





Article

Ultra-Low DNA Input into Whole Genome Methylation Assays and Detection of Oncogenic Methylation and Copy Number Variants in Circulating Tumour DNA

Celina Whalley¹, Karl Payne¹, Enric Domingo², Andrew Blake², Susan Richman³, Jill Brooks¹, Nikolaos Batis¹, Rachel Spruce¹, S-CORT Consortium[†], Hisham Mehanna¹, Paul Nankivell^{1,‡} and Andrew D. Beggs^{1,*}

¹ Institute of Cancer & Genomic Sciences, University of Birmingham, Vincent Drive, Birmingham B15 2TT, UK; c.m.whalley@bham.ac.uk (C.W.); k.payne.1@bham.ac.uk (K.P.); j.m.brooks@bham.ac.uk (J.B.); N.Batis@bham.ac.uk (N.B.); R.Spruce@bham.ac.uk (R.S.); h.mehanna@bham.ac.uk (H.M.); p.c.nankivell@bham.ac.uk (P.N.)

² Department of Oncology, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7BN, UK; enric.domingo@oncology.ox.ac.uk (E.D.); andrew.blake@oncology.ox.ac.uk (A.B.)

³ Pathology and Data Analytics, Leeds Institute of Medical Research, St James University Hospital, Leeds LS2 9JT, UK; s.d.richman@leeds.ac.uk

* Correspondence: a.beggs@bham.ac.uk

† The S-CORT Consortium members are listed in the Acknowledgments.

‡ These authors contributed equally to this work.



Citation: Whalley, C.; Payne, K.; Domingo, E.; Blake, A.; Richman, S.; Brooks, J.; Batis, N.; Spruce, R.; S-CORT Consortium; Mehanna, H.; et al. Ultra-Low DNA Input into Whole Genome Methylation Assays and Detection of Oncogenic Methylation and Copy Number Variants in Circulating Tumour DNA. *Epigenomes* **2021**, *5*, 6. <https://doi.org/10.3390/epigenomes5010006>

Academic Editor: Ernesto Guccione

Received: 30 December 2020

Accepted: 12 February 2021

Published: 19 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: Abnormal CpG methylation in cancer is ubiquitous and generally detected in tumour specimens using a variety of techniques at a resolution encompassing single CpG loci to genome wide coverage. Analysis of samples with very low DNA inputs, such as formalin fixed (FFPE) biopsy specimens from clinical trials or circulating tumour DNA is challenging at the genome-wide level because of lack of available input. We present the results of low input experiments into the Illumina Infinium HD methylation assay on FFPE specimens and ctDNA samples. Methods: For all experiments, the Infinium HD assay for methylation was used. In total, forty-eight FFPE specimens were used at varying concentrations (lowest input 50 ng); eighteen blood derived specimens (lowest input 10 ng) and six matched ctDNA input (lowest input 10 ng)/ fresh tumour specimens (lowest input 250 ng) were processed. Downstream analysis was performed in R/Bioconductor for quality control metrics and differential methylation analysis as well as copy number calls. Results: Correlation coefficients for CpG methylation were high at the probe level averaged $R^2 = 0.99$ for blood derived samples and $R^2 > 0.96$ for the FFPE samples. When matched ctDNA/fresh tumour samples were compared, $R^2 > 0.91$ between the two. Results of differential methylation analysis did not vary significantly by DNA input in either the blood or FFPE groups. There were differences seen in the ctDNA group as compared to their paired tumour sample, possibly because of enrichment for tumour material without contaminating normal. Copy number variants observed in the tumour were generally also seen in the paired ctDNA sample with good concordance via DQ plot. Conclusions: The Illumina Infinium HD methylation assay can robustly detect methylation across a range of sample types, including ctDNA, down to an input of 10 ng. It can also reliably detect oncogenic methylation changes and copy number variants in ctDNA. These findings demonstrate that these samples can now be accessed by methylation array technology, allowing analysis of these important sample types.

Keywords: circulating tumour DNA; formalin fixed paraffin embedded; epigenome

1. Introduction

Abnormal epigenetic change in the form of the loss or gain of normal 5-methylcytosine patterns is one of the most ubiquitous epigenetic changes in cancer epigenetics, controlling gene expression via changes in transcription factor binding [1].

Methylation in cancer is of particular interest due to its association with specific phenotypes such as the CpG Island Methylator Phenotype (CIMP) that possess unique characteristics [2]. In colorectal cancer, CIMP is associated with a hypermethylator phenotype [3], usually found in the right colon in female patients with advanced age, can be associated with microsatellite instability (due to DNA mismatch repair suppression), and has been suggested to be a biomarker for poor prognosis [4]. In fact, the epigenome has been shown to vary widely across different cancer types, and multiple methylation subtypes have been identified. The TCGA consortium identified four methylator subtypes associated with colorectal cancer, each associated with differing clinico-pathological characteristics [5]. In head and neck squamous cell carcinoma (HNSCC), analysis of TCGA tumour methylation data has revealed both hyper- and hypomethylated gene sets for diagnostic purposes [6]. In addition, six methylation clusters have been identified that correlate with clinicopathological characteristics such as HPV status and also three-year survival [7].

A particular focus of interest in study of the epigenome is in circulating tumour DNA [8] (ctDNA). This has attracted particular attention as a biomarker for disease, with the GRAIL consortium utilising a mutation based assay for circulating tumour DNA with over 90% accuracy for the detection of cancer [9]. However, the use of methylation assays on this sample type has been restricted due to low input concentrations available from ctDNA. The UroMark consortium [10] has studied the use of bisulphite amplicon resequencing in the analysis of free urinary DNA. The DNA input required for this is relatively plentiful (>30 ng of urinary DNA) compared to that available for circulating tumour DNA (typically 10–20 ng per 10 mls of whole blood).

In order to develop new biomarker panels for the methylation based detection of disease, a whole genome approach is required. Several technologies exist for this purpose—whole genome bisulphite sequencing [11] (WGBS), reduced representation bisulphite sequencing [12] (RBBS), and methylation arrays, typically the Illumina Infinium assay [13]. Both WGBS and RBBS require large input amounts of DNA (~3 µg) and are particularly expensive, the former especially so due to the increased depth of coverage required necessitating increased sequencing cost. Methylation arrays, the most popular of which is the Illumina Infinium assay, require at a minimum 200–300 ng of bisulphite converted DNA, which can be from formalin fixed, paraffin embedded (FFPE) archival samples, repaired by use of a proprietary “DNA restore” kit. The input requirements, as specified by the manufacturer, make this assay seemingly unusable for the requirements of low input and ctDNA experiments.

We hypothesised that as the Illumina Infinium assay utilises a whole genome [14] amplification (WGA) step, the theoretical input to this kit could encompass concentrations seen in ctDNA samples. We therefore aimed to study the effect of input concentration of DNA on assay performance.

2. Results

2.1. Sample QC Metrics

Sample workflows are shown in Supplementary Figure S1. All samples across both blood, FFPE, and fresh tissue passed QC metrics for their respective samples types. For samples C, 1 and 3 all had >99% CpG detection across the range of DNA input (10–500 ng). QC results are shown in Tables 1 and 2. For FFPE samples from the SCORT consortium ($n = 16$, across 50, 100, and 150 ng input mass), all samples passed QC with >95% CpG detection, reflecting the lower detection threshold provided by Illumina for lower quality samples. A Bland–Altman plot from 10 randomly selected CpG probes comparing fresh tissue and ctDNA in order to understand reproducibility was performed (Supplementary Figure S2) showing good correlation (Bias = -0.0074 , 95% CI = -0.07 – 0.06). The normalisa-

tion graph demonstrated an unusual bump in the beta values around 0.5 but were all from the same blood sample donor at varying input amounts, suggesting a sample specific issue. When investigated, it was found that this donor has recently completed chemotherapy, which may explain the unusual probe distribution.

