

A New Blood-Based Epigenetic Diagnostic Biomarker Test (Episwitch® NST) with High Sensitivity and Positive Predictive Value for Colorectal Cancer and Precancerous Polyps

Ewan Hunter , [Heba Alshaker](#) , Cicely Weston , Mutaz Issa , Shekinah Bautista , Abel Gebregzabhar , [Anyu Viridi](#) , [Ann Dring](#) , [Ryan Powell](#) , [Jayne Green](#) , [Roshan Lal](#) , Vamsi Velchuru , Kamal Aryal , Muhammad Radzi Bin Abu Hassan , [Goh Tiong Meng](#) , [Janisha Suriakant Patel](#) , [Shameera Pharveen Mohamed Gani](#) , [Chun Ren Lim](#) , [Thomas Guie](#) , [Alexandre Akoulitchev](#) , [Dmitri Pchejetski](#) *

Posted Date: 10 December 2024

doi: 10.20944/preprints202412.0872.v1

Keywords: Colorectal cancer; colonic polyp; chromosome conformations; epigenetics; blood-based biomarkers; 3D-genomic profiling; cancer diagnosis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

A New Blood-Based Epigenetic Diagnostic Biomarker Test (EpiSwitch® NST) with High Sensitivity and Positive Predictive Value for Colorectal Cancer and Precancerous Polyps

Ewan Hunter ¹, Heba Alshaker ², Cicely Weston ¹, Mutas Issa ¹, Shekinah Bautista ¹, Abel Gebregzabhar ¹, Anya Virdi ¹, Ann Dring ¹, Ryan Powell ¹, Jayne Green ¹, Roshan Lal ³, Vamsi Velchuru ³, Kamal Aryal ³, Muhammad Radzi Bin Abu Hassan ⁴, Goh Tiong Meng ⁵, Janisha Suriakant Patel ⁶, Shameera Pharveen Mohamed Gani ⁶, Chun Ren Lim ⁶, Thomas Guiel ⁷, Alexandre Akoulitchev ¹ and Dmitri Pchejetski ²

¹ Oxford BioDynamics plc, Oxford UK

² Norwich Medical School, University of East Anglia

³ James Paget University Hospitals NHS Trust, Lowestoft Rd, Gorleston-on-Sea, Great Yarmouth NR31 6LA

⁴ Clinical Research Center, Hospital Sultanah Bahiyah, Jalan Langgar, 05460 Alor Setar, Kedah, Malaysia.

⁵ Island Hospital Penang, 308, Jalan Macalister, 10450 George Town, Pulau Pinang

⁶ Oxford BioDynamics plc, Penang Reference Laboratory, Island 88 118 4 9, Jalan Tanjung Tokong, Penang 10470, Malaysia

⁷ Oxford BioDynamics Inc, Frederick, MD, USA

* Correspondence: d.pshezhetskiy@uea.ac.uk

Simple Summary: This should be written in one paragraph before the Abstract in layman's terms, to explain why the research is being suggested, what the authors aim to achieve, and how the findings from this research may impact the research community. Please use as few abbreviations as possible and do not cite references in the Simple Summary. The Simple Summary should not exceed 150 words. Submissions without a simple summary will be returned directly. An example can be found at <https://www.mdpi.com/2072-6694/12/9/2424>.

Abstract: Background/Objectives: Colorectal cancer (CRC) arises from the epithelial lining of the colon or rectum, often following a progression from benign adenomatous polyps to malignant carcinoma. Screening modalities such as colonoscopy, faecal immunochemical tests (FIT), and FIT-DNA are critical for early detection and prevention, but non-invasive methods lack sensitivity to polyps and early CRC. Chromosome conformations (CCs) are potent epigenetic regulators of gene expression. We have previously developed an epigenetic assay, EpiSwitch®, that employs an algorithmic-based CCs analysis. Using EpiSwitch® technology, we have shown the presence of cancer-specific CCs in peripheral blood mononuclear cells (PBMCs) and primary tumours of patients with melanoma and prostate cancer. EpiSwitch®-based commercial tests are now available to diagnose prostate cancer with 94% accuracy (PSE test) and response to immune checkpoint inhibitors across 14 cancers with 85% accuracy (CiRT test). **Methods/Results/Conclusions:** Using blood samples collected from n=171 patients with CRC, n=44 patients with colorectal polyps and n=110 patients with 'clear' colonoscopy we performed whole Genome DNA screening for CCs correlating to CRC diagnosis. Our findings suggest the presence of two Eight-marker CC signatures (EpiSwitch® NST) in whole blood that allow diagnosis of CRC and precancerous polyps, respectively. Independent validation cohort testing demonstrated high accuracy in identifying colorectal polyps and early versus late stages of CRC with an exceptionally high sensitivity of 79-90% and a high positive prediction value of 60-84%. Linking the top diagnostic CCs to nearby genes, we have built pathways maps that likely underline processes contributing to the pathology of polyp and CRC progression, including TGFβ, cMYC, Rho GTPase, ROS, TNFα/NFκB, and APC.

Keywords: Colorectal cancer; colonic polyp; chromosome conformations; epigenetics; blood-based biomarkers; 3D-genomic profiling; cancer diagnosis.

1. Introduction

Globally, colorectal cancer (CRC) is the third most common cancer type, accounting for 10% of all cancer cases. There were 1.9 million new cases and 930,000 deaths from the disease in 2020 [1]. CRC arises from the epithelial lining of the colon or rectum, often following a progression from benign adenomatous polyps to malignant carcinoma driven by genetic mutations, epigenetic alterations, and chronic inflammation [2]. More than 80% of CRC arise from adenomatous polyps and outgrowths of the lining of the colon or rectum, which are usually asymptomatic [3]. Some of the inherited genetic disorders, such as familial adenomatous polyposis and hereditary non-polyposis colon cancer, can cause CRC and are responsible for circa 5% of all CRC cases. However, 75–95% of CRC cases occur in people with little or no genetic risk [4].

CRC diagnosis is usually performed via colonoscopy and biopsy. The most common form of CRC is adenocarcinoma, constituting between 95% and 98% of all cases. Imaging such as computed tomography (CT), magnetic resonance imaging (MRI) or positron emission tomography (PET) scans is used to identify local and distal spread and to plan the optimal surgical approach [5]. Treatments for CRC include surgery, radiation therapy, chemotherapy, and immuno-oncology therapy with checkpoint inhibitors [6]. Early CRC stages (1 and 2) are confined within the wall of the colon and could be treated radically with a combined surgical and medical approach. Late stages (3 and 4) often spread widely and are not curable. The individual likelihood of survival depends on how advanced the cancer is.

In this context, screening from the age of 45 (50 in the UK) for early detection of CRC is considered an effective measure for preventing and decreasing deaths from CRC [7]. Screening for this cancer is effective for both early detection and prevention and allows diagnosis 2–3 years before the symptoms arise. Polyps often can be removed at the precancerous stage, and an effective screening programme has the potential to reduce CRC deaths by 60% [8]. Currently, the primary screening tests include colonoscopy, faecal occult blood testing (FOBT), and monitoring of cell-free DNA from CRC tumours in blood. Colonoscopy is considered a gold-standard diagnostic test for CRC, and its sensitivity is ~95%. However, up to 20% of colonoscopies are unsuccessful due to poor preparation or difficult anatomy and cancers in these patients are missed.

Furthermore, colonoscopy bears a significant risk of bleeding and bowel perforation – up to 4% each [9]. Two main non-invasive screening tests include guaiac-based or immunochemical-based detection of blood in stool, FOBT and faecal immunochemical test (FIT), respectively. They have high specificity and negative predictive value (NPV) but lower sensitivity and positive predictive value (PPV). The latest studies confirm that these tests may miss more than half of bowel cancer cases, especially at the lower stages [10]. Other screening options include virtual CT-colonoscopy and stool DNA screening testing (FIT-DNA). Colonoscopy via a CT scan is expensive, associated with radiation exposure, and cannot remove any detected abnormal growths as standard colonoscopy can. Stool FIT-DNA screening test also looks for altered DNA associated with CRC and precancerous lesions but has a high level of false positive results [9]. The UK bowel cancer screening programme includes an FOBT test every two years between the ages of 50 and 74. FOBT overdiagnosis ranges from 2.0% to 7.6%, leading to unnecessary colonoscopies (with or without biopsies), patient distress and extra costs [11].

With the advent of epigenetic research, it has become evident that epigenetic modifications like aberrant DNA methylation [12] and histone acetylation [13] are related to CRC onset. Three-dimensional chromatin conformations (CCs), as part of genomic regulatory architecture, are also potent epigenetic regulators of gene expression and cellular pathological phenotypes [14]. Long-range epigenetic alterations in CCs were found in primary CRC and circulating DNA from CRC patients [15].

