

1 Title Page:

2 **Insulin-like growth factor 2 hypermethylation in peripheral blood leukocytes and colorectal**
3 **cancer risk and prognosis: a propensity score analysis of multiple-centre populations**

4 **Short Title:** IGF2 and CRC risk and prognosis

5 HongRu Sun^{1#}, YanLong Liu^{2#}, YuXue Zhang³, Yibaina Wang¹, BinBin Cui², YaShuang Zhao^{1*}, YuPeng Liu^{1*}

6 **Author affiliations:**

7 **HongRu Sun** (sunhongru_1024@126.com), **Yibaina Wang** (zhulin_vip@yeah.net), **YaShuang Zhao**

8 (zhao_yashuang@263.net), and **YuPeng Liu** (liuyupeng@wmu.edu.cn), ¹Department of Epidemiology, Public Health
9 College, Harbin Medical University, 157 Baojian Street, Harbin 150081, Heilongjiang Province, China;

10 **YanLong Liu** (liuyanlong1979@163.com) and **BinBin Cui** (cuibinbin222@hotmail.com), ²Department of Colorectal
11 Surgery, The Third Affiliated Cancer Hospital of Harbin Medical University, 150 Haping Street, Harbin 150081,
12 Heilongjiang Province, China;

13 **YuXue Zhang** (zhang_yuxue@126.com), ³Department of Hygiene Microbiology, Public Health College, Harbin Medical
14 University, 157 Baojian Street, Harbin 150081, Heilongjiang Province, China;

15 [#]These authors have contributed equally to this work and share first authorship.

16 ^{*}These authors have contributed equally to this work and share last authorship.

17 ^{*}**Corresponding Author: YuPeng Liu**, Ph.D., **Postal address:** 157# Baojian Street, Nangang District, Harbin 150081,
18 Heilongjiang Province, China. **Telephone:** 86-(0)451-86611797, **Fax:** 86-(0)451-87502885, **E-mail:**
19 liuyupeng@wmu.edu.cn; ORCID: 0000-0002-8652-1987.

20 **Author Contributors:** HRS and YLL contributed equally to this work. YPL had full access to all of the data in the
21 initial study and take responsibility for the integrity of the data and the accuracy of the data analysis. YPL and YSZ
22 contributed to study conception and design. YPL, HRS, YBNW and YXZ contributed to DNA preparation and bisulfite
23 modification. YPL, HRS, YBNW, and YLL contributed to MS-HRM. YPL, HRS, YXZ, and YSZ contributed to
24 collection and assembly of data. YSZ, BBC, YPL, HRS, and YLL contributed to the analysis and interpretation of data.
25 YPL, YLL, BBC, YXZ and YSZ contributed to sample collection. YPL, HRS, YLL, YXZ and YSZ contributed to the
26 drafting of the initial versions of the manuscript. All authors contributed to the review and final approval of the
27 manuscript. YPL and YSZ were responsible for study supervision.

28 **Conflict of Interests:** All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts

29 of Interest. The authors have declared that no conflict of interest exists. The funders had no role in the design of the study;
30 in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the
31 results.

32 **Ethics approval:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by
33 the Medical Ethics Committee of Harbin Medical University (1 Mar, 2014). All participants in the initial cohort provided
34 written informed consent. Informed consent was obtained from all subjects involved in the initial cohort. Written
35 informed consent has been obtained from the patients to publish this paper.

36 **Grant support:** This work was supported by the China Postdoctoral Science Foundation (grant number 2018M641875 to
37 YPL); the Natural Science Foundation of Heilongjiang Province (grant number YQ2019H021 to YPL); the National
38 Natural Science Foundation of China (grant number 81473055 to YSZ), and by grant from the SCORE Foundation
39 (Y-MX2016-045 to YLL).

40 **Writing Assistance:** The authors thank American Journal Experts for English language polishing of the manuscript.

41 **Acknowledgments:** The authors thank the participants and staff of both the initial study and the validation studies from
42 the GEO and TCGA for providing data and their valuable contributions. The authors thank the GEO and TCGA databases
43 for the access to the validation datasets. The authors thank Dr. Carlotta Sacerdote (Città della Salute e della Scienza
44 University) and Dr. Giovanni Fiorito (University of Sassari) for providing critical help in validation stage. The authors
45 thank Professor Wei Chen (Tulane University) for critically reading this manuscript.

46 **Data Transparency Statement:** Authors are willing to share any data that are used in this work. The data that support
47 the findings of the initial study are included in the manuscript and supporting file. The raw datasets used in the validation
48 stage are publicly available on GEO (<http://www.ncbi.nlm.nih.gov/geo/>) and TCGA (CORD and READ)
49 (<https://cancergenome.nih.gov/>).

50 **Abbreviations:** body mass index (BMI), carcinoembryonic antigen (CEA), confidence intervals (CIs), colorectal cancer
51 (CRC), CRC-specific survival (CSS), disease free survival (DFS), Gene Expression Omnibus (GEO), Harbin Medical
52 University (HMU), hazard ratios (HRs), insulin-like growth factor 2 (IGF2), loss of imprinting (LOI),
53 methylation-sensitive high-resolution melting (MS-HRM), odd ratios (ORs), overall survival (OS), peripheral blood
54 leukocytes (PBLs), propensity score (PS), The Cancer Genome Atlas (TCGA).

55 **Word counts:** Main Text 4918 words; Abstract 303 words; Figures 9; Tables 1; References 39; Supplementary Files 6

56 **Keywords:** IGF2 hypermethylation; colorectal neoplasms; risk; prognosis; propensity score analysis

57 **ABSTRACT:** (303 words)

58 **Background:** To comprehensively assess and validate the associations between insulin-like growth
59 factor 2 (*IGF2*) gene methylation in peripheral blood leukocytes (PBLs) and colorectal cancer (CRC)
60 risk and prognosis.

61 **Methods:** The association between *IGF2* methylation in PBLs and CRC risk was initially evaluated
62 in a case-control study and then validated in a nested case-control study and a twins' case-control
63 study, respectively. Meanwhile, an initial CRC patient cohort was used to assess the effect of *IGF2*
64 methylation on CRC prognosis and then the finding was validated in the EPIC-Italy CRC cohort and
65 TCGA datasets. A propensity score (PS) analysis was performed to control for confounders, and
66 extensive sensitivity analyses were performed to assess the robustness of our findings.

67 **Results:** PBL *IGF2* hypermethylation was associated with an increased risk of CRC in the initial
68 study (OR_{PS-adjusted}, 2.57, 95% CI: 1.65 to 4.03, $P < 0.0001$), and this association was validated using
69 two independent external datasets (OR_{PS-adjusted}, 2.21, 95% CI: 1.28 to 3.81, $P = 0.0042$ and
70 OR_{PS-adjusted}, 10.65, 95% CI: 1.26 to 89.71, $P = 0.0295$, respectively). CRC patients with *IGF2*
71 hypermethylation in PBLs had significantly improved overall survival compared to those patients
72 with *IGF2* hypomethylation (HR_{PS-adjusted}, 0.47, 95% CI: 0.29 to 0.76, $P = 0.0019$). The prognostic
73 signature was also observed in the EPIC-Italy CRC cohort, although the HR did not reach statistical
74 significance (HR_{PS-adjusted}, 0.69, 95% CI: 0.37 to 1.27, $P = 0.2359$).

75 **Conclusions:** *IGF2* hypermethylation may serve as a potential blood-based predictive biomarker for
76 the identification of individuals at high risk of developing CRC and for CRC prognosis.

77 **Funding:** This work was supported by the China Postdoctoral Science Foundation (grant number
78 2018M641875 to YPL); the Natural Science Foundation of Heilongjiang Province (grant number
79 YQ2019H021 to YPL); the National Natural Science Foundation of China (grant number 81473055
80 to YSZ), and by grant from the SCORE Foundation (Y-MX2016-045 to YLL).

81 **Clinical trial number:** Not applicable.

82 **MAIN TEXT (4918 words)**

83 **1. INTRODUCTION**

84 Colorectal cancer (CRC) is the third-most common cancer in men and the second-most in women
85 worldwide, with an estimated 1,931,590 newly diagnosed cases and 935,173 deaths in 2020,
86 accounting for approximately 1 in 10 cancer cases and deaths.^[1 2] In China, the number of CRC cases
87 has rapidly increased since the 1980s, with an estimated 555,477 newly diagnosed patients and
88 286,162 deaths in 2020, accounting for approximately 30% of all annually diagnosed CRC cases and
89 CRC-related deaths worldwide.^[2 3] The initiation and progression of CRC is multifactorial and
90 gradual with progressive accumulation of both genetic and epigenetic abnormalities, including
91 aberrant DNA methylation and loss of imprinting (LOI).^[4-6] The insulin-like growth factor 2 (*IGF2*)
92 gene is one of the first imprinted genes identified in humans. The IGF2 protein has a
93 tumour-promoting effect on existing colorectal neoplasia^[7-10] and that LOI of IGF2 in either tissue or
94 peripheral blood leukocyte (PBL) samples is associated with an increased risk of CRC.^[11 12]
95 Furthermore, a recent cohort study showed a significant association between *IGF2* hypomethylation
96 in paraffin-embedded tissues and poor CRC prognosis.^[13] However, no study has evaluated whether
97 PBL IGF2 methylation, which can be determined using non-invasive techniques, is associated with
98 CRC risk or prognosis.

99 We therefore performed this study to comprehensively assess the association between PBL
100 IGF2 methylation status and CRC risk and prognosis. The propensity score (PS) method has been
101 increasingly used to reduce the likelihood of confounding bias in observational studies. It is a
102 powerful statistical tool to control for confounding bias and is often more practical and statistically
103 more efficient than conventional strategies including covariate matching, stratified analysis, and
104 multivariate statistical analysis.^[14 15] In this study, we used PS-based methods to assess the effect of
105 PBL IGF2 methylation on the risk of developing CRC, and then further validated our findings using
106 external datasets from EPIC-Italy CRC cohort and GEO.^[16-18] In addition, we used the same

107 PS-based methods to assess the association between IGF2 methylation and CRC prognosis using
108 CRC patient PBLs and tumour tissues, and then further validated our findings using external datasets
109 from EPIC-Italy CRC cohort and The Cancer Genome Atlas (TCGA).^[17 19]

110 **2. METHODS**

111 **2.1. Study design and participants**

112 **2.1.1. Initial study**

113 *2.1.1.1. IGF2 methylation in PBLs and CRC risk*

114 Descriptions of the study design and patient selection strategy have been published elsewhere.^[20 21]
115 Briefly, this hospital-based case-control study included 281 primary sporadic CRC patients
116 diagnosed (from June 2004 to May 2005 and May 2007 to January 2008) at the Cancer Hospital of
117 Harbin Medical University (HMU) and 147 CRC patients at the Second Affiliated Hospital of HMU
118 (from October 2010 to December 2011) in Harbin, China (**Figure 1**). During the same time period,
119 428 cancer-free controls were selected from the Second Affiliated Hospital of HMU by individual
120 matching on gender and age (± 2 years). The basic characteristics of the CRC cases and controls are
121 shown in **Supplementary File 1A**. All participants provided written informed consent prior to
122 participation in the study. Blood samples were collected before surgery for the CRC patients and
123 before any therapy for the controls. This study was approved by the Ethics Committee of HMU.

124 All CRC patients had operable stage I-IV CRC, and their CRC diagnosis was histologically
125 confirmed by a senior surgeon (YYL). Patients with adenomatous polyposis coli, who had a family
126 history of CRC in first-degree relatives according to the Amsterdam criteria,^[22] or who received
127 anti-cancer therapy before surgery were excluded.

128 *2.1.1.2. IGF2 methylation in PBLs and CRC prognosis*

129 The 281 CRC patients diagnosed at the Cancer Hospital of HMU were all included in the final CRC
130 cohort, while the 147 CRC patients from the Second Affiliated Hospital of HMU were excluded

131 because of the lack of follow-up information. For each patient, extensive demographic,
132 clinicopathological and treatment information were extracted from the electronic medical record
133 system. All surgical operations were performed by the same surgical oncologists (BBC and YLL)
134 and all patients had negative surgical margins. The basic characteristics of the CRC patients included
135 in this cohort are shown in **Supplementary File 1B**.

136 The primary outcomes were overall survival (OS) from CRC diagnosis to death and disease
137 free survival (DFS) from CRC diagnosis to disease recurrence, metastasis, or death from CRC,
138 whichever came first. Another outcome was CRC-specific survival (CSS), which was defined as the
139 time from disease diagnosis to CRC-specific death. Outcomes were observed during the follow-up
140 period through March 15, 2014 via an established protocol. Postoperative patients were followed up
141 at 3-6 months intervals for the first year and then annually. We used a telephone-delivery follow-up
142 questionnaire to collect information on the date and cause of death of the CRC patients. The recorded
143 date and cause of death of the CRC patients were validated using the medical certification of death
144 and the Harbin death registration system. Of the 281 eligible CRC patients included in this analysis,
145 127 died, 120 remained alive and 34 were lost to follow-up.

146 *2.1.1.3. IGF2 methylation in tumour tissues and CRC prognosis*

147 Fresh tumour tissues were collected during surgery from the CRC patients treated at the Cancer
148 Hospital of HMU, with written consent being obtained prior to surgery. Among the 298 eligible CRC
149 patients included in this analysis, 124 died, 141 remained alive and 32 were lost to follow-up. In all,
150 185 paired tumour tissue and PBL samples were obtained from the same patients.

151 *2.1.1.4. Sample size*

152 The sample size was estimated using PASS version 11.0.7 (NCSS LLC., USA). To assess whether
153 aberrant methylation of insulin-like growth factor 2 (IGF2) in peripheral blood leukocytes (PBLs) is
154 associated with the risk of developing colorectal cancer (CRC) in our present case-control study, we

155 estimated the sample size according to a logistic regression model. A sample size of 652 participants
156 was needed to achieve 90% power (at the 5% level of statistical significance) in order to detect an
157 odds ratio (OR) of 1.8 or more with a 20% prevalence in the control group. In addition, taking into
158 consideration incomplete questionnaires and the failure rate for MS-HRM detection, we included
159 approximately 20% more patients and finally targeted a total sample size of 800 participants (n = 400
160 for CRC patients and controls, respectively).

161 To assess whether aberrant IGF2 methylation in PBLs is associated with CRC patient
162 prognosis in our present cohort study, we estimated the sample size according to a Cox regression
163 model. A sample size of 184 CRC patients was needed to achieve 80% power (at the 5% level of
164 statistical significance) to detect a hazard ratio (HR) of 0.5 or 2.0 with an overall event rate of 50%
165 in this cohort. In addition, taking into consideration incomplete questionnaires and the failure rate for
166 MS-HRM detection, we included approximately 20% more patients and finally targeted a total
167 sample size of 221 CRC patients.

168 **2.1.2. Validation study**

169 A case-control study nested in the EPIC-Italy cohort (GSE51032) and a twins' case-control study
170 (GSE89093) from the GEO were used to validate the association between PBL IGF2 methylation
171 and CRC risk. We also used the EPIC-Italy CRC cohort and TCGA datasets (COAD and READ) to
172 validate the relationship between IGF2 methylation in PBLs or tumour tissues and CRC prognosis,
173 respectively. The basic characteristics of validation populations are shown in **Supplementary File 1**.

174 **2.2. Methylation analysis**

175 Sample processing, DNA extraction and bisulfite modification were performed as described
176 previously.^[20] Briefly, DNA was extracted from buffy coats and tumour tissues using a QIAamp
177 DNA Blood Mini Kit (Qiagen, Hilden, Germany, Cat#51106) and a DNeasy Blood & Tissue Kit
178 (Qiagen, Cat #69506), respectively. The DNA was then bisulfite-modified using an EpiTect Plus

179 DNA Bisulfite Kit (Qiagen, Cat#59826) according to the manufacturer's protocols. The
180 bisulfite-modified DNA sample was quantified using a NanoDrop 2000c bioanalyzer (Thermo-Fisher,
181 USA), diluted to a final concentration of 10 ng/ μ L and divided into aliquots for storage (-20 °C).

182 We designed a methylation-sensitive high-resolution melting (MS-HRM) assay for the IGF2
183 promotor region (101bp, GRCh38/hg38; chr11:2139870-2139971, including 11 CpG sites, **Figure 2**)
184 using Methprimer,^[23] and IGF2 methylation status was tested with the researchers being blinded to
185 patient outcome. A set of methylation standards (100, 25, 10, 5, 2, 1, and 0% methylated DNA) were
186 prepared by mixing commercially available methylated and unmethylated DNA (Zymo Research,
187 Irvine, USA, Cat#D5014); these standards were used to semi-quantitatively measure the methylation
188 level of the IGF2 target region in the samples. The MS-HRM analysis was performed as previously
189 described. Briefly, each PCR mixture had a total volume of 10 μ L and contained 2 \times LightCycler 480
190 High Resolution Melting Master Mix (Roche Applied Science, Mannheim, Germany,
191 Cat#4909631001), 3 mmol/L MgCl₂, 0.4 μ mol/L of each primer (forward primer:
192 GGGATTTGGTTGAGGTTTTAAG; reverse primer: TACGACTAAAAAACCCTAAACTC) and
193 1 μ L (approximately 10 ng) of bisulfite-modified template DNA. PCR conditions were as follows:
194 initial PCR activation (95 °C for 15 minutes); 55 3-step cycles (95 °C for 10 seconds, 56 °C for 30
195 seconds, and 72 °C for 20 seconds); and final extension (72 °C for 10 minutes). A blank control
196 (non-template control) sample was included in each batch, and all reactions were performed in
197 duplicate (as technical replicates). A third trial was conducted for the samples that presented
198 inconsistent results between the two trials. PCR amplification and MS-HRM analyses were
199 performed using the LightCycler 480 platform (Roche). After normalization of the melting curves
200 using the software module of Gene Scanning (Roche), two investigators (Y.P.L. and H.R.S.), who
201 were blinded to the sample groups, independently assessed the MS-HRM data. Discrepancies were
202 resolved by discussion and consensus with another investigator (Y.B.N.W.).

203 **2.3. Covariates, Missing data analysis and imputation**

204 All participants were interviewed face-to-face to complete a structured standard questionnaire, which
205 was partially adopted from Shu et al.^[24] The questionnaire queried information on demographic
206 characteristics, lifestyle factors (including family history, smoking, alcohol consumption,
207 occupational physical activity), and diet. Dietary consumption over the past year was assessed using
208 a validated food frequency questionnaire (FFQ).^[24] The FFQ included nine major food groups, which
209 represented most of the common foods in Northeast China. The food items included barbecued foods,
210 coarse grains, fish stewed with brown sauce, fresh fruits, fried foods, green vegetables, leftovers and
211 pork.

