

1 A proteogenomic analysis of the adiposity colorectal cancer relationship identifies GREM1 as a
2 probable mediator

3

4 Matthew A Lee^{1,2},
5 Charlie A Hatcher^{2,3},
6 Emma Hazelwood^{2,3},
7 Lucy J Goudswaard^{2,3},
8 Konstantinos K Tsilidis^{8,9}
9 Emma E Vincent^{2,3,10}
10 Richard M Martin^{2,3,11}
11 Karl Smith-Byrne¹²
12 Hermann Brenner^{13,14,15}
13 Iona Cheng¹⁶
14 Sun-Seog Kweon^{17,18}
15 Loic Le Marchand¹⁹
16 Polly A Newcomb^{6,7}
17 Robert E Schoen²⁰
18 Ulrike Peters^{6,7}
19 Marc J Gunter^{1,8},
20 Bethany Van Guelpen^{4,5*}
21 Neil Murphy^{1*}

22

23 1 International Agency for Research on Cancer (IARC/WHO), Nutrition and Metabolism Branch, Lyon, France

24 2 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

25 3 Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK

26 4 Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden

27 5 Wallenberg Centre for Molecular Medicine, Umeå University, Umeå, Sweden

28 6 Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA

29 7 Department of Epidemiology, University of Washington, Seattle, WA, USA

30 8 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

31 9 Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece

32 10 School of Translational Health Sciences, University of Bristol, Bristol, UK

33 11 National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University Hospitals Bristol and
34 Weston NHS Foundation Trust and the University of Bristol, Bristol, UK

35 12 Cancer Epidemiology Unit, University of Oxford, Oxford, UK

36 13 Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg,
37 Germany.

38 14 Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases
39 (NCT), Heidelberg, Germany.

40 15 German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.

41 16 Department of Epidemiology and Biostatistics, University of California-San Francisco, San Francisco, California, United
42 States of America

43 17 Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, Korea.

44 18 Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, Korea.

45 19 University of Hawaii Cancer Center, Honolulu, Hawaii, USA.

46 20 Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA.

47

48

49 "Contributed equally

50

51 *Corresponding author: leem@iarc.who.int

52 International Agency for Research on Cancer

53 World Health Organisation

54 25 avenue Tony Garnier, 69007, Lyon, France

55

56

57 1. Abstract

58 Adiposity is an established risk factor for colorectal cancer (CRC). However, the pathways
59 underlying this relationship, and specifically the role of the circulating proteome, is unclear.

60

61 Utilizing two-sample Mendelian randomization and colocalization, based on summary data from
62 large sex-combined and sex-specific genetic studies, we estimated the univariable (UV)
63 associations between: (I) adiposity measures (body mass index, BMI; waist hip ratio, WHR) and
64 overall and site-specific (colon, proximal colon, distal colon, and rectal) CRC risk, (II) adiposity
65 measures and plasma proteins, and (III) adiposity-associated plasma proteins and CRC risk. We
66 used multivariable MR (MVMR) to investigate the potential mediating role of adiposity- and
67 CRC-related proteins in the adiposity-CRC association.

68

69 BMI and WHR were positively associated with CRC risk, with similar associations by anatomical
70 tumour site. 6,591 adiposity-protein (2,628 unique proteins) and 33 protein-CRC (8 unique
71 proteins) associations were identified using UVMR and colocalization. 1 protein, GREM1 was
72 associated with BMI only and CRC outcomes in a manner that was consistent with a potential
73 mediating role in sex-combined and female-specific analyses. In MVMR, adjusting the BMI-CRC
74 association for GREM1, effect estimates were attenuated - suggestive of a potential mediating role
75 - most strongly for the BMI-overall CRC association in women.

76

77 These results highlight the impact of adiposity on the plasma proteome and of adiposity-
78 associated circulating proteins on the risk of CRC. Supported by evidence from *cis*-SNP UVMR
79 and colocalization analyses, GREM1 was identified as a potential mediator of the BMI-CRC
80 association, particularly in women, and warrants further experimental investigation.

82 2. Introduction

83 Adiposity is an established causal risk factor for the development of colorectal cancer (CRC)¹⁻³.
84 However, the underlying biological pathways are not fully understood. Identifying potentially
85 modifiable mediators of the relationship between adiposity and CRC development could uncover
86 targets for pharmacological and/or lifestyle interventions. Circulating proteins are strong
87 candidate mediators. Evidence from molecular epidemiological and genetic studies has linked
88 adiposity with broad changes in the human circulating proteome, including via effects on glucose
89 and lipid metabolism, and inflammatory and immune markers⁴⁻⁷. Whether these changes to the
90 proteome influence the association between adiposity and CRC risk is unclear.

91

92 Mendelian randomization (MR) uses genetic variants as instrumental variables which, under
93 specific assumptions, can be used to investigate the causal relationship between an exposure and
94 outcome. Given the random allocation of alleles during gametogenesis, across a large enough
95 population, findings of MR analyses are more robust to the effects of confounding and reverse
96 causation than traditional observational studies. Two-step/network MR^{8,9} and Multivariable (MV)
97 MR¹⁰ analyses can be used to investigate the intermediate effects of traits which may reside on the
98 causal mechanistic pathway from exposure to outcome¹¹. To date, MVMR analyses examining the
99 role of the proteome as an intermediate in the relationship between adiposity and CRC have not
100 been undertaken.

101

102 Here, we investigated whether circulating proteins act as intermediates in the association
103 between adiposity (BMI and WHR) and CRC risk. We conducted colocalization and MVMR
104 analyses using summary statistics from genome-wide association studies (GWAS) of adiposity

105 traits, circulating protein levels, and CRC risk, and examined if mediating proteins were expressed

106 in adipose and CRC tissue.

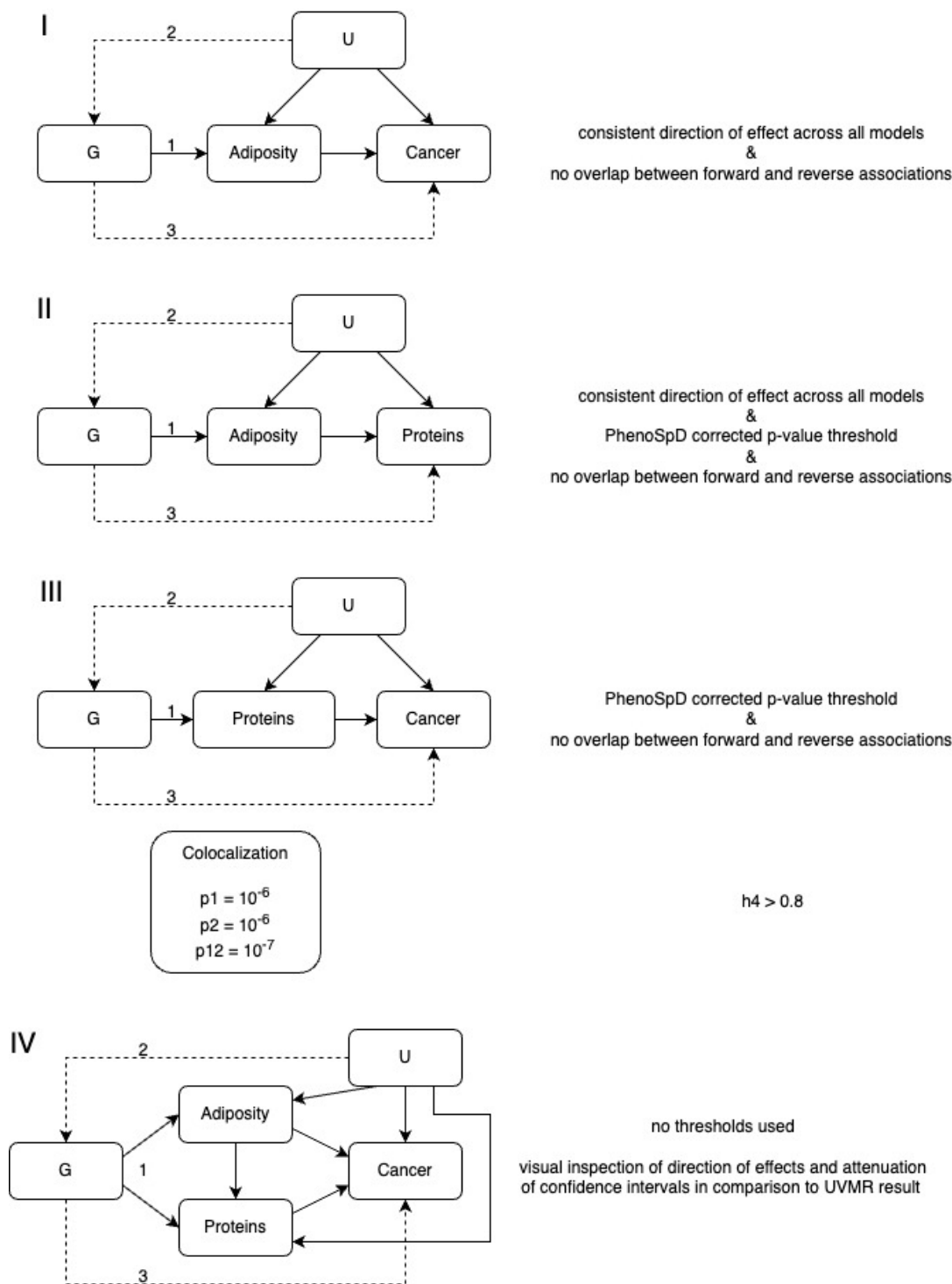
107 3. Methods

108 3.1. Study design

109 Four main analysis steps were performed sequentially (Figure 1) to estimate: (I) the causal
110 relationship between adiposity measures (BMI and WHR) and CRC risk, (II) the causal
111 relationship between adiposity measures and plasma proteins, (III) the causal relationship
112 between proteins and CRC risk, and (IV) the potential mediating effects of adiposity-associated
113 proteins in the adiposity-CRC association. We performed forward and reverse UVMR for steps I-
114 III and used MVMR for step IV. For step III we performed cis-SNP UVMR and colocalization. For
115 all steps, sex-combined and sex-specific analyses were performed. All analyses were performed
116 using R version 4.1.2 and the following packages: TwoSampleMR¹² (version 0.4.22), MVMR¹³
117 (version 0.3), and coloc¹⁴ (version 5.2.0). Forest plots were created using ggforestplot (version
118 0.1.0).