We have previously developed an epigenetic assay, EpiSwitch® [16], that employs an algorithmic-based CCs analysis. Using EpiSwitch® technology, we have shown the presence of cancer-specific CCs in peripheral blood mononuclear cells (PBMCs) and primary tumours of patients with melanoma [17,18] and prostate cancer [19]. In light of the regulatory role lately attributed to systemic exosome traffic, we have used indirect co-culture experiments or conditioned media and demonstrated horizontal transfer of CCs between cultured cancer cells and monocytes without direct contact [20]. EpiSwitch®-based commercial tests are now available to diagnose prostate cancer with 94% accuracy (PSE test) [21] and response to immune checkpoint inhibitors across 14 cancers with

85% accuracy (CiRT test) [22]. Interestingly, although the anchor sites associated with 3D genomic loops are scattered throughout genomes, by linking the top prognostic biomarkers to nearby genes (within 3Kb), it is possible to learn a great deal about the underlying processes contributing to the pathology of a disease and identify potential therapeutic strategies.

In this retrospective (with partial prospective collection) case-control study, we have used n=325 full blood samples collected from n=171 patients with CRC, n=44 patients with colorectal polyps and n=110 patients with ‘clear’ colonoscopy attending colorectal clinics and performed whole Genome DNA screening for CCs correlating to CRC diagnosis. Our findings suggest the presence of two Eight-marker CC signatures in whole blood that allow rapid and cost-effective diagnosis of CRC and precancerous polyps, respectively.

2. Materials and Methods

Patient Characteristics

In this retrospective case-control study with partial prospective recruitment, n=325 whole blood samples (n=110 controls (no polyp or cancer on colonoscopy), n=44 polyp and n = 171 CRC) were obtained from patients attending colorectal clinics at James Paget University Hospital, UK Hospital Sultanah Bahiyah, Malaysia and Island Hospital, Malaysia (Table 1). Inclusion criteria: clinical and histopathological diagnosis of CRC, precancerous lesion and normal colonoscopy, no prior history of any cancer, treatment naïve, and age range 18-75. A blood sample was taken prior to treatment. N=225 samples (n=68 control patients and n=157 CRC) were collected retrospectively, and N=100 patients were recruited through a prospective observational study yielding (n=42 controls, n=44 polyps, n=14 CRC).

All samples were collected at the time of diagnosis and randomly allocated for training and test cohorts. The study was approved by the UK National Research Ethics Committee and Medical Research as well as Ethics Committee Ministry of Health Malaysia, and conducted in accordance with Good Clinical Practice guidelines and the Declaration of Helsinki. All participants provided written informed consent. All data were pseudo-anonymised. All procedures and protocols were performed in accordance with the relevant guidelines and regulations.

Table 1. Summary of clinical characteristics for patient cohorts used for biomarker discovery.

Cohort	N (total)	Prospective	Retrospective	Male	Female	Age (mean)
Control	110	42	68	56	54	61
Polyp	44	44	0	29	15	63
CRC	171	14	157	89	82	64

Preparation of 3D Genomic Templates

A 5 mL full blood sample was collected from cancer patients and controls using BD Vacutainer® plastic EDTA tubes. The tubes were frozen and stored at -80 °C. Isolation of DNA from the whole cell lysate was performed as previously described, and DNA was fixed with formaldehyde. To identify interchromatin loops, fixed chromatin was digested into fragments with TaqI restriction enzyme, and the resulting DNA strands were joined, favouring cross-linked fragments. The cross-links were reversed, and PCR was performed using the primers designed using the algorithms of the EpiSwitch® software (as described in detail in [17–19,23]).

3C libraries were quantified using the Quant-iT™ Picogreen dsDNA Assay kit (Invitrogen) and normalised to 5 ng/μL prior to interrogation by PCR. The EpiSwitch® Explorer arrays were performed as published previously, with the modification of only one sample being hybridised to each array slide in the Cy3 channel. EpiSwitch® Explorer arrays, based on the Agilent SureSelect array platform, allow for the highly reproducible, non-biased interrogation of ~1.1 million anchor sites for 3D genomic interactions (964,631 experimental probes and 2500 control probes).

Custom Microarray Design

Custom microarrays were designed through the EpiSwitch® software that uses a pattern recognition algorithm based on DNA sequence, which operates on Bayesian modelling and yields a probability score of whether a region is involved in long-range chromatin interactions. GRCh38

human genome assembly was annotated across ~1.1 million sites, and the potential to form long-range chromosome conformations was quantified for each region [18,19,23–26]. The most probable interactions were identified and filtered on probabilistic score and proximity to protein, long non-coding RNA, or microRNA coding sequences. Predicted interactions were limited to EpiSwitch® sites larger than 10 kb and less than 300 kb apart. Repeat masking and sequence analysis were used to ensure unique marker sequences for each interaction. The EpiSwitch® Explorer array (Agilent Technologies, Product Code X-HS-AC-02), containing 60-mer oligonucleotide probes, was designed to interrogate potential 3D genomic interactions. 964,631 experimental and 2,500 control probes were added to a 1 × 1 M CGH microarray slide design. The experimental probes were placed on the design in singlicate with the controls in groups of 250. The control probes consisted of six different EpiSwitch® interactions generated during the extraction processes and used to monitor library quality. Four external inline control probe designs were added to detect non-human (*Arabidopsis thaliana*) spike-in DNA during the sample labelling protocol to provide a standard curve and control for labelling. The external spike DNA consists of 400 bp ssDNA fragments from genomic regions of *A. thaliana*. Array-based comparisons were performed as described previously, with the modification of only one sample being hybridised to each array slide in the Cy3 channel [18,19,23–26].

Microarray Statistical Analysis

Microarray readouts were normalised by background correction and quantile normalisation using the EpiSwitch® R analytic package, which is built on the Limma and dplyr libraries. Data was corrected for batch effects using the ComBat R script. Parametric (Limma R library, Linear Regression) and non-parametric (EpiSwitch® RankProd R library) statistical methods were performed to identify 3D genomic changes that demonstrated a difference in abundance between cancers and controls. The resulting data from both procedures were further filtered based on adjusted p-value (false discovery rate (FDR) correction) and abundance scores (AS). Only 3D genomic markers with adjusted p-value ≤ 0.05 and AS $-1.1 \leq$ or ≥ 1.1 were selected. Both filtered lists from Limma and RankProd analysis were compared, and the intersection of the two lists was chosen for further processing.

Step 1

Probes are selected based on their corrected p-value FDR, which is the product of a modified linear regression model. Probes below p-value ≤ 0.1 are selected and then further reduced by their fold change (FC); probes FC have to be ≤ -1.1 or ≥ 1.1 to be chosen for further analysis. The last filter is a coefficient of variation (CV); probes must be below ≤ 0.3 .

Step 2

The top 250 markers from the statistical lists are selected based on their FC for selection as markers for PCR translation.

Step 3

The resultant markers from step 1, the statistically significant probes, form the basis of enrichment analysis using hypergeometric enrichment (HE). This analysis enables marker reduction from the significant probe list and, along with the markers from step 2, forms the list of probes translated onto the EpiSwitch™ PCR platform.

The statistical probes are processed by HE to determine which genetic locations have an enrichment of statistically significant probes, indicating which genetic locations are hubs of epigenetic difference.

The most significant enriched loci based on a corrected p-value are selected for probe list generation. Genetic locations below a p-value of 0.3 or 0.2 are selected. The statistical probes mapping these genetic locations, with the markers from step 2, form the high-value markers for EpiSwitch™ PCR translation.

Translation of Array-Based 3D Genomic Markers to PCR Readouts

In the discovery cohort, we analysed the leading array-derived markers using Oxford BioDynamics (OBD's) proprietary primer design software. This process aimed to pinpoint genomic locations that are appropriate for a hydrolysis probe-based real-time PCR assay [27]. Briefly, the top array-derived markers associated with diagnostic potential were filtered on fold change and adjusted p-value. PCR primer probes were ordered from Eurofins genomics as salt-free primers. The probes

were designed with a 5' FAM fluorophore, 3' IABkFQ quencher and an additional internal ZEN quencher and ordered from iDT (Integrated DNA Technologies) [28]. Each assay was optimised using a temperature gradient PCR with an annealing temperature range from 58-68°C. Individual PCR assays were tested across the temperature gradient alongside negative controls, including soluble and unstructured commercial TaqMan human genomic DNA control (Life Technologies), and a TE buffer-only negative control was used. Assay performance was assessed based on Cq values, the reliability of detection, and efficiency based on the slope of the individual amplification curves. Assays that passed the quality criteria and presented reliable detection differences between cancers and controls were used to screen individual patient samples.

EpiSwitch® PCR

Each patient sample was interrogated using triplicate real-time PCR. Each reaction consisted of 50 ng of EpiSwitch® library template, 250 mM of each of the primers, 200 mM of the hydrolysis probe and a final 1X Kapa Probe Force Universal (Roche) concentration in a final 25 µL volume. The PCR cycling and data collection were performed using a CFX96 Touch Real-Time PCR detection system (Bio-Rad). The annealing temperature of each assay was changed to the optimum temperature identified in the temperature gradients performed during translation for each assay. Otherwise, the same cycling conditions were used: 98°C for 3 minutes followed by 45 cycles of 95°C for 10 seconds and 20 seconds at the identified optimum annealing temperature. The individual well Cq values were exported from the CFX manager software after baseline and threshold value checks.

