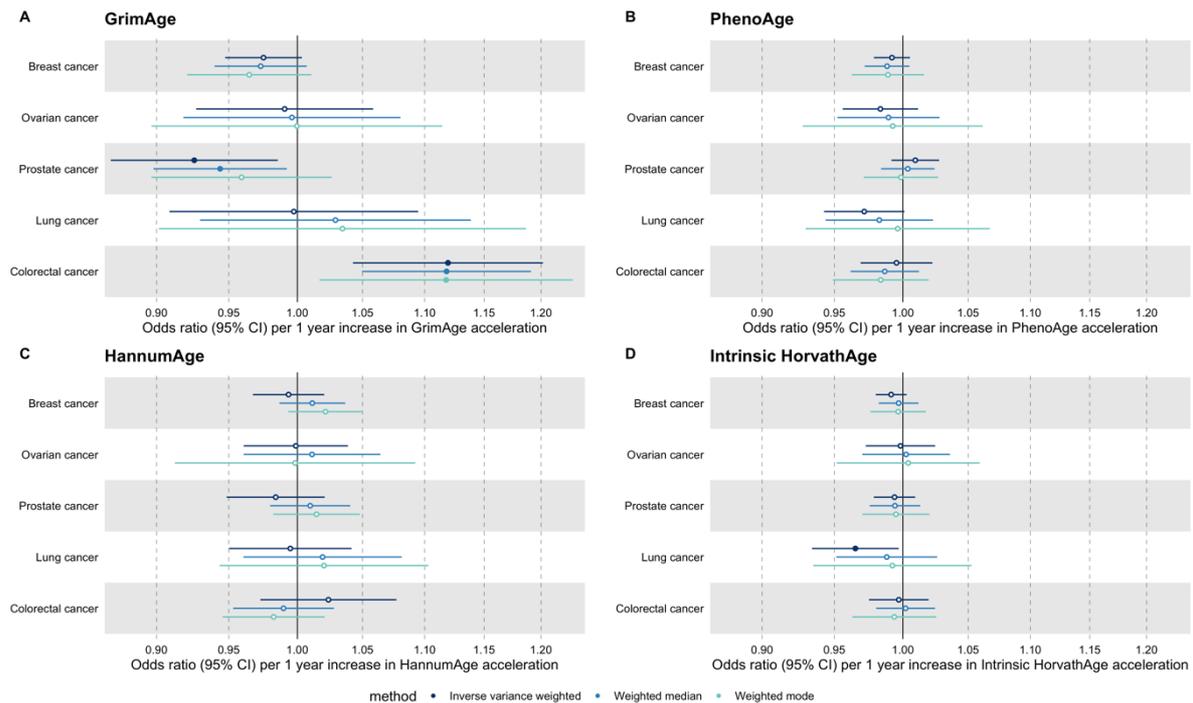
REFERENCES

1. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12(3):e1001779. doi: 10.1371/journal.pmed.1001779 [published Online First: 2015/04/01]
2. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562(7726):203-09. doi: 10.1038/s41586-018-0579-z [published Online First: 2018/10/12]
3. Mitchell R, Hemani G, Dudding T, et al. UK Biobank Genetic Data: MRC-IEU Quality Control, version 2. 2019 doi: <https://doi.org/10.5523/bris.1ovaau5sxunp2cv8rcy88688v>
4. Loh PR, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015;47(3):284-90. doi: 10.1038/ng.3190 [published Online First: 2015/02/03]
5. Elsworth BL, Mitchell R, Raistrick CA, et al. MRC IEU UK Biobank GWAS pipeline version 2. 2019 doi: <https://doi.org/10.5523/bris.pnoat8cxo0u52p6ynfaekeigi>
6. DeBoever C, Tanigawa Y, Aguirre M, et al. Assessing Digital Phenotyping to Enhance Genetic Studies of Human Diseases. *Am J Hum Genet* 2020;106(5):611-22. doi: 10.1016/j.ajhg.2020.03.007 [published Online First: 2020/04/11]
7. Richardson TG, Wang Q, Sanderson E, et al. Effects of apolipoprotein B on lifespan and risks of major diseases including type 2 diabetes: a mendelian randomisation analysis using outcomes in first-degree relatives. *The Lancet Healthy Longevity* 2021;2(6):e317-e26. doi: 10.1016/s2666-7568(21)00086-6
8. FinnGen. FinnGen Documentation of R5 Release. 2021 doi: <https://finngen.gitbook.io/documentation/> (accessed on 6 August 2021)
9. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017;551(7678):92-94. doi: 10.1038/nature24284 [published Online First: 2017/10/24]

10. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;45(4):353-61, 61e1-2. doi: 10.1038/ng.2563 [published Online First: 2013/03/29]
11. Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 2015;47(4):373-80. doi: 10.1038/ng.3242 [published Online First: 2015/03/10]
12. Zhang H, Ahearn TU, Lecarpentier J, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet* 2020;52(6):572-81. doi: 10.1038/s41588-020-0609-2 [published Online First: 2020/05/20]
13. Phelan CM, Kuchenbaecker KB, Tyrer JP, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet* 2017;49(5):680-91. doi: 10.1038/ng.3826 [published Online First: 2017/03/28]
14. Milne RL, Kuchenbaecker KB, Michailidou K, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet* 2017;49(12):1767-78. doi: 10.1038/ng.3785 [published Online First: 2017/10/24]
15. Schumacher FR, Al Olama AA, Berndt SI, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* 2018;50(7):928-36. doi: 10.1038/s41588-018-0142-8 [published Online First: 2018/06/13]
16. Wang Y, McKay JD, Rafnar T, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet* 2014;46(7):736-41. doi: 10.1038/ng.3002 [published Online First: 2014/06/02]
17. Timofeeva MN, Hung RJ, Rafnar T, et al. Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet* 2012;21(22):4980-95. doi: 10.1093/hmg/dds334 [published Online First: 2012/08/18]
18. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008;40(5):616-22. doi: 10.1038/ng.109 [published Online First: 2008/04/04]
19. Wang Y, Broderick P, Webb E, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet* 2008;40(12):1407-9. doi: 10.1038/ng.273 [published Online First: 2008/11/04]
20. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008;452(7187):633-7. doi: 10.1038/nature06885 [published Online First: 2008/04/04]
21. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet* 2019;51(1):76-87. doi: 10.1038/s41588-018-0286-6 [published Online First: 2018/12/05]
22. Huyghe JR, Harrison TA, Bien SA, et al. Genetic architectures of proximal and distal colorectal cancer are partly distinct. *Gut* 2021;70(7):1325-34. doi: 10.1136/gutjnl-2020-321534 [published Online First: 2021/02/27]
23. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44(2):512-25. doi: 10.1093/ije/dyv080 [published Online First: 2015/06/08]

