World Journal of Gastrointestinal Oncology

Online Submissions: http://www.wjgnet.com/esps/
wjgo@wjgnet.com
doi:10.4251/wjgo.v4.i8.193

World J Gastrointest Oncol 2012 August 15; 4(8): 193-201
ISSN 1948-5204 (online)

ORIGINAL ARTICLE

Leukocyte DNA methylation and colorectal cancer among male smokers

Ying Gao, Keith Killian, Hong Zhang, Kai Yu, Qi-Zhai Li, Stephanie Weinstein, Jarmo Virtamo, Margaret Tucker, Philip Taylor, Demetrius Albanes, Paul Meltzer, Neil Caporaso

Ying Gao, Kai Yu, Stephanie Weinstein, Margaret Tucker, Philip Taylor, Demetrius Albanes, Neil Caporaso, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20852, United States
Keith Killian, Paul Meltzer, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20852, United States
Hong Zhang, Institute of Biostatistics, School of Life Science, Fudan University, Shanghai 200433, China
Qi-Zhai Li, Department of Statistics, Chinese Academy of Sciences, Beijing 100190, China
Jarmo Virtamo, National Public Health Institute, Helsinki FI-00271, Finland
Author contributions: Gao Y and Caporaso N designed the study; Gao Y, Zhang H, Yu K and Li QZ analyzed the data; Gao Y and Caporaso N wrote the report; Killian K and Meltzer P conducted laboratory assays; all authors read, gave comments, and approved the final version of the manuscript; Gao Y had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Supported by Grants from the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health
Correspondence to: Ying Gao, MD, PhD, Professor, Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Boulevard, Bldg. EPS/Room 7110, NIH/NCI, Bethesda, MD 20892, United States. gaoying@mail.nih.gov
Telephone: +1-301-4969249 Fax: +1-301-4024489
Received: January 13, 2012 Revised: July 13, 2012
Accepted: July 20, 2012
Published online: August 15, 2012

Abstract

AIM: To explore the association between methylation in leukocyte DNA and colorectal cancer (CRC) risk in male smokers using the α-tocopherol, β-carotene cancer prevention study.

METHODS: About 221 incident CRC cases, and 219 controls, frequency-matched on age and smoking intensity were included. DNA methylation of 1505 CpG sites selected from 807 genes were evaluated using Illumina GoldenGate Methylation Cancer Panel I in prediagnostic blood leukocytes of study subjects. Tertiles of methylation level classified according to the distribution in controls for each CpG site were used to analyze the association between methylation level and CRC risk with logistic regression. The time between blood draw to cancer diagnosis (classifying cases according to latency) was incorporated in further analyses using proportional odds regression.

RESULTS: We found that methylation changes of 31 CpG sites were associated with CRC risk at $P < 0.01$ level. Though none of these 31 sites remained statistically significant after Bonferroni correction, the most statistically significant CpG site associated with CRC risk achieved a P value of 1.0×10^{-4}. The CpG site is located in *DSP* gene, and the risk estimate was 1.52 (95% CI: 0.91-2.53) and 2.62 (95% CI: 1.65-4.17) for the second and third tertile comparing with the lowest tertile respectively. Taking the latency information into account strengthened some associations, suggesting that the methylation levels of corresponding sites might change over time with tumor progression.

CONCLUSION: The results suggest that the methylation level of some genes were associated with cancer susceptibility and some were related to tumor development over time. Further studies are warranted to confirm and refine our results.

Key words: DNA methylation; Colorectal cancer; Susceptibility

Peer reviewer: Haruhiko Sugimura, MD, PhD, Department of Investigative Pathology I, Hamamatsu University School of Med-

icine, 1-20-1 Handayama, Higashi-ward, Hamamatsu 431-3192, Japan

Gao Y, Killian K, Zhang H, Yu K, Li QZ, Weinstein S, Virtamo J, Tucker M, Taylor P, Albanes D, Meltzer P, Caporaso N. Leukocyte DNA methylation and colorectal cancer among male smokers. *World J Gastrointest Oncol* 2012; 4(8): 193-201 Available from: URL: http://www.wjgnet.com/1948-5204/full/v4/i8/193.htm DOI: http://dx.doi.org/10.4251/wjgo.v4.i8.193

INTRODUCTION

Although its incidence and mortality have declined in the past decades, colorectal cancer (CRC) remains the third most common cancer worldwide[1]. Epigenetic changes, which affect gene expression and subsequent phenotypes by mechanisms other than DNA sequence change, have been shown to play important roles in etiology of various cancers[2]. Studying the roles of epigenetic and genetic factors, as well as their interaction with environmental exposures, should lead to better understanding on the pathogenesis of CRC, and ultimately contribute to the prevention and treatment.

DNA methylation, the covalent addition of a methyl group, is an epigenetic event that alters gene expression. It has a degree of dependence on host status, including age[3,4] and genetic background[5-7], as well as environmental exposures, including dietary availability of methyl groups[8], smoking[9], and other factors. Variation in DNA methylation helps explain genetic diversity, which might mediate gene-environment interaction; mechanistically it might represent a key mechanism whereby changing environment modulates gene action. Alterations of the normal DNA methylation pattern, which is unique for each gene, have been considered an important step in many human diseases and been associated with all cancers examined to date, including CRC[2]. Promoter hypermethylation of many genes, associated with silencing (tumor suppressor genes) or activation (oncogenes) of downstream genes, has been linked to CRC carcinogenesis[10-14]. The key pathways in CRC carcinogenesis, including loss of cell cycle regulation control ($p^{16INK4a}$), silencing of DNA mismatch repair genes (*MLH1, MGMT*), loss of function of apoptosis genes (*DAPK*), and abolishment of carcinogen metabolism (*GSTP1*), involve changes in promoter methylation[15]. Accumulating evidence suggests that epigenetic, including DNA methylation, abnormalities are a driving force of carcinogenesis[16]. Notably, a recent study induced cancer in Apc Min/+ mice through over expressing the *de novo* DNA methyltransferases *Dnmt3a1* and *Dnmt3b1* and thereby established a direct causal connection between DNA *hypermethylation* and the development of colon tumors[17]. However, evidence for the role that these mechanisms play in human populations remains incomplete. Case control studies are unable to differentiate methylation changes associated with tumor susceptibility from those changes that derive solely from tumor progression. For this reason, prospective studies are needed to examine whether certain DNA methylation changes precede cancer development.

