Development of digital PCR DNA methylation assays for blood plasma-based diagnosis of lung cancer

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Benjamin Russell Buckman Brown

November 2017

Supervisors: Dr. Triantafillos Liloglou and Prof. John K Field

ABSTRACT

Lung cancer is the leading cause of cancer-related death and is usually diagnosed at advanced stage leading to poor patient survival. Therefore there is a pressing need for early detection of disease. DNA methylation is an early event in carcinogenesis and a limited number of diagnostic markers have been developed for clinical use. This thesis seeks to address whether the development and application of novel DNA methylation assays can diagnose lung cancer at early stage.

Previously identified DNA methylation biomarkers, along with novel targets identified by methylation microarray, were developed in multiplex assay format. Twelve markers were used to screen 417 bronchoalveolar lavage specimens from Liverpool Lung Project (LLP) subjects divided into training and validation sets. The optimal biomarker panel (*CDKN2A*, *RARB* and *TERT*) demonstrated improved clinical sensitivity and specificity (Sensitivity/Specificity: 85.7%/93.8%, AUC: 0.91) compared to previous studies. The optimal methylation algorithm detected more than 60% of stage T1 tumours and 93 cytologically occult lung cancer cases. Eight methylated DNA assays were optimised for use with the newly developed Droplet Digital™ PCR (ddPCR) platform and a targeted pre-amplification technique, MethPlex enrichment, was developed. I established a comprehensive analytical framework to compare performance of methylation-specific ddPCR and quantitative methylation-specific PCR directly and in combination with MethPlex enrichment. ddPCR demonstrated greater precision and linearity, lower limit of detection (*WT1* MethPlex ddPCR LOD95 = 1.86 GE), and discriminated twofold differences in methylated DNA input. MethPlex ddPCR detected DNA methylation more frequently in lung cancer patient plasma than in controls in a retrospective case-control study. Technical methylation controls were consistently and precisely detected at inputs as low as 3 methylated copies. Discriminatory efficiency of marker combinations was inadequate, presumably due to limitations in DNA extraction methodology.

DNA methylation biomarker diagnostic performance in bronchoalveolar lavage merits further validation in a prospective trial. MethPlex ddPCR analysis showed great promise, demonstrating highly sensitive DNA methylation detection in technical assessment. It is expected that appropriate DNA extraction procedures and higher cfDNA yields will lead to much improved clinical discriminatory efficiency.

Table of contents

List of Figures

List of Tables

List of abbreviations

1D	one-dimensional
5methyl-dCTP	2'-deoxy-5-methylcytidine 5'-triphosphate
β_0	regression intercept
β_1	regression slope
ABCB1	ATP binding cassette subfamily B member 1
ACTB	actin beta
AIC	Akaike information criterion
ALK	anaplastic lymphoma kinase
ANOVA	analysis of variance
AUC	area under curve
BAL	bronchoalveolar lavage
BHQ	Black Hole Quencher®
BIC	Bayesian information criterion
BICq	Bayesian Information Criterion with Bernoulli prior
BIDD	Biomarkers and Imaging Discovery and Development
BM	biomarker
bp	base pair
BSR	best subset regression
C	cytosine
CDKN2A	cyclin dependent kinase inhibitor 2A
CDO1	cysteine dioxygenase type 1
CE	Conformité Européene
cfDNA	cell-free DNA
CGI	CpG island
CI	confidence interval
COPD	chronic obstructive pulmonary disease
CpG	5'-C-phosphate-G-3'
Cq	quantification cycle
CT	computed tomography

CV(%)	coefficient of variation
CYGB	cytoglobin
ddPCR	Droplet Digital™ PCR
df	degrees of freedom
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dsDNA	double-stranded DNA
E. coli	Escherichia coli
EDRN	Early Detection Research Network
EDTA	ethylenediaminetetraacetic acid
EV	Extracellular vesicle
EVL	Enah/Vasp-like
F primer	forward primer
F2R	coagulation factor II thrombin receptor
FAM	6-carboxyfluorescein
FDA	Food and Drug Administration
FFPE	formalin-fixed paraffin-embedded
FRET	Förster resonance energy transfer
gDNA	genomic DNA
GE	genome equivalents
GN model	optimal model identified in publication (first author: George Nikolaidis)
GPCR	G-protein-coupled receptor
H&E	hematoxylin and eosin
H_0	null hypothesis
HDACi	histone deacetylase inhibitor
HinP1I	Restriction endonuclease first identified in Haemophilus influenzae P1
HOXA1	homeobox A1
HOXA10	homeobox A10
HOXA17	homeobox A17

HPSF	High Purity Salt Free
HR	hazard ratio
IASLC	International Association for the Study of Lung Cancer
IQR	interquartile range
ISO	International Organization for Standardization
IVD	*In Vitro* Diagnostic
kb	kilobase pair
LB broth	Luria-Bertani broth
L_C	critical value, signal domain
L_D	limit of quantification, signal domain
LDCT	low-dose computed tomography
LLP	Liverpool Lung Project
lncRNA	long non-coding RNA
LOD	limit of detection
LOD95	limit of detection (detected with 95% confidence)
LOD_{qual}	limit of detection informally defined as lowest non-zero input at which all replicate measurements render positive qualitative results
L_Q	limit of quantification, signal domain
M.SssI	CpG methyltrasferase first identified in Spiroplasma sp. strain MQ1
MARCH11	membrane associated ring-CH-type finger 11
MDR	multifactor dimensionality reduction
MethPlex	methylation-specific multiplex PCR amplification
MGB	minor groove binder
MGMT	O-6-methylguanine-DNA methyltransferase
mRNA	messenger RNA
miRNA	microRNA
MS-NaME	Methylation Specific Nuclease-assisted Minor-allele Enrichment
MSP	methylation-specific PCR
MT1G	metallothionein 1G

NCI	National Cancer Institute
NED	2′-chloro-5′-fluoro-7′,8′-benzo-1,4-dichloro-6-carboxyfluorescein
NKK	nicotine-derived nitrosamine ketone
NLST	National Lung Cancer Screening Trial
NPY	neuropeptide Y
NSCLC	non-small cell lung cancer
NSCLC (NOS)	non-small cell lung cancer (not otherwise specified)
NTC	non-template control
NTRK3	neurotrophic receptor tyrosine kinase 3
OR	odds ratio
PBMC	peripheral blood mononuclear cell
PCR	polymerase chain reaction
PITX2	paired like homeodomain 2
pN	histopathological N stage (N = degree of spread to regional lymph nodes)
POCT	point-of-care testing
pT	histopathological T stage (T = size/extent of primary tumour)
PTGDR	prostaglandin D2 receptor
PTGER4	prostaglandin E receptor 4
pUC19	plasmid University of California 19
qMSP	quantitative methylation-specific PCR
qPCR	quantitative polymerase chain reaction
R primer	reverse primer
R^2	coefficient of determination
RARB	retinoic acid receptor beta
Ras	Ras family, named from Kirsten rat sarcoma oncogene homologue
RASSF1	Ras association domain family member 1
RASSF4	Ras association domain family member 4
RCF	Relative centrifugal force

RMSE	root-mean-square error
RNA	ribonucleic acid
ROC	receiver operating characteristic
r_s	Spearman's rank correlation coefficient
SAM	S-adenosyl methionine
SCLC	small-cell lung cancer
SD	standard deviation
Se	sensitivity, defined as: $\frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives}$
SEPT9	septin 9
SHOX2	short stature homeobox 2
SOP	standard operating procedure
SOX17	SRY-box 17
Sp	specificity, defined as: $\frac{number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ positives}$
s_r	repeatability standard deviation
s^2_r	repeatability variance
SspI	Restriction endonuclease first identified in Sphaerotilus species
SspI mCpG pUC19 DNA	SspI-digested, CpG methylated pUC19 DNA
T	thymine
T_a	annealing temperature
TAC1	tachykinin precursor 1
Taq	Thermus aquaticus
TBE	Tris-borate-EDTA
TERT	telomerase reverse transcriptase
T_m	melting temperature
Tris	tris(hydroxymethyl)aminomethane
TSG	tumour suppressor gene
UNCX	UNC homeobox
UNG	uracil N-glycosylase

VIC	2′-chloro-7′phenyl-1,4-dichloro-6-carboxy-fluorescein
WGA	whole genome amplification /whole genome amplified
WIF1	WNT inhibitory factor 1
WLS	Weighted Least Squares
WT1	Wilms tumour 1
x_C	critical value, DNA input quantity domain
x_D	limit of detection, DNA input quantity domain
x_Q	limit of quantification, DNA input quantity domain

ACKNOWLEDGEMENTS

I would like to offer my deepest thanks to my supervisors, Dr Lakis Liloglou and Prof John Field, for their guidance and support. In particular I would like to thank Lakis for maintaining his belief in me through difficult times and always having an open door.

I would also like to thank Dr Mike Davies for being a source of innovative ideas and insightful suggestions, Dr Michael Marcus for his advice in epidemiological and statistical matters, and Dr Elinor Chapman for her assistance in bronchoalveolar specimen screening.

I would like to say a personal thank you to my late mother, Andrea, who sadly passed away during the course of my studies. She gave me the best start in life one could ever hope for and was a constant inspiration.

A special thank you goes to my partner, Bonnie, for her unconditional love, support and encouragement.

Finally, I would like to acknowledge the Roy Castle Lung Cancer Foundation and the Institute of Translational Medicine, University of Liverpool for their generous funding.

Chapter 1

Introduction

1.1 General introduction

Lung cancer is the most common cancer worldwide and is also the leading cause of cancer related-death (Ferlay et al., 2015). Malignancy is usually diagnosed at an advanced stage leading to poor patient survival (Liloglou et al., 2014). In contrast, improved patient survival in the event of early detection has previously been demonstrated (McPhail et al., 2015). Previous work within our group identified a DNA methylation panel demonstrating very good discriminatory characteristics (82% sensitivity, 91% specificity) in bronchoalveolar lavage specimens (Nikolaidis et al., 2012). This thesis builds upon this previous work and was carried out as part of the Liverpool Lung Project (Field et al., 2005), with the ultimate goal of reducing lung cancer mortality through the early detection of this devastating disease.

The remainder of this chapter primarily serves as a concise presentation of background information regarding lung cancer and the detection of DNA methylation in cancer. Chapter 2 outlines the materials and methods used throughout this thesis. In Chapter 3, additional experimentally identified DNA methylation markers were added to the panel identified in the previous study (Nikolaidis et al., 2012). These were assessed in a retrospective case-control study of bronchoalveolar lavage samples, independent to those analysed previously. We hypothesized that the interrogation of additional markers would improve the diagnostic accuracy of the pre-existing marker panel. In addition, the assays used to detect DNA methylation have been further developed into multiplex reactions and it was anticipated that potential analytical improvements would also enhance the discriminatory ability of diagnostic algorithms. As a consequence, we also expected that DNA methylation analysis of bronchoalveolar lavage would be more sensitive than current gold-standard cytological diagnosis.

While bronchoalveolar lavage is classified as being a moderately invasive procedure, it has to be performed in a clinical environment, is rather unpleasant for the patient and is not without risk. In addition it is a costly clinical process. Non-invasive diagnostic procedures involving blood draws would be far more preferable. Previous experience within the group indicated that quantitative methylation-specific PCR assays (qMSP), such as those used in the Chapter 3, do not possess the analytical sensitivity required to reliably and precisely detect methylated DNA in blood plasma. At the start of this project, Droplet Digital™ PCR (ddPCR) was a newly developed technology that can detect nucleic acids with greater precision than real-time PCR methodologies (Hindson et al., 2011, 2013). Chapter 4 describes and documents the development and optimisation of assays and associated workflows for the detection of a number of methylated DNA targets. This was done at a moderately high level of total DNA input to emulate the likely yields from specimens such as bronchoalveolar lavage, and also at a very low level to imitate the low yields expected from plasma samples (Devonshire et al., 2014). Methylated DNA analysis at low level total DNA input also necessitated the development of a methylated target multiplex PCR pre-amplification method which I have designated MethPlex enrichment. The primary objective of this part of the study was to develop a number of highly optimised, robust methylation-specific ddPCR assays that would perform reliably in the detection of methylated DNA markers present in clinical samples.

A comprehensive technical comparison of ddPCR and qMSP technologies in the detection of methylated DNA is covered in Chapter 5. Two of the newly developed assays targeting methylated *RASSF1* and *WT1* promoters were initially evaluated on both platforms in the analysis of multiple methylated DNA standard curves.

Methylated DNA was diluted in nominally unmethylated peripheral blood mononuclear cell (PBMC) background at moderate total DNA quantity (66 ng/20,000 GE). A similar analysis of low total DNA inputs (10 ng/3030 GE) in conjunction with MethPlex enrichment was also performed, resulting in a moderate sized dataset for each marker/platform combination (N = 270). The data obtained was subjected to a rigorous, comprehensive analysis including the use of methods published as ISO standards (ISO, 1994a, 1994b, 1994c, 2000). The primary aim of this part of the study was to compare the performance of the two nucleic acid detection technologies. Droplet Digital™ PCR has previously demonstrated highly precise quantification of nucleic acids (Pinheiro et al., 2012) and indeed exhibited increased precision compared to qPCR in the analysis of diverse nucleic acid species (Hindson et al., 2013; Sedlak et al., 2014; Strain et al., 2013).

The final experimental chapter, Chapter 5, demonstrates the screening plasma specimens in a retrospective case-control study using MethPlex ddPCR assays. This study was intended to assess the utility and discriminatory ability of novel assays targeting methylated DNA in non-invasive clinical samples. The final chapter of the thesis is a general discussion and appraisal of the completed study.

1.2 Lung cancer

Lung cancer is the second most common cancer in men and women in the UK, accounting for 13% of new diagnoses (Cancer Research UK, 2015a). Between 2001 and 2005 the UK lung cancer five-year survival rate was as low as 7.5% (Holmberg et al., 2010). This has improved marginally and is still below 10% (Cancer Research UK, 2015b) (**Figure 1.2-1**). As a result, lung cancer is by far the most common cause of cancer-related death in the UK, with 35,895 deaths in 2014 at a mortality rate of 61 per 100,000 persons (Cancer Research UK, 2015c) (**Figure 1.2-2**).

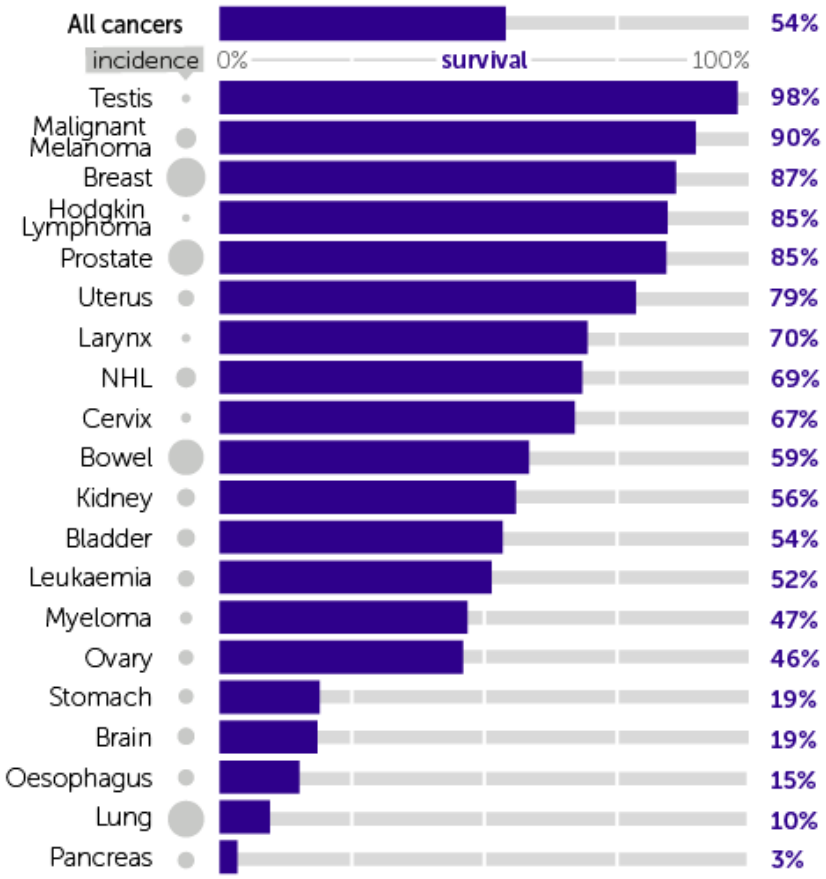


Figure 1.2-1 Age-Standardised Five-Year Net Survival, Selected Cancers, Adults (Aged 15-99), England and Wales, 2010-2011
(Cancer Research UK, 2015b)

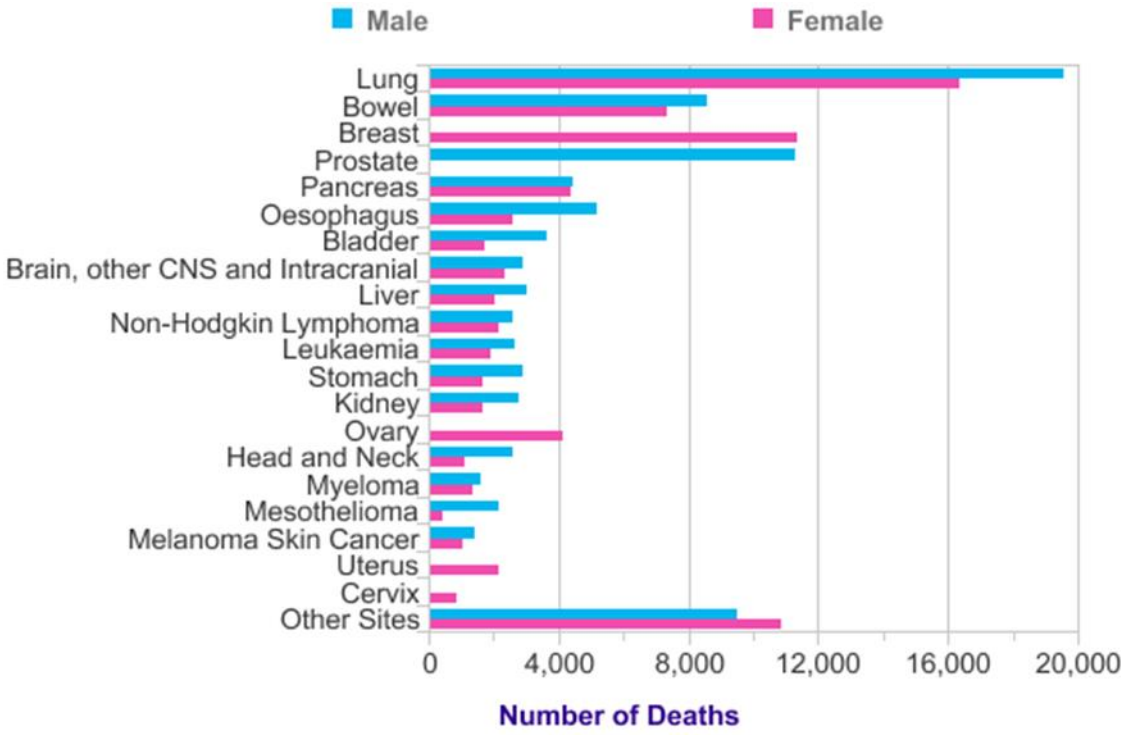


Figure 1.2-2 The 20 most common causes of cancer death, 2014
Lung cancer is by far the most common cause of cancer-relataed death in the UK **(Cancer Research UK, 2015c)**.

Worldwide, the lung cancer mortality rate exceeds that of colon, prostate and breast cancer combined, and, as previously stated, lung cancer is the most frequent cause of cancer-related death (Dela Cruz et al., 2011; Ferlay et al., 2015).

High mortality rates are primarily caused by diagnosis of disease at late stage when curative treatment is not feasible (Mulshine and van Klaveren, 2011). **Figure 1.2-3** definitively shows that lung cancer was most frequently diagnosed at Stage IV in the UK (48% in England in 2014) and only 13% to 18% diagnosed at Stage I (Cancer Research UK, 2015d). Diagnosis of disease at Stage I and subsequent early intervention can give rise to 5-year survival rates in the region of 70% (Field and Raji, 2010). There is therefore a genuine and pressing need to develop biomarkers for early detection of lung malignancy.

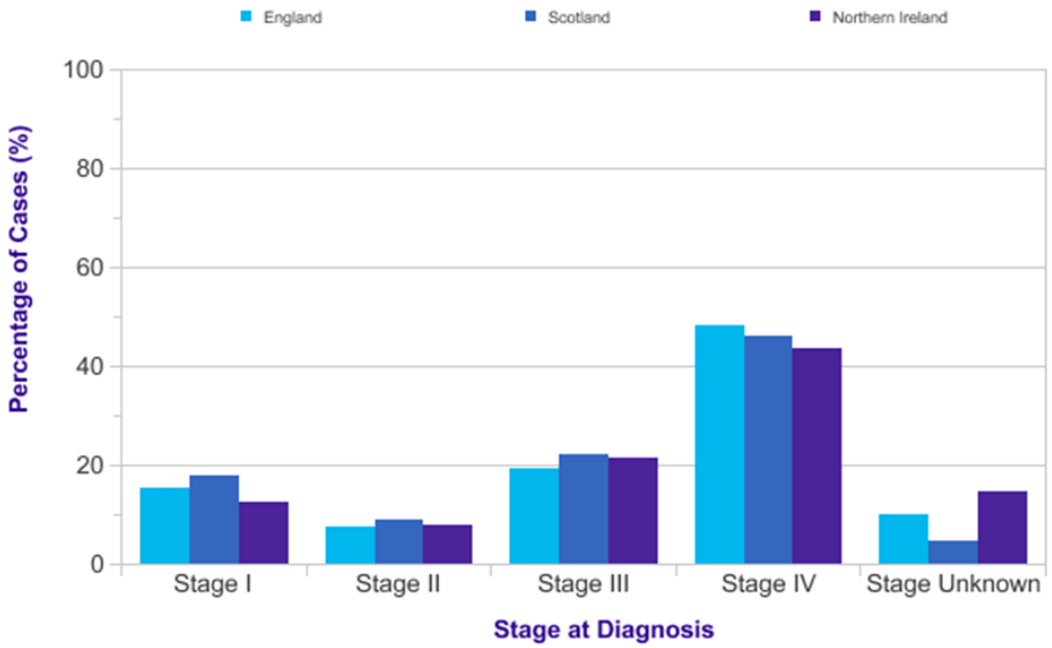


Figure 1.2-3 Proportion of Lung Cancer Cases Diagnosed at Each Stage, All Ages England 2014, Scotland 2014-2015, Northern Ireland 2010-2014 **(Cancer Research UK, 2015d)**

1.2.1 Lung cancer histological classification

Lung cancer is usually split into two classes based on histology, with non-small cell lung cancer (NSCLC) accounting for approximately 80-85% of all lung cancers and adenocarcinoma and squamous cell carcinoma composing about 80% of this category (Langer et al., 2010). Adenocarcinoma is the most frequently diagnosed lung cancer subtype, accounting for about 40% of lung cancer cases and is normally located peripherally. Squamous cell carcinoma, on the other hand, is usually centrally located in primary and secondary bronchi and is more common in smokers.

Small cell lung cancer (SCLC) is found in about 15% of lung cancer cases, is highly aggressive, is normally located centrally and is strongly associated with smoking (Tartarone et al., 2017). Extensive SCLC has a median survival of less than twelve months (Gaspar et al., 2012).

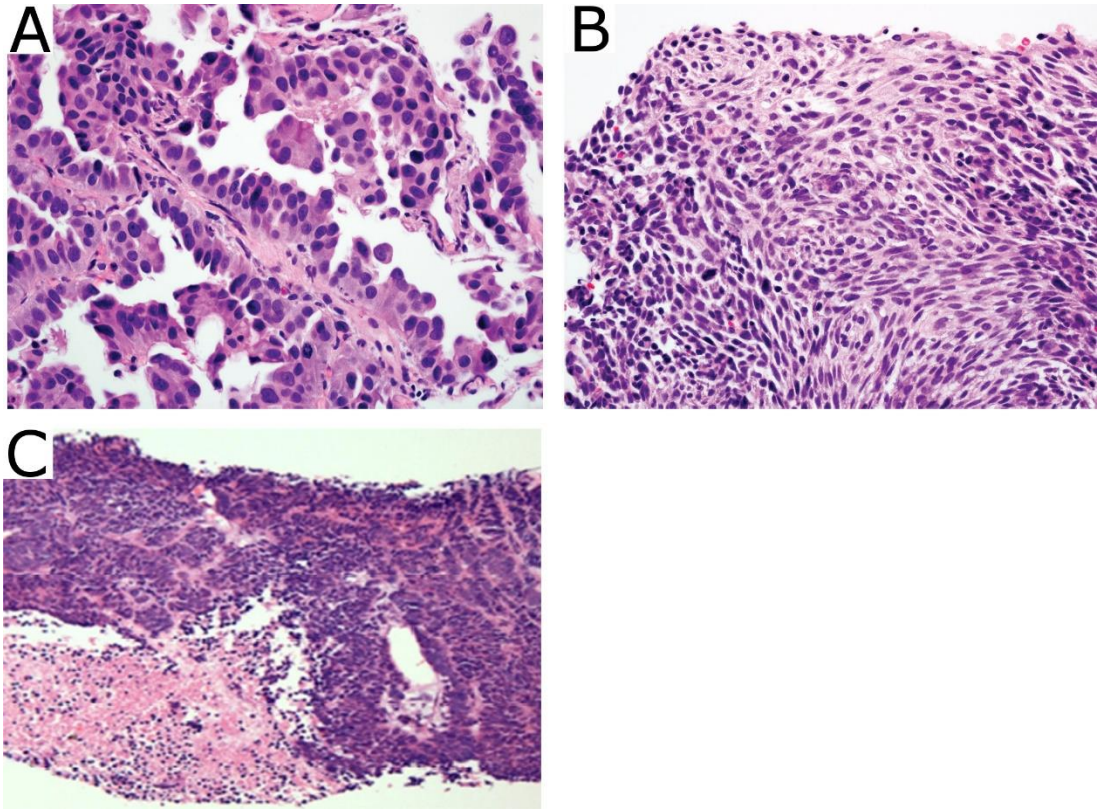


Figure 1.2-4 Example images of lung cancer histological types
H&E stained slides of (A) adenocarcinoma, (B) squamous cell carcinoma and (C) small cell carcinoma of the lung. Images adapted from Davidson, 2013 (C) and Travis, 2013 (A) and (B) (Davidson et al., 2013; Travis et al., 2013a).

1.2.2 Lung cancer staging

The international TNM staging system is used in the designation of disease stage, based upon the pathological assessment of tumour size, nodal status and metastasis classification. The proposed classifications for staging are organized by the International Association for the Study of Lung Cancer (IASLC) Staging Project. A revised eighth edition has been proposed, based upon a retrospective analysis of more than 94,000 cases of lung cancer collected from around the world between 1999 and 2010 (Goldstraw et al., 2016) and is presented in **Table 1.2-1**. This is due to be implemented in January 2018.

Table 1.2-1 Proposed stage groupings for the eighth edition of the TNM classification for lung cancer

Changes to the seventh edition are highlighted in bold and underlined.
TNM, tumour, node, metastasis; Tis, carcinoma *in situ*; T1a(mi), minimally invasive adenocarcinoma. Adapted from Goldstraw, 2016 (Goldstraw et al., 2016).

Clinical stage	**T**	**N**	**M**
Occult carcinoma	TX	N0	M0
Stage 0	Tis	N0	M0
Stage IA1	**T1a(mi)**	**N0**	**M0**
	T1a	**N0**	**M0**
Stage IA2	**T1b**	**N0**	**M0**
Stage IA3	**T1c**	**N0**	**M0**
Stage IB	T2a	N0	M0
	T2b	N0	M0
Stage IIA	**T1a-c**	**N1**	**M0**
Stage IIB	**T2a**	**N1**	**M0**
	T2b	N1	M0
	T3	N0	M0
Stage IIIA	**T1a-c**	**N2**	**M0**
	T2a-b	N2	M0
	T3	N1	M0
	T4	N0	M0
	T4	N1	M0
Stage IIIB	**T1a-c**	**N3**	**M0**
	T2a-b	N3	M0
	T3	**N2**	M0
	T4	N2	M0
Stage IIIC	**T3**	**N3**	**M0**
	T4	**N3**	**M0**
Stage IVA	**Any T**	**Any N**	**M1a**
	Any T	**Any N**	**M1b**
Stage IVB	**Any T**	**Any N**	**M1c**

1.2.3 Lung cancer diagnosis

Conclusive lung cancer diagnosis requires the analysis of an adequate tissue and is vitally important not only in diagnosis itself but also the planning of (Rivera et al., 2013). A relatively large amount of tissue is required or molecular testing. There a number of methods in routine use that can obtain cytology specimens or small biopsies. Bronchoscopy is more efficient at detecting central lesions. CT-guided transthoracic needle aspiration generally works better for detection in the periphery. Methods such as endobronchial ultrasound (Herth et al., 2006) and electromagnetic navigation (Lamprecht et al., 2009) can increase diagnostic yields in patients with peripheral lesions. A thorough review of lung cancer diagnosis in small biopsies and cytological samples published in 2013 engages with the implications of the 2011 International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society Classification (Travis et al., 2013b) and is recommended for an in depth understanding of this area.

1.2.4 Lung cancer treatment

1.2.4.1 Surgery

Surgery is the most common procedure in the management of lung cancer at early stage (Lackey and Donington, 2013). Surgical intervention may require pneumonectomy (whole lung excision), lobectomy (removal of a lobe), wedge resection or segmentectomy (both involving the removal of a smaller portion of the effected lung). This is dependent upon the size of the tumour and pulmonary function (Narsule et al., 2011).

1.2.4.2 Radiotherapy

Radiotherapy treatment is primarily used in the treatment of inoperable tumours or in patients were surgery would present complications (Zappa and Mousa, 2016). Stereotactic ablative radiotherapy (SABR), also known as stereotactic body radiotherapy (SBRT), is a highly precise method of high dose radiation treatment involving the delivery of multiple smaller doses of radiation from multiple positions around the body (Haasbeek et al., 2009), reducing the risk of normal organ toxicity (Louie et al., 2015). This technique results in a high rate of local tumour control with few side effects (Senan et al., 2013). A study involving 4605 elderly Dutch patients indicated that patients 75 years of age and older treated with SABR exhibited a reduction in the hazard ratio of death compared to the control group (HR 0.64, 95% CI 0.56–0.74), and median overall survival significantly increased from16.8 months to 26.1 months (Haasbeek et al., 2012).

1.2.4.3 Chemotherapy

Chemotherapeutic agents may be used as neoadjuvant therapy (before surgery) or as adjuvant therapy (after surgery), with postoperative treatment being more common in NSCLC (McElnay and Lim, 2014). Typical classes of chemotherapeutic drugs used in the treatment of lung cancer include antimitotic agents such as paclitaxel, platinum-based antineoplastic drugs such as cisplatin and carboplatin, and antimetabolites such as pemetrexed.

A large proportion of both SCLC and NSCLC patients are diagnosed at a late stage when therapeutic interventions are less likely to be effective (Heuvers et al., 2012). Treatment of patients diagnosed at late stage is directed towards improving patient

survival and reducing adverse effects (Zappa and Mousa, 2016). The 2017 ASCO Clinical Guideline Update recommends chemotherapy doublets of platinum-based therapy (cisplatin/carboplatin) in combination with docetaxel, paclitaxel, pemetrexed or vinorelbine as first-line therapy for NSCLC patients with negative or unknown tumour EGFR mutation or ALK/ROS1 gene rearrangement status and with performance status (PS) of 0 or 1 (Hanna et al., 2017). Platinum-based chemotherapy is also recommended for the treatment of locally advanced stage III NSCLC (Postmus et al., 2017). The standard treatment for metastatic SCLC consists of cisplatin/carboplatin and etoposide (Fruh et al., 2013).

1.2.4.4 Anti-angiogenic agents

The induction of angiogenesis is considered one of the hallmarks of cancer (Hanahan and Weinberg, 2011). The growth of the tumour-associated vascular network not only provides oxygen and nutrients that maintain tumour growth but can also be exploited by tumour cells to enable metastasis (Carmeliet and Jain, 2011). The monoclonal antibody bevacizumab inhibits VEGF (vascular endothelial growth factor), a growth factor that stimulates angiogenesis (Hicklin and Ellis, 2005), and overexpression of VEGF has been associated with disease progression and poor prognosis in lung cancer (Fontanini et al., 1997). Improved overall response and time to progression were observed in patients with advanced disease when bevacizumab was administered in combination with established chemotherapeutic agents (Johnson et al., 2004). The addition of bevacizumab to adjuvant cisplatin-based treatment regimens provided no additional benefit in surgically resected early stage NSCLC patients (Wakelee et al., 2016).

1.2.4.5 Targeted therapy

The identification of oncogenic driver mutations and the emergence of targeted therapies, such as precision tyrosine kinase inhibitors (TKIs), allows for the matching of systemic therapies with patient/tumour-specific genetic alterations. Epidermal growth factor receptor (EGFR) is overexpressed in more than half of NSCLC patients and EGFR TKIs, such as erlotinib and gefitinib, have provided impressive clinical benefits in patients with activating EGFR mutations (Mok et al., 2009; Rosell et al., 2012). The majority of patients develop resistance to EGFR TKIs and progress after a median of 10 to 14 months (Oxnard et al., 2011). Mechanisms of acquired resistance include tumour cells compensating for blockade of EGFR signalling pathways through the activation of alternative pathways (Ohashi et al., 2013). For example, MET overexpression can drive ERBB3-dependent activation of the PI3K/AKT pathway, abrogating the reliance upon activating *EGFR* mutation (Engelman et al., 2007). The most frequent mechanism of secondary resistance, however, is via the acquisition of the T790M mutation within exon 20 of *EGFR* (Blakely and Bivona, 2012; Pao et al., 2005; Yu et al., 2013). The use of irreversible EGFR inhibitors can potentially overcome this mechanism of resistance (Kobayashi et al., 2005). Indeed, the third-generation drug osimertinib is an irreversible EGFR tyrosine kinase inhibitor that is selective for both inhibition-sensitizing mutations and the T790M resistance mutation (Jänne et al., 2015) and is recommended for use in patients that develop T790M-mediated resistance (Goss et al., 2016; Yang et al., 2017).

Chromosomal rearrangement in the anaplastic lymphoma kinase (*ALK*) gene has been identified as an oncogenic driver in NSCLC (Rikova et al., 2007; Soda et al., 2007). Patients carrying *ALK* gene fusions are treated with crizotinib as first-line treatment

(Novello et al., 2016). Other potentially targetable oncogenic drivers include *KRAS*, *BRAF* and *ERBB2* mutations and *RET* and *ROS1* amplifications (Cardarella et al., 2013; Davies et al., 2012; Gadgeel, 2016; Mazières et al., 2013; Wang et al., 2012). An in-depth, comprehensive review of lung cancer molecular targeted therapies in non-small cell lung cancer was published in 2015 (Morgensztern et al., 2015) and is recommended for an in depth understanding of this area.

1.3 DNA methylation

DNA methylation is the most widely studied epigenetic modification (Kulis and Esteller, 2010). In adult mammalian cells, DNA methylation is predominantly confined to cytosines within CpG dinucleotides (Jones and Baylin, 2007). 5-methylcytosine is produced through the addition of a methyl group to the C_5 position of cytosine, contributed by the methyl donor S-adenosylmethionine (Bird, 2002), and catalysed by DNA methyltransferases (DNMTs) and accessory proteins such as UHRF1 (Bashtrykov et al., 2014; Daskalos et al., 2011). The *de novo* DNA methyltransferases, DNMT3A and DNMT3B, are essential for the establishment of DNA methylation patterns in early development (Okano et al., 1999), while the so-called maintenance DNA methyltransferase, DNMT1, is mainly responsible for the addition of methyl groups to the unmethylated, nascent daughter strand during DNA replication and has a preference for the hemimethylated DNA (Hermann et al., 2004). It is now apparent, however, that DNMT1 is not solely responsible for methylation preservation and DNMT3A/3B play a vital supporting role in enabling the efficient propagation of DNA methylation patterns (Jones and Liang, 2009). The catalytically inactive homologue of Dnmt3a/3b, Dnmt3L, interacts with Dnmt3a and Dnmt3b and

facilitates their action in *de novo* methylation and the establishment of maternal imprinting in mouse oocytes (Hata et al., 2002).

Spontaneous and/or enzymatic deamination of 5-methylcytosine to thymine induces a greater than 75% under-representation of CpG dinucleotides in the human genome (Lander et al., 2001). Epigenetic methylation marks can directly confer genetic alterations through this mechanism (Pfeifer, 2010; Rideout et al., 1990). Relatively short, predominantly unmethylated CpG-rich regions with high G + C content, termed CpG islands (CGIs), are interspersed throughout the mammalian genome, primarily coinciding with gene promoters (Bird et al., 1985; Cooper et al., 1983; Meissner et al., 2008; Weber et al., 2007). In practice, CGIs are predicted computationally and characterized as being longer than 200 bases, presenting at least a content of 50% of guanines and cytosines and an observed-to-expected CpG ratio greater than 0.6 (Wu et al., 2010). Actively transcribed genes that contain CGIs within their promoters normally have nucleosome-depleted regions at their transcription start sites, typically with adjacent H2A.Z-containing nucleosomes bearing the H3K4 trimethylation mark (Kelly et al., 2010). Inactive CGI-containing promoters do not commonly gain DNA methylation but acquire the repressive H3K27 trimethylation mark via a Polycomb-facilitated mechanism (Lynch et al., 2012; Tanay et al., 2007). Interestingly, in experiments using a colon cancer cell model system, histone deacetylase inhibitors (HDACi) transiently reactivated expression of methylated genes, with repression returning after two weeks, indicating that DNA methylation does not permanently silence gene expression but acts as a long term epigenetic molecular marker for gene silencing memory (Raynal et al., 2012).

1.3.1 DNA methylation and cancer

Cancer cells are typified by genome-wide DNA hypomethylation, site-specific aberrant hypermethylation of gene promoters, changes in covalent histone modifications and dysregulated expression of histone-modifying enzymes and chromatin-remodelling catalytic subunits (Portela and Esteller, 2010; Sharma et al., 2010). DNA methylation at gene promoters is generally associated with condensed chromatin structure and long-term stabilization of transcriptional repression (Jones, 2012) and is involved in various physiological processes, such as cell differentiation and development (Reik, 2007), X-chromosome inactivation (Reik and Lewis, 2005), gene imprinting (Kacem and Feil, 2009), alternative splicing (Lev Maor et al., 2015) and genomic stability (Daskalos et al., 2009). DNA methylation also occurs at regions of lower CpG density that flank CGIs, typically within c. 2 kb, known as CpG shores, and has also been demonstrated to strongly associate with silencing of gene expression (Irizarry et al., 2009). DNA methylation abnormalities have been demonstrated in human diseases including neurodegeneration, inflammation, metabolic syndromes and cancer (Portela and Esteller, 2010).

Hypermethylation of promoter regions can contribute to the promotion of carcinogenesis through the silencing of tumour suppressor genes (TSG) (Laird and Jaenisch, 1996). The tumour suppressor protein p16INK4A is encoded by *CDKN2A* and inhibits cyclin-dependent kinases CDK4 and CDK6, performing a crucial role in cell cycle arrest and senescence (Hara et al., 1996). *CDKN2A* is subject to transcriptional silencing in lung cancer, principally via aberrant hypermethylation (Merlo et al., 1995). Human bronchial epithelial cells demonstrated invasive characteristics when *RASSF1* was silenced, and suppression of invasive phenotype

was observed to be a result of RASSF1A-mediated activation of the GEF-H1/RhoB pathway and downstream inhibition of YAP (Dubois et al., 2016). Hypermethylation of the *RASSF1* promoter is frequently observed in lung cancer (Dammann et al., 2000). The tobacco smoke carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (nicotine-derived nitrosamine ketone, NNK) readily induces lung tumours in mice (Castonguay et al., 1983) and promotes methylation of *RARB* in tumours and preneoplastic hyperplasia (Vuillemenot et al., 2004). The cellular functions of these and other genes investigated in this thesis are presented in **Table 1.3-1**. Furthermore, a promoter methylation signature demonstrated prognostic utility in stage I NSCLC, distinguishing patients with high-risk disease and shorter relapse-free survival from patients with low-cancer disease (Sandoval et al., 2013a).

Table 1.3-1 Genes investigated in this thesis and their functions

Gene	Functional role	Reference
ACBC1	Drug efflux pump, multidrug resistance	(Higgins, 2007)
CDKN2A	Cell cycle arrest	(Hara et al., 1996)
CYGB	Globin protein, function not conclusively known	(Burmester et al., 2002)
F2R	Thrombin/GPCR signalling	(Coughlin, 2000)
HOXA1	Transcription factor, embryonic development	(Krumlauf, 1994)
HOXA10	Transcription factor, embryonic development	(Krumlauf, 1994)
MT1G	Metal homeostasis and detoxification	(Fu et al., 2013)
PITX2	Transcription factor, embryonic development	(Quentien et al., 2006)
RARB	Retinoic acid signalling, transcriptional regulator	(Liu et al., 1996)
RASSF1	Ras signalling	(Ortiz-Vega et al., 2002)
SHOX2	Transcription factor, embryonic development	(Clement-Jones et al., 2000)
TERT	Telomerase maintenance, immortalization	(Baird, 2010)
WT1	Transcription factor, embryonic development	(Wilm and Muñoz-Chapuli, 2016)

1.3.2 DNA methylation as a cancer biomarker in body fluids

DNA methylation has been observed in almost all cancer types (Baylin and Jones, 2011). Of particular relevance, epigenetic changes occur early in cancer development and have the potential to provide effective diagnostic biomarkers (Feinberg et al., 2006). DNA methylation is therefore an attractive target for the potential early diagnosis of cancer. It also possesses a number of additional favourable attributes for biomarker development. DNA methylation patterns are preserved through post-sampling processes, since DNA is a very stable biomolecule and methylation is a covalent modification. *In vivo* changes to CpG methylation necessitate a DNA replication step. It is, therefore, more robust to changes brought about by environmental shocks which cells may go through prior to freezing, as these most likely induce cell cycle arrest. Finally, DNA methylation signatures are present across broad regions of DNA, unlike point mutations, facilitating greater flexibility in assay design.

The presence of different methylation profiles in tumour and non-tumour cells resulting from cancer-specific hypermethylation has potential as a cancer biomarker. DNA methylation can be readily detected in biological fluids of lung cancer patients using various laboratory techniques (reviewed in Liloglou et al., 2014). Sufficiently sensitive and specific detection of abnormal DNA methylation in body fluids obtained through moderately invasive means, such as bronchoalveolar lavage or pleural effusion; minimally invasive plasma or serum liquid biopsy collected by venepuncture; or non-invasive specimens, such as sputum, could be of great clinical benefit in the time-effective and economical diagnosis of lung cancers. This would also greatly reduce discomfort and risk to patients compared to standard tissue biopsy.

1.3.3 Quantitative Methylation-specific PCR

Historically, methylation-specific PCR (MSP) has been the most commonly used technique in the detection of DNA methylation (Herman et al., 1996). This approach utilizes bisulfite-mediated hydrolytic deamination and subsequent alkaline desulphonation of cytosine to uracil in a single-stranded DNA context. 5-methylcytosine is resistant to this conversion and remains unchanged effectively translating an epigenetic DNA modification into a change in base sequence that can be detected to discriminate between methylated and unmethylated DNA. It should be noted that bisulfite-converted DNA is no longer complementary and MSP assays are designed to target one of the two strands. Bisulfite conversion also reduces sequence complexity and loss of input DNA can occur during preparation and purification. Methylation-specific primers containing guanines (R primer) and cytosines (F primer) at or proximal to the 3' end of each oligonucleotide specifically anneal to conserved sequence variants (resulting from non-conversion of 5-methylcytosine): amplicons will only be generated from methylated regions of interest under optimal PCR conditions (Herman et al., 1996). Amplification products are subsequently analyzed by agarose gel electrophoresis.

The original endpoint variation of MSP, although of great utility when developed, provides only qualitative or semi-quantitative outputs, lacks quantification efficiency and can often provide challenges regarding analytical sensitivity and specificity. The real-time PCR variant, quantitative methylation-specific PCR (qMSP or MethyLight) was developed and introduced by Peter Lairds's group (Eads et al., 2000). The most frequently adopted strategy uses two methylation-specific primers with the addition

of a methylation-specific fluorescent hydrolysis probe (**Figure 1.3-1**), overcoming many of the limitations associated with endpoint MSP.