212 Questionnaire-derived covariates included: age (< 60, ≥ 60), gender (male, female), BMI (< 24,
213 ≥ 24), family history of cancer other than CRC in first-degree relatives (no, yes), occupational
214 physical activity (blue-collar, white-collar), smoking status (no, yes), and consumption of barbecued
215 foods (< 1, ≥ 1 times/week), coarse grains (< 50, ≥ 50 g/week), fish stewed with brown sauce (< 1, ≥
216 1 times/week), fruits (< 2, ≥ 2 times/week), fried foods (< 1, ≥ 1 times/month), green vegetables (<
217 100, ≥ 100 g/day), leftovers (< 1, ≥ 1 times/week) and pork (< 250, ≥ 250 g/week). All
218 questionnaire-derived variables were analysed via missing value analysis and were imputed via
219 multiple imputations as described in our previous study.^[20]

220 **2.4. Statistical analysis**

221 Means and standard deviations were reported for continuous variables, while counts and frequencies
222 were reported for categorical variables. Covariate differences between groups were compared using
223 the standardised differences method with a significant imbalance level of standardised difference ≥
224 25%. In the CRC risk analysis, we first categorised individuals into two groups according to the
225 optimal cut-off point for the IGF2 methylation level (≤ 1% hypomethylation group; > 1%
226 hypermethylation group), which was determined using the receiver operating characteristic (ROC)
227 curve and the Youden index method with case-control status as the dependent variable (0 for controls;
228 1 for cases). We then conducted univariate and multivariate logistic regression analyses and reported

229 odd ratios (ORs) and 95% confidence intervals (CIs) to assess the association between IGF2
230 methylation status and CRC risk. In the CRC prognosis analysis, the cut-off point for the IGF2
231 methylation level was also 1% using the same method with overall survival time as the dependent
232 variable (0 for less than median survival time; 1 for longer than or equal to median survival time).
233 According to this cut-off point, CRC patients were categorised into IGF2 hypomethylation and IGF2
234 hypermethylation groups (206 and 75 cases, respectively). Kaplan-Meier curves and log-rank tests
235 were used to compare OS, DFS or CSS between groups. The associations between IGF2 methylation
236 and OS, DFS or CSS were estimated using univariate and multivariate Cox regression models and
237 reported as hazard ratios (HRs) and 95% CIs. Two-sided statistical significance was defined as
238 $P < 0.05$. ROC analyses were performed with MedCalc version 15.4 (Ostend, Belgium) and all other
239 statistical analyses were performed with SPSS Statistics version 23.0 (IBM, Inc., USA).

240 To minimise group differences on covariates, we performed a PS-based analysis. In the CRC
241 risk analysis, the PS was calculated with case-control status as the dependent variable using a
242 multivariate logistic regression model that included demographic and lifestyle factors
243 (**Supplementary File 2A**). In the survival analysis, the PS was calculated with IGF2 methylation as
244 the dependent variable using a multivariate logistic regression model that included demographic and
245 lifestyle factors and clinicopathological characteristics (**Supplementary File 2B**). To incorporate all
246 the patients in the analyses, we primarily employed the PS-adjustment method.^[14] Additionally,
247 several PS-based methods, including stratification by quintile of PS, weighting with inverse
248 probability of treatment weights, and individual PS matching, were also performed as sensitivity
249 analyses.

250 For PS-stratification analyses, all patients were stratified by PS quintile. HRs for PS-stratified
251 analyses were obtained by pooling effect estimates from each PS quintile. For PS-weighted analyses,
252 each patient was weighted by the inverse probability of being in the higher versus lower IGF2
253 methylation groups; this method is known as weighting with inverse probability of treatment weights

254 (IPTW). The weight for each patient was calculated according to the method described by Robins
255 and Hernan:^[25] Pt/PS and $(1-Pt)/(1-PS)$ for patients in the IGF2 hypermethylation and
256 hypomethylation groups, respectively; where Pt is the proportion of the patients with IGF2
257 hypermethylation in all participants. In addition to using PS as an adjustment or weighting factor in
258 the analysis model, PS was also used for individual matching. PS matching was performed using a
259 nearest-neighbour algorithm with a calliper of 0.2 (which means that the maximum allowable PS
260 difference between groups was no larger than 20% of the standard deviation of PS).

261 **2.5. Sensitivity analysis**

262 To assess the robustness of our results, we performed extensively predefined sensitivity analyses.
263 First, to evaluate the potential impact of the PS-adjusted confounders on our results in the initial
264 study, we performed confounding RR analysis, which was defined as the ratio of the PS-adjusted
265 effect estimates and the unadjusted effect estimates.^[26] To investigate whether potential residual
266 confounders could impact the results, we calculated the E-value for PS-adjusted effect estimates and
267 the limit of the CI closest to the null.^[27] Finally, we performed subgroup analyses according to the
268 tumour location (colon or rectum), UICC stage, gender (female vs. male), age (≥ 60 vs. < 60 years),
269 and body mass index (BMI, ≥ 24 vs. < 24).

270 Additionally, we performed several post hoc sensitivity analyses. To test whether tumour load
271 impacts IGF2 methylation level in PBLs, we assessed the associations between IGF2 methylation
272 and UICC stage and serum carcinoembryonic antigen (CEA) level. We also tested whether PBL
273 IGF2 methylation levels could have been impacted by leukocyte count and the percentage of certain
274 subpopulations. In addition, to comprehensively determine the association between IGF2
275 methylation and CRC risk, we combined the results from the initial case-control and the validation
276 studies using a meta-analysis method of random effect model. Finally, we explored whether PBL
277 IGF2 methylation levels differ between the CRC patients included in the survival analysis and those
278 excluded.

279 **3. RESULTS**

280 **3.1. IGF2 methylation in PBLs and CRC risk**

281 As shown in **Figure 3A**, subjects with IGF2 hypermethylation (15.54%), compared to subjects with
282 IGF2 hypomethylation (84.46%), had a significantly increased CRC risk (PS-adjusted OR, 2.57, 95%
283 CI: 1.64-4.03; $P < 0.001$). This association remained statistically significant even after subgroup
284 analyses. Notably, stratified analyses showed a positive association only in UICC stage I-III but not
285 in stage IV cancers.

286 In the EPIC-Italy dataset, IGF2 hypermethylation was also significantly associated with an
287 increased CRC risk (OR, 2.21, 95% CI: 1.28-3.81; $P = 0.004$). Subgroup analyses showed the
288 association was statistically significant in distal colon cancer, men, and older subjects (**Figure 3A**).
289 Using the GSE89093 dataset, IGF2 hypermethylation was strongly associated with an increased risk
290 of CRC or colon cancer alone. After pooling the results from these three datasets, the observed
291 association was still significant (**Figure 3B**).

292 **3.2. IGF2 methylation in PBLs and CRC prognosis**

293 The Kaplan-Meier survival curves for the PBL IGF2 hypermethylation and hypomethylation groups
294 are shown in **Figure 4**. The median OS was 73 months (95% CI: 66-80 months) in the IGF2
295 hypermethylation group versus 56 months (95% CI: 52-61 months) in the hypomethylation group.
296 The OS rate was 75% (95% CI: 60-80%) in the IGF2 hypermethylation group versus 62% (95% CI:
297 50-57%) in the hypomethylation group ($HR_{PS-adjusted}$, 0.47, 95% CI: 0.29-0.76; $P = 0.002$). The CSS
298 rate was 56% (95% CI: 45-68%) in the IGF2 hypermethylation group versus 38% (95% CI: 31-45%)
299 in the hypomethylation group ($HR_{PS-adjusted}$, 0.49, 95% CI: 0.30-0.80; $P = 0.004$). The median DFS was
300 66 months (95% CI: 57-74 months) in the IGF2 hypermethylation group versus 52 months (95% CI:
301 48-57 months) in the hypomethylation group ($HR_{PS-adjusted}$, 0.53, 95% CI: 0.32-0.85; $P = 0.009$). Based
302 on subgroup analyses, we found that the statistically significant association persisted in rectal cancers,

303 UICC stage I-III cancers, males, older, or normal body weight patients. In contrast, the effect
304 estimates did not reach statistical significance in colon cancers, stage IV cancers, females, younger,
305 and overweight or obese patients (**Table 1**).

306 Using the follow-up results of the 166 CRC patients from the EPIC-Italy cohort, we observed a
307 clear trend for a longer OS related with IGF2 hypermethylation, although the association was not
308 statistically significant. The univariate HR was 0.65 (95% CI: 0.35-1.19; $P=0.164$), and after PS
309 adjustment, the $HR_{PS\text{-adjusted}}$ was 0.69 (95% CI: 0.37-1.27; $P=0.236$; K-M survival curves are shown
310 in **Figure 5**).

311 **3.3. IGF2 methylation in tissues and CRC prognosis**

312 In our initial study, the association between IGF2 methylation in tumour tissues and CRC patient
313 survival did not reach statistical significance (**Figure 6A**), and this result was confirmed by the
314 TCGA dataset (**Figure 6B**). Using IGF2 mRNA expression data in tumour tissues, we found an
315 obvious negative correlation with IGF2 methylation ($r=-0.24$, $P<0.001$; **Figure 6C**). So we further
316 explored whether IGF2 mRNA expression levels impact the disease prognosis but found no
317 statistically significant association (**Figure 6D**).

318 **3.4. Sensitivity analysis**

319 We repeated analyses using other PS-based methods and the results were consistent with the
320 PS-adjustment results (**Supplementary File 3**). Based on the confounding RR analysis (**Figure 7**),
321 we did not find any substantial differences between the PS-adjusted effect estimates and the
322 corresponding unadjusted effect estimates. Furthermore, the E-value analysis showed that our
323 findings in both the initial and validation studies appear to be very robust (**Supplementary File 4**).

324 We found no evidence for an obvious impact of UICC stages on IGF2 methylation in PBLs. Of
325 note, IGF2 hypermethylation was significantly more frequent in the CEA-low group than in the
326 CEA-high group (34.15% vs. 20.89%, $P=0.013$). We found no significant relationship between the

327 PBL IGF2 methylation levels and the leukocyte count or the percentages of certain leukocyte
328 subpopulations. Finally, we analysed samples from 147 additional CRC patients that were excluded
329 from the survival analysis and found no differences in their PBL IGF2 methylation levels compared
330 to those of the 281 CRC patients included in the survival analysis (**Supplementary File 5**).

331 **4. DISCUSSION**

332 In this study, we first assessed the impact of the PBL IGF2 methylation status on the risk and
333 prognosis of CRC. We found that individuals with IGF2 hypermethylation in their PBLs were at
334 significantly higher risk of developing CRC than those with IGF2 hypomethylation. However, our
335 findings showed better survival rates in CRC patients with PBL IGF2 hypermethylation compared to
336 those in CRC patients with IGF2 hypomethylation.

337 In our initial case-control study, it is not possible to determine the aetiologically relevant time
338 window of IGF2 methylation relative to CRC development. Fortunately, the prospective nature of the
339 EPIC-Italy cohort is invaluable for confirming the temporal sequence of DNA methylation and CRC
340 onset and therefore helps to distinguish causal from consequential epigenetic changes.^[28 29] In this
341 nested case-control study, blood samples from CRC patients were collected 0.02-14.40 years
342 (average 6.16 years) before diagnosis. Using this dataset, we found a positive but non-significant
343 association between PBL IGF2 hypermethylation and an increased CRC risk. To account for reverse
344 causality, we repeated the analyses excluding subjects who developed CRC within two years after
345 their blood sampling and found that this positive association remained significant (PS-adjusted OR,
346 2.08, 95% CI: 1.18-3.69; $P=0.012$). Furthermore, our main finding was verified using the GSE89093
347 dataset. Finally, we pooled the results from these three datasets and found a 2.19-fold higher risk of
348 developing CRC in the IGF2 hypermethylation group compared to the hypomethylation group.

349 The exact mechanism that links alterations in the IGF2 methylation status in PBLs and the
350 susceptibility for CRC remains unclear. The PBL-derived DNA methylation profiles represent the

351 overall methylation status of an individual. Alterations in PBL-derived DNA may be constitutively
352 present before cancer occurs or represent an early response of the haematology system to the
353 presence of tumour cells after cancer develops. Imprinting of IGF2 is primarily maintained by DNA
354 methylation.^[30 31] It has been reported that imprinting and expression are controlled by CpG-rich
355 regions upstream of the IGF2 promoters. Normally, IGF2 is expressed from the paternal allele only,
356 while the maternal allele is methylated and imprinted. Several previous studies have reported that the
357 loss of imprinting (LOI) of IGF2 in peripheral blood (leukocytes or lymphocytes) may be a potential
358 biomarker for the risk of developing CRC,^[12 32] and IGF2 methylation alterations were suggested as
359 a surrogate marker for LOI of IGF2.^[32 33] If IGF2 aberrant methylation is merely a surrogate marker
360 for LOI of IGF2, then IGF2 hypermethylation would be expected to be beneficial for maintaining the
361 imprinting status of IGF2 and thus should decrease the risk of developing CRC. On the contrary, our
362 results showed that IGF2 hypermethylation increases the risk of developing CRC, indicating that
363 IGF2 participates in CRC tumorigenesis through two different modes of epigenetic alteration,
364 aberrant hypermethylation and LOI, which is supported by previous studies.^[34 35] Thus, our findings
365 demonstrate that aberrant IGF2 hypermethylation can be assayed with a blood-based test and that the
366 PBL IGF2 methylation status is likely to be a valuable predictive biomarker for CRC risk,
367 independent of LOI of IGF2.

368 Further analysis of the GEO and TCGA datasets provided additional evidence supporting the
369 association between IGF2 hypermethylation and the risk of developing CRC. For example, using
370 these datasets we found significant associations between IGF2 hypermethylation in tissues and an
371 increased risk of CRC or adenomas (**Figures 8A-D**). Interestingly, colorectal adenomas, a
372 precancerous condition, showed similar levels of IGF2 methylation compared to tumour tissues
373 (**Figure 8E**). Most colorectal cancers initially develop as benign precursor lesions (adenomas) that
374 can take as long as 10 to 15 years to develop into carcinomas, which underpins early detection and
375 removal of adenomas as an important strategy for preventing CRC.^[6 36] The findings from the GEO

376 and TCGA datasets indicate that IGF2 methylation in tissues can discern colorectal adenomas or
377 CRCs from normal intestinal mucosa, which also suggests that IGF2 methylation may prove valuable
378 during CRC screening for early cancer detection. We were unable to evaluate the association
379 between IGF2 methylation in PBLs and the risk of developing adenomas. This issue should be
380 explored in future studies.

381 In 2008, Ito and colleagues used peripheral blood samples to assess six CpG sites located in
382 the IGF2 gene and reported that IGF2 methylation is not statistically associated with CRC risk.^[34]
383 However, the region tested in Ito's study is different from the region examined in our study. As
384 shown in **Figure 2**, the differentially methylated region assessed in our study is located near the TSS.
385 To further validate our findings, we used the EPIC-Italy dataset to evaluate the methylation status of
386 a relatively large region of -1500 bp upstream to +1500 bp downstream from the TSS. This
387 sensitivity analysis again showed a statistically significant association between IGF2
388 hypermethylation and CRC risk (OR, 1.75, 95% CI: 1.15-2.66; $P=0.009$).

389 For the first time, we assessed the prognostic value of PBL IGF2 methylation status in CRC
390 patients and found that patients with IGF2 hypermethylation in PBLs had significantly improved
391 survival compared to patients with IGF2 hypomethylation. These findings were especially obvious
392 for stage I-III CRC patients. However, this association did not reach statistical significance in the
393 external EPIC-Italy CRC cohort. Given these inconsistent results, further cohorts with large sample
394 size are needed to validate this novel finding.

395 For sensitivity analysis, we also evaluated the association between PBL IGF2 methylation and
396 CRC prognosis during different follow-up periods. We observed a significantly strong association for
397 5-year relative survival, while the associations for both 1-year and 3-year relative survival were not
398 statistically significant. Based on these results, it is hypothesised that the beneficial effects of IGF2
399 hypermethylation on CRC survival might begin the fifth year following CRC diagnosis and might
400 persist for several years. In fact, the potentially beneficial effects may occur as early as the first year

401 after CRC diagnosis, even though this effect did not reach statistical significance. The limited sample
402 size of the cohort study may hinder the interpretation of these results; therefore, larger cohort studies
403 are required to further evaluate this issue.

404 We also evaluated the impact of IGF2 methylation in tumour tissues on CRC patient survival
405 and found no statistically significant association in both our initial cohort and the TCGA datasets.
406 This is in consistent with a recent research.^[7] However, two previous studies from Japan have
407 assessed the IGF2 methylation status in tumour tissues of CRC patients and the results were
408 inconsistent.^[13 37] The region tested in those two studies is different from that examined in our study.
409 The lack of consistency between the results from the PBL samples and tumour tissues may reflect the
410 fact that IGF2 methylation is merely a predictive marker rather than a prognostic marker. In addition,
411 the detection of tissue-based markers depends on material from a biopsy or tumour tissue from
412 resection. Because of intratumour heterogeneity, the detection of a biomarker from a single biopsy or
413 one section of a tumour tissue sample might not necessarily represent the IGF2 methylation status of
414 a given patient. Repeated biopsies and tests of multiple samples, however, are not feasible in routine
415 clinical practice. Fortunately, blood-derived biomarkers have the potential to overcome these
416 problems. In this respect, repeat blood sampling and detection of PBL IGF2 methylation is more
417 acceptable and feasible than repeat biopsies in the clinic. Given that CRC has a wide range of
418 long-term outcomes, PBL IGF2 methylation, as a DNA-based non-invasive blood test, could prove
419 beneficial during follow-up and help identify patients at high risk of disease recurrence and
420 progression.

421 An important aspect and potential concern of using PBL DNA methylation as a biomarker is
422 whether leukocyte subpopulations affect the methylation signature of an individual. To address this
423 concern, we collected patient clinical records including leukocyte counts and included these data as
424 covariates in the PS model. Theoretically, the PS adjustment would control for the potential impact
425 of different leukocyte counts and different subpopulations on our results. We compared the results

426 before and after including the leukocyte count data in the PS model and found similar results
427 (**Supplementary File 6**), suggesting that the effect of leukocyte counts and subpopulations on our
428 results is negligible. Additional evidence supporting these findings can be found in several recently
429 published studies which also showed that differences in leukocyte subpopulations were unlikely to
430 interfere with the results of PBL-derived DNA methylation.^[38 39]

431 Interestingly, our data indicates that PBL IGF2 hypermethylation correlates with serum CEA
432 levels before surgery. Therefore, we further assessed whether the association between PBL IGF2
433 methylation and CRC prognosis is impacted by CEA levels. Using the additive model, we found a
434 positive interaction between PBL IGF2 methylation and CEA on CRC prognosis (**Figure 9**). Given
435 the limited sample size of the subgroups, this aspect should be further validated in future studies. Of
436 note, after adjustment for CEA and the interaction between IGF2 methylation and CEA, the effect of
437 IGF2 hypermethylation itself on CRC prognosis remained statistically significant (HR, 0.44, 95% CI:
438 0.24-0.79; $P=0.006$), suggesting a robust and independent role for PBL IGF2 hypermethylation in
439 predicting the prognosis of CRC.

440 This study has several strengths. The findings from our initial studies were validated with
441 several external datasets. In our initial CRC patient cohort, the covariates were collected
442 prospectively and were blinded to patient outcome. We used PS techniques to control for multiple
443 potential confounding factors. Furthermore, we performed extensive sensitivity analyses to assess the
444 robustness of our findings. The confounding RR and the E-value sensitivity analyses demonstrated
445 that our results are unlikely to be substantially impacted by both the adjusted confounders included
446 in the PS models or a potential residual confounder.