Table 1. Shows CpG % detection levels from the fresh blood DNA samples processed on the Infinium® HumanMethylation 450 k BeadChip. There are 485,577 CpG sites on this array with the QC cut off for % CpG detection, at $\geq 99\%$.

Sample ID	DNA Input (ng)	Detected CpG (0.05)	% CpG Detection
C-10	10	483,570	99.59
C-50	50	485,395	99.96
C-100	100	485,475	99.98
C-150	150	485,466	99.98
C-200	200	485,473	99.98
C-250	250	485,471	99.98
C-300	300	485,455	99.97
C-500	500	485,446	99.97
3-2_10	10	484,491	99.78
3-2_50	50	485,417	99.97
3-2_100	100	485,406	99.96
3-2_200	200	485,435	99.97
3-2_250	250	485,445	99.97
1-1_10	10	484,158	99.71
1-1_50	50	485,346	99.95
1-1_100	100	485,446	99.97
1-1_200	200	485,458	99.98
1-1_250	250	485,485	99.98
199-ctDNA	10	457,792	94.28%
207-ctDNA	10	467,045	96.18%
264-ctDNA	10	480,752	99.01%
268-ctDNA	10	471,080	97.01%
276-ctDNA	10	471,041	97.01%
288-ctDNA	10	451,610	93.00%
199-tissue	250	485,046	99.89%
207-tissue	250	484,988	99.88%
264-tissue	250	484,727	99.82%
268-tissue	250	484,983	99.88%
276-tissue	250	484,442	99.77%
288-tissue	250	484,962	99.87%

Table 2. Shows qPCR quality and CpG % detection levels per sample for each FFPE DNA input amount. There are 866,895 CpG sites on the Infinium[®] MethylationEPIC BeadChip arrays with a QC cut off of $\geq 90\%$ detection with FFPE DNA.

Sample ID	qPCR Delta Cq Value < 5 (>5 Poor)	% CpG Detection ($p = 0.05$) 150 ng Input	% CpG Detection ($p = 0.05$) 100 ng Input	% CpG Detection ($p = 0.05$) 50 ng Input
SC00236A2	2.34	96.55	97.57	96.69
SC00264A2	1.59	97.95	98.73	97.93
SC00267A2	2.95	98.03	98.40	97.50
SC00273A2	3.10	97.99	98.22	97.66
SC00275A2	2.72	97.60	98.09	97.10
SC00292A2	3.61	96.99	97.97	96.30
SC00296A2	2.08	99.22	99.27	98.01
SC00357A2	0.89	99.40	99.44	99.22
SC00447A2	0.71	98.66	99.04	98.38
SC00476A2	1.87	99.18	99.17	99.19
SC00482A2	1.09	98.95	98.85	98.49
SC00486A2	1.97	98.85	98.94	98.63
SC00491A2	1.83	98.28	98.29	98.10
SC00496A2	3.05	93.07	95.28	94.96
SC00498A2	1.83	98.86	98.98	98.54
SC00529A2	1.08	99.14	99.20	98.75

2.2. Sample Correlations

In order to understand whether there was a relationship between reducing input and loss of correlation between identical samples, a correlation analysis was run using the *cor* command in the *psych* module of *R*.

For samples C, 1, and 3, all correlation coefficients were >0.99 across all comparators (Pearsons $p < 0.001$, Table 3). For the FFPE derived samples, the correlation coefficient was at least 96% in all samples. Correlation matrices and distributions for all sample types are shown in Figure 1. Correlation between sample inputs dispersed more widely as the input amount, with a particular widening around 10 ng input across the blood derived samples. In the FFPE sample group, these were more widely dispersed generally; however, less of an effect in terms of dispersion at low input concentrations was seen. For the ctDNA samples, correlation was performed between all the samples, and lower correlations (typically <0.95) and very wide dispersions were seen, as the point of this plot was to demonstrate that they were not similar.

Table 3. Correlation coefficients between all samples.

Comparison	Correlation Coefficient	Lower 95% CI	Upper 95% CI	p -Value
C-10-C-50	0.99	0.99	0.99	<0.01
C-10-C-100	0.99	0.99	0.99	<0.01
C-10-C-150	0.99	0.99	0.99	<0.01
C-10-C-200	0.99	0.99	0.99	<0.01
C-10-C-250	0.99	0.99	0.99	<0.01
C-10-C-300	0.99	0.99	0.99	<0.01
C-10-C-500	0.99	0.99	0.99	<0.01

Table 3. Cont.

Comparison	Correlation Coefficient	Lower 95% CI	Upper 95% CI	p-Value
C-50-C-100	1.00	1.00	1.00	<0.01
C-50-C-150	1.00	1.00	1.00	<0.01
C-50-C-200	1.00	1.00	1.00	<0.01
C-50-C-250	1.00	1.00	1.00	<0.01
C-50-C-300	1.00	1.00	1.00	<0.01
C-50-C-500	1.00	1.00	1.00	<0.01
C-100-C-150	1.00	1.00	1.00	<0.01
C-100-C-200	1.00	1.00	1.00	<0.01
C-100-C-250	1.00	1.00	1.00	<0.01
C-100-C-300	1.00	1.00	1.00	<0.01
C-100-C-500	1.00	1.00	1.00	<0.01
C-150-C-200	1.00	1.00	1.00	<0.01
C-150-C-250	1.00	1.00	1.00	<0.01
C-150-C-300	1.00	1.00	1.00	<0.01
C-150-C-500	1.00	1.00	1.00	<0.01
C-200-C-250	1.00	1.00	1.00	<0.01
C-200-C-300	1.00	1.00	1.00	<0.01
C-200-C-500	1.00	1.00	1.00	<0.01
C-250-C-300	1.00	1.00	1.00	<0.01
C-250-C-500	1.00	1.00	1.00	<0.01
C-300-C-500	1.00	1.00	1.00	<0.01
1-1_25-1-1_20	1.00	1.00	1.00	<0.01
1-1_25-1-1_100	1.00	1.00	1.00	<0.01
1-1_25-1-1_5	0.99	0.99	0.99	<0.01
1-1_25-1-1_10	0.98	0.98	0.98	<0.01
1-1_20-1-1_100	1.00	1.00	1.00	<0.01
1-1_20-1-1_5	0.99	0.99	0.99	<0.01
1-1_20-1-1_10	0.98	0.98	0.98	<0.01
1-1_100-1-1_5	0.99	0.99	0.99	<0.01
1-1_100-1-1_10	0.99	0.99	0.99	<0.01
1-1_5-1-1_10	0.99	0.99	0.99	<0.01
3-2_25-3-2_20	1.00	1.00	1.00	<0.01
3-2_25-3-2_100	1.00	1.00	1.00	<0.01
3-2_25-3-2_5	1.00	1.00	1.00	<0.01
3-2_25-3-2_10	0.97	0.97	0.97	<0.01
3-2_20-3-2_100	1.00	1.00	1.00	<0.01
3-2_20-3-2_5	1.00	1.00	1.00	<0.01
3-2_20-3-2_10	0.97	0.97	0.97	<0.01
3-2_100-3-2_5	1.00	1.00	1.00	<0.01
3-2_100-3-2_10	0.97	0.97	0.97	<0.01
3-2_5-3-2_10	0.97	0.98	0.98	<0.01

Table 3. Cont.

