Colorectal cancer risk in bowel adenomas based on lifestyle exposures, tissue preconditioning and DNA methylation

Jaim Sutton^{1,2*}, Morgan Moorghen³, Lai Mun Wang⁴, Christina Thirlwell⁵, Christodoulos Pipinikas⁵, Attila Lorincz^{1,6}

¹ Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Barts and The London School of Dentistry and Medicine, Queen Mary University of London, London, United Kingdom

² Institute of Reproductive and Developmental Biology, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, United Kingdom

³ St Mark's Hospital, London North West Healthcare NHS Trust, London, United Kingdom

⁴ John Radcliffe Hospital, Oxford University Hospitals NHS Trust, Oxford, United Kingdom

⁵ Medical Genomics Group, Cancer Institute, University College London, London, United Kingdom

⁶ Molecular Epidemiology Laboratory, Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Barts and The London School of Dentistry and Medicine, Queen Mary University of London, London, United Kingdom

* Corresponding author: Jaim Sutton email j.sutton@imperial.ac.uk or jaim.sutton@gmail.com

Abstract

Background: Colorectal cancer (CRC) is associated with patient demographics, lifestyle exposures and molecular alterations. However, it is not possible to determine which adenomas will progress to CRC, as ethically it is unacceptable to leave and follow adenomas. We hypothesised that certain lifestyle exposures at high levels could precondition exposed bowel tissue by changing and aging it, increasing the risks of deleterious DNA methylation and genetic alterations. We used a novel study design comparing adenomas with concurrent CRC (thus more likely exposed to deleterious lifestyle effects) to single adenomas in bowels with no history of CRC; we called these high (HR) and low-risk (LR) adenomas respectively.

Methods: We carried out a discovery and replication epigenome-wide association study (EWAS) on 106 HR and 111 LR adenomas, profiled with MethylationEPIC BeadChips. In order, to identify differentially methylated positions (DMP), regions (DMR), and DNAm (DNAmethylation) lifestyle exposures and risks, with adjustment for confounders, and gene ontology (GO) and pathway enrichment. Then, two open-source gene expression omnibus (GEO) validation datasets (52, 57 and 49, 48 HR and LR normal bowel tissues respectively) were analysed for these DNAm lifestyle exposures and risks, with adjustment for confounders.

Results: Our EWAS found 5 Bonferroni significant DMPs with absolute delta betas $\geq 5\%$, and 14 significant DMRs with absolute mean DMR delta betas $\geq 5\%$, replicated in the *GPX7*, *RGS3* and *TMEM135* cancer-associated genes. DNAm high alcohol exposures were strongly associated with increased risk of HR adenomas (odds ratio (OR) per standard deviation (SD) = 2.16 (95% confidence interval (CI) 1.55 - 3.09, p-value = 9.7×10^{-6})). In the validation datasets, DNAm high alcohol (ORperSD = 2.12 (95% CI 1.35 - 3.55, p-value = 2.0×10^{-3}) and ORperSD = 1.79 (95% CI 1.14 - 2.96, p-value = 1.7×10^{-2})), and high body mass index (BMI) exposures (ORperSD = 1.72 (95% CI 1.13 - 2.73, p-value = 1.5×10^{-2})) were associated with increased risk of HR normal bowel tissues.

Conclusions: High alcohol and BMI exposures may precondition normal bowel tissues and adenomas for increased risk of DNA methylation alterations associated with CRC progression. The DNAm exposure signatures and our newly identified genes may be useful epigenetic biomarkers for CRC prevention.

Keywords: colorectal cancer, adenoma, risk, epigenome-wide association study, epigenetics, DNA methylation, lifestyle exposures, alcohol, BMI, smoking

Background

CRC is the third most common cancer in males and females in the United Kingdom (UK) (1,2). It is a multifactorial disease thought to be caused by a combination of patient demographics, molecular alterations and lifestyle factors (3–5). Notably, numerous studies have found that high levels of alcohol consumption, BMI and tobacco smoking can be associated with increased risk of CRC (6–9) and poorer survival (10–12). Prevention is an important method for reducing CRC progression, and currently relies on the identification and removal of early bowel lesions such as adenomas, which are dysplastic growths in the bowel composed of abnormal cells with characteristic nuclear and cytoplasmic aberrations (13). As adenomas can progress into CRC through an adenoma to carcinoma sequence, that is commonly thought to start in abnormal epithelial stem cells, which grows into a small adenoma, then a larger adenoma and then to invasive cancer (14). Traditionally, this adenoma to carcinoma sequence is thought to occur through the selection of specific cancer gene alteration sweeps with each increase in adenoma size, low to high grade dysplasia histology and tubular to villous architecture stage (15,16).

Importantly, not all adenomas will progress into CRC, and it's difficult to predict those that will versus those that do not. As, it is unethical to leave adenomas in patients' bowels to follow their natural history and molecular changes, and cell culture and animal model systems are not able to fully define all CRC drivers. Therefore, current clinical, histopathological and somatic molecular predictors of CRC progression have poor accuracy. However, recent studies have started to identify putative cancer driver gene alterations in normal tissues that may clonally persist throughout pre-cancer lesion growth to invasive cancer, and therefore be sensitive biomarkers for cancer progression (17–20). Notably, such cancer gene alterations have been found in normal bowel mucosa epithelial cells (21), and also in skin, lung and gastrointestinal normal tissues that are more exposed to environmental and lifestyle exposures (20,22). Furthermore, a recent study has found that carcinogenic chemical exposures can also cause genetic alterations in mouse normal tissue that may create a high-risk tissue environment for tumour growth and progression (23).

Epigenetic changes are associated with lifestyle exposures and can be detected as altered DNA methylation patterns in blood and tissues. Notably, EWASs have been carried out that have identified DNAm signatures associated with high alcohol consumption, high BMI and current smoking behaviour exposures. These studies encompass genome-wide cytosine-phosphate-

guanine (CpG) methylation differences in the extreme levels of each exposure, compared to those without the exposure (24–26). Additionally, high alcohol and current smoking lifestyle exposure effects on DNA methylation have been found in normal and cancer tissues, especially in individuals more exposed to these lifestyle factors (27–29). Furthermore, a recent study has also found DNA methylation alterations in CRC, postulated to have originated in normal bowel tissue, and then maintained during adenoma to CRC progression (30).

We hypothesised that high levels of certain lifestyle exposures could precondition susceptible bowel tissue, by changing and aging these tissues, thereby increasing DNA methylation alterations and associated risk of CRC progression. Then, that in our novel case-control study design, patients' adenomas in a bowel environment with concurrent cancer are more likely subjected to high levels of deleterious lifestyle exposures compared to single adenomas in a bowel environment with no history of CRC (called HR and LR bowel adenomas, respectively), and may help determine DNA methylation biomarkers for CRC progression. We, therefore, carried out a EWAS in HR compared to LR bowel adenomas in discovery and stratified study-site replication cohorts. The samples were profiled with Illumina MethylationEPIC BeadChips to identify DMPs, DMRs, and DNAm lifestyle exposure levels and association with risk of HR bowel adenomas, with adjustment for confounders, and then GO and pathway enrichment. To compare our results to a partially similar validation study, HR and LR normal bowel tissue GEO DNA methylation datasets were then downloaded and also analysed for DNAm lifestyle exposure levels and association with risk of HR normal bowel tissues (31).

Methodology

Study design

The EWAS consisted of 106 HR and 111 LR patients' bowel adenomas collected retrospectively from formalin-fixed paraffin-embedded (FFPE) tissue blocks at St Mark's Hospital in London and the John Radcliffe Hospital in Oxford, England, as part of the CRC-adenoma-normal (CAN) tissue and clinical data collection (Figure 1). Only sporadic adenomas graded according to the World Health Organization (WHO) tumour classification guidelines (4,32), were included in this study. The adenomas were then further classified from these guidelines, as either having low-grade dysplasia (LGD) or high-grade dysplasia (HGD) histology and tubular adenoma (TA) or tubulovillous adenoma (TVA) or villous adenoma (VA) architecture (32). Patients were excluded if they were under 18 years of age, had known inherited CRC syndromes, inflammatory bowel disease (IBD), and if they

had undergone neoadjuvant therapy. An adenoma was classified as HR, if it came from a patients' bowel environment that contained at least one adenoma ≥ 8 mm with a fully separate (both lesions bordered by morphologically normal tissue) concurrent CRC, identified from 2010 to the end of 2015 in their hospital pathology records. While an adenoma was classified as LR if it came from a patients' bowel environment that contained a single adenoma ≥ 8 mm without any other concurrent neoplasia (adenomas, serrated lesions and CRCs), with no CRCs in their history and at least 4 years of follow-up, identified from 2010 to the end of 2012 from their hospital pathology records. It is important to note that if the LR adenomas contained HGD on histology, we retained their classification, but this was adjusted for in future analyses. Additionally, the HR bowel adenomas were originally individually matched to LR bowel adenomas by age at diagnosis, gender, study-site and adenoma lesion-size. However, some HR bowel adenomas were then later excluded, specifically if they came from bowels that also contained concurrent microsatellite instability positive (MSI +) CRCs, as recorded in their hospital pathology records. This was in-case these concurrent MSI + CRCs, which are thought to be associated with inflammation, directly caused a HR bowel environment that influenced the DNA methylation alterations in the adenomas and biased the analysis. Additionally, for each patient, their age, gender, study-site and adenoma bowel-side, lesion-size, dysplasia and architecture characteristics were also collated from their hospital pathology records. Then, these characteristics were compared between the patients with HR and LR bowel adenomas to identify any differences, using the student's t-test or chi-square test for the continuous and categorical variables respectively. The study was approved by proportional ethical review from the Northern Ireland ethics committee, Integrated Research Application System (IRAS) project identification 152955, approval reference 15/NI/0142.

