Supplementary Data

Comprehensive profiling of DNA methylation in Korean patients with colorectal cancer

Hyeran Shim[1,#], Kiwon Jang[2,#], Yeong Hak Bang[3,4,#], Hoang Bao Khanh Chu[1], Jisun Kang[1], Jin-Young Lee[1], Sheehyun Cho[1], Hong Seok Lee[1], Jongbum Jeon[2], Taeyeon Hwang[2], Soobok Joe[2], Jinyeong Lim[5], Ji-Hye Choi[5], Eun Hye Joo[5], Kyunghee Park[6], Ji Hwan Moon[6], Kyung Yeon Han[6], Yourae Hong[7], Woo Yong Lee[8], Hee Cheol Kim[8], Seong Hyeon Yun[8], Yong Beom Cho[8], Yoon Ah Park[8], Jung Wook Huh[8], Jung Kyong Shin[8], Dae Hee Pyo[8], Hyekyung Hong[8], Hae-Ock Lee[9,10], Woong-Yang Park[3,5,6,*], Jin Ok Yang[2,*] & Young-Joon Kim[1,11,*]

[1]Department of Biochemistry, College of Life Science and Biotechnology, Yonsei University, Seoul 03722, [2]Korea Bioinformation Center (KOBIC), Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, [3]Department of Digital Health, Samsung Advanced Institute for Health Science & Technology (SAIHST), Sungkyunkwan University, Seoul 06355, [4]Division of Hematology-Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, [5]Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul 06355, [6]Samsung Genome Institute, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea, [7]Department of Oncology, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium, [8]Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, [9]Department of Microbiology, College of Medicine, The Catholic University of Korea, Seoul 06591, [10]Department of Biomedicine and Health Sciences, Graduate School, The Catholic University of Korea, Seoul 06591, [11]LepiDyne Co., Ltd., Seoul 04779, Korea

*Corresponding authors. Young-Joon Kim, Tel: +82-2-2123-2628; Fax: +82-2-363-4083; E-mail: yjkim@yonsei.ac.kr; Jin Ok Yang, Tel: +82-42-879-8550; Fax: +82-42-879-8519; E-mail: joy@kribb.re.kr; Woong-Yang Park, Tel: +82-2-2148-9810; Fax: +82-2-2148-9819; E-mail: woongyang@skku.edu
#These authors contributed equally to this work.

MATERIALS AND METHODS

Statistical analysis

Prior to preprocessing, we compared the beta value distribution between raw and processed probes using PCA, to uncover sex- and batch-related biases in the raw beta values (Supplementary Fig. 5A and 6A). After normalization and filtering, we obtained high-quality harmonized data, effectively removing technical noise and sex-based biases, as demonstrated by the PC plots in Supplementary Fig. 5B and 6B. Using minfi's dmpfinder function (1, 2), we identified DMPs between tumor and normal samples, classifying them as either hyper- or hypomethylated based on our criteria (mean difference: 0.15, $q < 0.00001$). We calculated the odds ratio of enrichment for each DMP group according to genomic annotations using Fisher's exact test. To assess the prevalence of CIMP, we used a previously defined CIMP probe set (4,327 probes) (3) and the K-means algorithm to categorize tumor samples into CIMP-high, CIMP-low, or non-CIMP groups. In addition, Fisher's exact test was used to compare the significance of CIMP cluster proportions for each tumor characteristic. Survival analysis was performed by employing the log-rank test between the different CIMP groups. In order to select individual prognostic markers, we utilized a univariate Cox regression model, also known as proportional hazards regression, specifically for probes identified as CIMP markers (4). In this process, we selected the top 80 cg probes according to significant P-values ($<$ 0.006). For assessing methylation level of gene promoters, including that of the MLH1 gene, we calculated the mean of beta values of probes annotated as promoter-like regions (TSS1500, TSS200, 5' UTR, and first exon). To compare MSI-H and MSS subtypes within the CIMP-H group, we used selected genes with at least two probes in their promoter regions. Statistical analysis between groups was performed using t-test. Genes associated with the WNT signaling pathway were sourced from the Molecular Signatures Database(5). All statistical analyses were performed using MATLAB2022a (MathWorks Inc., USA) and R statistical software version 4.4.