Comparison	Correlation Coefficient	Lower 95% CI	Upper 95% CI	p-Value
FFPE-50-FFPE-10	0.96	0.96	0.96	<0.01
FFPE-50-FFPE-15	0.96	0.96	0.96	<0.01
FFPE-50-FFPE-20	0.96	0.96	0.96	<0.01
FFPE-50-FFPE-250	0.96	0.96	0.96	<0.01
FFPE-50-FFPE-300	0.96	0.96	0.96	<0.01
FFPE-50-FFPE-500	0.96	0.96	0.96	<0.01
FFPE-10-FFPE-15	0.97	0.97	0.97	<0.01
FFPE-10-FFPE-20	0.97	0.97	0.97	<0.01
FFPE-10-FFPE-25	0.97	0.97	0.97	<0.01
FFPE-10-FFPE-3	0.97	0.97	0.97	<0.01
FFPE-10-FFPE-500	0.97	0.97	0.97	<0.01
FFPE-15-FFPE-20	0.98	0.98	0.98	<0.01
FFPE-15-FFPE-25	0.98	0.98	0.98	<0.01
FFPE-15-FFPE-3	0.98	0.98	0.98	<0.01
FFPE-15-FFPE-500	0.97	0.97	0.97	<0.01
FFPE-20-FFPE-25	0.98	0.98	0.98	<0.01
FFPE-20-FFPE-3	0.98	0.98	0.98	<0.01
FFPE-20-FFPE-500	0.97	0.97	0.97	<0.01
FFPE-25-FFPE-3	0.98	0.98	0.98	<0.01
FFPE-25-FFPE-500	0.97	0.98	0.98	<0.01
FFPE-3-FFPE-500	0.98	0.98	0.98	<0.01
FFPE-15-FFPE-20	0.98	0.98	0.98	<0.01
199-ctDNA-207-ctDNA	0.93	0.93	0.93	<0.01
199-ctDNA-264-ctDNA	0.93	0.93	0.94	<0.01
199-ctDNA-268-ctDNA	0.93	0.93	0.93	<0.01
199-ctDNA-276-ctDNA	0.94	0.94	0.94	<0.01
199-ctDNA-288-ctDNA	0.92	0.92	0.92	<0.01
207-ctDNA-264-ctDNA	0.94	0.94	0.95	<0.01
207-ctDNA-268-ctDNA	0.94	0.94	0.94	<0.01
207-ctDNA-276-ctDNA	0.94	0.94	0.95	<0.01
207-ctDNA-288-ctDNA	0.92	0.92	0.92	<0.01
264-ctDNA-268-ctDNA	0.95	0.95	0.95	<0.01
264-ctDNA-276-ctDNA	0.96	0.96	0.96	<0.01
264-ctDNA-288-ctDNA	0.93	0.93	0.93	<0.01
268-ctDNA-276-ctDNA	0.94	0.94	0.95	<0.01
268-ctDNA-288-ctDNA	0.92	0.92	0.92	<0.01
276-ctDNA-288-ctDNA	0.93	0.93	0.93	<0.01

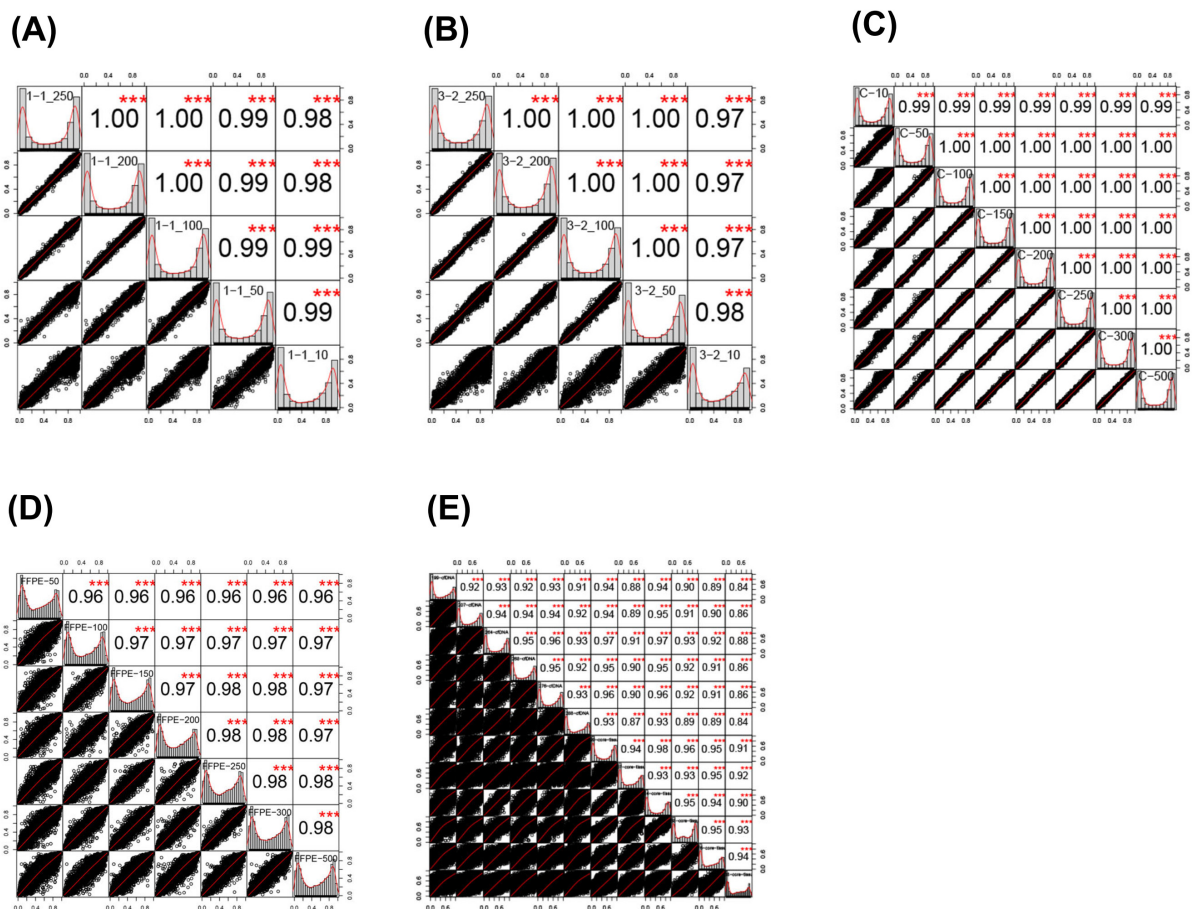


Figure 1. Pearson correlation plots of percentage methylation between probes in (A) Sample 1 (1-1_250 = 250 ng input, 1-2_200 = 200 ng, 1-1_100 = 100 ng, 1-1_50 = 50 ng input, 1-1_10 = 10 ng input); (B) Sample 3 (3-2_250 = 250 ng, 3-2_200 = 200 ng, 3-2_100 = 100 ng, 3-2_50 = 50 ng, 3-2_10 = 10 ng); (C) Sample C (C-10 = 10 ng input, C-50 = 50 ng, C-100 = 100 ng, C-150 = 150 ng, C-200 = 200 ng, C-250 = 250 ng, C-300 = 300 ng, C-500 = 500 ng); (D) FFPE samples (FFPE-50 = 50 ng input, FFPE-100 = 100 ng, FFPE-150 = 150 ng, FFPE-200 = 200 ng, FFPE-250 = 250 ng, FFPE-300 = 300 ng, FFPE-500 = 500 ng); (E) ctDNA samples vs. core tissue samples (ctDNA and core tissue samples labelled); Sample correlations are shown within box with labels down diagonal. Red stars represent statistical significance (t -test, Bonferroni correction (***) $p < 0.001$). Black dot plots represent correlation between two samples. Each box within the plot represents a single sample, and the axes within each box represent the correlations between the sample above and the sample to the right.