PCR Statistical Analysis

The 250 markers screened on 40 individual patient samples were subject to permuted logistic modelling with bootstrapping for 500 data splits and non-parametric Rank Product analysis (EpiSwitch® RankProd R library). Two machine learning procedures (eXtreme Gradient Boosting: XGBoost and CatBoost) were used to reduce the feature pool further and identify the most predictive/prognostic 3D genomic markers. The resulting markers were then used to build the final classifying models using CatBoost and XGBoost. All analyses were performed using R statistical language with Caret, XGBoost, SHAPforxgboost and CatBoost libraries.

Biological Network/Pathway Analysis

Network analysis for functional/biological relevance of the 3D genomic markers was performed using the Hallmark Gene Sets and BioCarta and Reactome Canonical Pathway gene sets from the Molecular Signatures Database (MSigDB) [29]. Protein interaction networks were generated using the Search Tool for the Retrieval of Interacting Proteins (STRING)

Results

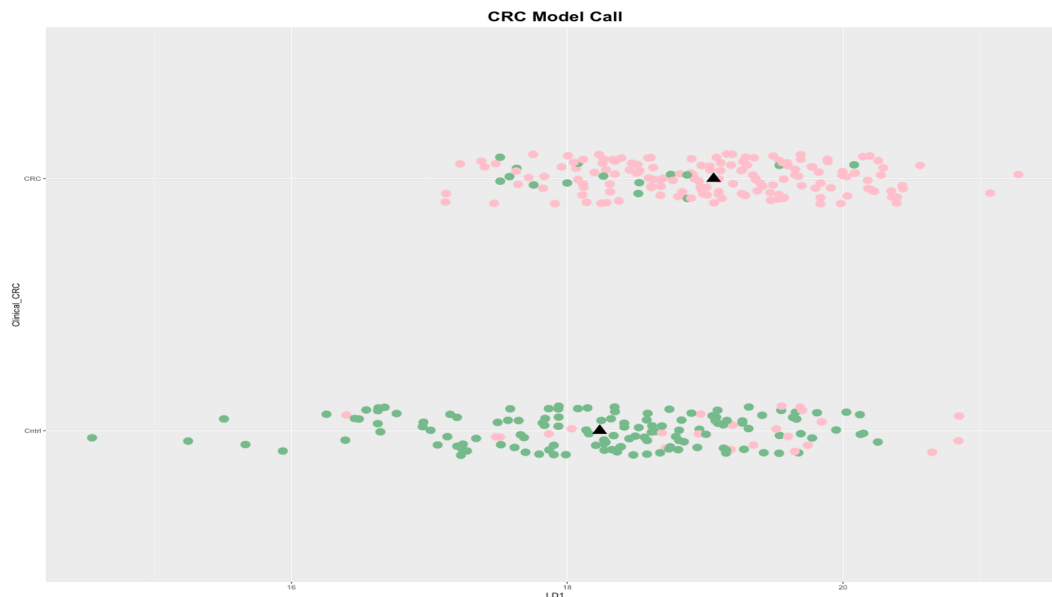
4 In this retrospective (with partial prospective collection) case-control study, n=325 whole blood samples (n=110 controls, n=44 polyps and n = 171 cancers) were obtained from patients attending colorectal clinics (Table 1). Patients were separated according to diagnosis (CRC, polyp and control), and blood samples were taken prior to treatment. All control subjects had a 'clear' colonoscopy. Pre-lesion polyps were confirmed by a biopsy and histopathology. Pre-lesion and control samples were collected through a prospective observational study.

Microarrays

To design custom microarrays, we used the EpiSwitch® pattern recognition algorithm, which predicts long-range chromatin interactions through Bayesian modelling and provides a probabilistic score for each region. ~1.1 million sites across the whole GRCh38 human genome assembly were identified as having the potential to form long-range chromosome conformations [18,19,23–26]. The most probable interactions were identified and filtered on probabilistic score and proximity to protein, long non-coding RNA, or microRNA coding sequences. Predicted interactions were limited to EpiSwitch® sites that were more significant than 10 kb and less than 300 kb apart. Repeat masking and sequence analysis were used to ensure unique marker sequences for each interaction.

Whole-genome EpiSwitch® Explorer arrays were used to screen PBMCs samples collected at the time of confirmed diagnosis. All cohorts showed separation by principal component analysis (PCA) for CRC and polyp without pre-selection or reduction of the 964,631 array markers (Figure 1 A and B), suggesting that 3D genomic profiles associated with different clinical outcomes exist and can be distinguished.

A



B

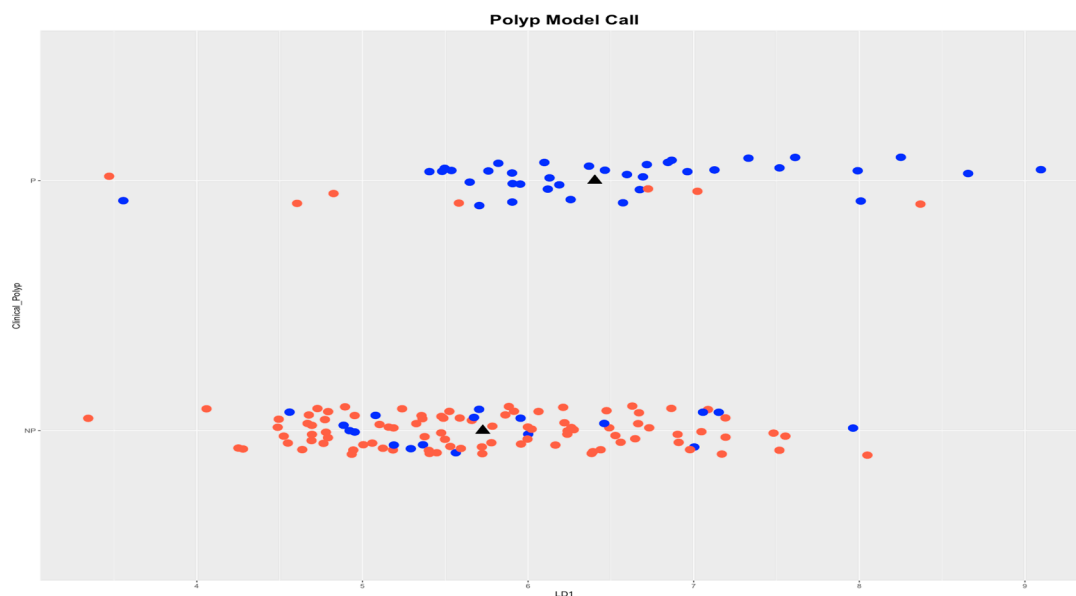


Figure 1. Principal component analysis (PCA) plot of all CRC patients versus control (A) and CRC versus polyp (B). A. The PCA plot of patients with CRC (pink circles) vs non-CRC (green circles). B. The PCA plot of patients with polyps (blue circles) vs control (orange circles). The analysis is based on whole genome profiling of all 964,631 3D genomic markers screened without any marker reduction.

To evaluate the biological relevance of the observed separation of patients with CRC, polyps and control patients, the 964,631 3D genomic markers from each patient were subject to statistical testing using both parametric testing (Limma) and non-parametric testing (EpiSwitch® RankProd), both procedures that correct for multiple testing by using FDR corrections. The RankProd approach also has a resampling step to control for random rank importance, adding another layer of statistical stringency in marker selection when testing many possibilities. The selected markers were filtered based on an adjusted FDR p-value ≤ 0.05 and high abundance scores (AS), $-1.1 \leq$ or ≥ 1.1 . Similar approaches and thresholds for FDR cut-offs have been used in previously published biomarker development studies [18,19,23–27]. Thus, starting with the 964,631 whole genome screened *cis*-interactions. After statistical filtering, the 250 3D genomic markers with the highest and lowest abundance scores were chosen for further analysis and PCR translation. The top 250 EpiSwitch® array markers were identified as statistically significant and consistently present, based on standard statistical analysis based on p-value and adjusted p-value, when screened on blood from patients with CRC, polyps or control (Figure 1, complete list in supplementary table 1). These top 250 markers were randomly distributed throughout the human genome (Figure 2).

