

Supplementary Table S1: Baseline characteristics of participants in the Hp-CRC Consortium, by cohort

	CLUE (N=982)	CPSII (N=722)	HPFS (N=302)	MEC (N=1510)	NHS (N=576)	NYUWHS (N=572)	PHS (N=360)	PLCO (N=1240)	SCCS (N=252)	WHI (N=1610)	Total (N=8126)
Age											
N	982	722	302	1510	576	572	360	1240	252	1610	8126
Mean (SD)	52 (12.2)	71 (5.5)	65 (8.7)	69 (8.2)	58 (6.9)	55 (7.7)	69 (8.2)	65 (5.1)	55 (8.8)	66 (6.9)	63 (10.1)
Median	53	71	66	69	59	58	69	64	55	66	64
Q1, Q3	44, 60	67, 75	58, 72	63, 76	53, 64	50, 62	64, 74	61, 69	47, 63	61, 71	58, 71
Range	(18-82)	(57-87)	(47-81)	(49-87)	(43-70)	(36-66)	(51-89)	(55-75)	(40-76)	(50-79)	(18-89)
Sex											
Female	554 (56%)	368 (51%)	0 (0%)	716 (47%)	576 (100%)	572 (100%)	0 (0%)	572 (46%)	144 (57%)	1610 (100%)	5112 (63%)
Male	428 (44%)	354 (49%)	302 (100%)	794 (53%)	0 (0%)	0 (0%)	360 (100%)	668 (54%)	108 (43%)	0 (0%)	3014 (37%)
Race											
White	974 (99%)	712 (99%)	290 (96%)	226 (15%)	576 (100%)	446 (78%)	342 (95%)	1098 (89%)	74 (29%)	1396 (87%)	6134 (75%)
Black	8 (1%)	6 (1%)	0 (0%)	286 (19%)	0 (0%)	62 (11%)	2 (1%)	98 (8%)	172 (68%)	164 (10%)	798 (10%)
Asian	0 (0%)	2 (0%)	0 (0%)	570 (38%)	0 (0%)	6 (1%)	12 (3%)	24 (2%)	0 (0%)	0 (0%)	614 (8%)
Latino	0 (0%)	0 (0%)	0 (0%)	332 (22%)	0 (0%)	20 (3%)	4 (1%)	16 (1%)	0 (0%)	50 (3%)	422 (5%)
Other/unknown/multiracial	0 (0%)	2 (0%)	12 (4%)	96 (6%)	0 (0%)	38 (7%)	0 (0%)	4 (0%)	6 (2%)	0 (0%)	158 (2%)
Education											
Missing	3 (0%)	2 (0%)	0 (0%)	9 (1%)	0 (0%)	73 (13%)	0 (0%)	1 (0%)	5 (2%)	15 (1%)	108 (1%)
Less Than High School	399 (41%)	22 (3%)	0 (0%)	215 (14%)	0 (0%)	35 (6%)	0 (0%)	119 (10%)	99 (39%)	83 (5%)	972 (12%)
Completed high school or GED	356 (36%)	153 (21%)	0 (0%)	343 (23%)	0 (0%)	156 (27%)	0 (0%)	308 (25%)	76 (30%)	276 (17%)	1668 (21%)
Post high school training other than college	0 (0%)	58 (8%)	0 (0%)	0 (0%)	0 (0%)	9 (2%)	0 (0%)	147 (12%)	6 (2%)	142 (9%)	362 (4%)
Some College	122 (12%)	180 (25%)	0 (0%)	494 (33%)	0 (0%)	120 (21%)	0 (0%)	261 (21%)	50 (20%)	445 (28%)	1672 (21%)
College graduate	53 (5%)	153 (21%)	0 (0%)	261 (17%)	531 (92%)	88 (15%)	0 (0%)	218 (18%)	10 (4%)	169 (10%)	1483 (18%)
Graduate School	49 (5%)	154 (21%)	302 (100%)	188 (12%)	45 (8%)	91 (16%)	360 (100%)	186 (15%)	6 (2%)	480 (30%)	1861 (23%)
BMI											
Missing	982	5	19	3	4	6	0	22	9	25	1075
<20	0 (0%)	260 (36%)	104 (37%)	443 (29%)	272 (48%)	268 (47%)	116 (32%)	330 (27%)	51 (21%)	470 (30%)	2314 (33%)
20- <25	0 (0%)	19 (3%)	4 (1%)	65 (4%)	18 (3%)	24 (4%)	3 (1%)	18 (1%)	10 (4%)	33 (2%)	194 (3%)
25- <30	0 (0%)	311 (43%)	144 (51%)	633 (42%)	182 (32%)	189 (33%)	193 (54%)	551 (45%)	70 (29%)	579 (37%)	2852 (40%)
>=30	0 (0%)	127 (18%)	31 (11%)	366 (24%)	100 (17%)	85 (15%)	48 (13%)	319 (26%)	112 (46%)	503 (32%)	1691 (24%)
Smoking											
Missing	0	1	10	2	3	41	1	0	5	26	89
Never smoker	460 (47%)	325 (45%)	125 (43%)	594 (39%)	254 (44%)	255 (48%)	155 (43%)	572 (46%)	84 (34%)	804 (51%)	3628 (45%)
Normer smoker	245 (25%)	375 (52%)	154 (53%)	675 (45%)	233 (41%)	183 (34%)	183 (51%)	555 (45%)	62 (25%)	693 (44%)	3358 (42%)
Current smoker	277 (28%)	21 (3%)	13 (4%)	239 (16%)	86 (15%)	93 (18%)	21 (6%)	113 (9%)	101 (41%)	87 (5%)	1051 (13%)
Family history of CRC											
Missing	982	0	0	17	2	572	54	44	252	129	2052
No	0 (0%)	609 (84%)	249 (82%)	1292 (87%)	495 (86%)	0 (0%)	276 (90%)	1063 (89%)	0 (0%)	1183 (80%)	5167 (85%)
Yes	0 (0%)	113 (16%)	53 (18%)	201 (13%)	79 (14%)	0 (0%)	30 (10%)	133 (11%)	0 (0%)	298 (20%)	907 (15%)