In order to understand the differences between ctDNA, tissue samples, and (unmatched) blood samples with leucocyte derived DNA, we carried out a three way analysis (Supplementary Figure S3). Both hierarchical clustering and a multi-dimensional scaling plot demonstrated clear separation of the plots, meaning that the observed differences between the ctDNA and tumour were not functions of the differences in methylation profiles seen between leucocytes and tumour cells.

We also carried out a genomewide plot (Supplementary Figures S4 and S5) of all samples in order to understand the differences and reproducibility in methylation across sample types. Unsurprisingly, in the blood derived DNA, there was a bimodal distribution of methylation in keeping with no methylation or methylation of both alleles. Both tumour and ctDNA samples showed a bimodal distribution reflecting the heterogeneity in this sample type.

2.3. Effect of Input Amount on Results

In order to understand whether reducing the sample amount impacted on results obtained through standard analysis, a differentially methylated position analysis was

performed with the DMP analysis of R/ChAMP/Bioconductor with each sample randomly assigned to one of two groups, and each sample group analysed separately.

In the fresh blood derived sample group, for sample “1”, the DMP analysis demonstrated no significant differences between any of the groups comparing between each concentration, with the highest ranking probe cg17735539 having a d-Beta-meth of -0.12 an adjusted p -value of 0.99 and a Bayes Factor (B) of 1.96. For sample “3”, the DMP analysis demonstrated no significant differences between any of groups comparing between each concentration, with the highest ranking probe cg14640149 having a d-Beta-meth of -0.11 , an adjusted p -value of 0.99, and a B of 0.41. For sample “C”, the DMP analysis demonstrated no significant differences between groups, with the highest ranking probe cg18253591 having a d-Beta-meth of 0.04, an adjusted p -value of 0.43, and a B of 5.61.

In the FFPE sample group, for comparisons between groups, there was no significant difference with the highest ranking probe cg06871764 having a d-Beta-meth of 0.18, an adjusted p -value of 0.99, and a B of 2.77.

2.4. ctDNA Analysis

In order to understand whether the observed low inputs would be suitable for detection of very fragmented, low input (typically 1–2 ng) DNA, an experiment was carried out on six samples of circulating tumour DNA matched to their primary head and neck tumours of origin, sourced from fresh tissue.

All samples hybridised successfully to the arrays with very low input amounts when restored using the FFPE restore kit. One sample (288-ctDNA) had only 93% of probes that passed QC, possibly due to the increased fragmentation inherent to ctDNA. Median correlation between the tumour of origin and what was observed in the ctDNA samples was 0.92 (max 0.98, min 0.84). In the sample the correlation was observed to be 0.84 (sample 288); this may be due to the poor quality of input sample.

In order to understand whether the observed differences were because of enrichment for “tumour” in the ctDNA samples over the contaminating normal stroma in the “tumour” samples, which would exaggerate methylation differences; rather than contaminating genomic DNA, a standard methylation pipeline analysis using ChAMP was carried out. This was because if the ctDNA was mainly genomic DNA, it was expected that methylated regions typically seen in leucocytes (which are the predominant contributor to contaminating ctDNA) would be observed rather than those typical for HNSCC.

A differential methylation analysis revealed 19,014 differentially methylated positions between the control tumour samples and the circulating tumour DNA samples. The top ranked probe cg09568464 (chr19: 56904901) tagged a region within 200 bp of the transcription start site of ZNF582 ($\Delta\text{Beta} = 0.40$, $\text{methcontrol} = 0.42$, $\text{methctDNA} = 0.02$, $\text{adj } p = 1.72 \times 10^{-4}$, $B = 11.6$). G:Profiler pathway analysis of the differentially methylated positions revealed the top enriched pathway was KEGG:04072 (Phospholipase-D pathway, $\text{adj } p \text{ value} = 6.21 \times 10^{-5}$) Several other tumourigenic pathways of interest were also enriched including RAP1 signalling ($\text{padj} = 1.14 \times 10^{-4}$), wnt signalling ($\text{padj} = 6.12 \times 10^{-4}$), HPV infection ($\text{padj} = 9.56 \times 10^{-3}$), and PIK3-Akt signalling ($\text{padj} = 0.02$).

Analysis of differentially methylated regions revealed 410 DMRs with adjusted p values < 0.05 . The top DMR was at chr6:29520965–29521803 ($p < 0.05$) and tagged a region just upstream of UBD (ubiquitin-D), a key regulator of NF- κ B. Pathway analysis utilising G:Profiler demonstrated two Kegg pathways that were enriched: KEGG:04080 (neuroactive ligand receptor interaction, $\text{padj} = 6.73 \times 10^{-4}$) and KEGG:04550 (signalling pathways regulating pluripotency of stem cells, $\text{padj} = 7.46 \times 10^{-3}$).

2.5. DNA Copy Number Calls

In order to ascertain whether copy number variation could be detected using this methodology in circulating tumour DNA, the copy prediction algorithm of conumee was used, as per the vignette.

Generally, there was good agreement between copy number predictions for the fresh sample and the ctDNA sample, with multiple copy number changes detected in both. There was dispersion of the copy number plot in the fresh samples, most likely due to contaminating normal and the presence of multiple clones within the tumour sample. An example plot is shown in Figure 2, where the copy numbers between a fresh sample (top) and ctDNA derived sample from the sample patient (bottom) are compared. Copy number gains in *CDKN2A* and *NOTCH1* are both observed in both samples, with high copy number seen in the ctDNA as compared to the fresh sample. A similar pattern was seen in the FFPE derived samples (Figure 3), where increasing dispersion of copy number was seen as input DNA mass decreased. In matched ctDNA and fresh tumour samples, plotting of residual copy number segmentation values (DQ plot, Supplementary Figure S5) demonstrated reasonably good concordance between copy number values but with dispersion towards the extremes of value, but no more than ± 0.7 maximally. The TCGA HNSCC dataset [15] demonstrated that loss of 3p and 8p, as well as gain of 3q, 5p, and 8q, were seen frequently in HNSCC, and copy number changes were seen in all these regions in the ctDNA samples.