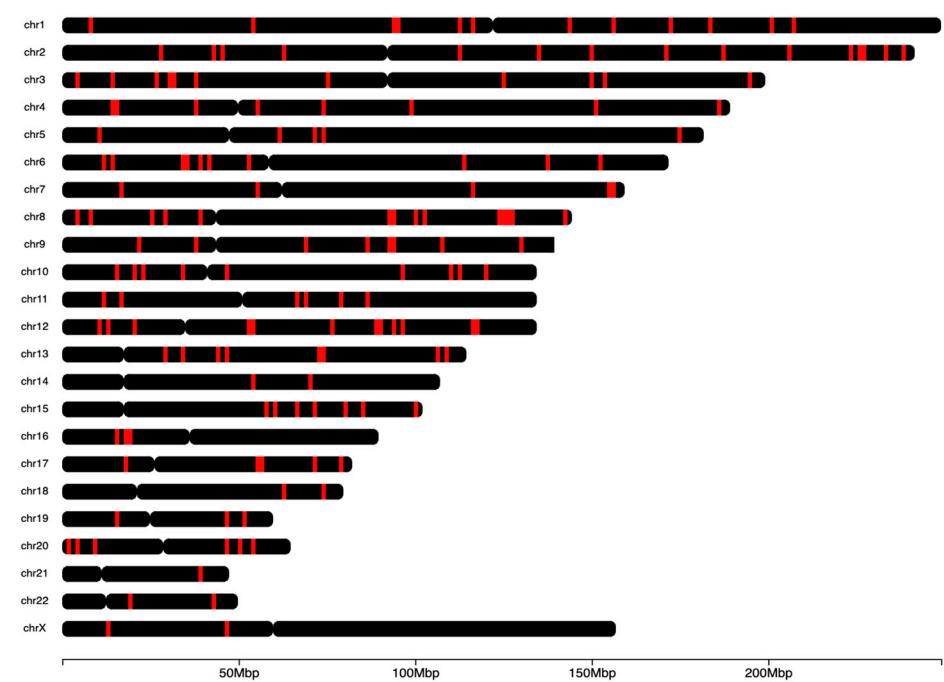


Figure 2. Genome-wide mapping of 250 3D genomic loci associated with CRC. Genomic locations and distribution of the top 250 3D genomic markers for CRC. Individual human chromosomes are shown on the y-axis (chr1–chr22 and the X chromosome). The heatmap shows the number of markers within a 0.3Mb genomic window, with black representing a low density of markers and red indicating a high density of markers.

qPCR Validation of Biomarkers

To translate the EpiSwitch® Explorer array markers to a PCR-detectable assay for clinical use, primers to detect individual 3D genomic markers were generated and validated. Starting with whole blood samples from the training set, we identified feature reduction using machine learning methods on the initial pool of 250 3D genomic biomarkers through feature reduction of 12 markers with predictive power to differentiate between CRC patients and controls. These 12 array markers had strong identifiers for CRC and polyps and were selected for qPCR translation.

The 12 qPCR markers were further refined on the training cohort of $n=74$ CRC (both early and late stages) and non-CRC (polyps + controls), and the top eight markers from 12 were built into a classifier model using machine learning package XGBoost (Table 2). In a blind validation study of 251 samples, this marker set (EpiSwitch® no stool test (NST)) correctly classified 125 samples as CRC and 89 as non-CRC, with the remaining 37 samples as false positives or false negatives. It demonstrated

high accuracy of 85%, 90% sensitivity, 79% specificity, and 84% PPV in stratifying patients, with and without CRC (Table 3).

Table 2. Eight PCR biomarkers for CRC classifier.

qPCR markers	Array marker	Probe sequence	P.Value	adj.P.Val	FC	Gene	GeneDist
obd156_q117_7_q1179	ORF1_1_116481182_116484855_116627241_116630872_RF	TTGACATAGGACCTCA GCAGAGAGCAGCTCG AGATCCACCCACGTTG TTGCATGTATCAA	0.0263014	1	1.29588 9575	RP5-1086K13.1;CD58;NAP1L4P1; 0;0;0;0;0;23 MIR548AC;IGSF3;AL355794. 41;40878 1;RP4-655J12.4;MIR320B1	
obd156_q131_3_q1315	ORF1_5_61009121_61015983_61116919_61125541_FR	GAGGCAGGCAGATCA CAAGGTCAAGAGATCG ATAAGTACATGAGAAA TAAACAAAATTCA	2.33E-07	8.13E-05	1.38950 2347	NDUFAF2;CTC-436P18.4;ERCC8;CTC-436P18.5	0;0;64049;20306
obd156_q130_1_q1303	ORF1_12_93013996_93019448_9310234_5_93106201_FR	TGATGGACTTATGGAC TCATTCACTGCATCGA TATGGCTCATGCCATT TTATGTGCTATC	2.02E-08	2.64E-05	1.44548 2624	RP11-511B23.1;RP11-511B23.2;Y_RNA;RP11-511B23.4;RPL41P5;RP11-202G11.2;AC138123.1;RNU6-1329P;NACAP3	0;0;0;0;0;0;61252;17863
obd156_q118_5_q1187	ORF1_1_201477609_201480715_201569360_201570965_RF	ACAAAGCTATCTCATT TCCTGAGCTTCATCGA GGTGAGGAGATCATGG ATGAGTTTTTTA	0.0247499	1	1.53825 2027	CSRP1;RP11-134G8.7;RP11-134G8.5;RP11-134G8.6;PHLDA3;NAV1	0;0;0;0;8373;51921
obd156_q124_5_q1247	ORF1_8_8307248_8309141_8529093_8530943_RF	CAATAATTCATTCTTCT TCATCAGTCCCTTCGAA CTCCTGACTCAGGAGA TCTATCCACCT	0.0161572	1	1.34291 3514	SGK223;CTA-398F10.1;CTA-398F10.2;FAM86B3P;CTD-3023L14.3	0;0;0;62384;24796
obd156_q121_7_q1219	ORF1_1_94060570_94064104_94081020_94084795_RF	TCTTGCCGGGAGTACT CTTCAAATCCTTCGA CATGATGGAGAAGCTG TCCAGGAACCAG	0.0000011	0.000163	1.53548 0941	ABCA4;RP5-837O21.2;RP11-78O9.1	0;125327;60317
obd156_q129_7_q1299	ORF1_15_71449255_71457687_71567140_71571578_RR	GTAAGTAATAATAGTG TATGTGTTTATGTCGA CTGTACTGGCGGACCC TATAAGAGGCAG	6.85E-06	0.000421	1.46248 0028	THSD4;RP11-1123I8.1;RP11-592N21.2;AC104938.1	0;0;100785;201033
obd156_q122_5_q1227	ORF1_15_67079527_67081854_67195948_67198335_RF	ATCTGTCCCAATCCTTT ATCCTTCTAGCTCGAG TCAGCAGTGTGACTG TTAGCAAATCA	1.8E-07	7.03E-05	1.65268 5053	SMAD3;RP11-342M21.2;RP11-798K3.2;AAGAB	0;0;20275;2699

Table 3. Eight-marker diagnostic set validation cancer (all stages) versus non-cancer.

Test	Present	n	Absent	n	Total
Yes	True positive	125	False positive	25	150
No	False negative	14	True negative	89	103
Total		139		114	
Statistic		Value (%)		95% CI	
Sensitivity		89.93		83.68 to 94.38	
Specificity		79.46		70.80 to 86.51	
Positive Likelihood Ratio		4.38		3.03 to 6.33	
Negative Likelihood Ratio		0.13		0.08 to 0.21	
Disease prevalence		55.38		49.00 to 61.63	
Positive Predictive Value		84.46		78.99 to 88.71	
Negative Predictive value		86.41		79.31 to 91.33	
Accuracy		85.26		80.26 to 89.40	

To assess the validity of this marker set for early-stage CRC detection, late-stage (stages 3 and 4) CRC were excluded from analysis. In the resulting early-stage cohort of n=149 samples, n=31 samples were correctly classified as cancer (CRC), and n=89 samples were correctly classified as non-cancer, with the remaining 29 samples being either false positives or negatives (table 4). The test showed an accuracy of 81% with 84% sensitivity, 79% specificity, 57% PPV and 94% NPV) in identifying patients with early stages of CRC.

Table 4. Eight-marker diagnostic set validation early cancer (stages 1 and 2) versus non-cancer.

Test	Present	n	Absent	n	Total
Yes	True positive	31	False positive	23	54
No	False negative	6	True negative	89	95
Total		37		112	
Statistic		Value (%)		95% CI	
Sensitivity		83.78		67.99 to 93.81	
Specificity		79.46		70.80 to 86.51	
Positive Likelihood Ratio		4.08		2.76 to 6.03	
Negative Likelihood Ratio		0.2		0.10 to 0.43	
Disease prevalence		24.83		18.13 to 32.57	
Positive Predictive Value		57.41		47.69 to 66.58	
Negative Predictive value		93.68		87.64 to 96.88	
Accuracy		80.54		73.26 to 86.56	

Non-invasive detection of precancerous polyps remains a significant clinical challenge. While not technically a cancer, they often have precancerous features and have the potential to develop into CRC. Currently, only colonoscopy has good diagnostic accuracy, while existing non-invasive tests cannot be used for this purpose [31]. For patients classified as non-CRC, a second assessment has been evaluated. Based on the data from the polyps training cohort (n=29), a separate set of top eight markers (with four markers overlapping with the CRC eight-marker set (supplementary table 2)) from 12 were built into a classifier model using the machine learning approach XGBoost (Table 5).

Using this new polyp-specific eight-marker set (Episwitch® NST), we have performed blinded validation on n=142 non-CRC samples. Of those, 27 samples were correctly classified as polyp, and 90 were correctly classified as control. The remaining 25 samples are either false positives or negatives. This test achieved a high overall diagnostic accuracy of 82% with 79% sensitivity, 83% specificity, 60% PPV and 92% NPV to detect the presence of adenomas/precancerous lesions/polyps (Table 6).