447 In our present cohort, adjuvant chemotherapy was not offered routinely to high risk individuals.
448 Thus, we did not analyse the clinical significance of IGF2 methylation as a predictive biomarker for
449 sensitivity to adjuvant chemotherapy. Further studies are needed to explore and clarify this issue.
450 Another potential limitation is the limited sample size used in the stratified analyses. Therefore, the

451 results from the stratified analyses should be interpreted with caution.

452 5. CONCLUSIONS

453 In summary, IGF2 methylation in PBLs is significantly associated with the risk and prognosis of
454 CRC, suggesting an important role for IGF2 methylation as a blood-based predictive biomarker to
455 identify of individuals at high risk of developing CRC; meanwhile, PBL IGF2 methylation might
456 also serve as a predictive biomarker for CRC prognosis.

457 REFERENCES

- 458 1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and
459 Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-49.
- 460 2. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2018). Global
461 Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer.
462 <https://gco.iarc.fr/today/home> (Accessed November 7, 2021).
- 463 3. Cao W, Chen HD, Yu YW, et al. Changing profiles of cancer burden worldwide and in China: a secondary analysis of
464 the global cancer statistics 2020. *Chin Med J (Engl)* 2021;134:783-91.
- 465 4. Dekker E, Tanis PJ, Vleugels JLA, et al. Colorectal cancer. *Lancet* 2019;394:1467-80.
- 466 5. Jones PA, Baylin SB. The epigenomics of cancer. *Cell* 2007;128:683-92.
- 467 6. Kuipers EJ, Grady WM, Lieberman D, et al. Colorectal cancer. *Nat Rev Dis Primers* 2015;1:15065.
- 468 7. Hidaka H, Higashimoto K, Aoki S, et al. Comprehensive methylation analysis of imprinting-associated differentially
469 methylated regions in colorectal cancer. *Clin Epigenetics* 2018;10:150.
- 470 8. Unger C, Kramer N, Unterleuthner D, et al. Stromal-derived IGF2 promotes colon cancer progression via paracrine
471 and autocrine mechanisms. *Oncogene* 2017;36:5341-55.
- 472 9. Rogers MA, Kalter V, Strowitzki M, et al. IGF2 knockdown in two colorectal cancer cell lines decreases survival,
473 adhesion and modulates survival-associated genes. *Tumour Biol* 2016;37:12485-95.
- 474 10. Bergman D, Halje M, Nordin M, et al. Insulin-like growth factor 2 in development and disease: a mini-review.
475 *Gerontology* 2013;59:240-9.
- 476 11. Cheng YW, Idrees K, Shattock R, et al. Loss of imprinting and marked gene elevation are 2 forms of aberrant IGF2
477 expression in colorectal cancer. *Int J Cancer* 2010;127:568-77.
- 478 12. Cui H, Cruz-Correa M, Giardiello FM, et al. Loss of IGF2 imprinting: a potential marker of colorectal cancer risk.
479 *Science* 2003;299:1753-5.
- 480 13. Baba Y, Noshio K, Shima K, et al. Hypomethylation of the IGF2 DMR in colorectal tumors, detected by bisulfite
481 pyrosequencing, is associated with poor prognosis. *Gastroenterology* 2010;139:1855-64.
- 482 14. Elze MC, Gregson J, Baber U, et al. Comparison of Propensity Score Methods and Covariate Adjustment:
483 Evaluation in 4 Cardiovascular Studies. *J Am Coll Cardiol* 2017;69:345-57.
- 484 15. Haukoos JS, Lewis RJ. The Propensity Score. *JAMA* 2015;314:1637-8.
- 485 16. Gene Expression Omnibus. Available at: <https://www.ncbi.nlm.nih.gov/geo/>. Accessed May 22, 2020.
- 486 17. Cordero F, Ferrero G, Polidoro S, et al. Differentially methylated microRNAs in prediagnostic samples of subjects
487 who developed breast cancer in the European Prospective Investigation into Nutrition and Cancer (EPIC-Italy)
488 cohort. *Carcinogenesis* 2015;36:1144-53.

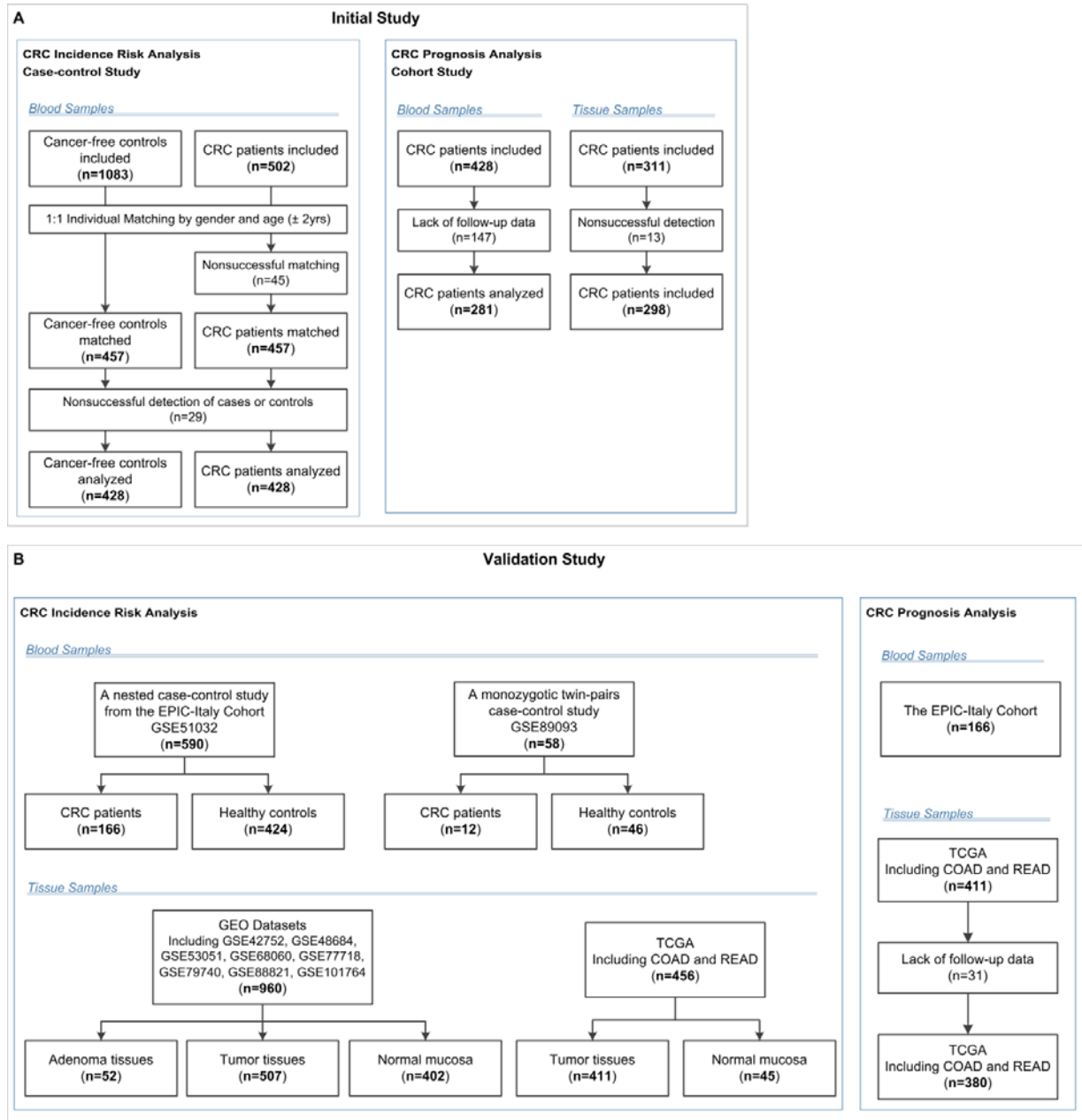
- 489 18. Roos L, van Dongen J, Bell CG, et al. Integrative DNA methylome analysis of pan-cancer biomarkers in cancer
490 discordant monozygotic twin-pairs. *Clin Epigenetics* 2016;8:7.
- 491 19. The Cancer Genome Atlas Program. Available at: <https://cancergenome.nih.gov/>. Accessed February 4, 2020.
- 492 20. Liu Y, Wang Y, Hu F, et al. Multiple gene-specific DNA methylation in blood leukocytes and colorectal cancer risk:
493 a case-control study in China. *Oncotarget* 2017;8:61239-52.
- 494 21. Sun H, Huang H, Li D, et al. PBX3 hypermethylation in peripheral blood leukocytes predicts better prognosis in
495 colorectal cancer: A propensity score analysis. *Cancer Med* 2019;8:4001-11.
- 496 22. Vasen HF, Watson P, Mecklin JP, et al. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC,
497 Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology*
498 1999;116:1453-6.
- 499 23. Li LC, Dahiya R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 2002;18:1427-31.
- 500 24. Shu XO, Yang G, Jin F, et al. Validity and reproducibility of the food frequency questionnaire used in the Shanghai
501 Women's Health Study. *Eur J Clin Nutr* 2004;58:17-23.
- 502 25. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology.
503 *Epidemiology* 2000;11:550-60.
- 504 26. Greenland S, Pearce N. Statistical foundations for model-based adjustments. *Annu Rev Public Health*
505 2015;36:89-108.
- 506 27. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med*
507 2017;167:268-74.
- 508 28. Langevin SM, Houseman EA, Accomando WP, et al. Leukocyte-adjusted epigenome-wide association studies of
509 blood from solid tumor patients. *Epigenetics* 2014;9:884-95.
- 510 29. Rakyan VK, Down TA, Balding DJ, et al. Epigenome-wide association studies for common human diseases. *Nat*
511 *Rev Genet* 2011;12:529-41.
- 512 30. Ransohoff DF. Cancer. Developing molecular biomarkers for cancer. *Science* 2003;299:1679-80.
- 513 31. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene.
514 *Nature* 2000;405:482-5.
- 515 32. Cruz-Correa M, Cui H, Giardiello FM, et al. Loss of imprinting of insulin growth factor II gene: a potential
516 heritable biomarker for colon neoplasia predisposition. *Gastroenterology* 2004;126:964-70.
- 517 33. Cui H. Loss of imprinting of IGF2 as an epigenetic marker for the risk of human cancer. *Dis Markers*
518 2007;23:105-12.
- 519 34. Ito Y, Koessler T, Ibrahim AE, et al. Somatically acquired hypomethylation of IGF2 in breast and colorectal cancer.
520 *Hum Mol Genet* 2008;17:2633-43.
- 521 35. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*
522 2012;487:330-7.
- 523 36. Bosman FT. Cancer by organ site: Colorectal cancer. In: Bernard W. Stewart CPW, ed. World Cancer Report 2014.
524 *International Agency for Research on Cancer, WHO* 2014:560-576.
- 525 37. Ohta M, Sugimoto T, Seto M, et al. Genetic alterations in colorectal cancers with demethylation of insulin-like
526 growth factor II. *Hum Pathol* 2008;39:1301-8.
- 527 38. Lonning PE, Berge EO, Bjornstlett M, et al. White Blood Cell BRCA1 Promoter Methylation Status and Ovarian
528 Cancer Risk. *Ann Intern Med* 2018;168:326-34.
- 529 39. Heiss JA, Brenner H. Impact of confounding by leukocyte composition on associations of leukocyte DNA
530 methylation with common risk factors. *Epigenomics* 2017;9:659-68.

531

532 **Figures**

533 **Figure 1.** Flow chart of participants included and analysed in the (A) initial and (B) validation
 534 studies. COAD, colon adenocarcinoma; CRC, colorectal cancer; EPIC-Italy, the European
 535 Prospective Investigation into Cancer and nutrition (EPIC cohort) in Italy; GEO, Gene Expression
 536 Omnibus; READ, rectum adenocarcinoma; TCGA, The Cancer Genome Atlas.

537



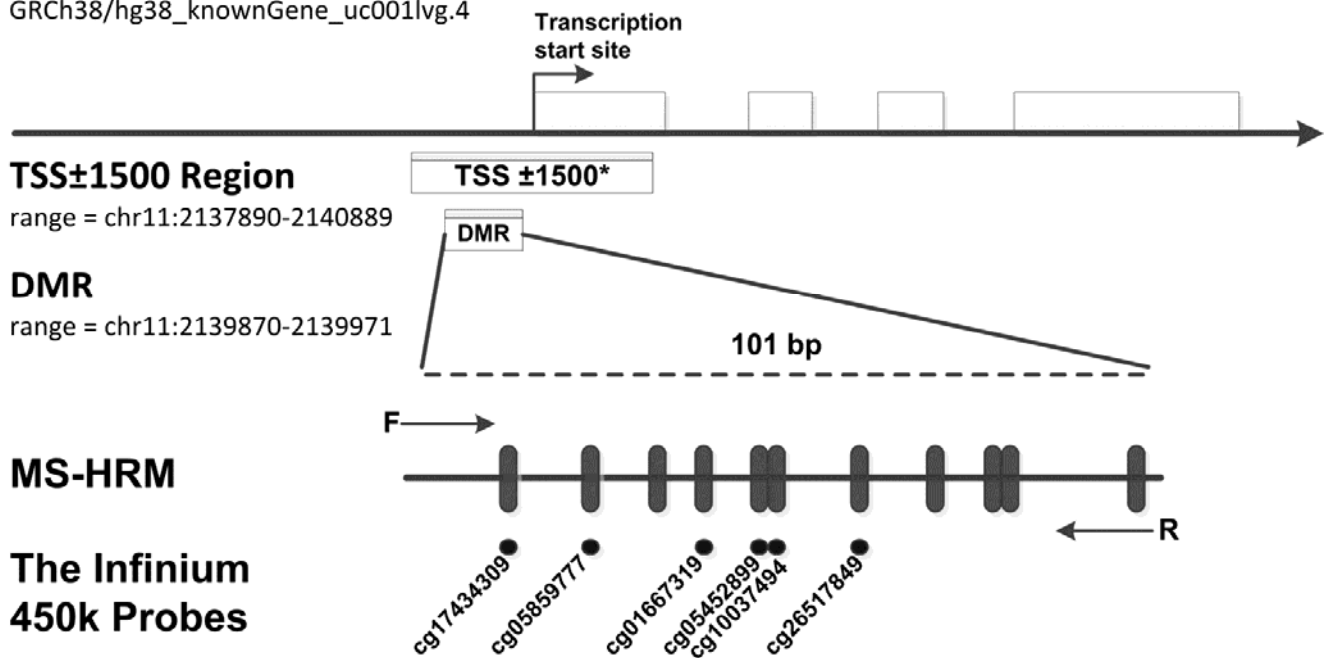
538

539 **Figure 2.** MS-HRM assay to detect methylation levels at the IGF2 promoter region. Our assay encompassed
540 a 101 bp region (range: chr11: 2139870-2139971, including 11 CpGs), which located in the promoter of
541 human IGF2 gene and near the transcription start site (TSS). Arrows indicate position and direction of
542 MS-HRM primers. **Black dashes (|)** indicate individual CpGs on DMR region tested in our study. **Black**
543 **Dots (●)** indicate six methylation probes (including cg17434309, cg05859777, cg01667319, cg05452899,
544 cg10037494, cg26517849) on the Illumina Infinium HumanMethylation450 BeadChip, which overlap CpG
545 sites detected in our DMR region. *The TSS±1500 region is an expanded region of -1500 bp upstream to
546 +1500 downstream from the TSS (range: chr11: 2137890-2140889). To test the robustness of our results, we
547 further assess the association of methylation levels of this expanded region with CRC risk. In the TSS±1500
548 region, a total of 32 CpG sites were detected by the Infinium HumanMethylation450 methylation probes.
549 The methylation probes are as follows: cg25163476, cg24917382, cg24366657, cg23905216, cg23676551,
550 cg22287492, cg21667878, cg20339650, cg19443075, cg19371526, cg19002337, cg18087943, cg17434309,
551 cg17037101, cg16415340, cg15393937, cg14608156, cg14188639, cg13756879, cg12773325, cg12614029,
552 cg11915650, cg10659464, cg10037494, cg09694722, cg08162473, cg05859777, cg05452899, cg04072545,
553 cg03760951, cg02835822, cg01667319.