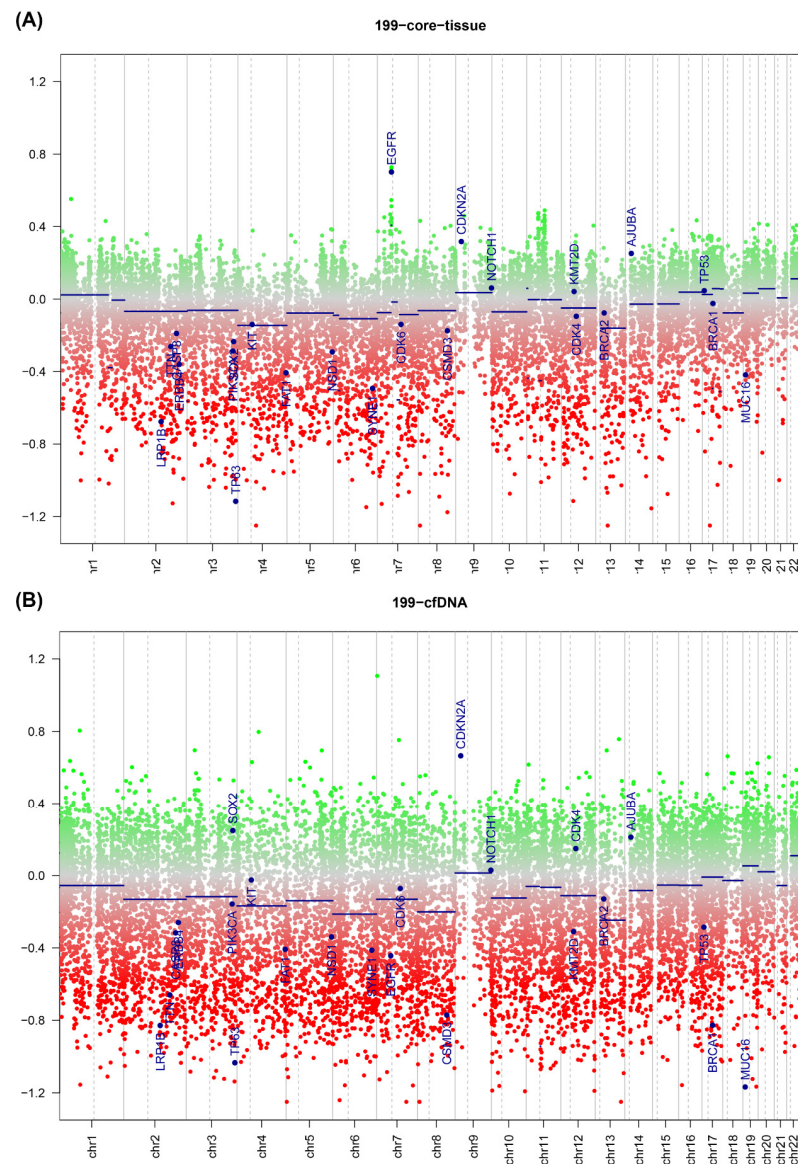


Figure 2. Copy number plot of reference fresh tissue **(A)** and ctDNA profile **(B)** from head and neck cancer derived samples. Green dots = copy number gain; red dots = copy number loss. Blue lines represent normalised copy number across segment represented by line.

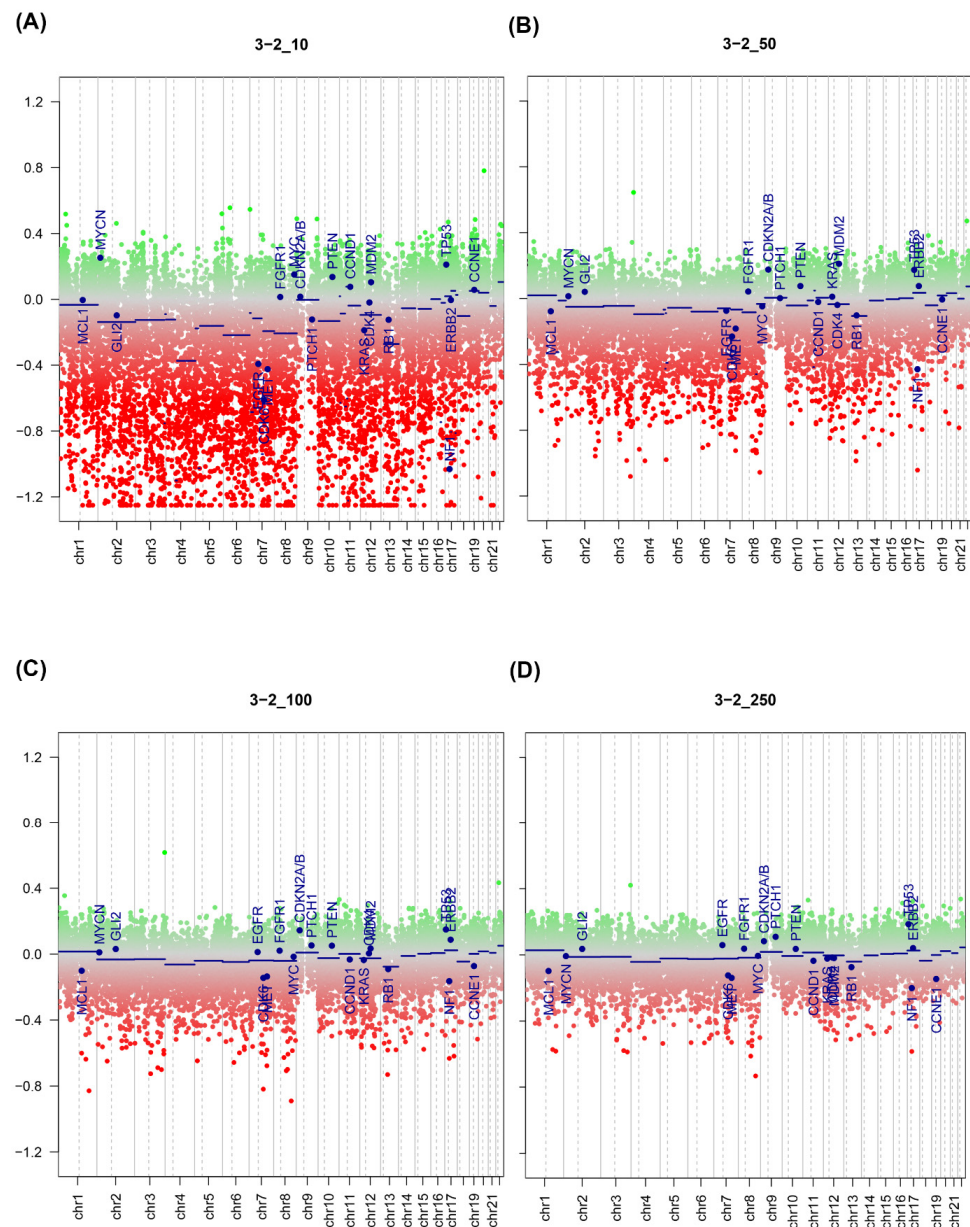


Figure 3. Copy number plots of a representative FFPE from colon cancer sample at (A) 10 ng input, (B) 50 ng input, (C) 100 ng input, and (D) 250 ng input. Green dots = copy number gain; red dots = copy number loss. Blue lines represent normalised copy number across segment represented by line.

3. Discussion

Our results demonstrate that lower input to the Illumina Infinium assay as utilised in the MethylationEPIC and now unavailable Methylation 450 K assay are feasible and provide reliable and reproducible results.

Because the core of the Infinium assay protocol is a whole genome amplification step, theoretical inputs into this kit could be as low as 1 ng [16], and we have shown by our experiments on fresh blood derived samples, FFPE tissue, and circulating tumour DNA that this target is achievable in future experiments. Methylation calls can be reliably and reproducibly carried out on ctDNA samples, including detection of copy number variation with its origins in the base tumour. A weakness of our study is the small numbers of ctDNA samples studied in this project; however, we believe our data illustrates the feasibility of the approach that could be examined in a larger study in the future. Ideally, we would have further validated reproducibility; however, this was limited because of the pilot nature

of the experiment. Despite our findings, low inputs into this assay should be used with caution, as they go beyond the manufacturers recommendations, and may not always work as intended.