Table 5. Eight PCR biomarkers for polyps classifier.

qPCR markers	Array marker	Probe sequence	P.Value	adj.P.Val	FC	Gene	GeneDist
obd156_q1205	ORF1_13_734350	ACACACAGTAGGTAATT	4.43E-08	3.46E-05	1.424866474	LINC00393;MAR K2P12;LINC00392	0;26702;77701
53_73437099	734	AATACGGTGGATCGAAG					
_q1207	84222_73486544	TACGCTCTAGTTATACG					
	RF	AGGCTTGTT					
obd156_q1213	ORF1_9_3791992	CCGAGGTCCCGAGACTA	2.49E-07	8.27E-05	1.373971322	RP11-613M10.9;SHB;RNU7-124P;SLC25A51;AL161448.1	0;0;0;15573;141722
5_37923489	3800	TCTGCCAATCCTCGATTC					
_q1215	2100_38004773	F TCTGGTTTTCAGTTTGT					
	R	CTATCTT					
obd156_q1273	ORF1_2_1132099	CCAACACCACCCCAAAT	0.0178566	1	1.456428166	PAX8-AS1;PAX8;RP11-65112.1;PSD4;IGKV10R2-108	0;0;0;507;128903
02_113215780	11	GCCGGGGCAGTCGAGC					
_q1275	3275966_1132774	GTCCCCGGTTATTGGGA					
	94_FR	AGGGTGCGC					
obd156_q1293	ORF1_9_9321863	TTTATATAACAATGTTTT	1.12E-07	5.53E-05	1.438635096	WNK2;RP11-370F5.4;C9orf129	0;70077;40134
2_93223726	9327	TTTCAAGGCTTCGAGCA					
_q1295	4460_93278066	R GACATTTCCTCGAGG					
	F	AAGTAACA					
obd156_q1245	ORF1_8_8307248	CAATAATTCATTCTTCTT	0.0161572	1	1.342913514	SGK223;CTA-398F10.1;CTA-398F10.2;FAM86B3P;CTD-3023L14.3	0;0;0;62384;24796
_q1247	8309141_852909	CATCAGTCCTTCGAACT					
	3_8530943	RF CCTGACTCAGGAGATCT					
		ATCCACCT					
obd156_q1217	ORF1_1_9406057	TCTTGCCGGGAGTACTC	0.0000011	0.000163	1.535480941	ABCA4;RP5-837O21.2;RP11-78O9.1	0;125327;60317
_q1219	94064104_9408	TTCAAACCTCTCGACAT					
	1020_94084795	R GATGGAGAAGCTGTCCA					
	F	GGAACCAG					
obd156_q1297	ORF1_15_714492	GTAAGAATAATAGTGT	6.85E-06	0.000421	1.462480028	THSD4;RP11-112318.1;RP11-592N21.2;AC104938.1	0;0;100785;201033
_q1299	71457687_715	ATGTGTTTATGTGCACTG					
	67140_71571578	TACTGGCGGACCCTATA					
	RR	AGAGGCAG					
obd156_q1225	ORF1_15_670795	ATCTGTCCCAATCCTTTA	1.8E-07	7.03E-05	1.652685053	SMAD3;RP11-342M21.2;RP11-798K3.2;AAGAB	0;0;20275;2699
_q1227	27_67081854	671 TCCTTCTAGCTCGAGTCA					

95948_67198335_	GCAGTGTGACTGTTAG
RF	CAAATCA

Table 6. Blinded validation of eight-marker ‘polyp’ diagnostic set for the presence of polyp versus control

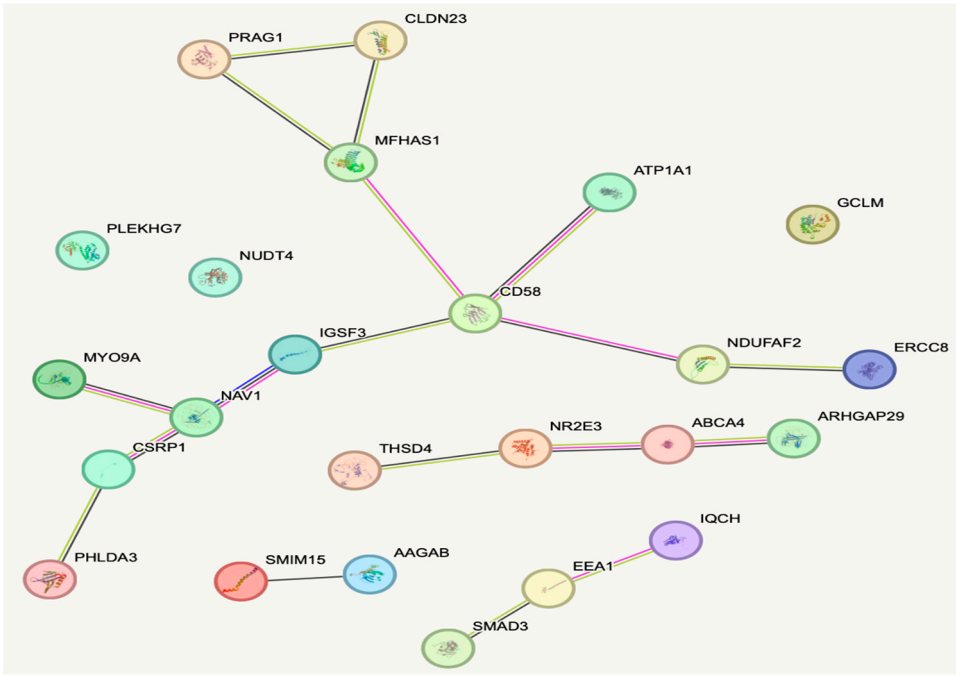
Test	Present	n	Absent	n	Total
Yes	True positive	27	False positive	18	45
No	False negative	7	True negative	90	97
Total		34		108	

Statistic	Value (%)	95% CI
Sensitivity	79.41	62.10 to 91.30
Specificity	83.33	74.94 to 89.81
Positive Likelihood Ratio	4.76	3.02 to 7.51
Negative Likelihood Ratio	0.25	0.13 to 0.48
Disease prevalence	23.94	17.19 to 31.82
Positive Predictive Value	60	48.76 to 70.28
Negative Predictive value	92.78	86.86 to 96.16
Accuracy	82.39	75.12 to 88.27

Previous analysis has indicated that changes in the 3D chromosome architecture captured using EpiSwitch® biomarkers are also reflected in the broader region surrounding each biomarker. Analysis of these regions can give insights into the causes of the observed phenotype [18,19,23–27]. The genomic positions of the 250 3D genomic markers were mapped to enable the identification of the three closest protein-coding loci. Potential functional roles for these loci were obtained using Hallmark Gene Sets, BioCarta and Reactome canonical pathway analysis. Pathway analysis for the **eight colorectal classifier markers** showed multiple pathways involved in CRC, including transforming growth factor beta (TGFβ), cMYC, Rho GTPase, reactive oxygen species (ROS) and adenomatous polyposis coli (APC) (supplementary table 3). A similar analysis of the eight-marker polyp classifier showed pathways related to TGFβ, PAX8, epithelial-mesenchymal transition (EMT), tumour necrosis factor alfa (TNFα)/nuclear factor kappa B (NFκB), ROS, and APC (supplementary table 4).

When evaluating the biological function of the genes within the genomic regions identified as being dysregulated between patients with CRC, polyps and controls, several biological pathways with known associations to cancer were identified. Analysis of the top 3D genomic markers associated with CRC using the Search Tool for Retrieval of Interacting Genes (STRING) database revealed eight marker CRC and polyp panels protein-protein interaction networks with hubs on cluster of differentiation (CD)58, Mothers against decapentaplegic homolog (SMAD)3 and interleukin 1 receptor (IL1R) (Figure 3A and B).

A.



B.

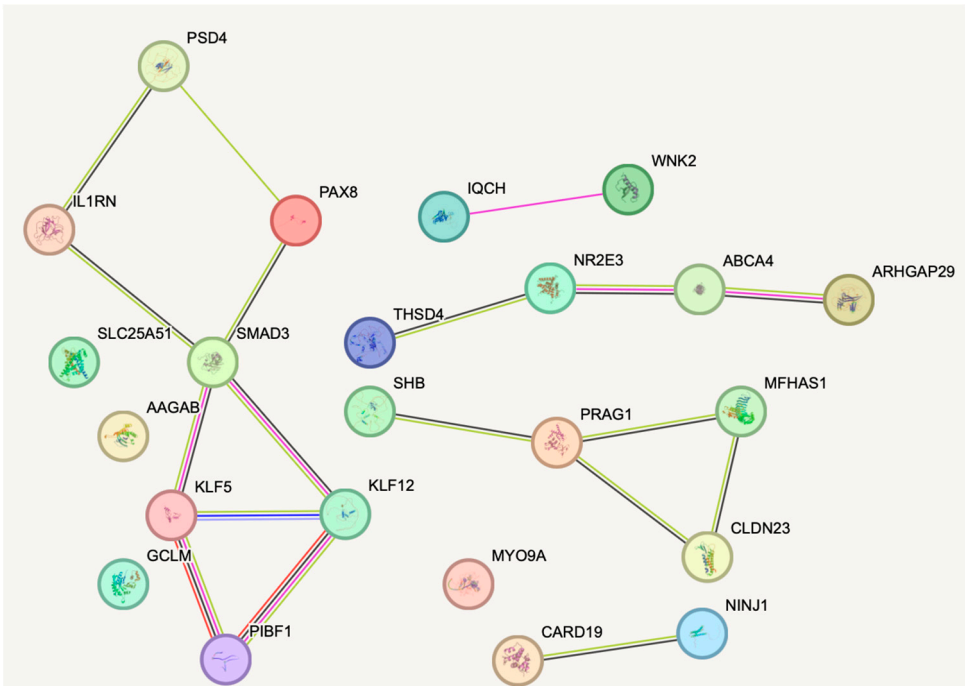


Figure 3. STRING network of the eight colorectal classifier markers (A) and eight polyp classifier markers (B). STRING Network associated with CRC and colorectal polyps. The proteins encoded by genes in the vicinity of the top 3D genomic markers related to CRC and polyps reveal a network with hubs, as shown.