DNA methylation profiling and quality control

Adenoma tissue was macro-dissected from pathologist annotated FFPE tissue sections for each patient. Sample DNA was then extracted and bisulfite converted using the Zymo EZ DNA Methylation Kit. Then profiled by the Illumina MethylationEPIC BeadChips at University College London (UCL) Genomics, Institute of Child Health, in London using standard Illumina protocols. The EPIC DNA methylation data was then pre-processed through quality control (QC) steps to produce pre-processed sample beta values, mainly through The Chip Analysis Methylation Pipeline (ChAMP) and minfi pipelines (33–35). These QC steps included single-nucleotide polymorphism (SNP), cross-hybridisation and X and Y chromosome probe filtering (36,37), gender checking, omicsPrint sample relationship genotyping (38), beta-mixture quantile normalization (BMIQ) (39),

singular value decomposition (SVD) identification of batch effects through dimension reduction, and ComBat batch effect correction (40,41). From the pre-processed DNA methylation data, the sample beta values were then calculated, by dividing each CpG site M intensity value by their combined target CpG site unmethylated (U) and methylated (M) intensity values ($\text{beta value} = M / (M + U)$). Where, the beta values are continuous variables between 0 (no methylation) and 1 (completely methylated). The sample beta values were then used in the following analyses.

High and low-risk bowel adenoma epigenome-wide association study

Differentially methylated positions

DMP analysis was then carried out using the Confounder Adjusted Testing and Estimation (CATE) package (42), with adjustment for age, gender, study-site and adenoma bowel-side, lesion-size, dysplasia, architecture and 19 latent confounders. The CATE package used a linear model to test for DMPs between the HR and LR bowel adenomas in multiple individual CpG site beta values. It then also identified latent confounders not correlated with case-control status, by surrogate variable analysis (SVA), that were included and adjusted for in the model (42,43). For each model, the DMP p-values were plotted against their expected p-values in quantile-quantile (Q-Q) plots and lambda values calculated, to visualise and identify any DMP p-value genomic inflation. A lambda value of 1.00 to 1.15 represented no or acceptable genomic inflation, respectively. While a lambda value above this, represented high genomic inflation and increased likelihood of DMP false positives, most likely due to un-accounted systematic differences between the HR and LR bowel adenomas. However, any models with high lambda values and inflated DMP p-values were corrected using the BACON (BC) package, that is especially designed to control for EWAS inflation (44). The DMP p-values were then adjusted for multiple testing, as the MethylationEPIC BeadChip includes many individual CpG site tests. With, the conservative Bonferroni and/or less conservative Benjamini-Hochberg false discovery rate (BH FDR) multiple testing correction methods used to determine significant DMP delta beta values with p-values $< 7 \times 10^{-8}$ and q-values < 0.05 , respectively (45,46). The DMP delta beta values were calculated from the mean CpG site beta value of all HR bowel adenoma samples minus the mean CpG site beta value of all LR bowel adenomas.

Differentially methylated regions

DMR analysis was then carried out using the DMRcate package, with adjustment for age, gender, study-site and adenoma bowel-side, lesion-size, dysplasia, architecture and with/without principal

components (PC) 1 and 2 (47). DMR analysis identifies regional CpG differences for groupings of CpG sites in the same direction. The DMRcate package computed a kernel-smoothed estimate against a null comparison to identify DMR mean delta values between the HR and LR bowel adenomas, across multiple CpG site beta values with a minimum of 7 CpG site probes. DMRs with Stouffer p-values < 0.05 and smoothed estimate FDR q-values < 0.05, were then used to determine significant DMR mean delta values. A Stouffer value is a summary transformation of individual CpG site FDRs used to rank DMR significance, and PCs 1 and 2 measured the variability in the dataset above 5%. In both the DMP and DMR analyses, positive and negative effect values represent hypermethylation and hypomethylation of the HR compared to the LR bowel adenomas, respectively.

Gene ontology and pathway enrichment

GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was then carried out using the MissMethyl package, to identify significantly enriched functions and biological pathways for the DMP findings with BH FDR p-values < 0.05 and then for p-values < 0.05. The MissMethyl package was used, as it takes into account the number of CpG site probes per gene in the MethylationEPIC BeadChip during analyses (48).

DNAmethylation lifestyle exposures

In, a previously published study, DNAm exposure signatures for overweight/obese BMI, moderate to heavy alcohol consumption and current smoking behaviour were identified from genome-wide CpG methylation differences in the extreme levels of each exposure, compared to those without the exposure. These DNAm high BMI, high alcohol and current smoking exposure signatures consisted of 1109, 450 and 233 CpG sites and corresponding regression coefficients, respectively. Additionally, these DNAm exposure signatures each consisted of a combination of hypermethylated and hypomethylated CpG sites. Notably, the DNAm high alcohol exposure signature included CpG sites for the *SLC7A11* gene, previously found in the literature to be associated with DNAm alterations representing high alcohol consumption (49). We then used these pre-defined CpG site specific DNAm exposure signatures, to further calculate DNAm BMI, alcohol and smoking exposure levels in our patients' HR and LR bowel adenomas. Where, for each exposure and patient, a DNAm exposure score was calculated from the total sum of each of their DNAm signature CpG site beta values multiplied by their corresponding published DNAm signature regression coefficients. The regression coefficients were used to weight their corresponding CpG

site beta values in our study according to their importance in each pre-defined DNAm exposure signature. DNAm exposure scores were then standardized to z-scores and used in unconditional logistic regression analyses, to determine if the DNAm lifestyle exposure levels measured in our patients' adenoma bowel tissue were associated with increased risk of HR bowel adenomas. In these predictive models, patients' HR and LR bowel adenoma status was used as the outcome variable and their EWAS measured adenoma DNAm exposure z-scores were used as the predictor variables with adjustment for age, gender, study-site and adenoma bowel-side, lesion-size, dysplasia, architecture and with/without PCs 1 and 2. These models were classified as significant if they had p-values < 0.05, as multiple testing correction was not required for these pre-defined DNAm signatures, as they are classified as a single test.

Stratification and study-site replication

Our 106 HR and 111 LR bowel adenoma EWAS dataset, was then also then stratified into St Mark's Hospital (total 138 bowel adenomas; 67 HR and 71 LR) and John Radcliffe Hospital (total 79 bowel adenomas; 39 HR and 40 LR) study-sites (Figure 1). Then, for each study-site dataset the following analyses were carried out. Firstly, DMP analysis using the CATE package (42), with adjustment for age, gender, and adenoma bowel-side, lesion-size, dysplasia, architecture and 8 latent confounders. Then, DMR analysis using the DMRcate package, with adjustment for age, gender, and adenoma bowel-side, lesion-size, dysplasia, architecture and with/without PCs 1 and 2 (47). The previously calculated HR and LR bowel adenoma DNAm lifestyle exposure z-scores, were then used in study-site unconditional logistic regression analyses, with adjustment for age, gender and adenoma bowel-side, lesion-size, dysplasia, architecture and with/without PCs 1 and 2. For all the study-site analyses, the same correction methods and p-value significant thresholds were used as in the combined analyses. Apart from the John Radcliffe Hospital study-site model analyses, where adenoma bowel-side was not included and adjusted for, due to an imbalance between the cases and controls.

Validation in high and low-risk normal bowel tissue

The open-source Gene Expression Omnibus (GEO) HR and LR normal bowel tissue datasets (GSE132804 - GPL21145 and GPL13534) (31), were downloaded from www.ncbi.nih.gov/gds and used for validation. This dataset included similarly classified HR and LR normal bowel tissue patient samples to the HR and LR bowel adenomas. Where, HR was defined as normal bowel tissue in a patients' bowel with concurrent CRC, and LR as normal bowel tissue in a patients' bowel with no

concurrent adenomas; there was no data on adenomas in this GEO dataset. From these datasets, Illumina MethylationEPIC BeadChip and clinical data from 109 normal bowel tissue samples (52 HR and 57 LR) and Illumina Methylation450K BeadChip and clinical data from 97 normal bowel tissue samples (49 HR and 48 LR), collected from 2017 to 2019 and 2012 to 2016 respectively, were selected. We, then applied the same pre-process (probe filtering, normalisation and relevant batch effects correction) and DNAm exposure and risk association analyses steps to these HR and LR normal bowel tissue datasets, as previously described for our HR and LR bowel adenoma study, so they could be more comparable. Notably, for the DNAm exposure analyses, we used the previously described published study (50) that identified the pre-defined DNAm high BMI, high alcohol and current smoking exposure signatures, to further calculate these DNAm exposure levels in patients' normal bowel tissue (Figure 1). Where, for each exposure and patient, a DNAm exposure score was calculated from the total sum of each of their DNAm signature CpG site beta values multiplied by their corresponding published DNAm signature regression coefficients. These DNAm lifestyle exposure scores were then standardized to z-scores for subsequent analyses and exposure comparison. Unconditional logistic regression analyses were then carried out to determine if the DNAm lifestyle exposure levels measured in patients' normal bowel tissue were associated with risk of HR normal bowel tissue. In these models, patients' HR and LR normal bowel tissue status was used as the outcome variable and their normal bowel tissue DNAm exposure z-scores as the predictor variable with adjustment for age, gender and with/without PCs 1 and 2. These models were classified as significant if they had p-values < 0.05. Additionally, both HR and LR normal bowel tissue datasets only included left normal bowel-sided tissue. However, these should compare well with the HR and LR bowel adenoma dataset, which was mostly left-bowel-sided.