Notable DNA methylation pattern changes have been observed in CRC tissue[18-22], which could classify CRC molecular subtypes. For example, the CpG island methylator phenotype (panels of four to eight CpG islands)[23-26] has distinct pathogenic features that could have prognostic implications and suggest avenues for prevention or therapy. However to date, prospective data on DNA methylation and risk of incident CRC have not been reported, and data from non-target tissue, including leukocytes, are sparse. Study in mice have shown that changes of methylation in leukocyte DNA were parallel those observed in other somatic tissues[27]. Emerging data suggest that leukocyte DNA methylation might be linked to susceptibility of bladder cancer[28], lung cancer[29], cervical intraepithelial neoplasia[30]. The study of DNA methylation status in peripheral leukocytes in relation to CRC risk has mainly been conducted among hereditary non-polyposis CRC patients to compare methylation status between tissues[31-35]. Limited epidemiological studies have only examined small sets of genes[36,37] in case-control settings. To date, no report has examined the association between DNA methylation and CRC risk in a prospective study setting.

We conducted a nested case-control study within the α-tocopherol, β-carotene (ATBC) cancer prevention study to explore the relationship between CRC risk and gene-specific DNA methylation from pre-diagnostic leukocytes.

MATERIALS AND METHODS

ATBC study

The ATBC cancer prevention study was a double-blinded, placebo-controlled, 2 × 2 factorial design trial, which assessed the effect of ATBC, or both, on the incidence of cancer in male smokers[38-40]. The prospective cohort comprised of 29 133 Finish men, aged 50-69 years at study entry (1985-1988), who smoked at least 5 cigarettes per day. Participants were randomly assigned to groups receiving ATBC, both supplements, or placebo for 5-8 years until April 30, 1993 or death. Incident cases were identified through the Finnish Cancer Registry. Medical records for each cancer case were reviewed centrally by study physicians or oncologists for diagnosis confirmation. Questionnaire based general risk factor, education, family history, medical history, detailed smoking, and dietary information, as well as anthropometry, were collected at baseline. Whole blood biospecimens were collected from subjects close to the end of the intervention (before 1993).

This trial was approved by the institutional review boards of the National Institute for Health and Welfare of Finland and the National Cancer Institute of USA. All participants provided written informed consent.

Study subjects

We conducted a nested case-control study within the

ATBC study. All participants were male smokers. All the incident cases with colon and rectal adenocarcinomas (International Classification of Diseases 9, codes 153 and 154) identified through the latest follow-up (April 2006) that had a pre-diagnostic blood sample were included in this study. Controls were cancer-free subjects identified through April 2006, who provided a blood sample, were frequency matched to the cases by age at study entry (± 5 years), date of blood draw (± 90 d), and smoking intensity (± 10 cigarettes/d). Leukocyte DNA was extracted from pre-diagnostic buffy coat sample of all study subjects.

DNA methylation assay

For the current study, 1 μg DNA from each study subject was treated with the bisulfate conversion kit from Zymo Research (D5008) (http://www.zymoresearch.com) to convert unemthylated cytosines to uracil and leave methylated ones intact. Bisulfite converted DNA was then genotyped using a commercially available chip, the Illumina GoldenGate Methylation Cancer Panel 1 (Appendix 3) (http://www.illumina.com/pages.ilmn?ID=193), to differentiate the methylated and unmethylated cytosine. The chip covers 1505 CpG sites, which are located in the promoter regions of 807 genes reflecting a broad spectrum of carcinogenic processes including tumor suppressor genes, oncogenes, genes involved in DNA repair, cell cycle control, differentiation, apoptosis, X-linked, and imprinted genes. All of the 454 samples (224 cases and 230 controls) were tested in five 96 well-plate chips/ batches. Matched case and control samples were placed within the same batch in random order. Duplicates from each of 2 QC subjects in each of 5 batches were randomly and blindly dispersed among study samples in order to assess variability of the assay.

Statistical analysis

Raw data for methylated (M) alleles and unmethylated (U) alleles of each CpG site were exported from GenomeStudio Data Analysis Software (Illumina). Ten samples were excluded because 5% or more of the markers exceeded the threshold for non-specific cross-hybridization (*P* value of detection > 0.1). Then quantile normalization[41] was conducted on both M and U, and β values for each CpG site were generated for statistical analysis. The β value was expressed as percentage of methylated cytosines as follows: β = min (M, 0)/[min (U, 0) + min (M, 0) + 100], and was used as the proxy of methylation level at each CpG site. Principal component analysis was conducted based on overall β values of each individual to detect outlier individuals: the first two (principal) components, which explained most of the variance between individuals, were selected. Four samples which were outside the range of ± 5 SD of each principal component were excluded. A total of 440 subjects were eligible for the final analysis.

Methylation levels for each CpG site were classified into tertiles according to the distribution among controls.

Table 1 Characteristics of cases and controls *n* (%)

	Controls (*n* = 219)	Cases (*n* = 221)	*P*[1]
Age of randomization (mean) (yr)	58 (54-62)	58 (54-61)	0.7
ATBC			0.33
Placebo	62 (28)	48 (22)	
B-carotene	56 (26)	60 (27)	
VitE	49 (22)	62 (28)	
Both	52 (24)	51 (23)	
Family history of CRC	6 (3)	6 (3)	0.98
Smoking age (mean) (yr)	20 (17-22)	19 (17-20)	0.18
Smoking years (mean) (yr)	36 (30-42)	37 (31-42)	0.26
BMI (kg/m^2)			0.03
< 25	80 (36)	75 (34)	
25-29.9	113 (52)	99 (45)	
⩾ 30	26 (12)	47 (21)	

[1]*P* from *T*-test for continuous variables or χ^2 test for categorical variables. CRC: Colorectal cancer; ATBC: α-tocopherol, β-carotene; BMI: Body mass index.