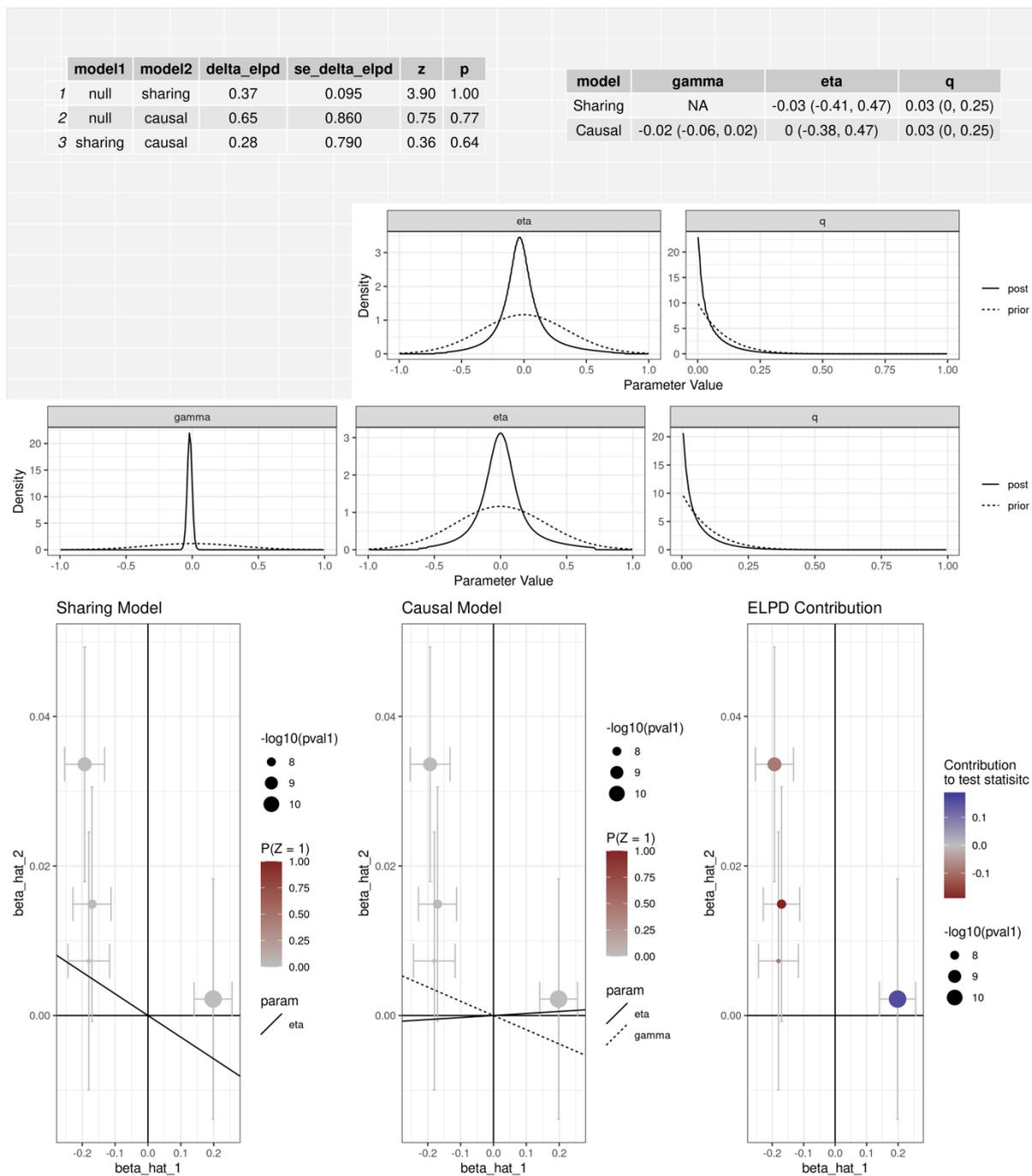
24. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017;46(6):1985-98. doi: 10.1093/ije/dyx102 [published Online First: 2017/10/19]