REFERENCES

1. Aryee MJ, Jaffe AE, Corrada-Bravo H et al (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30, 1363-1369
2. Ritchie ME, Phipson B, Wu D et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47
3. McInnes T, Zou D, Rao DS et al (2017) Genome-wide methylation analysis identifies a core set of hypermethylated genes in CIMP-H colorectal cancer. BMC Cancer 17, 228
4. Sheaffer KL, Elliott EN and Kaestner KH (2016) DNA hypomethylation contributes to genomic instability and intestinal cancer initiation. Cancer Prev Res (Phila) 9, 534- 546
5. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP and Tamayo P (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 1, 417-425
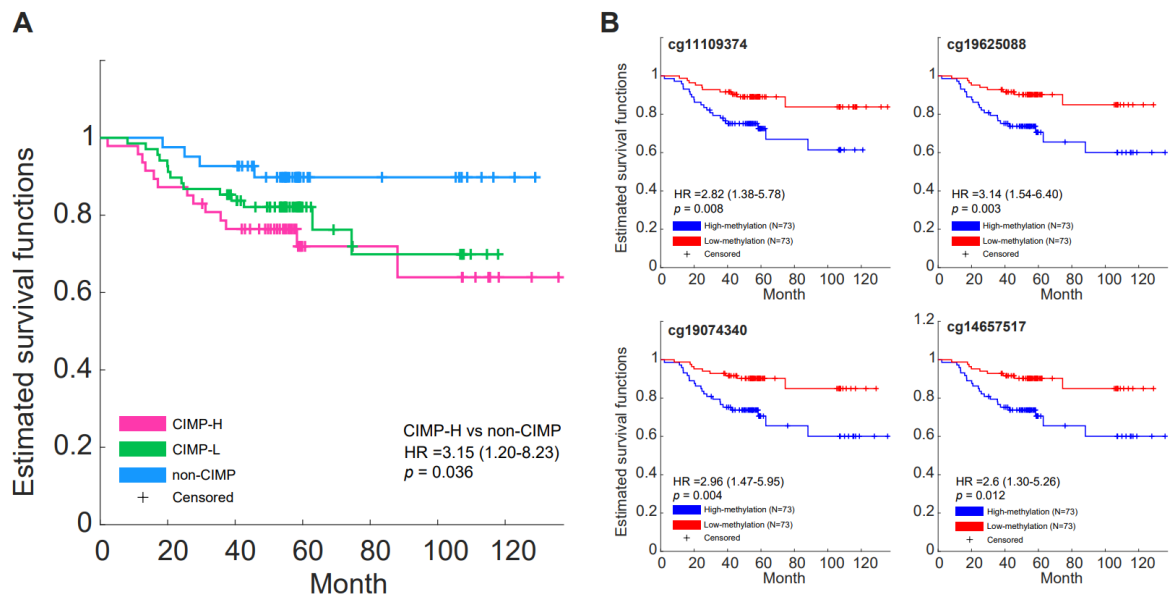
SUPPLEMENTARY FIGURES AND TABLE
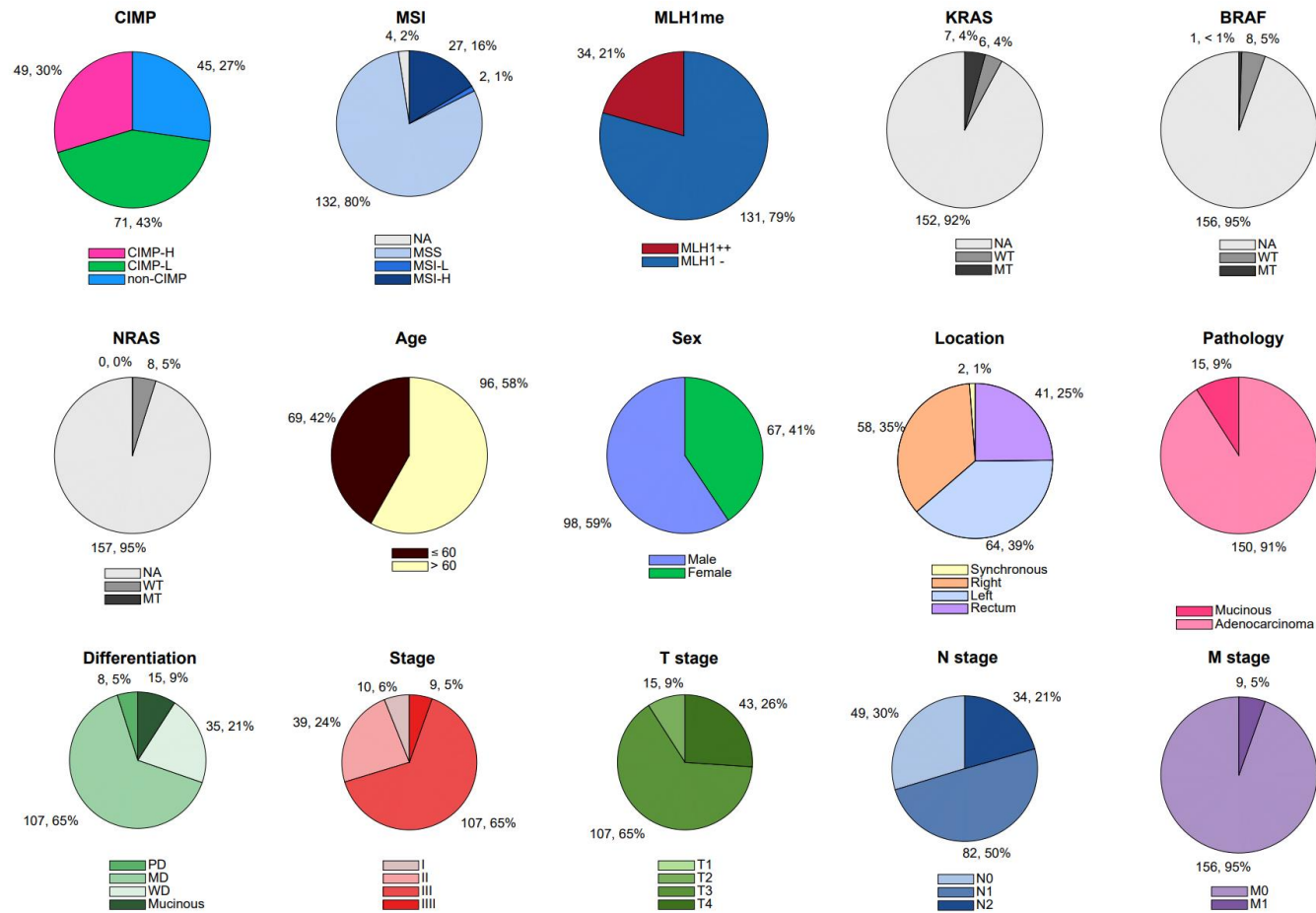