Discussion

One of the main challenges in CRC management is reliable early detection, with active treatments and prophylactics of precancerous lesions offering effective cures and reduced mortality. Colonoscopy with subsequent biopsy remains a gold-standard test for all types of CRC and precancerous lesions. However, it is invasive, costly, and requires secondary care settings and expertise. Most current early detection tests in CRC and polyps perform poorly in detecting early

stages of CRC, resulting in significant false positive readings. Most precancerous polyps are missed (table 7) [32–35].

Table 7. Comparative efficacy of non-invasive diagnostic tests for CRC and polyps.

Detection of precancerous lesions (polyps)						
	EpiSwitch® NST	Cologuard	FIT	Freenome PREEMPT CRC®	Guardant Shield®	Colonoscopy
Sensitivity	79%	43%	23%	13%	13%	75%
Specificity	83%	91%	95%	92%	90%	89%
PPV	60%	36%	35%	39%	17%	40%
NPV	93%	93%	91%	73%	86%	99%
Accuracy	82%	86%	87%	70%	77%	86%
Detection of early-stage (I/II) CRC						
	EpiSwitch® NST	Cologuard	FIT	Freenome PREEMPT CRC®	Guardant Shield®	Colonoscopy
Sensitivity	84%	90%	60%	79%	65%	75%
Specificity	79%	91%	95%	92%	90%	89%
PPV	57%	2%	3%	3%	1%	80%
NPV	94%	100%	100%	100%	100%	86%
Accuracy	81%	91%	95%	92%	90%	80%
Detection of CRC (stages I - IV)						
	EpiSwitch® NST	Cologuard	FIT	Freenome PREEMPT CRC®	Guardant Shield®	Colonoscopy
Sensitivity	90%	97%	71%	82%	83%	75%
Specificity	79%	91%	95%	92%	90%	89%
PPV	84%	2%	3%	3%	2%	80%
NPV	86%	100%	100%	100%	99%	86%
Accuracy	85%	91%	95%	92%	90%	80%
Abbreviations: positive predictive value, PPV ; negative predictive value, NPV						

Epigenetics is the study of heritable changes in phenotype that do not involve alterations in the DNA sequence. [36]. These changes are powered by the modification of gene expression and occur through three principal mechanisms: DNA methylation, histone modification, changes in 3D chromosome structures, and chromosomal looping (36). Chromosomal loops are the dynamic 3D chromatin structures that exist in all cells and determine gene expression in so-called “active cluster” regions. They are potent regulators of gene expression during tumorigenesis [37].

Genome-wide Association Studies (GWAS) have shown that epigenetic alterations are often more important than genetic mutations during oncogenic transformation. For example, in cancers, loss of gene expression occurs about ten times more frequently by epigenetic transcription silencing than by mutations [38]. One interesting observation from GWAS was that most allele regions of the genome that confer risk to cancer are outside known protein-coding regions [39]. In CRC, only ~5% of all cases are due to inherited genetic disorders (such as familial adenomatous polyposis), and 75–95% of CRC cases occur in people with little or no genetic risk [4]. There is a growing understanding of the role of epigenetic regulation in CRC progression (reviewed in [40]).

Chromatin conformations are often controlled by non-coding RNAs, which may also regulate tumour-specific conformations [41]. It was shown that tumour cells can secrete non-coding RNAs endocytosed by neighbouring or circulating cells and change their chromosomal conformations in a process called “horizontal transfer”[42,43].

We have performed a proof-of-concept study where we co-cultured prostate cancer cells and primary macrophages in a Boyden chamber (through a membrane of 0.45µM without direct contact) or exposed macrophages to prostate cancer cell-conditioned media. In both cases, we have detected new prostate cancer-specific chromosomal conformation changes in the macrophages [20]. We have confirmed these findings in patients with melanoma, where we have shown the presence of identical melanoma-specific chromatin conformations in primary tumours of melanoma patients and their PBMCs (17, 18). Blood cell fractionation showed that the detected signature comes from PBMCs, not

circulating tumour cells [17]. A similar approach was used to confirm the presence of prostate cancer-specific signatures in PBMCs from prostate cancer patients.

Using our EpiSwitch® technology, we have created a new diagnostic test for melanoma [17,18], thyroid cancer [26], and prostate cancer [19,21] (PSE test), capable of identifying prostate cancer presence with 94% overall accuracy. This test is currently available in the clinic.

In this retrospective (with partial prospective collection) case-control study, we have analysed n=325 whole blood samples from n=171 patients with CRC, n=44 patients with colorectal polyps and n=110 patients with 'clear' colonoscopy.

Initially, all three cohorts were compared using the whole-genome EpiSwitch® Explorer arrays, which showed a remarkable separation between CRC and non-cancer, and polyp and control, without pre-selection or reduction of the 964,631 array markers (Figure 1 A and B). This aligns with our previous findings in prostate cancer [19,21] and melanoma [17,18], where CCs profiles could accurately distinguish cancer patients from controls. Biomarker reduction and translation from DNA CHIPs to PCR is a crucial step in EpiSwitch® technology, where statistical filtering and machine learning methods allowed us to reduce initially to 250 and then to 12 3D genomic biomarkers.

Unlike other cancers, CRC often has a well-identifiable precancerous stage manifesting as colonic polyps. These polyps could be relatively easily found during colonoscopy, but other non-invasive methods have difficulties identifying them (Table 7). Since polyps are usually asymptomatic and can be easily removed (therefore reducing the cancer risk), their early non-invasive identification is the next challenge for CRC management. Bearing this in mind, we have used the machine learning package XGBoost to train our classifier system further to detect CRC and polyps separately. This resulted in establishing two independent but overlapping eight-marker signatures for a dual purpose (EpiSwitch® NST) based on 12 markers (Tables 2 and 5). The first eight-marker signature, when tested in a blind validation study of 251 samples, showed 90% sensitivity, 79% specificity, and 84% PPV in identifying all CRC (table 3) and 84% sensitivity, 79% specificity and 57% PPV in identifying patients, with stages 1 and 2 of CRC (table 4). This is a remarkable finding as most non-invasive tests, while having high specificity and NPV, have low sensitivity and PPV, resulting in both many missed diagnoses and false positive diagnoses, causing anguish and unnecessary procedures. Similarly, for polyp detection, a second eight-marker signature (Table 5) (EpiSwitch® NST) showed 79% sensitivity, 83% specificity and 60% PPV to detect the presence of adenomas/precancerous lesions/polyps. This is a remarkable accuracy compared to available analogues (Table 7). Thus, using a 12-marker set of EpiSwitch® NST could offer a consecutive two-step stratification of high accuracy (>80%) in detecting early stages of CRC and precancerous lesions/polyps for non-CRC cases.

Interestingly, the two developed eight-marker sets had four common CCs (supplementary table 2), likely reflecting the pathophysiological relationship between precancerous lesions and CRC. SGK223 (PRAG1) is involved in Notch signalling and epithelial mesenchymal transition (EMT) in CRC [44]. SMAD3 is a vital member of the TGFβ pathway (also a key player in EMT) and is deregulated in CRC [45].

The additional benefit of Whole-genome EpiSwitch® Explorer arrays is the ability to map the genomic positions of identified CCs, enabling the identification of the three closest protein-coding loci. Potential functional roles for these loci were obtained using Hallmark Gene Sets, BioCarta and Reactome canonical pathway analysis. Pathway analysis for the eight colorectal classifier markers showed several pathways involved in CRC, including TGFβ, cMYC, Rho GTPase, ROS and APC (supplementary table 3). A similar analysis of the eight-marker Polyp classifier showed TGFβ, PAX8, EMT, TNFα/NFκB, ROS, and APC (supplementary table 4). The pathway overlap is evident and likely represents a possible pathway of CRC progression, emphasising EMT and inflammation. Notably, the aberrations in these pathways are already present in precancerous lesions. This may significantly facilitate their detection, as at that stage, they rarely present with any symptoms or physical manifestations (such as bleeding due to extensive growth or neovascularisation) as CRC lesions do. Analysis of the top 3D genomic markers associated with CRC using the STRING database revealed the eight-marker CRC and polyp panels protein-protein interaction networks with hubs on CD58, SMAD3 and IL1R (Figure 3A and B). These three key signalling molecules have a well-documented association with CRC. Interestingly, they are all simultaneously involved in two fundamental mechanisms: EMT and inflammation (46-49). These identified pathways may suggest further

targeting therapies that can be used at various stages of the disease, including possible earlier stages in the adjuvant or neoadjuvant settings.