Logistic regression was used to estimate the association between the risk for CRC and each CpG site. Adjustment for date of blood draw and smoking intensity did not modify the results substantially so we only included age as covariate among the matching factors. Batch differences of QC samples were non-negligible and adjustment for batch variable modified the results significantly. Therefore, we reported the results from logistic regression adjusted for age and batch factor.

Taking advantage of the prospective design of the current study, we explored the effect of latency, which was defined as the difference in days between the date of blood draw to reported date of cancer diagnosis (for CRC cases). The range of latency (ranging 0-13.4 years) allowed us to classify cases into three categories: short (0-4.8 years), medium (4.9-8.9 years), and long latency groups (9.0-13.4 years). Proportional odds regression was used to estimate the association between DNA methylation and CRC risk by coding cases according to latency. In addition, another set of logistic regression analyses (partial logistic regression) were conducted with cases restricted to those having short and medium latencies. If we assume methylation of some CpG sites changes over time in relation to tumor progression, this model should enhance the detection of these markers by excluding subjects with long latency.

To evaluate the potential false positive findings due to multiple testing, we adjusted the P values using a Bonferroni correction for the total number of all the CpG sites tested in the current analysis (1505 sites). All the analyses were conducted with SAS9.1 or R software and all the tests are two sided.

RESULTS

A total of 221 CRC cases and 219 frequency matched controls with successful DNA methylation measurements were included in the current analysis. Cases and controls

Table 2 Methylation of CpG sites associated with colorectal cancer development (*P* value < 0.01 before multiple testing correction)[1]

Marker name	2nd tertile *vs* 1st tertile			3rd tertile *vs* 1st tertile			*P*[2]
	OR	L95	U95	OR	L95	U95	
DSP	1.09876	0.95772	1.26057	1.26539	1.11734	1.43306	0.000599826
MLH3	1.04349	0.90638	1.20134	1.26606	1.0994	1.45798	0.000731318
FLT4	1.25952	1.10631	1.43394	1.14513	0.98989	1.32473	0.00096231
INSR	1.16236	1.01962	1.32509	1.24011	1.09317	1.40681	0.001490452
GLA	1.37049	1.15665	1.62388	1.44612	1.19452	1.75072	0.001636963
PTCH2	1.18022	1.03475	1.34615	1.26374	1.11417	1.43339	0.001637435
GSTM2	1.24321	1.09656	1.40947	1.17084	1.02674	1.33516	0.001994473
HCK	1.18573	1.04409	1.34659	1.19691	1.03199	1.38819	0.002287131
PLAGL1	0.91619	0.80363	1.04451	0.87645	0.75745	1.01414	0.002326078
PARP1	1.07016	0.95	1.20553	0.86663	0.75535	0.9943	0.003005978
COL18A1	1.23987	1.08901	1.41162	1.20014	1.05351	1.36719	0.003163099
KCNK4	1.19498	1.05273	1.35645	1.15065	1.00785	1.31368	0.003250103
GJB2	1.16605	1.02831	1.32224	1.02785	0.89877	1.17547	0.003317356
SEMA3F	1.16703	1.01753	1.33848	1.16456	1.01456	1.33673	0.00380004
HBII	0.89709	0.79121	1.01713	0.86969	0.7624	0.99207	0.004064919
SEMA3C	1.08773	0.96383	1.22756	0.87597	0.76919	0.99757	0.004352617
MME_	1.18426	1.04194	1.34601	1.17187	1.02959	1.3338	0.005258668
WNT1_	1.18575	1.02807	1.36761	1.07102	0.91035	1.26004	0.005648004
SGCE	1.0712	0.93671	1.22501	1.22832	1.08144	1.39516	0.00629186
PDE1B_	1.24257	1.09182	1.41413	1.21384	1.04623	1.40831	0.008629968
HIC1	1.16348	1.01124	1.33864	1.22483	1.06765	1.40515	0.008997589
PODXL	1.16034	1.02143	1.31815	1.12456	0.98675	1.2816	0.009182736
IL18BP	1.17737	1.0254	1.35188	1.07299	0.91441	1.25908	0.009206805
DKC1	1.05822	0.92046	1.2166	1.22886	1.05463	1.43187	0.009315729
B3GALT5	0.95137	0.83621	1.08239	1.20365	1.05804	1.3693	0.009772808
SNRPN	0.85441	0.7514	0.97155	0.98115	0.86779	1.1093	0.00995797

[1]The markers detected here could be markers indicating susceptibility; [2]*P* from logistic regression model (all cases were included).

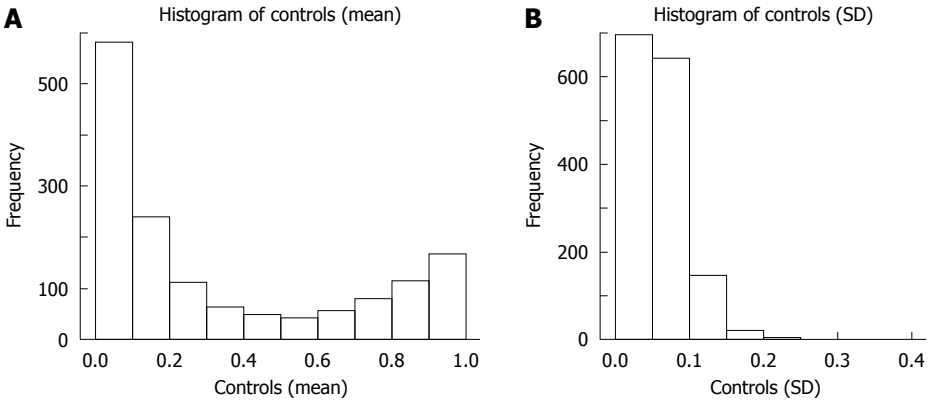

Figure 1 Histogram for data distribution. A: Means of β for all CpG sites in controls; B: SD of β for all CpG sites in controls.

were similar with regard to age, supplement treatment, family history of CRC, smoking starting age, and smoking duration (years) (Table 1). Cases had relatively higher body mass index than controls. The median age of cancer diagnosis of cases was 69.5 years old, which is over 10 years after study entry (median = 58 years old). The histograms (Figure 1) show the data distribution of β values for all CpG sites in controls, suggesting most CpG sites in controls were unmethylated.