Supplementary Figure 1. Proportions of characteristics in 165 patients with colon cancer. Twelve clinical characteristics, including the CIMP group and MLH1 promoter methylation statuses, are shown. Abbreviations: CIMP, CpG island methylator phenotype; MSI, Microsatellite instability
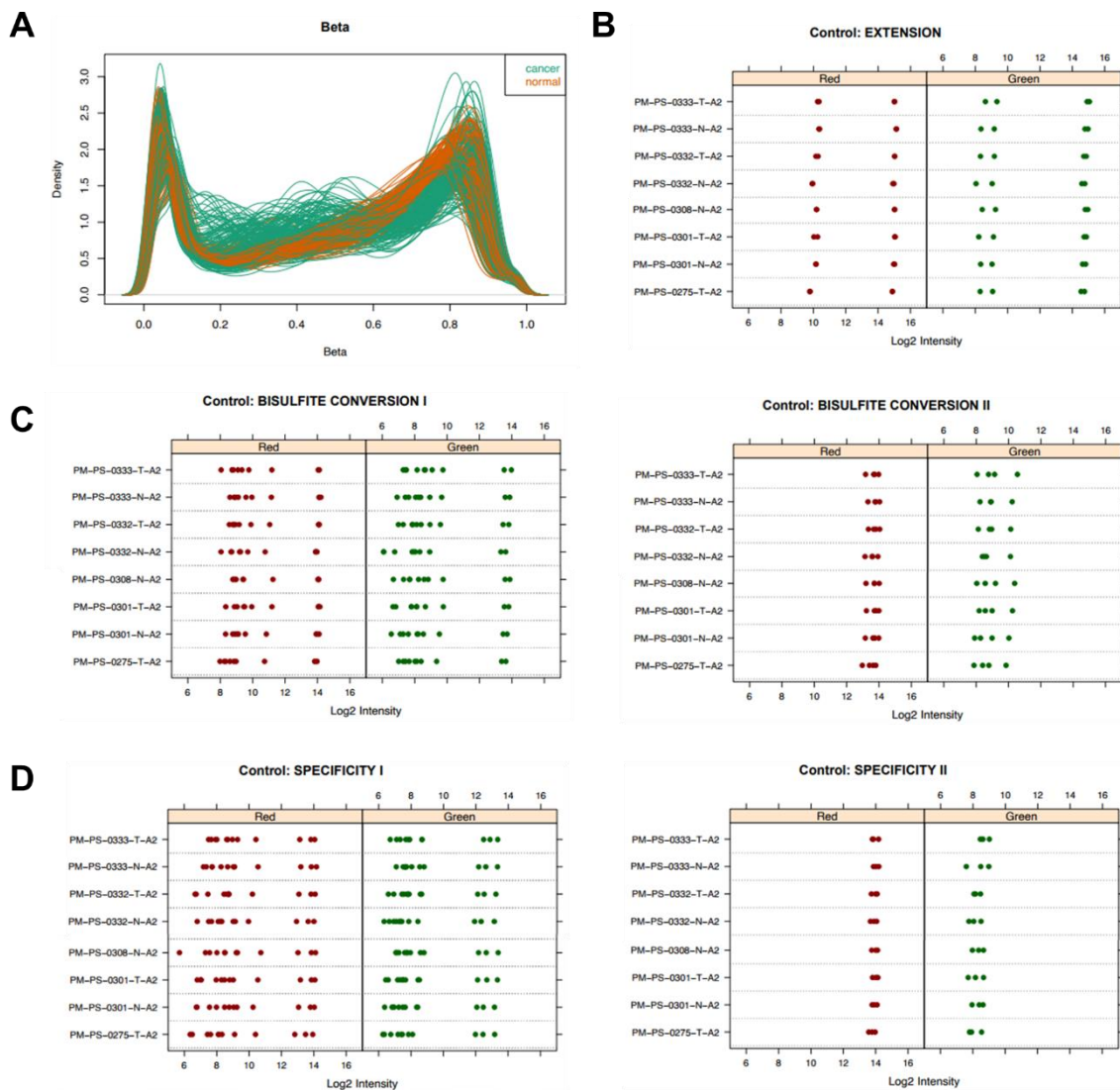
Supplementary Figure 2. Density plot of methylation beta values and control strip plots.