Statistical analyses

A flowchart outlining all of the study and GEO datasets and analyses are shown in Figure 1, and further dataset details can be found in the supplementary section and in the previously published paper. All statistical analyses were carried out with R version 3.6 and annotation with the IlluminaHumanMethylationEPICanno.ilm10b2.h19 package and GRCh37/hg19 build.

Results

Study population characteristics

The HR and LR bowel adenoma study characteristics are described in Table 1. Our study included 64% (n = 138) and 36% (n = 79) patients from the St Mark's and John Radcliffe Hospitals, respectively. The number of left versus right-sided bowel specimens was significantly different in the HR compared to LR bowel adenomas (chi-square test p-value < 0.001). With overall more adenomas detected on the left bowel-side but with a relative enrichment for HR bowel adenomas on the right bowel-side. Further details of the patients' HR bowels are also described in Table S1. Notably, it was found that 62% of the HR bowels contained 1 to 3 adenomas concurrently.

Methylation BeadChip quality control

The methylation levels of 866,836 CpG positions across the genome were measured using the MethylationEPIC BeadChip in the HR and LR bowel adenomas, then pre-processed through QC steps. During these steps CpG positions were removed that failed pre-specified QC criteria. The final number of CpG positions per sample were 713,255 with the following probes removed: 39,117 that failed hybridisation at detection p-value > 0.01, 5,537 with a bead count < 3 in at least 5% of samples, 2,457 non-CpG probes, 91,005 with SNPs close to the 3' end, 11 that could bind to multiple regions in the genome, and 14,566 in the X and Y chromosomes (34,35). Sample well position, BeadChip and plate number were found to be significant batch effects, and the sample beta values were then adjusted for these batch effects (40,41). All samples were then found to have good beta density distributions, some case-control clustering, identical clinical and bioinformatic calculated genders and were not related to each other by SNP relationship analysis (38) (Figure S1).

High and low-risk bowel adenoma epigenome-wide association study analysis

Differentially methylated positions

DMP analysis after confounder adjustment and genomic inflation correction, identified 5 DMPs with Bonferroni p-values < 7×10^{-8} and absolute delta beta values $\geq 5\%$, and 95 DMPs with BH FDR q-values < 0.05 and absolute delta beta values $\geq 10\%$ (Figure S3). Genomic inflation correction was able to reduce the model lambda from 2.10 to 1.09, resulting in low levels of DMP genomic inflation and thus relatively decreased false positives (Figure S2). Notably, three of the DMPs that passed Bonferroni significance were found in the previously described *DIAPH3*, *RGS3* and *TMEM135* cancer-associated genes (51–53). Additionally, the most significant annotated DMP was

hypomethylation in the *DIAPH3* gene, which plays a role in actin remodelling, epithelial-mesenchymal transition (EMT) and cancer signalling (delta beta = - 6.7%, BH FDR q-value = 1.6×10^{-3}). While the annotated DMP with the largest absolute delta beta value was hypomethylation in the *ANK2* gene, which plays a role in the structural part of cytoskeleton and enzyme binding (delta beta = - 8.1%, BH FDR q-value = 2.7×10^{-2}). Interestingly, most of our highly significant DMP delta beta values were hypomethylated in the HR versus the LR bowel adenomas, and generally spanned widely across the human genome (Figure S3). The DMPs with delta beta values > 5% and BH FDR q-values < 0.05 are shown in Table 3. Notably, some of these tabled DMPs were found to be associated with lifestyle factors, epithelium, adenomas, progression and CRC, and other cancers in the literature. With gene functions encompassing telomere length, EMT, and cancer growth, invasion and signalling pathways (Table S2).

Differentially methylated regions

DMR analysis after confounder adjustment, identified 2,074 DMRs, and of these 14 DMRs had absolute mean delta beta values $\geq 5\%$. The most significant DMR was hypermethylation in the *GPX7* gene, that plays a role in DNA damage and double-strand break protection (mean delta beta = 6.2%, Stouffer p-value = 4.26×10^{-12}). Additionally, the DMR with the largest absolute mean delta beta value was hypermethylation in the *MCIDAS* gene that plays a role in transcription coactivator activity (mean delta beta = 10.4%, Stouffer p-value = 1.83×10^{-9}). The DMRs with mean delta beta values > 5% and Stouffer p-values < 0.05, with/without PCs adjustment were similar and mostly hypermethylated for the HR compared to LR bowel adenomas (Tables S3 and 4). Notably, some of these tabled DMRs were found to be associated with lifestyle factors, epithelium, progression and CRC, and other cancers in the literature. With gene functions encompassing transcriptional factor activity, DNA damage, immunity, and cancer signalling pathways (Table S4).

Gene ontology and pathway enrichment

GO and KEGG pathway enrichment analysis of the DMPs with significant BH FDR p-values < 0.05, did not find any terms and pathways enriched in the HR bowel adenomas at BH FDR q-values < 0.05, most likely due to low sample size power. However, the GO and KEGG pathway enrichment analysis of the DMPs with significant p-values < 0.05, found many terms and pathways enriched in the HR bowel adenomas at BH FDR q-values < 0.05 (Table S5). Such as terms associated with metabolic, cellular and molecular processes, bacterial invasion of epithelium, metabolism, and

cancer and associated signalling. All of which, if aberrantly methylated, could potentially increase the risk of adenoma to CRC progression in HR bowel environments.

DNA methylation lifestyle exposures in bowel tissue

High and low-risk bowel adenoma tissue

In the HR and LR bowel adenoma dataset, the number of CpG sites and beta-values that were available in our data corresponding to the published DNAm exposure signature CpG sites were as follows: 1021/1109, 413/450 and 211/233 for the BMI, alcohol and smoking exposures respectively. DNAm exposure logistic regression analyses, after adjustment for confounders, identified significant associations for DNAm high alcohol and high BMI exposures with the risk of HR bowel adenomas, at p-values < 0.05. Notably, a one SD unit increase in DNAm high alcohol exposure score was found to be strongly significantly associated with an increased risk of HR bowel adenomas (ORperSD = 2.16 (95% CI 1.51 - 3.09, p-value = 9.7×10^{-6})). While a one SD unit increase in DNAm high BMI and then current smoking exposure scores were found to be weakly and not significantly associated with a decreased risk of HR bowel adenomas, respectively (Table 2).

Study-site stratification and replication

DMP analyses after confounder adjustment and genomic inflation correction, found significantly replicated DMPs with delta beta values in the same direction in both study-sites, at BH FDR p-values < 0.05 for the larger St Marks Hospital and at p-values < 0.05 for the smaller John Radcliffe Hospital study-sites, and previously in the combined analysis. These replicated DMPs were found in the previous RGS3 and TMEM135 cancer-associated genes (52,53) in a novel *SRP14-AS1* gene, and in the open sea cg22240312 and cg17332384 sites (Tables 3 and S2). Genomic inflation correction was also able to reduce the model lambdas from 1.89 to 1.09 for the St Mark's Hospital, and 1.62 to 1.15 for John Radcliffe Hospital study-sites (Figure 1). Resulting in low levels of DMP p-value genomic inflation and thus relatively decreased false positives for both study-sites. DMR analyses after confounder adjustment, then found the *GPX7*, *XKR6*, *LBX2* genes to be significantly replicated DMRs in both hospital study-sites, with mean delta beta values in the same direction and Stouffer p-values < 0.05, consistent with the earlier combined analysis (Tables 4 and S4). Notably strong effects were seen for the DMR in the *GPX7* gene, for both St Marks Hospital (DMR mean delta beta = 6.3%, Stouffer p-value = 8.37×10^{-6}) and John Radcliffe Hospital study-sites (DMR mean delta beta = 15.5%, p-value = 3.46×10^{-3}) (Table 4). This was also the only DMR gene

to be similarly and significantly replicated in both study-sites, after further PC adjustment (Table S3). Then, the DNAm exposure logistic regression analyses after adjustment for confounders only found DNAm high alcohol exposure to be significantly associated with the risk of HR bowel adenomas at p-values < 0.05, in both study sites and the previous combined analysis. With, a one SD unit increase in DNAm high alcohol exposure score found to be significantly associated with an increased risk of HR bowel adenomas for the St Mark's Hospital (ORperSD = 1.85 (95% CI 1.24 - 2.84, p-value = 3.5×10^{-3})), and John Radcliffe Hospital study-sites (ORperSD = 3.44 (95% CI 1.91 - 7.00, p-value = 1.6×10^{-4})). While a one SD unit increase in DNAm high BMI exposure score was found to be associated with a decreased risk of HR bowel adenomas for each study-site, but only significantly for the smaller John Radcliffe Hospital study-site. Then a one SD unit increase in DNAm current smoking exposure score was found to be associated with a decreased risk of HR bowel adenomas for each hospital study-site, but not significantly (Table 2). The study-site DNAm exposure logistic regression analyses with PCs adjustment are also shown in Table 2. Where only the DNAm high alcohol exposure score was found to be significantly associated with an increased risk of HR bowel adenomas, with and without PC adjustment, in both study-sites and the previous combined analysis.