Supplementary Figures



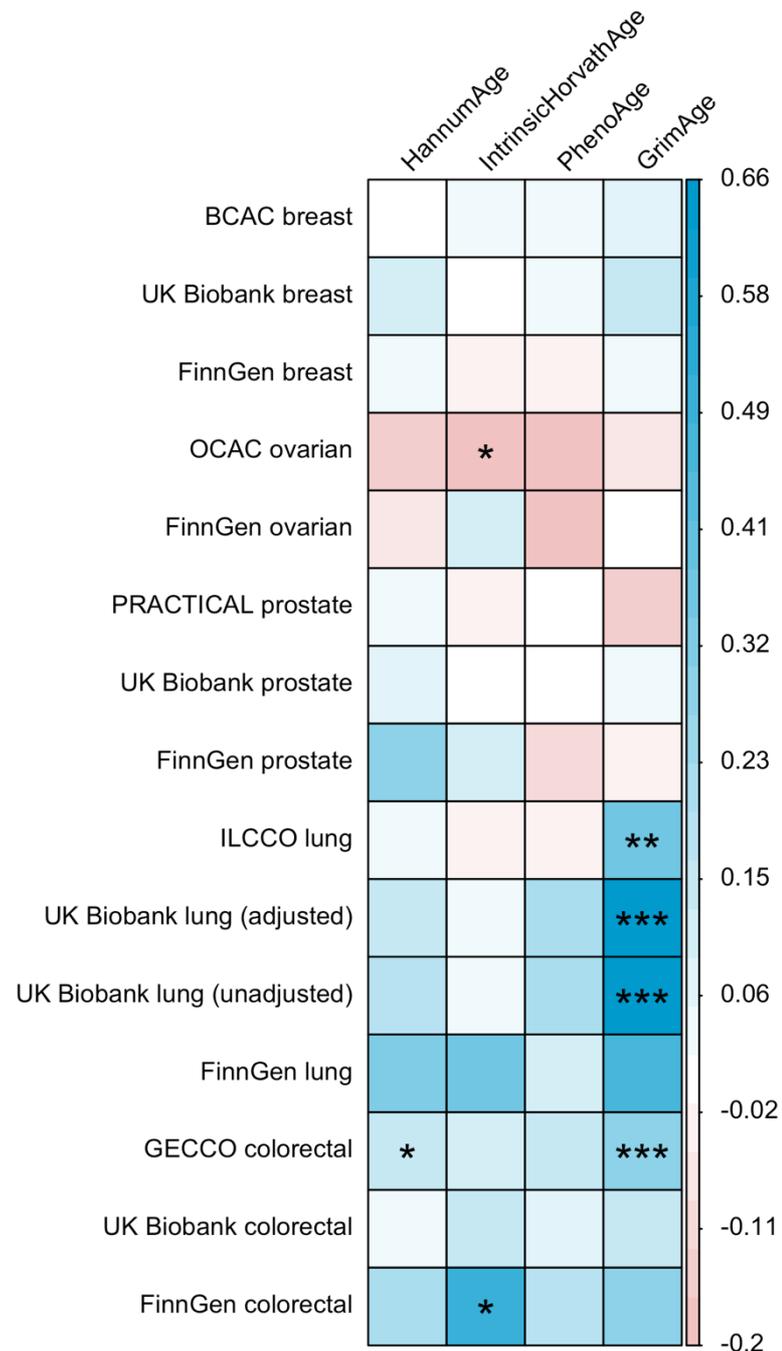
Supplementary Figure 1. Fixed effect meta-analysis of Mendelian randomization estimates for genetically predicted effects of epigenetic age acceleration on multiple cancers.

Odds ratios and 95% confidence intervals are reported per 1 year increase in (A) GrimAge acceleration, (B) PhenoAge acceleration, (C) HannumAge acceleration and (D) Intrinsic HorvathAge acceleration. Results were obtained using inverse variance weighted MR (dark blue), weighted median (sky blue) and weighted mode (turquoise) methods. All meta-analysis estimates were calculated using data from UK Biobank, FinnGen and international consortia, except for colorectal cancer estimates, which exclude UK Biobank data to avoid double counting.