554

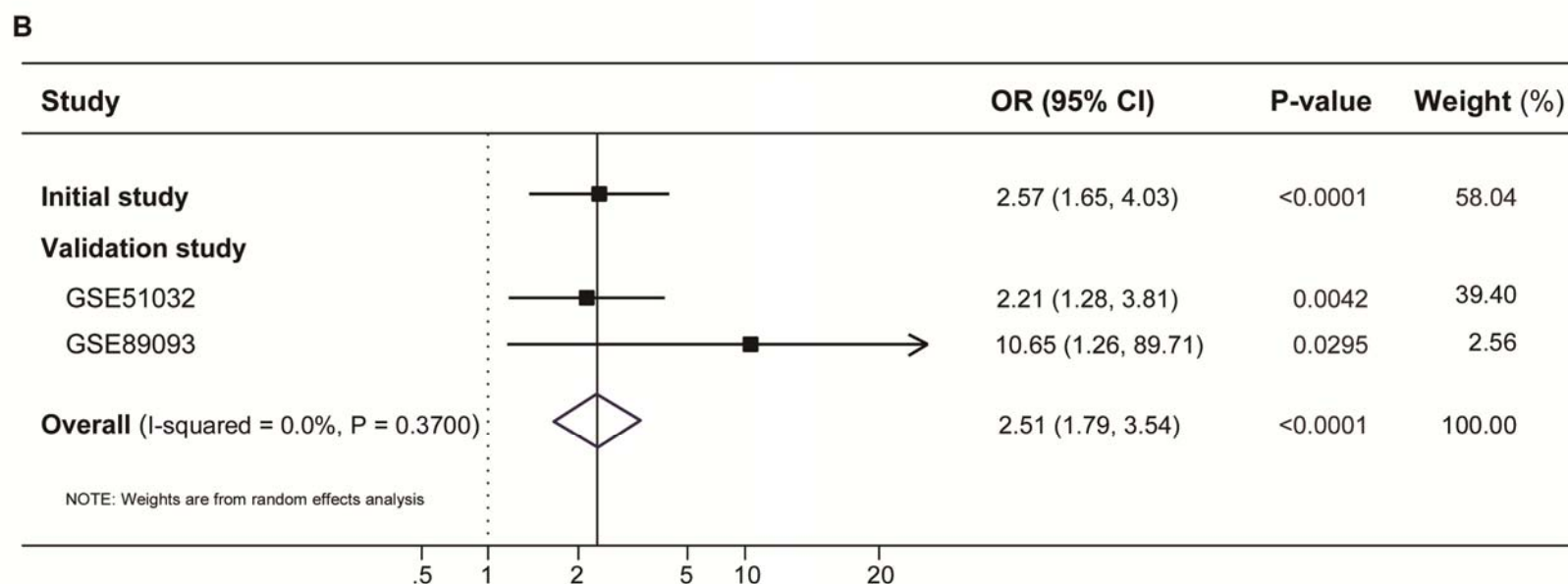
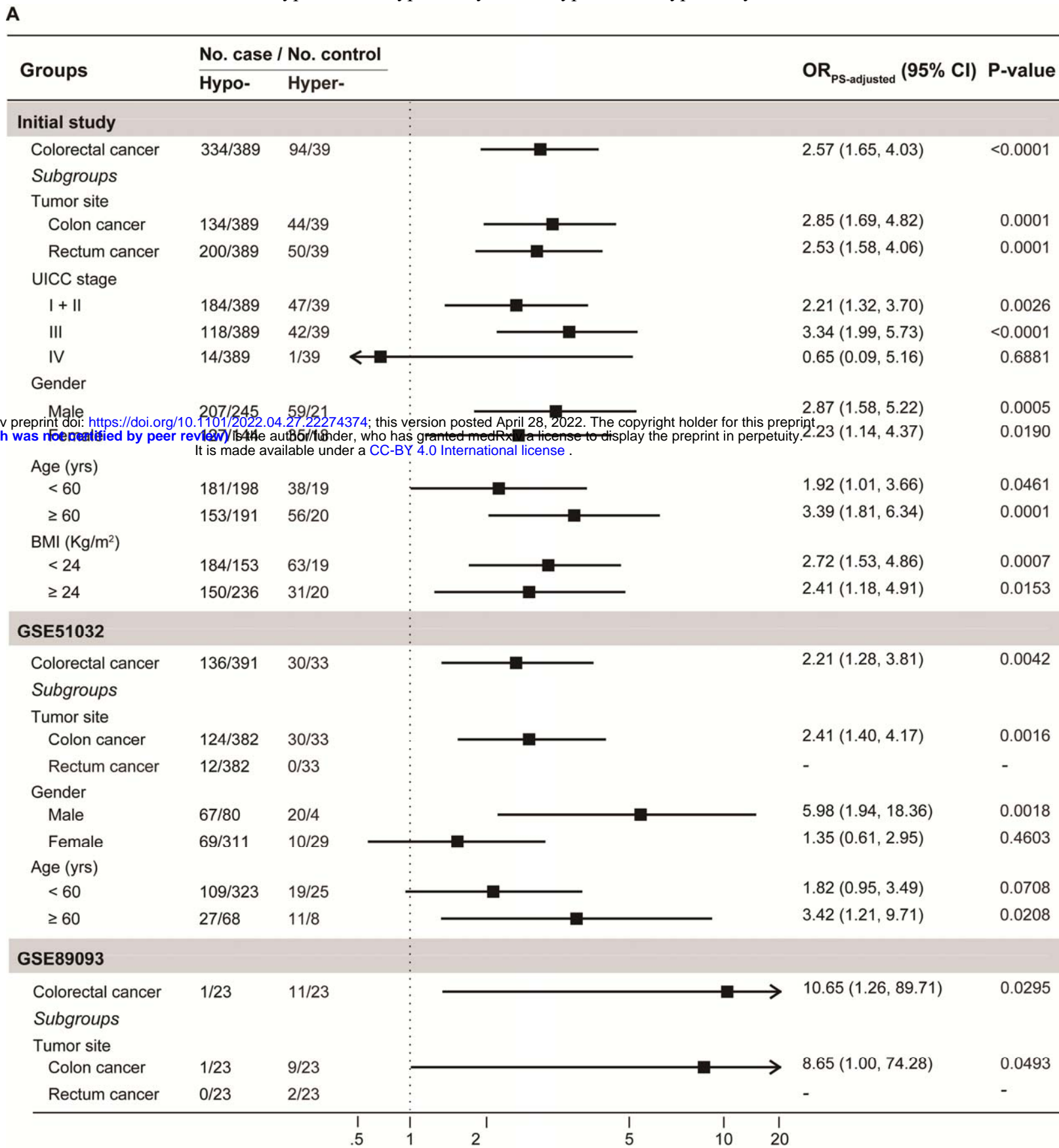
IGF2

GRCh38/hg38_knownGene_uc001lv.4



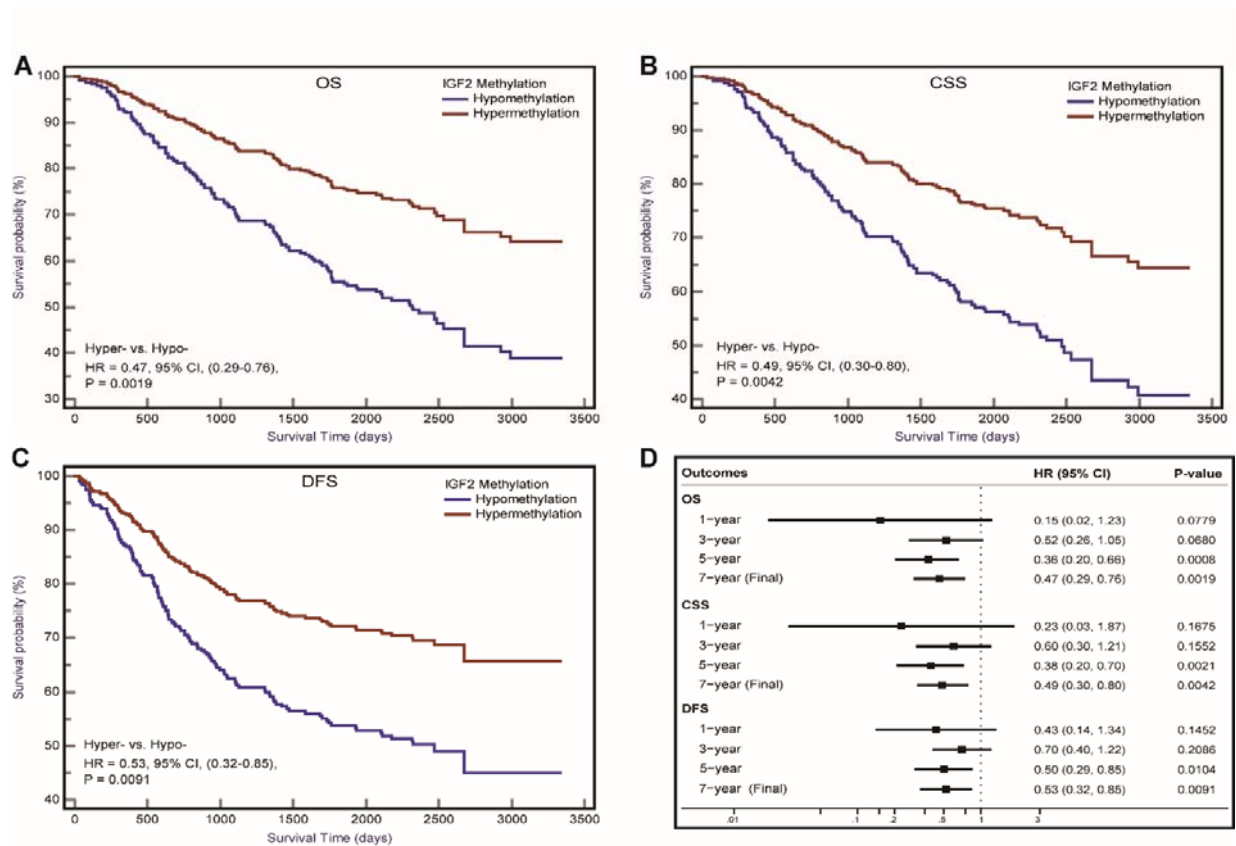
555

556 **Figure 3.** The results for the association between PBL IGF2 methylation and the risk of developing CRC (A) in the initial and validation
 557 studies, and (B) in the combined populations.
 558 *A total of 14 covariates were included in the propensity score model in the initial case-control study (see supplementary table S3). In the
 559 validation studies, gender and age were included in the the propensity score model for GSE51032 and age for GSE89093, respectively.
 560 To be conservative, we pooled the results from these three datasets with a random effect model in the meta-analysis. BMI, body mass
 561 index; CI, confidence interval; Hyper-, IGF2 hypermethylation; Hypo-, IGF2 hypomethylation; OR, odds ratio.



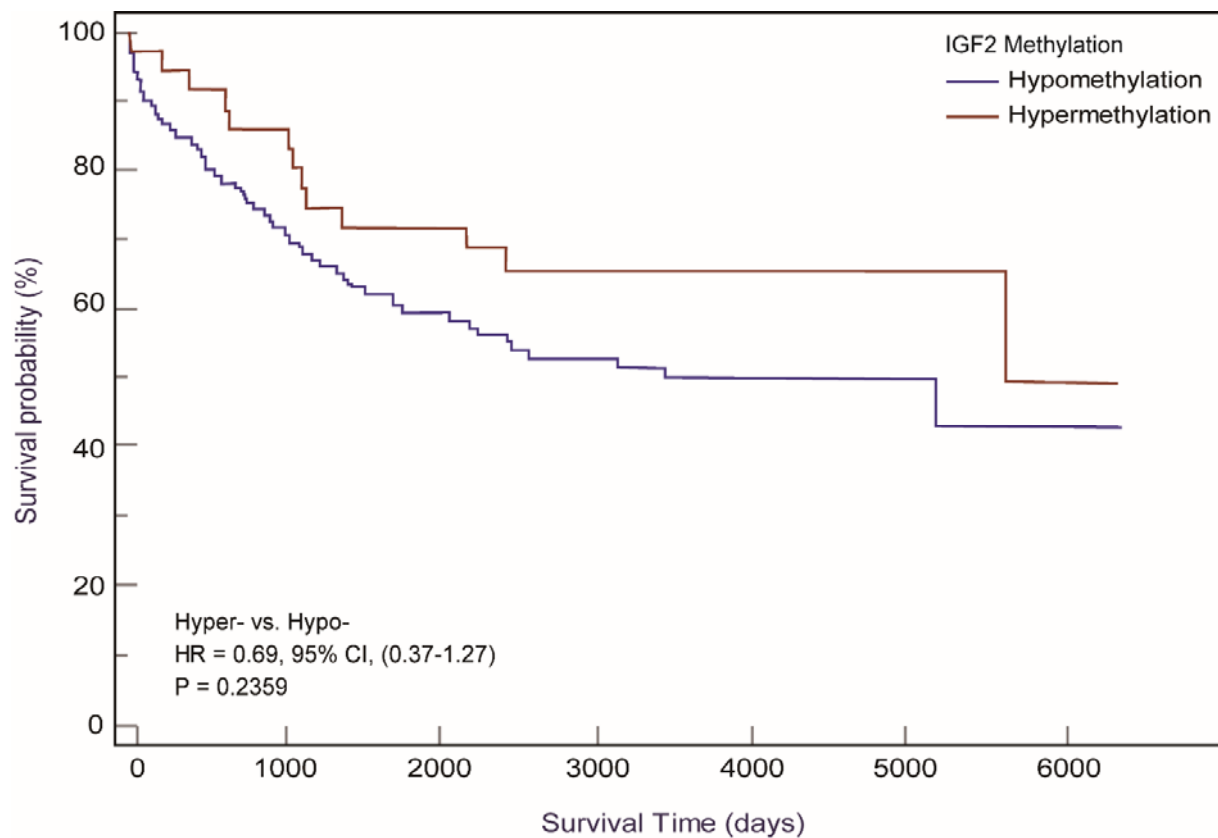
562

563 **Figure 4.** IGF2 methylation in PBLs and CRC prognosis in the initial cohort. Kaplan-Meier plots for
 564 (A) overall survival, (B) CRC specific survival, and (C) disease-free survival according to IGF2
 565 methylation status in PBLs in CRC patients. (D) Associations of *IGF2* methylation in PBLs with
 566 CRC prognosis during different follow-up periods. CI, confidence interval; CRC, colorectal cancer;
 567 CSS, cancer specific survival; DFS, disease free survival; HR, hazard ratio; Hyper-, IGF2
 568 hypermethylation; Hypo- IGF2 hypomethylation; OS, overall survival; PBL, peripheral blood
 569 leukocyte.
 570



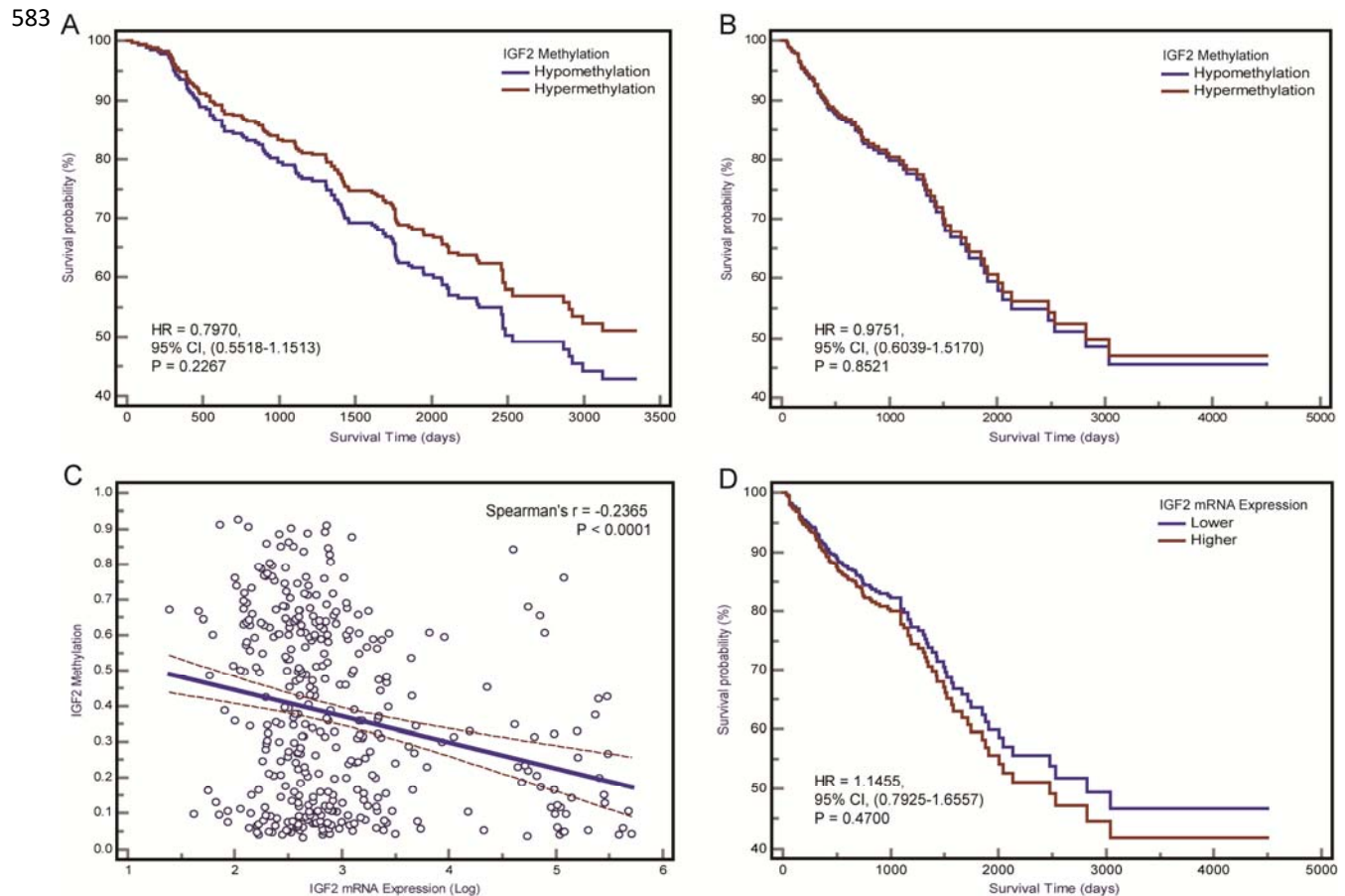
571

572 **Figure 5.** Kaplan-Meier curves for IGF2 methylation in PBLs and CRC overall survival in the
573 EPIC-Italy cohort. CI, confidence interval; HR, hazard ratio; Hyper-, IGF2 hypermethylation; Hypo-
574 IGF2 hypomethylation.

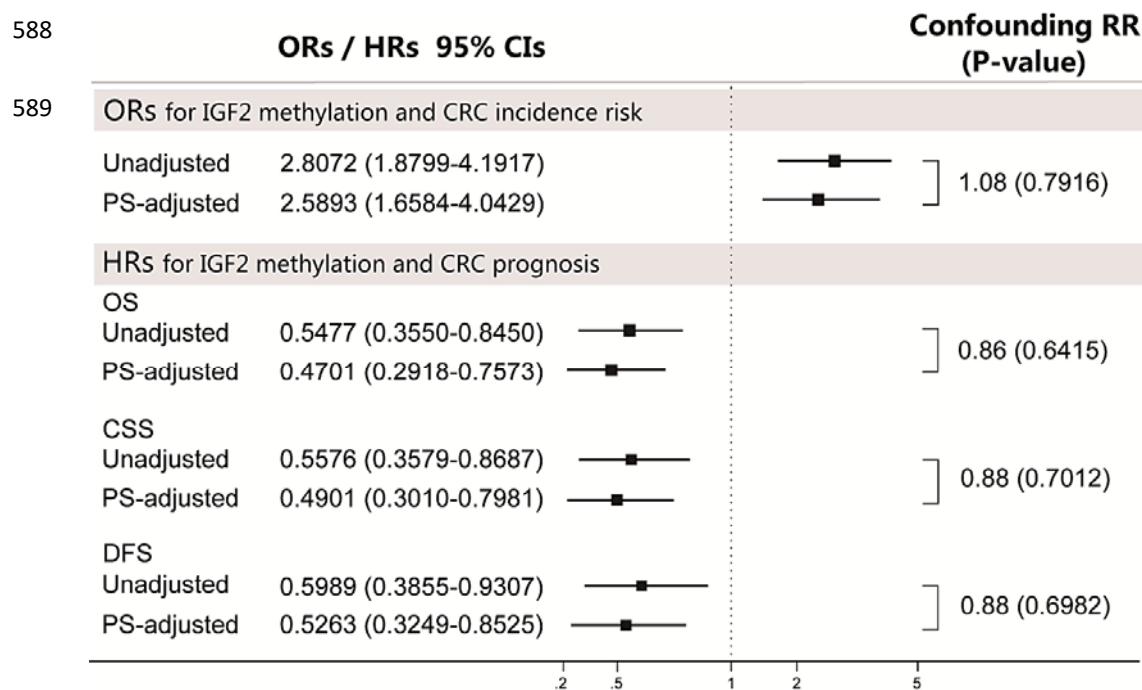


575

576 **Figure 6.** The association between IGF2 methylation in tumour tissues and CRC patient prognosis.
577 Kaplan-Meier curves for overall survival according to IGF2 methylation status in tumour tissues of
578 CRC patients in (A) the initial cohort and (B) TCGA datasets. (C) The correlation between IGF2
579 mRNA expression level and methylation level of the differentially methylated region of IGF2 tested
580 in our study using the TCGA dataset. (D) Kaplan-Meier curves for overall survival according to
581 IGF2 mRNA expression levels in tumour tissues of CRC patients in the TCGA cohort.
582 CI, confidence interval; CRC, colorectal cancer; HR, hazard ratio.