Validation in high and low-risk normal bowel tissue

In the GEO GSE132804 Illumina MethylationEPIC BeadChip HR and LR normal bowel tissue GPL21145 and GPL13534 validation datasets, the number of CpG sites and beta-values that were available out of the published DNAm exposure signature CpG sites were as follows: 1056/1109, 428/450 and 215/233 for GPL21145, and then 1092/1109, 445/450 and 228/233 for GPL13534, for the BMI, alcohol and smoking exposures respectively. Then, the DNAm exposure logistic regression analyses, after adjustment for confounders, only identified significantly replicated associations for DNAm high alcohol exposure with the risk of normal bowel tissue, at p-values < 0.05, in both validation datasets and all previous study-site analyses. With a one SD unit increase in DNAm high alcohol exposure score found to be significantly associated with an increased risk of HR normal bowel tissues for the GPL21145 (ORperSD = 2.12 (95% CI 1.35 - 3.55, p-value = 2.0×10^{-3})), and GPL13534 datasets (ORperSD = 1.79 (95% CI 1.14 - 2.96, p-value = 1.7×10^{-2})), and in the previous combined and study-site replicated analyses. While a one SD unit increase in DNAm current smoking exposure score was found to be significantly associated with a decreased risk of HR normal bowel tissue in both validation datasets, but only strongly for the GPL13534 dataset (ORperSD = 0.28 (95% CI 0.14 - 0.50, p-value = 8.1×10^{-5})), and not significantly in any of the

previous combined and study-site replicated analyses. Then, a one SD unit increase in DNAm high BMI exposure score was associated with an increased risk of HR normal bowel tissue in both validation datasets, but only significantly for the GPL21145 dataset (ORperSD = 1.72 (95% CI 1.13 - 2.73, p-value = 1.5×10^{-2})) (Table 2). The validation DNAm exposure logistic regression analyses with PCs adjustment for both datasets are also shown in Table 2.

Discussion

In summary, our HR versus LR bowel adenoma EWAS, found highly significant and replicated DMPs with absolute delta beta values $\geq 5\%$. We also identified a significantly replicated DMR with absolute mean DMR delta beta values $\geq 5\%$. These are large and significant effect sizes in what might be diagnostically important and druggable CRC genes. Our results are novel for potentially predicting HR adenomas using molecular methods. Some of our observed effects were seen in genes previously associated with cancer (52–54). Another important result, is that a previously derived and published DNAm high alcohol exposure signature (25), found that DNAm high alcohol bowel exposure levels were significantly associated with increased risk of HR bowel adenomas in our data. While, in the validation dataset, both DNAm high alcohol and high BMI bowel exposure levels were found to be significantly associated with increased risk of HR normal bowel tissue.

It could be, that high levels of DNAm measured lifestyle exposures in a patients' bowel could precondition the exposed normal and later adenoma tissue by changing and aging this tissue. This could then increase the risk and occurrence of early clonal or later sub-clonal DNA methylation alterations in the epithelium stem cells, other normal and adenoma bowel tissue, creating a high-risk tissue environment for tumour growth and progression. Notably, for high BMI and alcohol tissue exposures, there is evidence for such a proposed cancer progression model in the literature. As, high BMI exposure has been found to be associated with increased risk for breast and colorectal cancers and/or poorer survival (55,56), which originate in tissues more likely to be exposed to obesity-associated metabolites. Additionally, a recent study has identified a transcriptomic metabolic subtype in adjacent normal tissue to breast cancer that was also associated with poor breast cancer prognosis (57). Consequently, in our study the DNAm high BMI scores may be measuring the impact of high levels of obesity-associated metabolites related to changes in metabolic reprogramming, lipid levels and other processes, that could precondition the exposed normal bowel tissue and alter epigenetically controlled differentiation pathways (57,58). High alcohol exposure has also been found to be associated with increased risk and/or poorer

survival of head and neck, pancreatic, esophageal and colorectal cancers, that originate in tissues more likely to be exposed to high levels of alcohol metabolites (7,10,11,59,60). Additionally, somatic genetic alcohol mutational signatures have also been found to be associated with cancers likely to be exposed to high levels of alcohol metabolites, and to occur clonally in normal tissue before carcinogenesis (22,61,62). Therefore, high BMI and alcohol exposures could precondition bowel tissue for increased risk of CRC progression and reveal relevant DNA methylation biomarkers.

In our study, the significant differentially methylated positions and regions found between the HR and LR adenomas could be drivers of CRC progression. As many of our observed biomarkers are in plausible effector genes and previously associated with early bowel tissues, cancer progression and risk, in the literature. As for the DMPs, notably the *GUCY2C* gene, a hypermethylated study DMP, was previously found to be associated with obesity and gene signalling silencing in normal bowel epithelial cells, where clonal cancer driver alterations are expected to occur (63,64). Then, the *ANXA6* gene, a hypomethylated study DMP, was postulated to be a HR adenoma biomarker, in a study where patients' adenomas were classified as HR if had cancer associated copy-number alterations, and then found to express this protein in their stools with high sensitivity (65). While the *TPP1* gene, a hypermethylated study DMP, was previously found to be DNA altered and associated with telomere length dysfunction and increased CRC risk (66). Additionally, the *TMEM135* gene, a hypermethylated study DMP, was previously found to be near a SNP associated with increased breast cancer risk (53). Furthermore, the study-site replicated hypermethylated DMPs in the *DIAPH3* and *RGS3* genes, were previously found to be associated with EMT, which is important in cancer transformation, and with adenoma and cancer signalling pathways, such as the wingless-type (Wnt), and mitogen-activated protein kinase (MAPK) pathways, respectively (51,52).

Then, for the significant DMRs, notably, the *GPX7* gene, a study-site replicated hypermethylation DMR, was previously found to be associated with DNA damage repair and alterations in epithelial cells in normal bowel tissue compared to gastric cancer, again where clonal cancer driver alterations are expected to occur. While, the *MCIDAS* gene, a hypermethylated study DMR with a large effect size, was previously found to be hypermethylated in serrated CRC versus normal bowel tissue (67). Additionally, the *CCL28* gene, a hypermethylated study DMR, was previously found to be associated with immune response and transforming growth factor beta (TCFβ)

signalling pathway, and then with reduced expression in left CRC versus normal bowel tissue (68). Furthermore, the *PITX2* gene, a hypermethylated study DMR, was previously found to be associated with promoter hypermethylation in CRC versus normal bowel tissue and then reduced expression was also associated with poorer survival (69,70). These positional and regional differentially DNA methylated results were often methylated in a direction that matched their cancer associated literature, and taken in combination, strengthens the evidence that our observed alterations could be due to normal and/or adenoma bowel tissue CRC preconditioning due to high lifestyle bowel exposures, and causally associated with CRC progression. Therefore, these DNA methylation gene alterations could be possible HR adenoma CRC progression biomarkers and target genes suitable for interventions via lifestyle modification and preventative drug interventions.

There are several strengths to our study. Notably, EWAS differential methylation findings are commonly influenced by confounders, which are variables that if distributed unequally between the cases and controls, can bias the findings. Therefore, in this study model analyses were adjusted for patient demographic, adenoma and bowel confounders, by model inclusion of these collected variables. While other relevant latent confounders or PCs were then also included in the models to adjust for unaccounted confounders such as ethnicity, adenoma cell type composition, diagnosis route, and general FFPE block age. Selected findings were also corrected for genomic inflation, adjusted for multiple testing and stratified by study site, to reduce the occurrence of false positives and further increase validity. Additionally, the Infinium Methylation EPIC BeadChip technology has a relatively low error rate and has been shown to compare well with whole-genome bisulfite sequencing (WGBS), therefore our study findings do not need to be cross-validated by another molecular technique (71).

There are some limitations to our study. Firstly, the EWAS study numbers were relatively small and the EPIC methylation BeadChip has low genome coverage, including less than 3% of all CpG sites (72), which is just a fraction of the full methylome. The low coverage would impact the numbers of DMPs and DMRs we discovered but not the validity of their discovered effects. As, there are probably many other relevant DMPs and DMRs related to adenoma to CRC progression that remain to be discovered. Then, model analyses that include PCs, can lead to over adjustment and reduced measurement of the biological effect, especially with small sample sizes. However, in this study, model analyses with and without PCs were both run and presented, with more focus on the

latter. Additionally, in the stratified study-site analyses, bowel-side could not be adjusted for the John Radcliffe Hospital dataset. However, only findings that had been originally adjusted for bowel-side in the unstratified analyses were investigated for study-site replication. Furthermore, with this study design, causality can't be concluded, and therefore the concurrent cancer may be responsible for the HR adenoma DNA methylation alterations. However, this is unlikely, as the study adenomas were checked thoroughly for cancer by the study pathologists, and any patients' HR bowels with concurrent MSI+ CRC recorded in their hospital pathology records, were excluded, in-case they created a HR bowel environment. It is also expected that any concurrent cancers not previously selected and recorded for MSI testing and status, would have a low or negligible proportion of MSI + typing.