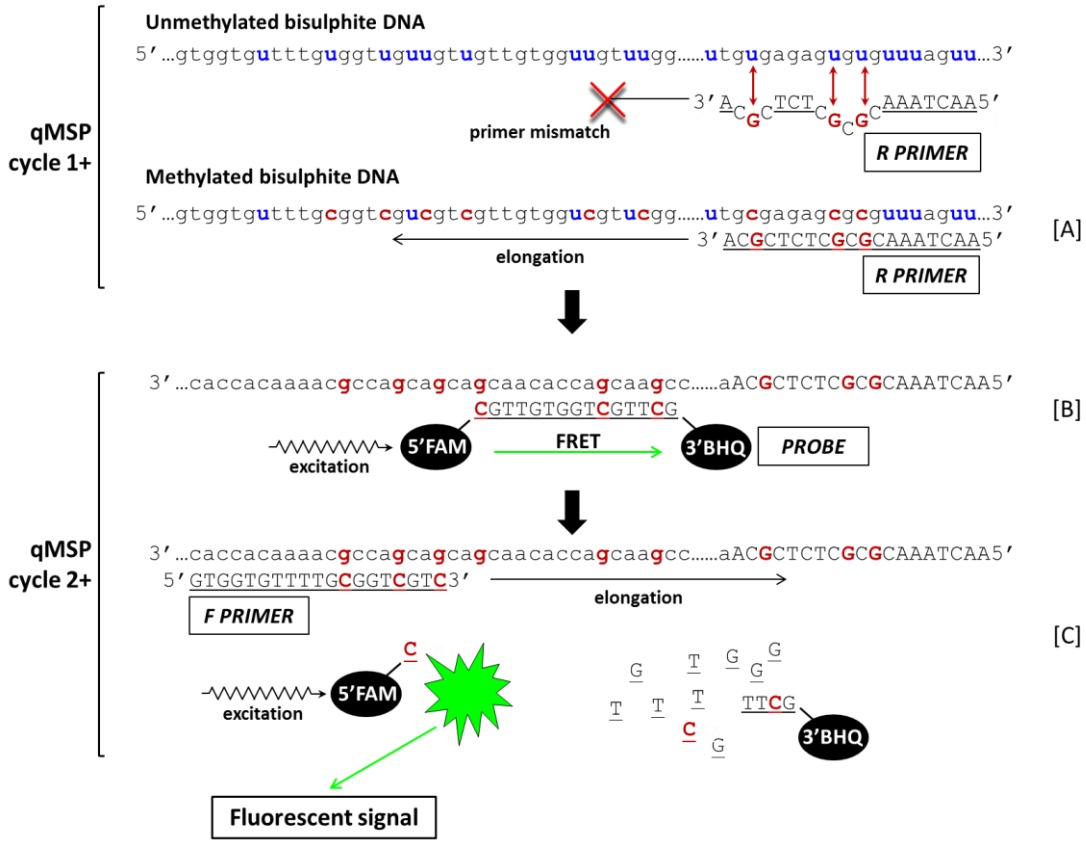


Figure 1.3-1 Methylation quantification by qMSP
u=uracil**, c=5**-methylcytosine
Methylation-specific R-primer anneals to 3'end of region of interest and Taq polymerase synthesizes complementary strand [A]. Methylation-specific probe incorporating 5'fluorescent dye and 3'quencher hybridizes to copied strand [B]. F-primer anneals to 3'end of copied sequence. Polymerase extends nascent strand and 5'exonuclease activity cleaves probe. Reporter dye is released from FRET quenching of fluorescence [C], generating increased fluorescent signal.

This methodology results in highly specific detection of DNA methylation that is both analytically sensitive and reproducible, detecting one methylated molecule in a background of 10,000 unmethylated molecules (Eads et al., 2000). qMSP enables rapid high throughput screening of hundreds to thousands of clinical specimens in one PCR step without requiring further analysis (Candiloro et al., 2011). Methylation

biomarkers from multiple loci can be analyzed rapidly using relatively small quantities of DNA (de Fraipont et al., 2005). Multiplexing with methylation independent internal controls to normalize for DNA input facilitates relative quantification of methylation (Weisenberger, 2005). The measurement of standard dilutions in concert with clinical specimens calibrates run-to-run and operator reproducibility for Quality Control and validation assessment (Tost, 2016). It is suitable for analysis of FFPE tissue samples, microdissected specimens, plasma and other body fluids (How Kit et al., 2012). However, qMSP is not without its disadvantages. DNA methylation is not quantified at the individual CpG dinucleotide level and only one DNA strand is targeted (Fraga and Esteller, 2002). Heterogeneous DNA methylation may not be reliably detected as qMSP primers and probes are designed to detect fully methylated DNA (Alnaes et al., 2015; Claus et al., 2012).

1.3.4 Detection of DNA methylation by qMSP in clinical samples

A multitude of qMSP cancer biomarker assays are in development for screening, risk stratification, diagnosis, prognosis and treatment response employing different sample types. I will here focus on a select number of studies utilizing clinical samples and adhering to guidelines for clinical biomarker development and validation or used in clinical trials.

A recent multicenter prospective cohort study sought to demonstrate that DNA methylation levels could predict risk of metachronous gastric cancer after endoscopic resection. DNA from 826 non-cancerous gastric mucosal samples was analyzed by qMSP and patients were categorized into quartiles corresponding to degree of methylation. Multivariate analysis adjusting for other risk factors confirmed that high

miR-124a-3 methylation was associated with increased risk of metachronous gastric cancer development (Asada et al., 2015).

Methylation-specific PCR *MGMT* gene promoter analysis of FFPE tissue showed within a phase II trial that clinical response of colorectal cancer patients to dacarbazine was restricted to the patient group exhibiting *MGMT* hypermethylation, thus characterizing a subgroup of patients for whom treatment with alkylating agents may be beneficial (Amatu et al., 2013).

A recent prospective study of 1544 participants evaluated the utility of colorectal cancer methylation biomarker *SEPT9* in blood plasma. The limit of detection was 7.8pg/mL for the specific assay (Epi proColon) employed in this study. The clinical sensitivity was 68% and specificity was 80% based on diagnosis by colonoscopy (Potter et al., 2014).

SHOX2, a candidate clinical biomarker for the diagnosis of lung cancer, achieved 68% clinical sensitivity at 95% clinical specificity in bronchial aspirates from 281 malignant lung cancer cases and 242 controls with benign lung diseases. 62% of aspirate cases classified as negative by routine cytology were positively identified by *SHOX2* methylation (Dietrich, 2011). This CE IVD (Conformité Européene *in vitro* Diagnostic) certified assay uses the HeavyMethyl variant of qMSP which facilitates specific amplification of bisulfite converted methylated sequences by employing methylation-specific oligonucleotides that block annealing of methylation-insensitive primers. Further validation was performed on a 250 patient-case control cohort. Analysis of six samples from the previous study with methylation levels close to the

cut-off providing 95% specificity informed the clinical decision point for this performance analysis. Reproducible detection was demonstrated at 0.8% methylation and clinical sensitivity (78%) and specificity (96%) were marginally superior (Kneip et al., 2011).

The use of panels of multiple methylation biomarkers can potentially increase clinical sensitivity and provide better diagnostic accuracy. A large retrospective case-control study comprising 655 bronchial lavage specimens used qMSP to evaluate a panel of ten lung cancer markers. Using a robust positive-control-determined dichotomizing cut-off of 0.5% methylation (c. 60 pre-bisulfite copies) and mathematical modelling to identify the optimal combination of markers, a combination of four genes (*TERT*, *WT1*, *CDKN2A* and *RASSF1*) correctly predicted lung cancer status in 85.9% of samples assayed. Overall, clinical sensitivity was 82% with 91% clinical. Importantly, this panel detected greater than 50% of T1 tumors (Nikolaidis et al., 2012).

1.3.5 Droplet Digital PCR

Droplet Digital PCR (ddPCR) partitions DNA molecules across c. 20,000 replicate reactions, using microfluidics and surfactant chemistry (Hindson et al., 2013). Due to random sampling effects, some reactions will contain one or more target copies and others will contain no template. Reactions are amplified by PCR to end-point and analysed by an automated flow-cytometer. Reactions containing amplified template molecules are designated positive on the basis of fluorescence amplitude and those without template, and hence lower amplitude, are discriminated as negative. The number of target molecules present can be calculated from the

fraction of positive droplets using the following equation, based on the Poisson distribution:

$$\lambda = -\ln(1-p)$$

The average number of target molecules per droplet is represented by λ and p represents the proportion of positive droplets. Using the average number of target molecules, the volume of the ddPCR partitions and the number of droplets successfully analysed we can estimate the absolute concentration of the target of interest. Since ddPCR depends on a binary classification of end-point PCR, variation in amplification efficiency is well tolerated. Furthermore, partitioning of "bulk" reactions into many thousands of individual reactions increases the tolerance to inhibitory substances (Dingle et al., 2013). ddPCR also uses the same chemistries widely employed in qPCR applications.

1.3.6 Detection of DNA methylation by ddPCR

Prior to this project, there were no peer-reviewed publications utilizing ddPCR for the detection of DNA methylation in circulation. However, a small number of studies have now been reported. MethyLight ddPCR showed higher sensitivity than qMSP in the detection of methylated *EVL* and *NTRK3* in a technical performance evaluation and demonstrated improved precision in primary colorectal cancer tissue (Yu et al., 2015). After identifying a number of potential biomarkers for the detection of breast cancer using methylation array analysis, a four methylation marker panel (in conjunction with two additional modelling parameters) achieved a sensitivity of 86% and specificity of 83% in case and control plasma samples

(Uehiro et al., 2016). The detection of methylated *WIF1* and *NPY* DNA markers in the tissue and plasma of colorectal cancer patients has also been demonstrated using a similar droplet-based digital platform, the RainDrop Digital PCR System (Garrigou et al., 2016).

1.4 Study aims and objectives

The aim of this study was to evaluate DNA methylation biomarkers in clinical samples to facilitate early detection of lung cancer. Specific objectives included:

(1) The validation of an extended marker panel, using enhanced assays, in moderately invasive bronchoalveolar lavage.

(2) The development and optimisation of novel ddPCR assays capable of detecting very low levels of DNA methylation in minimally invasive liquid biopsy samples such as blood plasma, with the intention of replacing invasive surgical/clinical procedures with a simple blood draw.

(3) The comprehensive assessment and comparison of ddPCR and qMSP technologies in the detection of moderate and very low quantities of methylated DNA to ensure reliability in analysis of clinical samples.

(4) The validation of DNA methylation assays and detection models using the best performing methodology in blood plasma samples in a retrospective case-control study.

Chapter 2

Materials and methods

2.1 Sample collection

2.1.1 Peripheral blood mononuclear cell (PBMC) control DNA

Genomic DNA previously extracted from many (> 96) non-cancer blood bank PBMC samples was pooled and split into two 50 ml polypropylene centrifuge tubes (Corning, Corning, NY, USA). RNase A (Sigma-Aldrich, St. Louis, MO, USA) was added to a final concentration of 50 µg/ml and incubated at 37 °C for 1 hour. Proteinase K was then added to a final concentration of 100 µg/ml and incubated at 56 °C for 1 hour. One volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the pooled DNA solution and thoroughly mixed by inversion for 30 seconds. The tubes were centrifuged at 3900 RCF for 10 minutes and the aqueous layer removed to a new centrifuge tubes (equal volumes in each). 0.5 volumes 7.5 M NH_4OAc was added each tube, followed by 2.5 volumes 100% ethanol (molecular biology grade). DNA was precipitated overnight at -20 °C. Samples were centrifuged at 3900 RCF for 10 minutes at 4°C and supernatant removed. 1 ml 70% ethanol was added to the tubes which were then centrifuged at 3900 RCF for 10 minutes at 4 °C and supernatant removed. The 70% ethanol wash was repeated, removing as much remaining ethanol as possible. The DNA pellets were then dried at room temperature for 10 minutes. 2 ml TE buffer was added to each tube. Solutions were left at 4 °C overnight to allow for rehydration of the DNA. Purity was assessed by ultraviolet spectrophotometry (see **2.3.1**) and concentration estimated by fluorescence quantification (see **2.3.2**). PBMC DNA was subsequently diluted to 100 ng/µl with TE buffer, aliquoted and stored at -20 °C.

2.1.2 Bronchoalveolar lavage specimen collection

Bronchoalveolar lavage samples were collected at participating hospitals according to clinical SOPs and a proportion (excess to diagnosis) was provided to LLP in Saccomano's fixative and stored in a temperature-controlled environment (18°C) until use.

2.1.3 Blood plasma specimen collection

Blood samples were collected and plasma isolated according to the SOP located in the Appendix.

2.2 DNA extraction

2.2.1 Bronchoalveolar lavage (BAL) DNA extraction

2 ml aliquots of bronchoalveolar lavage specimens were transferred into 2.0 ml SuperLock microcentrifuge tubes (STARLAB (UK) Ltd., Milton Keynes, UK). 15 µl 15 % dithiothreitol was added to each sample, mixed using a vortex mixer and incubated at room temperature for 30 minutes. Sample tubes were centrifuged at 14,000 RCF for 15 minutes at 4 °C. Supernatants were discarded and genomic DNA extracted from pelleted samples was extracted using DNeasy 96 Blood and Tissue Kit (Qiagen). Buffers ATL and AL were first inspected for the formation of precipitates. If precipitation was suspected, buffers were warmed to 56 °C for 5 – 10 minutes. A working lysis solution was prepared containing 180 µl Buffer ATL and 20 µl proteinase K (activity 600 mAU/ml solution) and mixed by pulse vortexing. 200 µl working solution was added to each sample, tubes securely closed and incubated at 56 °C for 16 hours with constant agitation. Lysate was collected from microcentrifuge tube caps by brief centrifugation. Complete sample lysis was confirmed by visual

inspection. Gelatinous or very viscous samples were incubated for a further six hours after the addition of 20 µl proteinase K. 410 µl premixed Buffer AL and ethanol (260 ml ethanol to 247 ml Buffer AL) was added to each sample. Samples were then mixed by pulse vortexing and briefly centrifuged to collect any solution form tube lids. Samples were transferred to DNeasy 96 plates placed on top of waste collection S-Blocks. The plates were sealed with AirPore Tape Sheet (Qiagen), to prevent cross-contamination during centrifugation, and centrifuged for 10 minutes at 5796 RCF. Flow-through was discarded, 500 µl Buffer AW1 added to sample columns, followed by centrifugation at 5796 RCF for 10 minutes. If the lysate had not passed through the membrane in all wells, plates were centrifuged for a further 10 minutes. Flow-through was discarded, 500 µl Buffer AW2 added to sample columns, followed by centrifugation at 5796 RCF for 15 minutes. The plates were placed on an Elution Microtubes RS (Qiagen) racks and incubated at 70 °C for 10 minutes to allow evaporation of any traces of Buffer AW2. The plates were placed on clean Elution Microtubes RS racks, 200 µl 0.1 x TE was added to the column matrix directly and the column left to stand for 10 minutes to enable greater solubilisation of matrix-bound DNA. DNA was eluted by centrifugation at 5796 RCF for 2 minutes. To increase yield, the eluate was reloaded into columns and centrifugation repeated. BAL DNA was either analysed immediately or stored at -20 °C.

2.2.2 Blood plasma DNA extraction

cfDNA was extracted from 1.8 ml blood plasma using DNeasy Blood and Tissue Kit (Qiagen). The manufacturer's protocols were modified to take into consideration the larger volumes required for extraction of DNA from plasma. 1.8 ml plasma was transferred into 15 ml centrifuge tubes (Corning). A working lysis solution was

prepared containing 1.8 ml Buffer ATL and 20 µl 20 mg/ml proteinase K (Promega, Madison, WI, USA) per sample and mixed well. 1.8 ml working solution was added to each sample, tubes were tightly closed and mixed by pulse vortexing. Samples were incubated at 56 °C for 16 hours. Tubes were centrifuged briefly to collect any lysate from tube caps. 1.8 ml Buffer AL was added to samples which were mixed by pulse vortexing and incubated at 56 °C for 10 minutes. 1.8 ml molecular biology grade ethanol was added to samples which were mixed by pulse vortexing. 3 ml of the mixture was transferred into 3ml Extension Tubes (Qiagen) fitted into a DNeasy Mini spin columns on a QIAvac 24 Plus vacuum manifold and the vacuum pump switched on. As the lysate was drawn through the spin columns further lysate was added until all had passed through the column. The vacuum pump was then switched off. 750 µl Buffer AW1 was added to the columns and vacuum reapplied. When all the Buffer AW1 had passed through the columns, the vacuum pump was again switched off, and 750 µl Buffer AW2 was added to the columns and vacuum reapplied. When all the Buffer AW2 had passed through the columns, the vacuum pump was again switched off, the DNeasy Mini spin columns placed in clean 2ml collection tubes and centrifuged at 16,000 RCF for one minute in order to dry the column membrane and prevent carry-over of Buffer AW2. DNeasy Mini spin columns were placed in 1.5 ml DNase-free microcentrifuge tubes (STARLAB (UK)), 110 µl 0.22 µm-filtered Tris-HCl, pH 8.0, 0.025% Tween-20 was added to the column matrix directly and the column left to stand for 10 minutes to enable greater solubilisation of matrix-bound DNA.

DNA was eluted by centrifugation at 12,000 RCF for 2 minutes. To increase yield, the eluate was reloaded into columns and centrifugation repeated. 10 µl plasma

cfDNA was aliquoted into PCR strips for real-time PCR genomic DNA quantification (see **2.3.3**), and stored at -20 °C. The remaining volume of plasma cfDNA was transferred to PCR strips for MethPlex ddPCR analysis. These DNA samples were dehydrated using a Genevac miVac DNA concentrator (Genevac Ltd., Ipswich, UK), rehydrated in 20 µl 0.22 µm-filtered Tris-HCl, pH 8.0, 0.025% Tween-20 at 4 °C overnight and stored at -20 °C.

2.3 DNA quantification and Quality Control

2.3.1 Ultraviolet spectrophotometry

DNA concentration and purity were assessed using a NanoDrop™ 2000 (Thermo Scientific, Waltham, MA, USA) following the manufacturer's instructions. Concentration was based on absorbance at 260 nm and the concentration extinction coefficient was set to 50 ("Type" = DNA-50). Purity was assessed by 260/280 and 260/230 ratios, and general observation of absorbance spectra.

2.3.2 Fluorescence quantification

PBMC DNA stocks and plasmid DNA samples were quantified using Quant-iT™ Broad-Range dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA). PBMC DNA standards used in the development and validation of *ACTB* qPCR genomic DNA quantification assay (see **2.3.3**) were quantified using Qubit® High-Sensitivity dsDNA Assay Kit (Invitrogen).

2.3.2.1 Quant-iT™ Broad-Range dsDNA Assay Kit

A working solution was made by diluting Quant-iT dsDNA BR Reagent 1:200 in Quant-iT dsDNA BR Buffer and mixing using a vortex mixer. 200 µl was pipetted

into the appropriate number of wells of 96 Well Black FLUOTRAC™ 200 Polystyrene Microplate (Greiner Bio-One International GmbH, Kremsmünster, Austria). 10 µl of each λ DNA standard (0, 5, 10, 20, 40, 60, 80 and 100 ng/µl) was added to individual wells in triplicate. 5 µl of DNA of unknown concentration was added to appropriate wells in triplicate. All samples were mixed well by thoroughly pipetting up and down. Fluorescence was read using a Tecan GENios™ plate reader (Tecan Group Ltd., Männedorf, Switzerland) using a 485 nm excitation filter and a 535 nm emission filter. A calibration curve was constructed in Microsoft Excel 2010 (Microsoft Corporation, Redmond, Washington, USA) by plotting λ DNA standard input amount as the independent variable and background-corrected mean fluorescence as the dependent variable. Unknown DNA amounts were calculated by solving the linear regression equation of a fitted straight line forced through the origin for x = background-corrected mean fluorescence of unknown DNA. Unknown DNA concentration (ng/µl) was calculated by dividing this number by the volume of added unknown DNA and multiplying by any appropriate dilution factor.

2.3.2.2 Qubit® High-Sensitivity dsDNA Assay Kit

A dilution series of dsDNA standards was constructed by dilution of Qubit® dsDNA HS Standard #1 (0 ng/µl in TE buffer) in 0.22 µm-filtered Tris-HCl, pH 8.0, 0.025% Tween-20 according to Table 2.3-1. A working solution was made by diluting Qubit® dsDNA HS Reagent 1:200 in Qubit® dsDNA HS Buffer and mixing using a vortex mixer. 10 µl of each dsDNA standard (0, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 1.5, 2 and 5 ng/µl) was added to individual wells in triplicate. 3 µl of PBMC DNA of unknown concentration was added to appropriate wells in triplicate. From this point

forward the protocol was the same as that for Quant-iT™ Broad-Range dsDNA Assay Kit (see **2.3.2.1**).

Table 2.3-1 Table of High Sensitivity dsDNA dilution series concentrations and volumes

Concentration (ng/µl)	Dilution	Volume previous dilution (µl)	Volume Tris-HCl, pH 8.0, 0.025% Tween-20 (µl)	Mixed Volume (µl)
10	-	-	-	-
5	1/2	125	125	250
2	2/5	80	120	200
1.5	3/4	135	45	180
1	2/3	100	50	150
0.5	1/2	75	75	150
0.2	2/5	60	90	150
0.1	1/2	75	75	150
0.05	1/2	75	75	150
0.02	2/5	60	90	150
0.01	1/2	75	75	150

2.3.3 Real-time PCR genomic DNA quantification

I developed a highly sensitive quantitative real time genomic DNA assay targeting the *ACTB* gene. This was prompted by the expectation of low circulating cell-free DNA yields from blood plasma. The final assay used for quantification of DNA extracted from patient plasma samples used duplicate seven step, 3-fold standard curves of PBMC DNA, ranging from 7.29 ng/µl to 0.01 ng/µl. Quantification assay runs also included duplicate non-template controls.

Please refer to Chapter 4 “ddPCR assay and workflow development and optimisation” for full assay details.

2.4 *In vitro* genomic DNA methylation

CpG dinucleotide methylation positive control DNA for methylation detection method development, optimisation and clinical screening was produced by *in vitro* methylation of pooled blood bank PBMC DNA using M.SssI CpG methyltransferase (New England Biolabs (UK), Hitchin, UK). Reactions were carried out in PCR tubes containing 5 µg genomic DNA, 1 x NEBuffer 2, 640 µM S-adenosyl methionine (SAM), 20 U M.SssI CpG methyltransferase and ddH_2O in 100 µl total volume. Reactions were thoroughly mixed by pipetting up and down and incubated at 37 °C using an Applied Biosystems GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA). Since SAM is consumed during the reaction, an additional 2 µl 32 mM SAM was added after two hours and the reaction was incubated for a further two hours followed by heat inactivation at 65 °C for 20 minutes. Discrete reactions were combined, mixed and aliquoted, and the resulting 100 ng/µl *in vitro* methylated genomic DNA was stored at -20 °C.

2.5 Whole genome amplification

DNA methylation demonstrates tissue specificity (Chen et al., 2016; Ghosh et al., 2010; Schilling and Rehli, 2007); therefore, we cannot assume that any given gene promoter in the various subpopulations of PBMCs and various individual donors of control specimens is unmethylated. Whole genome amplification (WGA) using REPLI-g® Screening Kit (Qiagen) was carried out to produce fully unmethylated control DNA. REPLI-g® Mini DNA Polymerase was thawed on ice. 17 µl Buffer SB1 was added to 50 ng PBMC DNA (5 µl), mixed by vortexing and centrifuged briefly. DNA was denatured by incubation at 65 °C for 5 minutes and allowed to cool to room temperature. A master mix containing Buffer SB2 and REPLI-g® Mini

DNA Polymerase was prepared in a 17:1 ratio. 18 µl master mix was added to each denatured DNA sample and WGA reactions were incubated at 30 °C for 16 hours, followed by enzyme inactivation at 65 °C for 3 minutes. All incubation steps were carried out using an Applied Biosystems GenAmp PCR System 9700 (Applied Biosystems). WGA DNA was stored at -20 °C.

2.6 HinP1I restriction digestion of genomic DNA

To increase specificity of methylation-specific qMSP/ddPCR assays, genomic DNA was first digested with methylation-sensitive restriction endonuclease HinP1I prior to bisulphite conversion (see Chapter 4). Reactions contained DNA (varying input amounts), 1 x CutSmart® Buffer, 10 U HinP1I (NEB (UK)) and ddH_2O in a 100 µl total volume. Reactions were thoroughly mixed by pipetting up and down and incubated at 37 °C for 16 hours using an Applied Biosystems GeneAmp PCR System 9700 (Applied Biosystems). Digestion was stopped by heat inactivation at 65 °C for 20 minutes. DNA was either bisulphite treated immediately or stored at -20 °C.

2.7 Bisulphite conversion of DNA

Genomic DNA was subjected to bisulphite conversion of unmethylated cytosine to uracil to enable discrimination between methylated and unmethylated DNA by sequence sensitive detection methods. Single format bisulphite conversion was carried out using the EZ DNA Methylation Gold™ kit (Zymo Research, Irvine, CA, USA) and the general protocol follows, with modifications specified in appropriate chapters. CT Conversion Reagent was prepared by adding 900 µl ddH_2O to the solid conversion reagent mixture, followed by 300 µl M-Dilution Buffer and 50 µl M-

Dissolving Buffer. The resulting solution was then vortexed thoroughly and mixed on a Stuart Scientific Blood Tube Rotator SB1 (Cole-Parmer, Stone, UK) for 10 minutes. 130 µl CT Conversion Reagent was added to 20 µl DNA sample in a PCR tube, mixed by pipetting up and down, and briefly centrifuged. Samples were incubated at 98 °C for 10 minutes and 64 °C for 2.5 hours using an Applied Biosystems GeneAmp PCR System 9700 (Applied Biosystems), followed by immediate desulphonation/deamination and clean-up steps. Incubated samples were centrifuged briefly to remove any condensate from PCR tube lids, pipetted into Zymo-Spin™ IC Columns containing 600 µl M-Binding Buffer and mixed by inversion. Columns were centrifuged at 14,000 RCF for 60 seconds and the flow-through was discarded. 200 µl M-Wash Buffer was then added to columns and centrifugation repeated. 200 µl M-Desulphonation Buffer was added to columns and incubated at room temperature for 20 minutes, followed by centrifugation at 14,000 RCF for 60 seconds. The flow-through was discarded and the columns were washed twice with 300 µl M-Wash Buffer, as above. Columns were then centrifuged at 14,000 RCF for five minutes, in clean waste collection tubes, to prevent ethanol carry-over. Spin columns were placed in clean 1.5 ml DNase-free microcentrifuge tubes (STARLAB (UK)), 50 µl M-Elution Buffer was added to the column matrix directly and the column left to stand for 10 minutes to enable greater solubilisation of matrix-bound DNA. DNA was eluted by centrifugation at 14,000 RCF for 60 seconds. To increase yield, the eluate was reloaded into columns and centrifugation repeated. Bisulphite treated DNA was either analysed immediately or stored at -20 °C.

2.8 Agarose gel electrophoresis

DNA was analysed by agarose gel electrophoresis. Restriction digests and Pyrosequencing PCR products were run on 1% and 2% (w/v) agarose gels respectively. Gels were produced by dissolving the appropriate mass of agarose in 0.5 x Tris-borate-EDTA (TBE) buffer (45 mM Tris-borate, 1mM EDTA, pH 8.3) in a conical flask, facilitated by microwave heating. SafeView Nucleic Acid Stain (NBS Biologicals, Huntingdon, UK) was then added to a concentration of 1/20,000 and the melted agarose was mixed on a Stuart Scientific Orbital Shaker SO3 (Cole-Parmer) for 3 - 5 minutes, allowing sufficient cooling prior to gel casting. The gel was then poured into a casting tray, a comb was inserted and the gel was left to set completely. The comb was removed from the gel which was then placed in a Sub-Cell GT electrophoresis tank (Bio-Rad, Hercules, USA) and immersed in 0.5 x TBE buffer. 5 µl Hyperladder I (Bioline, London, UK) was loaded as a marker for restriction digest runs; 5 µl Hyperladder IV (Bioline) was used as a marker for Pyrosequencing amplicons. Pyrosequencing PCR reactions contained CoralLoad PCR Buffer (Qiagen, Hilden, Germany) and 5 µl PCR product was loaded directly into the wells. Electrophoresis was carried out at 3 V/cm (distance between the electrodes of the unit) until the marker dye had run 3/4 the length of the gel. Gels were transferred to a UVP EC3 ChemiHR 410 imaging system (Ultra-Violet Products Ltd, Cambridge, UK) and bands were visualized at wavelength of 365 nm. Image analysis was carried out using VisionWorks LS image acquisition and analysis software (Ultra-Violet Products Ltd).

2.9 Pyrosequencing Methylation Analysis

The level of methylation of *in vitro* methylated DNA was assessed by Pyrosequencing Methylation Analysis. The results of this analysis informed the consequent adjustment in dilution of methylated DNA to desired percentages/ concentrations.

2.9.1 Pyrosequencing assay design

Target gene sequences were obtained from NCBI GenBank. CpG rich regions within the promoter region 1500 bp proximal to the transcription start site and the first exon were identified using CpG Island Finder (http://cpgislands.usc.edu/)[1] to predict potential for regulation by cytosine methylation.

Sequences were manipulated *in silico* to reflect post-bisulphite treatment sequence changes. Sequences were imported into Pyromark Assay Design 2.0 (Qiagen) to facilitate assay design. Regions of interest were defined and optimal forward, reverse and Pyrosequencing primers were designed. Desirable criteria adhered to, if possible, are:

- To ensure that template size is between 80 and 120 bp in order to reduce likelihood of secondary structure formation.
- To include at least four CpG dinucleotides within the sequencing region.
- To avoid thymine runs prior to CpGs within the assay region.
- PCR primers not to include CpGs to enable amplification of both methylated and unmethylated species.

[1] This bioinformatics tool is no longer available at this URL. Alternatives include:
- http://www.bioinformatics.org/sms2/cpg_islands.html
- http://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/

- PCR primers to include at least three cytosines not in a CpG context so that assays are specific to completely bisulphite converted DNA.
- To avoid mispriming of the Pyrosequencing primer.
- To avoid potential duplex formation between the Pyrosequencing and biotinylated PCR primer.
- To avoid 3' template loops.

It is essential to also include bisulphite conversion control sites within the assayed sequence. Multiple control sites should be included if possible. Favoured conversion controls are GCA and ACA trinucleotides, subsequently converted to GTA and ATA after bisulphite conversion. Inefficient conversion is indicated by incorporation of cytosine at these sites during the Pyrosequencing reaction. Dependent upon which DNA strand the sequencing primer is complementary to, the forward or reverse primer is modified with a 5'biotin label. This facilitates the efficient isolation of the strand selected for analysis (see **2.9.3**).

Oligonucleotides were obtained from Eurofins Genomics (Eurofins Genomics GmbH, Ebersberg, Germany) and are presented in Table 4.9.1. High Purity Salt Free (HPSF) purified, lyophilized oligonucleotides were dissolved in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 50 % glycerol and stored at -20 °C.

2.9.2 Pyrosequencing PCR amplification

Bisulphite treated DNA was amplified by PCR in a final volume of 25 µl. Reactions contained 1 x CoralLoad PCR Buffer, 200 µM dNTPs , 0.125 units HotStarTaq *Plus* DNA Polymerase (Qiagen), 300nM unbiotinylated primer, 150 nM biotinylated

primer (Eurofins Genomics) and ddH_2O. PCR thermal cycling was carried out using a Life Technologies Veriti 96-well Thermal Cycler (Applied Biosystems) using the following thermal profile(s): 95 °C for 5 minutes, followed by 40 cycles of 95 °C for 30 seconds, annealing temperature of 52 °C (*MT1G*), 54 °C (*RASSF4*, *RASSF5*) or 56 °C (*CDKN2A*) for 30 seconds and extension at 72 °C for 30 seconds. Partially synthesized products present from previous amplification cycles were completed by a final elongation at 72 °C for 10 minutes (**Table 2.9-1**). Amplification of bisulphite treated DNA was confirmed by agarose gel electrophoresis (see **2.8**) prior to Pyrosequencing.

Table 2.9-1 Pyrosequencing PCR thermal cycling conditions.

	Temperature (°C)	Time	Cycles
Initial denaturation/ Polymerase activation	95	5 minutes	1
Denaturation	95	30 seconds	40
Annealing	Target specific	30 seconds	
Extension	72	30 seconds	
Final extension	72	10 minutes	1

2.9.3 Pyrosequencing PCR product immobilisation, strand separation and sequencing primer annealing

Biotinylated PCR products were first immobilised on Streptavidin Sepharose™ High Performance beads (GE Healthcare Life Sciences, Little Chalfont, UK). 75 µl binding mixture containing 3 µl Streptavidin Sepharose™ HP, 50 µl PyroMark Binding Buffer (Qiagen) and 22 µl ddH_2O was added to each PCR reaction. PCR

plates were then sealed tightly and agitated at 300 rpm for 10 minutes using a vortex mixer with microtitre plate stand. Beads were immediately captured using a 96-pin vacuum tool on a PyroMark Q96 Vacuum Workstation (Qiagen) and filter probes/attached beads washed in 70% ethanol for 10 seconds (tray 1). The vacuum tool was transferred to a trough containing 0.2 M NaOH (tray 2) in order to denature PCR products, resulting in the retention of the biotin labelled strand (bound to the isolated beads) and the loss of the unbiotinylated PCR stand. The vacuum tool was then transferred to a trough containing 10 mM Tris-acetate (tray 3) for 10 seconds to wash the isolated strand. Filter probes were then carefully positioned above a PyroMark Q96 Plate Low containing 45 µl 0.3 µM sequencing primer (Eurofins Genomics) in PyroMark Annealing Buffer (Qiagen), the vacuum pump switched off and the Streptavidin-bound biotin-labelled single-stranded DNA released into the appropriate wells. The plate was then incubated at 80 °C for 2 minutes, allowed to cool to room temperature for 5 minutes and then placed in a PyroMark Q96 ID System (Qiagen) for Pyrosequencing analysis. Enzyme, substrate and nucleotides (PyroMark Gold Q96 Reagents, Qiagen) were pipetted into the Pyrosequencing reagent cartridge according to the volumes recommended by the PyroMark Q96 Software (Qiagen). The cartridge was then loaded into the PyroMark Q96 ID System and Pyrosequencing analysis performed.

2.10 Quantitative Methylation Specific PCR (qMSP)

qMSP assays amplify and specifically quantify methylated gene promoter DNA sequences after bisulphite conversion (**1.3.3**). I previously developed multiplex qMSP assays for the screening of clinical BAL DNA in a case-control study (Chapter 3). These triplex assays utilise hydrolysis probe technology, specifically

TaqMan® MGB probes (Applied Biosystems), and each individual assay targets two potentially methylated gene promoters, probes being labelled with FAM™ and VIC™ respectively. Assays also include a methylation-independent endogenous control primers and NED™-labelled probe, targeting a sequence within the *ACTB* gene promoter. Optimal thermal cycling conditions and primer/probe concentrations for qMSP assays had been established prior to this study (**Table 2.10-2**, **Table 2.10-3**). Two gene-specific assays, however, produced optimal results in duplex assays with *ACTB* control only and were used in this format for the present study. qMSP primer and probe sequences can be found in **Table 2.10-1**.

BAL qMSP reactions contained 1 x TaqMan® Universal Master Mix II (no UNG) (Applied Biosystems), 125 nM *ACTB* primers, 125 nM *ACTB* NED™ probe, 1.5 nM methylated FAM™ and VIC™ target probes, 250 nM – 900 nM methylated target primers (**Table 2.10-2**) and 40 ng bisulphite-treated DNA. Reactions were run on StepOne Plus, 7500 FAST and 7500 real-time PCR instruments with an initial DNA denaturation and enzyme activation at 95 °C for 10 minutes and 50 cycles of 95 °C for 15 seconds and annealing/extension steps as indicated in **Table 2.10-3**.

Table 2.10-1 Oligonucleotide sequences, theoretical T_m and modifications of qMSP assay primers and probes.

Gene promoter	Primer/probe name	Sequence (5'-3')	T_m (°C)	Modification
ABCB1	ABCB1qmsp_F	GGTTTAGAGAGTAGGTTTTCGTATCGC	61.6	
	ABCB1qmsp_R	CTAAAACAACGCCCAAACCG	61.2	
	ABCB1qmsp_P	CGTTTCGGGTAATATGG	70.5	VIC-MGB
CYGB	CYGBqmsp_F	GTGTAATTTCGTCGTGGTTTGC	60.2	
	CYGBqmsp_R	CCGACAAAATAAAAACTACGCG	59.5	
	CYGBqmsp_P	TGGGCGGGCGGTAG	70.0	FAM-MGB
CDKN2A	p16qmsp_F	GGAGGGGGTTTTTTCGTTAGTATC	61.0	
	p16qmsp_R	CTACCTACTCTCCCCCTCTCCG	61.2	
	p16qmsp_P	AACGCACGCGATCC	69.7	VIC-MGB
F2R	F2Rqmsp_F	TTTTTTCGCGTCGTTCGTC	60.7	
	F2Rqmsp_R	ACCTCACCGACAAACGCG	61.7	
	F2Rqmsp_P	TTTGCGTATTTGGGAGGAG	68.6	VIC-MGB
HOXA1	HOXA1qmsp_F	TCGGGTGGTCGTAAGTACGTC	60.8	
	HOXA1qmsp_R	CCTAAAAACCCGAAACGCG	61.3	
	HOXA1qmsp_P	TAGCGCGTTTAGTCGGAA	69.0	VIC-MGB
HOXA10	HOXA10qmsp_F	GGGTTTTGAGGGCGTTTTC	60.3	
	HOXA10qmsp_R	CCTAAAAACCCGAAACGCG	61.3	
	HOXA10qmsp_P	CGGTTTCGCGTTTAC	69.0	FAM-MGB
MT1G	MT1Gqmsp_F	GTATAGCGTTTTTTTCGCGAGTC	60.1	
	MT1Gqmsp_R	AAATCCCTTAAACGCAACTAAACG	60.5	
	MT1Gqmsp_P	ACCGCAAAACGACC	69.0	FAM-MGB
PITX2	PITX2qmsp_F	GATGTTAGCGGGTCGAAGAGTC	60.3	
	PITX2qmsp_R	AAAAATCCGTACTCCTACTCCTCG	61.3	
	PITX2qmsp_P	CCTCTCCTTTCGCTCC	69.0	VIC-MGB
RARB	RARbqmsp_F	GATTGGGATGTCGAGAACGC	61.4	
	RARbqmsp_R	ACTTACAAAAAACCTTCCGAATACG	59.7	
	RARbqmsp_P	AGCGATTCGAGTAGGGT	68.0	FAM-MGB
RASSF1	RASSF1qmsp_F	GTGGTGTTTTGCGGTCGTC	61	
	RASSF1qmsp_R	AACTAAACGCGCTCTCGCA	60.6	
	RASSF1qmsp_P	CGTTGTGGTCGTTCG	68.7	VIC-MGB
TERT	TERTqmsp_F	TGGGAGTTCGGTTTGGTTTC	60.7	
	TERTqmsp_R	ACACCCTAAAAACGCGAACG	60.8	
	TERTqmsp_P	AGCGTAGTTGTTTCGGG	69.2	FAM-MGB
WT1	WT1qmsp_F	GAGGAGTTAGGAGGTTCGGTC	57.7	
	WT1qmsp_R	CACCCCAACTACGAAAACG	57.5	
	WT1qmsp_P	AGTTCGGTTAGGTAGC	65.3	FAM-MGB
ACTB	ACTBmgb_F	GGGTGGTGATGGAGGAGGTT	61.9	
	ACTBmgb_R	TAACCACCACCCAACACACAAT	60.8	
	ACTBmgb_P	TGGATTGTGAATTTGTGTTTG	69.0	NED-MGB

Table 2.10-2 Optimal BAL multiplex qMSP assay primer and probe concentrations.

Assay	Gene	F/R primer concentration (nM)	Probe concentration (nM)	Probe dye
WT1-p16	*WT1*	900	150	FAM™
	CDKN2A	900	150	VIC™
	ACTB	125	125	NED™
TERT-RASSF1	*TERT*	900	150	FAM™
	RASSF1	900	150	VIC™
	ACTB	125	125	NED™
RARB-HOXA1	*RARB*	900	150	FAM™
	HOXA1	900	150	VIC™
	ACTB	125	125	NED™
MT1G-PITX2	*MT1G*	500	150	FAM™
	PITX2	750	150	VIC™
	ACTB	125	125	NED™
HOXA10-ABCB1	*HOXA10*	900	150	FAM™
	ABCB1	900	150	VIC™
	ACTB	125	125	NED™
CYGB	*CYGB*	500	150	FAM™
	ACTB	125	125	NED™
F2R	*F2R*	250	150	VIC™
	ACTB	125	125	NED™

Table 2.10-3 Optimal BAL multiplex qMSP assay annealing/extension temperatures and times.

Assay	Stage	Temperature (°C)	Time (s)
WT1-p16	Annealing	58	90
	Extension	72	15
TERT-RASSF1	Annealing	58	90
	Extension	72	15
RARB-HOXA1	Annealing	60	90
	Extension	72	15
MT1G-PITX2	Annealing	63	90
	Extension	-	-
HOXA10-ABCB1	Annealing	58	90
	Extension	72	15
CYGB	Annealing	58	90
	Extension	72	15
F2R	Annealing	60	90
	Extension	-	-

2.11 MethPlex enrichment

2.11.1 MethPlex PCR pre-amplification

Low cfDNA yields from blood plasma made it necessary to develop a pre-amplification strategy to enrich for methylated gene promoters within my prospective biomarker panel. Refer to Chapter 6 for an exposition of methodological development. The pre-amplification method entailed methylation-specific multiplex PCR amplification of bisulphite-converted DNA, targeting eight gene promoters of interest, a process that I have termed "MethPlex enrichment". MethPlex methylation-specific outer nested primers were designed to pre-amplify bisulphite-treated DNA external to methylation-specific ddPCR amplicons and are presented in **Table 2.11-1**. The final procedure utilised a hot start DNA polymerase and buffer optimised for use

with bisulphite-converted DNA, EpiMark® Hot Start Taq DNA Polymerase (NEB UK). A shortened initial denaturation of 30 seconds at 95°C was used to avoid further damage to bisulphite-converted DNA. PCR reactions contained 1 x EpiMark® Buffer, 1.25 units EpiMark® Hot Start Taq DNA Polymerase (NEB UK), 200 µM dNTPs (Qiagen), varying concentrations (**Table 2.11-2**) of MethPlex enrichment primers (Sigma-Aldrich), 10 ng bisulphite-treated DNA (technical experiments and screening controls) or 20 µl bisulphite-treated DNA (clinical screening), and ddH_2O in a total volume of 50 µl.

Table 2.11-1 MethPlex oligonucleotide sequences and properties

Gene promoter	Primer name	Sequence (5'-3')	Length	T_m (°C)
ABCB1	ABCB1nest_F	TATTTATCGTTCGCGGGTT	19	56.7
	ABCB1nest_R	AACTCAAAAACTCCTAAAACAACG	24	56.8
F2R	F2Rnest_F	GTGGAATAGGAAGAGCGGTT	20	56.8
	F2Rnest_R	CAAACCATTTCTCTCCTAAACG	22	56.6
MT1G	MT1Gnest_F	TGGTAGGGTGAGAGAAGTCG	20	56.3
	MT1Gnest_R	AATACAAAATCCCTTAAACGCA	22	56.5
CDKN2A	p16nest_F	AGAGGATTTGAGGGATAGGGT	21	56.6
	p16nest_R	GTAACCAACCAATCAACCGA	20	56.8
RASSF1	RASSF1nest_F	TTAGAAATACGGGTATTTTCGC	22	56.2
	RASSF1nest_R	AAAAACTATAAAACCCGAAAACG	23	56.6
SHOX2	SHOX2nest_F	GTTTTGTTGGTAGAGGTTGAGC	22	56.5
	SHOX2nest_R	GAACTACTACGATCGTCGCG	20	57.0
TERT	TERTnest_F	TTTTGGGCGTTTGTGTTC	18	56.5
	TERTnest_R	CCTACACCCTAAAAACGCG	19	56.8
WT1	WT1nest_F	TTTAAATAAGAGGGGTCGGC	20	56.8
	WT1nest_R	ACACCGACCAACTAAAAACG	20	56.2

Table 2.11-2 MethPlex primer stocks
Table indicating final MethPlex enrichment reaction primer concentrations, 20 x stock concentrations and volumes of individual 100 µM oligonucleotides required to produce 20 x stocks. 2.5 µl 20 x MethPlex primer stock was used per 50 µl MethPlex enrichment reaction.

Gene promoter	Primer	Final concentration	20 x stock concentration	Volume/250 µl 20x stock
ABCB1	ABCB1nest_F	200 nM	4 µM	10 µl
	ABCB1nest_R	200 nM	4 µM	10 µl
F2R	F2Rnest_F	200 nM	4 µM	10 µl
	F2Rnest_R	200 nM	4 µM	10 µl
MT1G	MT1Gnest_F	200 nM	4 µM	10 µl
	MT1Gnest_R	200 nM	4 µM	10 µl
CDKN2A	p16nest_F	100 nM	2 µM	5 µl
	p16nest_R	100 nM	2 µM	5 µl
RASSF1	RASSF1nest_F	200 nM	4 µM	10 µl
	RASSF1nest_R	200 nM	4 µM	10 µl
SHOX2	SHOX2nest_F	100 nM	2 µM	5 µl
	SHOX2nest_R	100 nM	2 µM	5 µl
TERT	TERTnest_F	200 nM	4 µM	10 µl
	TERTnest_R	200 nM	4 µM	10 µl
WT1	WT1nest_F	200 nM	4 µM	10 µl
	WT1nest_R	200 nM	4 µM	10 µl
ddH_2O	-	-	-	110 µl

PCR thermal cycling was carried out using a Life Technologies Veriti 96-well Thermal Cycler (Applied Biosystems) using the following thermal profile(s): 95 °C for 30 seconds, followed by 16 cycles of 95 °C for 15 seconds, annealing temperature of 56 °C for 30 seconds and extension at 68 °C for 60 seconds, and a final extension step of 68 °C for 5 minutes (**Table 2.11-3**). PCR products were purified immediately (see **2.11.2**).