(A) Density plot of methylation beta values from individual samples (orange=normal, green=tumor). (B) Sample- and hybridization-independent extension control hairpin probe signals. Clear separation of control probe signals from background signals indicates successful extension. (C) Bisulfite conversion control probe signals show conversion success based on intensities at non-CpG control sites. Clear separation of converted DNA signals from non-converted DNA signals in type I indicate successful conversion results. High red (converted) and low green (unconverted) signals in type II indicate successful conversion results. (D) Extension specificity control probe signals show match-specific extensions through

intensities at non-polymorphic T control sites. Clear separation of matched DNA signals from non-matched DNA signals in type I indicate specific extensions. High red (matched) and low green (unmatched) signals in type II indicate specific extensions.

Supplementary Figure 3. Comparative analysis of overall survival associated with CIMP status. (A) Kaplan‑Meier survival curves illustrating the estimated overall survival differences among patients with high‑level CIMP (CIMP-H, magenta), low‑level CIMP (CIMP-L, green), and those without CIMP (non‑CIMP, blue). (B) Comparisons of patient survival based on the methylation status of four specific cg probe markers—cg11109374, cg19625088, cg19074340, and cg14657517. Blue denotes patients showing hypermethylation (representing 50% higher methylation levels of total patients) at each respective cg probe; red denotes patients exhibiting hypomethylation. Abbreviations: CIMP, CpG island methylator phenotype.
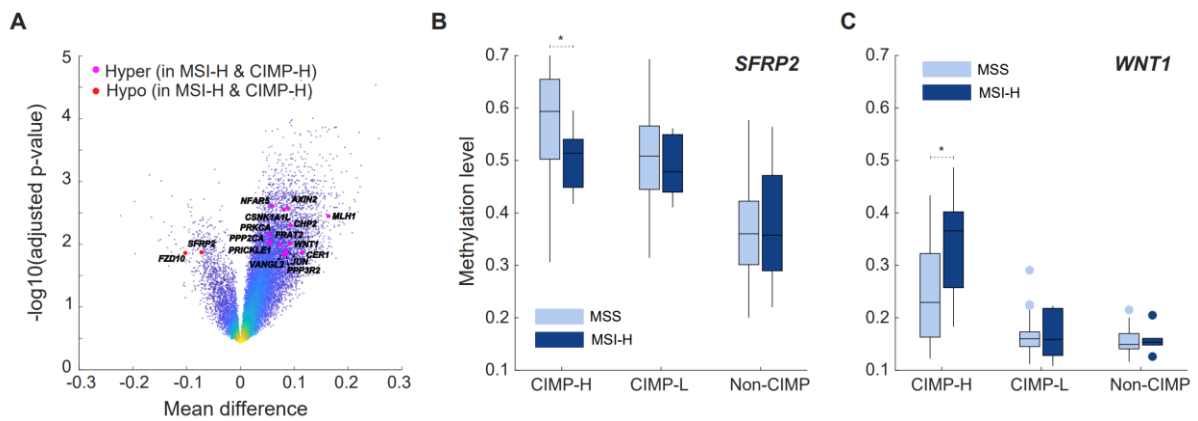
Supplementary Figure 4. Comparison of gene promoter methylation levels with respect to MSI status within CIMP-H samples. (A) Volcano plot for comparison of gene promoter methylation levels. Colored dots represent significantly different promoter methylation levels of WNT-related genes, including MLH1 (mean difference >5%, adjusted *p*-value <0.01). (B) Comparison