Next, the DNAm lifestyle exposure signatures used in this study were originally identified in blood and may not translate well and be maintained in bowel tissue. However, one published study has found that tissue from alveoli has a similar epigenetic profile with blood-derived samples at CpG sites associated with smoking (28). Early bowel progression tissue would also be expected to be less molecularly altered than in later invasive CRC tissue. Additionally, it is likely that there may be unmeasured environmental and lifestyle exposures in the HR bowel environment that may have also contributed to causing the DNA methylation alterations, some of which could be measured in the future, with further developed lifestyle exposure signatures and assays. Furthermore, the pre-defined DNAm current smoking exposure signature used in this study, may have also shown conflicting results because it is not suited to bowel tissue and may need re-development. Also, due to restrictions in the sample collection where follow-up data was required for the LR bowel adenomas, there was an in-balance of FFPE block age between the HR (2010 to 2015 block age years) and LR bowel adenomas (2010 to 2013 block age years), that made precise FFPE block age adjustment difficult. However, it has been found that there was no significant deterioration of methylation concordance between sample duplicates in archived FFPE tissue blocks up to 10 years of age in one study (73) and generally in other methylation BeadChip array studies (71,74,75). Lastly, additional external validations were not able to be carried out, but would be desirable, when such data is available.

Conclusions

High levels of alcohol and BMI bowel exposures may precondition normal bowel tissue, and also adenoma bowel tissue, for increased risk of DNA methylation alterations that aid high-risk

adenoma growth and CRC progression. The effect of BMI is less clear in our study, but high BMI is probably quite important and has been suggested as a risk factor for CRC in many earlier studies. Our inability to show an obvious association between HR adenomas and high BMI bowel exposure may be more a function of our study power than a lack of actual effects, and/or a stronger effect earlier in CRC progression. Finally, tobacco use did not manifest as an exposure that could be associated with DNA methylation and increased risk of HR bowel adenomas in our study. Overall, our findings could potentially be used in the future to risk stratify normal bowel tissue and adenomas that are more likely to progress to CRC and may be useful biomarkers to aid CRC early detection, screening and prevention. It also brings to light a new study design that may potentially be useful in identifying further HR adenoma molecular biomarkers for CRC progression. However, future investigations are needed to determine if DNAm high BMI and high alcohol exposure bowel tissue levels and the HR adenomas DNA methylation alterations can be further associated with risk of CRC progression.

Funding

This work was supported by the Trevor Collins Foundation project grant, then the Medical Research Council and infrastructure support from Cancer Research UK and Ovarian Cancer Action.

Acknowledgments

The authors would like to thank the study participants and study/healthcare staff who have contributed to the study cohorts. As well as, members of the Molecular Epidemiology and Evolution Cancer Laboratories, Barts and the London School of Medicine and Dentistry, Queen Mary and the Medical Genomics Group, Cancer Institute, University College London, University of London for their advice and support during the study. The study dataset was obtained from our larger CRC-adenoma-normal (CAN) tissue and clinical data collection, and collaborations are welcome.

Figures and Tables

Figure 1

Abbreviations: body mass index (BMI), colorectal cancer-adenoma-normal (CAN), differentially methylated position (DMP), differentially methylated region (DMR), DNAmethylation (DNAm), epigenome-wide association study (EWAS), gene expression omnibus (GEO), gene ontology (GO), high-risk (HR), Kyoto Encyclopaedia of Genes and Genomes (KEGG) and low-risk (LR)

Table 1

	Category	Details	HR bowel adenomas (n = 106)		LR bowel adenomas (n = 111)		P-value
			n	(%)	n	(%)	
Study site	Hospital	St Mark's John Radcliffe	67 39	(63) (37)	71 40	(64) (36)	0.908
Patient	Gender	Male	74	(70)	78	(70)	0.941
		Female	32	(30)	33	(30)	
Adenomas	Age at lesion diagnosis (years)	≤ 50	3	(3)	4	(4)	0.594
		51 to 60	9	(8)	12	(11)	
		61 to 70	36	(34)	41	(37)	
		71 to 80	42	(40)	39	(35)	
		≥ 81	16	(15)	15	(14)	
		Mean	71		71		
	Lesion-size (mm)	8 to 19	83	(78)	87	(78)	0.547
		20 to 29	18	(17)	18	(16)	
		30 to 59	5	(5)	6	(5)	
		Mean	15		15		
	Architecture	TA	48	(45)	36	(32)	0.151
		TVA	54	(51)	70	(63)	
		VA	4	(4)	5	(5)	
	Dysplasia	LGD	100	(94)	96	(86)	0.050
		HGD	6	(6)	15	(14)	
	Bowel-side	Left	61	(58)	92	(83)	< 0.001
		Right	45	(42)	19	(17)	

Abbreviations: high-grade dysplasia (HGD), high-risk (HR), low-grade dysplasia (LGD), low-risk (LR), millimetre (mm), number (n), percentage (%), tubular adenoma (TA), tubulovillous adenoma (TVA) and villous adenoma (VA)

Table 2

Datasets	Total (n HR, n LR)	DNAm exposures	ORperSD	95% CI	P-value
HR and LR bowel adenomas EWAS datasets					
All (EPIC 850K)	217 (106, 111)	Alcohol	2.16	1.55 - 3.09	9.7e-06
			2.12 ¹	1.51 - 3.04	2.7e-05
		BMI	0.70	0.51 - 0.95	2.4e-02
			0.72 ¹	0.52 - 0.99	4.6e-02
		Smoking	0.86	0.64 - 1.15	3.2e-01
			0.88 ¹	0.65 - 1.19	4.2e-01
SM	138 (67, 71)	Alcohol	1.85	1.24 - 2.84	3.5e-03
			1.85 ¹	1.21 - 2.89	5.3e-03
		BMI	0.84	0.57 - 1.24	3.9e-01
			0.87 ¹	0.58 - 1.30	4.9e-01
		Smoking	0.89	0.61 - 1.28	5.3e-01
			0.92 ¹	0.63 - 1.35	6.8e-01
JR	79 (39, 40)	Alcohol	3.44	1.91 - 7.00	1.6e-04
			5.24 ¹	2.57 - 13.01	5.1e-05
		BMI	0.36	0.17 - 0.66	3.0e-03
			0.37 ¹	0.17 - 0.70	5.4e-03
		Smoking	0.78	0.47 - 1.26	9.0e-01
			0.78 ¹	0.46 - 1.29	3.4e-01
GEO HR and LR normal bowel tissue EWAS dataset					
GLP22145 (EPIC 850K)	109 (52, 57)	Alcohol	2.12	1.35 - 3.55	2.0e-03
			1.37 ¹	0.78 - 2.50	2.9e-01
		BMI	1.72	1.13 - 2.73	1.5e-02
			2.73 ¹	1.52 - 5.39	1.7e-03
		Smoking	0.59	0.38 - 0.90	1.7e-02
			0.79 ¹	0.45 - 1.37	4.0e-01
GPL13534 (450K)	97 (49, 48)	Alcohol	1.79	1.14 - 2.96	1.7e-02
			0.80 ¹	0.39 - 1.61	5.4e-01
		BMI	1.11	0.85 - 1.44	4.5e-01
			0.70 ¹	0.46 - 1.02	7.7e-02
		Smoking	0.28	0.14 - 0.50	8.1e-05
			0.49 ¹	0.20 - 1.08	9.2e-02

Abbreviations: body mass index (BMI), confidence interval (CI), DNAmethylation (DNAm), epigenome-wide association study (EWAS), exponential notation (e), gene expression omnibus (GEO), high-risk (HR), John Radcliffe Hospital (JR), low-risk (LR), number (n), odds ratio (OR), standard deviation (SD) and St Mark's Hospital (SM)