This advance will allow access to previous inaccessible samples. Typical yields of ctDNA are between 1–10 ng/mL of plasma [17], and so a typical blood draw will yield 4–40 ng of ctDNA, which could now ideally be divided between mutational and epigenetic assays. We utilised 20 ng of input of ctDNA into the assay, as this was the maximum available to us, and we wanted to ensure that the assay would be successful. While previous strategies to overcome this low ctDNA input was to pool patient samples in risk defined groups [18] for exploratory purposes, this has no benefit for detecting prognostic signatures in individual patients for accurate risk stratification and/or treatment guidance. Additionally, biopsy samples of human disease states (e.g., cancer), especially when obtained from formalin-fixed, paraffin embedded specimens, are typically very small and yield small amounts of DNA [19]. These samples are now accessible to whole methylome analysis. For FFPE samples, we would recommend an input of at least 50 ng of DNA, although smaller amounts are possible, with increasing inaccuracy, but if input DNA is not a limiting factor, we would still suggest use at the recommended input into the assay. Therefore, in samples where copy number variants need to be detected, as high an input mass of DNA should be achieved as possible, as lower input leads to greater dispersion of probes and therefore greater uncertainty about precise copy number. A weakness of this study is that we were unable to demonstrate reproducibility of analysis of the same ctDNA sample, as the limited input material was expended fully in the initial experiment. However, there was good correlation between the original fresh frozen tumour and the associated ctDNA sample. Additionally, we cannot give a precise recommendation for recommended input to maximise CNV detection, as our sample size precludes this.

We have also observed differences in CNV between ctDNA and fresh tumour samples. It is difficult to know whether these are genuine CNV or artefacts caused by whole genome amplification; however, the use of methylation arrays to derive copy number has been extensively validated [20]. The ability to access CNV samples in the same assay as whole genome methylation is particularly attractive, especially in ctDNA samples, as it prevents using up precious ctDNA.

Little, if any, work has examined the whole epigenome of circulating tumour DNA. Several groups have demonstrated [21,22] single gene techniques across a variety of cancer types. We have demonstrated that oncogenic methylation changes can be detected as well as copy number variants that are present in the original tumour. Recent work in DNA mutation burden from ctDNA samples suggests that the ctDNA seen represent the dominant metastatic clone in the cancer [23] and can be used for disease tracking [24]. Several tumours demonstrate aberrant methylation as a key driver of their progression, including head and neck, brain, and colorectal cancer. Epigenetic changes can be demonstrated from plasma samples, but these have been restricted to single gene promoters and thus make discovery work difficult [25,26]. Similarly, discovery work in human disease samples, e.g., in pre-treatment biopsies taken at endoscopy, has also been limited due to the amount of input DNA present after extraction. This technique will allow access to these previously inaccessible disease states and allow temporal profiling of methylation i.e., pre-, during, and post-treatment.

In conclusion, we have demonstrated that low input is possible into the Illumina HumanMethylation series of arrays, allowing access to new sample types that would have previously been inaccessible as well as detection of copy number variants and oncogenic methylation changes.

4. Materials and Methods

4.1. Patient Samples

All patient samples were obtained under either ethics for the S-CORT project (ref 15/EE/0241), the ethics for the Birmingham Human Biomaterials Resource Centre (ref 15/NW/

0079), or the Accelerated (head and neck squamous cell carcinoma, HNSCC) tissue collection platform (REC ref: 16/NW/0265). In order to understand the effects of low input on FFPE derived samples (the SCORT cohort), DNA from the SCORT consortium was used. Sixteen biological FFPE DNA samples were processed on Infinium[®] MethylationEPIC BeadChips in a single batch (Illumina Inc., Cambridge, UK). Each biological sample had three varying DNA input amounts of 50, 100, and 150 ng to give a 48 sample comparison. In order to understand the effects of DNA input on fresh DNA, blood derived samples three biological genomic DNA samples (from three separate healthy donors) were processed on an Infinium[®] HumanMethylation 450 k BeadChip (the Blood cohort). For this one biological sample, input amount varied from 10, 50, 100, 150, 200, 250, 300, and 500 ng (labelled sample “C”). The other two biological DNA samples were processed using 10, 50, 100, 200, and 250 ng input (labelled samples “1” and “3”). For the head and neck ctDNA samples (HNSCC samples), six HNSCC patient biological samples were obtained pre-treatment from the Accelerated tissue collection platform—fresh tissue and blood (collected in Streck blood tubes)—and processed on an Infinium[®] HumanMethylation 450 k BeadChip.

For the Infinium[®] MethylationEPIC BeadChip and Infinium[®] HumanMethylation 450 k BeadChip, the Illumina recommended input amount is ≥ 250 ng. Previous data [27] has demonstrated the equivalence of the 450 k vs. 850 k array for high level analysis.

4.2. DNA Extraction

Human FFPE DNA was extracted from between six and nine 5-micron FFPE colorectal tumour blocks using the QiaAmp Micro Kit (Qiagen, Qiagen Ltd, Manchester, UK). A marked up H&E stained section was used to guide macrodissection of tumour tissue. An assessment of DNA quality and quantity was made initially using a Nanodrop 1000 3.3.0 (Thermo Scientific, Paisley, UK), then the quantity of double stranded DNA was determined using the Quant-iT Picogreen ds DNA Assay Kit (Life technologies, Paisley, UK). Sample concentrations were checked and normalised to 20 ng (six samples) and 250 ng (six samples).

Genomic DNA was extracted from HNSCC fresh tissue biopsies using the DNeasy blood and tissue kit (Qiagen). Cell-free DNA was extracted from 2 mL of plasma from HNSCC blood samples using the QIAamp circulating nucleic acid kit (Qiagen). Sample concentrations were checked and normalised to 20 ng (six ctDNA samples) and 250 ng (six tissue samples).

For the blood samples, a Promega Maxwell RSC instrument (AS4500) was used with the Maxwell RSC Whole Blood DNA Kit (AS1520) (Promega, Southampton, UK). This is a semi-automated DNA extraction method that utilises paramagnetic particles in pre-loaded cartridges to bind DNA and elute into 20 μ L volumes. The purified DNA was quantified using the Qubit 2.0 Fluorometric Quantitation instrument using the dsDNA BR (broad range) assay.

4.3. FFPE QC Assay

To determine FFPE DNA suitability for the Infinium HD FFPE methylation assay, the quality was tested in duplicate by real-time PCR following the Illumina FFPE QC protocol (Part # 15020981 Rev. C. Illumina Ltd, Cambridge, UK). Amplification of FFPE sample DNA was compared with the amplification of a Quality Control Template (QCT). The real-time PCR threshold cycle (Ct) was averaged, and a Delta-Ct (DCt) for each sample was calculated ($DCt = Ct_{FFPE} - Ct_{QCT}$). An FFPE DNA sample was deemed suitable if the DCt was <5 .

4.4. Bisulphite Conversion and Restoration of DNA

The FFPE DNA was bisulphite converted using the EZ DNA Methylation Kit (cat# D5002, Zymo Research, Irvine, CA, USA) following the manufacturer's instructions appropriate for the Illumina HD FFPE assay. A total of 8 μ L of DNA was eluted and taken through to the next stage FFPE DNA Restoration. For non-FFPE DNA, 4 μ L was taken

through directly to the Infinium HD FFPE Methylation Assay. FFPE DNA restoration was achieved by using the Infinium HD FFPE restore protocol. This restores degraded FFPE DNA to a state that is amplifiable. All eluted restored DNA (approx. 8 μ L) was taken through to the Infinium HD FFPE methylation assay.