In analysis comparing all cases to all controls using logistic regression model, the CpG sites with differential methylation are candidate sites potentially related to susceptibility. Of the 1505 CpG sites analyzed, 25 sites were associated with CRC risk at $P < 0.01$ level and are shown in Table 2. None of these 25 sites exceeded a Bonferroni threshold for significance (threshold is 3.3×10^{-5} for 1500 tests). However, the most statistically significant CpG site associated with risk was DSP_P440_R ($P = 6.0 \times 10^{-4}$) with the second tertile displaying a 1.01-fold risk (95% CI: 0.96-1.26) and the third tertile displaying a 1.26-fold risk (95% CI: 1.12-1.43) comparing with the lowest tertile of methylation.

To capture markers whose status changed over time (potentially reflecting tumor progression), we incorporated the time variable, latency, in analysis using proportional odds regression by coding the cases as three

Table 3 Methylation of CpG sites related to colorectal cancer development with different latencies (*P* value < 0.01 before multiple testing correction)[1]

Marker name	2nd tertile *vs* 1st tertile			3rd tertile *vs* 1st tertile			*P*[2]
	OR	L95	U95	OR	L95	U95	
DSP	1.52027	0.91054	2.53831	2.62423	1.65105	4.17103	0.000104474
PTCH2	1.86565	1.14946	3.02807	2.55851	1.60114	4.08832	0.000327688
GLA	3.22896	1.66054	6.27879	3.62345	1.70627	7.69478	0.000529374
FLT4	2.49882	1.55051	4.02713	1.7312	1.02958	2.91095	0.000651567
B3GALT5	0.91735	0.5746	1.46454	2.11128	1.31709	3.38434	0.000726807
INSR	1.88977	1.16281	3.07122	2.44371	1.52713	3.91043	0.0007899
GSTM2	2.29464	1.44903	3.63373	1.92103	1.18011	3.12712	0.001165928
FASTK	1.66117	1.02457	2.69331	2.26999	1.43313	3.59553	0.001992703
MLH3	1.12063	0.67503	1.86036	2.15187	1.31098	3.53213	0.002175877
HOXC6	1.2189	0.75795	1.96015	2.06861	1.32056	3.24043	0.002826542
MGMT	2.42747	1.34114	4.39373	2.71599	1.45027	5.08638	0.00448402
SYBL1	1.35661	0.89446	2.05754	0.63927	0.39971	1.02239	0.004618893
PDE1B	2.12203	1.33959	3.36151	1.74663	1.01704	2.9996	0.004839975
SEMA3C	1.34045	0.87915	2.04378	0.62126	0.38566	1.0008	0.004858838
AFF3	2.11473	1.30773	3.41973	1.88147	1.16588	3.03628	0.005422481
HCK	2.0406	1.27766	3.25911	2.0338	1.19416	3.46382	0.005503
PXN	1.92122	1.24421	2.9666	1.12493	0.66824	1.89374	0.005766585
PARP1	1.19708	0.79209	1.80913	0.55056	0.33191	0.91323	0.006004429
TK1	1.83172	1.07512	3.12075	2.24792	1.34064	3.76922	0.006184503
SGCE	1.21217	0.74619	1.96916	1.9685	1.25257	3.09363	0.006354949
MME	2.00858	1.25423	3.21663	1.94042	1.20308	3.12967	0.00664085
RARA	1.51526	0.94405	2.43209	2.07639	1.31635	3.27525	0.006899358
COL18A1	2.08647	1.28893	3.37752	1.90316	1.17926	3.07144	0.007110996
SEMA3F	1.98588	1.21715	3.24014	2.03976	1.22672	3.39166	0.007292652
ITGB4	2.08685	1.2669	3.43748	2.02919	1.20223	3.42495	0.007595246
PTPRH	0.47922	0.30082	0.76343	0.71909	0.44823	1.15361	0.007598903
SMO	1.52188	0.95276	2.43094	2.08604	1.3032	3.33915	0.008280298
DKC1	1.25326	0.76383	2.05632	2.20393	1.27785	3.80116	0.008427231
TRPM5	2.22962	1.33245	3.73087	1.80604	1.01389	3.21707	0.008971559
EVI1	1.62829	1.01415	2.61431	2.08373	1.29638	3.34927	0.009481166
EPHA3	1.93252	1.21576	3.07185	1.85827	1.15814	2.98165	0.009792546

[1]Latency: according to the time difference between blood draw and cancer diagnosis, cases were classified to three categories-short, medium, and long latency; [2]*P* from proportional odds regression model (all cases were included).

categories according to latency. Of the 1505 CpG sites analyzed, 31 sites were associated with CRC risk at $P < 0.01$ level (Table 3). Though none of these 31 sites remained statistically significant after Bonferroni correction, the most statistically significant CpG site associated with CRC risk achieved a *P* value of 1.0×10^{-4}. The CpG site is located in *DSP* gene, and the risk estimate was 1.52 (95% CI: 0.91-2.53) and 2.62 (95% CI: 1.65-4.17) for the second and third tertile comparing with the lowest tertile respectively.

To further explore the time-related DNA methylation change, we conducted partial logistic regression analysis by excluding the one third of the cases with the longest latency, and only comparing cases with short and medium latency to controls. In this analysis, 26 sites were associated with CRC risk at $P < 0.01$ level (Table 4). None of them remain statistically significant after Bonferroni correction. GLA_E98_R was the top hit with over 30% elevated risk ($P = 1.3 \times 10^{-4}$) for subjects with highest methylation level compared to those at the lower level.