Supplementary Table S2: *H. pylori* multiplex serology antigens

Sytematic name	Name	Full name	Acc. No ¹	Selected amino acids	Strain	Cut-off [MFI]	N+C
HP0010	GroEL	Chaperonin GroEL	AM_997163	1-547	G27	68	
HP0073	UreA	Urease alpha subunit	NP_206873	1-238	26695	95	
HP0231	HP0231	Hypothetical protein	NP_207029	1-265	26695	95	
HP0243	NapA	Neutrophil-activating protein	NP_207041	1-144	26695	66	
HP0305	HP0305	Hypothetical protein	NP_207103	1-184	26695	55	
HP0410	HpaA	Neuraminylactose-binding hemagglutinin homolog	NP_207208	1-249	26695	73	
HP0547	CagA N-Terminus CagA C-Terminus	Cytotoxin-associated antigen A	NP_207343	1-650 641-1186	26695 26695		654
HP0695	HyuA N-Terminus HyuA C-Terminus	Hydantoin utilization protein A	NP_207208	1-220 210-713	26695		114
HP0875	Catalase	Catalase	NP_207669	1-505	26695	107	
HP0887	VacA N-Terminus VacA C-Terminus	Vacuolating cytotoxin A	NP_207680	34-363 328-1008	26695 26695		64
HP1098	HcpC	Conserved hypothetical secreted protein	NP_207889	1-290	26695	30	
HP1104	Cad	Cinnamyl alcohol dehydrogenase ELI3-2	NP_207895	1-348	26695	88	
HP1564	Omp	Outer membrane protein	NP_208355	1-271	26695	137	

¹ NCBI reference sequence

Supplementary Table S3: Conditional logistic regression models, with adjustment for potential confounders

	Odds of CRC				
	Unadjusted model (1)	Model + education (2)	Model + BMI (3)	Model + smoking (4)	Fully adjusted (5)
ALL					
<i>N</i>	8,126	7,942	6,960	7,956	6,696
HP+VacA+ OR	1.11 ^{**}	1.10 [*]	1.11 [*]	1.10 [*]	1.09
HP+VacA+ CI	(1.01, 1.22)	(0.99, 1.21)	(1.00, 1.23)	(1.00, 1.21)	(0.98, 1.22)
p-value	0.038	0.069	0.062	0.059	0.117
BLACK ONLY					
<i>N</i>	798	762	748	776	716
HP+VacA+ OR	1.45 ^{**}	1.46 ^{**}	1.48 ^{**}	1.41 ^{**}	1.48 ^{**}
HP+VacA+ CI	(1.08, 1.95)	(1.08, 1.98)	(1.09, 2.02)	(1.05, 1.90)	(1.08, 2.03)
p-value	0.013	0.014	0.011	0.023	0.015

Notes:

^{***}Significant at the 1 percent level.

^{**}Significant at the 5 percent level.

^{*}Significant at the 10 percent level.