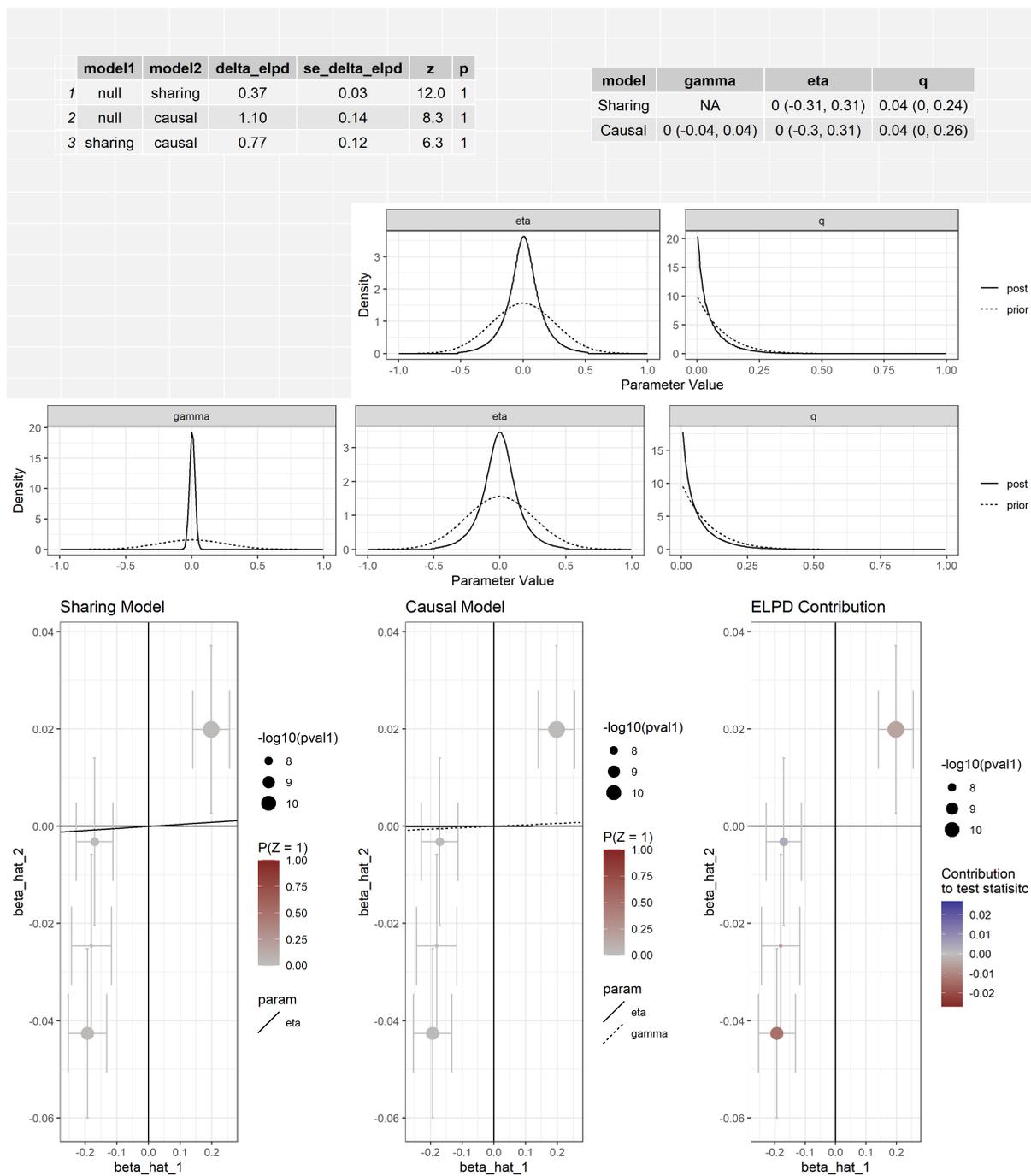
Supplementary Figure 2. CAUSE analysis for the genetically predicted effect of GrimAge acceleration on prostate cancer in PRACTICAL.

CAUSE estimates reported per 1 year increase in GrimAge acceleration. Data source: PRACTICAL, Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome.