As stated in methods, changes in methylation of CpG sites over time, can suggest a relationship to tumor progression; models incorporating time/latency information (the proportional odds model and the partial logistic regression model) are designed to detect such a pattern. The regular logistic regression model comparing all cases to all controls should select all risk-associated markers. These include both susceptibility-related markers (which differ between cases and controls and do not change over time), as well as progression-related, (which change over time with tumor progression). To differentiate the CpG sites whose methylation levels are susceptibility related from those that are potentially tumor progression- related, we compared the top CpG sites selected from the three models (full logistic regression, proportional odds, and partial logistic regression). We found that methylation levels of some CpG sites, like those in the *DSP* gene, were consistently associated with CRC risk across three models, suggesting these associations were unaltered by time prior to diagnosis, or tumor progression; the available evidence therefore suggests that they are related to cancer predisposition. On the other hand, the methylation levels of some other sites, such as *B3GALT5* and *GLA* genes, were associated with CRC risk at lower p values in analysis from models incorporating latency information than from standard model (the all logistic re-

Table 4 Methylation of CpG sites associated with colorectal cancer with short to mediate latency (*P* value < 0.01 before multiple testing correction)[1]

Marker name	2nd tertile *vs* 1st tertile			3rd tertile *vs* 1st tertile			*P*[2]
	OR	L95	U95	OR	L95	U95	
GLA	1.31618	1.12303	1.54256	1.31082	1.09072	1.57534	0.000126734
DSP	1.12406	0.9868	1.2804	1.25115	1.11146	1.4084	0.000305565
B3GALT5	1.00109	0.88664	1.1303	1.18931	1.05066	1.34625	0.001334159
MLH3	1.01719	0.89178	1.16025	1.2404	1.08916	1.41263	0.001421327
PTCH2	1.18044	1.04387	1.33487	1.24209	1.1026	1.39922	0.002140961
FLT4	1.26038	1.1159	1.42356	1.15518	1.01153	1.31924	0.002164413
SGCE	1.05149	0.92678	1.19299	1.19126	1.05786	1.34148	0.002535873
HOXC6	1.02303	0.90426	1.15741	1.16253	1.03342	1.30777	0.002602644
PDE1B	1.19845	1.06385	1.35009	1.10584	0.96046	1.27323	0.002706305
COL18A1	1.2126	1.07142	1.37239	1.19961	1.06058	1.35687	0.002968143
TRPM5	1.16951	1.02324	1.3367	1.10173	0.94879	1.27933	0.003084609
EVI1	1.06179	0.93967	1.19979	0.94512	0.8274	1.07959	0.003902443
GSTM2	1.24009	1.10312	1.39407	1.13952	1.0059	1.29089	0.003956314
FASTK	1.12763	0.99513	1.27778	1.17539	1.04318	1.32436	0.004798386
PARP1	1.08475	0.97053	1.2124	0.87907	0.77119	1.00203	0.005028903
PTGS2	1.03008	0.91145	1.16415	1.15319	1.02574	1.29648	0.00571389
WRN	0.97839	0.8614	1.11128	1.09867	0.96629	1.24918	0.005779497
EPO	1.14731	1.01334	1.299	1.13034	0.9854	1.2966	0.006090622
GPC3	1.19056	1.03981	1.36316	1.09698	0.94341	1.27556	0.006093671
TK1	1.13458	0.98846	1.30232	1.20862	1.05753	1.3813	0.006272625
EVI1	1.14882	1.01559	1.29953	1.17851	1.04142	1.33364	0.006474115
EPHB2	1.15596	1.02875	1.2989	1.05233	0.92701	1.19459	0.006504075
SEMA3C	1.10197	0.98427	1.23375	0.89579	0.7921	1.01304	0.00677389
HLA_DPA1	1.14664	1.00965	1.30223	1.09326	0.95971	1.24538	0.006960326
SYBL1	1.06835	0.9547	1.19553	0.90635	0.80226	1.02394	0.007029222
INSR	1.19021	1.05214	1.34639	1.25474	1.11332	1.41412	0.008045993

[1]The logistic regression analysis only includes cases with short and medium latency; [2]*P* from proportional odds regression model (all cases were included).

gression model); the time/latency pattern observed suggests that the markers are tumor progression- related.

DISCUSSION

In our study, more than 1500 CpG sites from 807 genes were evaluated among 440 study subjects, which is the first prospective study for CRC risk focusing on gene-specific DNA methylation to date. By comparing all the cases with all the controls, we found that the methylation levels of some CpG sites were associated with CRC risk. The methylation of these genes might be related to CRC susceptibility. By incorporating latency information in the analysis, we found some associations between DNA methylation and CRC risk were latency-related, suggesting the methylation in these genes altered with tumor progression. However, these results are cross-sectional, and further studies with repeated blood draws from the same subjects are needed to refine the temporal pattern prior to cancer diagnosis among same individuals.

Following the report from Feinberg and colleagues that colon cancer cells had different DNA methylation patterns in some genes compared to their normal counterparts[42], many studies have been conducted using colon cancer as a cancer model for methylation regulation in the past three decades. Methylation in several genes had been reported to be involved in the pathogenesis of colorectal precursor lesions and adenomas, as well as adenocarcinomas[43]. However, epidemiological studies to date have only tested limited sets of CpG sites in small numbers of subjects[36,44,45]. The reported associations require confirmation and methylation levels of other genes remained to be described. Although further replication is needed, our study is among the most comprehensive and large studies.