119

Main analyses



120

121

122

123

124

125

Figure 1 Analysis overview. Directed acyclic graph overview of main analyses: I-III; univariable Mendelian randomization analyses, IV; multivariable MR analysis. Text to the right of each analysis gives the requirements for an association. 1-3: MR assumptions; G: genetic variant(s); U: unmeasured confounders; p*: prior probability of a random SNP in the region (1) being (causally) associated with trait 1 and not trait 2, (2) trait 2 and not trait 1, or (12) both traits; h4: probability that there is an association with both traits in the region (shared causal variant).

126

127 3.2. Data sources and study populations

128 Details of datasets, study populations, and thresholds used in these analyses are available in
129 Extended Data 1. Briefly, data for adiposity measures were obtained from Pulit et al., (2019)¹⁵;
130 CRC (overall, colon, proximal, distal, and rectal) from Huyghe et al., (2019)¹⁶; and up to 4907
131 plasma proteins from Ferkingstad et al., (2021)¹⁷. BMI and WHR, available in European ancestries
132 (sex-combined and sex-specific), were calculated as $\frac{weight (kg)}{height (m)^2}$ and $\frac{waist circumference (cm)}{hip circumference (cm)}$,
133 respectively; CRC, available in European and East Asian ancestries (sex-combined and sex-
134 specific), was physician diagnosed; and protein concentrations, available in an Icelandic
135 population (sex-combined), were measured in ethylenediaminetetraacetic acid (EDTA) plasma
136 samples using SomaScan[®] (SomaLogic, version 4; 4,907 aptamers). The SomaScan platform uses
137 4,034 modified nucleotides known as Slow Off-rate Modified Aptamers (SOMAmers) which make
138 direct contact with proteins, enabling detection of 3,622 unique proteins or protein complexes
139 and quantifies them in relative fluorescence units (RFUs) using DNA microarray¹⁸. Separate
140 SOMAmers can bind to isoforms of the same protein but can also bind to the same protein at
141 different sites (which can be impacted by post-translational modifications or complexes formed
142 with other proteins) enabling a larger number of proteins and protein complexes to be quantified.

143

144 For all MR analyses, a minimum genome-wide significance threshold of 5×10^{-8} was used for all
145 data – more stringent genome-wide significance thresholds were used for adiposity measures and
146 proteins given wider genotyping coverage¹⁹ – and a linkage disequilibrium (LD) independence
147 threshold of 0.001 was used where applicable (Extended Data 1). For all exposures, F-statistics
148 were calculated for each SNP and a mean was calculated for each instrument, with an F-statistic >
149 10 indicating a strong instrument²⁰. Proteins were included in two UVMR analyses: *cis*- and

150 *trans*-SNPs were used for reverse MR analyses of the association between adiposity measures and
151 proteins to examine reverse causation; *cis*-SNPs were used in the forward MR analyses of the
152 association between proteins and CRC.

153

154 The *cis*-SNPs were obtained directly from the supplementary data of the Ferkingstad et al., paper
155 in which they identified *cis*-SNPs reaching the genome-wide significance threshold (p-value < 1.8
156 x 10⁻⁹) as those loci which were ≤1Mb from the transcription start site of the gene encoding the
157 measured protein. All SNPs within a 1Mb region of each *cis*-SNP were merged into a single
158 region, and the SNP with the lowest p-value was considered the 'sentinel' variant for that protein.
159 Sentinel variants in LD ($r^2 \geq 0.8$) with one another were considered to be a single 'pQTL variant'.
160 In total, Ferkingstad et al., identified a single *cis*-SNP for 1,490 of 4,907 proteins.

161

162 3.2.1. Units

163 Data for adiposity and protein measures were inverse rank normally transformed prior to
164 genome-wide analysis. Assuming the distribution of each trait was normal prior to transformation
165 and genome-wide analysis, we interpret these units to be approximately equivalent to a
166 normalized standard deviation (SD) of the respective trait. Estimates and odds ratios (ORs) are
167 therefore interpreted as the change in outcome per normalized SD unit change in the exposure.

168

169 3.3. Statistical analysis

170 MR relies upon three core assumptions: (1) the instrument must be associated with the exposure
171 of interest, (2) there are no confounders of the association between the instrument and the
172 outcome, and (3) the instrument is not related to the outcome except via its effect on the exposure
173 of interest. The same assumptions are extended to include the intermediate in MVMR: (1) the

174 instrument must be associated with the exposure given the presence of the mediator, (2) the
175 instrument must be independent of the outcome given both exposure and mediator, and (3) the
176 instrument is not related to the outcome except via its effect on the exposure given the presence
177 of the mediator. Assumption 1 may be satisfied by using a standard genome-wide significance
178 threshold of 5×10^{-8} and instruments with an F-statistic, or conditional F-statistic for MVMR
179 analyses²¹, > 10 . Assumptions 2 and 3 are unverifiable but were tested using sensitivity models
180 sensitive to the effects of pleiotropy and with *cis*-SNP UVMR and colocalization. Colocalization
181 attempts to differentiate between distinct causal variants and a single shared signal²². In
182 combination with UVMR, colocalization can be used to assess the validity of MR assumptions
183 (distinct causal variants likely result from LD) and strengthen evidence for a causal effect.

184

185 3.3.1. Identification of associations

186 An adiposity-CRC association (step I) was identified if there was a consistent direction of effect
187 across all MR models and there was no consistent direction of effect in the reverse MR analyses.
188 The same requirement, plus a PhenoSpD²³⁻²⁵ corrected p-value threshold (we used the more
189 conservative of the two approaches applied by PhenoSpD), were used to identify adiposity-
190 protein associations (step II). Only proteins with *cis*-SNP information were used in the
191 subsequent UVMR and MVMR analyses. A protein-CRC association (step III) was identified if the
192 PhenoSpD corrected p-value threshold was met, there was no consistent direction of effect across
193 all MR models in the reverse MR, and evidence of colocalization ($h4 \geq 0.8$) was observed. We
194 interpreted a protein as having a potential mediating role in the adiposity and CRC relationship if
195 the MVMR result adjusting for that protein attenuated towards the null when compared with the
196 UVMR result (step IV).

197

198 3.3.2. Univariable Mendelian randomization

199 For all exposures, the following summary-level data were obtained from the original GWAS: rsID,
200 effect allele, other/non-effect allele, effect allele frequency, effect estimate, standard error of the
201 effect estimate, p-value, sample size, and units. Genetic variants were extracted from each
202 outcome GWAS and, where these were not present, proxy SNPs were included if LD was ≥ 0.8 .
203 For proxy SNPs, the inclusion of SNPs where the reference strand was ambiguous (strand flips)
204 was allowed and the reference strand was inferred using a minor allele frequency (MAF)
205 threshold so long as the MAF was not ≥ 0.3 , in which case the proxy SNP was excluded. Exposure
206 and outcome summary statistics for each of the exposure-related SNPs were harmonised in
207 reference to the exposure effect allele being on the increasing scale. For included alleles where
208 the reference strand was ambiguous, the positive strand was inferred using effect allele frequency
209 (EAF) so long as the EAF was not within 0.3 – 0.7, otherwise the SNP was removed.

210

211 An inverse variance weighted (IVW), multiplicative random effects (IVW-MRE) model was used
212 to estimate the effect of each exposure on the outcome. The model assumes that the strength of
213 the association of the genetic instruments with the exposure is not correlated with the magnitude
214 of the pleiotropic effects and that the pleiotropic effects have an average value of zero²⁶. Where
215 only one SNP was present, the Wald ratio was used (e.g., *cis*-SNP UVMR).