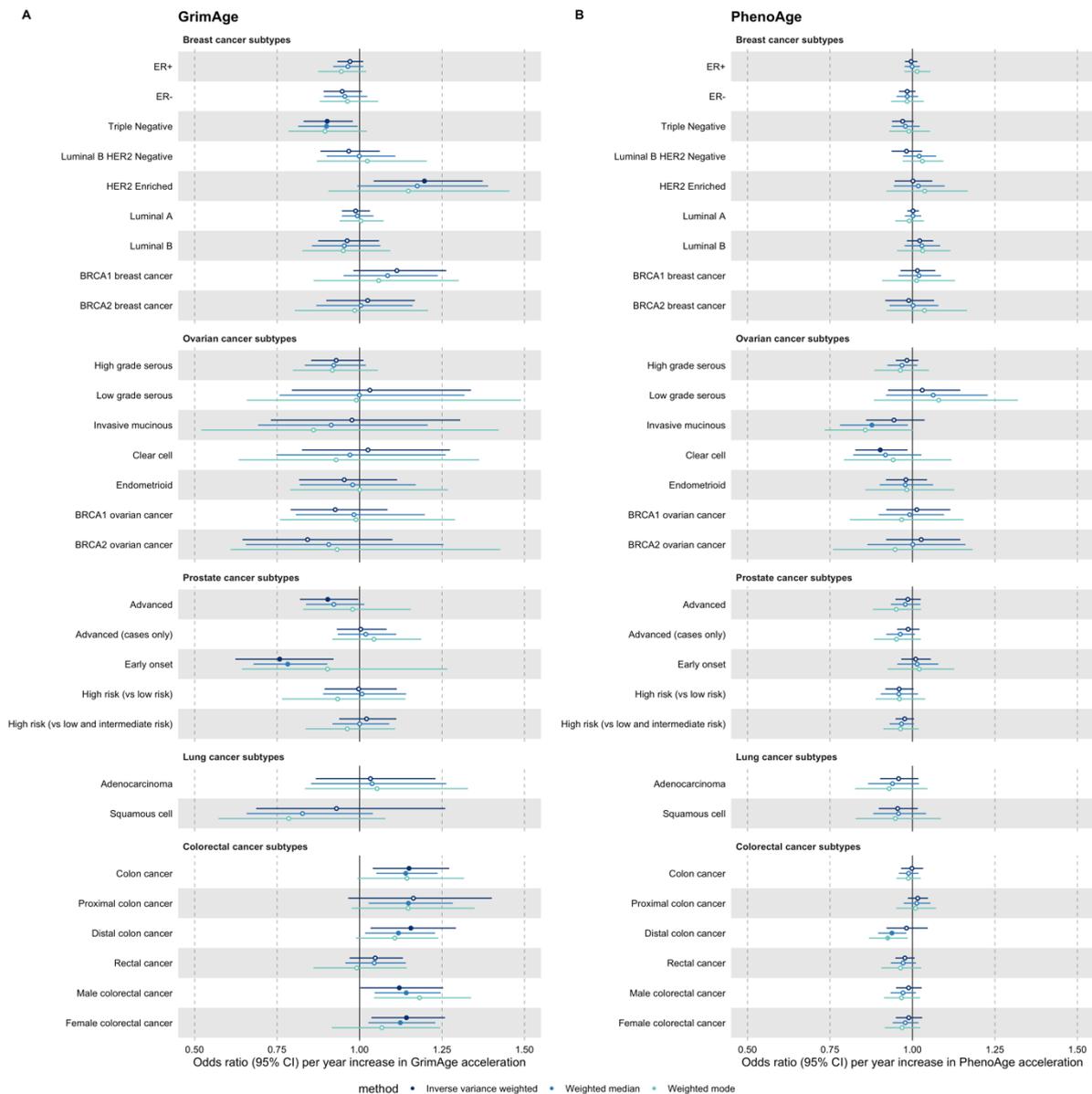
Supplementary Figure 3. Genetic correlation estimates for epigenetic age acceleration and multiple cancers.

Genetic correlation coefficients are reported per 1 year increase in epigenetic age acceleration. Results were obtained using LD Score regression. Abbreviations: BCAC, Breast Cancer Association Consortium; OCAC, Ovarian Cancer Association Consortium; PRACTICAL, Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome; ILCCO, International Lung Cancer Consortium; GECCO, Genetics and Epidemiology of Colorectal Cancer Consortium. For UK Biobank lung cancer results, adjusted means results have been adjusted for genotyping chip and unadjusted means results have not been adjusted for genotyping chip.