Epigenetic regulation of gene expression is specific for each cell type, within different tissues, according to stages of development or differentiation[11,46]. Most studies of DNA methylation in CRC have been conducted by comparing tumor tissue DNA with adjacent normal tissue DNA from cancer patients[47]. Circulating DNA from serum also has been studied[48-50], although this might reflect methylation status in DNA released from tumor tissue, and could be influenced by tumor volume, vascularity, stage, grade, metastasis, *et al*. Methylation patterns from other DNA sources have only had fragmentary study[36,51]. Selected epigenetic markers could be heritable from one generation to the next[52], or affected by aging or environmental exposures throughout life[53]. It is reasonable to hypothesize that DNA methylation status in some genes is associated with cancer susceptibility. Therefore, methylation patterns of germline tissue, including that derived from blood leukocyte DNA might be related to cancer risk. Changes of methylation in leu-

kocyte DNA have been shown to parallel other somatic tissues in mice[27], and have been linked to susceptibility of certain cancers[28]. Certainly, we could not rule out the possibility that different tissues have different response to environmental exposure and that methylation status in leukocytes may not fully reflect the changes in the target tissue[27]. In the current study, we found methylation levels of some genes, like *DSP*, were consistently associated with CRC risk across all three models, suggesting a relationship to cancer risk that is invariant over time, that is unaffected by tumor progression. We recognize that the procedure that favor tumor initiation and progression may have both similar (as we observed) and independent elements and specifically targeted populations will be required to explore the later hypothesis in future studies. Though further confirmation is required, our results provide some evidence that DNA methylation of some genes could be related to cancer susceptibility.

Aberrantly increased DNA methylation in the promoter region of *hMLH1* gene in germline DNA has been reported among nonpolyposis CRC patients[32-35]. Hypermethylation in the *MSH2* gene of germline DNA was also observed among early onset CRC patients[31]. Interestingly, the inverse relationship between leukocyte DNA genomic methylation and colorectal adenoma progression was found stronger for nonadvanced rather than advanced adenoma in a case-control study, which suggests that leukocyte DNA genomic methylation may be more important as an etiologic factor in early adenomas[54]. All these data suggest that the methylation levels of some genes either increase or decrease over time with tumor progression. Though further studies are required to replicate and confirm, our results suggest that methylation levels of some genes, i.e., *GLA, B3GALT5, et al*, could be markers of tumor progression, and methylation levels of some genes, i.e., *DSP et al*, were more likely markers of susceptibility. Also, some genes whose methylation levels were positively associated with CRC risk, i.e., *DSP, MLH3, et al,* suggest they might be tumor suppressing-like genes, and those showed negative association, i.e., *PLAL1, et al,* might be oncogenic-like genes.

Our study has several advantages. First, it is nested within the ATBC cohort, which offers a substantial, representative, well characterized, and relatively homogeneous population of male smokers. Second, data on cancer diagnoses and key covariates were systematically collected. Third, this is the first prospective population-based study to explore the role of gene-specific DNA methylation in colorectal carcinogenesis with a large number of CpG sites. DNA obtained from blood samples at baseline allows us to assess methylation status prior to cancer diagnosis, and explore the latency effect on the associations. Although to our knowledge the sample size is among the largest reported to date, variations of DNA methylation in leukocytes are less prominent than those observed in colorectal tissue/tumor comparisons; thus we had limited power to detect markers with moderate and small effect sizes. Another limitation is that this study is restricted to male Finnish smokers. The advantage of homogeneity implies the limitation that the results are not entirely generalizable to other populations. Furthermore, we only have one blood draw from each subject, which limits our power to precisely explore the direct effect of latency on the association between DNA methylation and cancer risk.

In summary, the results from this prospective study suggest that the methylation level of some genes were associated with cancer susceptibility and some were related with tumor progression. Further studies are warranted to confirm and refine our results.

ACKNOWLEDGMENTS

The authors thank Tim Sheehy and Dominick Parisi for sample preparation, and Kirk Snyder, Information Management Services, Silver Spring, MD, for data management. Finally, we acknowledge the study participants for donating their time and making this study possible.

COMMENTS

Background

Recent data suggest a link between leukocyte DNA methylation and cancer risk. However, data on DNA methylation from a prospective study, which may provide evidence for causality, are sparse.

Research frontiers

Results may provide new insights into the etiology of colorectal cancer (CRC) and suggest new cancer prevention strategies.

Innovations and breakthroughs

Data on DNA methylation from a prospective study, which could provide causality evidence, are missing. This is among the first prospective studies to examine gene-specific methylation status in germ-line DNA prior to CRC diagnosis.

Applications

Results may provide new insights into the etiology of CRC and suggest new cancer prevention strategies. The approach we propose, methylation profiling of PBMC DNA from prospectively obtained samples and its relation to CRC risk, should be relevant to other cancers.

Terminology

Methylation: Methylation is the covalent addition of a methyl group to cytosine residue, which is an epigenetic event that affects cell function by altering gene expression. Prospective study: Prospective study follows over time a group of similar individuals who differ with respect to certain factors under study, to determine how these factors affect rates of a certain disease.

Peer review

The comparison was made between ca. 200 pairs, thus small case-control study with comprehensive methylation pattern. There were no sites where methylation differences survived Bonferroni threshold of significance. Time-dependent DNA methylation was suspected, but no cases gave DNA at different times. However, the design is ambitious and the results are potentially important.

REFERENCES

1 **Jemal A**, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; **61**: 69-90
2 **Esteller M**. Epigenetics in cancer. *N Engl J Med* 2008; **358**: 1148-1159
3 **Cho YH**, Yazici H, Wu HC, Terry MB, Gonzalez K, Qu M, Dalay N, Santella RM. Aberrant promoter hypermethylation and genomic hypomethylation in tumor, adjacent normal tissues and blood from breast cancer patients. *Anticancer Res*