¹ Principal components 1 and 2 adjustment

Table 3

Details					DMPs (n = 217)			SM DMPs (n = 138)			JR DMPs (n = 79)		
CpG site	Chr	Gene	Gn reg	CpG reg	db	p-value	BH FDR	db	p-value	BH FDR	db	p-value	BH FDR
cg06962647	13	DIAPH3	body	opsea	- 6.7	<u>2.22e-09</u>	1.6e-03	-5.9	6.31e-06	6.84e-02	- 7.4	7.36e-02	9.53e-02
cg22240312	21	/	/	opsea	- 5.6	<u>1.32e-08</u>	4.7e-03	-5.7	1.30e-06	3.31e-02	- 5.6	5.98e-06	1.13e-01
cg07212584	9	RGS3	5'UTR	opsea	- 6.7	<u>1.99e-08</u>	4.7e-03	-7.0	8.30e-07	2.82e-02	- 5.0	7.81e-04	3.77e-01
cg17332384	3	/	/	opsea	- 7.9	<u>3.82e-08</u>	4.8e-03	-9.0	1.18e-07	1.41e-02	- 5.0	2.73e-03	4.75e-01
cg15905657	11	TMEM135	body	opsea	- 5.9	<u>4.57e-08</u>	4.8e-03	-6.5	3.40e-07	1.53e-02	- 4.4	7.81e-04	3.77e-01
cg17135493	15	SRP14-AS1	body	opsea	- 7.0	8.62e-08	7.3e-03	-7.6	3.31e-06	3.31e-06	- 4.8	4.89e-04	3.33e-01
cg26609976	12	ANO6	body	opsea	- 6.0	9.58e-08	7.3e-03	-6.9	2.76e-07	1.51e-02	- 3.8	4.48e-03	5.20e-01
cg23556886	11	TPP1	3'UTR	shore	- 5.6	2.68e-07	1.4e-02	-5.4	2.01e-05	1.16e-01	- 4.9	7.42e-04	3.72e-01
cg26987316	1	/	/	opsea	- 6.5	5.48e-07	1.7e-02	-5.9	1.21e-04	1.77e-01	- 7.4	1.06e-05	1.18e-01
cg07998983	7	CHCHD3	body	opsea	- 5.4	1.17e-06	2.1e-02	-5.7	1.93e-05	1.16e-01	- 4.5	1.32e-03	4.20e-01
cg08770311	17	TTC19	body	opsea	6.9	1.31e-06	2.3e-02	6.3	2.16e-04	2.19e-01	7.9	7.10e-05	2.07e-01
cg00785312	12	GUCY2C	body	opsea	-6.5	1.59e-06	2.6e-02	-5.9	4.59e-04	2.76e-01	- 6.8	3.53e-05	1.59e-01
cg11946449	11	SYT13	TSS1500	shore	-6.1	1.64e-06	2.6e-02	-5.8	5.30e-05	1.50e-01	- 6.1	3.91e-04	3.10e-01
cg10306664	5	ANXA6	TSS200	shore	7.4	1.86e-06	2.7e-02	7.7	3.37e-05	1.36e-01	6.8	1.54e-03	4.27e-01
cg07486252	4	ANK2	body	opsea	-8.1	1.86e-06	2.7e-02	-8.2	7.10e-06	7.34e-02	- 6.2	1.38e-02	6.23e-01
cg14242024	9	NACC2	5'UTR	island	-5.7	2.20e-06	3.0e-02	-7.2	9.54e-08	1.41e-02	- 1.9	2.27e-01	8.72e-01
cg05027594	4	TAPT1	body	shore	-6.2	2.89e-06	3.3e-02	-6.6	1.22e-05	9.09e-02	- 4.7	5.67e-03	5.41e-01
cg18068612	14	/	/	opsea	-5.6	2.90e-06	3.3e-02	-5.4	9.29e-05	1.74e-01	- 5.4	9.17e-04	3.90e-01
cg22264813	7	WDR60	body	shore	-5.1	3.20e-06	3.5e-02	-5.4	3.19e-05	1.35e-01	- 4.2	2.68e-03	4.72e-01
cg00063268	1	VWA5B1	5'UTR	opsea	-5.9	3.51e-06	3.6e-02	-6.0	2.19e-05	1.21e-01	- 5.0	5.63e-03	5.40e-01
cg02807786	12	/	/	shelf	-7.2	3.64e-06	3.7e-02	-7.4	7.60e-05	1.65e-01	- 5.9	1.64e-03	4.34e-01
cg08116386	5	TXNDC15	TSS200	shore	-6.3	4.32e-06	4.1e-02	-8.3	1.32e-07	1.41e-02	- 2.9	8.83e-02	7.90e-01
cg21042179	14	/	/	shelf	-6.3	4.83e-06	4.2e-02	-6.5	9.00e-05	1.74e-01	- 5.2	1.98e-03	4.51e-01
cg21562867	4	MAML3	body	opsea	-5.4	5.02e-06	4.3e-02	-5.5	1.47e-04	1.93e-01	- 4.7	1.29e-03	4.20e-01
cg01896926	17	GLD4	5'UTR	island	5.1	5.16e-06	4.3e-02	4.6	3.38e-04	2.53e-01	5.4	9.50e-04	3.94e-01
cg25558737	11	DDB1	body	opsea	-6.5	5.40e-06	4.4e-02	-7.4	7.91e-06	7.52e-02	- 3.2	1.09e-01	8.07e-01
cg06430870	10	WDR37	body	opsea	-6.1	5.75e-06	4.6e-02	-8.4	2.27e-07	1.43e-02	- 2.3	1.10e-01	8.08e-01
cg25069618	15	CEMIP	body	opsea	-5.3	5.82e-06	4.6e-02	-5.1	2.56e-04	2.28e-01	- 4.6	8.90e-04	3.88e-01
cg09740935	1	SLC26A9	body	opsea	-7.9	6.54e-06	4.9e-02	-8.6	4.37e-05	1.43e-01	- 5.4	1.25e-02	6.14e-01

Abbreviations: benjamini-hochberg false discovery rate q-values (BH FDR), chromosome (chr), cytosine-phosphate-guanine (CpG), CpG region (CpG reg), delta beta % (db), differentially methylated position (DMP), exponential notation (e), gene region (gn reg), John Radcliffe Hospital (JR), not annotated or known (/), open sea (opsea), St Mark's Hospital (SM), transcription start site (TSS) and untranslated region (UTR)

Table 4

Details		DMRs (n = 217)						SM DMRs (n = 138)				JR DMRs (n = 79)			
Chr	Gene	CpGs (n)	Length (bp)	Min FDR ¹	Stouffer p-value	Max db ²	Mean db ³	Length (bp)	Min FDR ¹	Stouffer p-value	Mean db ³	Length (bp)	Min FDR ¹	Stouffer p-value	Mean db ³
1	GPX7	13	1469	3.71e-31	4.26e-12	13.1	6.2	1469	6.05e-19	8.37e-06	6.3	1469	9.45e-15	3.46e-03	15.5
8	XKR6	16	1468	4.81e-20	4.61e-10	11.2	8.0	1020	4.42e-08	4.42e-03	7.3	1468	1.53e-26	5.33e-07	14.7
7	TRIL	7	2664	7.85e-19	1.36e-09	8.3	5.1	2440	9.08e-12	6.28e-05	5.6	/	/	/	/
5	MCIDAS	7	669	3.73e-16	1.83e-09	13.3	10.4	/	/	/	/	669	1.06e-15	6.30e-05	17.0
11	LUZP2	12	577	9.70e-17	3.31e-09	11.5	8.4	/	/	/	/	/	/	/	/
4	UCP1	13	1464	5.82e-11	2.35e-07	9.6	5.2	1464	1.88e-05	2.18e-02	4.6	/	/	/	/
2	LBX2	13	1511	2.78e-15	8.40e-07	12.4	7.6	843	9.79e-08	2.67e-02	8.0	843	1.15e-12	5.10e-03	13.3
10	FAM24B	14	1582	1.87e-11	1.60e-05	-12.6	-7.5	504	5.62e-05	5.04e-02	-7.6	/	/	/	/
4	PITX2	8	1518	2.54e-07	3.38e-05	10.8	8.0	/	/	/	/	/	/	/	/
5	PRDM6	12	1992	3.08e-07	7.17e-05	8.6	5.5	698	1.06e-04	3.35e-02	6.1	/	/	/	/

Abbreviations: base-pair (bp), chromosome (chr), cytosine-phosphate-guanine (CpG), delta beta percentage (db), differentially methylated region (DMR), exponential notation (e), family with sequence similarity 24 member B (FAM24B), false discovery rate q-value (FDR), glutathione peroxidase 7 (GPX7), John Radcliffe Hospital (JR), ladybird homeobox 2 (LBX2), leucine zipper protein 2 (LUZP2), maximum (max), minimum (min), multiciliate differentiation and DNA synthesis-associated cell cycle (MCIDAS), number (n), paired like homeodomain 2 (PITX2), PR/SET domain 6 (PRDM6), TLR4 interactor with leucine rich repeats (TRIL), St Mark's Hospital (SM), uncoupling protein 1 (UCP1) and XK related 6 (XKR6)

¹ Minimum FDR of the smoothed estimate

² Maximum delta beta is the maximum delta beta within DMR

³ Mean delta beta is the mean delta beta across DMR

Figure and Table legends

Figure 1 Flow diagram of datasets and analyses. The flowchart outlines the study and validation datasets analysed, and includes sample numbers, and types of analyses.

Table 1 High and low-risk bowel adenoma epigenome-wide association study characteristics.

Clinical, histopathological and demographic characteristics for the patients and high-risk (HR) and low-risk (LR) bowel adenomas are shown. The student's t-test and chi-square test was used to test for differences between continuous data and categorical data, respectively. Significant p-values < 0.05 are highlighted in bold.

Table 2 High and low-risk bowel tissue and DNAmethylation lifestyle exposures.

DNAmethylation (DNAm) alcohol, body mass index (BMI) and smoking exposure associated odds ratios (ORs) for the risk of high-risk (HR) bowel adenomas, stratified study-site HR bowel adenomas and external HR normal bowel tissue, adjusted for confounders and without/with principal components 1 and 2 are shown. The normal bowel tissue is left bowel-sided and obtained from the gene expression omnibus (GEO) public database. Significant associations are highlighted in bold.