Table 2.11-3 MethPlex enrichment PCR thermal cycling conditions.

	Temperature (°C)	Time	Cycles
Initial denaturation/ Polymerase activation	95	30 seconds	1
Denaturation	95	15 seconds	16
Annealing	56	30 seconds	
Extension	68	60 seconds	
Final extension	68	5 minutes	1

2.11.2 MethPlex PCR purification

MethPlex enrichment PCR products were cleaned in order to remove unincorporated oligonucleotides and EpiMark® Hot Start Taq DNA Polymerase enzyme. Low throughput enrichment in the early stages of the project development used using QIAquick PCR Purification Kit (Qiagen) (see **2.11.2.1**). High-throughput strategies utilised QIAquick 96 PCR Purification Kit (Qiagen) (see **2.11.2.2**).

2.11.2.1 QIAquick PCR Purification Kit

250 µl Buffer PB containing 1:250 pH Indicator I (Qiagen) was added to each 50 µl MethPlex sample and mixed by vortexing. If the solution was orange or violet, indicating pH > 7.5, 10 µl 3 M sodium acetate, pH 5.0 was added and the solution mixed by vortexing. Samples were then transferred to QIAquick spin columns, placed in collection tubes, and centrifuged at 16,000 RCF for 60 seconds. The flow-through was discarded, 750 µl Buffer PE added to the QIAquick columns which were centrifuged at 16,000 RCF for 60 seconds. The flow-through was discarded, the columns placed in clean collection tubes and centrifuged at 16,000 RCF for 60

seconds, in order to remove residual Buffer PE. Spin columns were then placed in clean 1.5 ml DNase-free microcentrifuge tubes (STARLAB (UK)), 50 µl 0.22 µm-filtered Tris-HCl, pH 8.0, 0.025% Tween-20 was added to the column matrix directly and the column left to stand for 10 minutes to enable greater solubilisation of matrix-bound DNA. DNA was eluted by centrifugation at 16,000 RCF for 60 seconds. To increase yield, the eluate was reloaded into columns and centrifugation repeated. Purified MethPlex enrichment products were either analysed immediately or stored at -20 °C.

2.11.2.2 QIAquick 96 PCR Purification Kit

A QIAvac 96 (Qiagen) vacuum manifold was assembled according to the manufacturer's instructions. A QIAquick 96 PCR Purification plate was placed in the QIAvac top plate. 150 µl Buffer PM was added to each 50 µl MethPlex sample and mixed by pipetting up and down before careful transfer to the QIAquick plate. The plate was then covered with an AirPore Tape Sheet (Qiagen) to avoid airborne contamination and the vacuum pump turned on. When all the liquid had passed through the plate, the vacuum pump was switched off, AirPore Tape Sheet peeled back, 900 µl Buffer PE added to each well and the vacuum reapplied. This wash step was repeated, followed by application of maximum vacuum pressure for 10 minutes to remove any residual Buffer PE. The vacuum was then switched off, the top plate removed from the base and tapped on absorbent paper to remove any remaining buffer from the plate. The QIAvac 96 waste tray was removed and replaced with a 1.2 ml microtube rack. 60 µl 0.22 µm-filtered Tris-HCl, pH 8.0, 0.025% Tween-20 was added to the centre of each well, AirPore Tape Sheet reapplied and the plate left to stand for 10 minutes to enable greater solubilisation of

matrix-bound DNA. DNA was eluted under maximum vacuum pressure for 5 minutes. Sample eluates were transferred into PCR strips and tightly capped. Purified MethPlex enrichment products were either analysed immediately or stored at -20 °C.

2.12 Droplet Digital™ PCR (ddPCR)

ddPCR was performed using the QX100™ Droplet Digital™ PCR System (Bio-Rad). ddPCR reaction preparation, DG8™ Cartridge loading and droplet transfer were all carried out in an AirClean® AC600 PCR Workstation (STARLAB (UK)) laminar flow PCR cabinet, previously decontaminated with Distel High Level Disinfectant (Tristel Solutions Ltd., Snailwell, UK) and UV light sterilisation.

2.12.1 ddPCR reaction preparation

20 µl PCR reactions contained 1 x ddPCR Supermix for Probes (No dUTP) (Bio-Rad), 750 nM primers (Eurofins Genomics), 250 nM TaqMan MGB probes (Applied Biosystems) and varying amounts of bisulphite-treated DNA or MethPlex enriched DNA (see experimental chapters for specific details). Probes targeting potentially methylated gene promoter sequences were labelled with FAM™; probes complementary to a custom bisulphite-treated CpG methylated pUC19 spike control (see Ch. 6) were labelled with VIC™. PCR reactions were prepared as master mixes of a volume corresponding to the number of PCR replicates multiplied by 20 µl (individual PCR reaction volume) plus 10 % in PCR strips held in 96 well cold blocks. Capped PCR strips were vortexed, briefly centrifuged and the reaction mixture allowed to equilibrate to room temperature prior to droplet generation. NTCs were included in all ddPCR plates.

2.12.2 ddPCR droplet generation

20 µl PCR reactions were transferred from PCR strips to the middle row of a DG8™ Cartridge placed in a DG8™ Cartridge Holder by multichannel pipette. It is important not to introduce air bubbles as this results in the formation of fewer droplets and poor data quality. Pipette tips were grounded against the bottom edge of sample wells at an angle of c. 15° and samples slowly dispensed using consistent pressure. After half the sample had been dispensed, pipette tips were slowly drawn up the well walls while dispensing the remainder of the sample (to the first 'stop' of the multichannel pipette). 70 µl Droplet Generation Oil for Probes (Bio-Rad) was added to the bottom wells of the cartridge and a DG8™ Gasket securely hooked over the cartridge holder. The cartridge assembly was then placed into the QX100™ Droplet Generator and droplet generation initiated. When this was complete, the assembly was removed and gasket discarded. 40 µl of droplets were aspirated, slowly and with consistent force, from the top row of the cartridge using a multichannel pipette. Pipette tips were placed at c. 45° against junction between the bottom and side of the wells. Droplets were dispensed into an Eppendorf twin-tec® PCR Plate, skirted (Eppendorf, Hamburg, Germany) by positioning pipette tips along the side of wells, near to the bottom of the wells, and dispensing slowly. I was vital to meticulously apply these droplet transfer techniques to avoid shearing or coalescence of droplets. The PCR plate was then sealed with Pierceable Foil Heat Seal (Bio-Rad) using a PX1™ PCR Plate Sealer (Bio-Rad) and thermal cycled immediately.

2.12.3 ddPCR thermal cycling

The PCR plate was immediately transferred to a C1000 Touch™ Thermal Cycler (Bio-Rad) and incubated under varying conditions. Optimal thermal cycling temperatures and times are shown in **Table 2.12-1**. A ramp rate of 2 °C/s was used to avoid droplet instability through rapid temperature changes and to ensure each individual droplets reached uniform temperature during cycling.

Table 2.12-1 ddPCR thermal cycling conditions.

	Temperature (°C)	Time	Cycles	Ramp Rate
Initial denaturation/ Enzyme activation	95	10 minutes	1	
Denaturation	94	30 seconds	45	2 °C/s
Annealing/extension	T_a	90 seconds		
Enzyme deactivation	98	10 minutes	1	
Hold	8	Infinite	1	

2.12.4 ddPCR droplet analysis

After PCR amplification, the 96-wellPCR plate was removed from the thermal cycler and secured in the QX100™ plate holder and loaded into the QX100™ Plate Reader. Droplet reading was then carried out after verification of sufficient ddPCR™ Droplet Reader Oil and sufficient volume for waste collection. Droplet count data was stored and analysed using Quantasoft™ Software, version 1.7.4 (Bio-Rad).

Chapter 3

Expanded DNA methylation panel validation in bronchoalveolar lavage samples

3.1 Introduction

The detection of lung cancer at early stage can improve patient survival (McPhail et al., 2015). Epigenetic alterations occur early in cancer development and have the potential to provide effective diagnostic biomarkers (Feinberg et al., 2006). A number of studies have demonstrated the detection of DNA methylation in patient body fluids (Hubers et al., 2014; Hulbert et al., 2017; Leng et al., 2012; Liloglou et al., 2014; Nikolaidis et al., 2012; Ostrow et al., 2010). It is of note that despite the plethora of published studies on cancer biomarkers, only a trivial proportion are further developed and moved forward to clinical validation, notwithstanding the volume of articles suggesting their utility. This is primarily because of a lack of assay optimisation, diversity in study design and screening methodologies, and low sample size and resultant statistical power (Liloglou and Field, 2010; Liloglou et al., 2014; Sandoval et al., 2013b). Previous work within our group sought to assess DNA methylation biomarkers in an adequately powered, retrospective case-control study using qMSP analysis of minimally invasive bronchoalveolar lavage samples from 655 Liverpool Lung Project recruited individuals (Nikolaidis et al., 2012). Compliance with NCI Early Detection Research Network (EDRN) guidelines (Pepe et al., 2001) and the Cancer Research UK Diagnostic Biomarker Roadmap were considered essential in the study design. This study demonstrated that a panel of four targets (*CDKN2A*, *RASSF1*, *WT1* and *TERT*) provided for a diagnostic algorithm exhibiting 82% clinical sensitivity at 91% specificity in diagnosing lung cancer, thus doubling the efficiency of cytology (Nikolaidis et al., 2012). Additional DNA methylation markers were identified by targeted methylation microarray (Pulverer et al., 2012) and validated in matched non-small cell lung cancer and adjacent normal tissue samples by Pyrosequencing methylation analysis (Tost and

Gut, 2007). The top six markers from the previous bronchial washings study and the six newly identified markers with highest methylation frequency in tumour samples and absence of hypermethylation in the normal lung tissue were developed and optimised as multiplex qMSP assays.

The present study describes the validation of this extended DNA methylation biomarker panel in an adequately powered retrospective case-control study, utilizing Liverpool Lung Project bronchoalveolar lavage samples (Field et al., 2005). This was intended to confirm the findings of the previous study and further validate the superior sensitivity of methylated DNA biomarkers compared to cytology. The evaluation of potential gains in sensitivity and specificity contributed by novel markers was also a key aim of this investigation. Our panel qMSP assays previously utilized singleplex hydrolysis probe chemistry and have been modified and optimised as multiplex assays. This may increase assay analytical performance in addition to improving the efficiency of sample use by decreasing the number of qMSP reactions required in methylation screening. Furthermore, this will provide vital information for the selection of markers to progress to truly non-invasive blood plasma DNA methylation screening.

3.2 Methods

3.2.1 Liverpool Lung Project patients and samples

Training and validation sets included 417 individuals in total (224 lung cancer cases – 112 in each study arm; 193 age/sex frequency matched controls – 96 in the training set and 97 in the validation set). All patients were recruited to the Liverpool Lung Project through the Liverpool Heart and Chest Hospital and were referred for bronchoscopy because of clinical suspicion of lung cancer. As such, control subjects were not healthy volunteers, but received diagnosis for various diseases including, but not limited to, bronchitis and chronic obstructive pulmonary disorder (COPD). Inclusion criteria for the study were as follows:

- participant resident within the Liverpool Lung Project study area
- participant < 80 years of age
- participant willing and able to provide informed consent
- biopsy sample and brochoalveolar lavage specimens collected on suspicion of lung cancer
- two or more years follow-up participant clinical and epidemiological information available through hospital records, Merseyside and Cheshire Cancer Registry (MCCR) and the Office for National Statistics (ONS).

Participants diagnosed with cancers other than lung cancer were not excluded from the study. Specimens were excluded if extracted DNA failed quality control procedures.

The LLP study protocol has been approved by the Liverpool Research Ethics Committee and all participants provided informed consent in accordance with the Declaration of Helsinki.

Bronchoalveolar lavage specimens were stored in Saccomanno's fixative in a temperature-controlled environment (18°C). Cytological adequacy was positively assessed by the presence of alveolar macrophages.

DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen) according to the protocol in Chapter 2 (**2.2.1**). Extracted DNA concentration was measured by fluorescence quantification (**2.3.2**) and samples were normalised to 1 µg in 20 µl 0.1x TE prior to bisulphite treatment (**2.7**) and qMSP analysis (**2.10**).

3.2.2 Quantitative methylation-specific PCR

Multiplex qMSP assays and their use for methylation detection in bronchoalveolar lavage are described in **2.10**.

3.2.3 Exploratory univariate analysis

Subjects' epidemiological, clinical and methylation characteristics were described by case-control status separately for training and testing sets. Descriptive statistics were obtained and compared by using Chi-square test or Fischer's exact test for categorical variables and t-tests for normally distributed variables. The Mann-Whitney nonparametric alternative was used where the normality assumption failed. Epidemiological and clinical characteristics are presented in **Table 3.2-1** and DNA methylation statistics are shown in **Table 3.3-1**. **Figure 3.2-1** provides additional evidence that there was no significant difference in the age distributions between case and control groups.

Table 3.2-1 Epidemiological and clinical characteristics separated by case-control status for training and validations sets

Numbers in the main body of the table represent frequencies. Numbers in parentheses represent percentages.

	Training set (N = 208)		Validation set (N = 209)	
Subject characteristics	**Case (n = 112)**	**Control (n= 96)**	**Case (n = 112)**	**Control (n= 97)**
Age group				
<60	11 (9.8)	16 (16.7)	14 (12.5)	15 (15.5)
60 – 79	93 (83.0)	76 (79.2)	92 (82.1)	75 (77.3)
80+	8 (7.1)	4 (4.2)	6 (5.4)	7 (7.2)
Age				
Mean ± SD	69.0 ± 7.54	67.5 ±7.74	68.3 ± 7.64	67.8 ± 7.90
Gender				
Male	70 (62.5)	53 (55.2)	65 (58.0)	54 (55.7)
Female	42 (37.5)	43 (44.8)	47 (48.0)	43 (44.3)
Smoking status[1]				
Never	5 (4.5)	19 (19.8)	6 (5.4)	20 (20.6)
Former	49 (43.8)	55 (57.3)	49 (43.8)	55 (56.7)
Current	58 (51.8)	22 (22.8)	57 (50.9)	22 (22.7)
Specimen storage duration (years)				
<5	15 (13.4)	19 (19.8)	13 (11.6)	19 (19.6)
5+	97 (86.6)	77 (80.2)	99 (88.4)	78 (80.4)
Cytology[1]				
Negative	57 (50.9)	95 (99.0)	62 (55.4)	97 (100.0)
Positive	43 (38.4)	0 (0.0)	37 (33.0)	0 (0.0)
Suspicious	12 (10.7)	1 (1.0)	13 (11.6)	0 (0.0)
Histological diagnosis				
Adenocarcinoma	27 (24.1)		28 (25.0)	
Squamous cell carcinoma	44 (39.3)		41 (36.6)	
NSCLC (NOS) [2]	16 (14.3)		14 (12.5)	
Small cell carcinoma	15 (13.4)		18 (16.1)	
Others[2]	10 (8.9)		11 (9.8)	

[1] Statistically significant in both datasets ($P < 0.05$)

[2] NSCLC (NOS): Non-small cell lung cancer (not otherwise specified)
Others (Large cell carcinoma, carcinoid, lung carcinoma unconfirmed pathology)

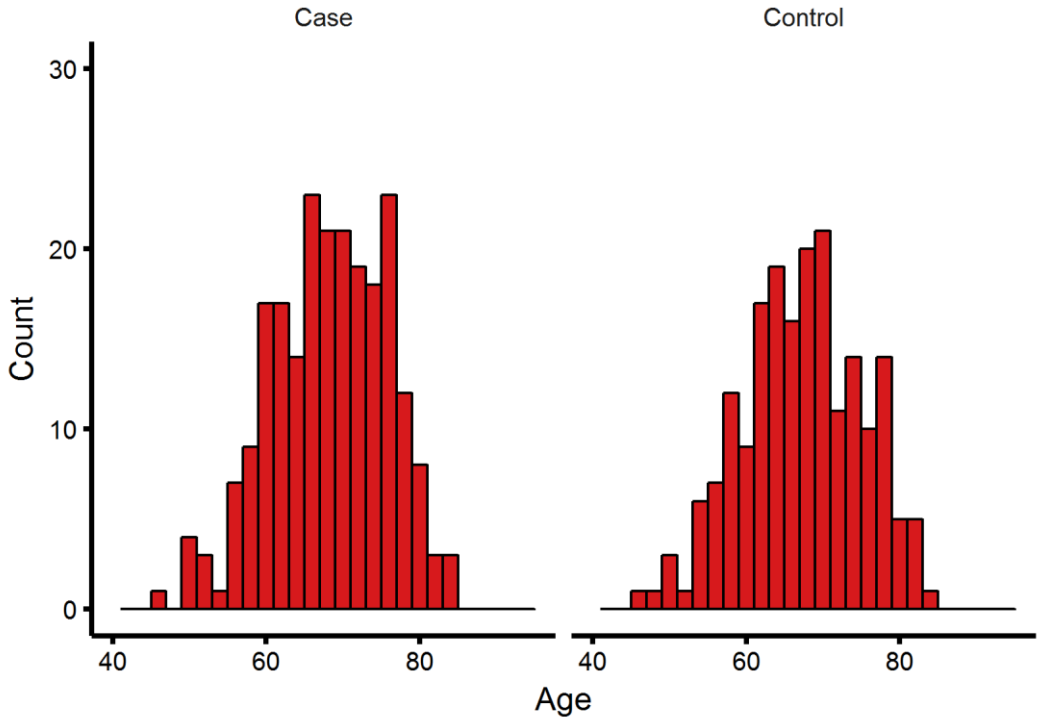


Figure 3.2-1 Distributions of subject ages in case and control groups

3.2.4 Statistical modelling for the identification of optimal markers

Three statistical models (Univariate association test, marker combination by Best Subset Regression (BSR) using glmulti/bestglm and marker combination by Multifactor Dimensionality Reduction (MDR)) were tested to identify the optimal marker panel and algorithm for diagnostic efficiency. Univariate analyses were used to explore the marginal effect of each marker on subject status. Because no single model selection criterion exhibits a uniformly superior performance over a wide range of scenarios, we implemented best model subset selection and multi-model inference using glmulti (Calcagno, 2013) and bestglm (McLeod and Xu, 2017) R packages to identify the best additive logistic regression combination most predictive of subject status using generalized linear models. The models were fitted using Akaike information criterion (AIC), Bayesian information criterion (BIC) and Bayesian Information Criterion with Bernoulli prior (BICq). Multifactor dimensionality

reduction (MDR) was used to develop predictive algorithms that examine non-additive combination of markers. MDR is a non-parametric data reduction procedure that searches exhaustively for the optimal combination for predicting subject status (Hahn et al., 2003; Winham, 2012; Winham and Motsinger-Reif, 2011). Each possible combination pools relevant markers into a single dimension with two groups classified as either high or low risk, and the best combination is chosen that maximises the internal cross-validation and minimises the prediction error. The MDR permutation test was used to test the significance of association between subjects' disease status and each marker interaction (Calle et al., 2010).

The predictive performance of each derived algorithm was evaluated in the independent test data. A logistic regression model was used to predict the probability of being a case for each observation using the optimal marker panel. The classification accuracy of this probability was assessed using diagnostic measures such as accuracy, sensitivity and specificity. The overall performance of the range of predicted probabilities was summarised using the area under the ROC curve (AUC) using the pROC R package (Robin et al., 2011). The diagnostic performance of the optimal marker panel, the 'Top 6 univariate' and GN models (see **3.3.3**) and their extended versions incorporating cytology were evaluated in both training and validation datasets and also in stratified analyses incorporating epidemiological and clinical risk factors such as age, gender, smoking status and lung cancer histological subtypes. ROC-AUCs obtained from the stratified analyses were compared using the DeLong test (DeLong et al., 1988). All analyses were performed using R 3.4.0 software (Vienna, Austria) and STATA® version 14.2 (StataCorp LP, College Station, Texas).

3.3 Results

The top six markers from the Nikolaidis study (Nikolaidis et al., 2012) were *TERT*, *RASSF1*, *WT1*, *p16* and *RARb* (designated *CDKN2A* and *RARB* respectively throughout this thesis, using the HGNC approved symbols (Wain et al., 2002)). Methylated gene promoters identified by methylation microarray and selected via Pyrosequencing validation were *ABCB1*, *F2R*, *HOXA1*, *HOXA10*, *MT1G* and *PITX2*. qMSP assays targeting the above gene promoters were previously designed, optimised and technically validated as triplex qMSP assays (two methylation targets plus methylation-independent *ACTB* endogenous control). The bronchoalveolar lavage case-control training set specimens were screened with these optimised assays. The most frequently methylated gene in lung cancer cases was *TERT* (82/112 subjects). *TERT* methylation was also detected in 3/97 control specimens. The detection frequencies for all screened methylation assays are depicted in **Figure 3.3-1**.

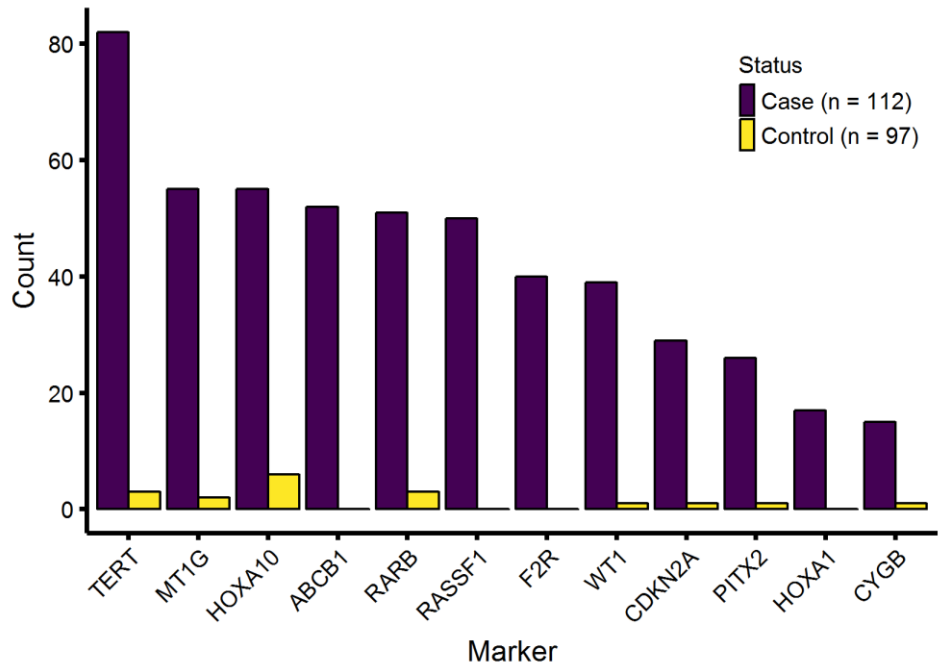


Figure 3.3-1 Detection of candidate marker methylated DNA in bronchoalveolar lavage training set.
Methylation positive subject counts realised by qMSP analysis of case and control bronchoalveolar lavage specimens are represented by the y-axis. Candidate genes are noted on the x-axis.

3.3.1 Individual markers classify cases and controls but do not discriminate adequately

Individual sensitivity and specificity were assessed, with the *TERT* assay showing moderate sensitivity (68.8%) and all other assays displaying poor performance (sensitivity <60%). All assays showed excellent specificity (>90%). Receiver operating characteristic (ROC) curve analysis was also implemented, with the area under the curve (AUC) being used as a measure of the ability of a test to discriminate between cancer and non-cancer classes. *TERT* showed good predictive ability to discriminate lung cancer from control subjects (AUC = 0.82), while other markers classified cases and controls less well (AUC <0.75). Assays targeting *TERT* and *HOXA10* correctly classified greater than 70% of training set subjects. However, these performance characteristics are not adequate for clinical implementation. Univariate analysis statistics for both the training and validation sets are presented in **Table 3.3-1**.

3.3.2 Identification of optimal marker combinations for classification of the training set

Twelve potential logistic regression model predictors (i.e. our twelve genes of interest) suggests 4096 potential additive models (including the null model). It was not desirable or feasible to manually test every possible marker combination. Therefore a number of automated model selection strategies were implemented to identify optimal candidate biomarker panels. Logistic regression models designated 'Top 6 univariate' and 'Top 4 univariate' included the appropriate number of markers ranked by AUC. Marker combinations were also selected by best subset logistic regression, with selection based upon AIC, BIC or BICq, and multifactor dimension reduction (MDR).

Table 3.3-1 Univariate statistics for candidate methylation biomarkers assessed by qMSP in bronchoalveolar lavage training and validation sets

	Training set					Validation set				
	Positives			Model-based classification[1]		Positives			Prediction using trained univariate logit model[1]	
Gene promoter	Case (n=112)	Control (n=96)	χ^2 *P* value	Accuracy (%)	AUC (95%CI)	Case (*n*=112)	Control (*n*=97)	χ^2 *P* value	Accuracy (%)	AUC (95% CI)
TERT	77	4	$<1 \times 10^{-4}$	81.3	0.82 (078, 0.87)	82	3	$<1 \times 10^{-4}$	84.2	0.85 (0.81, 0.90)
HOXA10	60	8	$<1 \times 10^{-4}$	71.2	0.73 (0.67, 0.78)	55	6	$<1 \times 10^{-4}$	69.9	0.71 (0.66, 0.77)
RARB	45	3	$<1 \times 10^{-4}$	66.3	0.69 (0.64, 0.73)	51	3	$<1 \times 10^{-4}$	69.4	0.71 (0.66, 0.76)
MT1G	40	2	$<1 \times 10^{-4}$	64.4	0.67 (0.62, 0.71)	55	2	$<1 \times 10^{-4}$	71.8	0.74 (0.69, 0.78)
WT1	25	2	$<1 \times 10^{-4}$	57.2	0.60 (0.56, 0.64)	39	1	$<1 \times 10^{-4}$	64.6	0.67 (0.62, 0.71)
CDKN2A	27	1	$<1 \times 10^{-4}$	58.7	0.61 (0.57, 0.65)	29	1	$<1 \times 10^{-4}$	59.8	0.62 (0.58, 0.67)
PITX2	24	1	$<1 \times 10^{-4}$	57.2	0.60 (0.56, 0.64)					
CYGB	12	1	4×10^{-3}	53.9	0.55 (0.52, 0.58)					
ABCB1	43	1	$<1 \times 10^{-4}$	66.4	0.69 (0.64, 0.73)					
RASSF1	49	0	$<1 \times 10^{-4}$	60.4	-					
F2R	48	3	$<1 \times 10^{-4}$	67.8	0.70 (0.65, 0.75)					
HOXA1	10	0	3×10^{-3}	51.5	-					

[1] Disease class prediction based on predicted $Pr(D) \geq 0.5$

Models including markers *ABCB1*, *RASSF1*, *HOXA1* and *F2R* resulted in complete separation in either the training set or both training and validation sets and were therefore dropped from the modelling process.

AIC-based best subset regression selected a large model with seven predictors (*CDKN2A*, *MT1G PITX2*, *HOXA10*, *CYGB*, *RARB* and *TERT*). Best BIC and BICq selection provided smaller, more parsimonious models with four and three predictors respectively. *TERT* and *RARB* were common to all models, regardless of the selection process.

3.3.3 Diagnostic DNA methylation algorithm performance in training and validation datasets

Performance of different discriminatory algorithms in training and validation datasets is displayed in **Table 3.3-2**. All logistic models classified well in the training set, with similar diagnostic accuracy. The best subset BICq and MDR models displayed identical statistics. The AIC-based model showed the highest sensitivity (83.9%), while the BIC and BICq based models attained the highest specificity (92.7%). The 'Top 6 univariate' model also conducted well, but specificity was reduced when markers were dropped from this model ('Top 4 univariate model'). Addition of the top three ranking two- and three-way marker interactions, identified by MB-MDR (Calle et al., 2010), to any of the above models did not improve their discriminatory efficiency by any measure. The best subset BICq logistic regression model including *TERT*, *RARB* and *CDKN2A* and the BIC-based model, also including *HOXA10*, had identical sensitivity, specificity and diagnostic accuracy. Despite the BIC-based model displaying a marginally higher AUC than the BICq-selected model, the latter was judged to be the optimal model by virtue of being the more parsimonious.

Table 3.3-2 Classification and predictive accuracies of DNA methylation models in training and validation datasets

Performance measure	Discriminatory algorithms				
	Top univariate markers		Best subset logit model/MDR		
	Top 6 univariate	Top 4 univariate	AIC	BIC	BICq/MDR
Classification performance in training dataset					
Se/Sp (%)	83.0/90.6	83.0/84.4	83.9/89.6	81.2/92.7	81.2/92.7
DA (%)	86.5	83.7	86.5	86.5	86.5
AUC (95% CI)	0.91 (0.87, 0.95)	0.89 (0.84, 0.93)	0.92 (0.88, 0.95)	0.90 (0.86, 0.94)	0.88(0.84, 0.93)
Predictive performance in validation dataset					
Se/Sp (%)[1]	91.1/91.8	87.5/91.8	90.2/90.7	85.7/93.8	85.7/93.8
DA (%)[2]	91.4	89.5	90.4	89.5	89.5
AUC (95% CI)[3]	0.94 (0.90, 0.97)	0.93 (0.89, 0.96)	0.94 (0.91, 0.97)	0.91 (0.88, 0.95)	0.91 (0.88, 0.95)

[1] Se, sensitivity = $\frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives}$

Sp, specificity = $\frac{number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ positives}$;

[2] DA, discriminatory accuracy;

[3] AUC, area under ROC curve

Optimal model performance in the validation dataset, broken down into cytology positive and negative cases, can be seen in **Table 3.3-3**. Higher sensitivity was observed in cytology-positive cases (89.5%) than cytology-negative individuals with lung cancer (78.2%). Nevertheless, DNA methylation panel lung cancer detection in cytology-negative subjects represents an increase from zero to 49 diagnoses of individuals with lung tumours. The utility of DNA methylation markers in cytologically occult samples is indicated in **Figure 3.3-2**. The optimal methylation algorithm detected an additional 93 lung cancer cases in cytology-negative subjects across both data sets. Overall, model sensitivity was 85.7% while maintaining excellent specificity (93.8%).

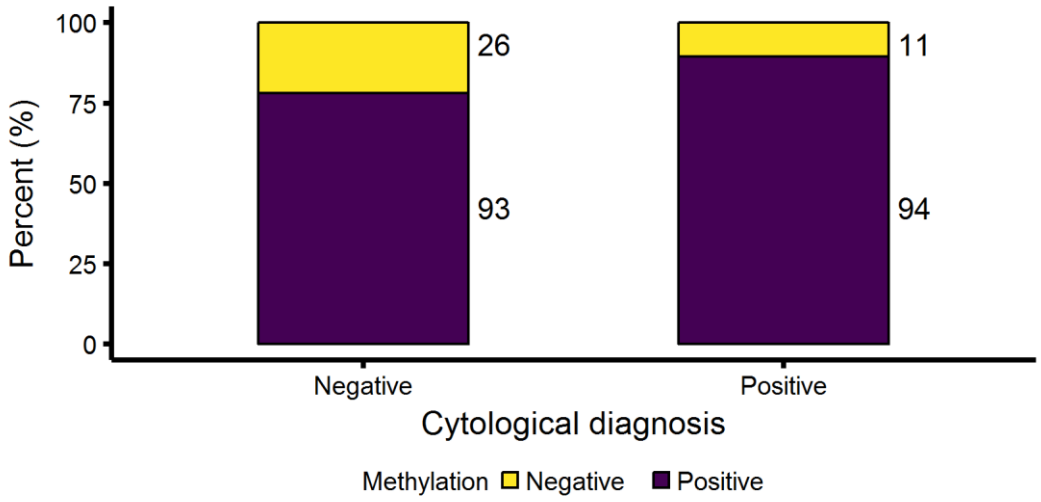


Figure 3.3-2 DNA methylation classified distributions case in cytology-negative and cytology-positive groups
The total number of subjects in each subgroup is annotated to the right of the bars. 93 lung cancer cases classified as negative by cytological diagnosis were classified positive by the optimal methylation model.

The optimal marker set identified in the previous study, *TERT*, *WT1*, *CDKN2A* and *RASSF1* (Nikolaidis et al., 2012) (designated the GN model, using the initials of the paper's first author) was also evaluated in the validation set, exhibiting higher

specificity (95.9%) but lower sensitivity (83.9%) than the best subset BICq panel. Diagnostic accuracy and AUC were the same as the new diagnostic algorithm.

The optimal model did not include any of the new markers identified by methylation microarray. Therefore, the 'Top 6 univariate' and 'Top 4 univariate' DNA methylation models, incorporating *MT1G* and *HOXA10*, were tested on the validation set, also showing improved operational characteristics (**Table 3.3-2**). In fact, the 'Top 6 univariate' showed higher specificity (91.1%) but slightly lower sensitivity (91.8%) than the best subset BICq panel. Diagnostic accuracy and AUC showed noticeable improvement in comparison to the smaller models (diagnostic accuracy = 91.4%, AUC = 0.94). ROC curves for the BICq and 'Top 6 univariate' models are presented in **Figure 3.3-3**.

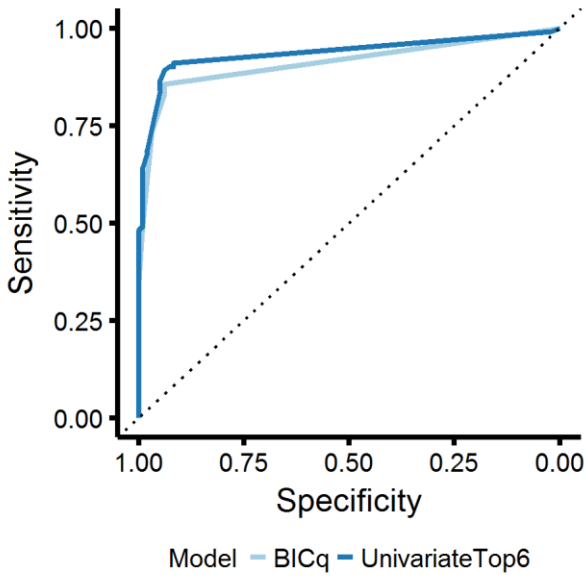


Figure 3.3-3 BICq and 'Top 6 univariate' model ROC curves

Addition of cytological diagnostic result as an additional predictor in DNA methylation panel models slightly improved diagnostic parameters. For example best subset BICq sensitivity was 85.7% in the validation set and increased to 88.4%

with the addition of cytology result. No changes were seen in specificity, but diagnostic accuracy and AUC increased marginally. However, the optimal methylation panel was clearly superior to cytology result alone in terms of diagnostic efficiency. Cytology sensitivity was 44.6% at 100% specificity.

Table 3.3-3 Best subset logistic regression model performance in the bronchoalveolar lavage validation set
The optimal methylation marker panel comprises *CDKN2A*, *RARB* and *TERT*. This model is compared with its extension, including cytological diagnosis, and cytological diagnosis alone.

Prediction model	Status	Cytology	+	-	Sensitivity (%)	Specificity (%)
Methylation marker panel	Lung cancer	+	47	3	94.0	
		-	49	13	79.0	
		Overall	96	16	85.7	
	Controls	+	0	0		(100)
		-	6	91		(93.8)
		Overall	6	91		(93.8)
Cytology only	Lung cancer		50	62	44.6	
	Controls		0	97		(100)
Methylation marker panel + cytology	Lung cancer		99	13	88.4	
	Controls		6	91		(93.8)

3.3.4 Identification of potential biases in epidemiological and clinical subgroups

Methylation panel signatures were tested across a unified dataset, including all test and validation data. This analysis was carried out to identify any differences between groups when diagnosed using DNA methylation algorithms and the details for the optimal (BICq) model are displayed in **Table 3.3-4**. No differences in model performance were observed between age or gender groups.

Table 3.3-4 Performance of the optimal discriminatory model in both training and validation datasets, stratified by epidemiological and clinical charateristics

Subject characteristics	Number of specimens	%		AUC (95% CI)	*P* value[2]			
		Se	Sp		(1)	(2)	(3)	(4)
Diagnosis								
Lung cancer	224	83.5						
No malignancy	193		(93.3)	0.90 (0.87, 0.93)				
Age								
<60	56	64	(96.8)	0.81(0.71, 0.91)				
60 – 79	336	85.4	(92.7)	0.91 (0.88, 0.94)	0.078			
80+	25	92.9	(90.9)	0.94 (0.85, 1)	0.068	0.513		
Gender								
Male	242	85.2	(90.7)	0.90 (0.86, 0.94)				
Female	175	80.9	(96.5)	0.90 (0.85, 0.94)	0.925			
Smoking status								
Never	50	81	(100)	0.91 (0.79, 1)				
Former	208	81.6	(90.9)	0.91 (0.86, 0.95)	0.955			
Current	159	85.2	(93.2)	0.88 (0.84, 0.93)	0.681	0.459		
Specimen storage duration (years)								
<5	65	67.9	(100)	0.84 (0.75, 0.92)				
≥5	351	85.7	(91.6)	0.90 (0.87, 0.93)	0.175			
Lung cancer cases only								
Cytology								
Negative	119	78.2		0.87 (0.83, 0.91)				
Positive	105	89.5		0.95 (-)	**<0.001**			
Histological diagnosis								
Adenocarcinoma	55	78.2		0.87 (0.81, 0.92)				
Squamous cell carcinoma	85	88.2		0.93 (0.89, 0.96)	0.084			
NSCLC (NOS)	30	76.7		0.87 (0.78, 0.95)	0.986	0.174		
Small cell lung cancer	33	93.9		0.95 (0.91, 0.99)	**0.022**	0.398	0.07	
Others[1]	21	71.4		0.82 (0.72, 0.93)	0.486	0.063	0.539	**0.024**
Stage (pT)								
1	26	61.5		0.78 (0.68, 0.88)				
2	62	88.7		0.93 (0.89, 0.97)	**0.007**			
3	15	90.0		0.88 (0.77, 0.99)	0.172	0.450		
4	43	93.0		0.95 (0.91, 0.99)	**0.002**	0.477	0.264	
Nodal status (pN)								
0	64	94.4		0.90 (0.85, 0.95)				
1	25	80.0		0.88 (0.80, 0.97)	0.756			
2	43	83.7		0.90 (0.84, 0.96)	0.942	0.729		
3	11	90.9		0.93 (0.84, 1)	0.54	0.452	0.595	

[1] NSCLC (NOS): Non-small cell lung cancer (not otherwise specified)
Others (Large cell carcinoma (n = 2), carcinoid (n = 3), lung carcinoma unconfirmed pathology (n = 18); [2] DeLong test.

Duration of specimen storage had no effect on diagnostic efficiency. Sensitivity and specificity, and AUC were similar in groups with different smoking status. The methylation marker panel displayed higher specificity (89.5% vs 78.2%) and significantly higher AUC (P <0.001) in cytology positive specimens. The optimal methylation model classified 93/119 (78.2%) cytology-negative samples, 20/25 (80%) suspicious cytology specimens and 74/80 (92.5%) lung cancer cytology-positive samples as disease positive with significant differences between groups (χ^2 = 0.025). Interestingly, if the data are classified using the 'Top 6 univariate' model, six addition cytology-negative subjects classify as disease positive and the difference between groups is no longer statistically significant (χ^2 = 0.12). The significance of group differences remains if discrimination of subject class uses the GN model from the Nikolaidis study (Nikolaidis et al., 2012), (χ^2 = 0.002).

Stage T1 tumours were detected with lower frequency (61.5%) than T2, T3 and T4 (>80%) using the optimal model from this study. This was similar to the findings of the Nikolaidis study (T1 = 63%, T2 – T4 >80%). When the GN model was used for disease discrimination, four fewer T2 tumours were correctly diagnosed resulting in a decrease in sensitivity from 88.7% to 82.3%. Furthermore, the 'Top6 univariate' model increased stage T4 sensitivity from 93.0% to 97.7%, while also increasing T2 sensitivity to greater than 90%. Inspection of **Figure 3.3-4** clearly indicates that cytological diagnostic sensitivity is considerably lower in all pathological stages and relatively small differences between DNA methylation diagnostic algorithms for any given stage can also be observed. Cytology shows markedly higher sensitivity for stage T4 tumours, whereas DNA methylation classification was similar for stages T2 – T4.

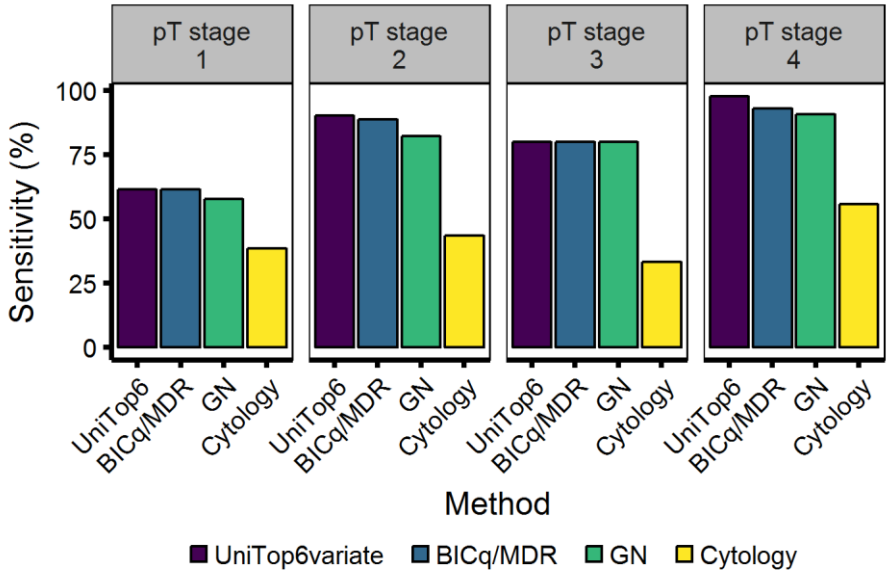


Figure 3.3-4 Sensitivities of DNA methylation lung cancer discrimination models and cytological diagnosis in different pathological pT stages of lung cancer.

Data were also stratified by histological subtype. The optimal model demonstrated higher efficiency of detection for small cell lung cancer (93.9%) and squamous cell carcinoma (88.2%) over adenocarcinoma (78.2%). The GN model demonstrated 100% efficiency in small cell lung cancer.

3.4 Discussion

Our group's previous study of DNA methylation detection in bronchoalveolar lavage yielded promising data (Nikolaidis et al., 2012). It was patently clear that a four-marker panel provided for superior disease detection compared to cytological diagnosis. We hypothesized that diagnostic performance could be improved through the interrogation of additional candidate markers, identified experimentally through methylation microarray screening of tumour and normal tissue samples and subsequent validation. Furthermore, improvements in discriminatory model characteristics were anticipated following the development of multiplex assays.

In the present study, the majority of markers performed moderately well when evaluated individually. Diagnostic performance was improved through their consideration as panels of multiple markers, modelled by logistic regression. The optimal selected model included data obtained from assays targeting *CDKN2A*, *RARB* and *TERT*, none of which were part of the newly established marker set. However, the 'Top 6 univariate' model, including *HOXA10* and *MT1G*, in addition to legacy markers (Nikolaidis et al., 2012), presented very strong performance characteristics in the validation set (Se 91.1%, Sp 91.8%, AUC 0.94). It would be justified, therefore, to investigate the utility of a subset of these novel markers, alongside more established biomarkers, in future blood plasma studies.

The above measures of diagnostic accuracy are superior to those reported in our group's previous study (Se 82.0%, Sp 90.8%, AUC 0.89) (Nikolaidis et al., 2012). The GN model also showed improved diagnostic efficacy in the present study (Se 83.9%, Sp 95.9%, AUC 0.91), presumably due to analytical improvements arising

from assay multiplexing. However, it should be noted that these changes in disease detection could arise for any number of reasons and may simply be peculiar to this dataset.

Validation of the CE marked *in vitro* diagnostic Epi proLung assay in bronchial aspirates demonstrated 78% sensitivity and 96% specificity (Dietrich, 2011), marginally more specific and notably less sensitive than the markers evaluated in this chapter. An appraisal of *SHOX2* methylation, in combination with other methylated markers, utilizing plasma of patients with malignant and non-malignant lung disease recently indicated that *SHOX2* and *PTGER4* realized 67% sensitivity at a fixed specificity of 90%, and 79% specificity at a fixed sensitivity of 90% (Weiss et al., 2017), lower than the present study on both benchmarks. However, it is not surprising to see higher performance characteristics in the assessment of locally sampled specimens such as bronchoalveolar lavage. A three-marker combination (*TAC1*, *HOXA17* and *SOX17*), evaluated in the sputum of stage I and IIa non-small cell lung cancer patients (i.e. node-negative) and patients with non-cancerous lesions, displayed higher sensitivity (98%) and lower specificity (71%) (Hulbert et al., 2017) than the final marker panel in this chapter. More recently, a four-gene model (*CDO1*, *PTGDR*, *MARCH11* and *UNCX*) conferred sensitivity and specificity of 70.3% and 84.8% when promoter methylation was analysed in pleural effusions (Ooki et al., 2017). Both of these diagnostic determinants were lower than our current best bronchial lavage model.

It is of critical importance that diagnostic algorithms detected disease independently, displaying no biases associated with smoking status, indicating that DNA methylation differences observed were cancer-specific and not smoking-related alterations.