2010; **30**: 2489-2496
4 **Gaston V**, Le Bouc Y, Soupre V, Burglen L, Donadieu J, Oro H, Audry G, Vazquez MP, Gicquel C. Analysis of the methylation status of the KCNQ1OT and H19 genes in leukocyte DNA for the diagnosis and prognosis of Beckwith-Wiedemann syndrome. *Eur J Hum Genet* 2001; **9**: 409-418
5 **Friso S**, Girelli D, Trabetti E, Olivieri O, Guarini P, Pignatti PF, Corrocher R, Choi SW. The MTHFR 1298A& gt; C polymorphism and genomic DNA methylation in human lymphocytes. *Cancer Epidemiol Biomarkers Prev* 2005; **14**: 938-943
6 **Shelnutt KP**, Kauwell GP, Gregory JF, Maneval DR, Quinlivan EP, Theriaque DW, Henderson GN, Bailey LB. Methylenetetrahydrofolate reductase 677C--& gt; T polymorphism affects DNA methylation in response to controlled folate intake in young women. *J Nutr Biochem* 2004; **15**: 554-560
7 **Singal R**, Das PM, Manoharan M, Reis IM, Schlesselman JJ. Polymorphisms in the DNA methyltransferase 3b gene and prostate cancer risk. *Oncol Rep* 2005; **14**: 569-573
8 **Lamprecht SA**, Lipkin M. Chemoprevention of colon cancer by calcium, vitamin D and folate: molecular mechanisms. *Nat Rev Cancer* 2003; **3**: 601-614
9 **Sadikovic B**, Rodenhiser DI. Benzopyrene exposure disrupts DNA methylation and growth dynamics in breast cancer cells. *Toxicol Appl Pharmacol* 2006; **216**: 458-468
10 **Peters U**, Hayes RB, Chatterjee N, Shao W, Schoen RE, Pinsky P, Hollis BW, McGlynn KA. Circulating vitamin D metabolites, polymorphism in vitamin D receptor, and colorectal adenoma risk. *Cancer Epidemiol Biomarkers Prev* 2004; **13**: 546-552
11 **Lopes EC**, Valls E, Figueroa ME, Mazur A, Meng FG, Chiosis G, Laird PW, Schreiber-Agus N, Greally JM, Prokhortchouk E, Melnick A. Kaiso contributes to DNA methylation-dependent silencing of tumor suppressor genes in colon cancer cell lines. *Cancer Res* 2008; **68**: 7258-7263
12 **Flood A**, Peters U, Jenkins DJ, Chatterjee N, Subar AF, Church TR, Bresalier R, Weissfeld JL, Hayes RB, Schatzkin A. Carbohydrate, glycemic index, and glycemic load and colorectal adenomas in the Prostate, Lung, Colorectal, and Ovarian Screening Study. *Am J Clin Nutr* 2006; **84**: 1184-1192
13 **Peters U**, Sinha R, Chatterjee N, Subar AF, Ziegler RG, Kulldorff M, Bresalier R, Weissfeld JL, Flood A, Schatzkin A, Hayes RB. Dietary fibre and colorectal adenoma in a colorectal cancer early detection programme. *Lancet* 2003; **361**: 1491-1495
14 **Reik W**, Dean W, Walter J. Epigenetic reprogramming in mammalian development. *Science* 2001; **293**: 1089-1093
15 **Arasaradnam RP**, Commane DM, Bradburn D, Mathers JC. A review of dietary factors and its influence on DNA methylation in colorectal carcinogenesis. *Epigenetics* 2008; **3**: 193-198
16 **Brower V**. Epigenetics: Unravelling the cancer code. *Nature* 2011; **471**: S12-S13
17 **Sinha R**, Peters U, Cross AJ, Kulldorff M, Weissfeld JL, Pinsky PF, Rothman N, Hayes RB. Meat, meat cooking methods and preservation, and risk for colorectal adenoma. *Cancer Res* 2005; **65**: 8034-8041
18 **Issa JP**. The epigenetics of colorectal cancer. *Ann N Y Acad Sci* 2000; **910**: 140-153; discussion 153-155
19 **Kondo Y**, Issa JP. Epigenetic changes in colorectal cancer. *Cancer Metastasis Rev* 2004; **23**: 29-39
20 **Wheeler JM**. Epigenetics, mismatch repair genes and colorectal cancer. *Ann R Coll Surg Engl* 2005; **87**: 15-20
21 **Wong JJ**, Hawkins NJ, Ward RL. Colorectal cancer: a model for epigenetic tumorigenesis. *Gut* 2007; **56**: 140-148
22 **Zitt M**, Zitt M, Müller HM. DNA methylation in colorectal cancer--impact on screening and therapy monitoring modalities? *Dis Markers* 2007; **23**: 51-71
23 **Jass JR**. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology* 2007; **50**: 113-130
24 **Jass JR**. Molecular heterogeneity of colorectal cancer: Implications for cancer control. *Surg Oncol* 2007; **16** Suppl 1: S7-S9
25 **Ogino S**, Goel A. Molecular classification and correlates in colorectal cancer. *J Mol Diagn* 2008; **10**: 13-27
26 **Søreide K**. [Genetics and molecular classification of colorectal cancer]. *Tidsskr Nor Laegeforen* 2007; **127**: 2818-2823
27 **McKay JA**, Xie L, Harris S, Wong YK, Ford D, Mathers JC. Blood as a surrogate marker for tissue-specific DNA methylation and changes due to folate depletion in post-partum female mice. *Mol Nutr Food Res* 2011; **55**: 1026-1035
28 **Marsit CJ**, Koestler DC, Christensen BC, Karagas MR, Houseman EA, Kelsey KT. DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. *J Clin Oncol* 2011; **29**: 1133-1139
29 **Liu Z**, Li W, Lei Z, Zhao J, Chen XF, Liu R, Peng X, Wu ZH, Chen J, Liu H, Zhou QH, Zhang HT. CpG island methylator phenotype involving chromosome 3p confers an increased risk of non-small cell lung cancer. *J Thorac Oncol* 2010; **5**: 790-797
30 **Piyathilake CJ**, Macaluso M, Alvarez RD, Chen M, Badiga S, Siddiqui NR, Edberg JC, Partridge EE, Johanning GL. A higher degree of LINE-1 methylation in peripheral blood mononuclear cells, a one-carbon nutrient related epigenetic alteration, is associated with a lower risk of developing cervical intraepithelial neoplasia. *Nutrition* 2011; **27**: 513-519
31 **Stumpel DJ**, Schneider P, van Roon EH, Boer JM, de Lorenzo P, Valsecchi MG, de Menezes RX, Pieters R, Stam RW. Specific promoter methylation identifies different subgroups of MLL-rearranged infant acute lymphoblastic leukemia, influences clinical outcome, and provides therapeutic options. *Blood* 2009; **114**: 5490-5498
32 **Hitchins M**, Williams R, Cheong K, Halani N, Lin VA, Packham D, Ku S, Buckle A, Hawkins N, Burn J, Gallinger S, Goldblatt J, Kirk J, Tomlinson I, Scott R, Spigelman A, Suter C, Martin D, Suthers G, Ward R. MLH1 germline epimutations as a factor in hereditary nonpolyposis colorectal cancer. *Gastroenterology* 2005; **129**: 1392-1399
33 **Hitchins MP**, Wong JJ, Suthers G, Suter CM, Martin DI, Hawkins NJ, Ward RL. Inheritance of a cancer-associated MLH1 germ-line epimutation. *N Engl J Med* 2007; **356**: 697-705
34 **Suter CM**, Martin DI, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* 2004; **36**: 497-501
35 **Valle L**, Carbonell P, Fernandez V, Dotor AM, Sanz M, Benitez J, Urioste M. MLH1 germline epimutations in selected patients with early-onset non-polyposis colorectal cancer. *Clin Genet* 2007; **71**: 232-237
36 **Kaaks R**, Stattin P, Villar S, Poetsch AR, Dossus L, Nieters A, Riboli E, Palmqvist R, Hallmans G, Plass C, Friesen MD. Insulin-like growth factor-II methylation status in lymphocyte DNA and colon cancer risk in the Northern Sweden Health and Disease cohort. *Cancer Res* 2009; **69**: 5400-5405
37 **Ally MS**, Al-Ghnaniem R, Pufulete M. The relationship between gene-specific DNA methylation in leukocytes and normal colorectal mucosa in subjects with and without colorectal tumors. *Cancer Epidemiol Biomarkers Prev* 2009; **18**: 922-928
38 The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. The ATBC Cancer Prevention Study Group. *Ann Epidemiol* 1994; **4**: 1-10
39 **Albanes D**, Malila N, Taylor PR, Huttunen JK, Virtamo J, Edwards BK, Rautalahti M, Hartman AM, Barrett MJ, Pietinen P, Hartman TJ, Sipponen P, Lewin K, Teerenhovi L, Hietanen P, Tangrea JA, Virtanen M, Heinonen OP. Effects of supplemental alpha-tocopherol and beta-carotene on colorectal cancer: results from a controlled trial (Finland) *Cancer Causes Control* 2000; **11**: 197-205
40 **Malila N**, Virtamo J, Virtanen M, Albanes D, Tangrea JA,