Table 3 Differentially methylated positions. High-risk (HR) compared to low-risk (LR) bowel adenoma differentially methylated positions (DMPs) adjusted for confounders and genomic inflation corrected, with delta betas values $\geq 5\%$ and Benjamini-Hochberg false discovery rate (BH FDR) q-values < 0.05, are shown. These were ordered by decreasing p-value significance, and then matched to their corresponding stratified St Mark's Hospital and John Radcliffe Hospital study-site DMP values. The Bonferroni significant p-values < 7×10^{-8} after genomic control are underlined and significant replicated study-site DMPs are highlighted in bold.

Table 4 Differentially methylated regions without principal component adjustment.

High-risk (HR) compared to low-risk (LR) bowel adenoma differentially methylated regions (DMRs) adjusted for confounders, with mean delta beta values $\geq 5\%$, Stouffer p-values < 0.05 and minimum of 7 cytosine-phosphate-guanines (CpGs) are shown. These were ordered by decreasing Stouffer p-values and then matched to their corresponding stratified St Mark's Hospital and John Radcliffe Hospital study-site DMR values. Significant replicated study-site DMRs are highlighted in bold.

References

1. Cancer registration statistics, England Statistical bulletins - Office for National Statistics [Internet]. [cited 2019 Jan 26]. Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/previousReleases>
2. Bowel cancer incidence statistics [Internet]. Cancer Research UK. 2015 [cited 2019 Jan 26]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence>
3. Cunningham D, Atkin W, Lenz H-J, Lynch HT, Minsky B, Nordlinger B, et al. Colorectal cancer. *Lancet*. 2010 Mar 20;375(9719):1030–47.
4. Snover D, Ahnen DJ, Burt RW. Serrated polyps of the colon and rectum and serrated (“hyperplastic”) polyposis. In: Bozman FT, Carneiro F, Hruban RH, et al., editors. WHO classification of tumours. Pathology and genetics. Tumours of the digestive system. Vol. 4. Berlin: Springer-Verlag; 2010.
5. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst*. 2004 Feb 18;96(4):261–8.
6. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D, et al. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ*. 2007 Dec 1;335(7630):1134.
7. Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V, et al. Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. *Br J Cancer*. 2015 Feb 3;112(3):580–93.
8. Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, et al. Meta-analyses of colorectal cancer risk factors. *Cancer causes & control: CCC*. 2013 Jun;24(6):1207–22.
9. Rumgay H, Shield K, Charvat H, Ferrari P, Sornpaisarn B, Obot I, et al. Global burden of cancer in 2020 attributable to alcohol consumption: a population-based study. *Lancet Oncol*. 2021 Jul 13;S1470-2045(21)00279-5.
10. Walter V, Jansen L, Ulrich A, Roth W, Bläker H, Chang-Claude J, et al. Alcohol consumption and survival of colorectal cancer patients: a population-based study from Germany. *Am J Clin Nutr*. 2016;103(6):1497–506.
11. Jayasekara H, English DR, Haydon A, Hodge AM, Lynch BM, Rosty C, et al. Associations of alcohol intake, smoking, physical activity and obesity with survival following colorectal cancer diagnosis by stage, anatomic site and tumor molecular subtype. *Int J Cancer*. 2018 15;142(2):238–50.

12. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *Lancet*. 2014 Aug 30;384(9945):755–65.
13. Wheeler PR, Burkitt HG. Wheeler's basic histopathology. Churchill Livingstone; 1996.
14. Morson BC. The evolution of colorectal carcinoma. *Clin Radiol*. 1984 Nov;35(6):425–31.
15. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990 Jun 1;61(5):759–67.
16. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. *N Engl J Med*. 1988 Sep 1;319(9):525–32.
17. Gerstung M, Jolly C, Leshchiner I, D'entrop SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. *Nature*. 2020;578(7793):122–8.
18. Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, et al. A Big Bang model of human colorectal tumor growth. *Nat Genet*. 2015 Mar;47(3):209–16.
19. Kang H, Salomon MP, Sottoriva A, Zhao J, Toy M, Press MF, et al. Many private mutations originate from the first few divisions of a human colorectal adenoma. *J Pathol*. 2015 Nov;237(3):355–62.
20. Yizhak K, Aguet F, Kim J, Hess JM, Kübler K, Grimsby J, et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science*. 2019 Jun 7;364(6444).
21. Lee-Six H, Olafsson S, Ellis P, Osborne RJ, Sanders MA, Moore L, et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature*. 2019;574(7779):532–7.
22. Wei R, Li P, He F, Wei G, Zhou Z, Su Z, et al. Comprehensive analysis reveals distinct mutational signature and its mechanistic insights of alcohol consumption in human cancers. *Brief Bioinformatics*. 2020 Jun 1;
23. Riva L, Pandiri AR, Li YR, Droop A, Hewinson J, Quail MA, et al. The mutational signature profile of known and suspected human carcinogens in mice. *Nature Genetics*. 2020 Sep 28;1–9.
24. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature*. 2017 05;541(7635):81–6.
25. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet*. 2016 Oct;9(5):436–47.
26. Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai P-C, Reynolds LM, et al. A DNA methylation biomarker of alcohol consumption. *Mol Psychiatry*. 2018;23(2):422–33.
27. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition

- (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum Mol Genet.* 2013 Mar 1;22(5):843–51.
28. Stueve TR, Li W-Q, Shi J, Marconett CN, Zhang T, Yang C, et al. Epigenome-wide analysis of DNA methylation in lung tissue shows concordance with blood studies and identifies tobacco smoke-inducible enhancers. *Hum Mol Genet.* 2017 01;26(15):3014–27.
 29. Teschendorff AE, Yang Z, Wong A, Pipinikas CP, Jiao Y, Jones A, et al. Correlation of Smoking-Associated DNA Methylation Changes in Buccal Cells With DNA Methylation Changes in Epithelial Cancer. *JAMA Oncol.* 2015 Jul;1(4):476–85.
 30. Bormann F, Rodríguez-Paredes M, Lasitschka F, Edelmann D, Musch T, Benner A, et al. Cell-of-Origin DNA Methylation Signatures Are Maintained during Colorectal Carcinogenesis. *Cell Rep.* 2018 12;23(11):3407–18.
 31. Wang T, Maden SK, Luebeck GE, Li CI, Newcomb PA, Ulrich CM, et al. Dysfunctional epigenetic aging of the normal colon and colorectal cancer risk. *Clin Epigenetics.* 2020 Jan 3;12(1):5.
 32. Kleihues P, Sobin LH. World Health Organization classification of tumors. *Cancer.* 2000 Jun 15;88(12):2887.
 33. Fortin J-P, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics.* 2017 15;33(4):558–60.
 34. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics.* 2014 Feb 1;30(3):428–30.
 35. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics.* 2017 Dec 15;33(24):3982–4.
 36. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* 2017 Feb 28;45(4):e22–e22.
 37. Nordlund J, Bäcklin CL, Wahlberg P, Busche S, Berglund EC, Eloranta M-L, et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biology.* 2013 Sep 24;14:r105.
 38. van Iterson M, Cats D, Hop P, BIOS Consortium, Heijmans BT. omicsPrint: detection of data linkage errors in multiple omics studies. *Bioinformatics.* 2018 15;34(12):2142–3.
 39. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics.* 2013 Jan 15;29(2):189–96.
 40. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007 Jan;8(1):118–27.

41. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS ONE*. 2009 Dec 18;4(12):e8274.
42. Wang J, Zhao Q, Hastie T, Owen AB. CONFOUNDER ADJUSTMENT IN MULTIPLE HYPOTHESIS TESTING. *Ann Stat*. 2017 Oct;45(5):1863–94.
43. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:Article3.
44. van Iterson M, van Zwet EW, BIOS Consortium, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol*. 2017 27;18(1):19.
45. Haynes W. Bonferroni Correction. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H, editors. *Encyclopedia of Systems Biology* [Internet]. New York, NY: Springer; 2013 [cited 2019 Dec 30]. p. 154–154. Available from: https://doi.org/10.1007/978-1-4419-9863-7_1213
46. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990 Jul;9(7):811–8.
47. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin*. 2015;8:6.
48. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*. 2016 Jan 15;32(2):286–8.
49. Xu K, Montalvo-Ortiz JL, Zhang X, Southwick SM, Krystal JH, Pietrzak RH, et al. Epigenome-Wide DNA Methylation Association Analysis Identified Novel Loci in Peripheral Cells for Alcohol Consumption Among European American Male Veterans. *Alcohol Clin Exp Res*. 2019 Oct;43(10):2111–21.
50. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, et al. Epigenetic prediction of complex traits and death. *Genome Biology*. 2018 Sep 27;19(1):136.
51. Dong L, Li Z, Xue L, Li G, Zhang C, Cai Z, et al. DIAPH3 promoted the growth, migration and metastasis of hepatocellular carcinoma cells by activating beta-catenin/TCF signaling. *Mol Cell Biochem*. 2018 Jan;438(1–2):183–90.
52. Wang J, Zhou Y, Fei X, Chen X, Zhu Z. Regulator of G-protein signaling 3 targeted by miR-126 correlates with poor prognosis in gastric cancer patients. *Anticancer Drugs*. 2017;28(2):161–9.
53. Natrajan R, Mackay A, Lambros MB, Weigelt B, Wilkerson PM, Manie E, et al. A whole-genome massively parallel sequencing analysis of BRCA1 mutant oestrogen receptor-negative and -positive breast cancers. *J Pathol*. 2012 May;227(1):29–41.