4.5. Infinium HD FFPE Methylation Assay

All DNA samples were processed following the Infinium HD FFPE Methylation Assay, Manual Protocol. For the Infinium[®] MethylationEPIC BeadChips, the appropriate RA1 resuspension and hybridising volume were used as provided in the Infinium HD Methylation Assay Manual Protocol. The Illumina iScan was used to scan the arrays recording high resolution images of light emitted by the excited fluorophores at each CpG site on the arrays. The raw intensity data from these images were stored as .idat files and used for analysis. Illumina GenomeStudio v2011.1 was used to process the .idat files. Sample dependent and independent controls were analysed to ensure the assay procedure had been successful. The % of CpG detection was determined at a confidence level of $p = 0.05$.

All samples passed QC. The analysis had no sample groupings or normalisation, and the background was subtracted.

4.6. Downstream Analysis

All Idat files were imported into R/Bioconductor (R Version 3.6.3) and analysed using ChAMP (v2.13.2) [28] pipeline up to DMP/DMR processing. ChAMP is a R/Bioconductor pipeline that reads raw data from Illumina iDat files, carries out filtering and quality control, and can carry out all downstream processing. Initially, ChAMP loads in probe intensities direct from iDAT files using a sample sheet provided by the end user. Probe filtering was then carried using ChAMP default settings with probes being dropped if they had a detection p -value of >0.01 , had <3 beads in at least 5% of samples per probe, were a non CpG probe, or were in the list of SNPs as provided in the Zhou et al. paper [29] or in the Nordlund et al. [30] paper describing multi-hit probes, and finally, all X/Y chromosome probes were filtered out. Normalisation was carried out using Beta-Mixture Quantile (BMIQ) Normalization. Although samples were done in the same batch, and sample size was relatively small, SVN analysis did not demonstrate any confounding batch effect, so batch correction was not performed. For intra-group comparisons, the probes for each concentration level were compared to every other concentration using multilevel regression. For copy number calling, the conumee algorithm [20] was used. For between sample correlations, the *cor* command of the *psych* module from CRAN (<https://cran.r-project.org/web/packages/psych/index.html>, accessed on 6 February 2021) was used. All correlations were adjusted for multiple testing using the Bonferroni correction. For pathway analysis, top ranked differentially expressed DMP/DMR were exported to GProfiler [31].

Supplementary Materials: The following are available online at <https://www.mdpi.com/2075-4655/5/1/6/s1>.

Author Contributions: Manuscript: A.D.B. and K.P.; molecular analysis: C.W. and A.B.; sample acquisition and patient cohort: C.W., K.P., E.D., A.B., S.R., J.B., N.B., and R.S., S-CORT Consortium: H.M., P.N., and A.D.B.; bioinformatics: A.D.B., A.B., and E.D. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the CRUK and MRC Stratified Medicine award for the Stratification in Colorectal Cancer (S-CORT) project (ref MR/M016587/1) as well as a Wellcome Trust Institutional Support Fund Award (P.N.). A.D.B. is currently supported by a Cancer Research UK Advanced Clinician Scientist award (ref C31641/A23923).

Institutional Review Board Statement: All patient samples were obtained under either ethics for the S-CORT project (ref 15/EE/0241, East of England Ethics committee, approval 1/10/15), the ethics for the Birmingham Human Biomaterials Resource Centre (ref 15/NW/0079, North West England Ethics Committee, approval 01/15), or the Accelerated (head and neck squamous cell carcinoma,

HNSCC) tissue collection platform (REC ref: 16/NW/0265, North West England Ethics Committee, approval 01/16). This specific study was not reviewed by an ethics committee, as the work for it was planned for the S-CORT study.

Informed Consent Statement: Informed written consent was obtained for all participants in the study.

Data Availability Statement: All data will be made available on GEO on publication of manuscript.

Acknowledgments: S-CORT Consortium membership list: Louise Brown (University College London), Peter Campbell (Wellcome Trust Sanger Institute), Phillip Dunne (Queens University Belfast), Richard Kaplan (University College London), Richard Kennedy (Queens University Belfast), Mark Lawler (Queens University Belfast), Tim Maughan (University of Oxford), Jonathan Murray (Queens University Belfast), Phillip Quirke (University of Leeds), Susan Richman (University of Leeds), Manuel Salto-Tellez (Queens University Belfast), Ian Tomlinson (University of Edinburgh), Steven Walker (Almac Diagnostics), Richard Wilson (Queens University Belfast).

Conflicts of Interest: A.D.B. has received travel funding and speaker fees from Illumina Inc.

Abbreviations

DNA	Deoxyribonucleic acid
ctDNA	Circulating tumour DNA
FFPE	Formalin fixed, paraffin embedded
WGA	Whole genome amplification
CIMP	CpG Island methylator phenotype
DMP	Differentially methylated position
DMR	Differentially methylated region

References

- Kulis, M.; Esteller, M. DNA methylation and cancer. *Adv. Genet.* **2010**, *70*, 27–56. [[CrossRef](#)] [[PubMed](#)]
- Toyota, M.; Issa, J.P. CpG island methylator phenotypes in aging and cancer. *Semin. Cancer Biol.* **1999**, *9*, 349–357. [[CrossRef](#)] [[PubMed](#)]
- Ogino, S.; Cantor, M.; Kawasaki, T.; Brahmandam, M.; Kirkner, G.J.; Weisenberger, D.J.; Campan, M.; Laird, P.W.; Loda, M.; Fuchs, C.S. CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies. *Gut* **2006**, *55*, 1000–1006. [[CrossRef](#)]
- Puccini, A.; Berger, M.D.; Naseem, M.; Tokunaga, R.; Battaglin, F.; Cao, S.; Hanna, D.L.; McSkane, M.; Soni, S.; Zhang, W.; et al. Colorectal cancer: Epigenetic alterations and their clinical implications. *Biochim. Biophys. Acta Rev. Cancer* **2017**, *1868*, 439–448. [[CrossRef](#)] [[PubMed](#)]
- Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **2012**, *487*, 330–337. [[CrossRef](#)] [[PubMed](#)]
- Zhou, C.; Ye, M.; Ni, S.; Li, Q.; Ye, D.; Li, J.; Shen, Z.; Deng, H. DNA methylation biomarkers for head and neck squamous cell carcinoma. *Epigenetics* **2018**, *13*, 398–409. [[CrossRef](#)] [[PubMed](#)]
- Colacino, J.A.; Dolinoy, D.C.; Duffy, S.A.; Sartor, M.A.; Chepeha, D.B.; Bradford, C.R.; McHugh, J.B.; Patel, D.A.; Virani, S.; Walline, H.M.; et al. Comprehensive analysis of DNA methylation in head and neck squamous cell carcinoma indicates differences by survival and clinicopathologic characteristics. *PLoS ONE* **2013**, *8*, e54742. [[CrossRef](#)] [[PubMed](#)]
- Warton, K.; Mahon, K.L.; Samimi, G. Methylated circulating tumor DNA in blood: Power in cancer prognosis and response. *Endocr. Relat. Cancer* **2016**, *23*, R157–171. [[CrossRef](#)]
- Razavi, P.; Li, B.T.; Brown, D.N.; Jung, B.; Hubbell, E.; Shen, R.; Abida, W.; Juluru, K.; De Bruijn, I.; Hou, C.; et al. High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. *Nat. Med.* **2019**, *25*, 1928–1937. [[CrossRef](#)] [[PubMed](#)]
- Feber, A.; Dhimi, P.; Dong, L.; de Winter, P.; Tan, W.S.; Martinez-Fernandez, M.; Paul, D.S.; Hynes-Allen, A.; Rezaee, S.; Gurung, P.; et al. UroMark-a urinary biomarker assay for the detection of bladder cancer. *Clin. Epigenetics* **2017**, *9*, 8. [[CrossRef](#)] [[PubMed](#)]
- Olova, N.; Krueger, F.; Andrews, S.; Oxley, D.; Berrens, R.V.; Branco, M.R.; Reik, W. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol.* **2018**, *19*, 33. [[CrossRef](#)]
- Meissner, A.; Gnirke, A.; Bell, G.W.; Ramsahoye, B.; Lander, E.S.; Jaenisch, R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **2005**, *33*, 5868–5877. [[CrossRef](#)] [[PubMed](#)]
- Moran, S.; Arribas, C.; Esteller, M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* **2016**, *8*, 389–399. [[CrossRef](#)] [[PubMed](#)]