Conclusions

In this retrospective study using whole Genome DNA screening for CCs correlating to CRC and polyp diagnosis we have identified two Eight-marker CC signatures (Episwitch® NST) in whole blood that allow rapid and cost-effective diagnosis of CRC and precancerous polyps, respectively. Both diagnostic sets demonstrate high sensitivity and PPV which is particularly vital since existing blood and stool tests lack those attributes, especially in detection of pre-cancerous polyps and early-stage CRC. Genomic pathway analysis revealed signalling pathways related to the identified CCs, many of which have already been associated with CRC. Considering the significant signature/pathway overlap between polyps and CRC, these pathways are likely to play a key role in CRC pathophysiological progression and may suggest further targeted therapies in CRC management.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: title; Table S1: title; Video S1: title.

EH: CW: MI, SB, AG, AV, AD, RP, JG, TG and AA are full-time employees at Oxford BioDynamics plc and have no other competing financial or other interests. None of the remaining authors have competing interests.

Author Contributions: **EH:** TG: AA and DP conceived the study. RL, VV, KA, MH and GM supervised clinical samples collection, provided clinical insight in the cohort stratification and reviewed the manuscript. **CW:** MI, SB, AG, AV, AD, RP, JG, HSP, SPMG and CRL conducted clinical sample analysis. EH, HA, TG, AA and DP wrote and reviewed the manuscript.

Consent for publication: Written informed consent for publication was obtained from all authors.

Ethical consent and guidelines: All patients signed informed consent forms prior to providing blood samples. All ethical guidelines were followed.

Funding: This work was funded by Oxford BioDynamics plc.

Acknowledgments.: The authors would like to thank: All members (especially Serene Ng) of OBD Penang Reference Laboratory, Island 88 118 4 9, Jalan Tanjung Tokong, Penang 10470, Malaysia for help and support in preparing this manuscript. **Dr. Mohamad Fadli bin Abd Rahman** and **Dr. Beh Yuan Ye**, Hospital Pulau Pinang, Jalan Resideni, 10450 George Town, Pulau Pinang, who helped with patient recruitment and sample collection. **Dr. Muhammad Firdaus bin Md Salleh** and **Dr. Asha Gayatheri Balakrishnan** Hospital Sultanah Aminah, Jalan Persiaran Abu Bakar Sultan, 80100 Johor Bahru, Johor 16, who helped with patient recruitment and sample collection. **Dr. Fabian Lee Wei Luen:** Dr. Rakesh Raman and Dr. Ho Kean Fatt, Mount Miriam Cancer Hospital, 23, Jalan Bulan, Fettes Park, 11200 Tanjung Bungah, Pulau Pinang, who helped with patient recruitment and sample collection. **Dr. Siti Maisarah Md Ali:** Dr. Daniel Ching Lee Zhi, Dr. Kiew Kuang Kiat, Dr. Leow Voon Meng, Dr. Manisekar A/L K.Subramaniam, Dr. Mohd Azri Bin Mohd Suan, Dr. Mohd Razeen Bin Mohd Hassan, Dr. Muhd Yusairi Bin Kamarulzaman, Dr. Naveen A/L Ramasami, Dr. Ng Ying Zhuang, Dr. Nil Amri Bin Mohamed Kamil, Dr. R.Puthashanan A/L Rajamanickam, Dr. Thanamogan A/L Kerison @ Krishnan, Dr. Zalwani Zainuddin, Dr. Yeo Tian Ming, Hospital Sultanah Bahiyah, Km 6, Jalan Langgar, Bandar, 05460 Alor Setar, Kedah, who helped with patient recruitment and sample collection.

References

1. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, Vignat J, Ferlay J, Murphy N, and Abou-Kheir W. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut*. 2023 Feb;72(2):338-44. PubMed PMID: 36604116
2. Nguyen LH, Goel A, Chung DC. Pathways of Colorectal Carcinogenesis. *Gastroenterology*. 2020 Jan;158(2):291-302. PubMed PMID: 31622622. Pubmed Central PMCID: 6981255.
3. Sullivan BA, Lieberman DA. Colon Polyp Surveillance: Separating the Wheat From the Chaff. *Gastroenterology*. 2024 May;166(5):743-57. PubMed PMID: 38224860.
4. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. *N Engl J Med*. 2020 Aug 27;383(9):874-82. PubMed PMID: 32853499.
5. Fadlallah H, El Masri J, Fakhereddine H, Youssef J, Chemaly C, Doughan S, and Abou-Kheir W. Colorectal cancer: Recent advances in management and treatment. *World journal of clinical oncology*. 2024 Sep 24;15(9):1136-56. PubMed PMID: 39351451. Pubmed Central PMCID:.

6. Gharib E, Robichaud GA. From Crypts to Cancer: A Holistic Perspective on Colorectal Carcinogenesis and Therapeutic Strategies. *International journal of molecular sciences*. 2024 Aug 30;25(17). PubMed PMID: 39273409. Pubmed Central PMCID: 11395697.
7. Force USPST, Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling JW, Jr, arcía FAR, Gillman MW, Harper DM, Kemper AR, et al. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. *Jama*. 2016 Jun 21;315(23):2564-75. PubMed PMID: 27304597FHe J, Efron JE. Screening for colorectal cancer. *Advances in surgery*. 2011;45:31-44. PubMed PMID: 21954677.
8. He J, Efron JE. Screening for colorectal cancer. *Advances in surgery*. 2011;45:31-44. PubMed PMID: 21954677
9. Seager A, Sharp L, Neilson LJ, Brand A, Hampton JS, Lee TJW, Evans R, Vale L, Whelpton J, Bestwick N et al. Polyp detection with colonoscopy assisted by the GI Genius artificial intelligence endoscopy module compared with standard colonoscopy in routine colonoscopy practice (COLO-DETECT): a multicentre, open-label, parallel-arm, pragmatic randomised controlled trial. *The Lancet Gastroenterology & hepatology*. 2024 Oct;9(10):911-23. PubMed PMID:
10. Li SJ, Sharples LD, Benton SC, Blyuss O, Mathews C, Sasieni P, and Duffy SW. Faecal immunochemical testing in bowel cancer screening: Estimating outcomes for different diagnostic policies. *Journal of medical screening*. 2021 Sep;28(3):277-85. PubMed PMID: 33342370. Pubmed Central PMCID:.
11. Wieszczyni P, Kaminski MF, Loberg M, Bugajski M, Bretthauer M, Kalager M. Estimation of overdiagnosis in colorectal cancer screening with sigmoidoscopy and faecal occult blood testing: comparison of simulation models. *BMJ open*. 2021 Apr 14;11(4):e042158. PubMed PMID: 33853794. Pubmed Central PMCID: 8054108.
12. Fatemi N, Tierling S, Es HA, Varkiani M, Mojarad EN, Aghdaei HA, Walter J, and Totonchi M. DNA methylation biomarkers in colorectal cancer: Clinical applications for precision medicine. *Int J Cancer*. 2022 Dec 15;151(12):2068-81. PubMed PMID: 35730647.
13. Qin J, Wen B, Liang Y, Yu W, Li H. Histone Modifications and their Role in Colorectal Cancer (Review). *Pathology oncology research : POR*. 2020 Oct;26(4):2023-33. PubMed PMID: 31055775. Pubmed Central PMCID: 7471167.
14. Mukhopadhyay S, Ramadass AS, Akoulitchev A, Gordon S. Formation of distinct chromatin conformation signatures epigenetically regulate macrophage activation. *International immunopharmacology*. 2014 Jan;18(1):7-11. PubMed PMID: 24211766.
15. Seaman L, Chen H, Brown M, Wangsa D, Patterson G, Camps J, Omenn G.S., Ried T., and Rajapakse I. Nucleome Analysis Reveals Structure-Function Relationships for Colon Cancer. *Molecular cancer research : MCR*. 2017 Jul;15(7):821-30. PubMed PMID: 28258094. Pubmed Central PMCID: 7934948.
16. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002 Feb 15;295(5558):1306-11. PubMed PMID: 11847345.
17. Bastonini E, Jeznach M, Field M, Juszczak K, Corfield E, Dezfouli M, Ahmat N, Smith A, Womersley H, Jordan P, et al. Chromatin barcodes as biomarkers for melanoma. *Pigment cell & melanoma research*. 2014 Sep;27(5):788-800. PubMed PMID: 24807349.
18. Jakub JW, Grotz TE, Jordan P, Hunter E, Pittelkow M, Ramadass A, Akoulitchev A, and Markovic S. A pilot study of chromosomal aberrations and epigenetic changes in peripheral blood samples to identify patients with melanoma. *Melanoma research*. 2015 Oct;25(5):406-11. PubMed PMID: 26225582.
19. Alshaker H, Mills R, Hunter E, Salter M, Ramadass A, Skinner BM, Westra W, Green J, Akoulitchev A, Winkler M et al. Chromatin conformation changes in peripheral blood can detect prostate cancer and stratify disease risk groups. *Journal of translational medicine*. 2021 Jan 28;19(1):46. PubMed PMID: 33509203. Pubmed Central PMCID: 7845038.
20. Alshaker H, Hunter E, Salter M, Ramadass A, Westra W, Winkler M, Green J, Akoulitchev A, and Pchejetski D. Monocytes acquire prostate cancer-specific chromatin conformations upon indirect co-culture with prostate cancer cells. *Front Oncol*. 2022;12:990842. PubMed PMID: 36059613. Pubmed Central PMCID: 9437316.
21. Pchejetski D, Hunter E, Dezfouli M, Salter M, Powell R, Green J, Naithani T, Koutsothanasi C, Alshaker H, Jaipuria J, et al. Circulating Chromosome Conformation Signatures Significantly Enhance PSA Positive Predicting Value and Overall Accuracy for Prostate Cancer Detection. *Cancers*. 2023 Jan 29;15(3). PubMed PMID: 36765779. Pubmed Central PMCID: 9913359.
22. Hunter E, Salter M, Powell R, Dring A, Naithani T, Chatziioannou ME, Gebregzabhar A, Issa M, Green J, Ng S et al. Development and Validation of Blood-Based Predictive Biomarkers for Response to PD-1/PD-L1 Checkpoint Inhibitors: Evidence of a Universal Systemic Core of 3D Immunogenetic Profiling across Multiple Oncological Indications. *Cancers*. 2023 May 10;15(10). PubMed PMID: 37345033. Pubmed Central PMCID: 10216232.
23. Salter M, Corfield E, Ramadass A, Grand F, Green J, Westra J, Lim CR, Farrimond L, Feneberg E, Scaber J, et al. Initial Identification of a Blood-Based Chromosome Conformation Signature for Aiding in the