216

217 3.3.2.1. Sensitivity analysis

218 The assumptions of no pleiotropy among genetic instruments and outcomes were explored using:
219 MR-Egger²⁷, weighted median²⁸, and weighted mode²⁹ based estimators where ≥ 3 SNPs were
220 available. MR-Egger provides an estimate of unbalanced or directional horizontal pleiotropy via
221 the intercept of a linear regression of the SNP-exposure and SNP-outcome association²⁷. The

222 weighted median provides consistent estimates when at least 50% of included instruments are
223 invalid²⁸. The weighted mode assumes the true causal effect is the most common effect and it is
224 robust when most effect estimates are derived from valid instruments²⁹. In addition, a single-SNP
225 MR using the Wald ratio and a “leave-one-out” MR sensitivity analysis were conducted to assess
226 the influence of individual variants on the observed associations.

227

228 3.3.3. Colocalization

229 For each protein, the sentinel cis-SNP identified by Ferkingstad et al., was extracted along with a
230 1Mb window. This region was then extracted from each CRC GWAS and colocalization was
231 implemented using the single causal variant approach of Giambartolomei et al., (2014)³⁰. The LD
232 matrix was generated using the 1000 genomes reference panel (phase 3) and priors were set at $p^1 =$
233 10^{-6} , $p^2 = 10^{-6}$, and $p^{12} = 10^{-7}$ based on a window of 5,000 SNPs using
234 <https://chr1swallace.shinyapps.io/coloc-priors/> (accessed 15/05/2023).

235

236 3.3.4. Multivariable Mendelian randomization

237 SNPs associated with the exposure (adiposity) and proposed intermediate proteins (*cis*- and *trans*-
238 SNPs) were extracted and combined. This combined SNP list was extracted from the adiposity
239 GWAS. The resulting instrument was then clumped to remove duplicate SNPs and SNPs in LD
240 with one another using the same clumping thresholds as with the UVMR analysis. An IVW
241 model was used to obtain the direct causal effect of each adiposity measure adjusted for each
242 protein on CRC risk. Instrument strength for each exposure was estimated using a generalized
243 version of Cochran’s Q ³¹ assuming a pairwise covariance of zero¹³.

244

245 3.3.5. Protein expression analyses

246 To investigate whether proteins included in the MVMR analyses were expressed in adipose and

247 CRC tissue we used GTEx³² (version 8) data to compare protein tissue expression relative to whole

248 blood using the Wilcoxon rank sum test and visualised expression profiles using violin plots.

249

250 4. Results

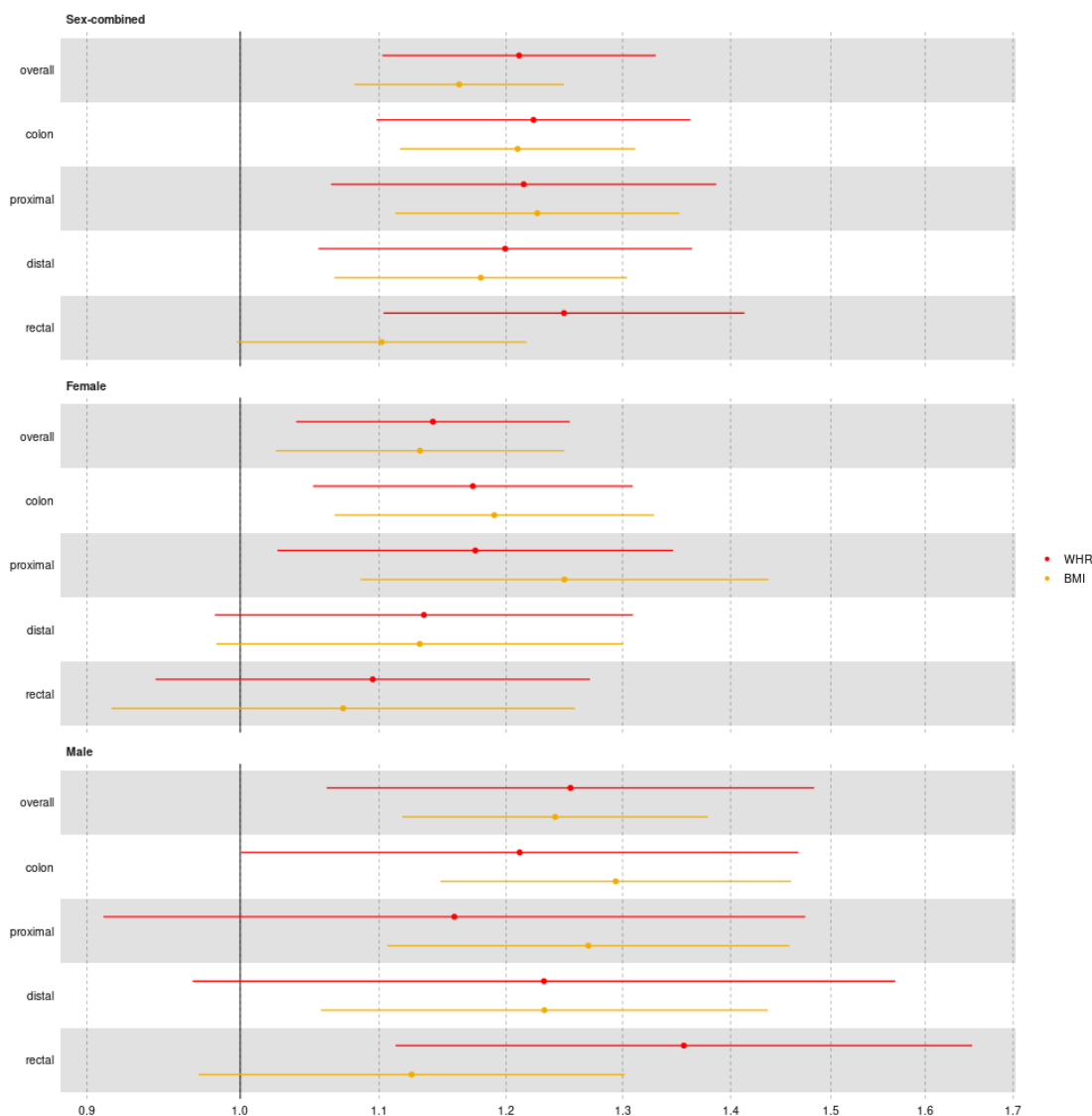
251 An overview of the datasets used, protein information, and all results from the MR and
252 colocalization analyses are presented in Extended Data 1. PhenoSpD identified a total of 1,293
253 independent variables, and a PhenoSpD corrected p-value threshold of 3.97×10^{-5} was, therefore
254 used (Extended Data 2).

255

256 4.1. Association between adiposity measures and colorectal cancer

257 In step I, BMI and WHR were positively associated with overall and site-specific CRC risk in men
258 and women (Figure 2). Sensitivity models were broadly consistent (directions of effect estimates
259 were in the same direction as the IVW-MRE model and CIs overlapped) with the main IVW-
260 MRE model. However, inconsistent directions were observed for the association between WHR
261 and distal colon cancer and overall CRC in men and rectal cancer in women (Extended data 3).
262 Furthermore, the reverse UVMR analyses (Extended data 4) showed an increasing effect of
263 proximal and distal colon cancer on WHR. In both cases, sensitivity analyses produced generally
264 consistent results, particularly for proximal colon cancer in men and for distal colon cancer in
265 women. These pairings were, therefore, excluded from subsequent MVMR analyses.

266



267
268
269
270

Figure 2 Association between adiposity measures and colorectal cancer outcomes. Odds ratios and 95% confidence intervals shown for the main analysis using the inverse variance weighted multiplicative random effects (IVW-MRE). BMI = body mass index; WHR = waist hip ratio

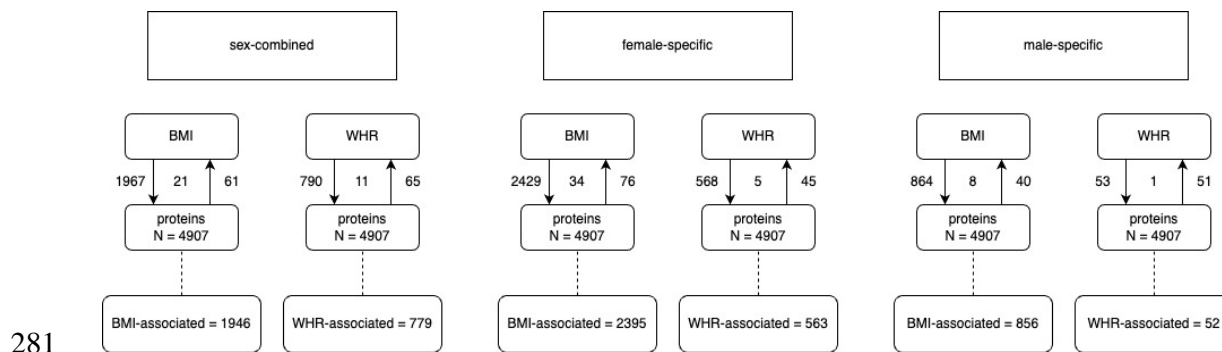
271

272 4.2. Association between adiposity measures and proteins

273 In step II, 6,591 adiposity-protein (2,628 unique proteins) associations were identified (consistent
274 directions of effect across MR models, PhenoSpD corrected p-value threshold reached, and no
275 conflicting association identified in the reverse UVMR; **Error! Reference source not found.**) across
276 the sex-combined and sex-specific analyses. The largest number of associations were identified for
277 the female-specific analysis of BMI (N = 2,395 BMI-associated proteins), with more analyses