14. Gyanchandani, R.; Kvam, E.; Heller, R.; Finehout, E.; Smith, N.; Kota, K.; Nelson, J.R.; Griffin, W.; Puhalla, S.; Brufsky, A.M.; et al. Whole genome amplification of cell-free DNA enables detection of circulating tumor DNA mutations from fingerstick capillary blood. *Sci. Rep.* **2018**, *8*, 17313. [[CrossRef](#)] [[PubMed](#)]
15. Cancer Genome Atlas, N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **2015**, *517*, 576–582. [[CrossRef](#)] [[PubMed](#)]
16. Liu, H.E.; Triboulet, M.; Zia, A.; Vuppapalaty, M.; Kidess-Sigal, E.; Collier, J.; Natu, V.S.; Shokoohi, V.; Che, J.; Renier, C.; et al. Workflow optimization of whole genome amplification and targeted panel sequencing for CTC mutation detection. *Npj Genom. Med.* **2017**, *2*, 34. [[CrossRef](#)] [[PubMed](#)]
17. Johansson, G.; Andersson, D.; Filges, S.; Li, J.; Muth, A.; Godfrey, T.E.; Stahlberg, A. Considerations and quality controls when analyzing cell-free tumor DNA. *Biomol. Detect. Quantif* **2019**, *17*, 100078. [[CrossRef](#)]
18. Gallardo-Gomez, M.; Moran, S.; Paez de la Cadena, M.; Martinez-Zorzano, V.S.; Rodriguez-Berrocal, F.J.; Rodriguez-Girondo, M.; Esteller, M.; Cubiella, J.; Bujanda, L.; Castells, A.; et al. A new approach to epigenome-wide discovery of non-invasive methylation biomarkers for colorectal cancer screening in circulating cell-free DNA using pooled samples. *Clin. Epigenetics* **2018**, *10*, 53. [[CrossRef](#)] [[PubMed](#)]
19. Alderdice, M.; Richman, S.D.; Gollins, S.; Stewart, J.P.; Hurt, C.; Adams, R.; McCorry, A.M.; Roddy, A.C.; Vimalachandran, D.; Isella, C.; et al. Prospective patient stratification into robust cancer-cell intrinsic subtypes from colorectal cancer biopsies. *J. Pathol.* **2018**, *245*, 19–28. [[CrossRef](#)]
20. Capper, D.; Jones, D.T.W.; Sill, M.; Hovestadt, V.; Schrimpf, D.; Sturm, D.; Koelsche, C.; Sahm, F.; Chavez, L.; Reuss, D.E.; et al. DNA methylation-based classification of central nervous system tumours. *Nature* **2018**, *555*, 469–474. [[CrossRef](#)] [[PubMed](#)]
21. Leygo, C.; Williams, M.; Jin, H.C.; Chan, M.W.Y.; Chu, W.K.; Grusch, M.; Cheng, Y.Y. DNA Methylation as a Noninvasive Epigenetic Biomarker for the Detection of Cancer. *Dis. Markers* **2017**, *2017*, 3726595. [[CrossRef](#)] [[PubMed](#)]
22. Warton, K.; Samimi, G. Methylation of cell-free circulating DNA in the diagnosis of cancer. *Front. Mol. Biosci.* **2015**, *2*, 13. [[CrossRef](#)] [[PubMed](#)]
23. Abbosh, C.; Birkbak, N.J.; Wilson, G.A.; Jamal-Hanjani, M.; Constantin, T.; Salari, R.; Le Quesne, J.; Moore, D.A.; Veeriah, S.; Rosenthal, R.; et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **2017**, *545*, 446–451. [[CrossRef](#)]
24. Rosell, R.; Karachaliou, N. Lung cancer: Using ctDNA to track EGFR and KRAS mutations in advanced-stage disease. *Nat. Rev. Clin. Oncol.* **2016**, *13*, 401–402. [[CrossRef](#)] [[PubMed](#)]
25. Schrock, A.; Leisse, A.; de Vos, L.; Gevensleben, H.; Droge, F.; Franzen, A.; Wachendorfer, M.; Schrock, F.; Ellinger, J.; Teschke, M.; et al. Free-Circulating Methylated DNA in Blood for Diagnosis, Staging, Prognosis, and Monitoring of Head and Neck Squamous Cell Carcinoma Patients: An Observational Prospective Cohort Study. *Clin. Chem.* **2017**, *63*, 1288–1296. [[CrossRef](#)] [[PubMed](#)]
26. Mydlarz, W.K.; Hennessey, P.T.; Wang, H.; Carvalho, A.L.; Califano, J.A. Serum biomarkers for detection of head and neck squamous cell carcinoma. *Head Neck* **2016**, *38*, 9–14. [[CrossRef](#)] [[PubMed](#)]
27. Solomon, O.; MacIsaac, J.; Quach, H.; Tindula, G.; Kobor, M.S.; Huen, K.; Meaney, M.J.; Eskenazi, B.; Barcellos, L.F.; Holland, N. Comparison of DNA methylation measured by Illumina 450K and EPIC BeadChips in blood of newborns and 14-year-old children. *Epigenetics* **2018**, *13*, 655–664. [[CrossRef](#)]
28. Tian, Y.; Morris, T.J.; Webster, A.P.; Yang, Z.; Beck, S.; Feber, A.; Teschendorff, A.E. ChAMP: Updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* **2017**, *33*, 3982–3984. [[CrossRef](#)]
29. Zhou, W.; Laird, P.W.; Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* **2017**, *45*, e22. [[CrossRef](#)]
30. Nordlund, J.; Backlin, C.L.; Wahlberg, P.; Busche, S.; Berglund, E.C.; Eloranta, M.L.; Flaegstad, T.; Forestier, E.; Frost, B.M.; Harila-Saari, A.; et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.* **2013**, *14*, r105. [[CrossRef](#)] [[PubMed](#)]
31. Raudvere, U.; Kolberg, L.; Kuzmin, I.; Arak, T.; Adler, P.; Peterson, H.; Vilo, J. g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **2019**, *47*, W191–W198. [[CrossRef](#)] [[PubMed](#)]