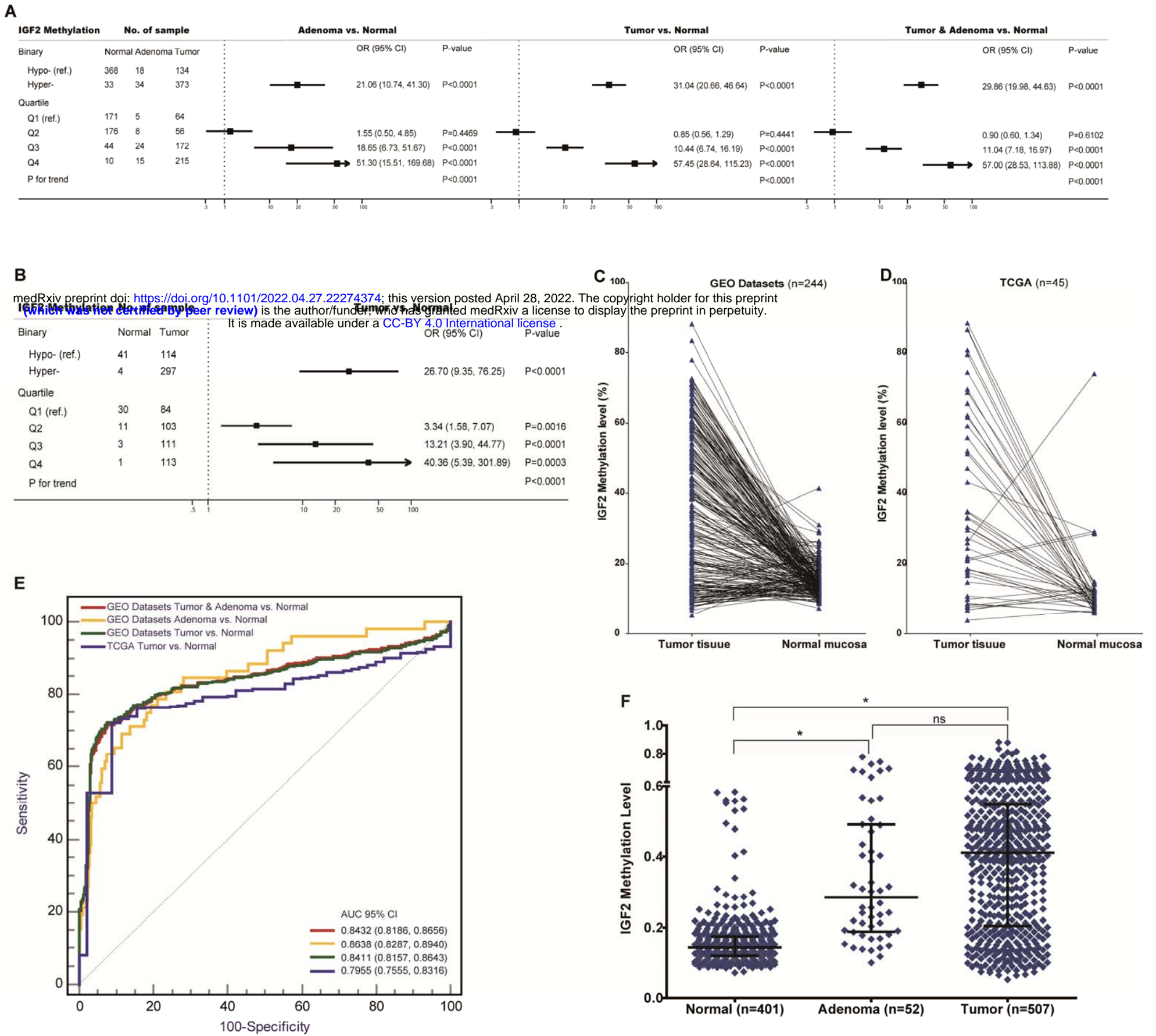


584 **Figure 7.** Sensitivity analyses using the confounding RR method. The confounding RR was defined
 585 as the ratio of the PS-adjusted effect estimates and the unadjusted effect estimates.
 586 CI, confidence interval; CRC, colorectal cancer; CSS, cancer specific survival; DFS, disease free
 587 survival; HR, hazard ratio; OR, odds ratio; OS, overall survival; PS, propensity score.



590 **Figure 8.** The associations between IGF2 methylation in tissue samples and CRC risk in the datasets from GEO and TCGA. (A) Adenoma or tumour
 591 tissues versus normal colorectal mucosa in the GEO datasets. (B) Tumour tissues versus normal colorectal mucosa in the TCGA datasets. (C) Tumour
 592 tissues versus matched adjacent normal tissues in the datasets from (C) GEO and (D) TCGA. (E) The discrimination performance of IGF2 methylation
 593 in tissue samples to identify CRC tumour or adenoma tissue and normal colorectal mucosa in the datasets from GEO and TCGA. (F) The IGF2
 594 methylation levels in normal colorectal mucosa, adenoma and tumour tissue samples in the GEO datasets. *indicates $P < 0.0001$, and ns indicates not
 595 statistical significance. AUC, the area under receiver operating characteristic curves; CI, confidence interval; OR, odds ratio; Hyper-, IGF2
 596 hypermethylation; Hypo-, IGF2 hypomethylation;

597

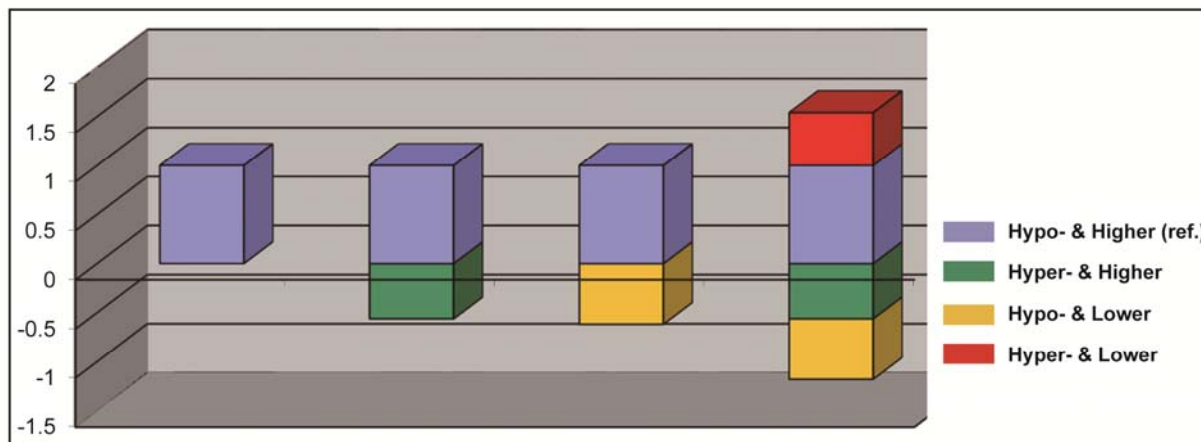


598

599 **Figure 9.** The interaction between PBL IGF2 methylation and serum CEA levels on CRC patient
 600 prognosis using the additive and multiplicative models. **(A)** The interaction analysed with the
 601 additive model. **(B)** The effect estimates for interaction in respected to both the additive and
 602 multiplicative models. AP, attributable proportion due to interaction; CI, confidence interval; HR,
 603 hazard ratio; Hyper-, IGF2 hypermethylation; Hypo-, IGF2 hypomethylation; PBL, peripheral blood
 604 leukocyte; RERI, relative excess risk due to interaction; SI, synergy index; V index, the
 605 multiplicative interaction index.

606

A



B

CEA Level (ng/mL)	IGF2 Methylation in PBLs		Effect for Interaction (95% CI)			
	Hypomethylation	Hypermethylation	Additive interaction		Multiplicative interaction	
	HR _{eg} (95% CI)		RERI	AP	SI	V index
Higher (> 5)	1	0.44 (0.24-0.79)	0.54	1.49	0.54	2.10
Lower (≤ 5)	0.38 (0.24-0.61)	0.36 (0.20-0.65)	(0.18-0.90)	(0.46-2.52)	(0.37-0.80)	(0.86-5.15)

607 **Table 1.** Subgroup analysis for the association of PBL IGF2 methylation with CRC prognosis.

Subgroups		Univariate HR, 95% CI	P-value	PS-adjusted HR, 95% CI	P-value	P-value for heterogeneity*
Tumour Site						
OS	Colon	0.65 (0.33-1.27)	0.2049	0.59 (0.29-1.17)	0.1309	0.4117
	Rectum	0.49 (0.28-0.86)	0.0132	0.39 (0.20-0.77)	0.0065	
CSS	Colon	0.60 (0.29-1.21)	0.1545	0.55 (0.27-1.13)	0.1062	0.6706
	Rectum	0.53 (0.30-0.94)	0.0293	0.45 (0.22-0.88)	0.0204	
DFS	Colon	0.51 (0.25-1.05)	0.0661	0.48 (0.23-1.00)	0.0514	0.7539
	Rectum	0.68 (0.39-1.18)	0.1714	0.56 (0.29-1.09)	0.0895	
UICC Stage						
OS	I + II + III	0.58 (0.37-0.90)	0.0158	0.45 (0.28-0.74)	0.0017	0.5840
	I + II	0.57 (0.27-1.21)	0.1427	0.35 (0.15-0.82)	0.0155	
	III	0.54 (0.31-0.95)	0.0327	0.54 (0.30-0.98)	0.0440	
	IV	1.04 (0.13-8.38)	0.9726	0.94 (0.11-8.08)	0.9542	
CSS	I + II + III	0.59 (0.37-0.93)	0.0243	0.47 (0.29-0.78)	0.0037	0.7135
	I + II	0.59 (0.27-1.30)	0.1880	0.39 (0.16-0.97)	0.0427	
	III	0.55 (0.31-0.97)	0.0375	0.54 (0.30-0.98)	0.0441	
	IV	1.04 (0.13-8.38)	0.9726	0.94 (0.11-8.08)	0.9542	
DFS	I + II + III	0.64 (0.40-1.00)	0.0508	0.51 (0.31-0.84)	0.0084	0.7337
	I + II	0.63 (0.29-1.37)	0.2435	0.41 (0.17-0.99)	0.0471	
	III	0.61 (0.35-1.06)	0.0812	0.60 (0.33-1.10)	0.0974	
	IV	0.67 (0.08-5.34)	0.7048	0.75 (0.09-6.55)	0.7982	
Gender						
OS	Female	0.72 (0.39-1.32)	0.2910	0.50 (0.24-1.04)	0.0626	0.6172
	Male	0.40 (0.21-0.76)	0.0054	0.39 (0.20-0.77)	0.0066	
CSS	Female	0.74 (0.39-1.40)	0.3552	0.55 (0.26-1.16)	0.1163	0.5527
	Male	0.41 (0.22-0.79)	0.0071	0.40 (0.20-0.79)	0.0086	
DFS	Female	0.75 (0.40-1.42)	0.3789	0.55 (0.26-1.16)	0.1184	0.7691
	Male	0.49 (0.26-0.90)	0.0226	0.48 (0.25-0.92)	0.0277	
Age, (years)						
OS	< 60	0.54 (0.27-1.06)	0.0719	0.53 (0.26-1.11)	0.0948	0.6705
	≥ 60	0.54 (0.30-0.95)	0.0339	0.43 (0.23-0.81)	0.0082	
CSS	< 60	0.55 (0.28-1.08)	0.0804	0.56 (0.27-1.17)	0.1211	0.6533
	≥ 60	0.56 (0.31-1.01)	0.0555	0.45 (0.23-0.85)	0.0144	
DFS	< 60	0.60 (0.30-1.18)	0.1379	0.68 (0.30-1.30)	0.2123	0.5624
	≥ 60	0.59 (0.33-1.05)	0.0747	0.47 (0.25-0.89)	0.0197	
BMI, (Kg/m ²)						
OS	< 24	0.52 (0.30-0.88)	0.0144	0.47 (0.27-0.85)	0.0119	0.9947
	≥ 24	0.56 (0.25-1.23)	0.1473	0.48 (0.21-1.07)	0.0726	
CSS	< 24	0.56 (0.33-0.97)	0.0370	0.53 (0.29-0.95)	0.0333	0.7016
	≥ 24	0.49 (0.21-1.14)	0.0984	0.43 (0.18-1.02)	0.0557	
DFS	< 24	0.61 (0.36-1.04)	0.0689	0.54 (0.30-0.97)	0.0378	0.8742
	≥ 24	0.51 (0.22-1.20)	0.1234	0.49 (0.21-1.18)	0.1127	

608 *Test for heterogeneity between ORs was conducted by using random effect models with Comprehensive Meta
609 Analysis (version 2.2.046). BMI, body mass index; CI, confidence interval; CRC, colorectal cancer; CSS, cancer
610 specific survival; DFS, disease free survival; HR, hazard ratio; OS, overall survival; PBL, peripheral blood leukocyte;
611 PS, propensity score.

612 **Supplementary Files**

613 Supplementary File 1A. Main characteristics of participants of the initial and validation studies in the
614 CRC risk analysis.

615 Supplementary File 1B. Main characteristics of patients of the initial and validation cohorts in the CRC
616 prognosis analysis.

617 Supplementary File 2A. Comparisons of participant characteristics and covariates between CRC cases
618 and controls before and after propensity score adjustment in the initial case-control study in CRC risk
619 analysis.

620 Supplementary File 2B. Comparisons of baseline characteristics of CRC patients before and after
621 propensity score adjustment in the initial cohort in CRC prognosis analysis.

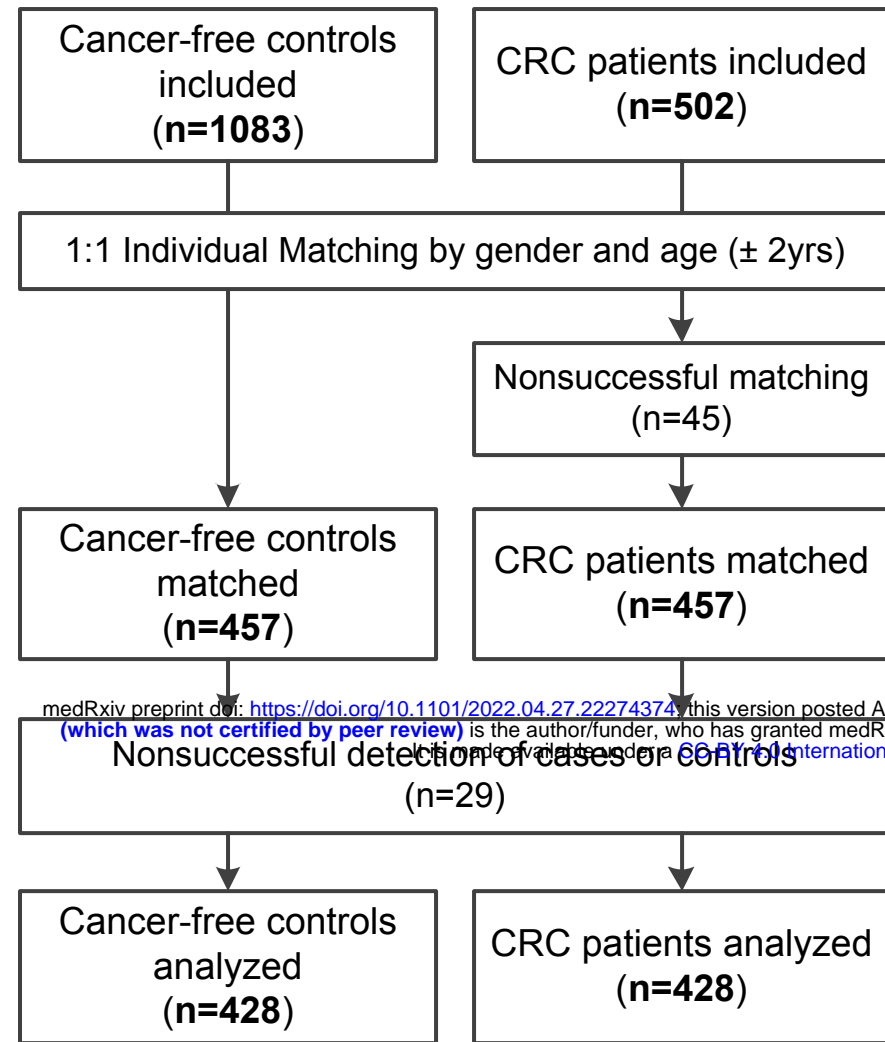
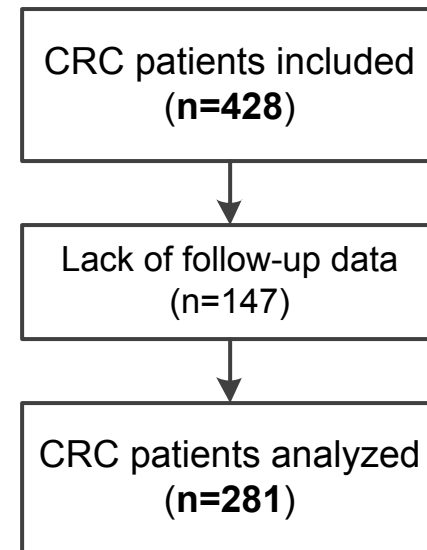
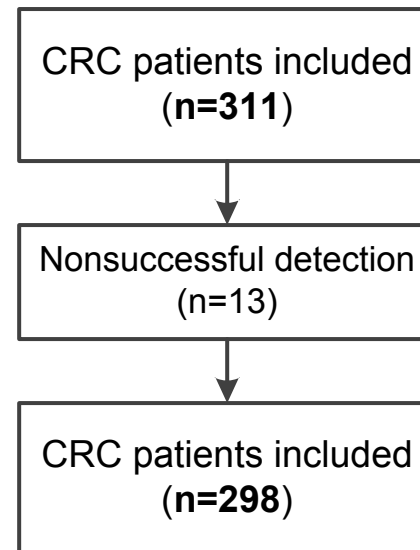
622 Supplementary File 3A. Sensitivity analyses using additional propensity score based methods for CRC
623 risk analysis in the initial case-control study.

624 Supplementary File 3B. Sensitivity analyses using additional propensity score based methods for CRC
625 survival analysis in the initial CRC cohort.