Analysis of all subject data stratified by histological diagnosis demonstrated 100% sensitivity in the detection of small cell lung cancer when using the GN model. Similarly, the 'Top 6 univariate' algorithm, including methylation detection results for *TERT*, *MT1G*, *HOXA10*, *RARB*, *WT1* and *CDKN2A*, had very high sensitivity (97.0%) for the same disease subtype. Furthermore, this model exhibited exceptionally high sensitivity for stage T4 disease. Unfortunately, this may not be of great clinical utility. Small cell carcinoma has the most aggressive development of any lung cancer and exhibited median survival of 11.6 months in UK patients who received both chemotherapy and radiotherapy (Khakwani et al., 2014). Stage T4 disease is by definition extensive and likely to be highly symptomatic, and therefore diagnosis by other means is likely to be straightforward. However, it is noteworthy that this DNA methylation algorithm provided a positive result in 18 out of 19 cytology-negative lung cancer cases with T4 pathological staging. Detection of these aggressive and late stage tumours by DNA methylation analysis when other diagnostic modalities fail could facilitate appropriate end-of-life care. The highest priority at this stage may be to reduce tumour burden and improve health-related quality of life (Gaertner et al., 2015).

DNA methylation panels demonstrated higher sensitivity for squamous cell carcinoma and small cell lung cancer than for other histological subtypes. This is to be expected as these cancers tend to develop centrally in the lung, and are therefore more likely to

be better sampled by bronchoalveolar lavage with a bronchoscope positioned in one of the major bronchi. Furthermore, it is not surprising that stage T1 disease was detected less readily than more advanced tumours, since T1 tumours are smaller and doubtless contribute fewer cells to any given sample. Their detection may be improved by further improved assay analytical sensitivity.

Statistically significant interactions between markers were found in MDR classification. These did not improve the diagnostic accuracy of the models under test and hence were not pursued, being beyond the scope of this study. However, investigation of potential epistatic relationships between methylated gene promoters may be of interest and should be investigated further.

The experimental work within this chapter is not without its limitations. The exclusion of markers exhibiting perfect prediction in univariate modelling is inefficient. In effect, markers were dismissed because of limitations in the chosen statistical methodology when they could be interesting or informative. It is possible that this characteristic of certain markers was simply an idiosyncrasy of this sample set and may not emerge in future studies. Data also should be reanalysed using additional classification methods where this problem does not arise, such as random forest classification. A recent study of plasma DNA methylation markers in the detection of breast cancer used support vector machines for disease classification and prediction (Uehiro et al., 2016). This and other machine learning approaches also should be considered.

In conclusion, this study demonstrates that the optimal, parsimonious three-marker diagnostic algorithm attained by analysis of methylated DNA in bronchoalveolar lavage specimens presents a minimally invasive method of discriminating between lung cancer cases and controls. It offers vast improvements over standard cytological diagnosis, classifying the majority of cytology-negative lung cancer cases positively. Assessment in a large, multicentre prospective trial should be a priority. Implementation of related methylation panel discriminatory testing in non-invasive specimens, such as blood plasma, should be investigated and is indeed the focus of the remaining chapters of this thesis. Eventual clinical validation and implementation could decidedly aid in meeting the long term goal of reducing lung cancer mortality.

Chapter 4

ddPCR assay and workflow development and optimisation

4.1 Introduction

This chapter outlines the experimental strategies pursued in the development of sensitive and specific ddPCR assays to facilitate DNA methylation detection and quantification. Droplet Digital™ PCR has previously demonstrated highly precise quantification of nucleic acids (Pinheiro et al., 2012) and indeed exhibited increased precision compared to qPCR in the analysis of diverse nucleic acid species (Hindson et al., 2013; Sedlak et al., 2014; Strain et al., 2013). Noting the potentially improved analytical characteristics of ddPCR, we decided to investigate its utility in the detection of methylated DNA with the expectation that it would increase analytical sensitivity and measurement precision. These properties would be required for the successful transition of DNA methylation marker panels to non-invasive clinical samples such as blood plasma. Prior to the commencement of this project, there were no reviewed communications regarding the utility of ddPCR for DNA methylation detection. A small number of papers have now been published and some of these are described in **1.3.6**.

Digital PCR conditions were heavily optimised and validated in order to provide reproducible and robust data resulting from assays showing clear differentiation between positive and negative ddPCR droplets. Thorough optimisation of ddPCR assays should ensure adequacy of data quality, precision and reproducibility (Taylor et al., 2015). This multifaceted process was realised to provide assays conducive to the analysis of CpG dinucleotide methylation of cfDNA in blood plasma.

Evaluation of extraction efficiency and material losses in downstream processing steps in clinical methods is critical in establishing analytical sensitivity and requires

adequate quality controls. Since plasma yields low concentration DNA (Devonshire et al., 2014; Tamkovich et al., 2005), and it is necessary to divide extracted DNA multiple times to assess multiple methylation markers, quantification by UV spectroscopy or fluorescence spectroscopy is not appropriate. Based on our previous experience with the low amounts of cfDNA extracted from plasma, which is the ultimate targeted sample type, we decided to quantify the extracted DNA by qPCR. Such assays are frequently used for measurements of cfDNA (Fleischhacker and Schmidt, 2007) and have the advantage of measuring amplifiable DNA. The qPCR assay targeting *ACTB* was designed and optimised for the quantification of DNA extracted from plasma samples (calling it hereafter *ACTB* cfDNA assay). However, estimated genomic DNA yields from plasma have been observed to vary by over twofold within the same subject when using different qPCR genomic DNA assays (Devonshire et al., 2014). The ddPCR workflows in this project, therefore, in addition to the *ACTB* assay, used an exogenous pUC19 plasmid DNA spike-in process control to assess losses in DNA extraction and manipulation. Assay design was complicated by the fact that methylation-specific assays require bisulphite-treatment of DNA and the resulting changes in sequence made it necessary to manipulate the pUC19 vector to provide a suitable control. Control production and engineering is covered in the methods section of this chapter and the optimisation of an assay for its quantification is presented in the results.

Validation of *ACTB* cfDNA quantification assay in a moderately large blood plasma sample set revealed that the majority of specimens did not yield sufficient DNA to enable methylated DNA screening of multiple markers. It was necessary, therefore to

devise a strategy to overcome this challenge, leading to the development and validation of MethPlex enrichment ddPCR.

Schematic representations of optimisation workflows follow in **Figure 4.1-1, Figure 4.1-2** and **Figure 4.1-3**.

Due to the large amount of data reported in this chapter, I have combined the results and discussion sections, as this may be helpful to the reader.

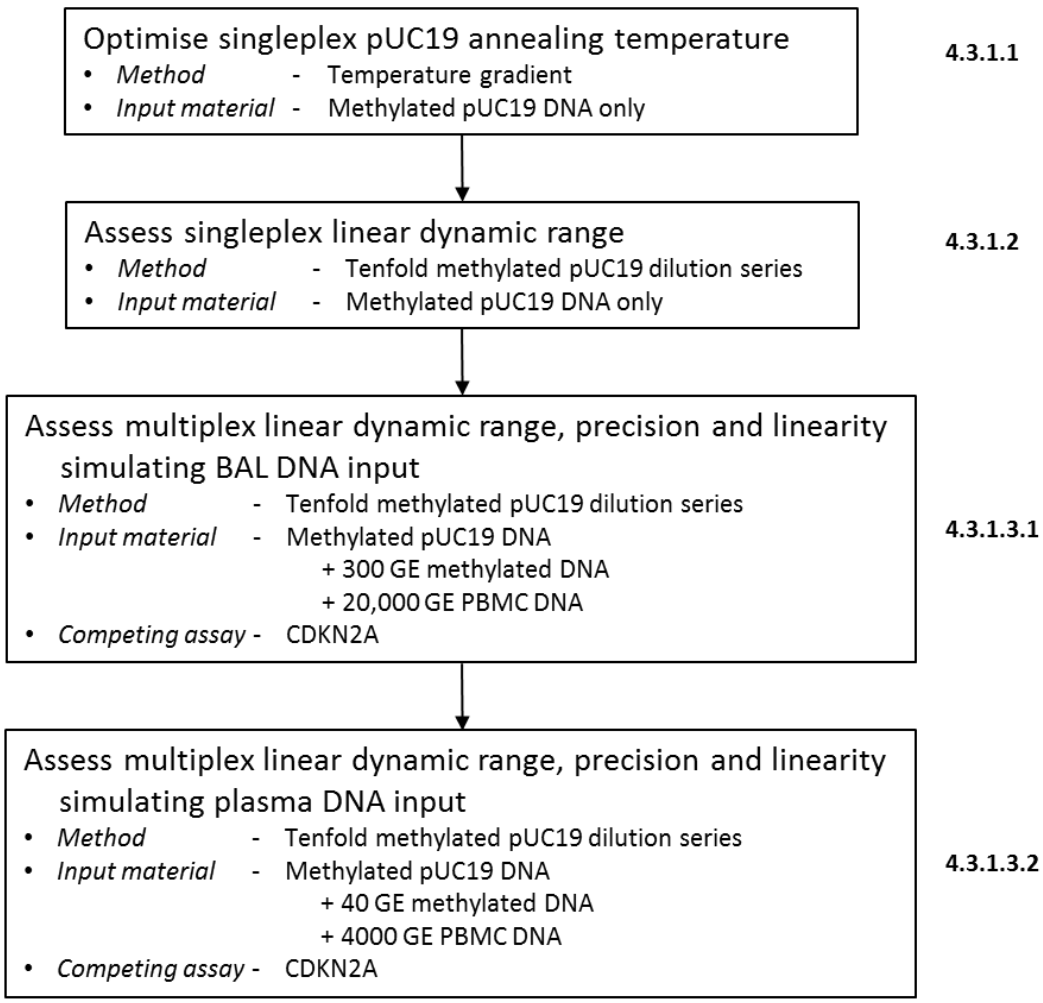


Figure 4.1-1 Control (spike-in) assay optimisation workflow
Schematic showing key steps in the optimisation of the methylated pUC19 ddPCR assay. Corresponding chapter sections are noted in bold to the right of the figure.

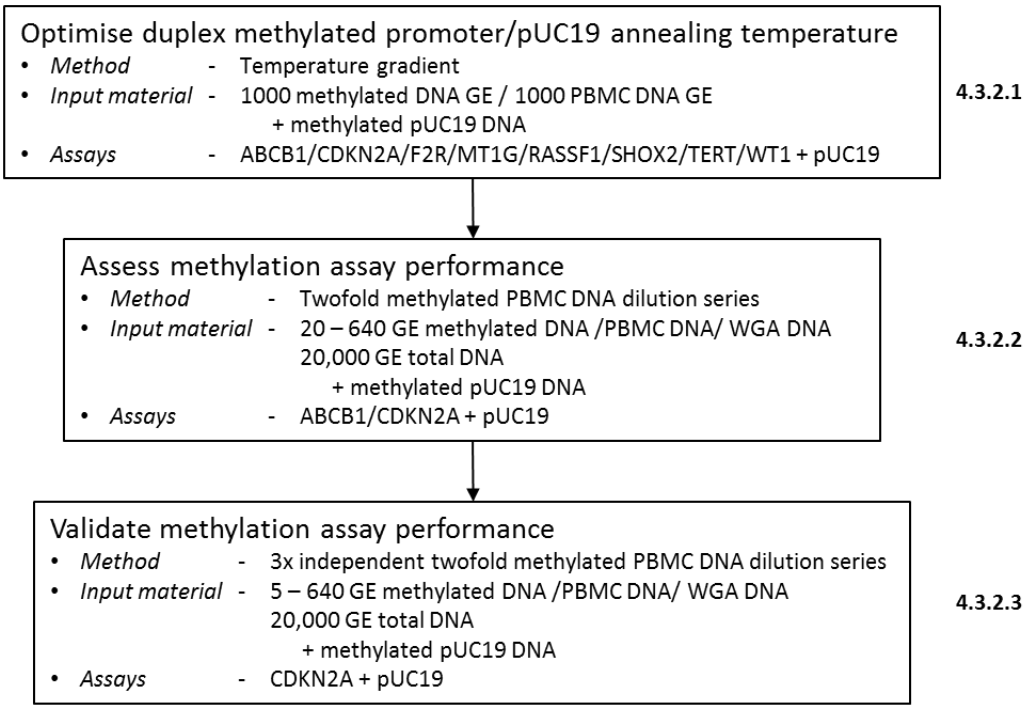


Figure 4.1-2 Target gene promoter ddPCR optimisation workflow
Schematic showing key steps in the optimisation of the methylated DNA biomarker duplex ddPCR assays. Corresponding chapter sections are noted in bold to the right of the figure.

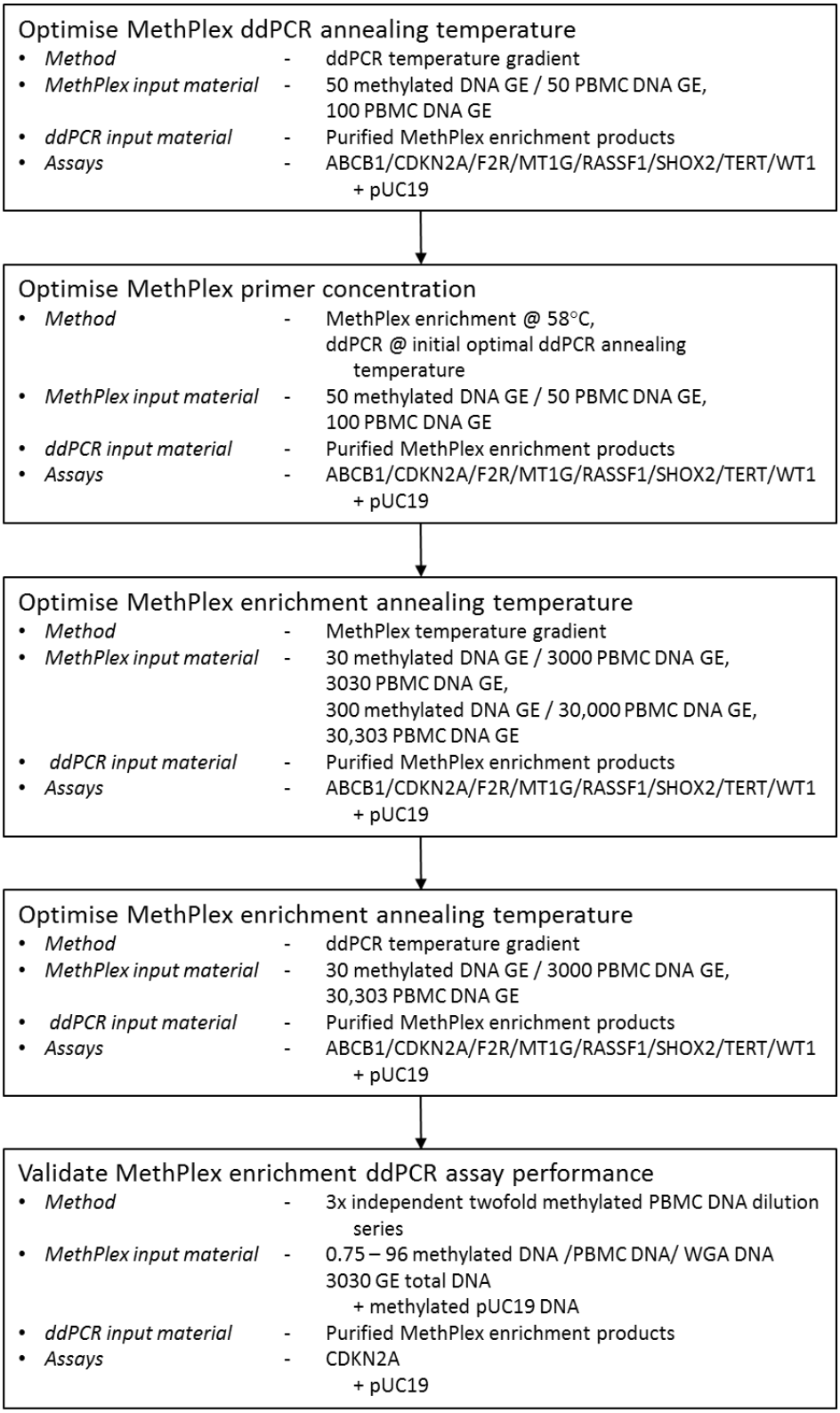


Figure 4.1-3 Target gene promoter MethPlex ddPCR optimisation workflow
Schematic showing key steps in the optimisation of the MethPlex DNA enrichment and subsequent detection by methylation-specific ddPCR. Corresponding chapter sections are noted in bold to the right of the figure.

4.2 Materials and methods

4.2.1 pUC19 spike-in control production and engineering

The assay in development required digestion of the target DNA sample with the DNA methylation sensitive endonuclease HinP1I and subsequent bisulphite conversion. Therefore, the pUC19 vector control required appropriate modification to accommodate this pre-treatment. In addition, our previous experience in similar control assay development indicated that efficient PCR amplification requires linearization of the plasmid. Hence, the plasmid was initially linearized by SspI endonuclease and then subjected to in vitro methylation by SssI methylase. The latter is necessary to block the conversion of cytosines in CpGs in order to increase the sequence complexity of the control and facilitate assay design. ddPCR primers and hydrolysis probe were designed to target the bisulphite-converted sequence containing methylated CpG. Primers (pUC19-dd_F: 5'-AGCGGATGTCGGGAGTAGATAAG-3', pUC19-dd_R: 5'-GCATCTATACGATATTTCACACCGC-3') were obtained from Eurofins Genomics and VIC™ dye-labelled TaqMan® MGB probe (pUC19-mgb_P: 5' -VIC™- TTAGGGCGCGTTAGC-3') was from Applied Biosystems and were designed to anneal to and amplify a 140bp region of modified pUC19 (see **4.2.1.2** and **Figure 4.2-1**).

4.2.1.1 pUC19 plasmid amplification in bacterial culture

One Shot™ TOP10 Chemically Competent *E. coli* (Invitrogen) transformed with pUC19 vector (NEB) were grown at 37°C for 16 hours in Luria-Bertani (LB) broth containing 100µg/ml ampicillin. Isolation of pUC19 plasmid was performed using the Zyppy™ Plasmid Midiprep Kit (Zymo Research) following the manufacturer's vacuum protocol. The plasmid DNA was eluted with 0.1x TE, and purity and

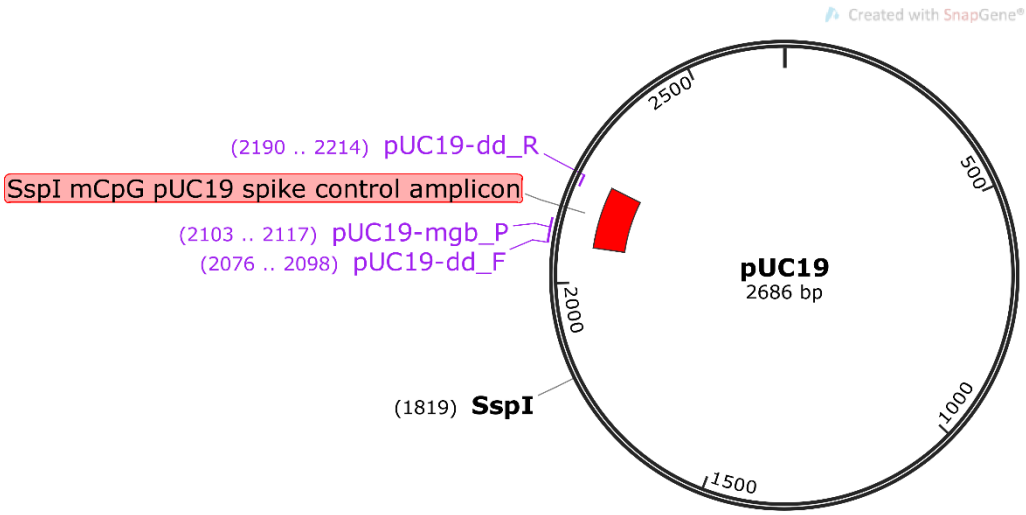


Figure 4.2-1 pUC19 restriction map
Restriction map of pUC19 plasmid vector indicating SspI restriction site and CpG methylated amplicon position.

concentration were measured using a Nanodrop 2000 (Thermo Scientific) (see **2.3.1**) and Quant-iT™ Broad-Range dsDNA Assay Kit (Invitrogen) (see **2.3.2.1**) respectively. Purified pUC19 DNA was stored at -20 °C.

4.2.1.2 Linearization and *in vitro* methylation of pUC19 plasmid

pUC19 DNA (1 µg) was linearized by restriction endonuclease digestion with SspI in reactions containing 1 x NEBuffer™ 2 Buffer, 10 U SspI (NEB (UK)) and ddH_2O in a 50 µl total volume. Reactions were incubated at 37 °C for 1 hour and the enzyme was heat inactivated at 65 °C for 20 minutes. Reaction volume was then increased to 100 µl by the addition of *in vitro* methylation reagents, resulting in a reaction containing 640 µM S-adenosyl methionine (SAM), 1x NEBuffer™ 2 Buffer, 20 U M.SssI CpG methyltransferase DNA (NEB (UK)) . Reactions were incubated at 37 °C. Since SAM is consumed fast during the reaction, an additional 2 µl 32 mM SAM was added after two hours and the reaction was incubated for a further two hours

followed by heat inactivation at 65 °C for 20 minutes. SspI-digested, CpG methylated pUC19 DNA (SspI mCpG pUC19 DNA) was then purified using QIAquick PCR Purification Kit (**2.11.2.1**). Purified DNA was either analysed immediately or stored at -20 °C. DNA purity and concentration was measured using a Nanodrop 2000 (Thermo Scientific) (**2.3.1**) and Quant-iT™ Broad-Range dsDNA Assay Kit (Invitrogen) (**2.3.2.1**). Linearization of plasmid was confirmed by agarose gel electrophoresis (**2.8**).

For ease of comprehension SspI mCpG pUC19 DNA is referred to as "methylated pUC19 DNA" from this point forward.

4.2.2 Measurement of methylation levels of *in vitro* methylated PBMC DNA

CpG dinucleotide methylation positive control DNA for assay optimisation and production of methylated DNA dilution series was produced by *in vitro* methylation of pooled blood bank PBMC DNA using M.SssI CpG methyltransferase (NEB (UK)) as outlined in **2.4**. Efficacy of CpG methylation by the enzyme was assessed by Pyrosequencing Methylation Analysis (**2.9**).

Pyrosequencing assays targeting two gene promoters within the prospective ddPCR biomarker panel, *CDKN2A* and *MT1G*, and one gene external to the panel, *RASSF4*, were used to analyse M.SssI-treated and untreated DNA. Pyrosequencing PCR amplification and analysis was carried out in triplicate and representative pyrograms are presented in **Figure 4.2-2** and average levels of methylation are shown in **Table 4.2-1**. The mean average of methylation of the three promoter assays was used to inform the preparation of methylated DNA dilution series and calculation of methylated copies present in positive controls.

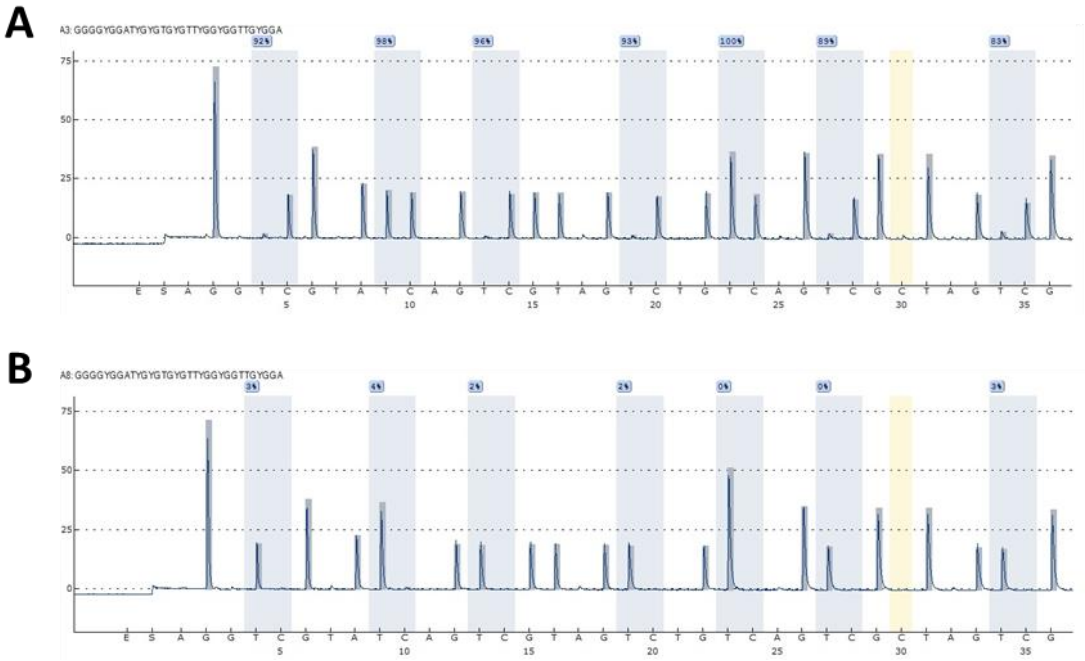


Figure 4.2-2 Representative pyrograms
CDKN2A pyrograms from **A.** M.SssI-treated DNA and **B.** PBMC DNA. The x-axis shows the nucleotide dispensation order. The y-axis shows the amount of light by the chemiluminescent reactions when each nucleotide is incorporated onto the growing DNA template strand. The light blue columns indicate the interrogated CpG dinucleotides. A bisulphite conversion control is highlighted in pale yellow. Percentage methylation of each CpG is indicated in the small blue boxes towards the top of each panel. The percentage of methylation is calculated as the C/(C + T) peak ratio per CpG. Taking the first examined CpG as an example, in panel **A.** there is a clear peak corresponding with the C dispensation, indicating (92%) methylation. In panel **B**. the peak coincides with a T dispensation, indicating C-T bisulphite conversion and the absence of methylation.

Table 4.2-1 Methylation % of M.SssI-treated DNA and PBMC DNA measured by Pyrosequencing

	M.SssI-treated DNA		PBMC DNA	
Assay	**Methylation%**	**SD**	**Methylation%**	**SD**
CDKN2A	91.2	5.3	1.9	1.8
MT1G	90.2	5.6	6.3	5.3
RASSF4	88.1	5.4	0.9	1.7
Mean average	89.8	2.4	3.0	1.6

4.2.3 *ACTB* cfDNA real-time PCR quantification assay optimisation

ACTB cfDNA qPCR assay annealing temperature and primer/probe concentrations were optimised using a temperature gradient and 500 nM, 750nM and 900nM primer concentrations and probe concentrations ranging from 100 nM to 250 nM. Primers (*ACTB*re_F: 5'- GACTGTGAACCTGTGTCTGCCA-3', *ACTB*re_R: 5'-CCAGTTCACGGTACAAGGCTG-3') and FAM™ dye-labelled TaqMan® probe (*ACTB*re_P: FAM™- 5'-CTCAGCCAATGGGACCTGCTCCTC-3'-BHQ1) were obtained from Eurofins Genomics. Reactions contained 1 x TaqMan® Universal Master Mix II (Applied Biosystems), variable concentrations of primers and probe as indicated above, and 50 ng PBMC DNA. Triplicate qPCR reactions were thermal cycled on a 7500 FAST real-time PCR instrument (Applied Biosystems) with an initial DNA denaturation and enzyme activation at 95 °C for 10 minutes and 45 cycles of 95 °C for 15 seconds and annealing/extension steps ranging from 57°C to 62°C for 60 seconds. Non-template controls were run in parallel and produced no amplification. Optimal annealing temperature was selected on the basis of the highest temperature resulting in the lowest consistent Cq value and no nonspecific amplification. The lowest primer and probe concentrations resulting in no increase in Cq value were considered to be optimal.

4.2.4 HinP1I restriction endonuclease digestion of DNA

Previous work in our group indicated that digestion of DNA with methylation-sensitive restriction endonuclease HinP1I can increase analytical specificity of qMSP assays containing HinP1I restriction sites (**Figure 4.2-3)**. All methylation-specific target amplicons tested within study contained at least one HinP1I restriction site.

Prior to bisulphite conversion and ddPCR analysis, all samples were digested with HinP1I according to the methodology laid out in **2.4**.

Figure 4.2-3 HinP1I restriction site

4.2.5 MethPlex enrichment PCR amplification

MethPlex pre-amplification was performed according to general procedures outlined in section **2.11**.

4.2.6 Droplet digital™ PCR (ddPCR)

ddPCR was performed according to procedures outlined in section **2.12**.

Methylated target primers and probes had the same sequences as qMSP oligonucleotides presented in **Table 2.10-1** with the addition of an assay targeting a 130 bp region of the *SHOX2* gene promoter:

*SHOX2*_F: 5'- TTTGTTGGTAGAGGTTGAGCGTC -3'

*SHOX2*_R: 5'- CTCCAACACCTCCCGATACG -3'

*SHOX2*_P: FAM™- 5'- AATCGCCTCCTTCTTCTCCTT -3'

TaqMan® probes were labelled with FAM™. pUC19 control primers and VIC™ probe sequences were as above (**4.2.1**).

4.3 Results & Discussion

4.3.1 Methylated pUC19 spike-in control assay optimisation and validation

4.3.1.1 pUC19 temperature gradient with methylated pUC19 DNA input only

Initial experiments assessed the optimal annealing temperature for the pUC19 ddPCR control assay in singleplex format with three different input quantities (copy numbers) of linearized, methylated pUC19 DNA template, namely, 1.5×10^5, 1.5×10^4 and 1.5×10^3. These tenfold dilutions were based on methylated pUC19 concentrations determined by dsDNA fluorescence quantification (**2.3.2.1**). Annealing temperatures ranged from 57°C to 62°C (**Figure 4.3-1** and **Figure 4.3-2**). Visualisation of droplet fluorescence amplitude in 1D format displayed single, discrete positive and negative populations indicative of single amplification products and high assay specificity. Manual selection of optimal annealing temperature is a somewhat subjective process based on the separation of positive droplets from the negative cluster. **Figure 4.3-1** and **Figure 4.3-2** show that separation of positive and negative droplet clusters for the pUC19 ddPCR assay did not increase to any great extent within the gradient range of this experiment.

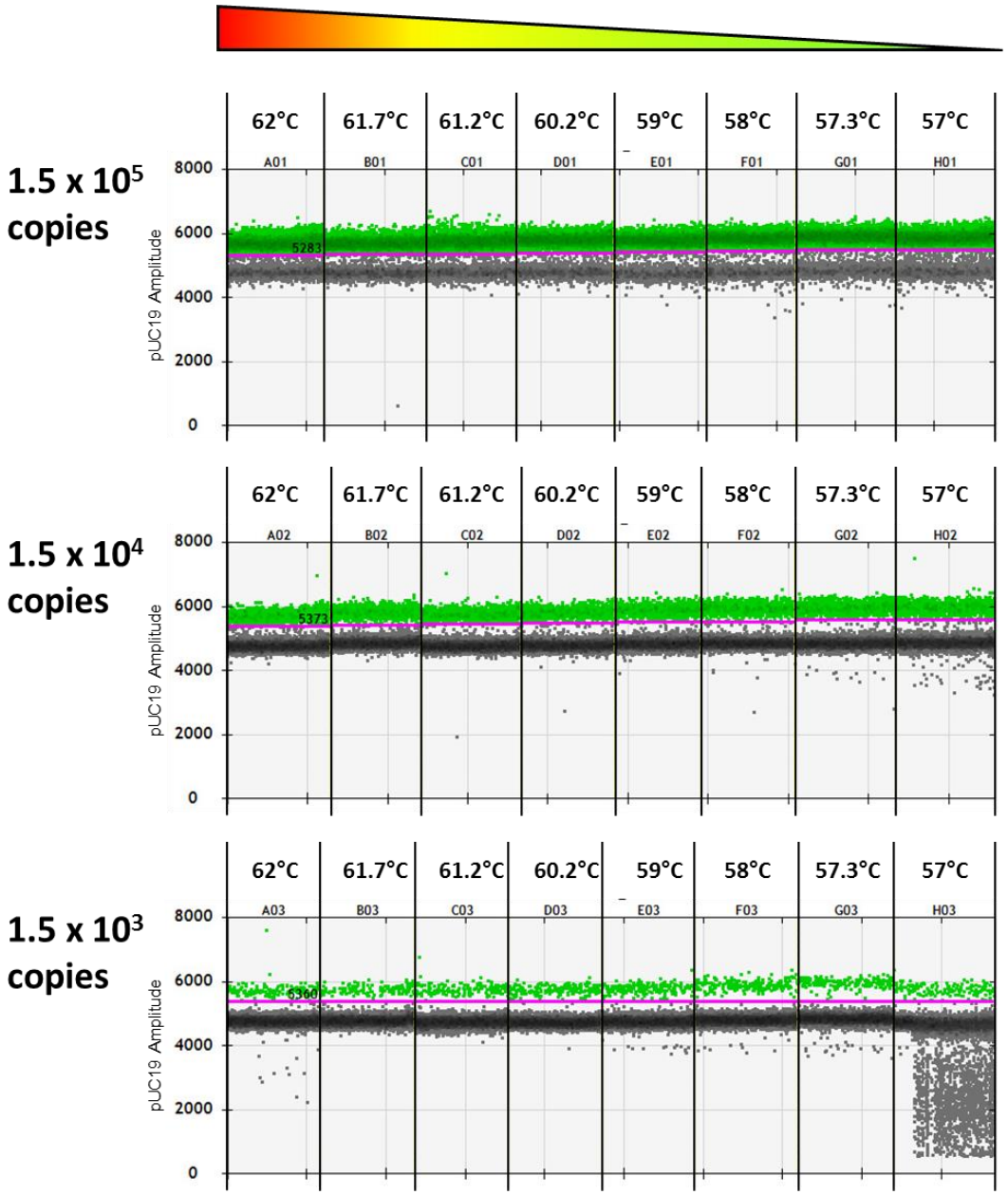


Figure 4.3-1 Annotated QuantaSoft ddPCR 1D output of methylated pUC19 temperature gradient at eight different temperatures and three different nominal input amounts

pUC19 positive droplets (VIC) are indicated in green and negative droplets appear as dark grey. Solid fuchsia horizontal line indicates fluorescent amplitude threshold level. Each segment between solid vertical black lines contains data points for >10, 000 individual ddPCR partitions.

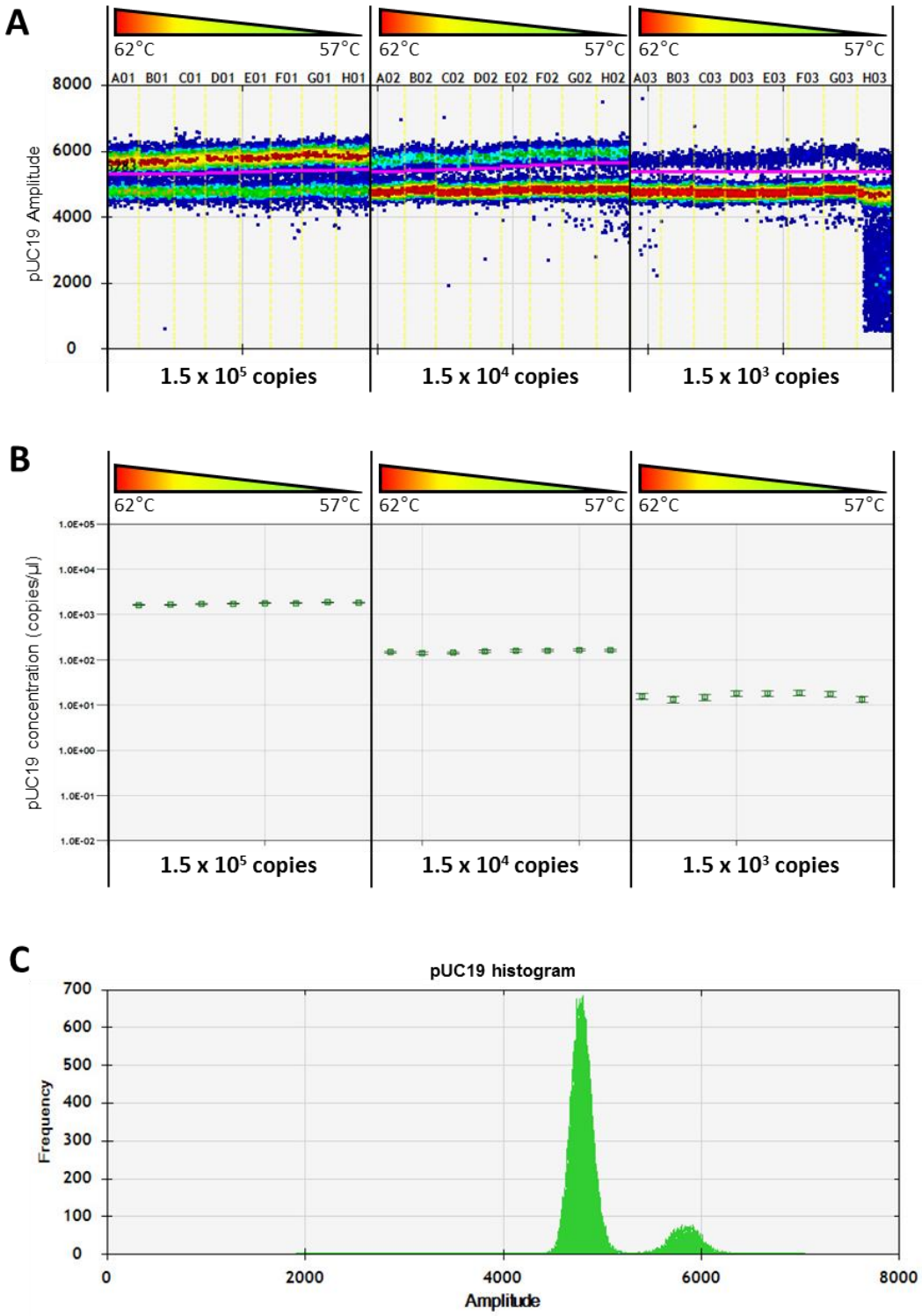


Figure 4.3-2 Modified QuantaSoft ddPCR graphical outputs of methylated pUC19 temperature gradient

A. Modified QuantaSoft ddPCR 1D output with false colour heat mapping illustrating the same information conveyed in **Figure 4.3-1** with red indicating high density of droplets and blue indicating lower droplet density. Error bars represent 95% confidence intervals derived from the Poisson distribution. **B.** ddPCR concentration measurements with varying methylated pUC19 input and temperature. **C.** Density histogram of ddPCR measured amplitude for 1.5 x 10^4 nominal copies methylated pUC19 input at an annealing temperature of 60.2°C.

False colour heat mapping of droplet density in QuantaSoft software (**Figure 4.3-2A**) showed clear differences in positive droplet density between different methylated pUC19 inputs but no clear difference between temperatures.

The representative density histogram in **Figure 4.3-2C** indicates that the assay exhibits good resolution in terms of the separation of the positive and negative clusters. Graphical plots of temperature against measured concentration (facetted by input amount) similarly indicate clear differences in measured pUC19 concentration between different input amounts (**Figure 4.3-2B**). Similar concentration measurements at different temperatures were observed for the same input quantity (**Figure 4.3-2B**). For the lowest template input, there was some variability in the measured concentration, exhibiting a coefficient of variation of 12.4%. The higher template amounts, namely 1.5×10^5 and 1.5×10^4, showed greater precision with coefficients of variation of 4.7% and 5.9% respectively. A 15% CV tolerance level for this control assay was selected *a priori* and the level of precision was therefore satisfactory. For all template inputs differences in measured concentration did not follow any observable trend, appearing to be randomly distributed.

Since separation of positive and negative clusters did not increase with decreasing annealing temperature, the annealing temperature (60.2°C) producing the measured concentration closest to the mean average across temperatures for 1.5×10^5 and 1.5×10^4 copies input was selected as being optimal (**Table 4.3-1**).

Table 4.3-1 pUC19 concentrations measured by ddPCR and absolute difference from their mean average across annealing temperatures at different inputs.

	Methylated pUC19 input (copies)			
	1.5×10^5		1.5×10^4	
Annealing temperature (°C)	Measured pUC19 concentration (copies/µl)	Absolute difference from mean (copies/µl)	Measured pUC19 concentration (copies/µl)	Absolute difference from mean (copies/µl)
62	1639	113.9	147	6.6
61.7	1652	100.9	141	12.6
61.2	1715	37.9	143	10.6
60.2	1734	18.9	153	0.6
59	1794	41.1	158	4.4
58	1797	44.1	160	6.4
57.3	1868	115.1	165	11.4
57	1824	71.1	162	8.4
mean	1752.9	-	153.625	-

4.3.1.2 Methylated pUC19 assay singleplex performance in the pure template context

To assess pUC19 assay linear dynamic range and potential inherent assay background, duplicate ddPCR reactions (T_a = 60.2°C) were carried out using a seven point, tenfold methylated pUC19 dilution series ranging from 2×10^6 to 2×10^0 copies reaction input, diluted in 0.1x TE. No negative droplets were detected at the 2×10^6 copies input level, making it impossible to measure target concentration (i.e. reactions were saturated). However, as all droplets were positive at this level, this simplified the setting of the threshold for target positive droplets. Only one positive droplet in one of two replicates was observed for 2×10^0 copies input (**Figure 4.3-3**). No amplification was detected in non-template controls.

Linear regression analysis of measured concentration against methylated pUC19 input in the 2 x 10^0 to 2 x 10^5 range (log_{10}-transformed data) revealed more favourable parameters than those exhibited in the 2 x 10^1 to 2 x 10^5 range (**Table 4.3-2**). The slope for the limited regression range was close to unity (1.01, 95% CI [0.94, 1.08]) indicating a tenfold increase in measured concentration when methylated pUC19 input increased tenfold. Limiting the range also increased the R^2 value to 0.993 and halved the root mean squared error (**Figure 4.3-3B** and **Table 4.3-2)**, indicating that this model explains a greater degree of variance than the extended range and provides a better fit to the data.

Table 4.3-2 Methylated pUC19 linear regression parameters from tenfold dilution series

Parameter	Input range (copies)	
	2 x 10^0 – 2 x 10^5	**2 x 10^1 – 2 x 10^5**
β_0 (95% CI)	-1.86	-1.94
	(-2.18, -1.54)	(-2.19, -1.69)
β_1 (95% CI)	0.99	1.01
	(0.89, 1.09)	(0.94, 1.08)
R^2	0.981	0.993
RMSE	0.236	0.123

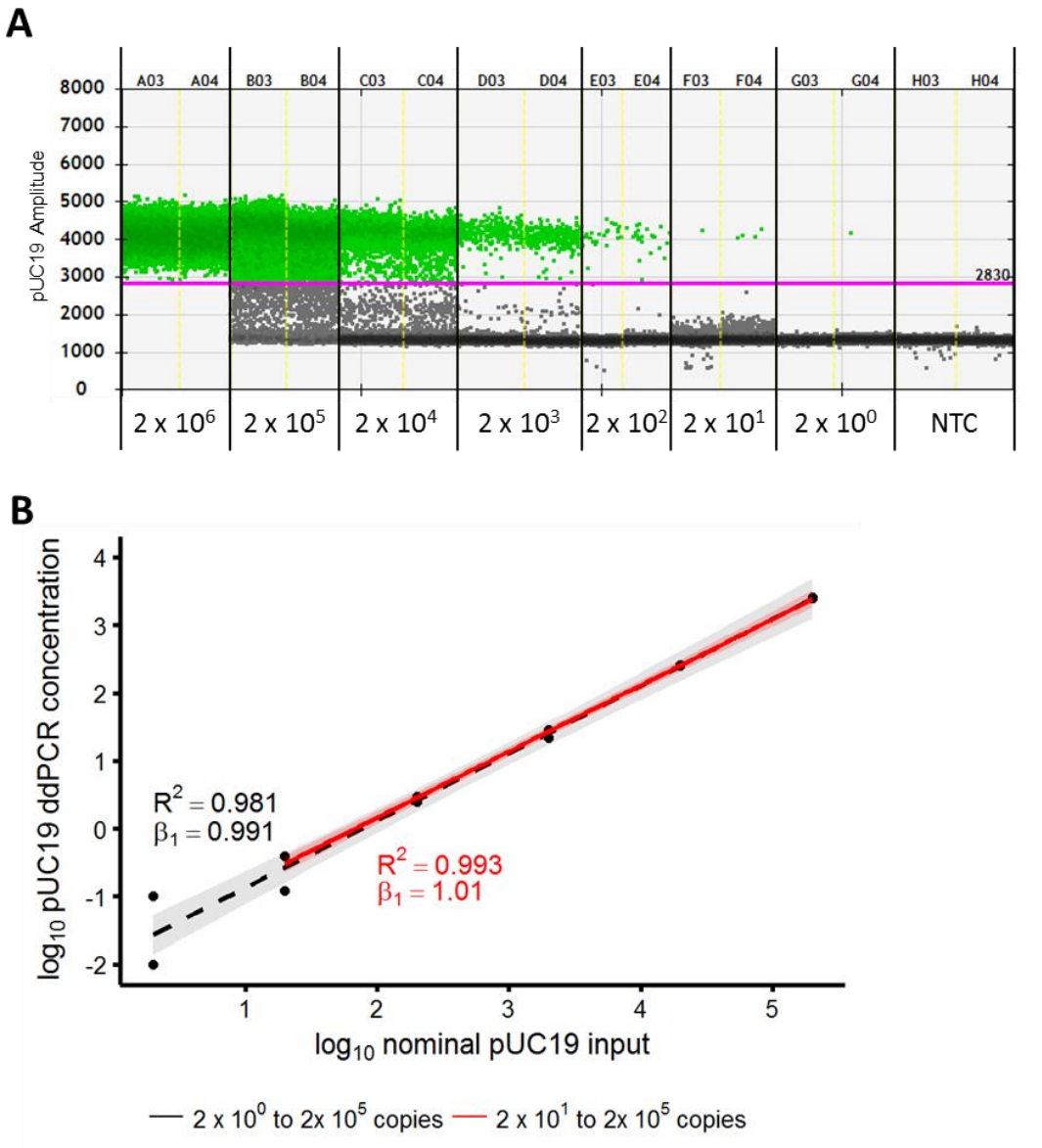


Figure 4.3-3 Tenfold methylated pUC19 dilution series QuantaSoft 1D output, scatterplots and linear regression

A. Methylated pUC19 positive droplets (VIC) are indicated in green and negative droplets appear as dark grey. Solid fuchsia horizontal line indicates fluorescent amplitude threshold level. Each segment between solid vertical black lines contains data points for two wells containing >10, 000 individual ddPCR partitions each. **B.** Scatterplot of $\log_{10}$-transformed ddPCR measured concentration against $\log_{10}$-transformed nominal methylated pUC19 input with linear regression lines superimposed. Shaded areas around regression lines indicate 95% confidence band.