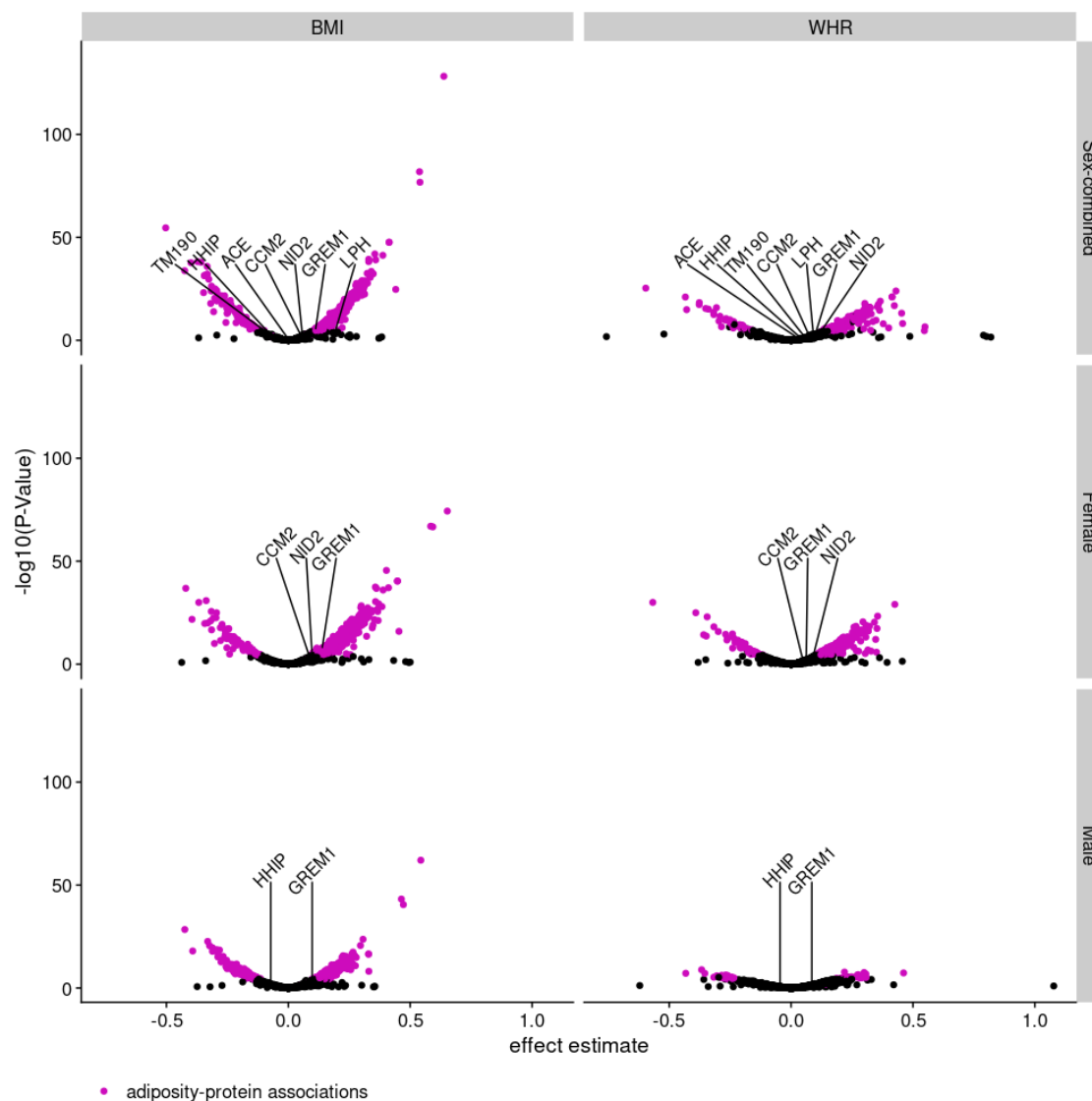
278 reaching the PhenoSpD-corrected p-value threshold in the female analysis compared to the male
 279 analysis (Figure 4).

280



282 *Figure 3 Overview of associations between adiposity measures and plasma proteins (step II in the main analysis plan).*
 283 *Arrows show the direction of the univariable Mendelian randomization (UVMR) analysis. Values on the outside of the lines*
 284 *indicate the number of associations identified in that direction; values in between the lines indicate the number of*
 285 *associations identified in both directions and for which there is, therefore, conflicting evidence of association. N = gives the*
 286 *number of proteins available for analysis. BMI- WHR-associated gives the number of proteins associated with the adiposity*
 287 *measure. BMI = body mass index; WHR = waist hip ratio.*

288



289

290 *Figure 4 Association between adiposity measures and proteins in step II of the main analysis plan. The volcano plot shows*
291 *effect estimates and $-\log_{10}(\text{pval})$. Adiposity-protein associations are highlighted in purple (analyses reaching the PhenoSpD*
292 *corrected p-value (0.05/1293), consistent directions of effect across MR models, and no conflicting association identified in*
293 *the reverse UVMR). Protein labels highlight those proteins which were associated with colorectal cancer outcomes in the*
294 *UVMR analysis in step III of the main analysis plan. BMI = body mass index; WHR = waist hip ratio.*

295

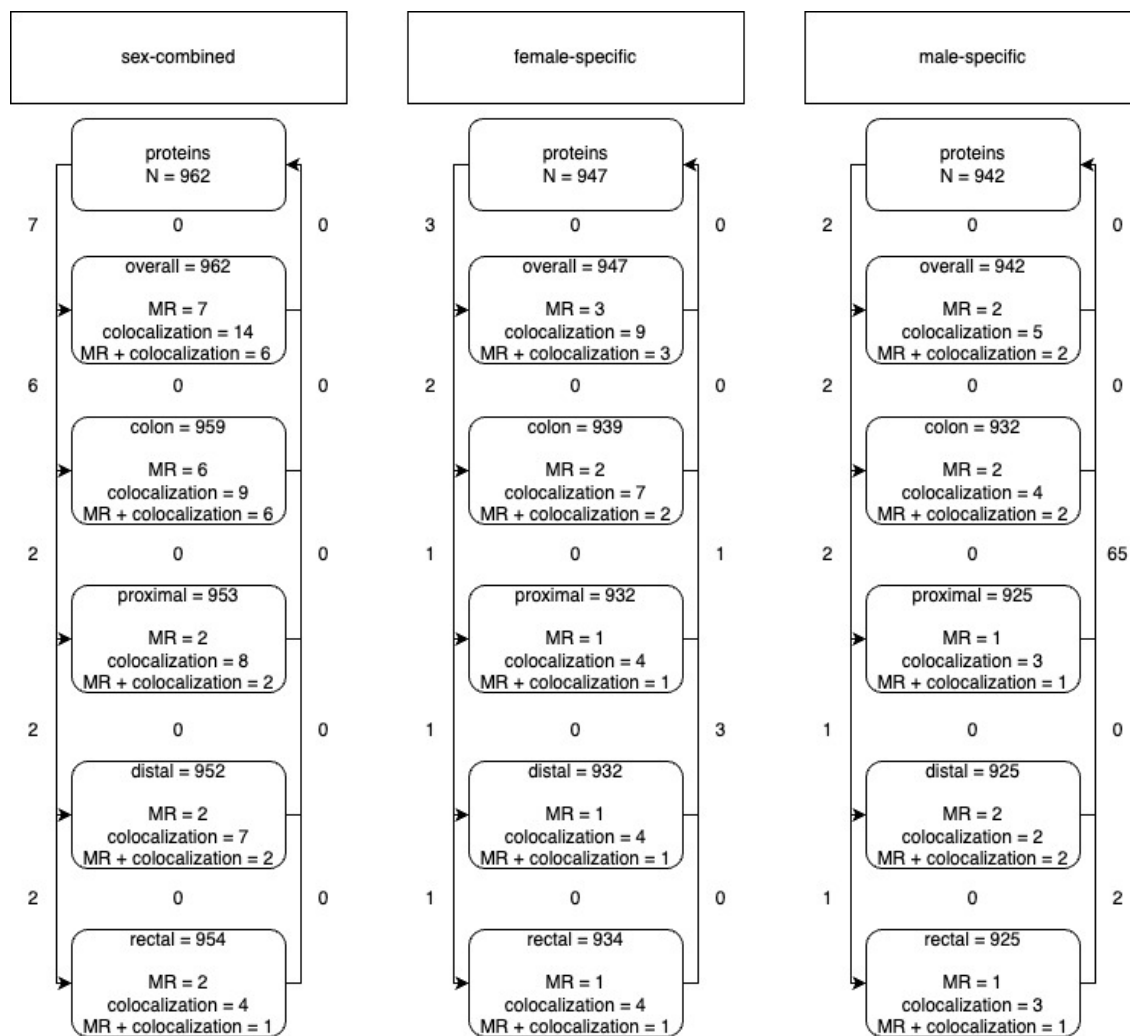
296 4.3. Association between proteins and colorectal cancer