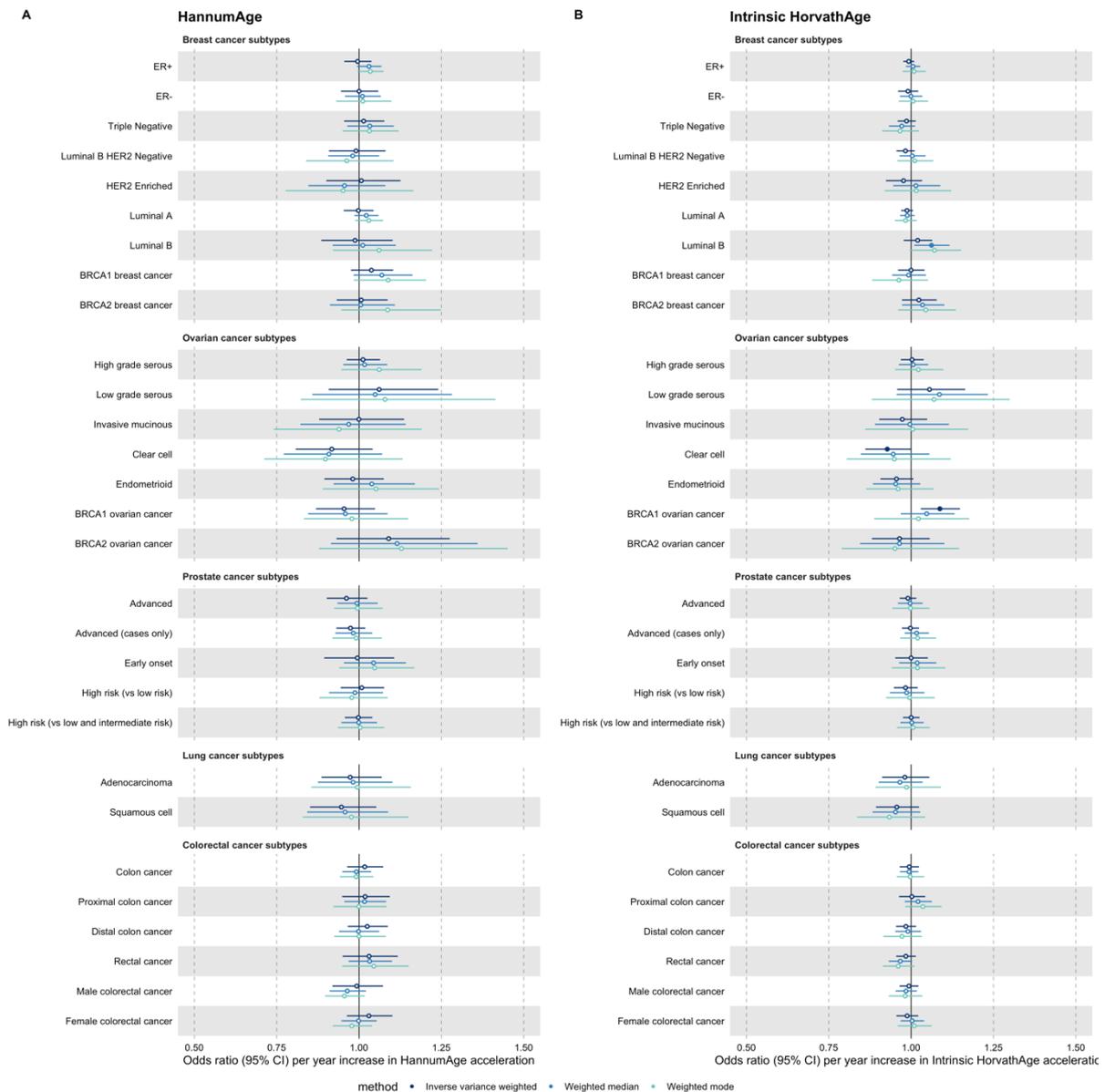
Supplementary Figure 4. CAUSE analysis for the genetically predicted effect of GrimAge acceleration on colorectal cancer in GECCO.

CAUSE estimates reported per 1 year increase in GrimAge acceleration. Data source: GECCO, Genetics and Epidemiology of Colorectal Cancer Consortium.