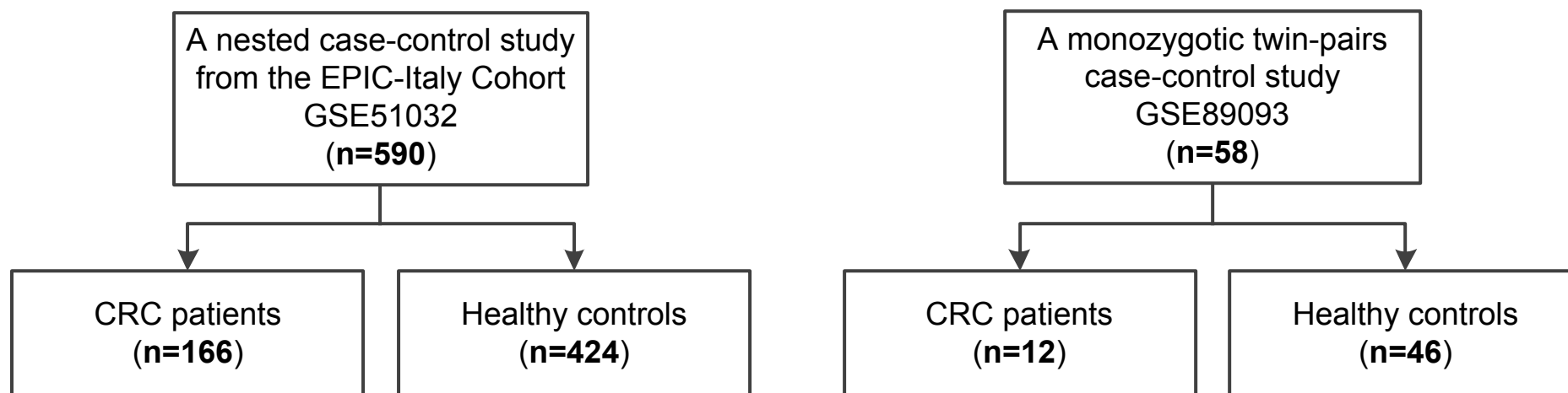
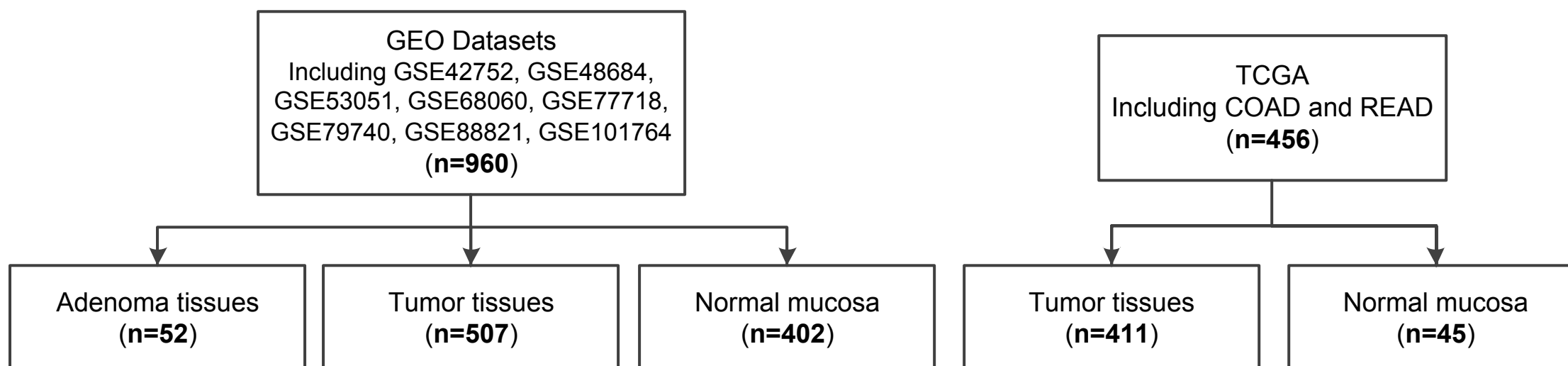
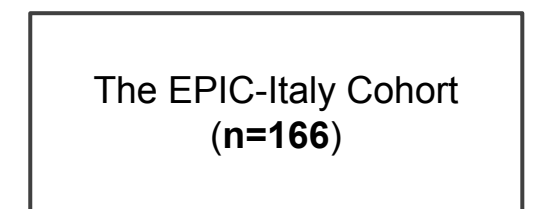
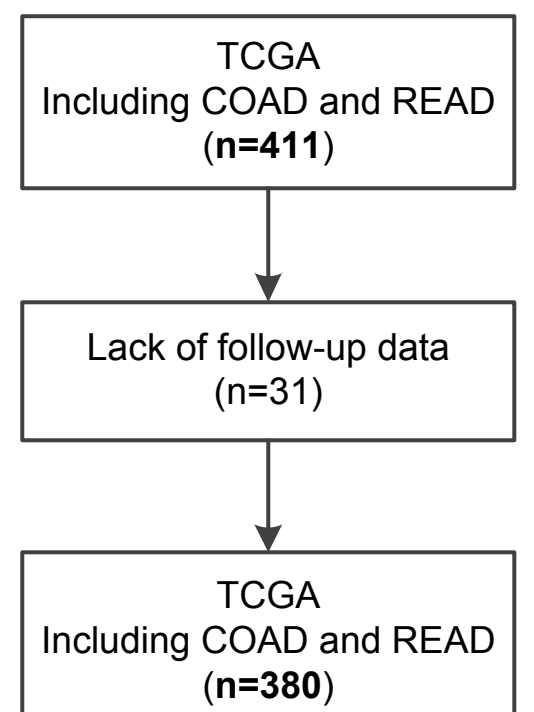
626 Supplementary File 4. E-values for PS-adjusted effect estimates.

627 Supplementary File 5. Associations between PBL IGF2 methylation status and UICC stage, serum CEA
628 level, WBC counts, and whether or not included in the survival analysis among CRC patients.

629 Supplementary File 6. Associations of PBL IGF2 methylation and CRC patient prognosis with or
630 without WBC counts included in PS models.

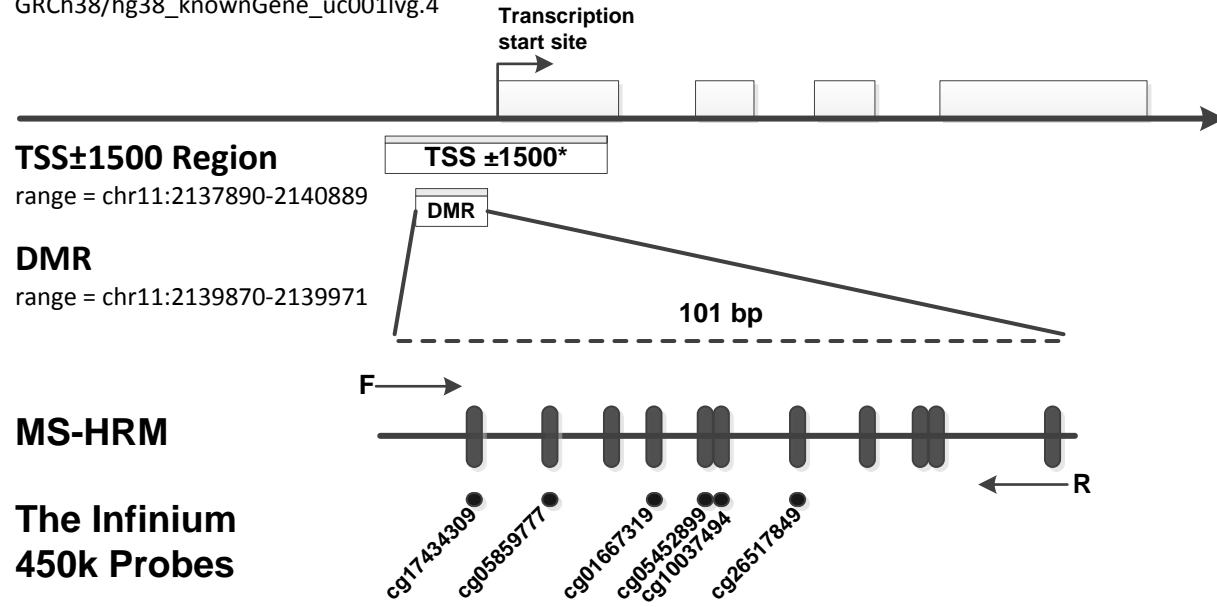
A**Initial Study****CRC Incidence Risk Analysis
Case-control Study**Blood Samples**CRC Prognosis Analysis
Cohort Study**Blood SamplesTissue Samples

medRxiv preprint doi: <https://doi.org/10.1101/2022.04.27.22274374>; this version posted April 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

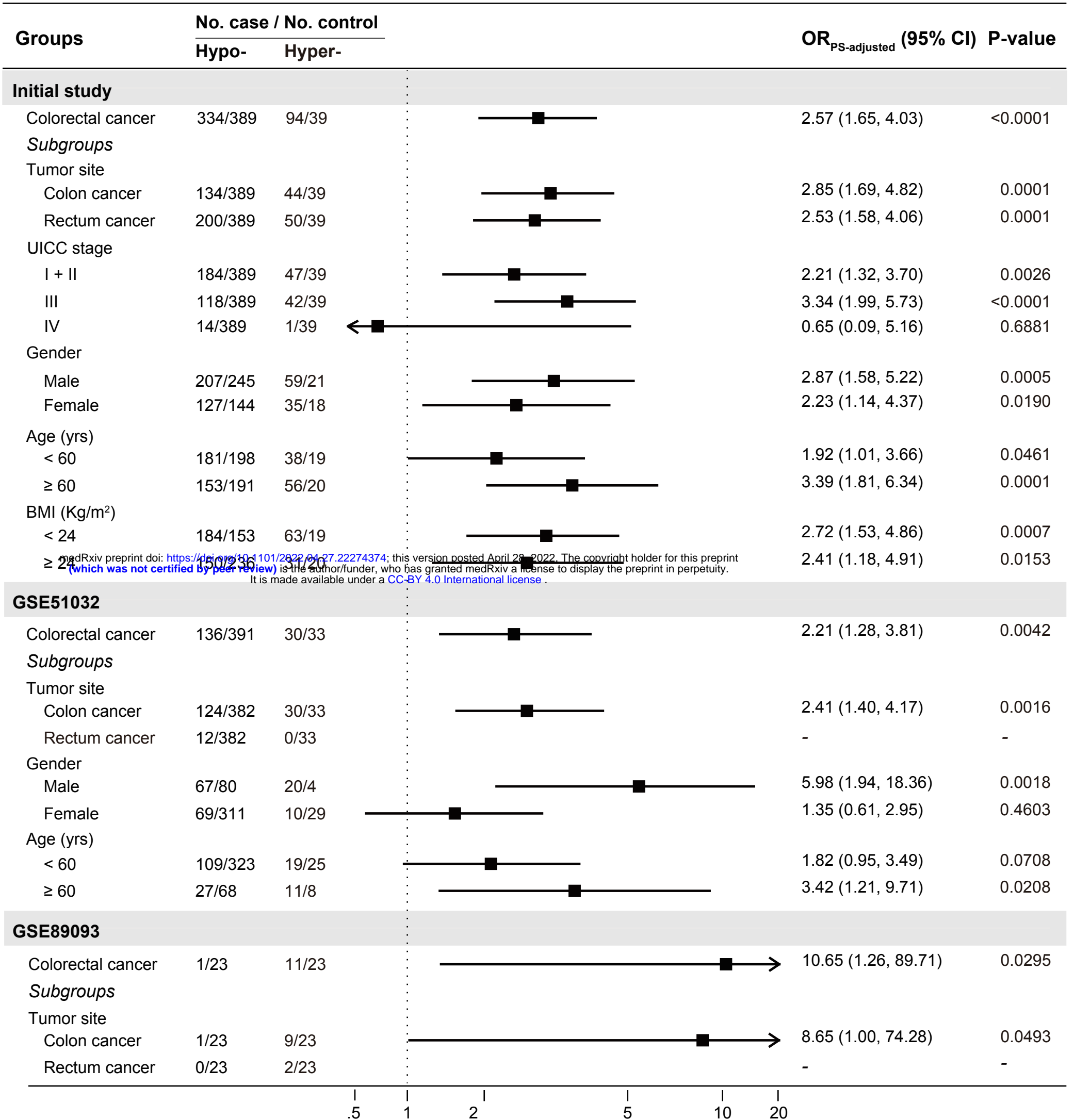
B**Validation Study****CRC Incidence Risk Analysis**Blood SamplesTissue Samples**CRC Prognosis Analysis**Blood SamplesTissue Samples

IGF2

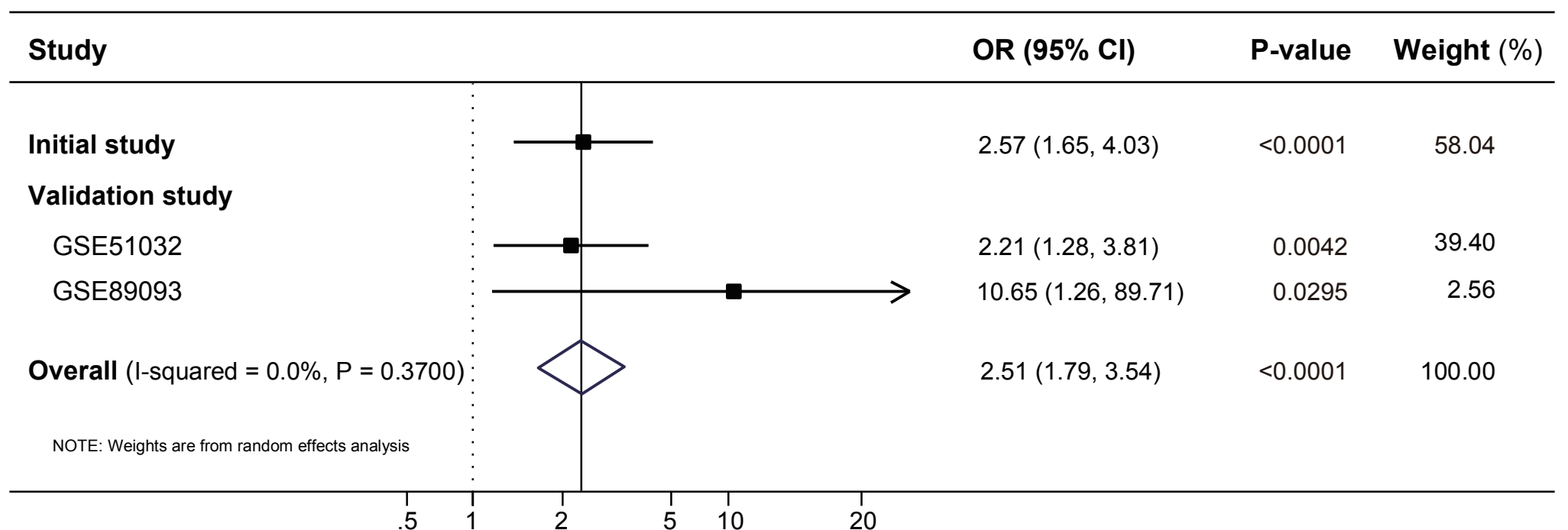
GRCh38/hg38_knownGene_uc001lvq.4

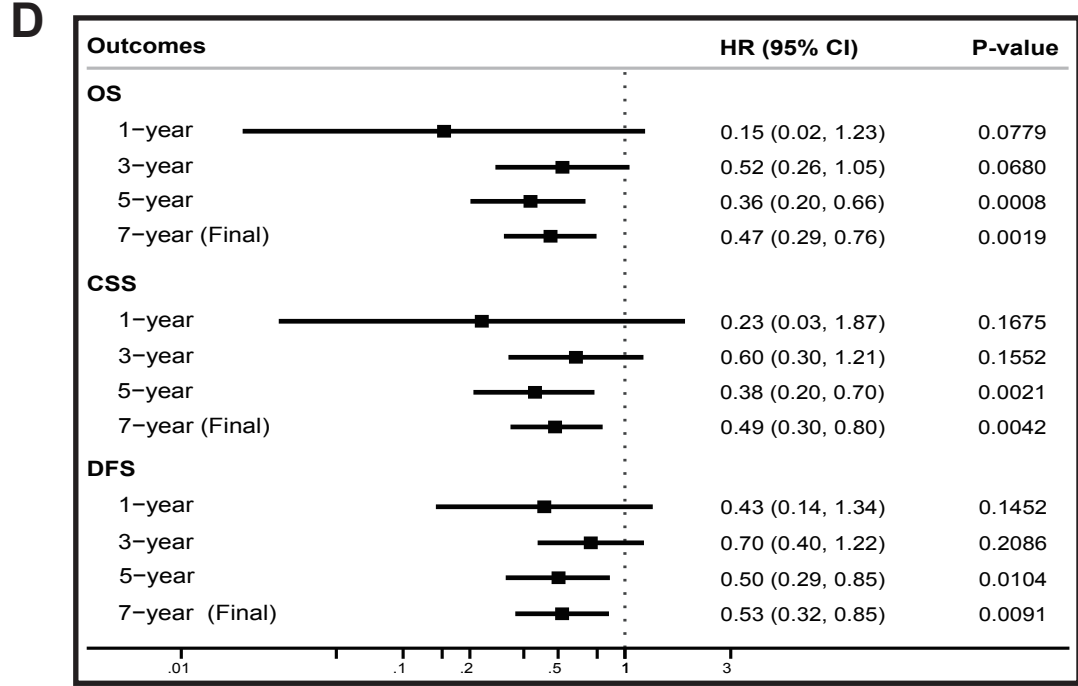
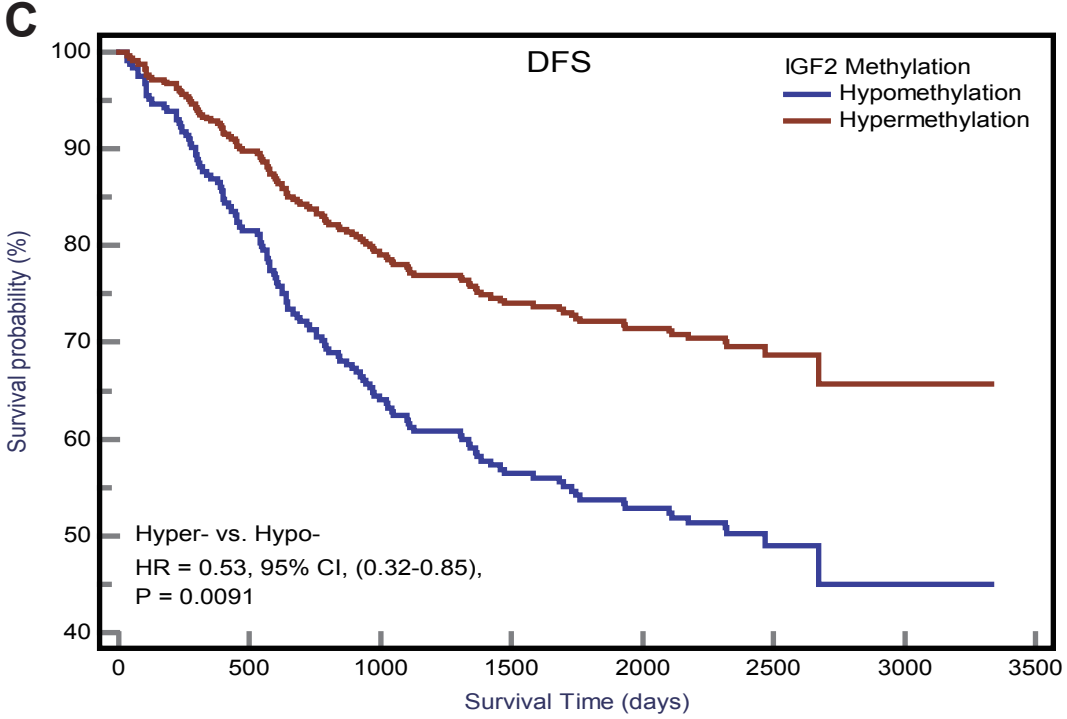
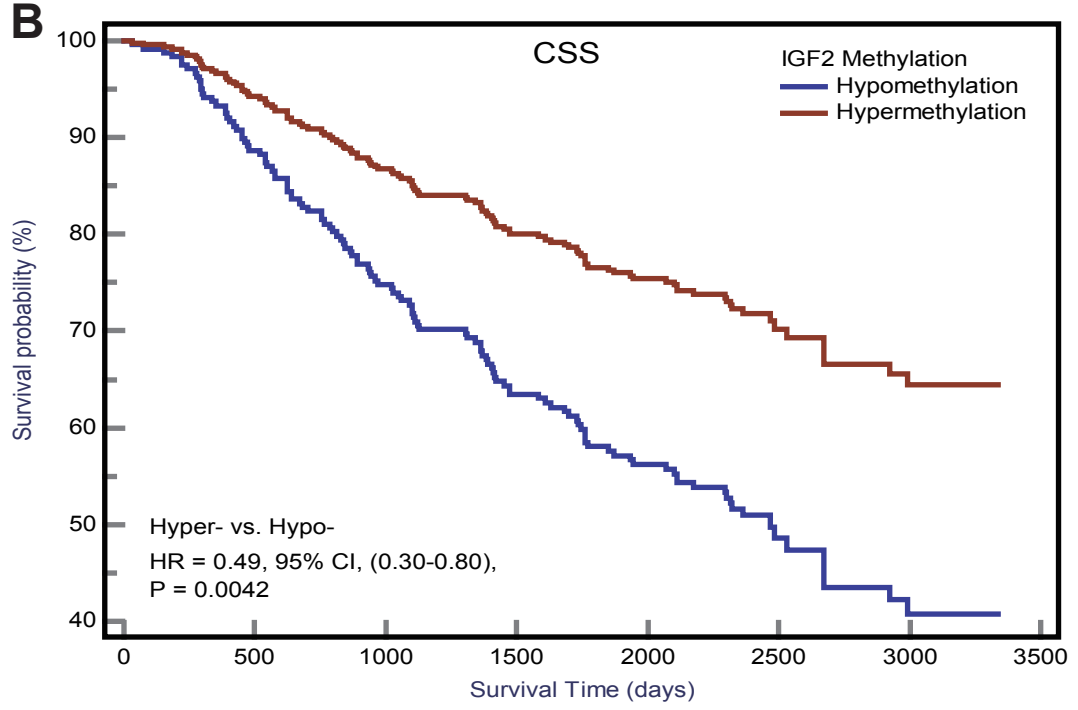
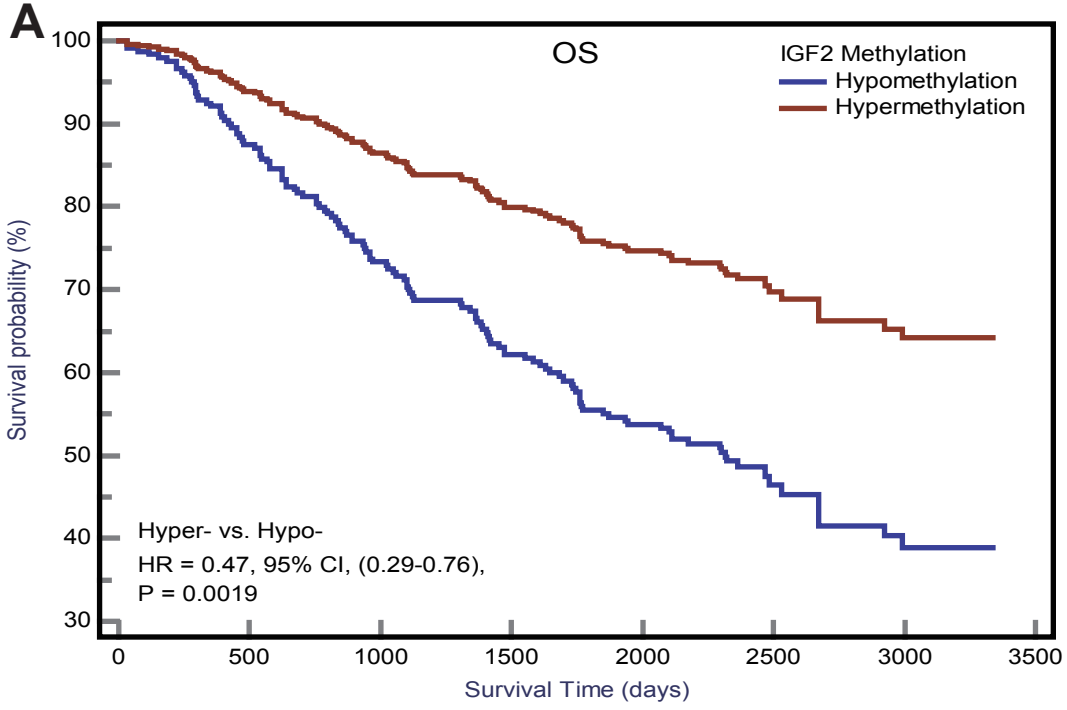