4.3.1.3 Methylated pUC19 multiplex assay performance in the presence of methylated/unmethylated PBMC DNA background and a methylation-specific ddPCR assay

To evaluate the methylated pUC19 assay when competing with methylation-specific ddPCR assay in a complex background matrix more representative of clinical samples, triplicate duplex ddPCR reactions targeting methylated pUC19 and methylated *CDKN2A* gene promoter (T_a = 59°C, *CDKN2A* optimal annealing temperature (**Table 4.3-5**)) were carried out using a seven point, tenfold methylated pUC19 dilution series in the presence of:

- 66 ng 1.5% methylated PBMC DNA, equivalent to 300 methylated GE in a total of 20,000 GE PBMC DNA (see **4.3.1.3.1**)
- 13.3 ng 1% methylated PBMC DNA, equivalent to 40 methylated DNA GE in a total of 4000 GE PBMC DNA (see **4.3.1.3.2**).

4.3.1.3.1 Methylated pUC19 multiplex assay performance: 66 ng 1.5% methylated PBMC background

Measurement of methylated pUC19 concentration by ddPCR was assessed with a moderately high methylated/untreated PBMC DNA input to mimic assay conditions in clinical samples from body fluids yielding relatively high DNA yields (e.g. bronchoalveolar lavage). Methylated pUC19 inputs ranged from 10^0 to 10^6 copies per well. Graphical output of ddPCR measurements from QuantaSoft software is shown in **Figure 4.3-4A** and visually indicates a linear relationship between methylated pUC19 input and measured concentration, and stable *CDKN2A* concentration across pUC19 inputs. Initial regression analysis of data obtained using inputs from 10^1 to

10^6 showed good characteristics (**Figure 4.3-4B**, **Table 4.3-3**). Reduction in the range of methylated pUC19 analysed by linear regression to include only data above

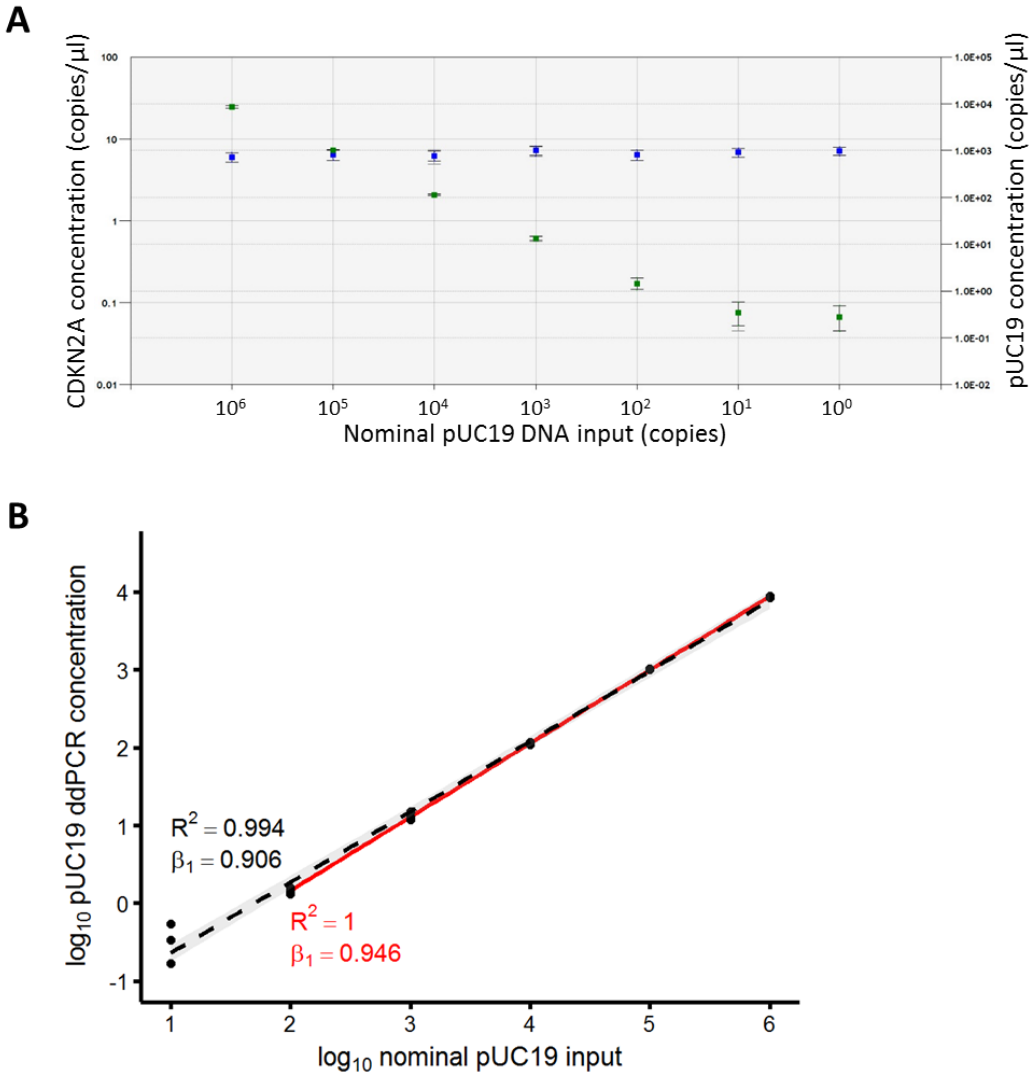


Figure 4.3-4 Tenfold methylated pUC19 dilution series, 66ng 1.5% methylated PBMC DNA background: *CDKN2A* and pUC19 ddPCR measured concentration and linear regression
A. QuantaSoft concentration output. Coloured point indicate measured concentration at given input calculated from counts in merged meta-wells. Darker error bars represent 95% Poisson confidence interval, lighter error bars represent overall 95% confidence interval.
B. Scatterplots of $\log_{10}$-transformed ddPCR measured methylated pUC19 concentration against nominal methylated pUC19 input with linear regression lines superimposed. Black dashed regression line and text correspond to 10^1 – 10^6 input range, red regression line and text correspond to 10^2 – 10^6 input range. Shaded areas around regression lines indicate 95% confidence bands.

10^2 copies input brought the regression slope closer to unity, while narrowing the confidence interval of the slope (β_1 = 0.95 (95% CI [0.93, 0.96]), indicating that a

tenfold increase in input results in a measurement closer to tenfold higher in the reduced model. Reduction in range also increased the coefficient of determination to 0.999 and reduced the root mean squared error fourfold (**Table 4.3-3**), indicating better linear regression model fit. Reducing the range to include 10^2 to 10^6 copies input also reduced the within-subject CV (i.e. within the same methylated pUC19 input) from 22.7% to 7.7% and the between-subject CV (i.e. across methylated pUC19 inputs) from 10.9% to 10.0%. *CDKN2A* concentration measured by ddPCR remained stable across the range of methylated pUC19 template inputs with a mean value of 6.60 copies/µl (95% CI [6.27, 7.03]) and CV of 7.1%.

Table 4.3-3 Methylated pUC19 linear regression parameters from tenfold dilution series in the presence of 20,000 GE PBMC (DNA 300 GE methylated DNA) in duplex ddPCR targeting *CDKN2A* and pUC19

Parameter	Input range (copies)	
	$10^1 – 10^6$	**$10^2 – 10^6$**
β_0 (95% CI)	-1.54	-1.79
	(-1.68, -1.40)	(-1.78, -1.68)
β_1 (95% CI)	0.91	0.95
	(0.87, 0.94)	(0.93, 0.96)
R^2	0.994	0.999
RMSE	0.118	0.028

Taken together, the above indicates that the methylated pUC19 control assay has a linear dynamic range from 10^2 to 10^6 copies input in the presence of 20,000 GE PBMC DNA when simultaneously assaying for methylated *CDKN2A* with a nominal methylated DNA target input of 300 GE. Within this range the pUC19 assay showed good linearity and expected tenfold increases in measured concentration when

template input was increased tenfold. In addition, the pUC19 assay did not affect the detection of methylated *CDKN2A* at this input level.

4.3.1.3.2 Methylated pUC19 multiplex assay performance: 13.3 ng 1% methylated PBMC background

In order to better emulate assay conditions expected when using blood plasma as a clinical sample of interest, determination of methylated pUC19 concentration by ddPCR was assessed in a low methylated/untreated PBMC DNA background (40 methylated DNA GE in a total of 4000 GE PBMC DNA input per well). Methylated pUC19 inputs ranged from 2×10^0 to 2×10^6 copies per well. Non-template controls were run in parallel and no amplification was observed.

Graphical output of ddPCR measurements from QuantaSoft software is shown in **Figure 4.3-5A** and visual inspection implied a linear relationship between methylated pUC19 input and measured concentration, and fairly stable *CDKN2A* concentration across pUC19 inputs. No negative droplets were detected at the 2×10^6 copies input level, making it impossible to measure target concentration (i.e. reactions were saturated, QuantaSoft software returns 1,000,000 copies/µl by default). ddPCR wells nominally containing two methylated pUC19 copies were also negative. Initial regression analysis of data obtained using the full range of inputs showed excellent characteristics with an R^2 of 0.999 and a slope indistinguishable from unity ($\beta_1 = 0.99$ (95% CI [0.969, 1.01]) (**Figure 4.3-5B** and **Table 4.3-4**). Model parameters did not change upon reducing the range to $2 \times 10^2 – 2 \times 10^5$ but their confidence intervals were slightly reduced.

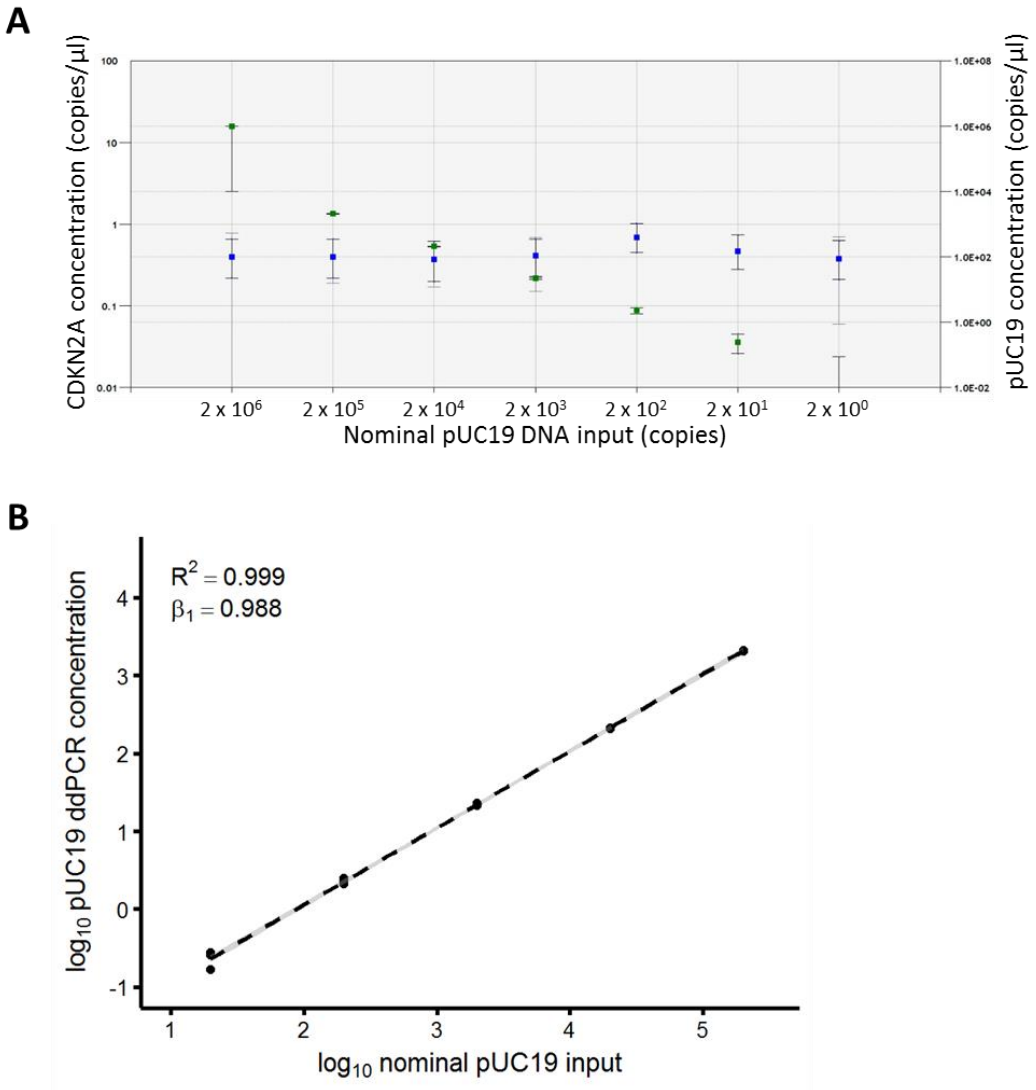


Figure 4.3-5 Tenfold methylated pUC19 dilution series, 13.3ng 1% methylated PBMC DNA background: *CDKN2A* and pUC19 ddPCR measured concentration and linear regression

A. QuantaSoft concentration output. Coloured point indicate measured concentration at given input calculated from counts in merged meta-wells. Darker error bars represent 95% Poisson confidence interval, lighter error bars represent overall 95% confidence interval.
B. Scatterplots of log_{10}-transformed ddPCR measured methylated pUC19 concentration against nominal methylated pUC19 input with linear regression lines superimposed. Black dashed regression line and text correspond to $2 \times 10^1 – 2 \times 10^6$ input range. Grey shaded area around regression lines indicates 95% confidence band.

Table 4.3-4 Methylated pUC19 linear regression parameters from tenfold dilution series in the presence of 4,000 GE PBMC (DNA 40 GE methylated DNA) in duplex ddPCR targeting *CDKN2A* and pUC19

Parameter	Input range (copies)	
	$2 \times 10^1 - 2 \times 10^5$	**$2 \times 10^2 - 2 \times 10^5$**
β_0 (95% CI)	-1.921	-1.920
	(-1.991, -1.852)	(-1.967, -1.874)
β_1 (95% CI)	0.988	0.988
	(0.969, 1.01)	(0.976, 1)
R^2	0.999	1
RMSE	0.046	0.019

Reduction in the range of linear regression increased R^2 to 1 and reduced RMSE from 0.046 to 0.019 (**Table 4.3-4**), indicating a marginally better fit for the model with reduced pUC19 input range. Reduction in analysed input range also reduced the within-subject CV (i.e. within the same methylated pUC19 input) from 12% to 5.2% and the between-subject CV (i.e. across methylated pUC19 inputs) from 8.4% to 6.8%. This indicates that the methylated pUC19 control assay has a linear dynamic range from 2×10^1 to 2×10^5 copies input in the presence of 4,000 GE PBMC DNA when simultaneously assaying for methylated *CDKN2A* with a nominal methylated DNA target input of 40 GE. The pUC19 assay did not affect the detection of methylated *CDKN2A* at this input level.

Considering these findings and those from **4.3.1.1**, **4.3.1.2** and **4.3.1.3.1**, it is apparent that the methylated pUC19 control assay functions well in the 10^2 to 2×10^5 copy ddPCR input range, displaying linearity and high precision. The assay also performs well across a range of different annealing temperatures. The assay is also specific and

neither interferes with nor is affected by multiplexing with the *CDKN2A* assay. Use of methylated pUC19 spike in ddPCR workflows at amounts resulting in downstream ddPCR inputs within this range should be acceptable for use as a control to assess extraction recovery and amplifiable DNA losses induced by sample manipulations required in the analysis of methylated DNA.

4.3.2 Methylated DNA gene promoter ddPCR assay optimisation

4.3.2.1 Duplex assay annealing temperature optimisation

Optimal annealing temperatures for methylation-specific ddPCR assays in duplex format in conjunction with methylated pUC19 control assay (see **4.3.1**) were assessed using temperature gradients. Initial experiments assessed the amplification detected in ddPCR reactions containing 6.6 ng HinP1I digested, bisulphite-treated PBMC DNA. Methylation positive wells contained 50% *in vitro* methylated DNA (1000 GE methylated DNA / 1000 GE untreated PBMC DNA); methylation negative wells contained 2000 GE untreated PBMC DNA alone. All reactions were also spiked with 6000 copies bisulphite-treated methylated pUC19 DNA. The temperature gradient comprised eight annealing temperatures ranging from 57°C to 62°C (**Figure 4.3-6**). Optimal annealing temperatures were selected considering the following:

- No amplification in methylation negative reactions
- Separation of positive droplets for the methylated target assay
- Separation of positive droplets for the methylated pUC19 assay
- Maximum concentration of methylated target measured by ddPCR
- Consistent methylated pUC19 concentration between methylation positive and negative wells.

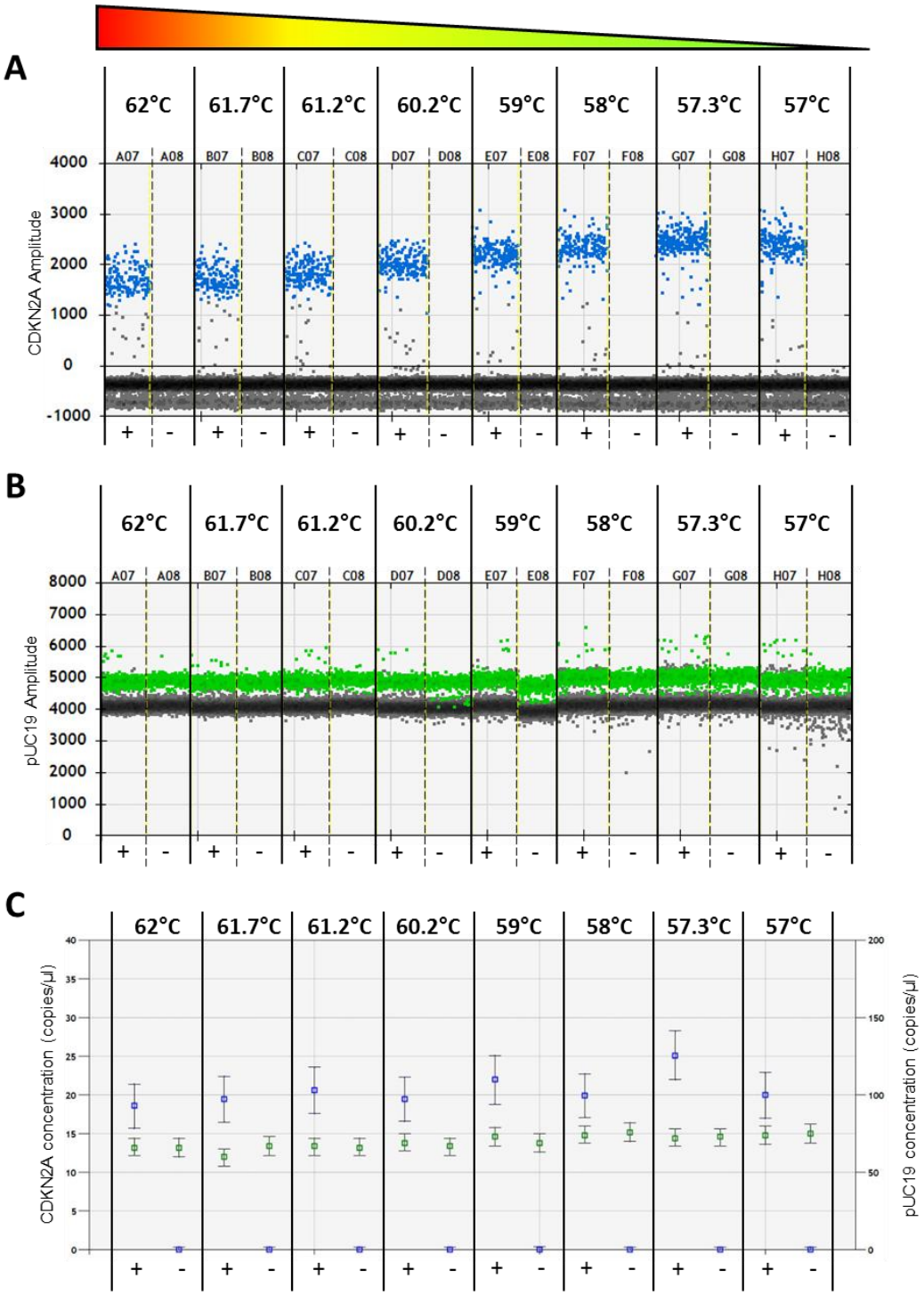


Figure 4.3-6 Modified QuantaSoft ddPCR 1D and concentration output of *CDKN2A*-pUC19 duplex temperature gradient at eight different temperatures
ddPCR fluorescence amplitude and positive/negative droplet identification for **A.** *CDKN2A* and **B.** pUC19. *CDKN2A* positive (FAM) positive droplets are indicated in blue, methylated pUC19 positive droplets (VIC) are indicated in green and negative droplets appear as dark grey. Each segment between solid and dashed vertical black lines contains data points for >10, 000 individual ddPCR partitions. **C.** QuantaSoft concentration output. Blue points correspond to *CDKN2A* concentration, green points correspond to pUC19 concentration. Error bars represent 95% Poisson confidence interval.

Representative QuantaSoft 1D amplification plots for the *CDKN2A*-pUC19 duplex ddPCR assay (**Figure 4.3-6A**) indicate no amplification in methylation negative wells and increasing separation of positive and negative droplet populations in methylation positive wells. *CDKN2A* concentration was consistent across the temperature range, when taking the margin of error into account, while the maximum *CDKN2A* concentration measurement was observed at 57.3°C (**Figure 4.3-6C**). Methylated pUC19 positive/negative cluster separation was consistent across the temperature gradient range (**Figure 4.3-6B**) and measured concentration was consistent between methylation positive and negative wells respectively (**Figure 4.3-6C**). The initial optimal annealing temperature assigned to the *CDKN2A*-pUC19 assay was 59°C.

Modified 1D amplification plots and QuantaSoft ddPCR concentration plots for all target gene promoters can be found in **Appendix, Figure 8.1** and **Appendix, Figure 8.2**. Optimal annealing temperatures for all targets were verified by further temperature gradient experiments (six annealing temperatures [57°C, 58°C, 59°C, 60.2°C, 61.2°C, 62°C) in duplicate. All target assay annealing temperatures were successfully verified, with the exception of *WT1*-pUC19 and *MT1G*-pUC19, both of which produced low levels of amplification in methylation negative reactions (**Figure 4.3-7, Appendix, Figure 8.1** and **Appendix, Figure 8.2**). In the case of these two genes, increasing annealing temperature resulted in no amplification in methylation negative reactions with no accompanying difference in amplification in methylation positive reactions. Optimal annealing temperatures are presented in **Table 4.3-5**.

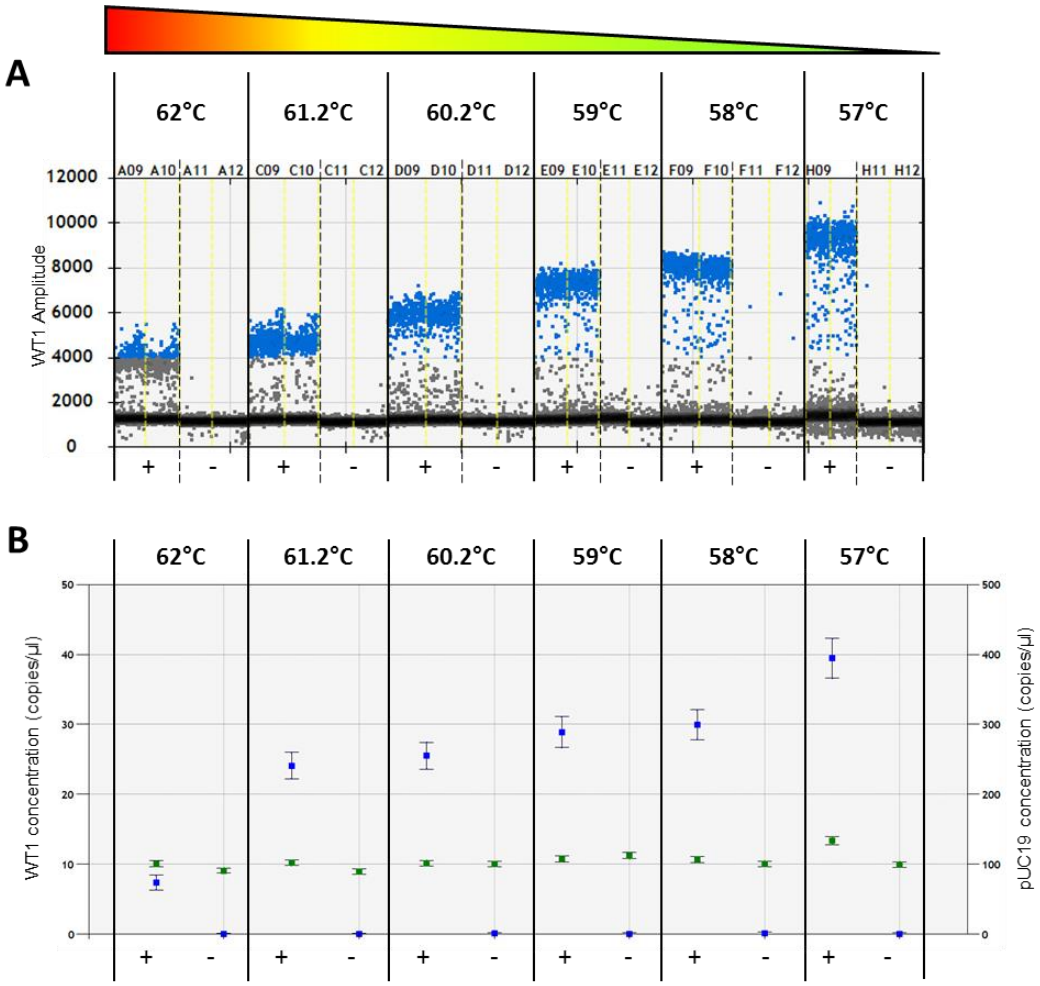


Figure 4.3-7 Modified QuantaSoft ddPCR 1D and concentration output of *WT1*-pUC19 duplex verification temperature gradient at six different temperatures

A. ddPCR fluorescence amplitude and positive/negative droplet classification for *WT1*. *WT1* positive (FAM) positive droplets are indicated in blue, and negative droplets appear as dark grey. Each segment between solid and dashed vertical black lines contains data points from > 22, 000 individual ddPCR partitions. Note positive (blue) droplets in negative reactions at 57°C and 58°C.

B. QuantaSoft concentration output. Blue points correspond to *WT1* concentration, green points correspond to pUC19 concentration. Error bars represent 95% Poisson confidence interval. Concentration values are derived from the merging of duplicate wells as a single meta-well containing > 22,000 droplets.

Table 4.3-5 Initial and final methylation-specific duplex ddPCR optimal annealing temperatures

Target gene	Optimal annealing temperature (°C)	
	Initial	Final
ABCB1	60.2	60.2
CDKN2A	59.0	59.0
F2R	58.0	58.0
MT1G	61.2	62.0
RASSF1	59.0	59.0
SHOX2	60.2	60.2
TERT	60.2	60.2
WT1	58.0	59.0

4.3.2.2 Evaluation of methylated DNA detection and quantification using methylated target-pUC19 duplex assays and complex background matrix

Having identified optimal annealing temperatures for multiplex methylation-specific ddPCR assays, two assays (*CDKN2A*-pUC19 and *ABCB1*-pUC19) from the potential methylation biomarker panel were evaluated further. Duplex assay performance was assessed using a six point, twofold methylated DNA dilution series resulting in ddPCR inputs from 20 GE to 640 GE. ddPCR reactions were run in triplicate. *In vitro* methylated PBMC DNA (**2.4**) was diluted in untreated PBMC DNA of the same concentration. PBMC DNA and WGA DNA were used as biological and technical controls respectively. PBMC DNA was used as the diluent since the most likely potential "contaminating" DNA in bronchoalveolar lavage and blood plasma is that originating from leukocytes, and dilution in water would simply dilute the methylated DNA target without a concomitant increase in potentially competing "unmethylated" DNA copies. While most CpG islands are unmethylated in leukocytes, a large number

can become methylated under certain circumstances (Bock, 2006). It was therefore important to include technical unmethylated controls produced by whole genome amplification (**2.5**) to assess assay specificity.

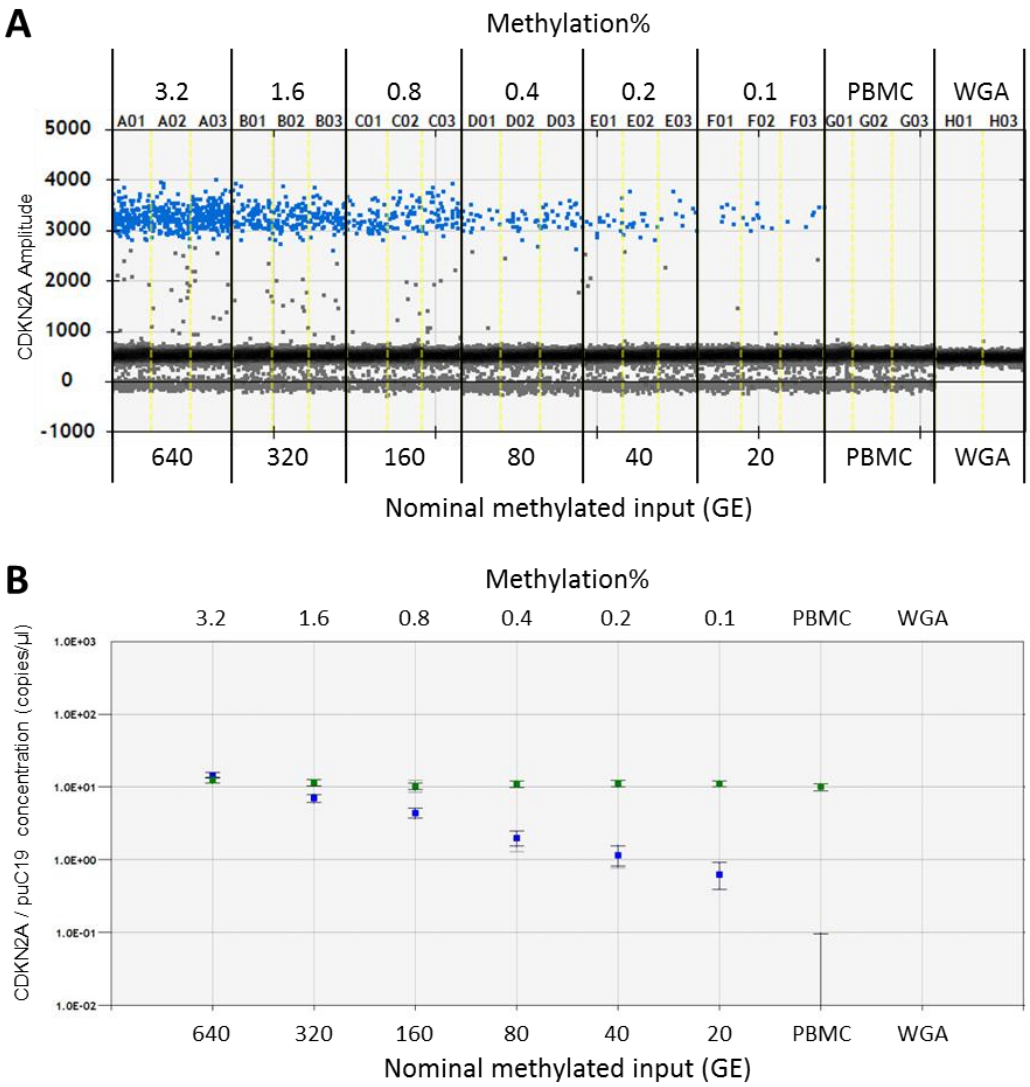


Figure 4.3-8 *CDKN2A*-pUC19 duplex assay performance evaluation using twofold methylated DNA dilution series
A. Modified QuantaSoft ddPCR 1D plot indicating positive/negative droplet classification for *CDKN2A*. *CDKN2A* positive (FAM) positive droplets are indicated in blue, and negative droplets appear as dark grey. Each segment between solid vertical black lines contains data points from > 27, 500 individual ddPCR reactions.
B. QuantaSoft concentration output. Blue points correspond to *CDKN2A* concentration, green points correspond to pUC19 concentration. Light grey error bars represent overall 95% confidence intervals between replicate ddPCR wells and darker error bars represent 95% Poisson confidence intervals. Concentration values are derived from the merging of duplicate wells as a single meta-well containing > 27,500 droplets.

No amplification was observed in either technical or biological methylation negative DNA controls for either gene promoter, indicating good analytical specificity. Amplification was detected at all levels of methylated DNA serial dilution, indicating a limit of detection (LOD) below 20 GE (0.1% methylated) for both assays (see Chapter 5 for a more rigorous evaluation of LOD). Initial inspection of 1D ddPCR plots indicated that the number of positive droplets decreased as methylated DNA was serial diluted with unmethylated DNA (**Figure 4.3-8A** and **Figure 4.3-9A**). Concentration plots in QuantaSoft software suggested a linear relationship between methylated DNA input and ddPCR measured methylated gene promoter concentration (**Figure 4.3-8B** and **Figure 4.3-9B**). Methylated pUC19 concentration was also consistent across reactions (*CDKN2A*: mean pUC19 concentration = 11.07 copies/µl, SD = 1.10; *ABCB1*: mean pUC19 concentration = 10.80 copies/µl, SD = 1.41). Quantitative assessment of linearity by regression analysis of technical replicates including linear and quadratic terms, as suggested by Vynck et al. (Vynck et al., 2017), confirmed a linear relationship between nominal methylated DNA input and measured concentration across the full dilution series for *CDKN2A* ($F(1, 15) = 0.155$, $P = 0.699$). Similar evaluation of the *ABCB1*-pUC19 assay indicated deviation from linearity when considering the full dilution series range ($F(1, 15) = 5.76$, $P = 0.030$) but supported a linear relationship in the range from 40 to 640 methylated DNA GE input (0.2% to 3.2%) ($F(1, 12) = 0.45$, $P = 0.516$) (**Figure 4.3-10**). Regression analysis of log2 transformed data provided slopes approaching unity and indicated that a doubling in methylated DNA input resulted in a near doubling in measured concentration (**Table 4.3-1**). Both assays also returned models with coefficients of determination greater than 0.95.

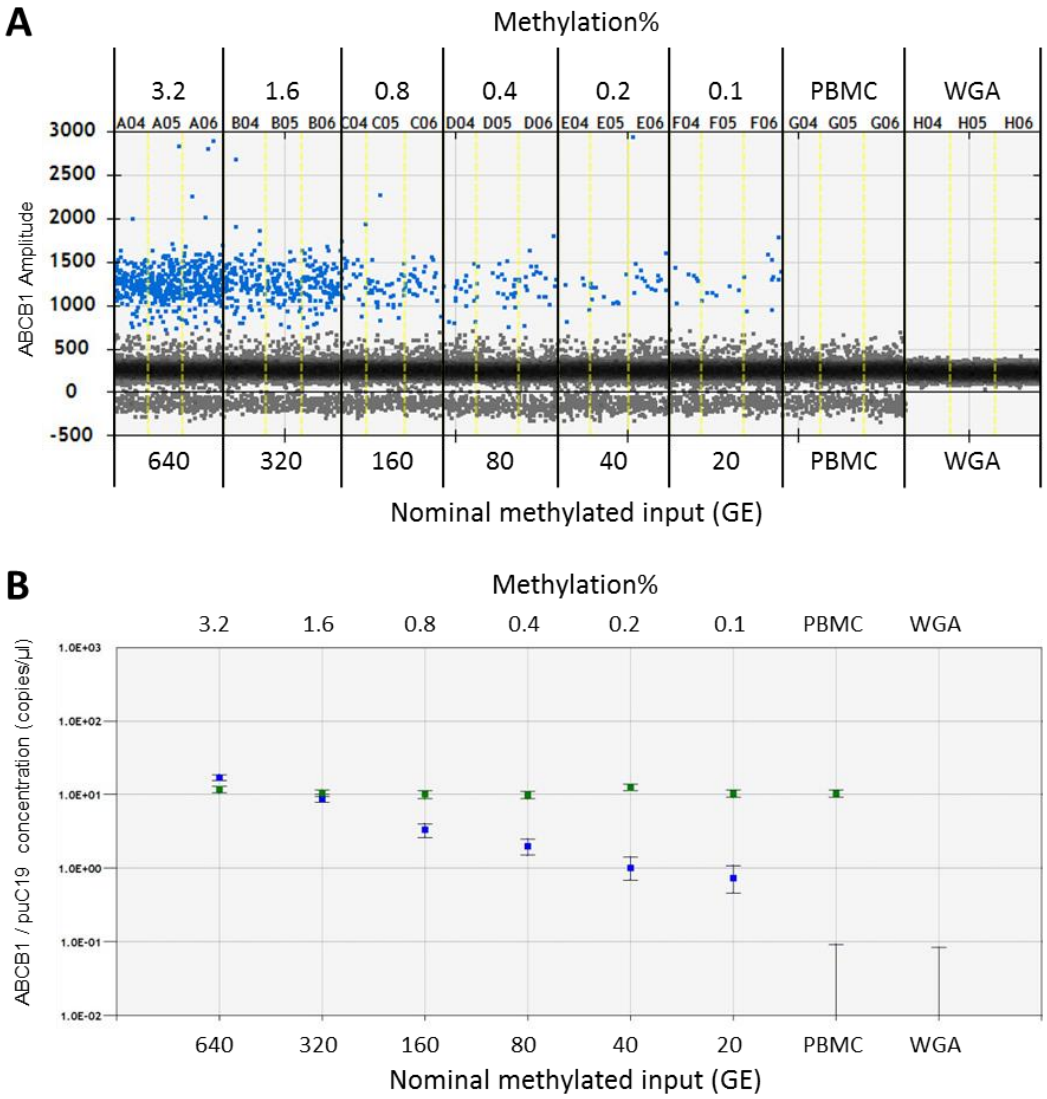


Figure 4.3-9 *ABCB1*-pUC19 duplex assay performance evaluation using twofold methylated DNA dilution series
A. Modified QuantaSoft ddPCR 1D plot indicating positive/negative droplet classification for *ABCB1*. *ABCB1* positive (FAM) positive droplets are indicated in blue, and negative droplets appear as dark grey. Each segment between solid vertical black lines contains data points from > 27,500 individual ddPCR reactions.
B. QuantaSoft concentration output. Blue points correspond to *ABCB1* concentration, green points correspond to pUC19 concentration. Light grey error bars represent overall 95% confidence intervals between replicate ddPCR wells and darker error bars represent 95% Poisson confidence intervals. Concentration values are derived from the merging of duplicate wells as a single meta-well containing > 27,500 droplets.

Methylated DNA was consistently detected at the 20 GE input level in both of the evaluated assays. This is comparable to the absolute limits of detection for methylated DNA diluted in water (*EVL*: 19 GE, *NTRK3*: 38 GE) observed in one of the earliest

published methylation-specific ddPCR peer-reviewed papers (Yu et al., 2015). Both assays also displayed excellent linearity and were specific, with no amplification detected in methylation negative controls.

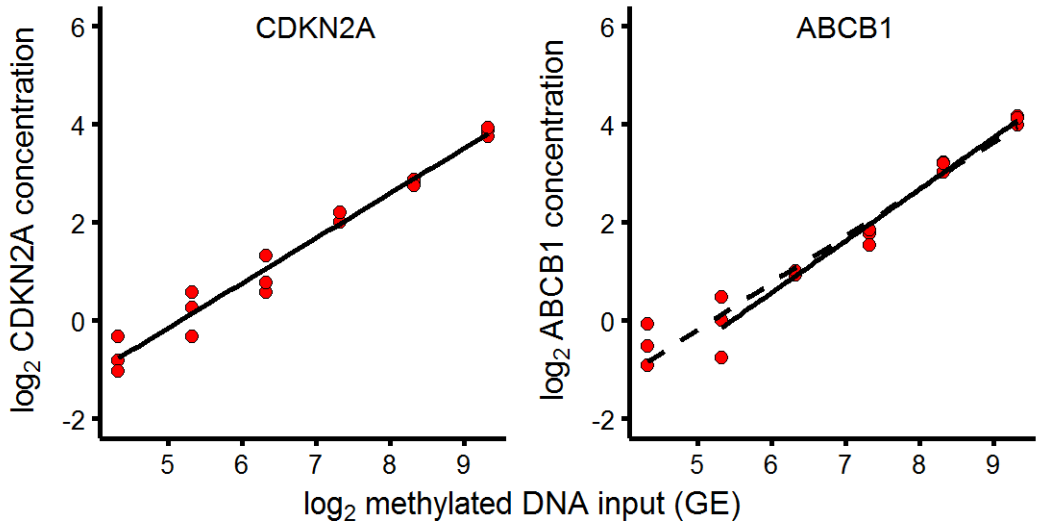


Figure 4.3-10 Evaluation of methylated DNA detection and quantification – linear regression plots
Scatterplot of $\log_{10}$-transformed **A.** *CDKN2A* and **B.** *ABCB1* ddPCR measured concentration against $\log_{10}$-transformed nominal methylated DNA input with linear regression lines superimposed. Dotted and solid regression lines in the *ABCB1* plot encompass the full and limited (40 – 640) range of inputs respectively.

Table 4.3-6 Duplex ddPCR assay regression parameters for twofold methylated DNA: PBMC DNA dilution series

Parameter	Methylated DNA target gene	
	CDKN2A	***ABCB1****
β_0 (95% CI)	-4.718	-5.75
	(-5.288, -4.146)	(-6.623, -4.878)
β_1 (95% CI)	0.9149	1.054
	(0.833, 0.996)	(0.937, 1.171)
R^2	0.973	0.967
RMSE	0.262	0.276

**ABCB1* regression analysis range limited to 40 – 640 GE methylated DNA input

4.3.2.3 *CDKN2A*-pUC19 ddPCR assay technical validation

The performance of methylation-specific ddPCR was further evaluated in the detection of *CDKN2A* in three independent twofold methylated DNA dilution series. *In vitro* methylated PBMC DNA (**2.4**) was diluted in untreated PBMC DNA of the same concentration. The methylated DNA dilution series was extended, when compared to the standards used in **4.3.2.2**, ranging from 5 GE to 640 GE. PBMC DNA and WGA DNA were again used as biological and technical controls respectively. ddPCR reactions were carried out in triplicate and contained 66 ng (20,000 GE) total DNA.

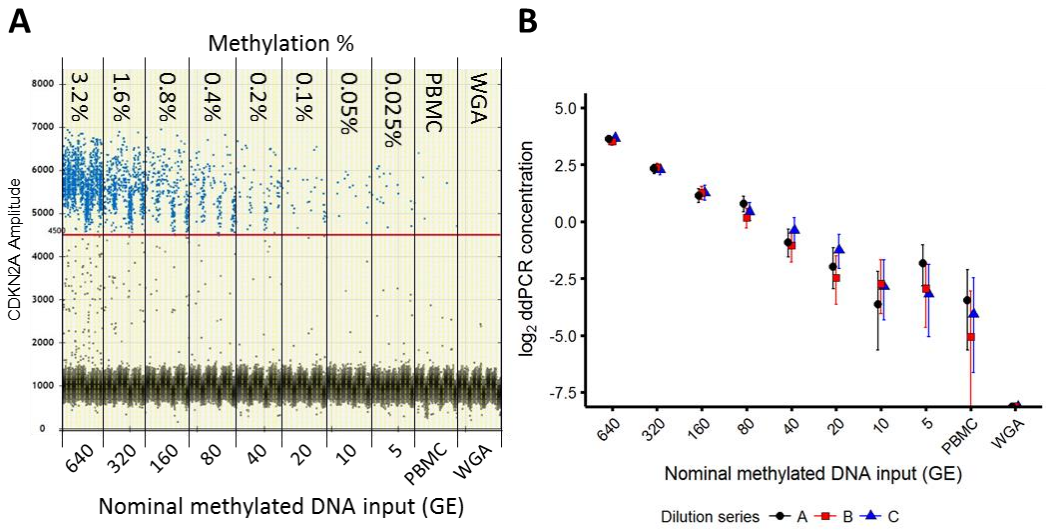


Figure 4.3-11 *CDKN2A*-pUC19 duplex assay performance evaluation using three independent twofold methylated DNA dilution series
A. Modified QuantaSoft ddPCR 1D plot indicating positive/negative droplet classification for *CDKN2A*. *CDKN2A* positive (FAM) positive droplets are indicated in blue, and negative droplets appear as dark grey. Each segment between solid vertical black lines contains data points from > 122,000 individual ddPCR reactions.
B. Scatter plot of three independent methylated DNA dilution series. Coloured data points are derived from the merging of triplicate wells as a single meta-well containing > 40,000 ddPCR partitions associated error bars represent 95% confidence intervals.