- Diagnosis of Amyotrophic Lateral Sclerosis. *EBioMedicine*. 2018 Jul;33:169-84. PubMed PMID: 29941342. Pubmed Central PMCID: 6085506.
24. Carini C, Hunter E, Scottish Early Rheumatoid Arthritis Inception cohort I, Ramadass AS, Green J, Akoulitchev A, et al. Chromosome conformation signatures define predictive markers of inadequate response to methotrexate in early rheumatoid arthritis. *Journal of translational medicine*. 2018 Jan 29;16(1):18. PubMed PMID: 29378619. Pubmed Central PMCID: 5789697.
 25. Hunter E, McCord R, Ramadass AS, Green J, Westra JW, Mundt K, and Akoulitchev A. Comparative molecular cell-of-origin classification of diffuse large B-cell lymphoma based on liquid and tissue biopsies. *Translational Medicine Communications*. 2020 2020/03/24;5(1):5.
 26. Yan H, Hunter E, Akoulitchev A, Park P, Winchester DJ, Moo-Young TA, and Prinz RA. Epigenetic chromatin conformation changes in peripheral blood can detect thyroid cancer. *Surgery*. 2019 Jan;165(1):44-9. PubMed PMID: 30377001.
 27. Hunter E, Koutsothanasi C, Wilson A, Santos FC, Salter M, Powell R, et al. 3D genomic capture of regulatory immuno-genetic profiles in COVID-19 patients for prognosis of severe COVID disease outcome. *bioRxiv*. 2021:2021.03.14.435295.
 28. Tsourkas A, Behlke MA, Xu Y, Bao G. Spectroscopic features of dual fluorescence/luminescence resonance energy-transfer molecular beacons. *Anal Chem*. 2003 Aug 1;75(15):3697-703. PubMed PMID: 14572032.
 29. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2005 Oct 25;102(43):15545-50. PubMed PMID: 16199517. Pubmed Central PMCID: PMC1239896. Epub 20050930.
 30. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017 Jan 4;45(D1):D362-D8. PubMed PMID: 27924014. Pubmed Central PMCID: PMC5210637. Epub 20161018.
 31. Jain S, Maque J, Galoosian A, Osuna-Garcia A, May FP. Optimal Strategies for Colorectal Cancer Screening. *Current treatment options in oncology*. 2022 Apr;23(4):474-93. PubMed PMID: 35316477. Pubmed Central PMCID: 8989803.
 32. Knudsen AB, Zauber AG, Rutter CM, Naber SK, Doria-Rose VP, Pabiniak C, et al. Estimation of Benefits, Burden, and Harms of Colorectal Cancer Screening Strategies: Modeling Study for the US Preventive Services Task Force. *Jama*. 2016 Jun 21;315(23):2595-609. PubMed PMID: 27305518. Pubmed Central PMCID: 5493310.
 33. Lin JS, Perdue LA, Henrikson NB, Bean SI, Blasi PR. Screening for Colorectal Cancer: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *Jama*. 2021 May 18;325(19):1978-98. PubMed PMID: 34003220.
 34. Imperiale TF, Porter K, Zella J, Gagrut ZD, Olson MC, Statz S, et al. Next-Generation Multitarget Stool DNA Test for Colorectal Cancer Screening. *N Engl J Med*. 2024 Mar 14;390(11):984-93. PubMed PMID: 38477986.
 35. Chung DC, Gray DM, 2nd, Singh H, Issaka RB, Raymond VM, Eagle C, et al. A Cell-free DNA Blood-Based Test for Colorectal Cancer Screening. *N Engl J Med*. 2024 Mar 14;390(11):973-83. PubMed PMID: 38477985.
 36. Dupont C, Armant DR, Brenner CA. Epigenetics: definition, mechanisms and clinical perspective. *Seminars in reproductive medicine*. 2009 Sep;27(5):351-7. PubMed PMID: 19711245. Pubmed Central PMCID: 2791696.
 37. Feinberg AP. The Key Role of Epigenetics in Human Disease Prevention and Mitigation. *N Engl J Med*. 2018 Apr 5;378(14):1323-34. PubMed PMID: 29617578.
 38. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. *Science*. 2013 Mar 29;339(6127):1546-58. PubMed PMID: 23539594. Pubmed Central PMCID: 3749880.
 39. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nature reviews Genetics*. 2009 Apr;10(4):241-51. PubMed PMID: 19293820.
 40. Yan W, Guo M. Epigenetics of colorectal cancer. *Methods Mol Biol*. 2015;1238:405-24. PubMed PMID: 25421673.
 41. Beckedorff FC, Amaral MS, Deocesano-Pereira C, Verjovski-Almeida S. Long non-coding RNAs and their implications in cancer epigenetics. *Bioscience reports*. 2013;33(4). PubMed PMID: 23875687. Pubmed Central PMCID: 3759304. Epub 2013/07/24. eng.
 42. Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature*. 2013 Feb 28;494(7438):497-501. PubMed PMID: 23417068. Pubmed Central PMCID: PMC4109059.
 43. Kosaka N, Ochiya T. Unraveling the Mystery of Cancer by Secretory microRNA: Horizontal microRNA Transfer between Living Cells. *Frontiers in genetics*. 2011;2:97. PubMed PMID: 22303391. Pubmed Central PMCID: 3262223. Epub 2012/02/04. eng.
 44. Vinson KE, George DC, Fender AW, Bertrand FE, Sigounas G. The Notch pathway in colorectal cancer. *Int J Cancer*. 2016 Apr 15;138(8):1835-42. PubMed PMID: 26264352.

45. Fleming NI, Jorissen RN, Mouradov D, Christie M, Sakthianandeswaren A, Palmieri M, et al. SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer. *Cancer Res.* 2013 Jan 15;73(2):725-35. PubMed PMID: 23139211.
46. Xu S, Wen Z, Jiang Q, Zhu L, Feng S, Zhao Y, et al. CD58, a novel surface marker, promotes self-renewal of tumor-initiating cells in colorectal cancer. *Oncogene.* 2015 Mar 19;34(12):1520-31. PubMed PMID: 24727892.
47. Ding N, Luo H, Zhang T, Peng T, Yao Y, He Y. Correlation between SMADs and Colorectal Cancer Expression, Prognosis, and Immune Infiltrates. *International journal of analytical chemistry.* 2023;2023:8414040. PubMed PMID: 36969909. Pubmed Central PMCID: 10038740.
48. Wu B, Zhan X, Jiang M. CD58 defines regulatory macrophages within the tumor microenvironment. *Communications biology.* 2024 Aug 21;7(1):1025. PubMed PMID: 39164573. Pubmed Central PMCID: 11335740.
49. Koncina E, Nurmik M, Pozdeev VI, Gilson C, Tsenkova M, Begaj R, et al. IL1R1(+) cancer-associated fibroblasts drive tumor development and immunosuppression in colorectal cancer. *Nature communications.* 2023 Jul 17;14(1):4251. PubMed PMID: 37460545. Pubmed Central PMCID: 10352362.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.