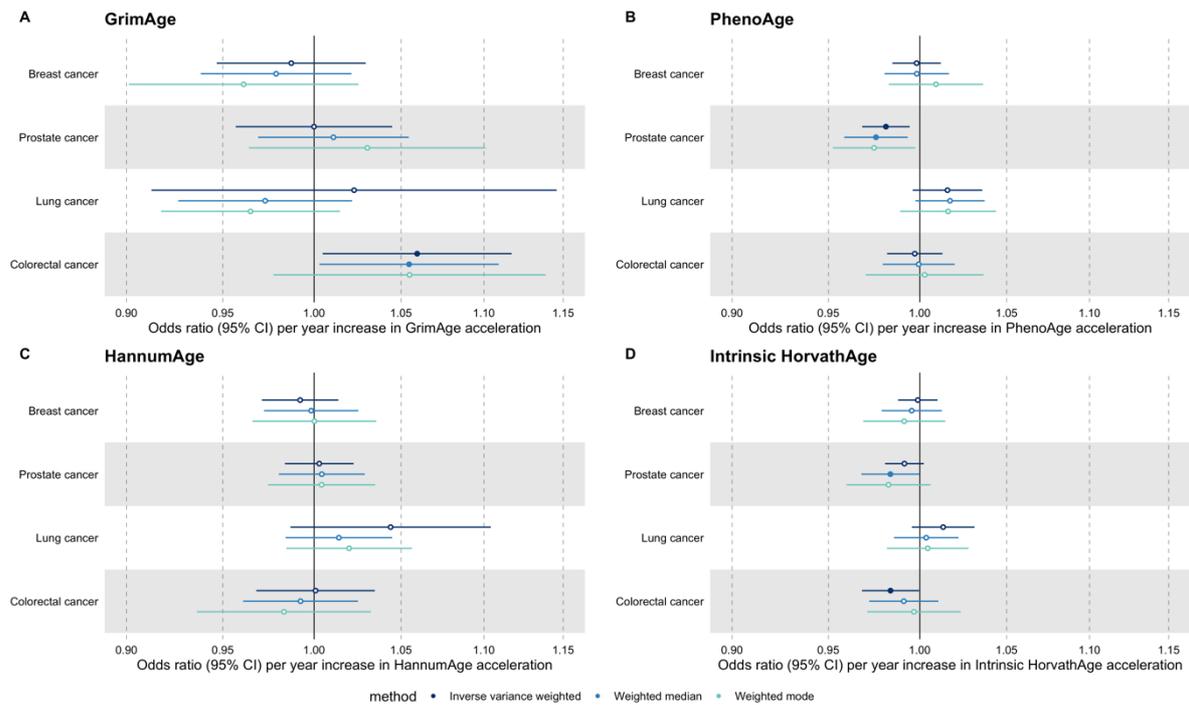
Supplementary Figure 5. Mendelian randomization estimates for genetically predicted effects of GrimAge and PhenoAge acceleration on multiple cancer subtypes.

Odds ratios and 95% confidence intervals are reported per 1 year increase in (A) GrimAge acceleration, (B) PhenoAge acceleration. GrimAge and PhenoAge acceleration were instrumented by four and 11 genetic variants, respectively. Results were obtained using inverse variance weighted MR (dark blue), weighted median (sky blue) and weighted mode (turquoise) methods. Data sources: BCAC, OCAC, CIMBA, PRACTICAL, ILCCO and GECCO.



Supplementary Figure 6. Mendelian randomization estimates for genetically predicted effects of HannumAge and Intrinsic HorvathAge acceleration on multiple cancer subtypes.

Odds ratios and 95% confidence intervals are reported per 1 year increase in (A) HannumAge acceleration, (B) Intrinsic HorvathAge acceleration. HannumAge and Intrinsic HorvathAge acceleration were instrumented by nine and 24 genetic variants, respectively. Results were obtained using inverse variance weighted MR (dark blue), weighted median (sky blue) and weighted mode (turquoise) methods. Data sources: BCAC, OCAC, CIMBA, PRACTICAL, ILCCO and GECCO.



Supplementary Figure 7. Mendelian randomization estimates for genetically predicted effects of epigenetic age acceleration on parental history of multiple cancers.

Odds ratios and 95% confidence intervals are reported per 1 year increase in (A) GrimAge acceleration, (B) PhenoAge acceleration, (C) HannumAge acceleration and (D) Intrinsic HorvathAge acceleration. GrimAge, PhenoAge, HannumAge and Intrinsic HorvathAge acceleration were instrumented by 4, 11, 9 and 24 genetic variants, respectively. Results were obtained using inverse variance weighted MR (dark blue), weighted median (sky blue) and weighted mode (turquoise) methods. Data source: UK Biobank.