DESCRIPTION OF THE METHODS

Please use following references for the database and the web-platform :

> Müller, D. and B. Győrffy, ***EpigenPlot: An interactive web platform for DNA methylation-based biomarker and drug target discovery in colorectal cancer.*** Br J Pharmacol. 2025 Apr;182(7):1452-1465. doi: 10.1111/bph.17455. Epub 2025 Jan 27.

DNA methylation database construction

We compiled a DNA methylation database using publicly available colorectal tissue data from the Genomic Data Commons Data Portal (GDC) and Gene Expression Omnibus (GEO). Within GEO, we searched for series containing the terms "CRC," "colon," or "colorectal," restricting results to Homo sapiens. We selected studies with ≥10 samples that employed "Methylation profiling by array" or "Methylation profiling by genome tiling array," excluding datasets involving prior therapy or genetic disorders. Only series providing raw methylation data for normal colorectal mucosa, adenomas, and adenocarcinomas were included. Data from the Illumina HumanMethylation450 and HumanMethylationEPIC platforms were processed separately to mitigate batch effects arising from their differing sensitivity, specificity, and dynamic ranges. The resulting database contains CpG site and UCSC-defined gene region methylation levels (β-values).

Data processing

Raw signal intensity data was filtered based on detection p-values. Probes containing SNPs or with potential cross-reactivity were excluded. CpG sites were annotated using the manufacturer's manifest file. β-values were calculated as $\beta = M/(M+U+100)$, where M and U represent methylated and unmethylated signal intensities, respectively. Probe type bias (II vs I) was corrected using the BMIQ (beta-mixture quantile normalization) method.

Statistics

Global methylation differences between normal tissue, adenomas, and adenocarcinomas (CRC) were assessed using the Kruskal–Wallis test. For gene regions, the Kruskal–Wallis test was

followed by pairwise Mann–Whitney post hoc comparisons, with p-values adjusted using Bonferroni correction.

Data visualization

Analysis results and plots for both methylation arrays were generated using the EpigenPlot R Shiny-based web application (www.epigenplot.com).