of *SFRP2* gene promoter methylation on the basis of MSI status. (C) Comparison of *WNT1* gene promoter methylation on the basis of MSI status. Color bar for box plots: sky-blue = microsatellite stability (MSS), dark-blue = microsatellite instability-high (MSI-H).
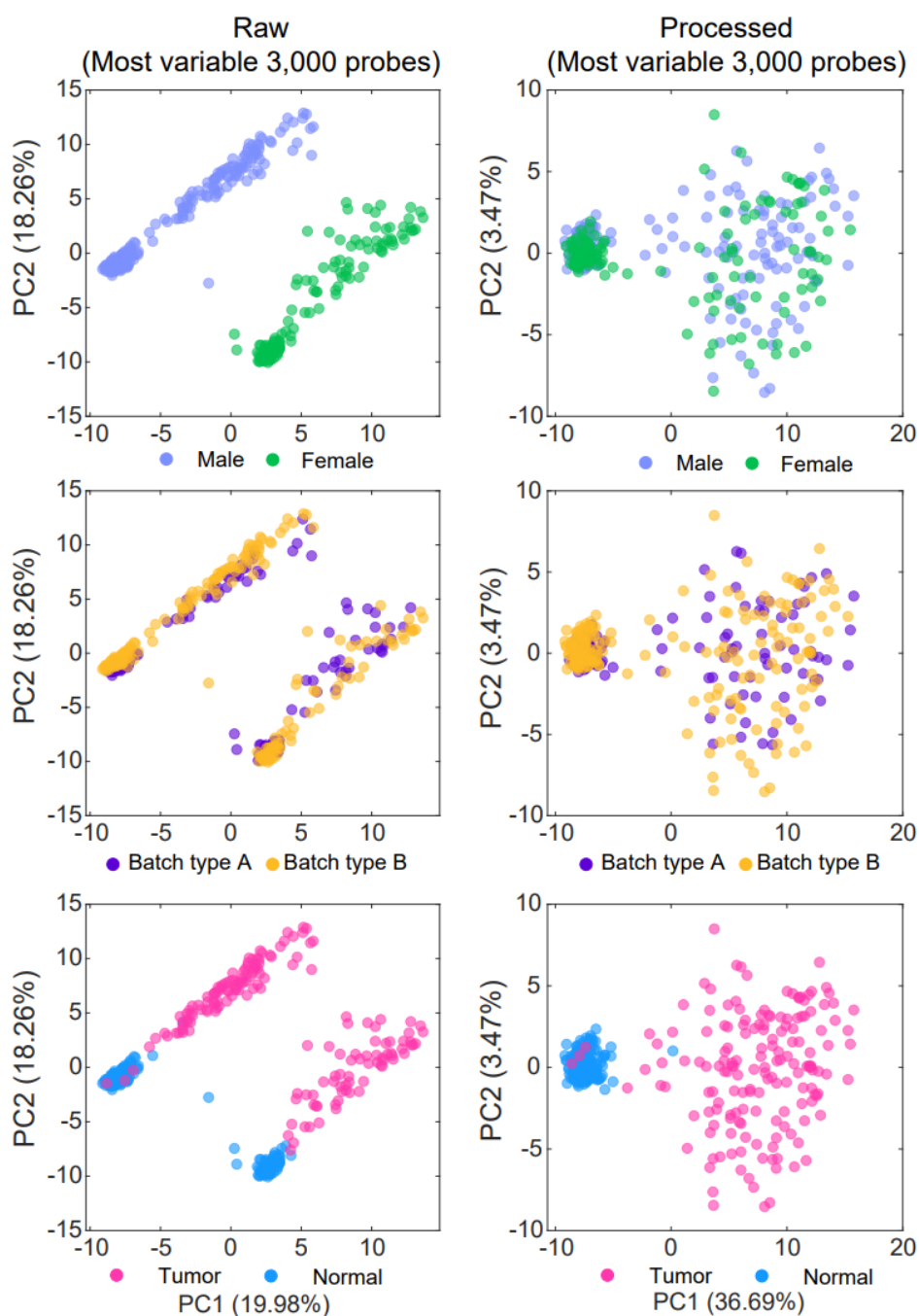
Supplementary Figure 5. Principal component (PC) plot for PC1 and PC2. Among the total probes, the top 3,000 variable beta values were used for PCA plot. We tested raw (left) and processed (right) probes according to gender (top: Male and Female), batch number (middle: batch types), and tumor status (bottom: Tumor and Normal) with PC1 and PC2.
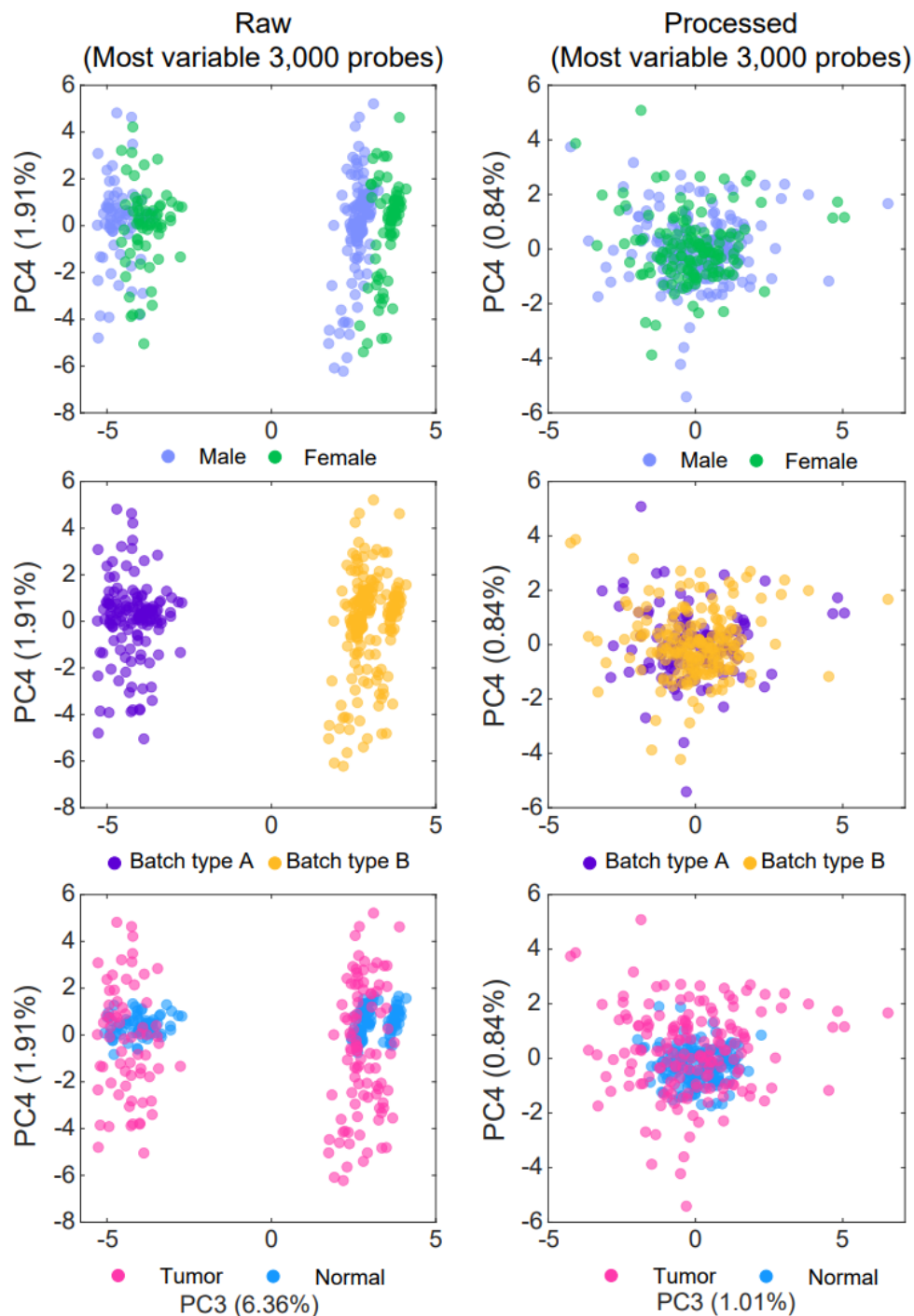
Supplementary Figure 6. Principal component (PC) plot for PC3 and PC4. Among the total probes, the top 3,000 variable beta values were used for PCA plot. We tested raw (left) and processed (right) probes according to gender (top: Male and Female), batch number (middle: batch types), and tumor status (bottom: Tumor and Normal) with PC3 and PC4.