Non-template controls and WGA technical unmethylated controls produced no amplification. Very low concentrations of methylated DNA were detected in six out of nine PBMC control wells (0 to 0.09 copies/µl). Since no methylation positive

droplets were detected in any NTC or WGA control wells, this is unlikely to be non-specific amplification but as a result of very low levels of methylated DNA in the pooled PBMC DNA stock.

Methylated DNA was detected in all replicate wells for all three dilution series down to and including a nominal methylated DNA input of 20 GE (0.1%). Qualitative detection began to drop out at an input of 10 GE with methylated DNA detected in only a proportion of replicate wells (**Table 4.3-7**). However, methylation-specific amplification positive droplets were detected in at least one well per dilution series at the lowest tested input of 5 GE (**Table 4.3-7**).

Table 4.3-7 Proportion of methylation positive ddPCR wells at low input quantities

Nominal methylated DNA input (GE)	Methylation %	Dilution series	Number of positive wells/ total wells
20	0.1%	A	3/3
		B	3/3
		C	3/3
10	0.05%	A	1/3
		B	3/3
		C	1/3
5	0.025%	A	3/3
		B	2/3
		C	1/3
PBMC	PBMC	A	3/3
		B	1/3
		C	1/3
WGA	WGA	A	0/3
		B	0/3
		C	0/3

Intermittent detection of methylated DNA at low copy number is not surprising. Quoted methylated DNA inputs are nominal and are calculated based upon genomic DNA quantification prior to bisulphite treatment and assume 100% conversion and recovery. In fact, degradation of DNA during bisulphite treatment and clean-up is considerable and can extend to a reduction in intact, full-length DNA of 90% (Grunau et al., 2001). Recovery of bisulphite converted DNA using different commercially available kits can range from 29% to 92% (Holmes et al., 2014; Leontiou et al., 2015). At low target inputs, sampling variation is also an issue. I will illustrate this using a theoretical nominal ddPCR input of 8 GE as an example. This originates from bisulphite treatment of 80 GE DNA, eluted in a volume of 50 µl, 5 µl of which is added per ddPCR well, equivalent to splitting the eluate into ten aliquots. If we now, not unreasonably, assume a processing loss of 75% of material, we have a true average expected ddPCR input of 2 GE per well. However, due to random sampling effects, there is variation in the copy number between wells and they will not all contain exactly two *CDKN2A* copies, some wells receiving more, some fewer and some none, with probabilities dictated by the Poisson distribution. In this example, the probability of obtaining exactly two *CDKN2A* copies is 27%, three copies is also 27% and five copies is 18%. Furthermore, the probability that a well receives no *CDKN2A* copies is 14% and the probability a ddPCR well actually contains the target is only 86%. The probability of target positivity for a random well from a sample containing an average of one copy per well falls to 63% and has a corresponding 37% probability of target absence. The probability distributions of wells contains a particular number of target molecules are shown in **Figure 4.3-12**.

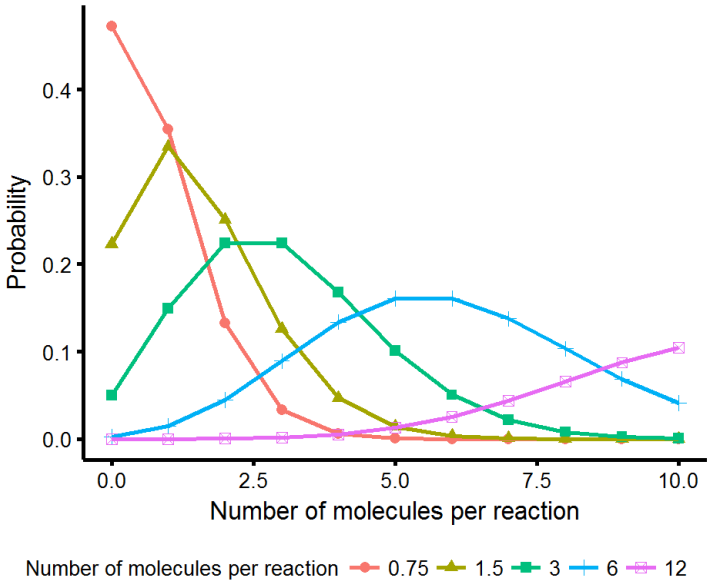


Figure 4.3-12 Sampling variation due to Poisson distribution of low target inputs
The probability that ddPCR wells contain a certain number of target molecules, when the average sample concentration is 0.75, 1.5, 3, 6 and 12 methylated target molecules per well.

4.3.2.3.1 Linear regression modelling of *CDKN2A*-pUC19 validation data

Data limited to the 20 to 640 GE input range was centred to the average log2-transformed nominal methylated DNA input so that intercept parameters would be more interpretable, applying at the geometric mean of the methylated DNA inputs (113.1 GE) rather than at an input not included in the model (i.e. 1 GE input without centring). Linear regression modelling of centred data including an explanatory variable controlling for different dilution series, with contrasts set to compare to the grand mean, indicated that the most appropriate model included an interaction, reflecting differences in slopes between dilution series. This model explained almost all of the variance in log2-transformed measured *CDKN2A* concentration [adjusted R^2 = 0.987, $F(5, 12) = 266.6$, $P < 0.001$]. After back-transformation to the measurement scale, ddPCR measured *CDKN2A* concentration increased by a factor of 2.108 (95% CI [2.015, 2.204]) when nominal methylated DNA input was doubled.

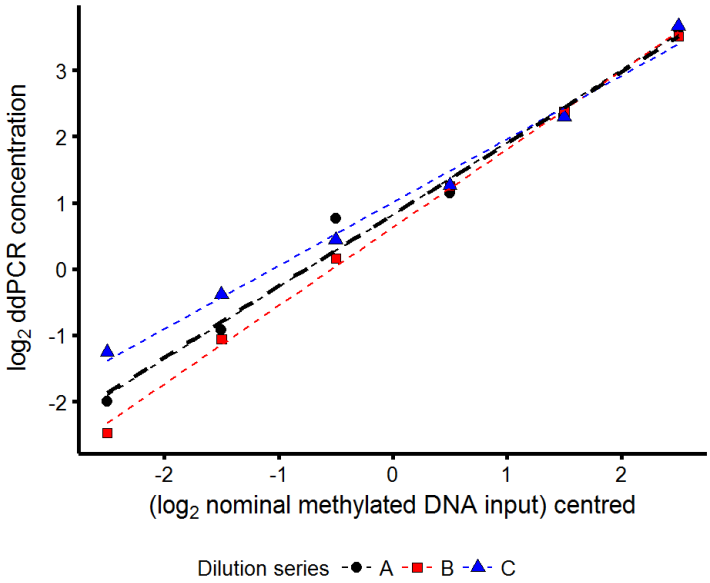


Figure 4.3-13 *CDKN2A*-pUC19 duplex assay scatterplot and linear regression plot
Scatter plot of three independent methylated DNA dilution series (20 – 640 GE nominal methylated DNA input) superimposed linear regression lines. Plotted points are derived from the merging of triplicate wells as a single meta-well containing > 40,000 ddPCR partitions. The grand mean regression for all three dilution series is shown in black and superimposes almost directly over that for dilution series A.

While the interaction between methylated DNA input and dilution series B is significant at the 95% confidence level, it accounts for a 2.27-fold increase in measured concentration for a doubling in input when compared to the average of the three dilution series. This equates to an absolute increase above the average of less than 8%, which is acceptable. Similarly, statistically significant shifts in the intercept attributable to different dilution series are of small effect size (<1%).

The *CDKN2A*-pUC19 assay produced excellent analytical performance with relatively high methylated DNA input.

4.3.3 *ACTB* cfDNA quantification assay optimisation and validation

ACTB TaqMan® PCR quantification assay was optimised on the basis of annealing temperature and primer and probe concentrations. Final optimal conditions included 61°C annealing temperature and 500 nM forward and reverse primers and 200nM probe. PCR efficiency calculated from triplicate measurement of a standard curve (0.25 ng to 32 ng per qPCR well) of PBMC DNA (concentration determined by fluorescent quantification) was 97.0% (R^2 = 0.999, *P* <0.001). The pooled intra-assay standard deviation was 0.067 C_q (minimum SD: 0.02 C_q, maximum SD: 0.11 C_q) indicating high precision of measurement.

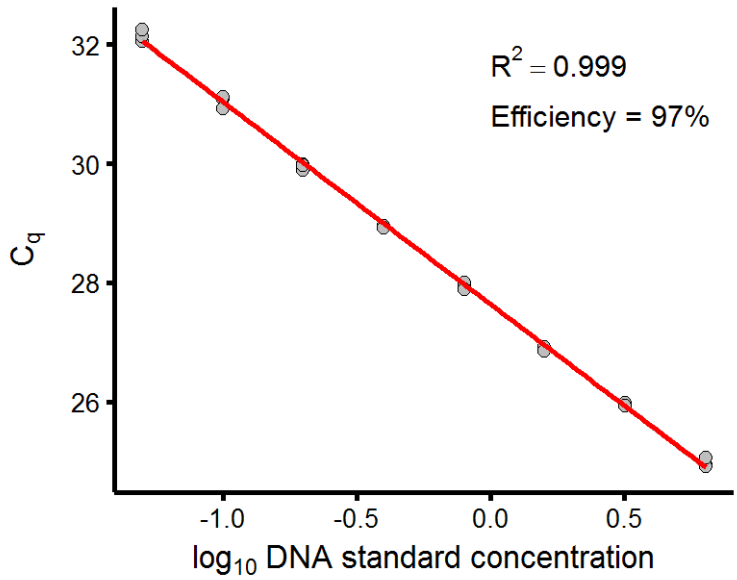


Figure 4.3-14 *ACTB* cfDNA qPCR quantification assay standard curve

It was important that this assay exhibited efficiency close to 100%, high precision and good linearity in order to quantify amplifiable cfDNA with confidence.

Utility of the quantification assay was tested on cfDNA extracted from 1 ml plasma samples from 134 cancer-free control subjects. cfDNA was successfully quantified in all samples and the distribution of yields is shown in **Figure 4.3-15**.

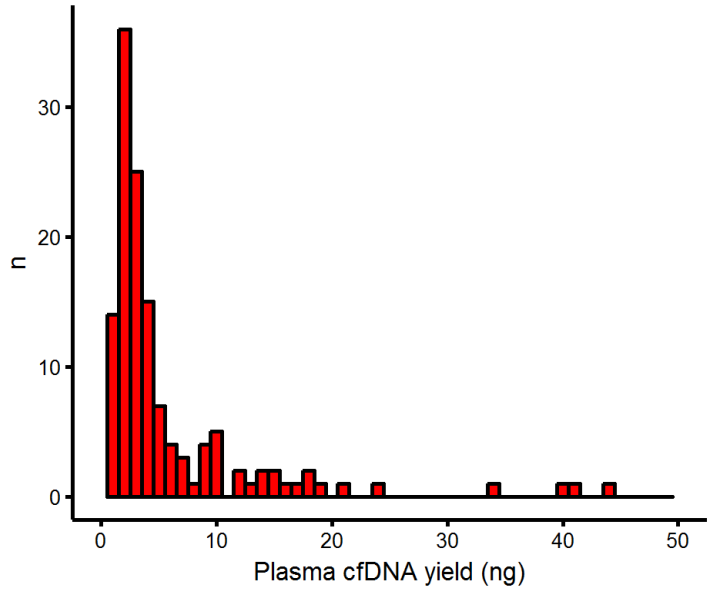


Figure 4.3-15 Plasma cfDNA yield
Histogram showing frequency of DNA yield based on concentration measured using *ACTB* cfDNA quantification assay (n = 134).

Yields were lower than anticipated (median: 3.0 ng, Interquartile range: 2.0 ng - 5.6 ng) and 94/134 samples presented yields lower than 5 ng. Clinical sensitivity is improved through the use of panels consisting of multiple markers (Laird, 2003) and low cfDNA yields could prohibit the use of multiple assays. Suppose a minimum of 10 methylated DNA copies per well are required for reliable detection. DNA methylation present in the plasma at 2% of total cfDNA, assayed for eight markers in duplicate wells, would call for in excess of 25 ng cfDNA extracted per patient sample. In this set of samples, only four specimens fulfilled this criterion. It is also important to note that this calculation does not take into consideration any DNA losses occurring in sample handling and treatment, or the effect of random sampling (see **4.3.2.3**). In order to successfully screen patient samples, it is therefore necessary to extend the utility of an individual specimen. This precipitated the development of methylation-specific multiplex PCR pre-amplification of bisulphite-converted DNA, or MethPlex enrichment.

4.3.4 MethPlex enrichment ddPCR optimisation and validation

Global pre-amplification of bisulphite-treated DNA using unselective techniques such as whole genome amplification utilising multiple displacement amplification, can lead to deviations from original bisulphite-modified DNA (Bundo et al., 2012). Multiplex PCR using primer pools is the most popular method of targeted pre-amplification (Ståhlberg and Kubista, 2014). Specific pre-amplification, solely enriching a panel of targets of interest (Livak et al., 2013; Tang, 2006), MethPlex enrichment, was therefore developed to increase the amount of target DNA available for ddPCR-based detection.

4.3.4.1 MethPlex enrichment ddPCR optimisation and validation

Although the same methylation-specific hydrolysis probe assays used in the analysis of moderate input (66ng / 20, 000 nominal GE) bisulphite-treated DNA were used to detect methylated DNA sequences after pre-amplification, it was considered necessary to determine optimal ddPCR annealing temperatures for MethPlex enrichment products. Methylation-specific qMSP/ddPCR uses bisulphite-treated DNA as template for amplification and cytosine to uracil conversion produces non-complementary, partially single-stranded DNA (Clark et al., 2006). Bisulphite-treated DNA is also highly degraded (Grunau et al., 2001), in particular fragmentation via depurination (Raizis et al., 1995) and chain breakage (Suzuki et al., 1994). This results in a reduction in the functional amount of amplifiable target DNA available (Leontiou et al., 2015). Taq DNA polymerase also has lower amplification efficiency when presented with templates containing uracil (Millar et al., 2015). MethPlex enrichment entails PCR amplification of bisulphite-treated DNA resulting in double-stranded DNA products that are templates for downstream methylation-specific ddPCR, albeit

with reduced sequence complexity compared to genomic DNA. It is feasible, therefore, that the reaction kinetics of methylation-specific ddPCR using MethPlex enrichment products and bisulphite-treated DNA will be different and reasonable to expect different optimal ddPCR annealing temperatures.

Duplicate ddPCR annealing temperature gradient experiments for all methylated DNA ddPCR assays were carried out using purified MethPlex enrichment products (T_a = 58°C) at a one in eight dilution. Methylation positive MethPlex reactions contained a nominal input of 50 methylated DNA copies in a background of 50 unmethylated DNA copies (330pg 50% methylated PBMC DNA) and methylation negative MethPlex reactions 100 unmethylated DNA copies (330pg untreated PBMC DNA). Optimal annealing temperatures were selected on the basis of criteria previously described (see **4.3.2**). Representative figures for the *F2R*-pUC19 assay are shown in **Figure 4.3-16** and clearly indicate no amplification in methylation negative wells across the temperature range. Separation of positive and negative of droplets was evident at 62°C, increasing with decreasing temperature to a maximum separation at 59°C, and *F2R* concentration measurements were consistent at temperatures with similar positive-negative cluster separation. Methylated pUC19 positive/negative cluster separation was consistent across the temperature gradient range and ddPCR estimated concentration was consistent between methylation positive and negative wells. Initial optimal MethPlex ddPCR annealing temperatures are presented in **Table 4.3-8**.

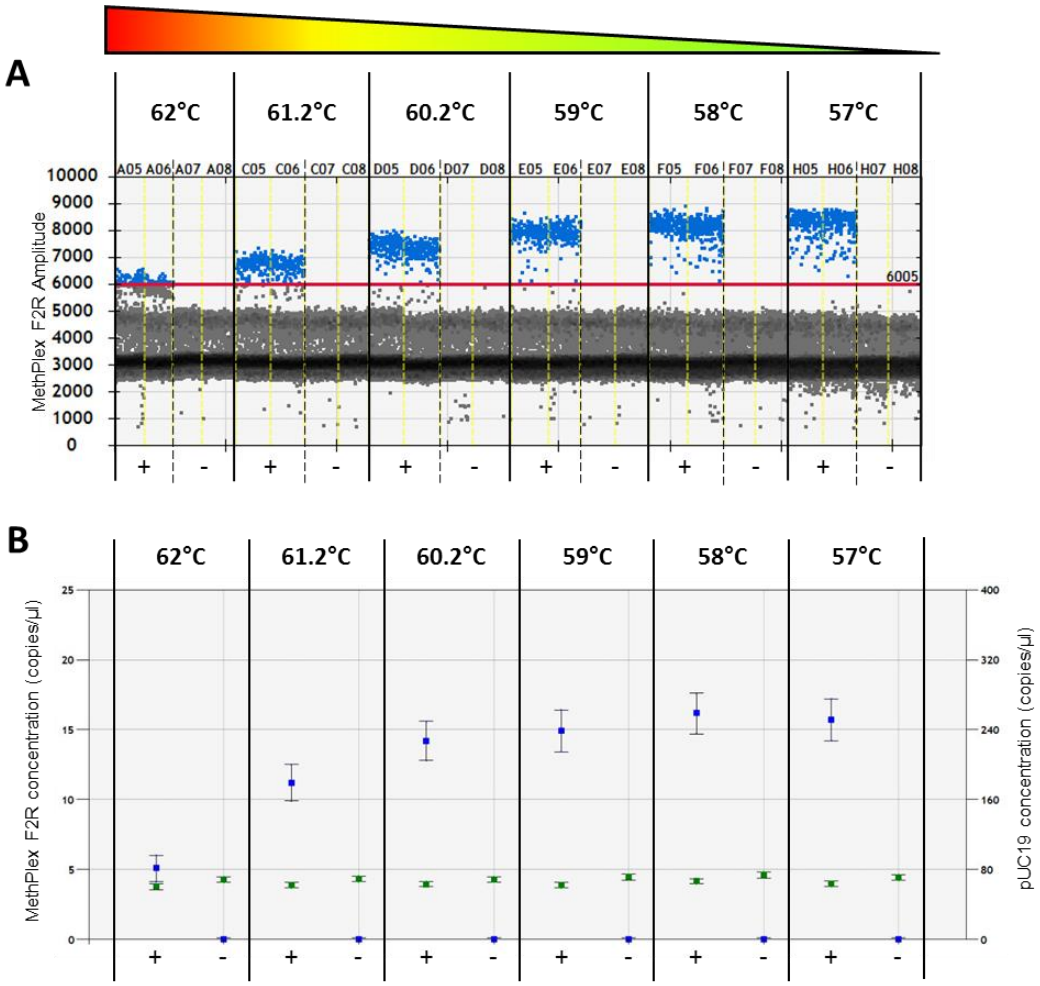


Figure 4.3-16 *F2R* MethPlex ddPCR annealing temperature optimisation
A. ddPCR fluorescence amplitude and positive/negative droplet identification for *F2R* positive (FAM) positive droplets are indicated in blue and negative droplets appear as dark grey. Solid red horizontal line indicates fluorescent amplitude threshold level. Each segment between solid vertical black lines contains data points for > 26,600 ddPCR reactions/metawell.
B. QuantaSoft concentration output. Blue points correspond to *F2R* concentration, green points correspond to pUC19 concentration. Error bars represent 95% Poisson confidence intervals. Concentration values are derived from the merging of duplicate wells as a single meta-well containing > 26,600 droplets.

Table 4.3-8 Initial optimal MethPlex ddPCR annealing temperatures

Target gene	Initial optimal MethPlex ddPCR annealing temperature (°C)	
	Initial	ddPCR concentration (copies/ul)
ABCB1	61.2	2.3
CDKN2A	59	109.6
F2R	59	14.9
MT1G	58	10.3
RASSF1	59	2.9
SHOX2	59	240
TERT	59	17.2
WT1	59	15.6

4.3.4.2 MethPlex primer concentration optimisation

Methylation-specific PCR enrichment reactions exhibited clear amplification bias, with methylated *RASSF1* and *ABCB1* detected by methylation-specific ddPCR at concentrations of only 19% and 15% of the median for all methylated targets (**Table 4.3-8**). Furthermore, *CDKN2A* and *SHOX2* were over-represented sevenfold and sixteen fold, respectively, compared to the median for all methylated targets (**Table 4.3-8**). To rectify this, final *CDKN2A* and *SHOX2* MethPlex primer concentrations were reduced to 100 nM. Target enrichment and ddPCR detection carried out as before. *RASSF1* pre-amplified methylated DNA ddPCR measurement increased to 15.9 copies/µl (95% CI [13.8, 18]) and *ABCB1* increased to 45.7 copies/µl (95% CI [42.6, 48.8]), while *CDKN2A* and *SHOX2* reduced to 46.8 copies/µl (95% CI [43.6, 50]) and 30.4 copies/µl (95% CI [27.4, 33.3]), respectively. Amplification bias is to be expected in a multiplex reaction of this nature (Polz and Cavanaugh, 1998) and it

is important to note that although all but one MethPlex primer were designed with a T_m = 56±0.6°, this is simply a software-based prediction and does not reflect the true complex nature of the reaction dynamics and kinetics. Observed improvements were deemed sufficient and future MethPlex enrichment reactions used 100 nM *CDKN2A* and *SHOX2* MethPlex primers and 200 nM for remaining target pre-amplification primers.

4.3.4.3 MethPlex enrichment annealing temperature optimisation

Duplicate MethPlex pre-amplification reactions containing the following nominal bisulphite DNA inputs were subject to annealing temperature gradients from 50°C to 60°C:

- 30 methylated DNA GE, 3000 untreated PBMC DNA GE
  (10ng 1% methylated PBMC DNA)
- 3030 untreated PBMC DNA GE
  (10ng untreated PBMC DNA)
- 303 methylated DNA GE, 30,000 untreated PBMC DNA GE
  (100ng 1% methylated PBMC DNA)
- 30,303 untreated PBMC DNA GE
  (100ng untreated PBMC DNA)

MethPlex enrichment of 10 ng inputs was performed to emulate expected yields from patient blood plasma samples, while pre-amplification of 100 ng inputs served to simulate preparation of ddPCR templates from high yielding specimens. Purified PCR products were diluted 1/8 and analysed in duplicate by methylation-specific ddPCR for each individual gene promoter target in the panel. Annealing temperatures for ddPCR reactions were as listed in **Table 4.3-8**.

An optimal MethPlex annealing temperature of 56°C was selected based upon a combination of magnitude and similarity of measured concentration for MethPlex pre-amplification replicates from methylation positive inputs and no/low signal measured in 10 ng PBMC DNA methylation negative controls. Low level amplification was present in the majority of wells following pre-amplification of 100ng PBMC DNA. A higher degree of DNA methylation was detected for the *MT1G* assay (mean at 56°C: 30.75 copies/µl). This is not surprising since Pyrosequencing Methylation Analysis of pooled PBMC DNA carried out in the validation of *in vitro* methylated DNA preparation indicated some background *MT1G* methylation (see **4.2.2**), a fact that may require careful consideration when drawing conclusions regarding *MT1G* DNA methylation.

4.3.4.4 Verification of MethPlex ddPCR optimal annealing temperature

Optimal ddPCR annealing temperatures using purified, diluted PCR products resulting from MethPlex enrichment (T_a = 56°C) were verified by further ddPCR temperature gradient experiments (eight annealing temperatures, 57°C - 62°C). MethPlex pre-amplification inputs were as follows:

- 30 methylated DNA GE, 3000 untreated PBMC DNA GE
  (10ng 1% methylated PBMC DNA)
- 30,303 untreated PBMC DNA GE
  (100ng untreated PBMC DNA)

10 ng inputs into MethPlex reactions were used to replicate expected yields from patient blood plasma samples, while 100 ng inputs represented high yielding samples, in order to assess potential background signal induced by high DNA input.

All optimal annealing temperatures were verified as the same as presented in **Table 4.3-8**, exhibiting good separation of positive and negative droplets. High PBMC DNA inputs resulted in amplification detection by ddPCR and may potentially result from pre-amplification of previously undetected methylated DNA copies in pooled blood bank PBMC stocks. For all genes, methylated samples could be distinguished from "unmethylated", despite the presence of signal in PBMC samples.

4.3.4.5 *CDKN2A*-pUC19 MethPlex ddPCR assay technical validation

The three independent twofold methylated DNA dilution series used in the technical validation of the *CDKN2A*-pUC19 assay with moderately high ddPCR reaction inputs were again used to evaluate the performance of the same ddPCR assay, facilitated by MethPlex enrichment with low methylated DNA input. MethPlex pre-amplification was carried out in triplicate with a total DNA input of 10 ng (~3030 GE). MethPlex enrichment products were diluted one in eight and analysed by methylation-specific ddPCR.

The results of this analysis are depicted as a scatter plot in **Figure 4.3-17**. No amplification was observed in either technical or biological methylation negative DNA controls, indicating good analytical specificity. *CDKN2A* methylation was reliably detected across all ddPCR and MethPlex replicates at 12 GE nominal methylated DNA input and above. Below this input quantity drop-out was observed at the MethPlex level with all three ddPCR replicates from one of three MethPlex enrichment reactions exhibiting detectable *CDKN2A* methylation at 1.5 GE and 3 GE nominal methylated DNA inputs. Pre-amplification of low inputs introduced

substantial variation in ddPCR analysis, making quantification unreliable due to a high degree of uncertainty.

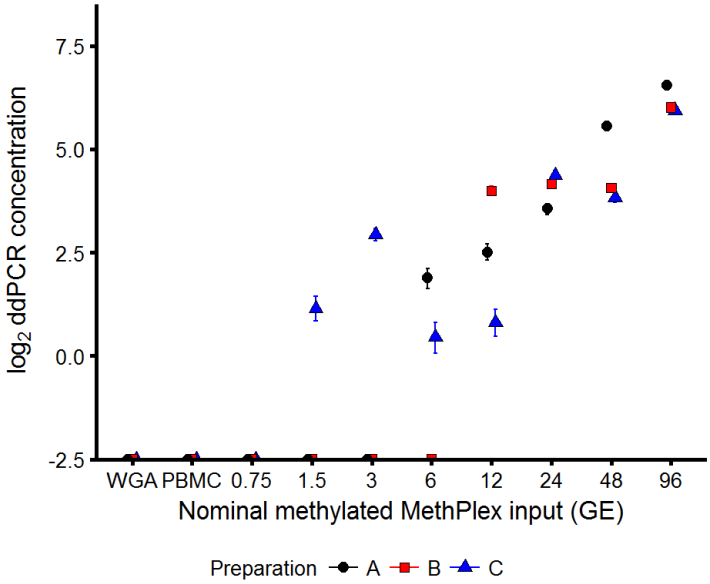


Figure 4.3-17 *CDKN2A*-pUC19 MethPlex ddPCR assay performance evaluation using three independent twofold methylated DNA dilution series
Scatter plot of three independent methylated DNA dilution series. Coloured data points are derived from the merging of triplicate wells as a single meta-well containing > 44,000 ddPCR partitions associated error bars represent 95% confidence intervals.

Increasing the amount of MethPlex enrichment product analysed in methylation-specific ddPCR analysis could possibly remedy this issue. In fact, a higher quantity of purified pre-amplification product was used for method comparison in the following chapter.

4.4 General discussion

Digital PCR has the potential for use in the analysis of nucleic acid biomarkers in clinical management of patients (Day et al., 2013). Furthermore, digital PCR has demonstrated greater precision and sensitivity in the detection of nucleic acids at low copy number than qPCR methods (Hindson et al., 2013; Jones et al., 2014; Strain et al., 2013). However, the ability to implement robust and relevant experiments and to exploit the latent enhanced precision of digital PCR analysis requires appropriate experimental design and execution (Huggett et al., 2013; Whale et al., 2012). Poor quality assay design in combination with inadequate performance of reaction condition optimisation is likely to lead to lower technical precision and false positive and/or negative detection of intended target amplicons (Bustin and Huggett, 2017). Therefore the extensive empirical optimisation described in this chapter was necessary and vital to the success of this project overall.

The production of appropriate control material and the optimisation of total cfDNA quantification methodology, the MethPlex targeted pre-amplification process and methylation-specific ddPCR assays that comprise this thesis chapter were the most time-consuming and labour-intensive elements of this entire body of work. While perhaps somewhat unexciting and uninformative, these contributions formed the foundation for the remainder of this thesis by ensuring that robust process controls were available and yields of amplifiable cfDNA could be reliably determined. Importantly, thorough optimisation and validation of ddPCR assays should facilitate the production of valid, reproducible data when these assays are used in the analysis of clinical samples. This will hopefully ensure that minimal further assay development will be required in order to meet stringent regulatory requirements should these assays progress to validation in clinical trial.

Potential methylation biomarkers, such as *CDKN2A*, are more likely to be effective in the detection of lung cancer as part of multiple marker panels in preference to analysis as single gene indicators of disease (Warton and Samimi, 2015). Validation of *ACTB* cfDNA quantification assay using cfDNA extracted from blood plasma revealed yields that would preclude the use of multiple methylated DNA marker assays; a limiting quantity of methylated DNA copies would require distribution among multiple ddPCR reactions resulting in stochastic effects due to random sampling error. The development of the MethPlex enrichment method effectively extends the utility of samples that yield limited cfDNA through the targeted methylation-specific amplification of genomic regions of interest. This could also potentially increase overall analytical sensitivity of workflows that include downstream ddPCR detection.

At this stage in the project we had confidence that DNA methylation detection workflows including process controls, cfDNA quantification, MethPlex pre-amplification and ddPCR assays had been sufficiently optimised for use with clinical plasma samples. However, superior performance of MethPlex ddPCR over and above that of legacy qMSP assays remained yet to be demonstrated and comparison of these methods is described in the following chapter.

Chapter 5

Comparison of Droplet Digital™ PCR and real-time PCR for the detection of methylated DNA

5.1 Introduction

Having satisfactorily developed and optimised methylation-specific assays for PCR-based analysis of bisulphite-treated DNA, the experimental work in this chapter was performed to compare the detection and quantification of methylated DNA by ddPCR and qMSP analysis. This was carried out both directly, at moderately high input, and at low input, assisted by MethPlex enrichment of methylated targets of interest. The experimental design employed in this method comparison (**5.2.1**) facilitated a comprehensive analysis of the data obtained using both detection methods, including the analysis of variance components at different levels within the experimental workflow (Tichopad et al., 2009). Methods and procedures from ISO technical standards (ISO, 1994a, 1994b, 1994c, 2000) were used in the analysis of relevant data in order to incorporate standardised assessment techniques. These assays are intended for use in pre-clinical and, ultimately, clinical study and will require stringent validation. Evaluating assay performance in this manner, at this early stage, should go some way to ensuring data acquired in their use are sufficiently accurate to meet strict requirements in pre-clinical assessment.

The experimental design structure was hierarchical (or nested). In order to account for subsequent non-independence of data and appropriately analyse data produced, I have used linear mixed modelling (Bates et al., 2015; Gelman and Hill, 2007; Laird and Ware, 1982; Pinheiro and Bates, 2000; Zuur, 2009).

5.2 Materials and methods

5.2.1 Experimental design

Methylated DNA dilution series were used to assess methylation-specific ddPCR and qMSP performance, testing assays targeting the *RASSF1* and *WT1* gene promoters. High total DNA inputs (66ng/20,000 GE) were used as a technical representation of assay behaviour when carried out with high yielding clinical samples such as tissue samples or bronchoalveolar lavage specimens; low total DNA inputs (10ng/3030 GE) were used as a facsimile of blood plasma liquid biopsy and were pre-amplified by MethPlex target enrichment.

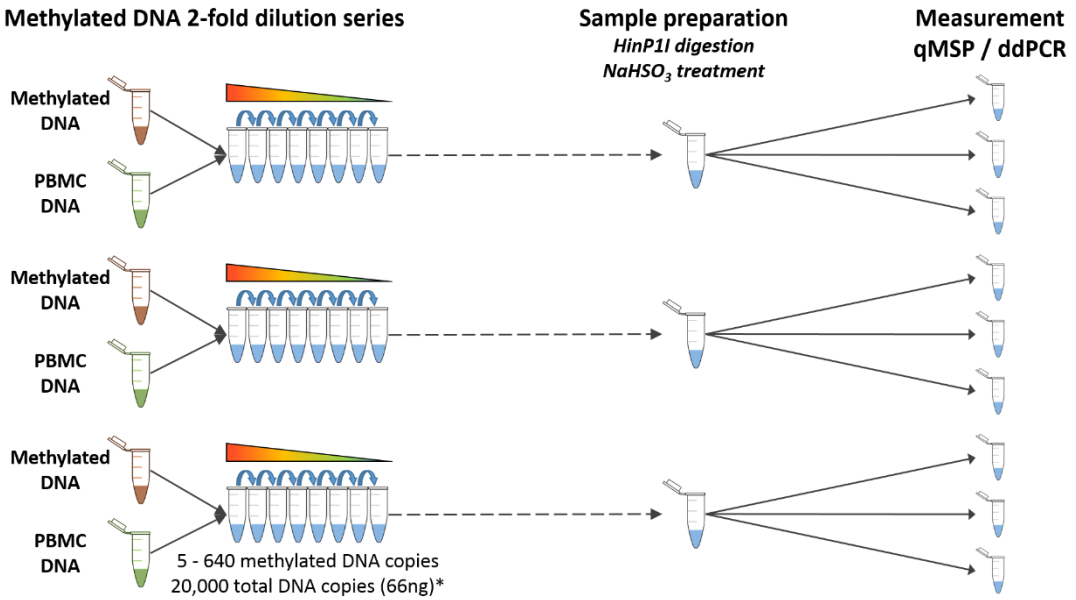


Figure 5.2-1 Hierarchical structure of high total DNA input experimental design
In vitro methylated DNA (50 ng/µl) was diluted to 3.2% with untreated PBMC DNA (50 ng/µl). Three aliquots of 3.2% methylated DNA were serially diluted to produce three independent two-fold dilution series also including PBMC (biological) and WGA (technical) controls. Each dilution series was independently HinP1I/bisulphite treated resulting in 30 "Prep" level samples in total. ddPCR/qMSP measurements were performed in triplicate resulting in 9 "Measure" level data points per methylated DNA input quantity per measurement method.
* Methylated and total DNA quantities refer to ddPCR/qMSP reaction inputs.

Three eight-point, two-fold *in vitro* methylated PBMC DNA dilution series (with unmethylated PBMC DNA as the background matrix/diluent) were constructed, ranging from 0.025% to 3.2% methylation, incorporating PBMC DNA matrix controls and WGA DNA technical methylation negative controls. These were then digested with methylation-sensitive restriction nuclease HinP1I and bisulphite treated. The three dilution series allowed us to assess the random error resulting from dilution, digestion and bisulphite treatment (designated "preparation" error) in addition to measurement/PCR error.

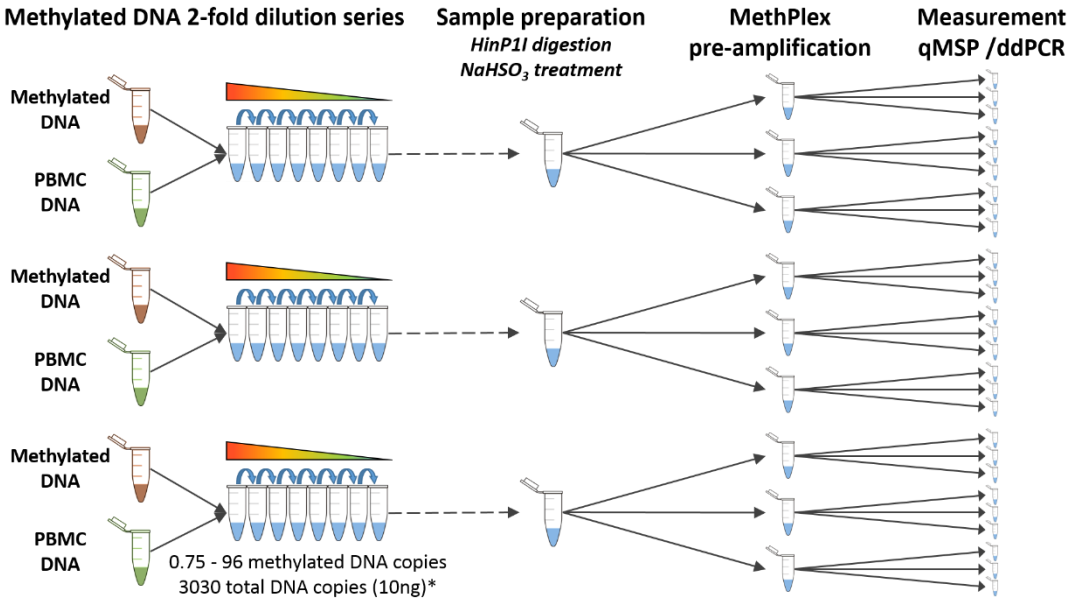


Figure 5.2-2 Hierarchical structure of low total DNA input experimental design including MethPlex enrichment
In vitro methylated DNA (50 ng/µl) was diluted to 3.2% with untreated PBMC DNA (50 ng/µl). Three aliquots of 3.2% methylated DNA were serially diluted to produce three independent two-fold dilution series also including PBMC (biological) and WGA (technical) controls. Each dilution series was independently HinP1I/bisulphite treated resulting in 30 "Prep" level samples in total. "Prep" level samples were pre-amplified by MethPlex enrichment in triplicate resulting in 90 "MethPlex" level samples in total. ddPCR/qMSP measurements were performed in triplicate resulting in 27 "Measure" level data points per methylated DNA input quantity per measurement method.
* Methylated and total DNA quantities refer to MethPlex enrichment reaction inputs.

High DNA input assessment was performed in triplicate with nominal methylated DNA inputs ranging from 5 to 640 GE (0.025 – 3.2% methylation). Assay

performance was evaluated at low DNA inputs by MethPlex pre-amplification of each dilution in triplicate, purification and dilution of PCR products and triplicate ddPCR and qMSP analysis. Nominal pre-amplification inputs ranged from 0.75 to 96 GE (0.025 – 3.2% methylation). Triplicate target enrichment enabled estimation of random error resulting from pre-amplification and clean-up (designated "MethPlex" error) in addition to "preparation" and measurement/PCR error.

5.2.2 Statistical methods

5.2.2.1 Outlier treatment

Data were screened for outliers through various mean, e.g. data points in linear regression with Studentized residuals with a value greater than three were considered outliers; Grubbs' test was used to identify outliers in precision and trueness analysis. When outliers were present, data collection and analyses were audited for potential procedural errors. Genuine potential outliers were judged to be representative of the study population and the uncertainty in measurement at low inputs and were therefore retained.

5.2.2.2 Significance of model parameters

The significance of fixed effect parameters in linear mixed models was assessed using the Kenward-Roger approximate F test for fixed effects in linear mixed models performed using the pbkrtest package in R (Halekoh and Højsgaard, 2014; Kenward and Roger, 1997). Random effects were evaluated by likelihood ratio test (LRT) with *P* value correction appropriate for a 50:50 mixture of chi-squared distributions with 0 and *x* degrees of freedom (Self and Liang, 1987; Stram and Lee, 1994).

5.2.2.3 Other statistical procedures

Other statistical tests and methods used in this chapter are noted in the text of the relevant results sections.

5.3 Results

5.3.1 Analysis of methylated *RASSF1* and *WT1* by ddPCR and qMSP at high assay input

Bisulphite-treated methylated DNA dilution series were analysed by ddPCR and real-time PCR to facilitate assessment of analytical characteristics. Total DNA input per PCR reaction was 20,000 GE (66ng) and methylated DNA inputs ranged from 5 to 640 GE (0.025 – 3.2% methylation). PBMC DNA and WGA DNA were included as matrix and technical methylation negative controls respectively. Methods were evaluated on the following criteria: sensitivity, precision, trueness and resolution.

Contributions of dilution series construction, HinP1I digestion and bisulphite treatment (combined as "preparation"), and PCR to the total variance across the measured range and for individual nominal methylated DNA input quantities was also assessed.

Scatter plots of measured methylated DNA abundance are presented in **Figure 5.3-1**.

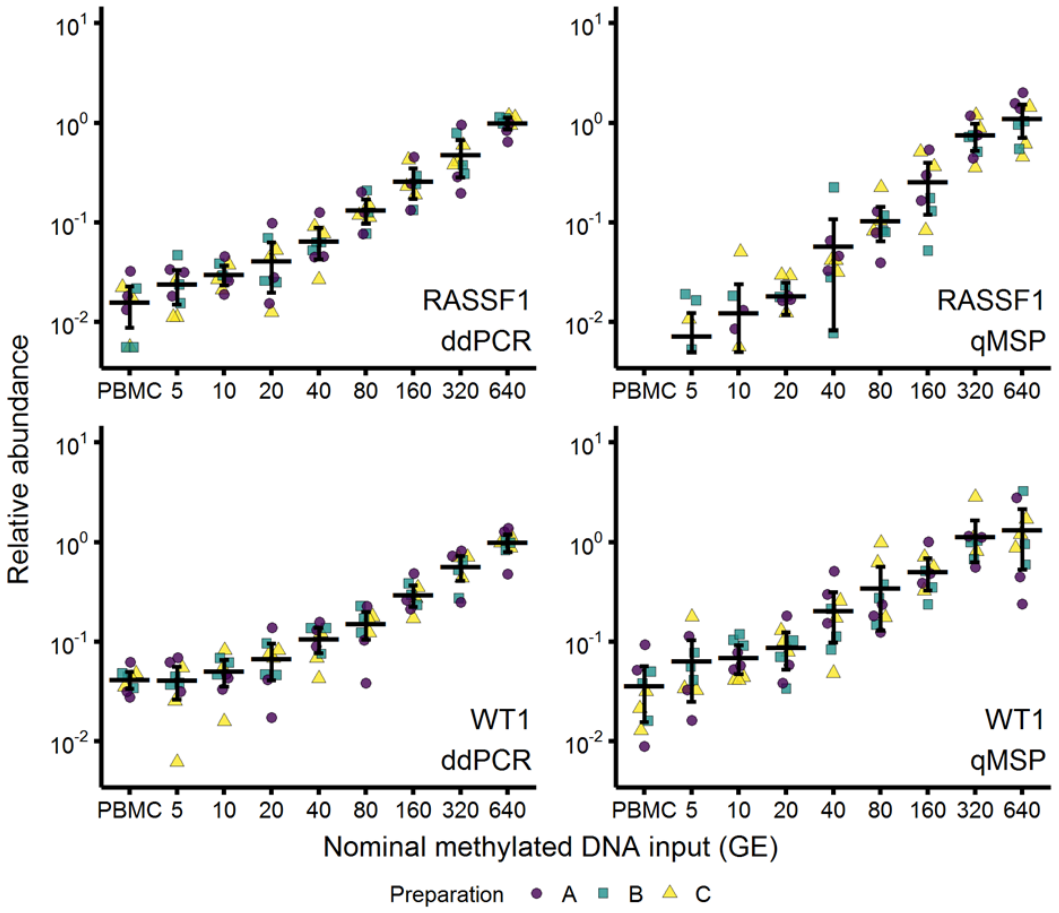


Figure 5.3-1 Comparative analysis of methylated DNA dilution series and PBMC DNA at high DNA input
Scatter plots of ddPCR/qMSP measured concentration against methylated DNA input amounts from analysis of three independent dilution series ('Preparation'). Data are presented with a $\log_2$-transformed y-axis for more readily interpretable visualisation. Each point represents an individual ddPCR/qMSP replicate, with nine PCRs for each point of the standard curve. Wide bars and error bars represent the grand mean of all replicates and 95% confidence intervals respectively. qMSP data represent relative abundance of target normalised to methylated pUC19 control and the mean average of all data for 640 GE nominal methylated DNA input using the Pfaffl method (Pfaffl, 2001). ddPCR data are displayed as relative concentration normalised to the mean average of target assay data for 640 GE nominal methylated DNA input to enable direct comparison with qMSP data.

5.3.1.1 Limit of detection

Both methods and both promoter assays returned no signal for water NTCs (n = 9) or WGA DNA (n = 9) indicating a zero theoretical bound to instrumental limit of detection. Using an informal definition of the limit of detection (LOD_{qual}) as the lowest non-zero assay input at which all replicate measurements render positive

qualitative results (Yu et al., 2015), both methods displayed LOD_{qual} of 5 nominal methylated DNA GE (0.025% methylation) for both assays. A more formal capability of detection assessment was performed using methods outlined in technical standard ISO 11843-2 (ISO, 2000) and summarised by Lavagnini and Magno (Lavagnini and Magno, 2007).