297 In step III, it was possible to perform UVMR analysis of up to 962 proteins in relation to CRC

298 outcomes using *cis*-SNPs. In total, 35 protein-CRC (8 unique proteins) associations were identified

299 (PhenoSpD-corrected p-value threshold reached and no conflicting association identified in the

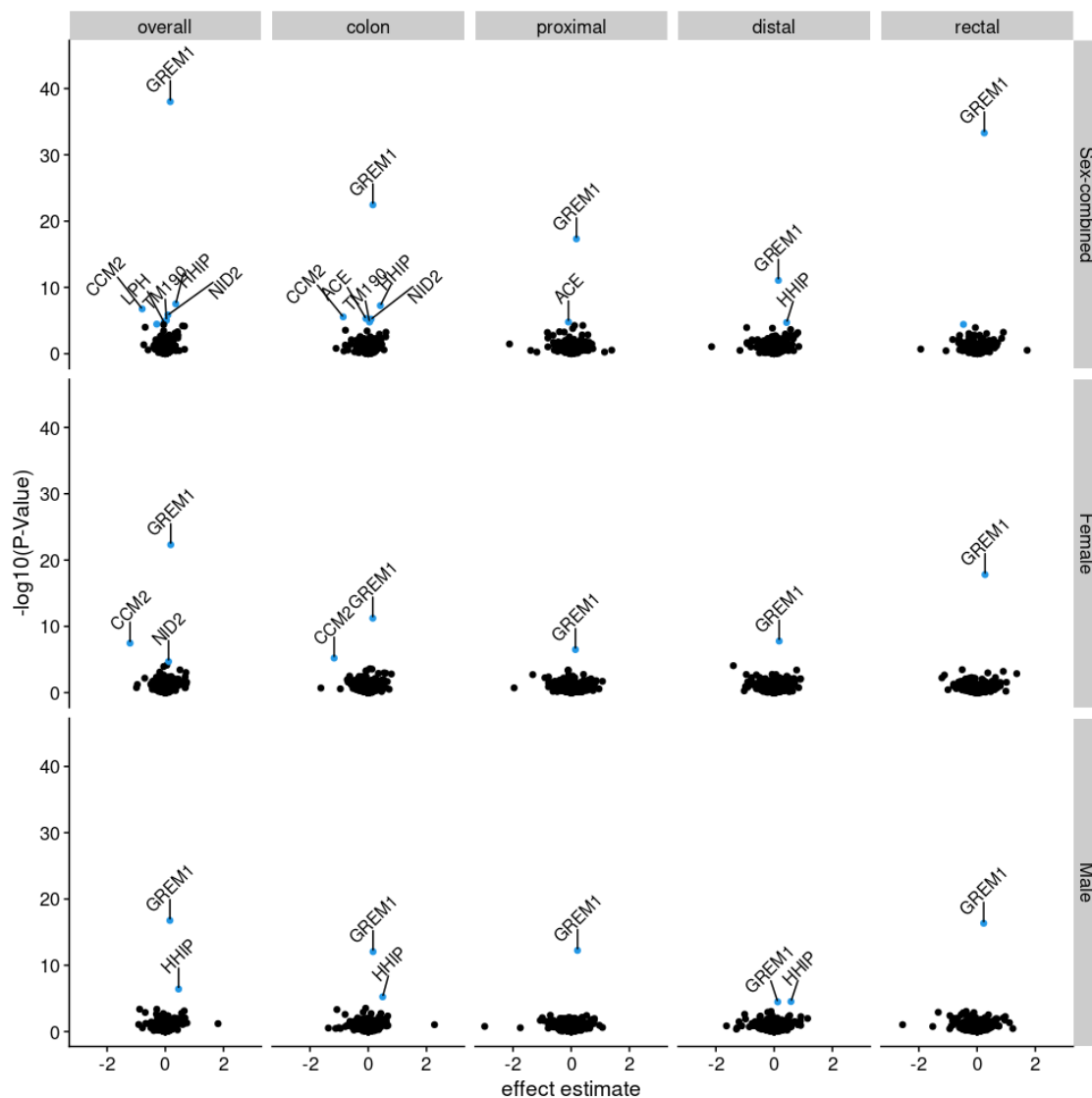
300 reverse UVMR; Figure 5) across the sex-combined and sex-specific UVMR analyses. In the male-
301 specific analyses, 65 reverse associations were identified between proteins and distal colon cancer
302 risk. However, there were no conflicts with the forward UVMR analysis in this or any other
303 protein-CRC analyses. There was evidence ($h4 \geq 0.8$) of colocalization for 87 protein-CRC pairs
304 (25 unique proteins) across the sex-combined and sex-specific analyses of the CRC outcomes
305 (Extended Data 4). Of the 35 protein-CRC UVMR associations, 2 were not corroborated by
306 colocalization ($h4 < 0.5$ for the protein GRFAL in relation to overall CRC and rectal cancer risk in
307 sex-combined analysis).
308



309

310 *Figure 5 Overview of associations between proteins and colorectal cancer (CRC) outcomes (step III in the main analysis*
 311 *plan). Arrows show the direction of the univariable Mendelian randomization (UVMR) analysis. Values on the outside of the*
 312 *lines indicate the number of associations identified in that direction; values in between the lines indicate the number of*
 313 *associations identified in both directions and for which there is, therefore, conflicting evidence of association. N = gives the*
 314 *number of proteins available for analysis. MR = gives the number of cis-SNP UVMR analyses which reached the PhenoSpD*
 315 *pvalue threshold for that analysis. Colocalization = gives the number of proteins which colocalized with that CRC outcome.*
 316 *MR + colocalization = gives the overlap between the cis-SNP UVMR and colocalization analyses and indicates the protein-*
 317 *CRC associations.*

318



319

320 *Figure 6 Association between adiposity-related proteins and colorectal cancer (CRC) outcomes in step III of the main*
 321 *analysis plan. The volcano plot shows effect estimates and $-\log_{10}(p\text{-value})$ with analyses reaching the PhenoSpD corrected p-*
 322 *value (0.05/1293) highlighted in blue and analyses reaching the PhenoSpD corrected-pvalue and with evidence of*
 323 *colocalization labelled with the protein name. The X-axis has been constrained to -3 – 3, excluding 3 analyses which did not*
 324 *meet any association thresholds: PTP4A2 and proximal colon cancer in males (effect estimate = 102) and NANS and distal*
 325 *colon cancer in males (effect estimate = 19) and females (effect estimate = 19).*

326

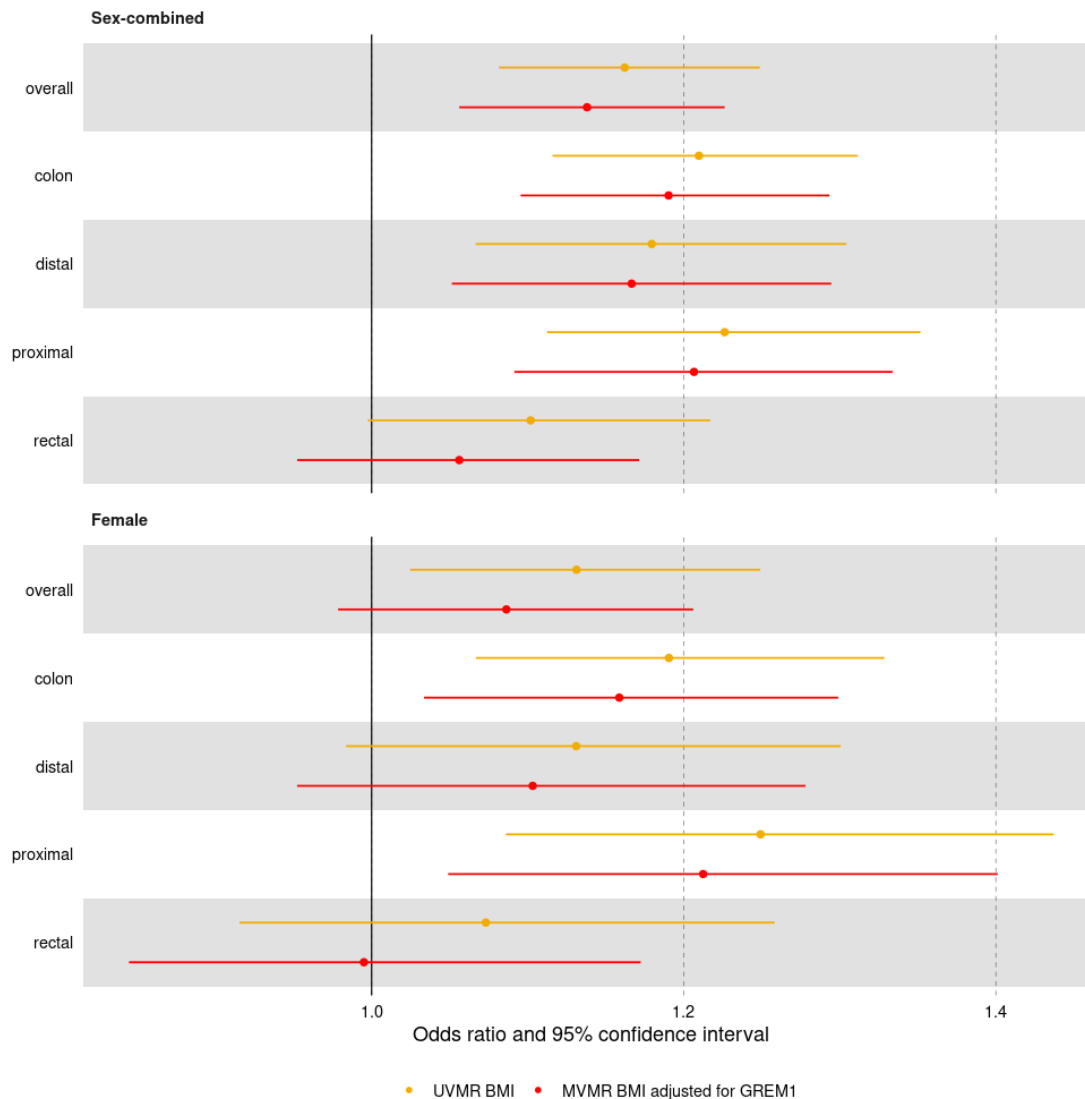
327 4.4. Multivariable Mendelian randomization