Supplementary Table 1. List of CIMP probes highly correlated with CRC prognosis

cg probes	Coefficient of Cox univariate model (i.e. ln(hazard ratio))	P-value	Standard deviation of coefficient
cg11109374	2.769325881	0.000470193	0.79187257
cg18755829	2.688377406	0.00065692	0.789085182
cg19625088	2.834409552	0.000905637	0.854172458
cg19074340	3.072897982	0.000909664	0.926388884
cg04074140	2.628897152	0.001276858	0.816141021
cg14657517	3.1517635	0.001327182	0.98184671
cg24897255	2.836813484	0.001593904	0.898573177
cg23210365	2.710375901	0.001657673	0.861650398
cg05336094	2.768382433	0.001666599	0.880530934
cg13450005	2.846794479	0.001705544	0.907424
cg09987562	3.063413815	0.001943411	0.988597693
cg27567401	2.600267048	0.002012619	0.84195617
cg26780404	2.605847535	0.002073823	0.84620515
cg07193766	2.646192651	0.002124549	0.861321994
cg26985666	2.060553695	0.0022002	0.672990551
cg18061259	2.756900309	0.002436681	0.909539208
cg17754876	2.530300572	0.002698909	0.843405373
cg01952234	3.456854497	0.002727797	1.153493759
cg15107670	3.610645825	0.002806215	1.208299344
cg01557547	2.543898245	0.00285114	0.852699253
cg13663793	2.496847533	0.002950841	0.83989389
cg03595100	2.681562977	0.003013919	0.904004886
cg05587282	2.190557869	0.003243445	0.744158506
cg11283700	2.233018186	0.003367114	0.761584327
cg12714180	2.544605848	0.003554522	0.872872984
cg01229860	2.612082483	0.003644208	0.898418662
cg24825966	2.675636088	0.003817121	0.924899992
cg08382235	2.342341092	0.003947002	0.812645104
cg23655107	2.618223837	0.00403663	0.910596297
cg20457275	3.405004947	0.004155028	1.188008275
cg05663341	2.155109701	0.004220639	0.753223409
cg06287318	2.66061099	0.004306522	0.931981149
cg03669907	3.210042387	0.00441367	1.127527061
cg18552861	2.521057892	0.004488186	0.887183495
cg07724977	1.808536198	0.004703782	0.639816286
cg19570244	2.622199036	0.004705874	0.927717208
cg04999026	1.971601789	0.004731683	0.697972757
cg12789062	2.155415608	0.004747661	0.763337208

cg05781968	2.182745013	0.004853189	0.774951866
cg25668236	2.195753273	0.004985499	0.78197198
cg07565441	3.758741973	0.00508966	1.341783132
cg25667335	2.227173765	0.005101973	0.795270411
cg23636606	2.447750551	0.005214759	0.876243494
cg05899618	2.68052352	0.0052741	0.960830825
cg09786985	2.579591377	0.005521006	0.92959896
cg20518889	2.269517351	0.005552485	0.818404087
cg26379156	2.909260646	0.005804845	1.054612554
cg19126300	2.776128322	0.005840679	1.007086326
cg27610486	2.594362341	0.005899423	0.942266977
cg18146873	2.224021038	0.006104321	0.811057667
cg23105839	2.684688288	0.00627929	0.982384254
cg01177906	2.192361618	0.006441244	0.804704486
cg10470588	2.43981819	0.006534969	0.897105156
cg14215586	2.048206363	0.006888474	0.757980624
cg05490712	2.228236068	0.00719737	0.829089045
cg25566352	2.037511199	0.007463963	0.761571907
cg01615018	2.900526337	0.007620416	1.086974137
cg15190013	2.302288136	0.007642275	0.863095246
cg18319921	2.267932407	0.007657107	0.850423507
cg13690989	2.036072271	0.007736323	0.76447334
cg24190415	2.171477362	0.007743525	0.815409036
cg02767613	2.650037346	0.007779158	0.995690302
cg17439757	2.35655307	0.007832567	0.886187612
cg17617527	1.845594904	0.008056083	0.696527072
cg04599026	2.102000995	0.008122737	0.794129413
cg24631913	2.444110697	0.008378384	0.92705364
cg23719318	2.319289113	0.008411483	0.880155358
cg06500714	2.627132535	0.008691865	1.001216036
cg10814856	2.301765714	0.008711097	0.877468551
cg06566669	3.120428855	0.008722757	1.189762251
cg03753481	2.147434558	0.008770885	0.819363261
cg25892837	2.502977811	0.008838808	0.9559821
cg00095976	2.194881941	0.008913815	0.839233241
cg14650577	2.154830334	0.009054341	0.825606906
cg26168324	2.305368588	0.009412809	0.887812052
cg25577463	2.418717291	0.009482298	0.932370949
cg20134916	2.067865992	0.009505022	0.797377215
cg04431629	2.142774603	0.009538346	0.826645976
cg19380001	2.328394232	0.009561644	0.898545832
cg15585318	2.226113108	0.009867023	0.86268242