Huttunen JK. The effect of alpha-tocopherol and beta-carotene supplementation on colorectal adenomas in middle-aged male smokers. *Cancer Epidemiol Biomarkers Prev* 1999; **8**: 489-493

41 **Teschendorff AE**, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs IJ, Widschwendter M. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One* 2009; **4**: e8274

42 **Feinberg AP**, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 1983; **301**: 89-92

43 **Carmona FJ**, Esteller M. Epigenomics of human colon cancer. *Mutat Res* 2010; **693**: 53-60

44 **Ye C**, Shrubsole MJ, Cai Q, Ness R, Grady WM, Smalley W, Cai H, Washington K, Zheng W. Promoter methylation status of the MGMT, hMLH1, and CDKN2A/p16 genes in non-neoplastic mucosa of patients with and without colorectal adenomas. *Oncol Rep* 2006; **16**: 429-435

45 **Fujii S**, Tominaga K, Kitajima K, Takeda J, Kusaka T, Fujita M, Ichikawa K, Tomita S, Ohkura Y, Ono Y, Imura J, Chiba T, Fujimori T. Methylation of the oestrogen receptor gene in non-neoplastic epithelium as a marker of colorectal neoplasia risk in longstanding and extensive ulcerative colitis. *Gut* 2005; **54**: 1287-1292

46 **Gargiulo G**, Minucci S. Epigenomic profiling of cancer cells. *Int J Biochem Cell Biol* 2009; **41**: 127-135

47 **Ogino S**, Chan AT, Fuchs CS, Giovannucci E. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. *Gut* 2011; **60**: 397-411

48 **Greenland S**. Principles of multilevel modelling. *Int J Epidemiol* 2000; **29**: 158-167

49 **Wallner M**, Herbst A, Behrens A, Crispin A, Stieber P, Göke B, Lamerz R, Kolligs FT. Methylation of serum DNA is an independent prognostic marker in colorectal cancer. *Clin Cancer Res* 2006; **12**: 7347-7352

50 **Ji BT**, Weissfeld JL, Chow WH, Huang WY, Schoen RE, Hayes RB. Tobacco smoking and colorectal hyperplastic and adenomatous polyps. *Cancer Epidemiol Biomarkers Prev* 2006; **15**: 897-901

51 **van Roon EH**, van Puijenbroek M, Middeldorp A, van Eijk R, de Meijer EJ, Erasmus D, Wouters KA, van Engeland M, Oosting J, Hes FJ, Tops CM, van Wezel T, Boer JM, Morreau H. Early onset MSI-H colon cancer with MLH1 promoter methylation, is there a genetic predisposition? *BMC Cancer* 2010; **10**: 180

52 **Kile ML**, Baccarelli A, Tarantini L, Hoffman E, Wright RO, Christiani DC. Correlation of global and gene-specific DNA methylation in maternal-infant pairs. *PLoS One* 2010; **5**: e13730

53 **Christensen BC**, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, Sugarbaker DJ, Yeh RF, Wiencke JK, Kelsey KT. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 2009; **5**: e1000602

54 **Lim U**, Flood A, Choi SW, Albanes D, Cross AJ, Schatzkin A, Sinha R, Katki HA, Cash B, Schoenfeld P, Stolzenberg-Solomon R. Genomic methylation of leukocyte DNA in relation to colorectal adenoma among asymptomatic women. *Gastroenterology* 2008; **134**: 47-55

S- Editor Wang JL **L- Editor** A **E- Editor** Zheng XM