328 Of the 7 proteins associated with CRC outcomes across the sex-combined and sex-specific

329 analyses, the results for only 1 (GREM1) were directionally consistent with a possible mediating

330 role of the association between adiposity and CRC risk. That is, an increase in BMI was associated

331 with an increase in GREM1 and an increase in GREM1 was associated with an increase in CRC
332 risk. In MVMR (step IV) we considered a protein as having a potential mediating role if, in
333 comparison to the UVMR estimate of the association between adiposity and CRC, the adjusted
334 estimate was attenuated (i.e., the effect estimate tends towards the null and the CI overlaps the
335 null). For GREM1, effect estimates and CIs for all MVMR analyses tended towards the null.
336 However, evidence of attenuation was observed solely for the female-specific analysis of overall
337 CRC risk (Figure 7). Attenuation of the sex-combined analysis of rectal CRC was also observed,
338 but the CI of the UVMR effect of BMI on rectal cancer risk also overlapped the null. Conditional
339 F-statistics were > 10 (Extended Data 4). There was evidence that all but the female-specific
340 analysis of distal colon cancer used invalid instruments (Q statistic p-value < 0.05).
341



342

343 *Figure 7 Association between body mass index (BMI) and colorectal cancer outcomes using univariable (UV; orange line)*
344 *and multivariable (MV; red lines) Mendelian randomization (MR). In these MVMR analyses (step IV), the effect of BMI on*
345 *colorectal cancer outcomes is estimated after adjusting for the effect of GREM1. Odds ratios for the inverse variance*
346 *weighted multiplicative random effects model shown alongside 95% confidence intervals. No adiposity-protein-CRC*
347 *associations were identified in the male UVMR analyses and as such MVMR was not performed.*

348

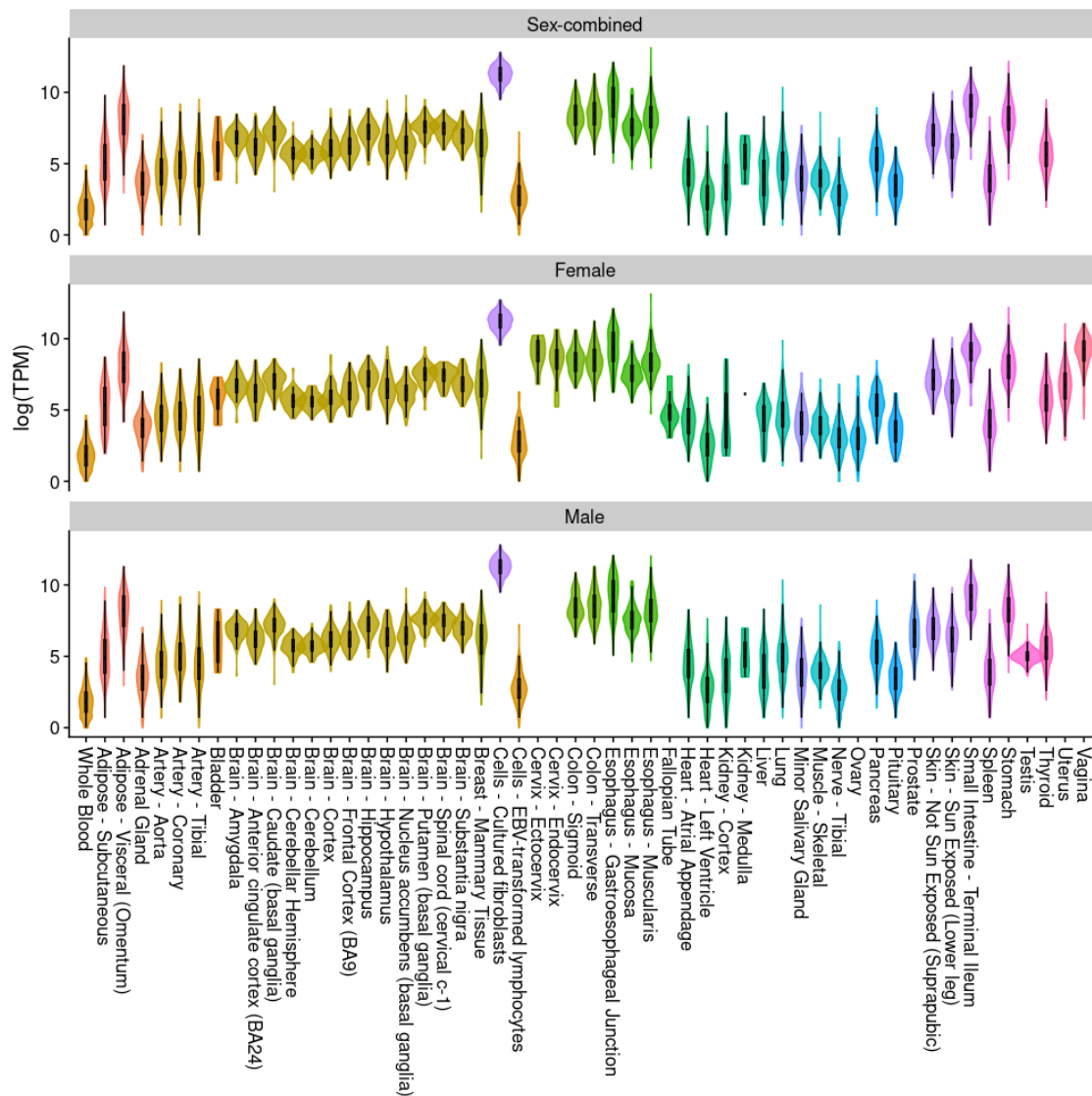
349 4.1. Follow-up analysis

350 Using GTEx data, GREM1 was found to be differentially expressed (Bonferroni corrected p-value

351 = 0.05/53) in most tissues compared to whole blood, with some of the highest levels noted in the

352 gastrointestinal tract and visceral adipose tissue in both sexes (Figure 8).

353



354

355 *Figure 8 Tissue gene expression profile of GREM1. The violin plot presents expression levels as log transcripts per million*
 356 *(TPM). Data are from GTEx version 8³². Box plots are shown with the interquartile range (25th and 75th percentiles).*
 357 *Colours distinguish different tissue types (e.g., adipose and brain).*

358

359 5. Discussion

360 Using complementary MR and colocalization analyses we conducted the largest and most
361 comprehensive study to date examining the role of the plasma proteome as an intermediate in the
362 relationship between adiposity and CRC development. We found novel evidence of a potential
363 mediating role of the protein GREM1, primarily in the association between BMI and overall CRC
364 risk in women. For the remainder of our analyses, we found little evidence of individual proteins
365 acting as intermediates in the adiposity-CRC relationship.

366

367 GREM1, a bone morphogenic protein (BMP) antagonist³³, has previously been linked with
368 CRC^{16,17,34,35}, and is expressed in many tissues, including adipose and colon tissue. BMPs play an
369 important role in embryonic development and morphogenesis and may be considered tumour
370 repressors as inhibition of BMP signalling has been associated with a number of cancers^{36,37}.
371 GREM1 is associated with proliferation, angiogenesis, and epithelial-to-mesenchymal transition
372 of cancer cells³⁶ and may be involved in colon cancer tumour progression³⁸, with a number of
373 studies linking GREM1 to CRC development^{34,35,39}. In addition, a recent gene-environment
374 interaction analysis identified a locus in the *FMNI/GREM1* gene region that interacted with BMI
375 on the positive association with CRC risk⁴⁰. The exact mechanisms underlying the GREM1-CRC
376 relationship are unclear but may be related to expression in the tumour microenvironment given
377 that GREM1 expression is lower in CRC tissue than adjacent non-cancerous and normal tissue³⁴.
378 Increased GREM1 expression in CRC tissue has also been associated with low tumour stage and a
379 more favourable prognosis⁴¹, and increased GREM1 expression is found in the tumour
380 microenvironment, such as in visceral adipose tissue⁴² and colonic crypt bases via cancer
381 associated fibroblasts⁴³.