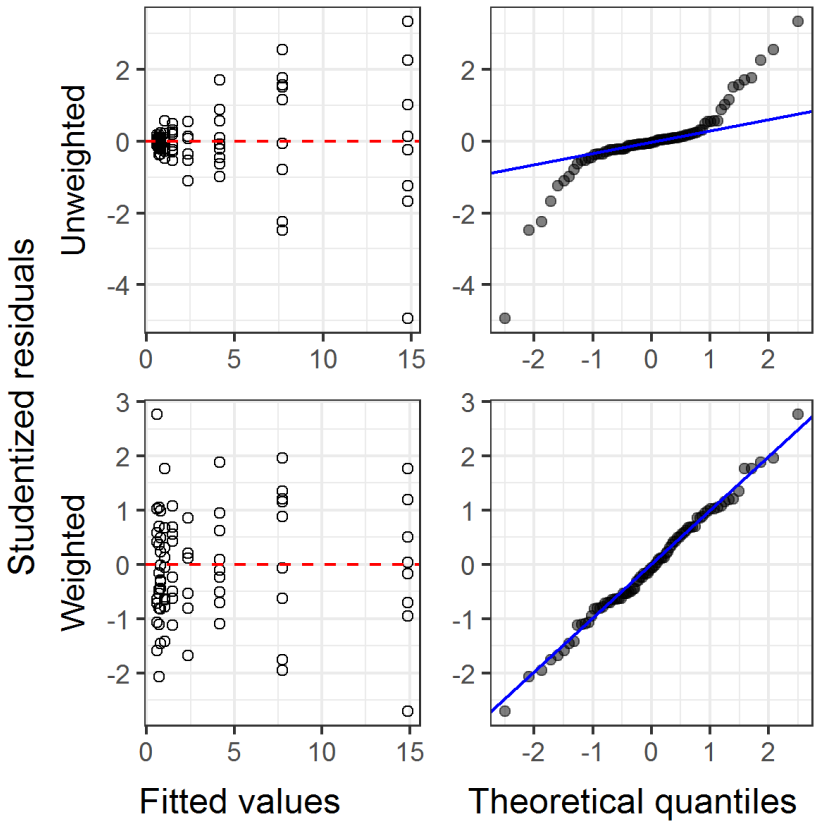


Figure 5.3-2 Capability of detection regression diagnostics
Representative diagnostic plots relate to *WT1* ddPCR data. Studentized residuals are plotted against model fitted values in the **left-hand panels** and are indicative of scedasticity (constancy of variance) and should not reveal any clear trends. The upper plot, with a clear "fanning-out" pattern is typical of heteroscedastic data. Distributions of residuals are represented by Q-Q plots in **right-hand panels** with points lying along the plotted line indicating a normal distribution. **Upper panels** show diagnostics for unweighted linear regression, while **lower panels** display those for weighted least squares regression.

This approach involves linear calibration of the analytical response of matrix blank, in this case untreated PBMC DNA, and methylated DNA inputs in the range expected in

clinical testing, and thus controls for the probability of committing both false positives and false negatives ($\alpha = \beta = 0.05$).

Inspection of residuals resulting from naïve linear regression reveals an increasing trend in variance, proportional to fitted values, indicating heteroscedasticity (**Figure 5.3-2 (upper left panel)** and **Appendix, Figure 8.3**). Q-Q plots also suggest that residuals are non-normally distributed and imply a "heavy-tailed" distribution (**Figure 5.3-2 (upper right panel)** and **Appendix, Figure 8.3**).

Table 5.3-1 Regression parameters for WLS calibration models
RSD = Residual standard deviation of the model

Parameter	*RASSF1*		*WT1*	
	ddPCR	**qMSP**	**ddPCR**	**qMSP**
β_0 (95% CI)	0.22	-0.0020	0.58	0.0406
	(0.16, 0.28)	(-0.0083, 0.0048)	(0.51, 0.65)	(0.0251, 0.0562)
β_1 (95% CI)	0.021	0.0018	0.022	0.0027
	(0.019, 0.023)	(0.0016, 0.0021)	(0.020, 0.025)	(0.0022, 0.0032)
R^2	0.85	0.79	0.83	0.59
RSD	0.11	0.011	0.12	0.027

Weighted least squares regression (WLS), with weights inversely proportional to nominal methylated DNA copies input, gives a more symmetrical distribution of the residuals around the zero values and thus corrects for the non-constant variance (**Figure 5.3-2 (bottom left panel)** and **Appendix, Figure 8.3**). Residual Q-Q plots provide strong evidence that weighting produces normally distributed residuals for both ddPCR assays (**Figure 5.3-2 (bottom right panel)** and **Appendix, Figure 8.3**). WLS provided only marginal improvements in qMSP residuals (**Appendix, Figure 8.3**). The addition of quadratic covariates to account for potential curvature in

regression line was unnecessary. **Table 5.3-1** shows the parameters of the four calibration models.

This more formal approach indicated higher limits of detection in the DNA input domain than LOD_{qual}, resulting from variability in measurement and from uncertainty in the calibration curve linear regression model (Lavagnini and Magno, 2007). The ddPCR limit of detection (x_D) was significantly lower than that for qMSP for both assays (*RASSF1*: $P = 0.04$, df = 42.1; *WT1*: $P = 1.7 \times 10^{-7}$, df = 32.1 (Welch's t-test, two tailed)), with ddPCR reducing the limit of detection by 20% and 47% for *RASSF1* and *WT1* respectively (**Table 5.3-2**).

Table 5.3-2 Capability of methylated DNA detection characteristics in the case of moderately high total DNA input (66 ng/20,000 GE)
Signal domain characteristics are representative of the response variable (units – ddPCR (copies/ul), qMSP (relative concentration).
L_C: critical value, signal domain; L_D: limit of quantification, signal domain; L_Q: limit of quantification, signal domain.
DNA input quantity domain characteristics give a read-out in trems of DNA input copies required for reliable detection and quantification.
x_C: critical value, DNA input quantity domain; x_D: limit of detection, DNA input quantity domain; x_Q: limit of quantification, DNA input quantity domain.
95% confidence intervals for assay parameters in the DNA input quantity domain are enclosed in parentheses.

		Signal domain			DNA input quantity domain (GE)		
Gene	Method	L_C	L_D	L_Q	x_C	x_D (LOD)	x_Q (LOQ)
RASSF1	ddPCR	0.40	0.58	1.28	8.27 (4.8, 11.74)	16.85 (13.39, 20.32)	49.61 (46.14, 53.08)
RASSF1	qMSP	0.02	0.04	0.11	10.28 (5.98, 14.59)	20.95 (16.65, 25.26)	61.68 (57.37, 65.99)
WT1	ddPCR	0.78	0.99	1.80	9.05 (5.25, 12.84)	18.44 (14.64, 22.23)	54.27 (50.48, 58.06)
WT1	qMSP	0.09	0.13	0.32	17.44 (9.75, 25.13)	34.50 (26.81, 42.19)	101.56 (93.87, 109.24)

Absence of statistical significance in differences between measurements for PBMC DNA and methylated DNA inputs below the calculated limits of detection was confirmed by Dunnett's multiple comparison test **Appendix, Table 8.1**. *RASSF1* ddPCR, however, exhibited a significant difference between 10 GE and PBMC input ($P = 0.005$), as did *WT1* qMSP at 20 GE input ($P = 0.028$).

Significant differences were observed between measurements for inputs above x_D and PBMC DNA (**Figure 5.3-1** and **Appendix, Table 8.1**). Differences between PBMC DNA and *RASSF1* qMSP 40 GE input, and PBMC DNA and *WT1* ddPCR 20 GE input, however, were not significant at the 95% significance level.

5.3.1.2 ddPCR has superior precision considering high total DNA inputs

Precision and trueness of the two methods were compared for data in the 20 – 640 GE input range, for both methylated DNA assays, using standard deviation and coefficient of variation as measures of precision, and bias as an indicator of trueness. This range was chosen as it encompasses all inputs above the highest calculated limit of detection. In order to evaluate precision between methods on a comparable relative scale, ddPCR concentration (copies/µl) was transformed by dividing the response variable by the mean average for 640 copies input. Precision analyses were performed using procedures described in technical standard ISO 5725 (parts 2, 3 and 6) (ISO, 1994a, 1994b, 1994c).

Inspection of scatter plots (**Figure 5.3-1**) indicates differences between digital and real-time methodologies, with ddPCR exhibiting tighter grouping. Repeatability standard deviations (s_r) at different methylated DNA input levels were consistently

lower for ddPCR measurements (**Appendix, Table 8.2**). *WT1* ddPCR at all methylated DNA quantities, and RASFF1 ddPCR at the maximum input, provided significantly better precision at the 95% confidence level than qMSP (**Appendix, Table 8.2**). Average variances (s_r^2) across input levels (root mean square average, using $\log_2$ transformed data to account for variances increasing in proportion to input DNA quantity) were also significantly lower for methylated DNA measurement by ddPCR (*RASSF1* $F_r = 0.472$, *WT1* $F_r = 0.464$, $F_{(0.025),\ 30,\ 30} = 0.482$).

Since s_r and intermediate standard deviations ($s_{i(prep)}$) were proportional to methylated DNA input, I also compared the coefficient of variation (CV%) between methods. **Figure 5.3-3** clearly shows that CV% values cluster above the line of identity, indicating consistently higher qMSP imprecision.

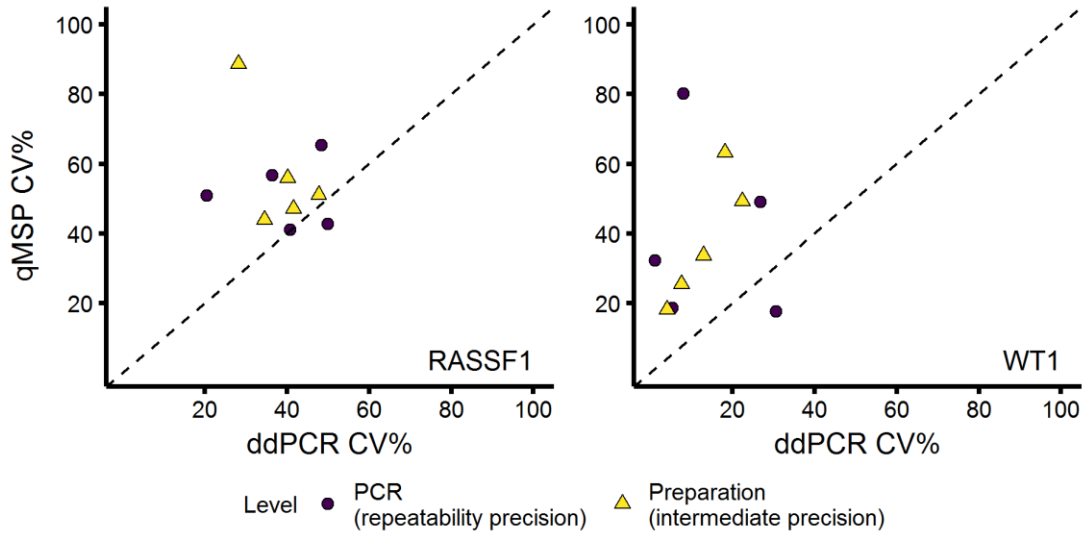


Figure 5.3-3 Coefficient of variation comparison plots for high total DNA inputs
Coefficients of variation were calculated for both methods for all dilution points above 20 GE and were plotted against each other. PCR CV% (within-preparation) was calculated using direct root mean square approach (Bland, 2006). Preparation CV% (between-preparation) was determined as the standard deviation of mean measurement values at the preparation level divided by the measurement grand mean for a given methylated DNA input.

ddPCR reduced *RASSF1* and *WT1* average PCR CV% by 34% and 35% respectively. Similarly preparation level CV% was reduced by 78% and 66% (**Table 5.3-3**).

Table 5.3-3 Coefficient of variation comparison table
Coefficients of variation, calculated at the PCR and preparation level, are presented for nominal methylated DNA inputs above 20 GE.
PCR CV% (within-preparation) was calculated using direct root mean square approach (Bland, 2006). Preparation CV% (between-preparation) was determined as the standard deviation of mean measurement values at the preparation level divided by the measurement grand mean for a given methylated DNA input. Reduction in CV = $(CV_{qMSP} - CV_{ddPCR})/(CV_{qMSP})$. Averages for method and promoter assay combinations are root mean square averages.

		PCR			Preparation		
Gene	**Nominal methylated DNA input (GE)**	**CV ddPCR (%)**	**CV qMSP (%)**	**Reduction in CV (%)**	**CV ddPCR (%)**	**CV qMSP (%)**	**Reduction in CV (%)**
RASSF1	640	20.04	50.14	60.04	3.71	17.22	78.44
	320	48.61	41.77	-16.38	28.61	16.34	-75.14
	160	38.09	56.06	32.05	25.55	49.23	48.10
	80	41.41	41.06	-0.86	2.01	34.30	94.13
	40	48.38	66.09	26.79	9.66	79.38	87.84
	Average	-	-	33.53	-	-	78.34
WT1	640	29.03	89.71	67.64	4.49	17.67	74.60
	320	40.25	46.92	14.21	12.48	33.41	62.66
	160	35.48	43.06	17.61	9.22	25.95	64.45
	80	49.36	49.84	0.96	17.31	62.93	72.49
	40	38.64	57.63	32.95	23.62	49.04	51.84
	Average	-	-	35.14	-	-	65.71

5.3.1.3 qMSP displays proportional bias across the measured range

Bias was assessed by linear regression between the observed relative methylated DNA abundance and the expected relative abundance, normalised to the maximum methylated DNA input. For an ideal assay, this approach should produce a straight line that is not significantly different to the line of equality, having a gradient equal to one and an intercept of zero (y = x). Deviation from the equality line indicates a bias in the method and statistical evaluation of regression parameters provides a formal test

of this. Disparity in intercepts and slopes demonstrate constant and proportional bias respectively. Data were modelled in log_2 space in order to correct for the lognormal distribution of response data and heteroscedasticity. Model residuals were assessed for normality and homogeneity of variance and were satisfactory (**Appendix, Figures 8.4 and 8.5**).

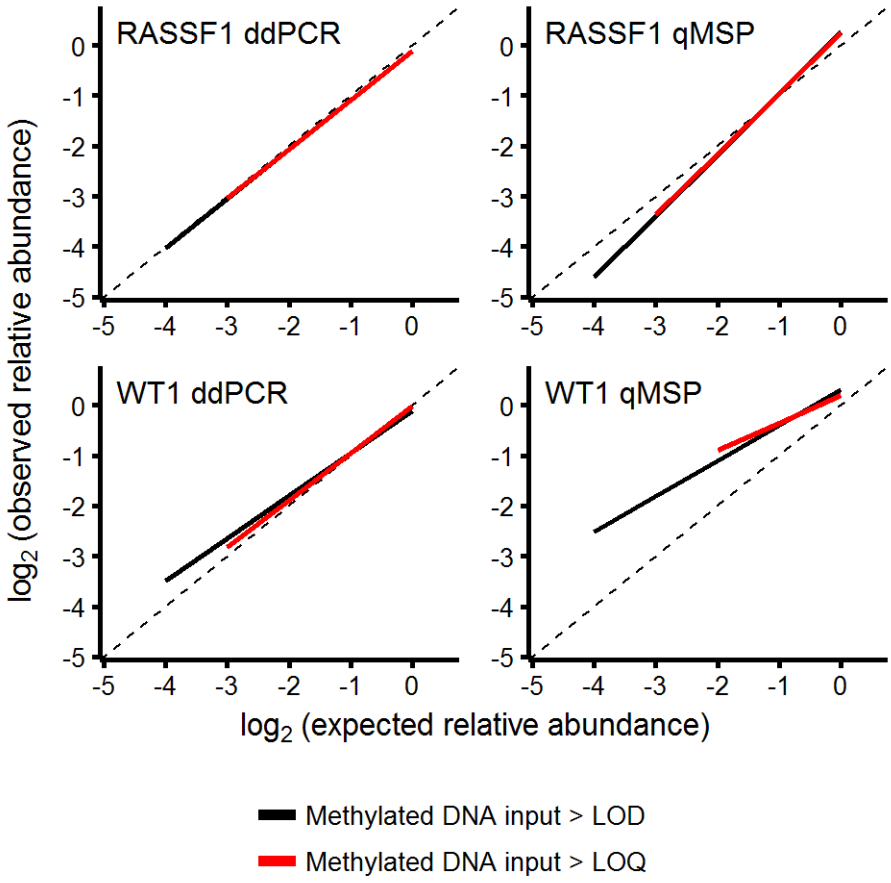


Figure 5.3-4 Linear regression plots of expected vs observed relative methylated DNA abundance to assess assay bias
Expected and obderved relative abundance data were log2-transformed, modelled by linear regression and plotted. This should yield a straight line not significantly different to the equality line. Slopes and intercepts statistically different to one and zero indicate the presence of linear proportional bias and constant bias respectively (**Table 5.3-4** and **Table 5.3-5**).

Models did not display constant bias for either measurement technique or promoter methylation assay in the 40 – 640 GE nominal methylated DNA input range (**Figure**

5.3-4 and **Table 5.3-4**), nor across ranges limited above respective limits of quantification (**Figure 5.3-4** and **Table 5.3-5**). Similarly, measurement of *RASSF1* DNA methylation by ddPCR did not reveal any proportional bias. Apparent proportional bias present in *RASSF1* qMSP and *WT1* ddPCR assays was ameliorated by limiting the lower end of the modelled range to the respective limits of quantification (**Figure 5.3-4** and **Table 5.3-4**). Similar adjustment of *WT1* ddPCR data did not provide any improvement **(Figure 5.3-4** and **Table 5.3-5**).

Table 5.3-4 Assay bias modelling parameters and statistical values for inputs above 20 GE nominal methylated input

		Constant bias (H_0: $\beta_0 = 0$)					Proportional bias (H_0: $\beta_1 = 1$)				
Gene	Method	β_0	95% CI	df	F	p-value	β_1	95% CI	df	t-value	p-value
RASSF1	ddPCR	-0.10	(-0.4, 0.19)	1	0.50	0.483	0.98	(0.86, 1.10)	43	-0.29	0.770
	qMSP	0.26	(-0.22, 0.75)	1	1.17	0.284	1.22	(1.02, 1.42)	43	2.21	0.032
WT1	ddPCR	-0.10	(-0.41, 0.21)	1	0.42	0.520	0.85	(0.72, 0.98)	43	-2.41	0.020
	qMSP	0.31	(-0.17, 0.80)	1	1.69	0.201	0.71	(0.51, 0.9)	43	-2.99	0.005

Table 5.3-5 Assay bias modelling parameters and statistical values for inputs above limit of quantification

		Constant bias (H_0: $\beta_0 = 0$)					Proportional bias (H_0: $\beta_1 = 1$)				
Gene	Method	β_0	95% CI	df	F	p-value	β_1	95% CI	df	t-value	p-value
RASSF1	ddPCR	-0.12	(-0.43, 0.20)	1	0.56	0.460	0.97	(0.8, 1.1)	34	-0.36	0.770
	qMSP	0.25	(-0.24, 0.73)	1	1.07	0.307	1.20	(0.94, 1.42)	34	1.59	0.115
WT1	ddPCR	-0.02	(-0.35, 0.31)	1	0.01	0.925	0.93	(0.76, 0.98)	34	-0.77	0.446
	qMSP	0.20	(-0.37, 0.77)	1	0.51	0.484	0.55	(0.11, 0.9)	25	-2.11	0.046

A more exhaustive analysis of relative bias, defined as:

$$\frac{(mean\ relative\ methylated\ DNA\ abundance - expected\ relative\ methylated\ DNA\ abundance)}{expected\ relative\ methylated\ DNA\ abundance}$$

at individual methylated DNA input quantities, indicated that for inputs greater than 20 GE, *RASSF1* ddPCR mean measured abundance closely matched expected relative abundance, resulting in relative biases ranging from -4.1% to 7.3%, while qMSP measurements exhibited larger systematic error, ranging from -16.3% to 51.7%. Observational error was higher for both *WT1* assays and indicated a clear trend, inversely proportional to methylated DNA input. *WT1* qMSP displayed considerable positive relative bias, ranging from 34.2% to 230.8%, vastly overestimating target relative abundance (**Appendix, Table 8.2**).

5.3.1.4 ddPCR can discriminate twofold differences in methylated DNA inputs

Linear mixed-effects regression of measured relative abundance on nominal methylated DNA input above 20 GE was used to assess the ability of assays to discriminate between different methylated DNA inputs. Post-hoc pairwise comparisons (Tukey's honest significant difference test, using the multcomp R package (Hothorn et al., 2008)) (**Figure 5.3-5**, **Table 5.3-6** and **Appendix, Table 8.3**) reveal that measurement of methylated *RASSF1* DNA by ddPCR distinguished all twofold differences across the tested range, while qMSP analysis showed limited significant differences, with fourfold differences evaluated as significant in the 80 – 640 GE nominal methylated DNA input range. *WT1* ddPCR differentiated two-fold differences above 40 GE nominal methylated DNA input whereas the equivalent qMSP assay data did not support any two-fold differences. *WT1* qMSP data also failed

to reject the hypothesis that 160 and 640 GE inputs (a fourfold difference) were equivalent at the 95% confidence level.

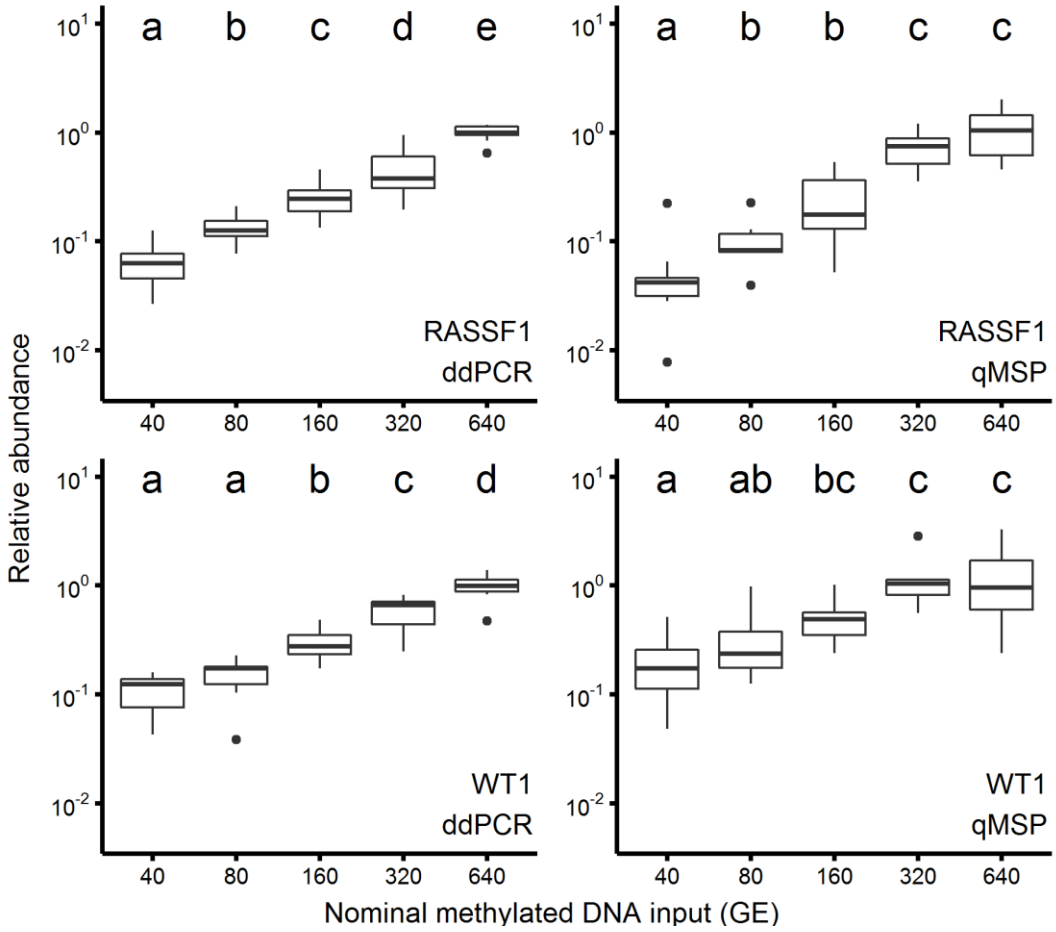


Figure 5.3-5 High DNA assay input boxplots in the 40 – 640 GE nominal methylated DNA input range
Box and whisker plots depict distributions of data for measurement of methylated DNA abundance by ddPCR and qMSP. Box limits represent first and third quartiles and the central line shows the median. Whiskers denote 1.5 x interquartile range and outliers beyond the limits of the whiskers appear as individual points. Inputs with the same letter at the top of the boxplots are not significantly different by Tukey's honest significant difference test (Hothorn et al., 2008) (**Table 5.3-6**). Quantification method and methylated gene promoter assay are indicated in the bottom right corner of individual boxplots.

Table 5.3-6 Table of statistics resulting from Tukey's honest significant difference test
This limited table presents statistical values limited to twofold comparisons. Additional comparisons can be found in the full post-hoc statistical table in **Appendix, Table 8.3**.

		ddPCR				qMSP			
Gene	**Comparison (Methylated DNA input) [GE]**	**Estimate**	**95% CI**	**Tukey's q statistic**	**Adjusted *P* value**	**Estimate**	**95% CI**	**Tukey's q statistic**	**Adjusted *P* value**
RASSF1	80 - 40	1.08	(0.33, 1.83)	3.91	0.001	1.24	(0.05, 2.43)	2.84	0.036
	160 - 80	0.93	(0.17, 1.68)	3.36	0.007	1.08	(-0.11, 2.27)	2.47	0.097
	320 - 160	0.83	(0.08, 1.58)	3.01	0.022	1.82	(0.63, 3.00)	4.17	2.8×10^{-4}
	640 - 320	1.21	(0.45, 1.96)	4.38	1.2×10^{-4}	0.51	(-0.68, 1.69)	1.16	0.774
WT1	80 - 40	0.46	(-0.32, 1.23)	1.61	0.493	0.71	(-0.49, 1.91)	1.62	0.486
	160 - 80	1.05	(0.28, 1.83)	3.71	0.002	0.77	(-0.43, 1.97)	1.74	0.407
	320 - 160	0.89	(0.12, 1.67)	3.13	0.015	1.14	(-0.06, 2.34)	2.58	0.073
	640 - 320	0.87	(0.09, 1.64)	3.06	0.019	-0.04	(-1.24, 1.16)	-0.10	1

5.3.1.5 Preparation error does not contribute significantly to total experimental variance

The linear mixed models characterized in **5.3.1.4** not only attempt to explain the variation in methylated DNA measurement as a function of the methylated DNA input, but also estimate the unexplained variation (i.e. the changes in y not explained by the methylated DNA input quantity) associated with "preparation". This is represented by an additional "preparation" variance term in the model, also known as a random effect. Recall that preparation includes variability due to dilution series construction, HinP1I digestion and bisulphite treatment. In all cases, the random effect for preparation, whether defined as an intercept, slope or multiple random intercept and random effects, was not significant (P = 0.75 to 1, Likelihood ratio tests). This is illustrated most clearly by the almost direct superimposition of the regression lines for

each preparation and the average, or marginal, regression line in linear mixed effects regression plots (**Figure 5.3-6**).

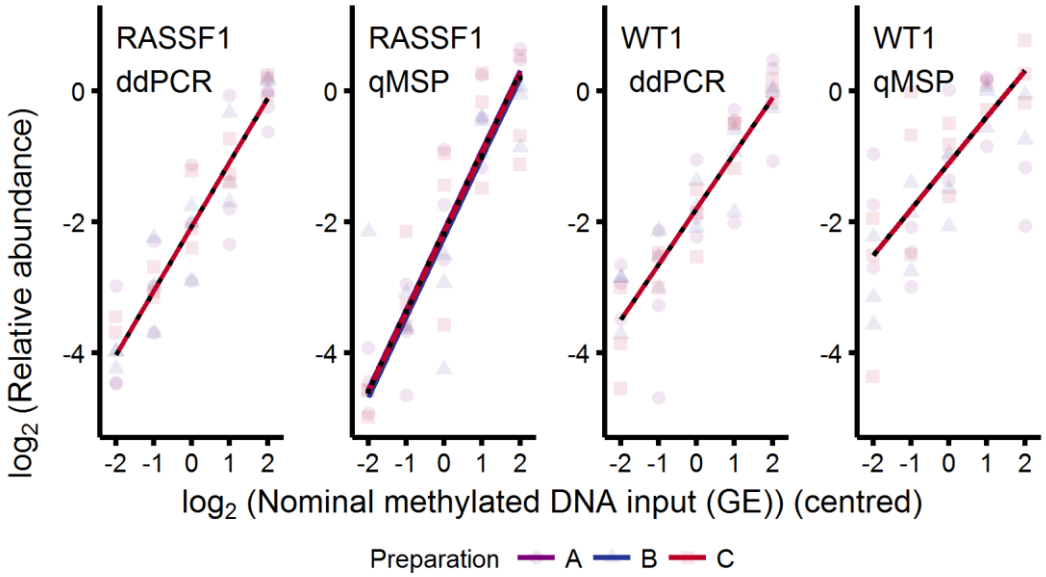


Figure 5.3-6 Linear mixed effects regression plots generated from high total DNA input methylation measurement data
Log_2-transformed relative abundance data were modelled in the 40 – 640 GE nominal methylated DNA input range, with log2-transformed methylated DNA input quantities centred on the 160 GE input for more informative interpretation of the intercept (x = 0). The three preparation group regression lines, indicated by different colours shown in the legend below the plot, do not vary from the marginal regression line (black dashed line).

Further evidence is provided by plots of the conditional means of the preparation random effect (i.e. the deviation from the average log_2 (relative abundance) for individual preparation group effects at x = 0). There was a very small variance contributed by preparation random effect in the *RASSF1* qMSP assay but this was not significantly different to zero (**Figure 5.3-7**).

Random intercepts were retained, however, in order to control for the non-independence and pseudoreplication stemming from multiple repeated measures for each preparation and nominal methylated DNA input combination. Furthermore,

assessment of linearity by quadratic regression did not indicate a deviation from linearity.

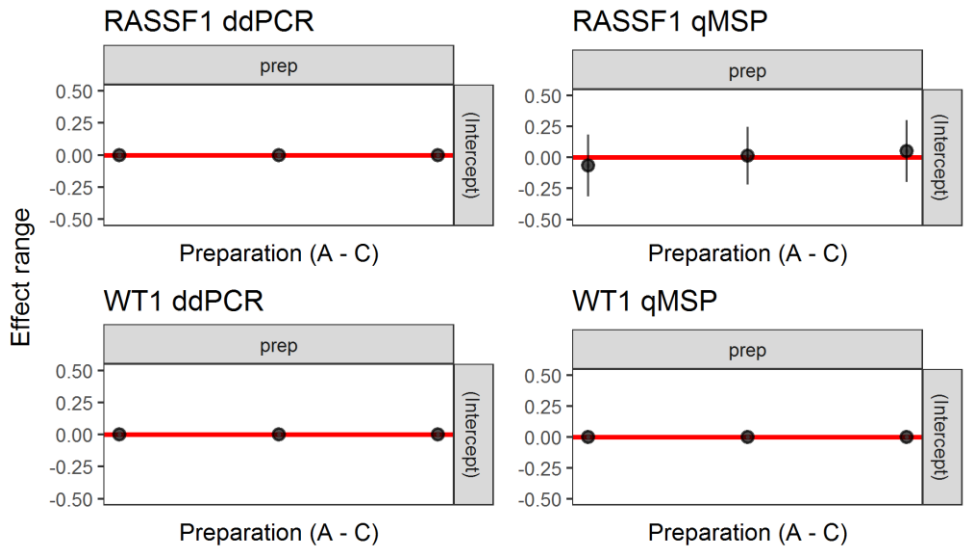


Figure 5.3-7 Linear mixed model conditional means of preparation group-level effects
Plots of conditional means of random effects give a view of the variation between preparation groups and the magnitude of these effects. 95% confidence intervals were constructed by simulation (n = 1000) and indicate a non-significant random effect when they cross zero (red horizontal line). In all cases there was no significant difference from zero.

We can also assess the contribution of preparation groups to the variation in quantification at the individual nominal methylated DNA input quantities. Results of variance component analysis presented graphically in the upper panel of **Figure 5.3-8** clearly indicate that there is little variation in the measured response that is attributable to the different preparation groupings, with the largest preparation contribution present for 640 GE *RASSF1* input for both detection techniques. Similarly, evaluation of the between-group variation in linear mixed models, controlling for the nominal methylated DNA input, indicated that there was essentially no preparation contribution to the experimental variance and fundamentally no measurable difference imparted upon similar DNA samples by HinP1I digestion and bisulphite treatment when measured using either technique (**Figure 5.3-8, lower panel**). Furthermore, when the contributions of the fixed effect (nominal methylated DNA input), random

effect (preparation) and residual (predominantly PCR measurement error) to the total modelled variance were analysed, it was again apparent that preparation effects were minimal, and that the methylated DNA input explained most of the variation in response. In fact, quantification by ddPCR explained substantially more of the experimental variance than qMSP (**Figure 5.3-8, right-hand panel**).

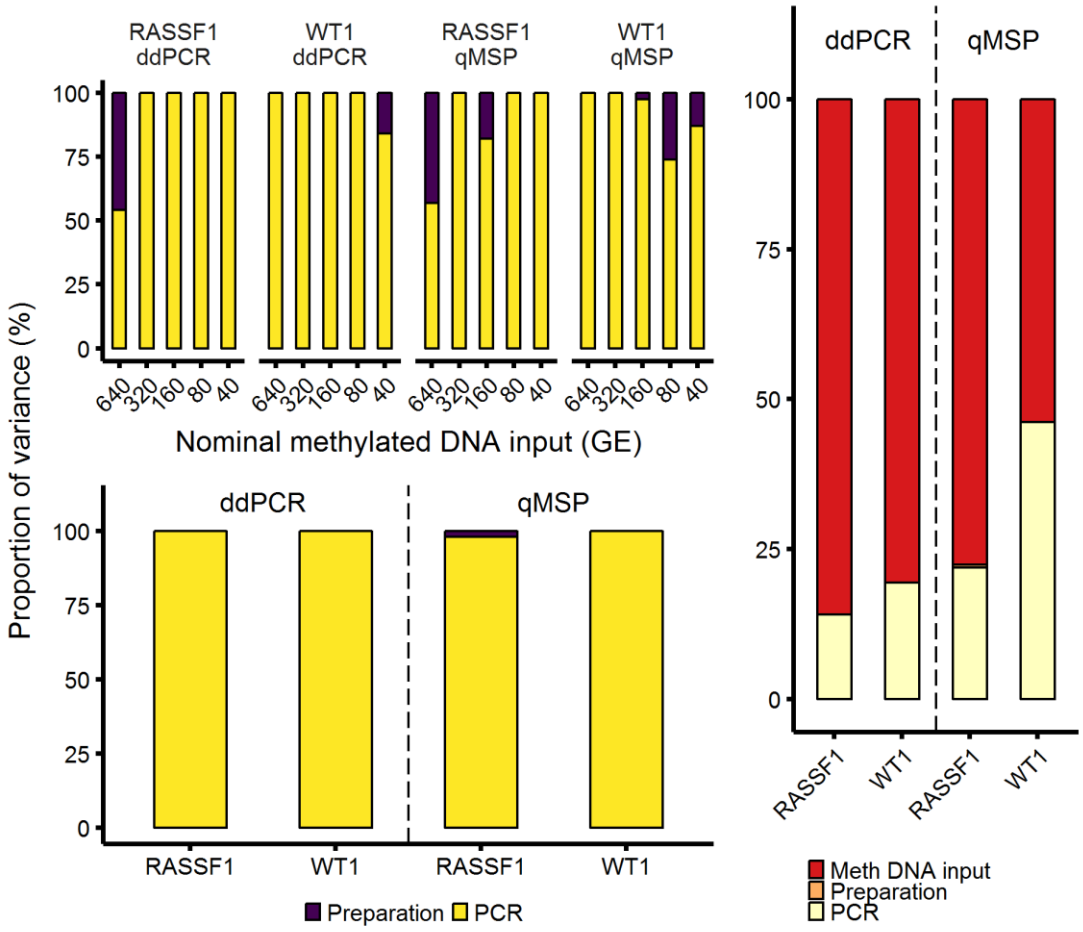


Figure 5.3-8 Variance component analysis of high total DNA input methylated DNA quantification data
The **upper panel** shows the contribution of the different levels of the experimental nested hierarchy to the total variance at the different nominal methylated DNA input quantities. Results were acquired using ANOVA-methodology implemented in the VCA R package (Schuetzenmeister, 2017). The four methylated DNA assay/ detection method combinations are displayed. The **lower panel** shows these same contributions modelled across the methylated DNA input range, through analysis by linear mixed effects regression and assessment of intra-class correlation coefficients, controlling for the variation due to the methylated DNA input, using the rptR package (Stoffel et al., 2017). The **right-hand panel** shows the contributions of the methylated DNA input (fixed effect), preparation (random effect) and PCR measurement error (residual variance) to the total experimental variance.

5.3.1.6 ddPCR models provide for more accurate and precise prediction of methylated DNA abundance

Further interrogation of model parameters reveals that residual variation (i.e. changes in y neither explained by methylated DNA input quantity nor by variability associated with "preparation") was greater than twofold higher for qMSP assays than for DNA methylation quantification by ddPCR. Marginal (fixed effects) and conditional (fixed effects and random effects) R^2_{GLMM} (Johnson, 2014; Nakagawa and Schielzeth, 2013) values also demonstrate that ddPCR models explain a greater proportion of the variance than qMSP measurement, with and without any preparation effects respectively ($R^2_{GLMM\ (conditional)}$: *RASSF1* ddPCR = 0.859, *WT1* ddPCR = 0.806, *RASSF1* qMSP = 0.781, *WT1* qMSP = 0.538). ddPCR slope parameters (β_1) are close to unity, indicating that a twofold increase in input results in a near twofold increase in measurement, and are also more precise, signified by narrower confidence intervals (**Table 5.3-7**).

Table 5.3-7 Linear mixed effects model parameters

Parameter	*RASSF1*		*WT1*	
	ddPCR	**qMSP**	**ddPCR**	**qMSP**
Fixed				
β_0 (95% CI)[1]	-2.07	-2.17	-1.80	-1.10
	(-2.23, -1.90)	(-2.53, -1.81)	(-1.97, -1.62)	(-1.41, -0.78)
β_1 (95% CI) [1]	0.98	1.22	0.85	0.71
	(0.86, 1.10)	(1.02, 1.41)	(0.73, 0.97)	(0.51, 0.90)
Random				
σ^2 (Preparation)	0.000	0.017	0.00	0.000
σ^2 (Residual)	0.323	0.856	0.355	0.874
Observations	45	45	45	45
R^2m / R^2c	0.859 / 0.859	0.776 / 0.781	0.805 / 0.805	0.538 /0.538

[1] (95% CI) = Bootstrap confidence intervals

5.3.1.6.1 Mixed model predictions

We can also predict the (marginal) measured relative abundance at specified input quantities for the average preparation using these models, providing predicted estimates for the possible population of preparations. By resampling the observed data with replacement many times (bootstrapping), we can also build a sampling distribution for a given prediction, and associated 95% confidence intervals.

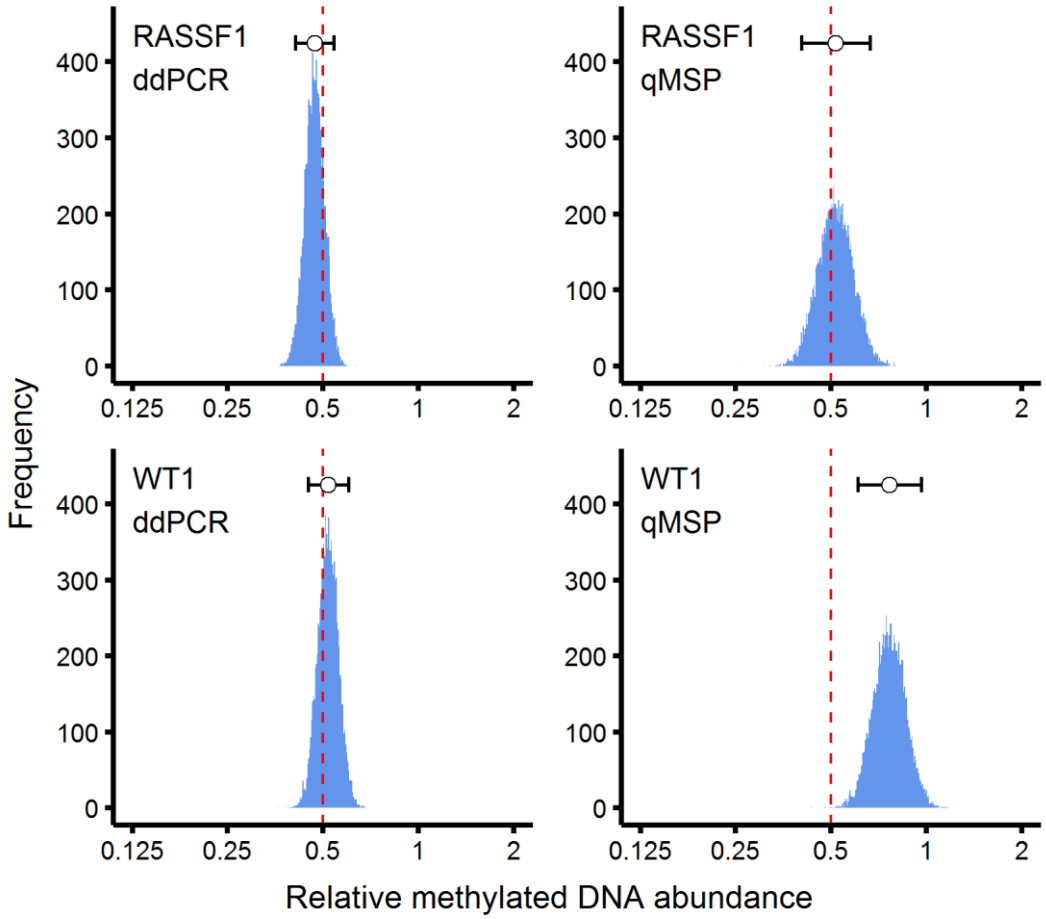


Figure 5.3-9 Histograms of bootstrap estimates for prediction at 320 GE nominal methylated DNA input
Sampling distributions from 10,000 bootstrap model simulations presented for each gene promoter methylation assay and measurement method. Model output has been back-transformed into the relative measurement scale. If models and measurement assays are accurate distributions should peak at 0.5 relative methylated DNA abundance, indicated by the red dashed line. Point estimate of predictions are shown as empty circles and error bars represent 95% confidence intervals.

Predictions at x = 1 for the centred models correspond to a nominal methylated DNA input of 320 GE or, alternatively stated, expected relative methylated DNA abundance of 0.5. Prediction estimates for ddPCR models at this input, back-transformed into the measurement scale, are similar to 0.5 (*RASSF1*: 0.47 (95% CI [0.41, 0.54]), *WT1*: 0.52 (95% CI [0.45, 0.60])). Across the measured range, estimates based upon qMSP models are, in contrast, more removed from expected values for given inputs, especially in the case of *WT1* which overestimates relative methylated DNA abundance by 50% at 320 GE input (*WT1*: 0.76 (95% CI [0.61, 0.96])). Moreover, the evident overestimation of the *WT1* qMSP assay increases to almost 100% at the 160 GE input level (*WT1*: 0.47 (95% CI [0.39, 0.56])) (**Appendix, Figure 8.6**). This is clearly demonstrated by histograms of predictions presented in **Figure 5.3-9** and reflects the greater accuracy of ddPCR, and offers further credence to bias results provided in **5.3.1.3**. Furthermore, ddPCR prediction distributions are decidedly narrower than those for qMSP assay models, providing for more restricted confidence intervals and attesting to superior ddPCR assay precision.

5.3.2 Analysis of methylated *RASSF1* and *WT1* by MethPlex ddPCR and qMSP at low assay input

Analytical characteristics of ddPCR and qMSP methylated DNA detection and quantification were also assessed through the measurement of methylated DNA dilution series following HinP1I digestion, bisulphite treatment and MethPlex enrichment. Total DNA input per MethPlex pre-amplification reactions was 3,030 GE (10ng) and methylated DNA inputs ranged from 0.75 to 96 GE (0.025 – 3.2% methylation). PBMC DNA and WGA DNA were again included as matrix and technical methylation negative controls respectively. Methods were evaluated same criteria as in **5.3.1**. Contributions of dilution series construction, HinP1I digestion and bisulphite treatment (combined as "preparation"), MethPlex enrichment and purification (combined as "MethPlex") and PCR to the total variance across the measured range and for individual nominal methylated DNA input quantities was also assessed. This experimental design resulted in a total of 270 observations for each promoter methylation assay/quantification method combination. Scatter plots of measured methylated DNA abundance are presented in **Figure 5.3-10**.