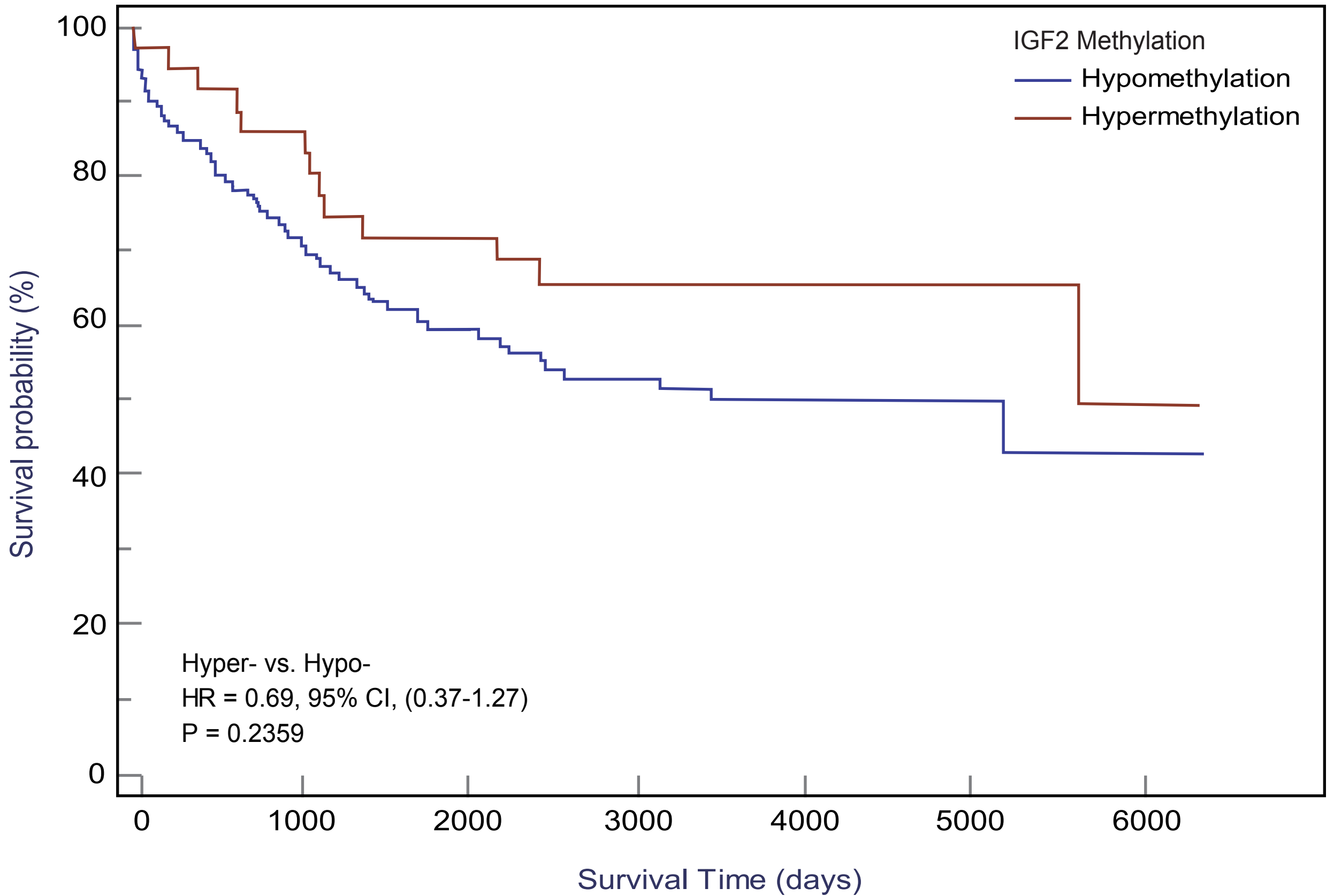
A

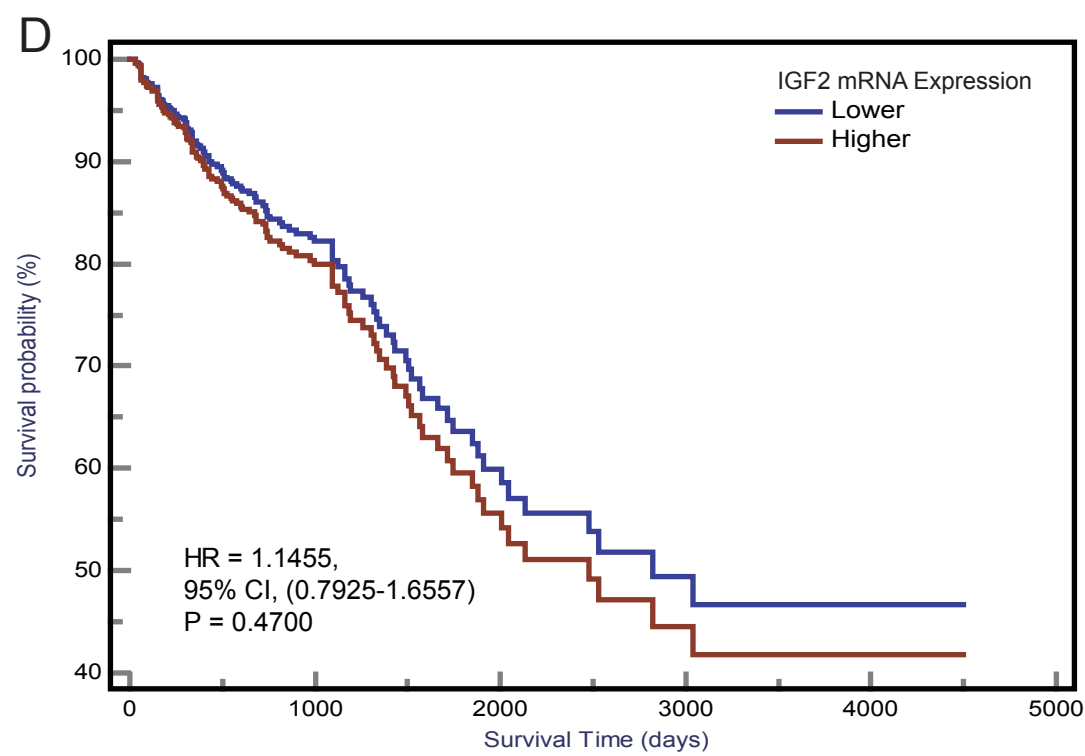
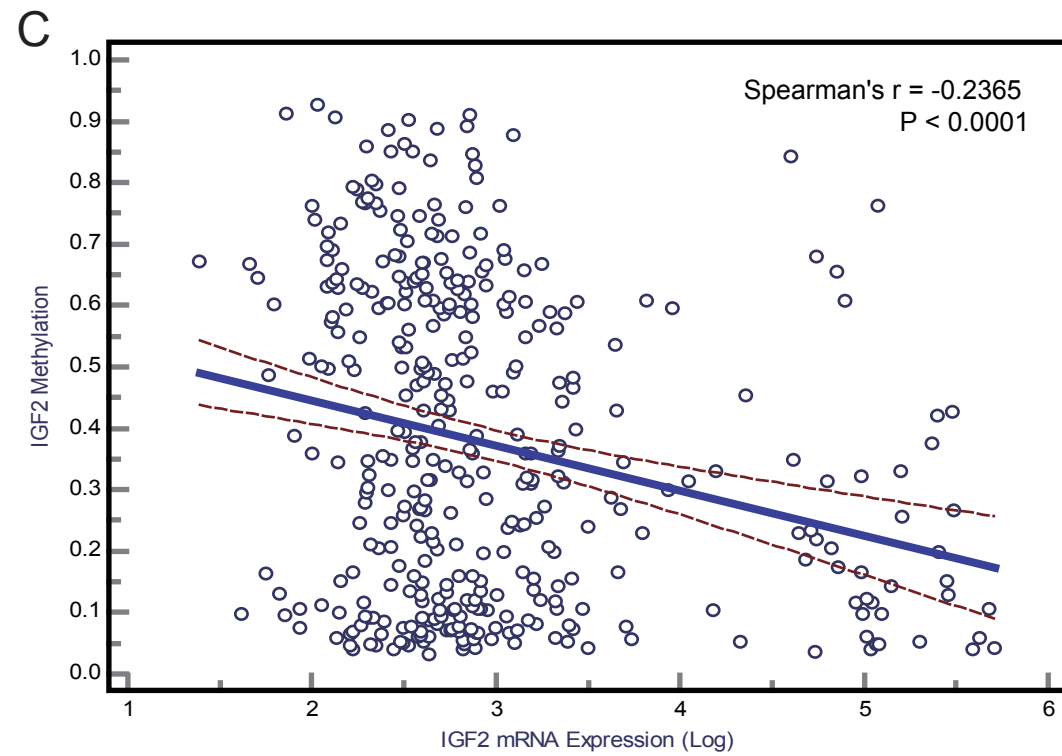
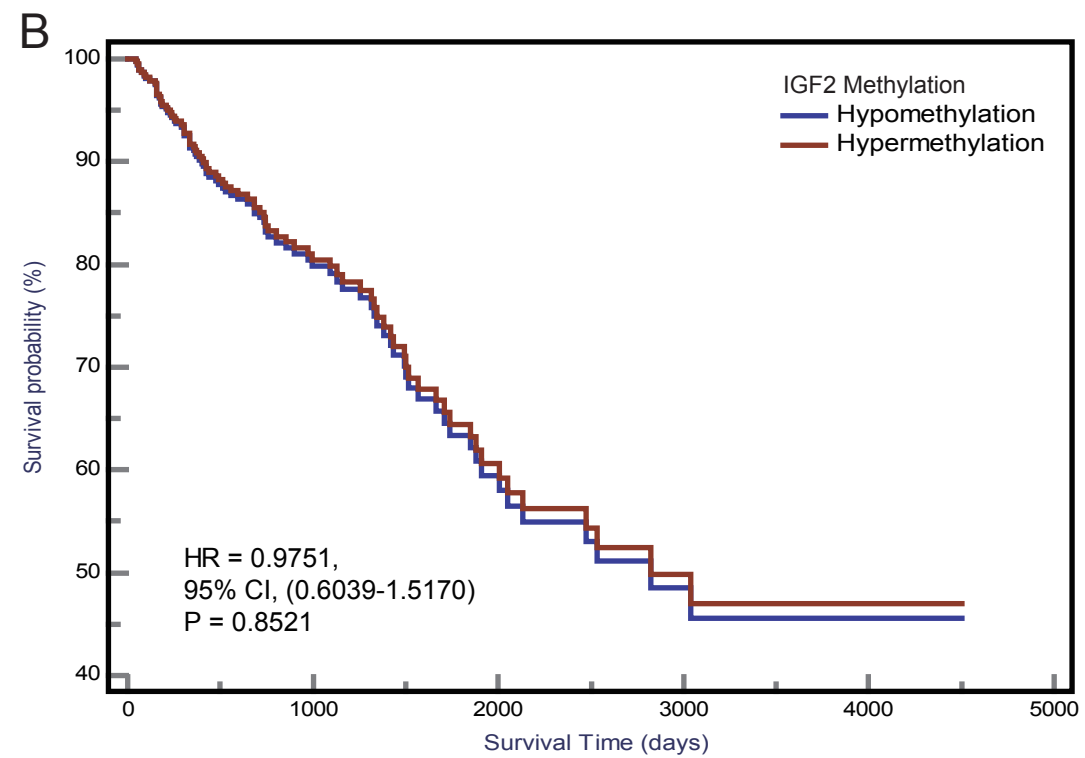
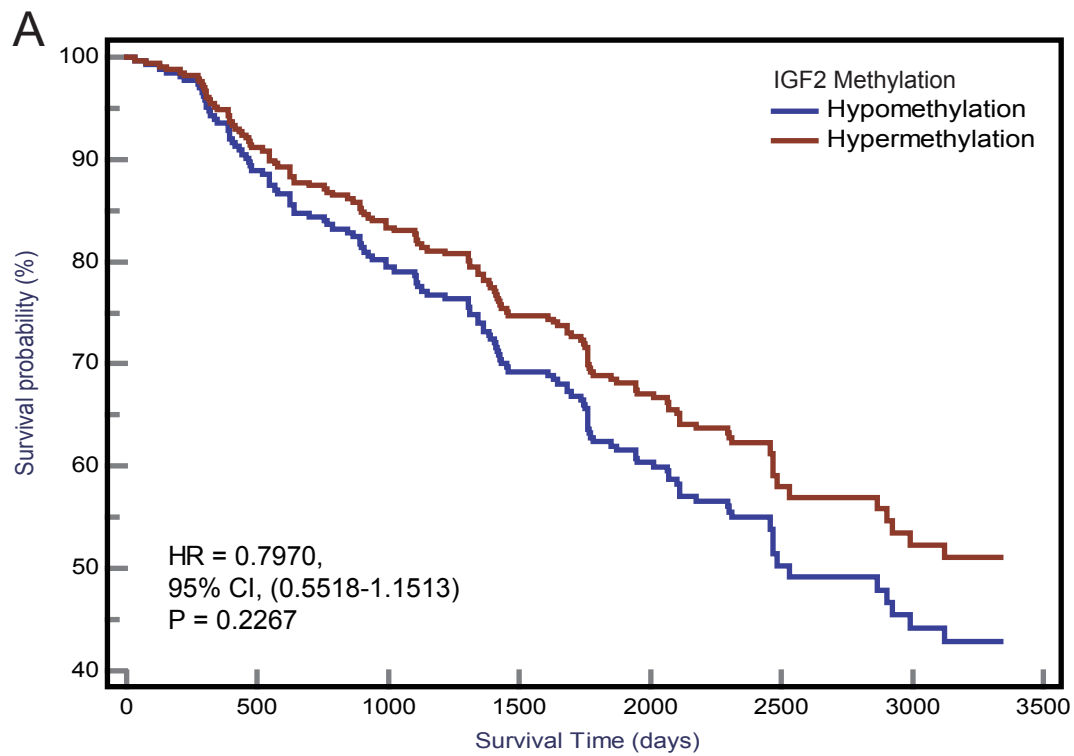


B









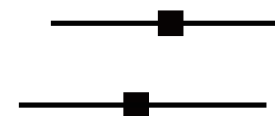
ORs / HRs 95% CIs

Confounding RR (P-value)

ORs for IGF2 methylation and CRC incidence risk

Unadjusted 2.8072 (1.8799-4.1917)

PS-adjusted 2.5893 (1.6584-4.0429)



1.08 (0.7916)

HRs for IGF2 methylation and CRC prognosis

OS

Unadjusted 0.5477 (0.3550-0.8450)

PS-adjusted 0.4701 (0.2918-0.7573)



0.86 (0.6415)

CSS

Unadjusted 0.5576 (0.3579-0.8687)

PS-adjusted 0.4901 (0.3010-0.7981)



0.88 (0.7012)