382

383 To our knowledge, this was the first comprehensive proteogenomic analysis conducted so far to
384 estimate the role of the plasma proteome as an intermediate of the positive association found
385 between adiposity and CRC. Our study has several additional notable strengths. Adiposity-
386 protein-CRC triples included in the MVMR analyses were supported by consistent directions of
387 effect across the main IVW-MRE model and sensitivity models (MR-Egger, weighted median, and
388 weighted mode). We also conducted reverse UVMR analyses for all adiposity-CRC, adiposity-
389 protein, and protein-CRC analyses to identify possible reverse causation and exclude pairs where
390 evidence was conflicting. Several limitations should also be acknowledged. Although the
391 availability of sex-specific summary statistics from large genome-wide association studies were
392 available for adiposity measures and CRC, only sex-combined protein data were available which
393 may have led to biased estimates, especially for some of the MVMR analyses, in which the
394 adiposity exposure and outcome were both sex-specific. Instrument strength, measured via F-
395 statistics, was appropriate for most analyses, including the MVMR analyses. However, there was
396 evidence of weak instrument bias across most MVMR analyses which can lead to estimates for the
397 exposure and intermediate moving towards and away from the null¹³. In addition, *cis*-SNPs
398 identified by Ferkingstad *et al.*, were available for less than half of the 4,907 proteins included in
399 the adiposity-protein UVMR analysis.

400

401 Our results highlight the broad impact of adiposity on the plasma proteome and of adiposity-
402 associated circulating proteins on the risk of developing CRC. We found robust evidence of a
403 causal effect of one adiposity-related protein, GREM1, as a likely mediator of the adiposity and
404 CRC relationship, particularly for women. Given evidence from previous studies highlighting the
405 relationships between GREM1 and CRC, our results suggest that the GREM1 pathway may be a

406 potential mechanism underlying the obesity-CRC relationship and future experimental

407 investigation is warranted.

408

409 6. Author contributions

410 MAL: Data curation, Formal analysis, Investigation, Methodology, Project administration,

411 Visualisation, Writing – original draft, Writing – review and editing

412 CAH: Methodology, Writing – review and editing

413 EH: Methodology, Writing – review and editing

414 LG: Writing – review and editing

415 UP: Writing – review and editing

416 KKT: Writing – review and editing

417 EEV: Writing – review and editing

418 RMM: Writing – review and editing

419 MJG: Conceptualization, Funding acquisition, Methodology, Project administration, Resources,

420 Supervision, Writing – review and editing

421 HB: Writing – review and editing

422 IC: Writing – review and editing

423 SK: Writing – review and editing

424 LLM: Writing – review and editing

425 PAN: Writing – review and editing

426 RES: Writing – review and editing

427 BVG: Conceptualization, Funding acquisition, Methodology, Project administration, Resources,

428 Supervision, Writing – review and editing

429 NM: Conceptualization, Funding acquisition, Methodology, Project administration, Resources,

430 Supervision, Writing – review and editing

431

432 7. Conflict of interest

433 None declared.

434

435 8. Acknowledgements & Funding

436 Funding (IIG_FULLL_2021_030) was obtained from Wereld Kanker Onderzoek Fonds (WKOF), as
437 part of the World Cancer Research Fund International grant programme, and Cancer Research
438 UK (C18281/A29019). NM is supported by the French National Cancer Institute (INCa SHSESP20,
439 grant No. 2020-076). BVG is supported by the Swedish Cancer Society (fellowship No. 21 0467 FE
440 01 H and project grant No. 20 1154 PjF). EH is supported by a Cancer Research UK Population
441 Research Committee Studentship (C18281/A30905 and is part of the Medical Research Council
442 Integrative Epidemiology Unit at the University of Bristol which is supported by the Medical
443 Research Council (MC_UU_00032/03) and the University of Bristol. LJG is supported by a
444 Transition Fellowship as part of the British Heart Foundation Accelerator Award (AA/18/1/34219)
445 and an Academic Career Development Fund (University of Bristol). RMM is a National Institute
446 for Health Research Senior Investigator (NIHR202411). RMM is supported by a Cancer Research
447 UK 25 (C18281/A29019) programme grant (the Integrative Cancer Epidemiology Programme).
448 RMM is also supported by the NIHR Bristol Biomedical Research Centre which is funded by the
449 NIHR (BRC-1215-20011) and is a partnership between University Hospitals Bristol and Weston
450 NHS Foundation Trust and the University of Bristol. RMM is affiliated with the Medical Research
451 Council Integrative Epidemiology Unit at the University of Bristol which is supported by the
452 Medical Research Council (MC_UU_00011/1, MC_UU_00011/3, MC_UU_00011/6, and
453 MC_UU_00011/4) and the University of Bristol. Department of Health and Social Care disclaimer:

454 The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or
455 the Department of Health and Social Care.

456

457 9. Disclaimer

458 Where authors are identified as personnel of the International Agency for Research on
459 Cancer/World Health Organization, the authors alone are responsible for the views expressed in
460 this article and they do not necessarily represent the decisions, policy or views of the
461 International Agency for Research on Cancer/World Health Organization.

462

463 10. Data availability

464 10.1. Underlying data

465 This work is supported by a GitHub repository
466 (https://github.com/mattlee821/adiposity_proteins_colorectal_cancer) which is archived on
467 Zenodo (<https://zenodo.org/record/7780822#.ZCQ3U-xBz0o>). Here, all publicly available data,
468 code, and results used in this work are available. The full summary statistics for BMI and WHR
469 are publicly available from Zenodo (<https://zenodo.org/record/1251813#.Yk7O25PMIUE>). The full
470 summary statistics for all proteins are publicly available from DECODE
471 (<https://download.decode.is/form/folder/proteomics>). The full summary statistics for CRC are not
472 publicly available but can be obtained from GECCO
473 ([https://www.fredhutch.org/en/research/divisions/public-health-sciences-
474 division/research/cancer-prevention/genetics-epidemiology-colorectal-cancer-consortium-
475 gecco.html](https://www.fredhutch.org/en/research/divisions/public-health-sciences-division/research/cancer-prevention/genetics-epidemiology-colorectal-cancer-consortium-gecco.html)).

476

477 10.2.Extended Data

478 This project contains the following extended data available from Zenodo
479 <https://zenodo.org/record/7780822#.ZCQ3U-xBz0o>):

480

481 Extended data 1 is an Excel file of 8 tables:

482 1. Overview of all data used, including doi of source. Columns: N = overall sample size or case
483 sample size; pvalue_threshld = genome-wide significance threshold used to identify associated
484 single nucleotide polymorphisms (SNPs); LD_r2 = linkage disequilibrium independence threshold;
485 LD_window = LD independence base window; SNPs = associated SNPs identified using pvalue
486 and LD thresholds; f_stat = mean f-statistics for the associated SNPs; measure = how the trait was
487 measured in the source; adjustment = how the trait was adjusted during genome-wide analysis;
488 transformation = transformation of the trait prior to genome-wide analysis; unit = how the trait is
489 expressed given the transformation.

490

491 2. Extended information on proteins sourced from Ferkingstad *et al.*

492

493 3. All publicly available exposure data used in Mendelian randomization analyses – this does not
494 include colorectal cancer data. Columns: CHR = chromosome; POS = base position; SNP = rsID;
495 eaf.exposure = effect allele frequency; beta.exposure = effect estimate; se.exposure = standard error
496 of the beta; pval.exposure = pvalue of the effect estimate and standard error.

497

498 4-8. All results produced in the Mendelian randomization (MR) and colocalization analyses, split
499 by analysis: adiposity-cancer, adiposity-protein, protein-cancer, colocalization, multivariable MR.

500 Columns: nsnp = number of SNPs used for the exposure instrument; b = effect estimate; se =

501 standard error; pval = pvalue of the effect estimate and standard error; forward always refers to
502 column 1 as the exposure and column 2 as the outcome; reverse always refers to column 2 as the
503 exposure and column 1 as the outcome; Qstat/Qpval = Cochranes Q statistic and associated pvalue;
504 fstat_adiposity/fstat_protein = the F-statistic associated with adiposity and protein instrument
505 used in the analysis; nsnp_colocalization = number of SNPs included in the 1Mb window used for
506 colocalization; h0-4 = hypothesis priors; h0 = neither trait has a genetic association in the region;
507 h1 = only trait 1 has a genetic association in the region; h2 = only trait 2 has a genetic association
508 in the region; h3 = both traits are associated, but with different causal variants; h4 = both traits are
509 associated and share a single causal variant.

510

511 Extended data 2. PhenoSpD results. File gives the eigenvalue associated with each factor in the
512 correlation matrix and the variance of these values. We use the effective number of independent
513 variables calculated using the method of Li and Ji (2005)²⁵.

514 Extended data 3. Association between adiposity measures and colorectal cancer outcomes. Effect
515 estimates and 95% confidence intervals shown for the main analysis using the inverse variance
516 weighted multiplicative random effects (IVW-MRE) model and 3 sensitivity models. BMI = body
517 mass index; WHR = waist hip ratio.

518

519 Extended data 4. Association between colorectal cancer measures and adiposity measures. Effect
520 estimates and 95% confidence intervals shown for the main analysis using the inverse variance
521 weighted multiplicative random effects (IVW-MRE) model and 3 sensitivity models. BMI = body
522 mass index; WHR = waist hip ratio.