54. Chen Z, Hu T, Zhu S, Mukaisho K, El-Rifai W, Peng D-F. Glutathione peroxidase 7 suppresses cancer cell growth and is hypermethylated in gastric cancer. *Oncotarget*. 2017 Aug 15;8(33):54345–56.
55. Moghaddam AA, Woodward M, Huxley R. Obesity and risk of colorectal cancer: a meta-analysis of 31 studies with 70,000 events. *Cancer Epidemiol Biomarkers Prev*. 2007 Dec;16(12):2533–47.
56. Taghizadeh N, Boezen HM, Schouten JP, Schröder CP, de Vries EGE, Vonk JM. BMI and Lifetime Changes in BMI and Cancer Mortality Risk. *PLoS One* [Internet]. 2015 Apr 16 [cited 2020 Oct 28];10(4). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4399977/>
57. Gadaleta E, Fourgoux P, Pirrò S, Thorn GJ, Nelan R, Ironside A, et al. Characterization of four subtypes in morphologically normal tissue excised proximal and distal to breast cancer. *NPJ Breast Cancer*. 2020;6:38.
58. Johnson KE, Siewert KM, Klarin D, Damrauer SM, VA Million Veteran Program, Chang K-M, et al. The relationship between circulating lipids and breast cancer risk: A Mendelian randomization study. *PLoS Med*. 2020;17(9):e1003302.
59. Ma Q, Liu W, Jia R, Long H, Zhang L, Lin P, et al. Alcohol and survival in ESCC: prediagnosis alcohol consumption and postoperative survival in lymph node-negative esophageal carcinoma patients. *Oncotarget*. 2016 Jun 21;7(25):38857–63.
60. Chen Y-P, Zhao B-C, Chen C, Lei X-X, Shen L-J, Chen G, et al. Alcohol drinking as an unfavorable prognostic factor for male patients with nasopharyngeal carcinoma. *Sci Rep*. 2016 Jan 18;6:19290.
61. Li XC, Wang MY, Yang M, Dai HJ, Zhang BF, Wang W, et al. A mutational signature associated with alcohol consumption and prognostically significantly mutated driver genes in esophageal squamous cell carcinoma. *Ann Oncol*. 2018 01;29(4):938–44.
62. Chang J, Tan W, Ling Z, Xi R, Shao M, Chen M, et al. Genomic analysis of oesophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and genomic alterations. *Nat Commun*. 2017 26;8:15290.
63. Lin JE, Colon-Gonzalez F, Blomain E, Kim GW, Aing A, Stoecker B, et al. Obesity-Induced Colorectal Cancer Is Driven by Caloric Silencing of the Guanylin-GUCY2C Paracrine Signaling Axis. *Cancer Res*. 2016 Jan 15;76(2):339–46.
64. Yarla NS, Gali H, Pathuri G, Smriti S, Farooqui M, Panneerselvam J, et al. Targeting the paracrine hormone-dependent guanylate cyclase/cGMP/phosphodiesterases signaling pathway for colorectal cancer prevention. *Semin Cancer Biol*. 2019;56:168–74.
65. Komor MA, Bosch LJ, Coupé VM, Rausch C, Pham TV, Piersma SR, et al. Proteins in stool as biomarkers for non-invasive detection of colorectal adenomas with high risk of progression. *J Pathol*. 2019 Nov 29;

66. Li J, Chang J, Tian J, Ke J, Zhu Y, Yang Y, et al. A Rare Variant P507L in TPP1 Interrupts TPP1-TIN2 Interaction, Influences Telomere Length, and Confers Colorectal Cancer Risk in Chinese Population. *Cancer Epidemiol Biomarkers Prev.* 2018;27(9):1029–35.
67. Andrew AS, Baron JA, Butterly LF, Suriawinata AA, Tsongalis GJ, Robinson CM, et al. Hyper-Methylated Loci Persisting from Sessile Serrated Polyps to Serrated Cancers. *Int J Mol Sci.* 2017 Mar 2;18(3).
68. Dimberg J, Hugander A, Wågsäter D. Protein expression of the chemokine, CCL28, in human colorectal cancer. *Int J Oncol.* 2006 Feb;28(2):315–9.
69. Semaan A, Uhl B, Branchi V, Lingohr P, Bootz F, Kristiansen G, et al. Significance of PITX2 Promoter Methylation in Colorectal Carcinoma Prognosis. *Clin Colorectal Cancer.* 2018;17(2):e385–93.
70. Hirose H, Ishii H, Mimori K, Tanaka F, Takemasa I, Mizushima T, et al. The significance of PITX2 overexpression in human colorectal cancer. *Ann Surg Oncol.* 2011 Oct;18(10):3005–12.
71. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* 2016 07;17(1):208.
72. Chatterjee A, Rodger EJ, Morison IM, Eccles MR, Stockwell PA. Tools and Strategies for Analysis of Genome-Wide and Gene-Specific DNA Methylation Patterns. *Methods Mol Biol.* 2017;1537:249–77.
73. de Ruijter TC, de Hoon JPJ, Slaats J, de Vries B, Janssen MJFW, van Wezel T, et al. Formalin-fixed, paraffin-embedded (FFPE) tissue epigenomics using Infinium HumanMethylation450 BeadChip assays. *Lab Invest.* 2015 Jul;95(7):833–42.
74. Thirlwell C, Eymard M, Feber A, Teschendorff A, Pearce K, Lechner M, et al. Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. *Methods.* 2010 Nov 1;52(3):248–54.
75. Siegel EM, Berglund AE, Riggs BM, Eschrich SA, Putney RM, Ajidahun AO, et al. Expanding epigenomics to archived FFPE tissues: An evaluation of DNA repair methodologies. *Cancer Epidemiol Biomarkers Prev.* 2014 Dec;23(12):2622–31.
76. Hesson LB, Ng B, Zarzour P, Srivastava S, Kwok C-T, Packham D, et al. Integrated Genetic, Epigenetic, and Transcriptional Profiling Identifies Molecular Pathways in the Development of Laterally Spreading Tumors. *Mol Cancer Res.* 2016;14(12):1217–28.
77. Liao C, Huang X, Gong Y, Lin Q. Discovery of core genes in colorectal cancer by weighted gene co-expression network analysis. *Oncol Lett.* 2019 Sep;18(3):3137–49.
78. Wang X, Zhang S, Zhang J, Lam E, Liu X, Sun J, et al. Annexin A6 is down-regulated through promoter methylation in gastric cancer. *Am J Transl Res.* 2013;5(5):555–62.
79. Grewal T, Enrich C, Rentero C, Buechler C. Annexins in Adipose Tissue: Novel Players in Obesity. *Int J Mol Sci.* 2019 Jul 13;20(14).

80. Fink SP, Myeroff LL, Kariv R, Platzer P, Xin B, Mikkola D, et al. Induction of KIAA1199/CEMIP is associated with colon cancer phenotype and poor patient survival. *Oncotarget*. 2015 Oct 13;6(31):30500–15.
81. Kim GW, Lin JE, Waldman SA. GUCY2C: at the intersection of obesity and cancer. *Trends Endocrinol Metab*. 2013 Apr;24(4):165–73.
82. Bauerschlag DO, Ammerpohl O, Bräutigam K, Schem C, Lin Q, Weigel MT, et al. Progression-free survival in ovarian cancer is reflected in epigenetic DNA methylation profiles. *Oncology*. 2011;80(1–2):12–20.
83. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet*. 2019 Feb;51(2):237–44.
84. Yang XL, Liu KY, Lin FJ, Shi HM, Ou ZL. CCL28 promotes breast cancer growth and metastasis through MAPK-mediated cellular anti-apoptosis and pro-metastasis. *Oncol Rep*. 2017 Sep;38(3):1393–401.
85. Uhl B, Gevensleben H, Tolkach Y, Sailer V, Majores M, Jung M, et al. PITX2 DNA Methylation as Biomarker for Individualized Risk Assessment of Prostate Cancer in Core Biopsies. *J Mol Diagn*. 2017;19(1):107–14.
86. Alnabulsi A, Cash B, Hu Y, Silina L, Alnabulsi A, Murray GI. The expression of brown fat-associated proteins in colorectal cancer and the relationship of uncoupling protein 1 with prognosis. *Int J Cancer*. 2019 15;145(4):1138–47.
87. Half E, Bercovich D, Rozen P. Familial adenomatous polyposis. *Orphanet J Rare Dis*. 2009 Oct 12;4:22.
88. Ulrich CM, Gigic B, Böhm J, Ose J, Viskochil R, Schneider M, et al. The ColoCare Study: A Paradigm of Transdisciplinary Science in Colorectal Cancer Outcomes. *Cancer Epidemiol Biomarkers Prev*. 2019 Mar;28(3):591–601.
89. Luo Y, Wong C-J, Kaz AM, Dzieciatkowski S, Carter KT, Morris SM, et al. Differences in DNA methylation signatures reveal multiple pathways of progression from adenoma to colorectal cancer. *Gastroenterology*. 2014 Aug;147(2):418-429.e8.