DFS

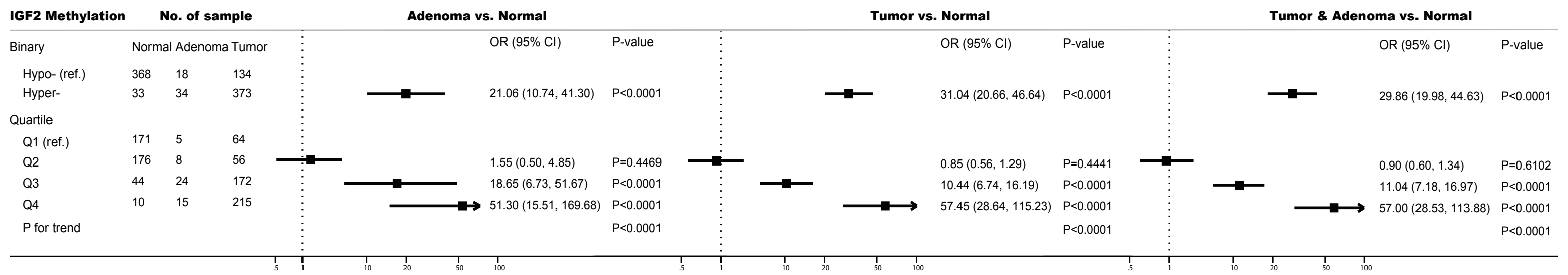
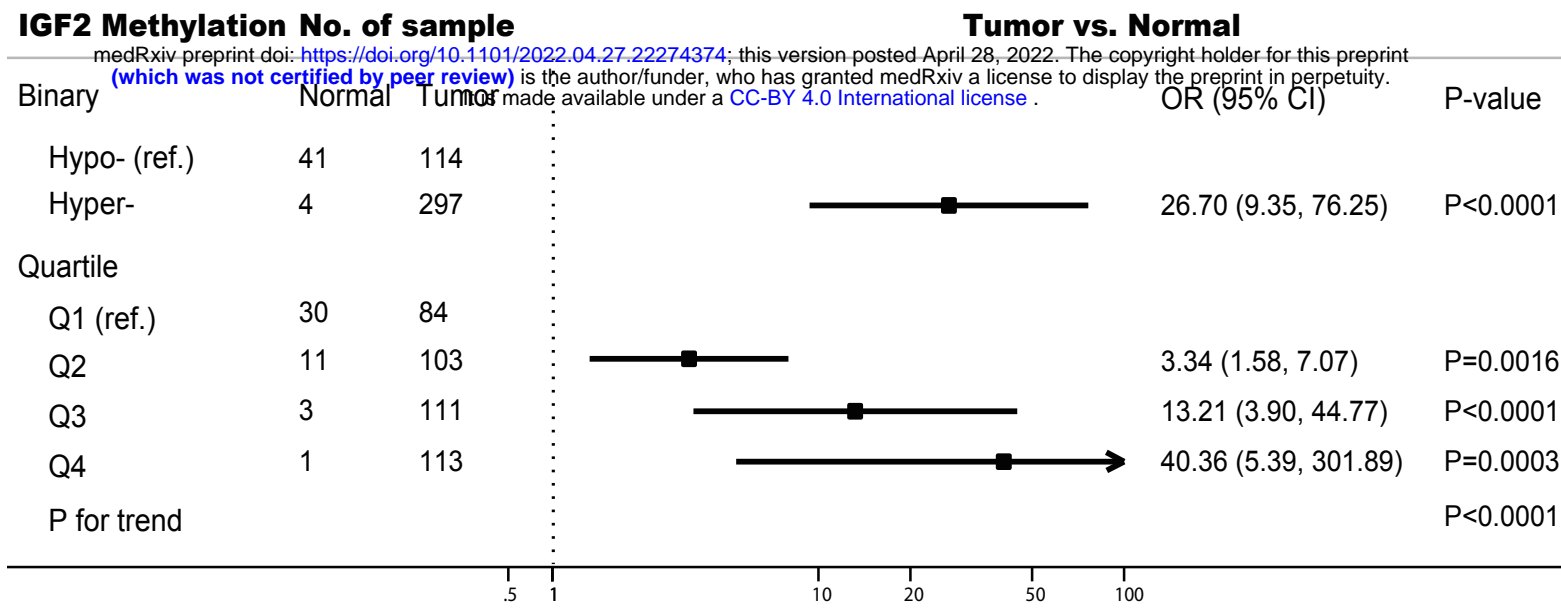
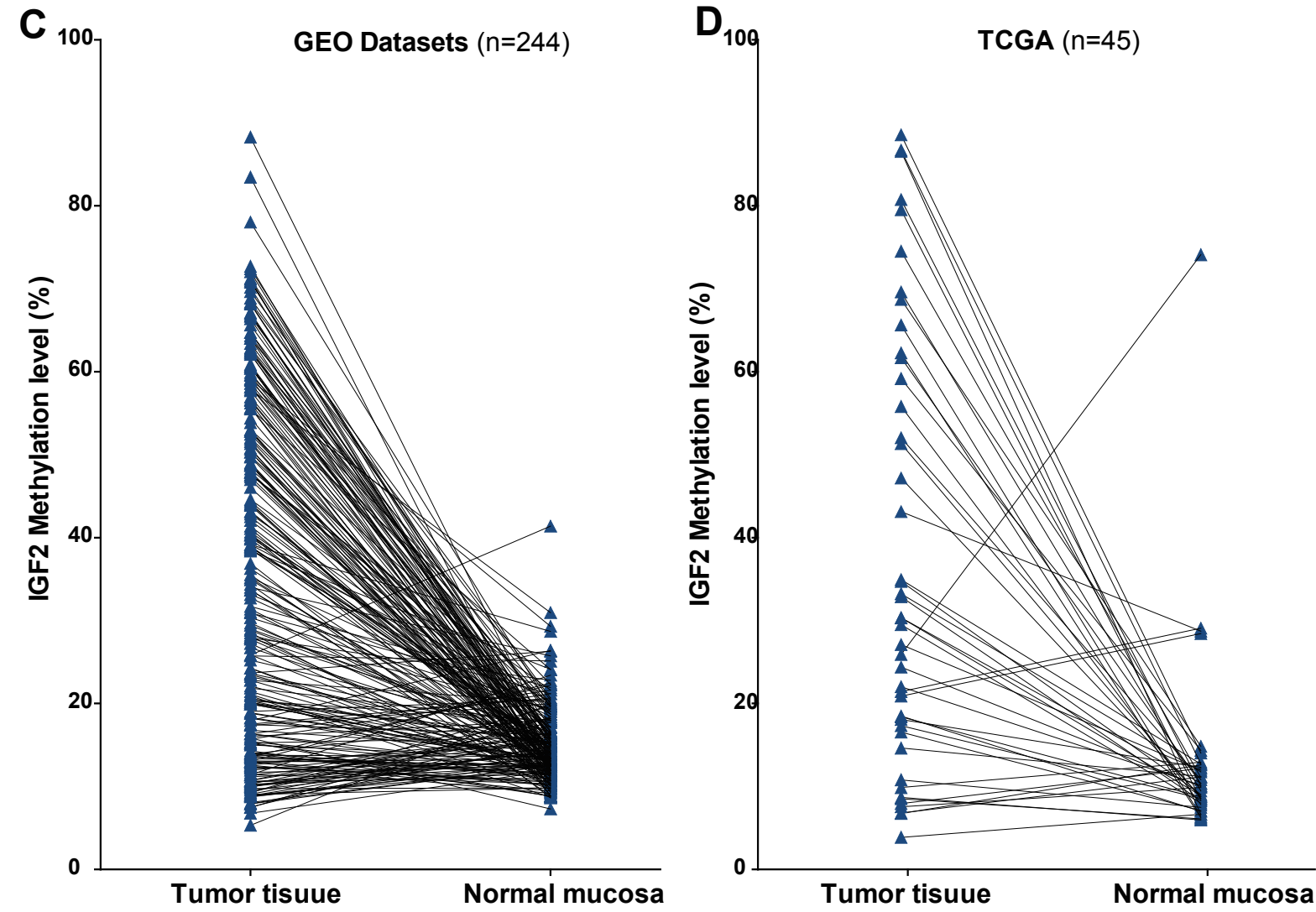
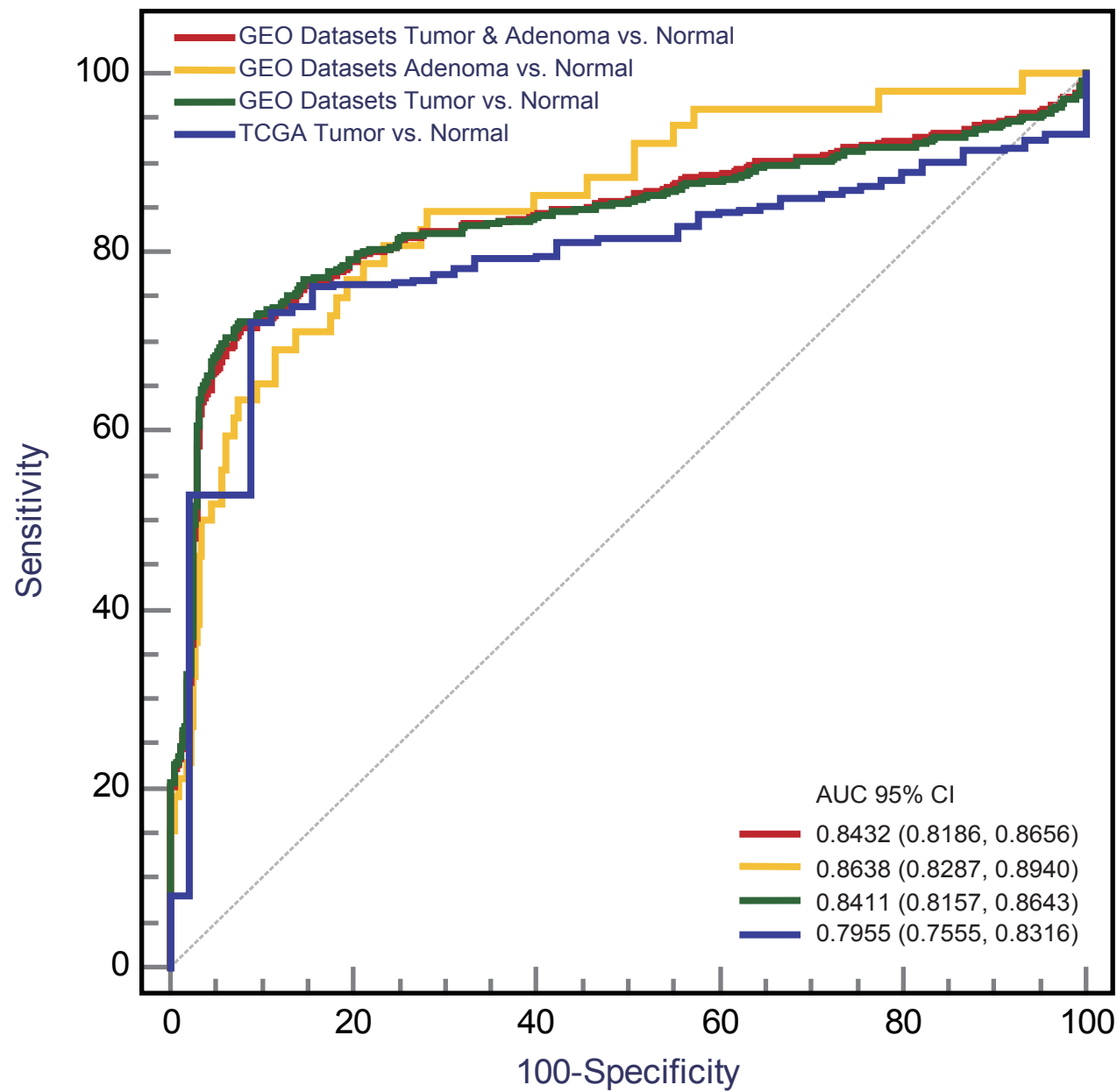
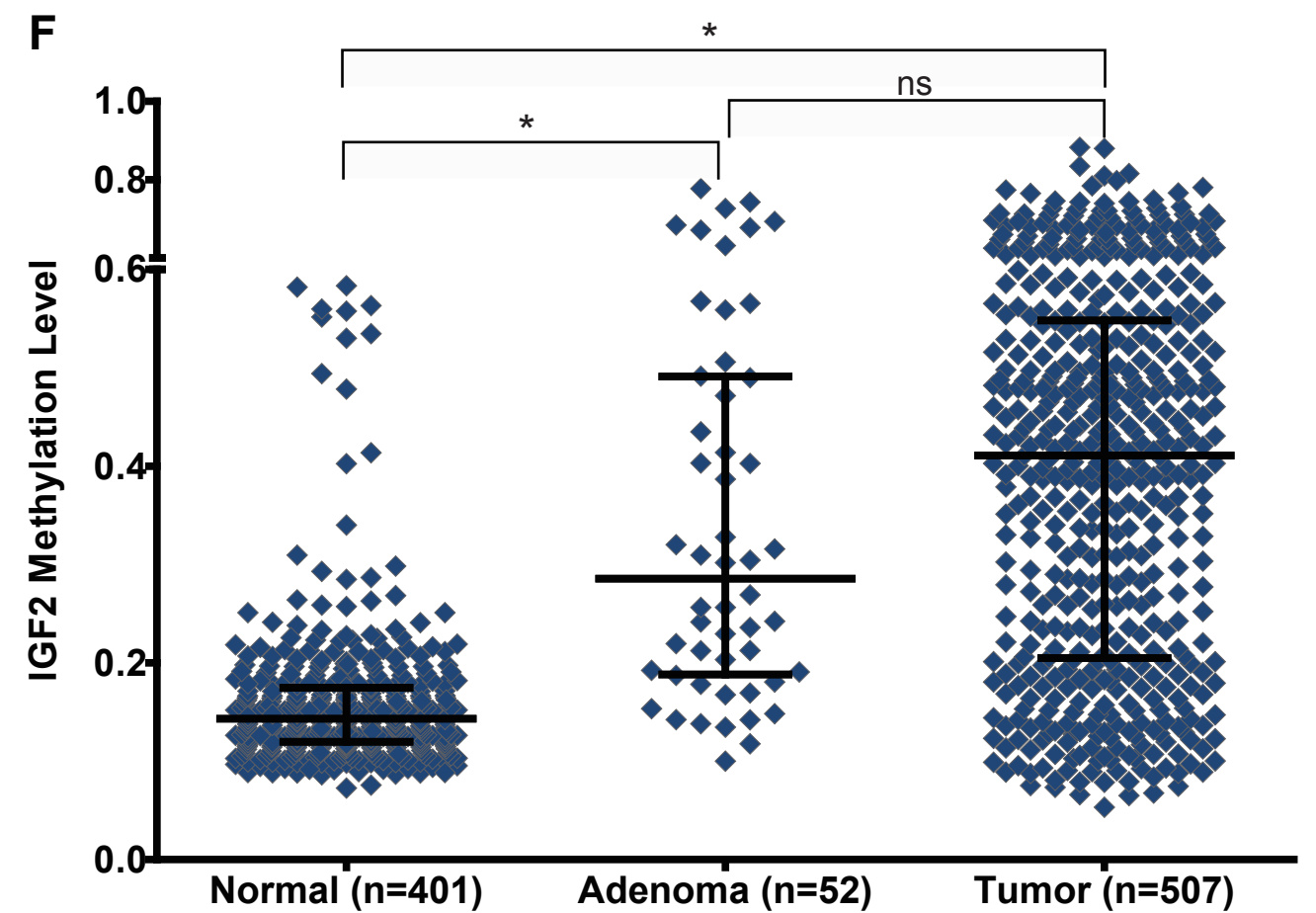
Unadjusted 0.5989 (0.3855-0.9307)

PS-adjusted 0.5263 (0.3249-0.8525)

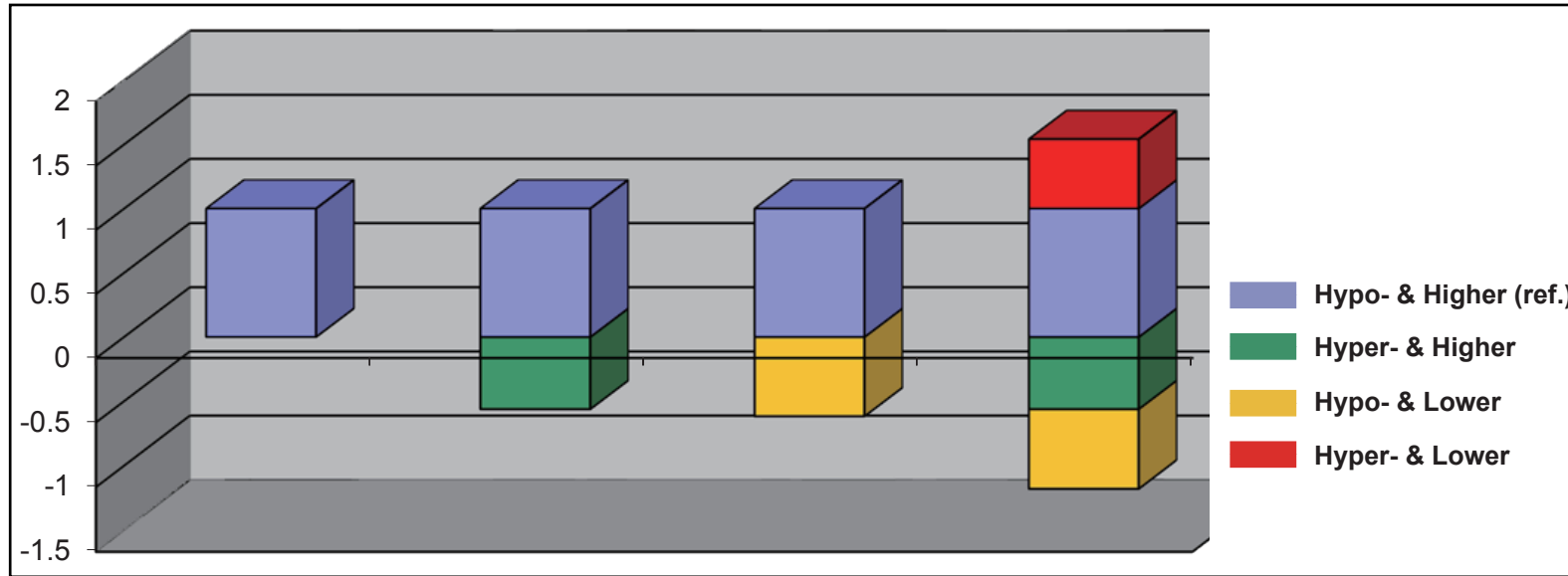


0.88 (0.6982)

.2 .5 1 2 5

A**B****C****E****F**

A



B

CEA Level (ng/mL)	IGF2 Methylation in PBLs		Effect for Interaction (95% CI)			
	Hypomethylation	Hypermethylation	Additive interaction			Multiplicative interaction
	HR _{eg} (95% CI)		RERI	AP	SI	V index
Higher (> 5)	1	0.44 (0.24-0.79)	0.54	1.49	0.54	2.10
Lower (≤ 5)	0.38 (0.24-0.61)	0.36 (0.20-0.65)	(0.18-0.90)	(0.46-2.52)	(0.37-0.80)	(0.86-5.15)