523

524 Extended data 5. STROBE-MR checklist^{44,45}

525

526

527 11. References

- 528 1. Lee, M. *et al.* Systematic review and meta-analyses: What has the application of
529 Mendelian randomization told us about the causal effect of adiposity on health outcomes?
530 [version 1; peer review: awaiting peer review]. *Wellcome Open Res.* **7**, (2022).
531 2. Bull, C. J. *et al.* Adiposity, metabolites, and colorectal cancer risk: Mendelian
532 randomization study. *BMC Med.* **18**, 396 (2020).
533 3. Lauby-Secretan, B. *et al.* Body Fatness and Cancer — Viewpoint of the IARC Working
534 Group. *N. Engl. J. Med.* **375**, 794–798 (2016).
535 4. Goudswaard, L. J. *et al.* Effects of adiposity on the human plasma proteome:
536 observational and Mendelian randomisation estimates. *Int. J. Obes.* **2005** **45**, 2221–2229
537 (2021).
538 5. Kibble, M. *et al.* An integrative machine learning approach to discovering multi-level
539 molecular mechanisms of obesity using data from monozygotic twin pairs. *R. Soc. Open*
540 *Sci.* **7**, 200872 (2022).
541 6. Rajan, M. R. *et al.* Comparative analysis of obesity-related cardiometabolic and renal
542 biomarkers in human plasma and serum. *Sci. Rep.* **9**, 15385 (2019).
543 7. Zaghlool, S. B. *et al.* Revealing the role of the human blood plasma proteome in obesity
544 using genetic drivers. *Nat. Commun.* **12**, 1279 (2021).
545 8. Burgess, S., Daniel, R. M., Butterworth, A. S., Thompson, S. G. & Consortium, E.-I.
546 Network Mendelian randomization: using genetic variants as instrumental variables to
547 investigate mediation in causal pathways. *Int. J. Epidemiol.* **44**, 484–495 (2015).
548 9. Relton, C. L. & Davey Smith, G. Two-step epigenetic mendelian randomization: A
549 strategy for establishing the causal role of epigenetic processes in pathways to disease.
550 *Int. J. Epidemiol.* (2012) doi:10.1093/ije/dyr233.
551 10. Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of
552 multivariable Mendelian randomization in the single-sample and two-sample summary
553 data settings. *Int. J. Epidemiol.* **48**, 713–727 (2019).
554 11. Carter, A. R. *et al.* Mendelian randomisation for mediation analysis: current methods and
555 challenges for implementation. *Eur. J. Epidemiol.* **36**, 465–478 (2021).
556 12. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the
557 human phenome. *eLife* **7**, e34408 (2018).
558 13. Sanderson, E., Spiller, W. & Bowden, J. Testing and correcting for weak and pleiotropic
559 instruments in two-sample multivariable Mendelian randomization. *Stat. Med.* **n/a**,
560 (2021).
561 14. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in
562 colocalisation analyses. *PLOS Genet.* **16**, e1008720 (2020).
563 15. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat
564 distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174
565 (2019).
566 16. Huyghe, J. R. *et al.* Discovery of common and rare genetic risk variants for colorectal
567 cancer. *Nat. Genet.* **51**, 76–87 (2019).
568 17. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and
569 disease. *Nat. Genet.* **53**, 1712–1721 (2021).
570 18. Rohloff, J. C. *et al.* Nucleic Acid Ligands With Protein-like Side Chains: Modified
571 Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Mol. Ther. Nucleic Acids*
572 **3**, e201 (2014).
573 19. Pulit, S. L., de With, S. A. J. & de Bakker, P. I. W. Resetting the bar: Statistical
574 significance in whole-genome sequencing-based association studies of global
575 populations. *Genet. Epidemiol.* **41**, 145–151 (2017).

- 576 20. Haycock, P. C. *et al.* Best (but oft-forgotten) practices: The design, analysis, and
577 interpretation of Mendelian randomization studies. *Am. J. Clin. Nutr.* **103**, (2016).
- 578 21. Sanderson, E. & Windmeijer, F. A weak instrument [Formula: see text]-test in linear IV
579 models with multiple endogenous variables. *J. Econom.* **190**, 212–221 (2016).
- 580 22. Zuber, V. *et al.* Combining evidence from Mendelian randomization and colocalization:
581 Review and comparison of approaches. *Am. J. Hum. Genet.* **109**, 767–782 (2022).
- 582 23. Zheng, J. *et al.* PhenoSpD: an integrated toolkit for phenotypic correlation estimation and
583 multiple testing correction using GWAS summary statistics. *GigaScience* **7**, (2018).
- 584 24. Nyholt, D. R. A simple correction for multiple testing for single-nucleotide
585 polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74**, 765–
586 769 (2004).
- 587 25. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a
588 correlation matrix. *Heredity* **95**, 221–227 (2005).
- 589 26. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between
590 imprecisely measured traits using GWAS summary data. *PLOS Genet.* **13**, e1007081
591 (2017).
- 592 27. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid
593 instruments: effect estimation and bias detection through Egger regression. *Int. J.*
594 *Epidemiol.* **44**, 512–525 (2015).
- 595 28. Burgess, S. *et al.* Dissecting Causal Pathways Using Mendelian Randomization with
596 Summarized Genetic Data: Application to Age at Menarche and Risk of Breast Cancer.
597 *Genetics* **207**, 481–487 (2017).
- 598 29. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data
599 Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*
600 (2017) doi:10.1093/ije/dyx102.
- 601 30. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic
602 Association Studies Using Summary Statistics. *PLOS Genet.* **10**, e1004383 (2014).
- 603 31. Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of
604 multivariable Mendelian randomization in the single-sample and two-sample summary
605 data settings. *Int. J. Epidemiol.* **48**, 713–727 (2018).
- 606 32. THE GTEx CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects
607 across human tissues. *Science* **369**, 1318–1330 (2020).
- 608 33. GREM1 gremlin 1, DAN family BMP antagonist [Homo sapiens (human)] - Gene -
609 NCBI. <https://www.ncbi.nlm.nih.gov/gene/26585>.
- 610 34. Jang, B. G. *et al.* Prognostic significance of stromal GREM1 expression in colorectal
611 cancer. *Hum. Pathol.* **62**, 56–65 (2017).
- 612 35. Li, R. *et al.* Gremlin-1 Promotes Colorectal Cancer Cell Metastasis by Activating ATF6
613 and Inhibiting ATF4 Pathways. *Cells* **11**, 2136 (2022).
- 614 36. Park, S.-A. Role of Gremlin-1 in Cancer. *Biomed. Sci. Lett.* **24**, 285–291 (2018).
- 615 37. Bach, D.-H., Park, H. J. & Lee, S. K. The Dual Role of Bone Morphogenetic Proteins in
616 Cancer. *Mol. Ther. Oncolytics* **8**, 1–13 (2018).
- 617 38. Karagiannis, G. S. *et al.* Bone morphogenetic protein antagonist gremlin-1 regulates
618 colon cancer progression. *Biol. Chem.* **396**, 163–183 (2015).
- 619 39. Li, J. *et al.* A functional variant in GREM1 confers risk for colorectal cancer by
620 disrupting a hsa-miR-185-3p binding site. *Oncotarget* **8**, 61318–61326 (2017).
- 621 40. Aglago, E. K. *et al.* A genetic locus within the FMN1/GREM1 gene region interacts with
622 body mass index in colorectal cancer risk. *Cancer Res.* CAN-22-3713 (2023)
623 doi:10.1158/0008-5472.CAN-22-3713.

- 624 41. Gremlin1 expression associates with serrated pathway and favourable prognosis in
625 colorectal cancer - Pelli - 2016 - Histopathology - Wiley Online Library.
626 <https://onlinelibrary.wiley.com/doi/abs/10.1111/his.13006>.
- 627 42. The Novel Adipokine Gremlin 1 Antagonizes Insulin Action and Is Increased in Type 2
628 Diabetes and NAFLD/NASH | Diabetes | American Diabetes Association.
629 [https://diabetesjournals.org/diabetes/article/69/3/331/39768/The-Novel-Adipokine-](https://diabetesjournals.org/diabetes/article/69/3/331/39768/The-Novel-Adipokine-Gremlin-1-Antagonizes-Insulin)
630 [Gremlin-1-Antagonizes-Insulin](https://diabetesjournals.org/diabetes/article/69/3/331/39768/The-Novel-Adipokine-Gremlin-1-Antagonizes-Insulin).
- 631 43. Dutton, L. R. *et al.* Fibroblast-derived Gremlin1 localises to epithelial cells at the base of
632 the intestinal crypt. *Oncotarget* **10**, 4630–4639 (2019).
- 633 44. Skrivankova, V. W. *et al.* Strengthening the reporting of observational studies in
634 epidemiology using mendelian randomisation (STROBE-MR): explanation and
635 elaboration. *BMJ* **375**, n2233 (2021).
- 636 45. Skrivankova, V. W. *et al.* Strengthening the Reporting of Observational Studies in
637 Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. *JAMA*
638 **326**, 1614–1621 (2021).
639
640