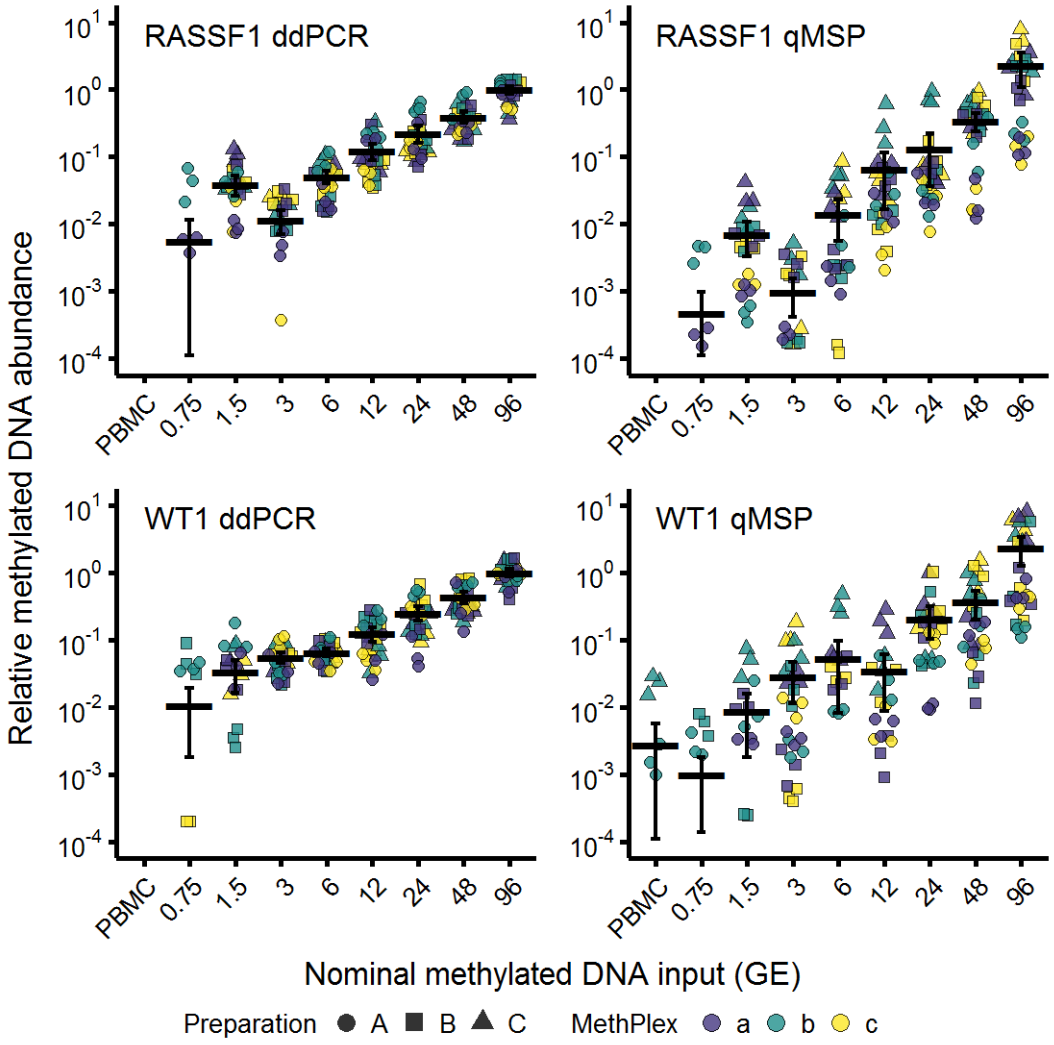


Figure 5.3-10 Comparative analysis of methylated DNA dilution series and PBMC DNA at low DNA input

Scatter plots of MethPlex ddPCR/qMSP measured concentration against methylated DNA input amounts from analysis of three independent dilution series ('Preparation'). Each dilution point was pre-amplified in triplicate, resulting in three MethPlex level samples for each preparation/dilution point. Each graphical point represents an individual ddPCR/qMSP replicate, with three PCRs for each MethPlex level sample nine PCRs for each preparation and 27 PCRs for each point of the standard curve. Data are presented with a $\log_2$-transformed y-axis for more readily interpretable visualisation. Horizontal bars and error bars represent the grand mean of all replicates and 95% confidence intervals respectively. qMSP data represent relative abundance of target normalised to methylated pUC19 control and the mean average of all data for 96 GE nominal methylated DNA input using the Pfaffl method (Pfaffl, 2001). ddPCR data are displayed as relative concentration normalised to the mean average of target assay data for 96 GE nominal methylated DNA input to enable direct comparison with qMSP data.

5.3.2.1 ddPCR methylation assays displayed lower limits of detection than their real-time PCR counterparts

5.3.2.1.1 Model selection

To evaluate limits of detection, MethPlex ddPCR assay data were modelled with mixed effects logistic regression using the lme4 package (Bates et al., 2015) in R (R Core Team, 2017), employing Laplace approximation. Prospective models were compared by likelihood ratio test with *P* value correction appropriate for a 50:50 mixture of chi-squared distributions with 0 and *x* degrees of freedom (Self and Liang, 1987; Stram and Lee, 1994). Model diagnostics were implemented using the DHARMa (Hartig, 2017) and arm (Gelman and Su, 2016) R packages and modelling assumptions were not violated in any of the selected models.

The most appropriate models for *RASSF1* gene promoter assays incorporated random slope terms for preparation and MethPlex:

```
glmer(pos ~ log2methcopies +
      (0 + log2methcopies|prep/methplex),
      Family = binomial(link = "logit")
```

in R programming terms.

Models selected for *WT1* assays included random preparation and MethPlex intercepts:

```
glmer(pos ~ log2methcopies +
      (1|prep/methplex),
      Family = binomial(link = "logit")
```

in R programming terms.

Table 5.3-8 LOD95 for MethPlex DNA detection and model parameter statistics

Gene	Method	LOD95 (95% CI)[1]	Fixed effects		Random effects	
			-2 log Λ^2	*P* value	-2 log Λ^2	*P* value
RASSF1	ddPCR	3.16 (1.88, 5.00)	11.3	3.8×10^{-4}	3.96	9.2×10^{-2}
	qMSP	3.32 (1.78, 5.70)	10.4	6.2×10^{-4}	16.59	1.5×10^{-4}
WT1	ddPCR	1.86 (1.27, 2.59)	202.1	3.7×10^{-46}	12.08	1.4×10^{-3}
	qMSP	13.44 (5.67, 30.25)	108.7	9.6×10^{-26}	26.4	1.1×10^{-6}

[1](95% CI) = Profile confidence intervals
[2]Likelihood Ratio test

In all four models random effects were significant (**Table 5.3-8**), as where MethPlex random effects alone (**Appendix, Table 8.4**). Preparation random effects were not significant (**Appendix, Table 8.4**) but were retained, however, because of the hierarchical experimental design and to control for dependence due to repeated measures and pseudoreplication. It should also be noted that fixed effect parameter estimates and standard errors were equivalent with and without preparation random effects, indicating that results are robust under differing assumptions.

5.3.2.1.2 Limits of detection (LOD95)

The limits of detection of methylated *RASSF1* promoter DNA by ddPCR and qMSP were similar ($P = 0.90$; Welch's t-test, two tailed), at just above 3 nominal methylated DNA GE input, with LOD95 qMSP being less precise, considering its wider confidence intervals (**Table 5.3-8** and **Figure 5.3-11**). qMSP also exhibited greater variance attributable to random effects (ddPCR $\sigma^2_{prep:methplex} = 0.58$ (95% CI [0, 1.80]), qMSP $\sigma^2_{prep:methplex} = 1.07$ (95% CI [2.46×10^{-13}, 3.04])). *WT1* MethPlex qMSP performed poorly compared to MethPlex ddPCR, exhibiting a much flatter probability

of detection curve (**Figure 5.3-11**) and significantly higher LOD95 (P = 9.2 x 10^{-5}; Welch's t-test, two tailed) of 13.4 GE nominal methylated input (**Table 5.3-8**).

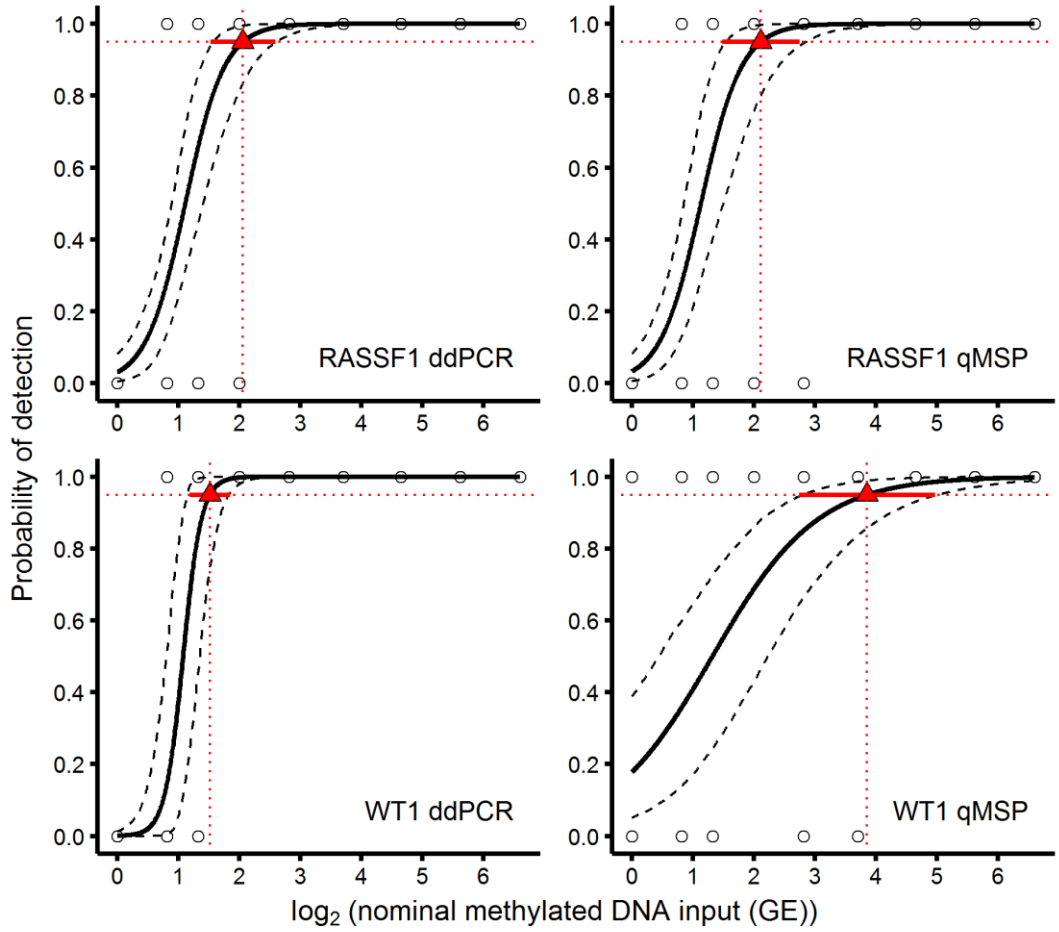


Figure 5.3-11 LOD95 for MethPlex enrichment methylation detection
Binary qualitative data from ddPCR/qMSP analysis of samples contributing on average 0 to 96 methylated DNA molecules to MethPlex enrichment reactions were modelled by mixed effects logistic regression. Predicted responses were converted to the probability scale and plotted. At 95% confidence (red horizontal dotted line), the LOD95 is the corresponding value on the x-axis (red vertical dotted line). Point estimates for LOD95 (red triangles) and corresponding confidence intervals were calculated using a custom version of the dose.p.glm function from the MASS R package (Venables and Ripley, 2002), modified to allow functionality with generalized linear mixed models. 95% confidence intervals of predicted plot lines were estimated by bootstrap resampling (n =1000).

Table 5.3-9 Mixed effects logistic regression final model parameters

Parameter	*RASSF1*		*WT1*	
	ddPCR	**qMSP**	**ddPCR**	**qMSP**
Fixed	**OR**	**OR**	**OR**	**OR**
β_0 (95% CI) [1]	0.03	0.03	0	0.22
	(0.01 – 0.12)	(0.01 – 0.12)	(0.00 – 0.03)	(0.06 – 0.73)
β_1 (95% CI) [1]	22.42	19.95	799.19	3.19
	(6.33 – 79.39)	(5.69 – 69.99)	(32.64 – 9568.77)	(2.30 – 4.44)
Random				
σ^2 (prep:methplex)	0.578	1.068	5.243	2.299
σ^2 (prep)	0	0	0	0
$N_{methplex:prep}$	9	9	9	9
N_{prep}	3	3	3	3
Observations	243	243	243	243
Deviance	85	90.455	37.242	150.513

[1] (95% CI) = Profile confidence intervals

5.3.2.2 MethPlex ddPCR measurement at low DNA inputs is more precise than qMSP

I compared precision and trueness of the two methods using data in the 6 – 96 GE input range, for both methylated DNA assays. As above, standard deviation and coefficient of variation as measures of precision, and bias as an indicator of trueness. This range was chosen as it encompasses all inputs above LOD95 for three out of the four methylated gene promoter assay/detection methodology combinations without excessively reducing the range of data available for analysis.

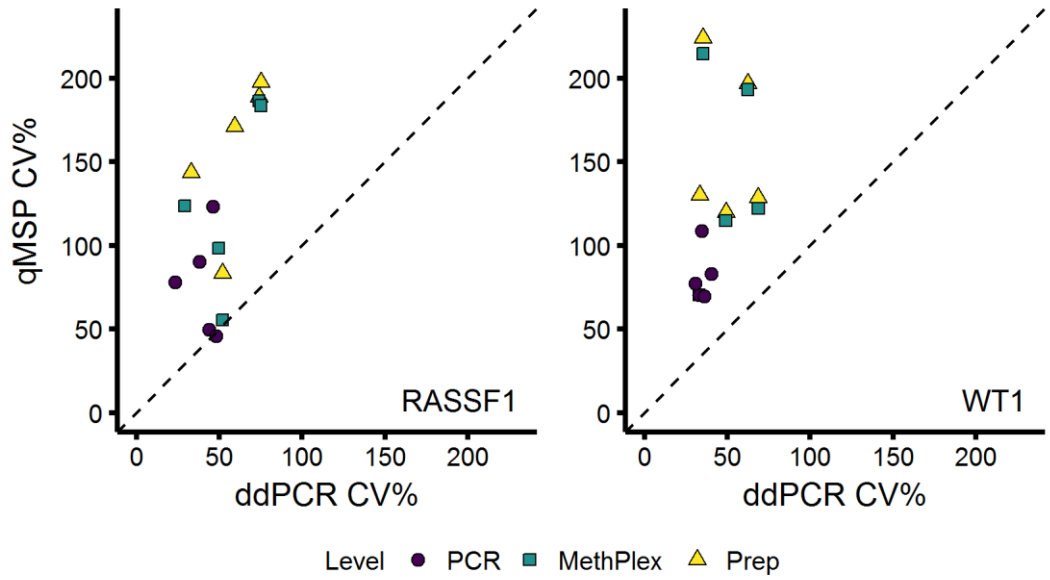


Figure 5.3-12 Coefficient of variation comparison plots for low total DNA MethPlex inputs
Coefficients of variation were calculated for both methods for all dilution points above 6 GE and were plotted against each other. CV% values were calculated by variance component analysis using the VCA R package (Schuetzenmeister, 2017). Data points clustered to one side of the line of identity indicate greater imprecision for the method associated with the nearest plot axis.

As in **5.3.1.2**, ddPCR concentration (copies/µl) was transformed by dividing the response variable by the mean average response for maximum methylated DNA input, in this case 96 GE, to enable comparison of precision between methods on a comparable relative scale. Precision analyses were performed using procedures described in technical standard ISO 5725 (parts 2, 3 and 6) (ISO, 1994a, 1994b, 1994c).

It is evident immediately from examination of scatter plots (**Figure 5.3-10**) that there was far greater variation in methylated DNA measurement by qMSP, with large 95% confidence intervals and inconsistency in the pattern of means. There is a clear implication that the response does not follow a linear relationship due to inherent randomness, imprecision and inaccuracy of measurement, despite a general positive

trend. These initial observations were further evaluated using more formal methodologies.

Repeatability standard deviations (s_r), gauging the precision of measurement at the PCR level, were generally lower when using MethPlex ddPCR assays and, therefore, equivalence of variances (s_r^2) was also tested. MethPlex ddPCR displayed significantly better precision at four out of the five inputs quantities above 3 GE for measuring *WT1* methylation (**Appendix, Table 8.5**). Average variances across input levels (root mean square average, using $\log_2$ transformed data to account for variances increasing in proportion to input DNA quantity) were also significantly lower for methylated DNA measurement by ddPCR (*RASSF1* $F_r = 0.654$, *WT1* $F_r = 0.375$, $F_{(0.025),\ 90,\ 90} = 0.659$).

Much the same as when evaluating high total DNA assay input analyses (**5.3.1.2**), repeatability standard deviations and intermediate standard deviations ($s_{i(prep)}$) were proportional to methylated DNA input. The coefficient of variation (CV%) was therefore compared between methods. When MethPlex ddPCR and qMSP CV% values were plotted against one another, plotted points distinctly clustered above the line of identity, indicating consistently higher qMSP imprecision at all nominal methylated DNA inputs above 6 GE and across hierarchical design levels. Measurement by MethPlex ddPCR reduced *RASSF1* and *WT1* average PCR CV% by 50% and 57% respectively. Similarly MethPlex level CV% was reduced by 56% and 63%, and preparation level CV% was reduced by 62% and 68% (**Table 5.3-10**).

Table 5.3-10 MethPlex coefficient of variation comparison

Coefficients of variation, calculated at the PCR, MethPlex and preparation level, are presented for nominal methylated DNA inputs above 20 GE. CV% values were calculated by variance component analysis using the VCA R package (Schuetzenmeister, 2017). Reduction in CV = (CV_{qMSP} – CV_{ddPCR})/ (CV_{qMSP}). Averages for method and promoter assay combinations are root mean square averages.

Gene	Nominal methylated DNA input (GE)	PCR			MethPlex			Preparation		
		CV ddPCR (%)	CV qMSP (%)	Reduction in CV (%)	CV ddPCR (%)	CV qMSP (%)	Reduction in CV (%)	CV ddPCR (%)	CV qMSP (%)	Reduction in CV (%)
RASSF1	96	23.3	78.1	70.2	29.0	123.6	76.5	33.0	143.5	77.0
	48	43.6	49.7	12.2	52.0	55.5	6.4	52.0	83.6	37.8
	24	48.0	46.0	-4.3	73.9	186.5	60.4	73.9	188.8	60.9
	12	46.3	123.3	62.4	75.1	183.6	59.1	75.1	197.5	62.0
	6	38.1	90.3	57.8	49.5	98.5	49.7	59.2	171.3	65.4
	Average	N/A	N/A	49.7	N/A	N/A	55.7	N/A	N/A	61.9
WT1	96	32.9	70.5	53.3	32.9	70.5	53.3	33.5	130.3	74.3
	48	36.0	69.8	48.4	49.0	114.9	57.3	49.2	119.8	58.9
	24	34.7	108.7	68.1	68.7	122.1	43.7	68.7	128.7	46.6
	12	40.5	83.0	51.2	62.5	193.2	67.7	62.5	196.8	68.3
	6	30.6	77.4	60.5	35.3	214.7	83.6	35.3	224.1	84.3
	Average	N/A	N/A	56.7	N/A	N/A	62.6	N/A	N/A	67.7

5.3.2.3 MethPlex qMSP displays substantial proportional bias across the measured range which is not improved by limiting the modelled range

Bias was assessed using a similar approach to the high total DNA input case (**5.3.1.3**). In this instance, however, linear regression between observed and expected relative methylated DNA abundance was undertaken using linear mixed models to account for the hierarchical, nested data structure and account for lack of independence of observations that would otherwise affect parameter standard errors and inferences made relating to the biases of interest. The procedure produces a straight line ($y = x$), if the assay is optimal and exhibits no bias, and statistical hypothesis testing of the intercept and slope assesses whether any observed biases are significant. Data were modelled in $\log_2$ space in order to correct for the lognormal distribution of response data and heteroscedasticity. Model residuals were assessed for normality and homogeneity of variance and were satisfactory (**Appendix, Figure 8.7**).

Table 5.3-11 MethPlex enrichment assay bias modelling parameters and statistical values for inputs above 3 GE nominal methylated input

		Constant bias (H_0: $\beta_0 = 0$)					Proportional bias (H_0: $\beta_1 = 1$)				
Gene	Method	β_0	Bootstrap 95% CI	df	F[1]	P-value	β_1	Bootstrap 95% CI	df	t[2]	P-value
RASSF1	ddPCR	-0.23	(-0.55, 0.12)	(1, 4)	1.78	0.25	1.07	(0.91, 1.23)	10.1	0.88	0.401
	qMSP	-0.27	(-2.26, 1.9)	(1, 2.2)	0.06	0.83	1.76	(1.4, 2.13)	10.1	4.05	0.002
WT1	ddPCR	-0.24	(-0.58, 0.12)	(1, 3.3)	1.99	0.25	0.98	(0.9, 1.05)	127	-0.56	0.576
	qMSP	-0.22	(-1.85, 1.33)	(1, 2.5)	0.07	0.81	1.81	(1.3, 2.37)	10.1	2.93	0.015

[1] Kenward-Roger adjusted F-test [2] Wald t-test

Initial inspection of plots of the marginal (average) fitted models, clearly shows that both MethPlex ddPCR assays provide for regression lines closely following the line of equality (**Figure 5.3-13, left-hand panels**). MethPlex qMSP assays, on the other hand, were clearly suboptimal, producing modelled plots with far steeper gradients

than the identity line, implying severe proportional bias (**Figure 5.3-13, right-hand panels**). Formal qualification of these observations revealed that neither measurement technique exhibited constant bias (**Table 5.3-11**), corresponding to a vertical shift in the regression line. Indeed, statistical testing revealed that neither MethPlex ddPCR assay projected proportional bias, while both MethPlex qMSP assays displayed proportional bias significant at the 95% confidence level (**Table 5.3-11**). Proportional biases, where present, were not removed on reduction of the modelled range (**Figure 5.3-13**).

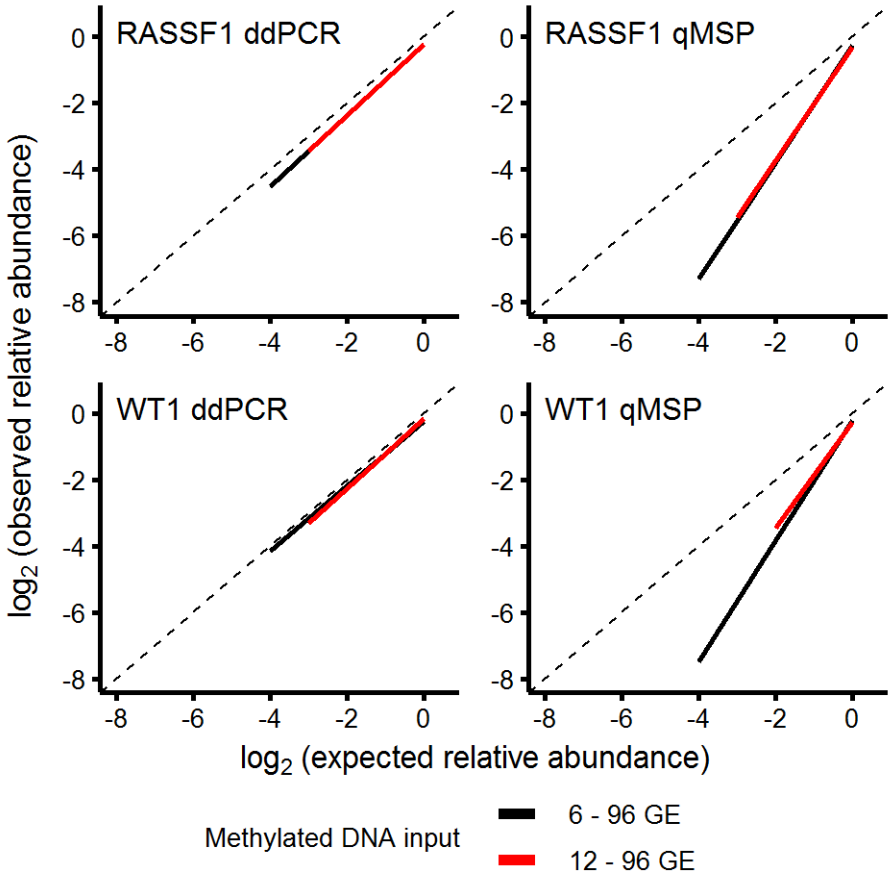


Figure 5.3-13 Linear mixed effects regression plots of expected vs observed relative methylated DNA abundance to assess assay bias
Expected and observed relative abundance data were log2-transformed, modelled by linear mixed effects regression and representative marginal fits are plotted reproduced above. This should yield a straight line not significantly different to the equality line. Slopes and intercepts statistically different to one and zero indicate the presence of linear proportional bias and constant bias respectively (**Table 5.3-11**).

5.3.2.4 MethPlex ddPCR and *RASSF1* qMSP can discriminate twofold differences in methylated DNA inputs in the 6 to 96 nominal methylated DNA GE range

Linear mixed effects regression of measured relative abundance on nominal methylated DNA input above 3 GE was used to evaluate the ability of assays to differentiate different methylated DNA inputs.

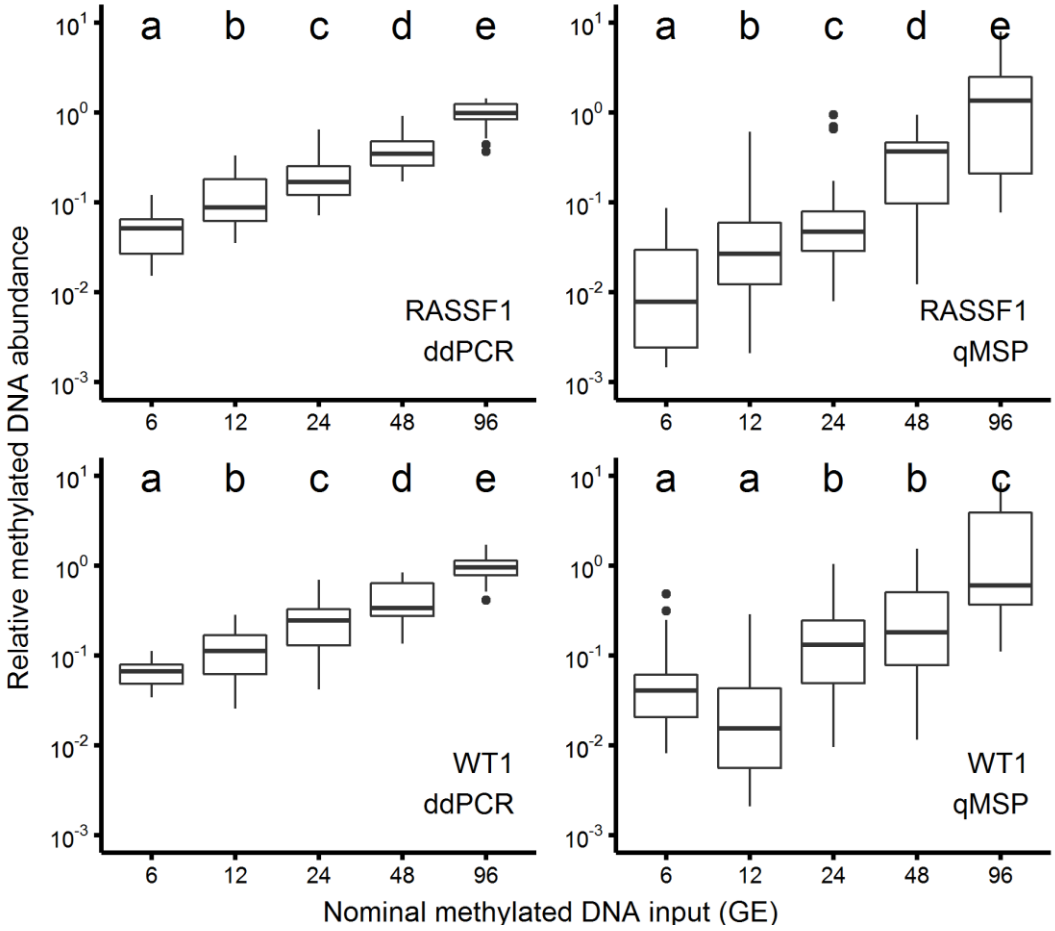


Figure 5.3-14 10ng DNA MethPlex assay input boxplots in the 6 – 96 GE nominal methylated DNA input range
Box and whisker plots depict distributions of data for measurement of methylated DNA abundance by ddPCR and qMSP. Box limits represent first and third quartiles and the central line shows the median. Whiskers denote 1.5 x interquartile range and outliers beyond the limits of the whiskers appear as individual points. Inputs with the same letter at the top of the boxplots are not significantly different by Tukey's honest significant difference test (Hothorn et al., 2008). Quantification method and methylated gene promoter assay are indicated in the bottom right corner of individual boxplots.

Post-hoc pairwise comparisons (Tukey's honest significant difference test, using the multcomp R package (Hothorn et al., 2008)) (**Figure 5.3-14** and **Table 5.3-12**) reveal that measurement of methylated *RASSF1* DNA by both MethPlex ddPCR and MethPlex qMSP, and methylated *WT1* DNA by MethPlex ddPCR distinguished all twofold differences across the tested range. MethPlex qMSP quantification of WT, on the other hand, was limited to fourfold differences at best in the 6 to 48 GE range. Median *WT1* qMSP measurements were, in fact, higher for 6 GE nominal methylated DNA input than for 12 GE. This difference was not, however, statistically significant (**Figure 5.3-14** and **Table 5.3-12**).

Table 5.3-12 Table of statistics resulting from Tukey's honest significant difference test performed upon MethPlex enrichment ddPCR and qMSP data
This limited table presents statistical values limited to twofold comparisons.

		ddPCR				qMSP			
Gene	**Comparison (Methylated DNA input) [GE]**	**Estimate**	**95% CI**	**Tukey's q statistic**	**Adjusted *P* value**	**Estimate**	**95% CI**	**Tukey's q statistic**	**Adjusted *P* value**
RASSF1	12 - 6	1.19	(0.67, 1.71)	6.26	2.8×10^{-9}	2.29	(1.39, 3.19)	6.93	1.5×10^{-11}
	24 - 12	0.91	(0.39, 1.43)	11.03	2.0×10^{-5}	1.02	(0.12, 1.93)	3.1	0.017
	48 - 24	0.94	(0.42, 1.45)	15.95	7.6×10^{-6}	1.77	(0.87, 2.67)	5.35	6.6×10^{-7}
	96 - 48	1.41	(0.89, 1.93)	23.35	6.6×10^{-13}	2.33	(1.43, 3.23)	7.05	9.7×10^{-12}
WT1	12 - 6	0.71	(0.22, 1.19)	4.00	5.7×10^{-4}	0.2	(-1.19, 1.59)	0.39	0.995
	24 - 12	1.04	(0.56, 1.52)	9.88	2.4×10^{-8}	3.57	(2.18, 4.96)	7.00	8.3×10^{-12}
	48 - 24	0.9	(0.42, 1.38)	15.00	2.9×10^{-6}	0.72	(-0.67, 2.11)	1.41	0.618
	96 - 48	1.28	(0.8, 1.76)	22.25	2.0×10^{-12}	2.44	(1.05, 3.83)	4.8	1.7×10^{-5}

5.3.2.5 Random effects variances are greater when DNA methylation is measured using MethPlex qMSP

Linear mixed models also characterized the variation in methylated DNA measurement in terms of the clustering of preparation and MethPlex samples resulting from the dependence structure dictated by the nested experimental design. Recall that "preparation error" includes variability due to dilution series construction, HinP1I digestion and bisulphite treatment, while "MethPlex error" incorporates variability resulting from MethPlex pre-amplification and purification. In modelling the data in this way, the variation in measured response not attributable to the methylated DNA input quantity was (partially) explained by the differences in preparation and MethPlex clusters from the average population level response and informatively enabled the provision of variance estimates for these random effects.

Table 5.3-13 Statistical significance of MethPlex enrichment DNA methylation detection preparation and MethPlex random effects

		Preparation random effects			MethPlex random effects		
Gene	**Method**	**df**	**-2 log Λ[1]**	**P value[2]**	**df**	**-2 log Λ[1]**	**P value[2]**
RASSF1	ddPCR	1	0.00	1	3	34.08	1.51×10^{-7}
	qMSP	1	8.07	0.0023	3	51.99	1.76×10^{-11}
WT1	ddPCR	1	0.00	1	1	21.95	1.40×10^{-6}
	qMSP	1	6.06	0.0069	3	47.93	1.29×10^{-10}

[1]Likelihood Ratio Test
[2]Corrected P value (Self and Liang, 1987; Stram and Lee, 1994)

In the assessment of both methylated gene promoter assays by MethPlex ddPCR, assessment of linearity by quadratic regression did not indicate a deviation from linearity. Random preparation intercept effects were not significant (**Table 5.3-13**),

but were retained to account for the data structure. MethPlex qMSP assays, however, did produce significant random preparation intercepts (**Table 5.3-13**). This is more intuitively demonstrated by plots of the individual fitted regression lines at the MethPlex level displayed in **Figure 5.3-15** and **Figure 5.3-16**.

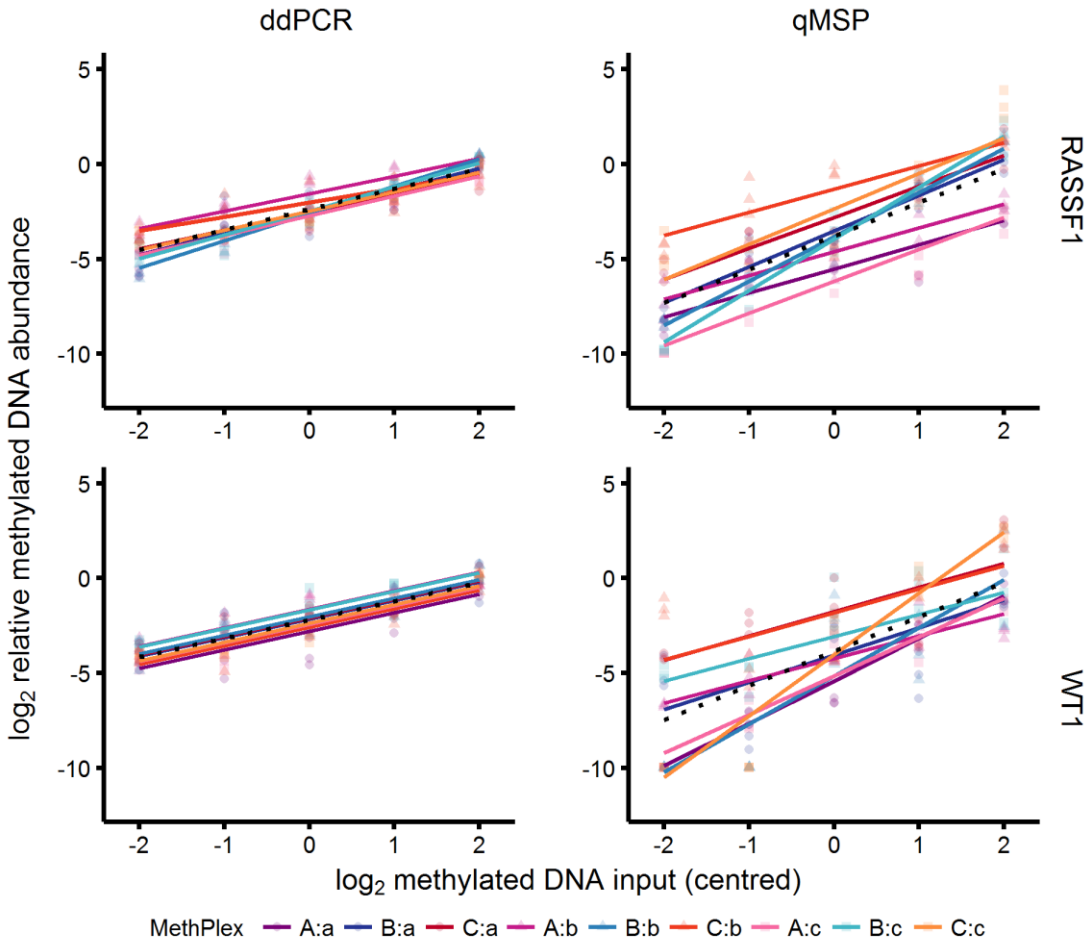


Figure 5.3-15 Linear mixed effects regression plots generated from 10ng total DNA input MethPlex enrichment methylated DNA measurement data
Log_2-transformed relative abundance data were modelled in the 6 – 96 GE nominal methylated DNA input range, with log2-transformed methylated DNA input quantities centred on the 24 GE input for more informative interpretation of the intercept (x = 0). The regression lines, indicated by different colours shown in the legend below the plot, are plots of fitted values at the MethPlex level and give an indication of data variability. The marginal (average) regression line is also plotted (black dotted line).

They unequivocally show deviation of the qMSP regression lines from the marginal (average) line indicated by the black dotted lines. This is apparent both in terms of vertical shifts (equivalent to random intercepts) and differences in line gradients

(corresponding to random slopes), and is evident at both the preparation and MethPlex level. While some divergence from the population average is noticeable for ddPCR plots, the magnitude of qMSP variation is much greater.

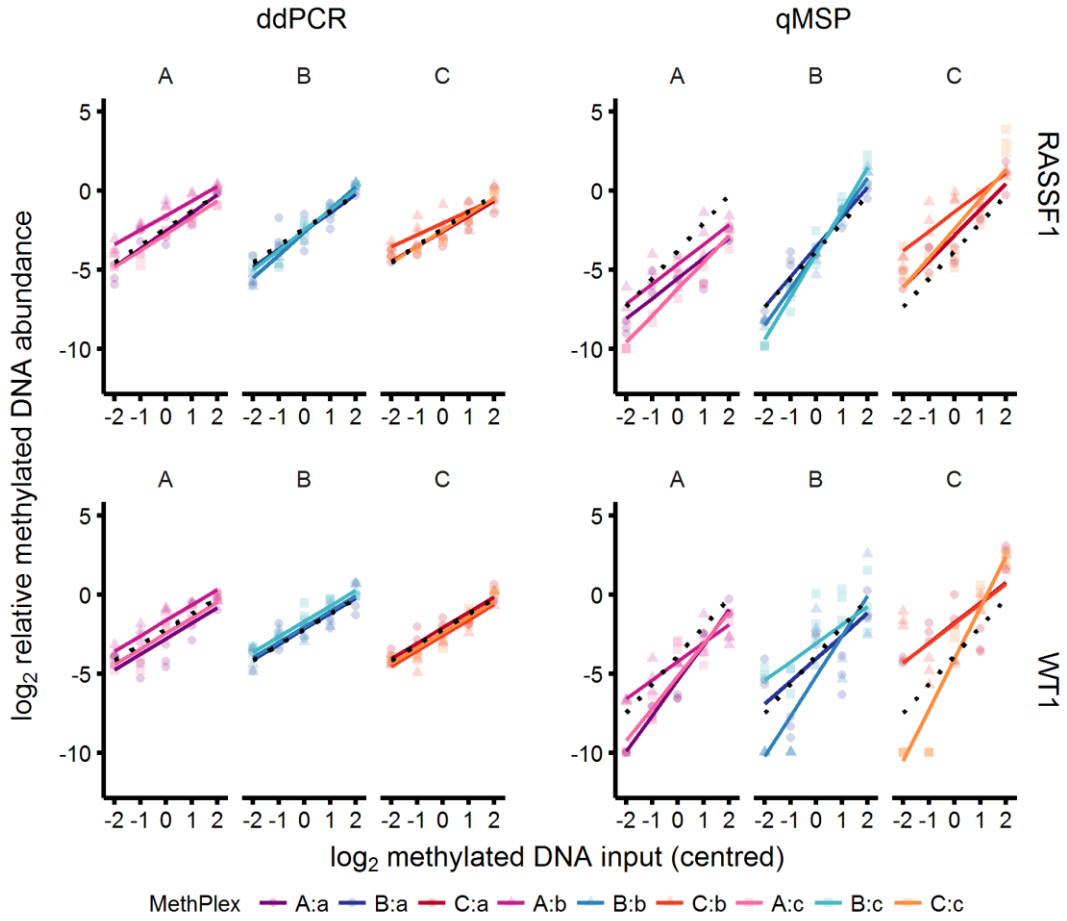


Figure 5.3-16 MethPlex enrichment ddPCR and qMSP linear mixed effects regression plots, facetted by preparation
Log_2-transformed data, modelled and plotted in **Figure 5.3-15**, are replotted with different preparation groups separated out into different facets to better enable contrast of prep level and MethPlex level regressions. Again, regression lines are at the MethPlex level, indicated by different colours shown in the legend below the plot, and the marginal (overall average) regression line is represented by the black dotted line.

This is confirmed by examination of the linear mixed model regression parameters (**Table 5.3-14**).

MethPlex qMSP assay models produce random effects variances that are more considerable than those for MethPlex ddPCR. For example, the *RASSF1* qMSP

MethPlex random intercept variance is more than five times larger than its ddPCR counterpart (**Table 5.3-14**). A more intuitive representation of this can be seen in random effects plots (**Figure 5.3-17**), showing the magnitude and confidence intervals of random effects predictors. MethPlex qMSP random effect magnitudes are larger and more variable, with a greater number of significant individual random effect groups, indicated by dots and error bars highlighted in black.

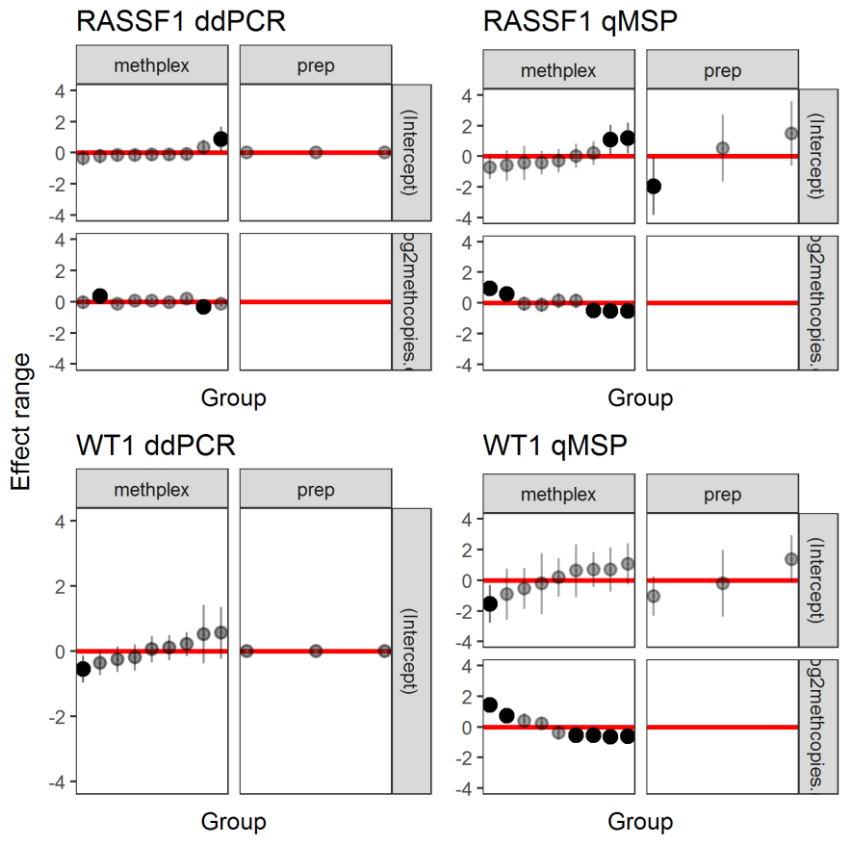


Figure 5.3-17 Linear mixed model conditional means of preparation and MethPlex group-level effects
Plots of conditional means of random effects give a view of the variation between preparation groups and that between MethPlex clusters and the magnitude of these effects. 95% confidence intervals were constructed by simulation (n = 1000) and indicate a non-significant random effect when they cross zero (red horizontal line). Points that are distinguishable from zero (i.e. the confidence interval does not cross the red line) represented in black. Within an individual plot (e.g. "*RASSF1* ddPCR") upper panels show the random effects for intercepts and the lower left panel displays the conditional means for MethPlex random slopes (lower right panels are empty and uninformative). Note that the *WT1* ddPCR model included random intercepts only for both preparation and MethPlex level nesting.

Further inspection of model parameters (**Table 5.3-14**), informs that MethPlex ddPCR fixed effect slope parameters (β_1) are close to unity, indicating that a twofold increase in input results in a near twofold increase in measurement, while qMSP slopes approach two, with evident proportional bias (**5.3.2.3**). MethPlex ddPCR parameters are also more precise, signified by narrower confidence intervals.

Table 5.3-14 Summary of MethPlex enrichment ddPCR/qMSP methylation detection linear mixed model regression parameters

	RASSF1		***WT1***	
	ddPCR	**qMSP**	**ddPCR**	**qMSP**
Fixed				
β_0 (95% CI)[1]	-2.38	-3.79	-2.19	-3.85
	(-2.64, -2.09)	(-5.73, -1.83)	(-2.49, -1.87)	(-5.43, -2.09)
β_1 (95% CI) [1]	1.07	1.76	0.978	1.81
	(0.90, 1.24)	(1.40, 2.13)	(0.89, 1.06)	(1.32, 2.36)
Random				
σ^2 ($Preparation_{intercept}$)	0	3.159	0	1.711
σ^2 ($MethPlex_{intercept}$)	0.155	0.599	0.177	0.916
σ^2 ($MethPlex_{slope}$)	0.050	0.2857	NA	0.606
σ^2 (Residual)	0.397	0.990	0.428	2.712
Observations	135	135	135	135
R^2m / R^2c	0.781 / 0.867	0.541 / 0.914	0.761 / 0.831	0.503 /0.794

[1] (95% CI) = Bootstrap confidence intervals

The residual variance (i.e. the experimental variance not explained by the methylated DNA input quantity, preparation grouping and MethPlex clustering, and predominantly accounted for by measurement error) is far greater for MethPlex qMSP assays (**Table 5.3-14**).

5.3.2.5.1 Decomposition of variance

The contribution of preparation and MethPlex level clustering of data was also assessed at the individual nominal methylated DNA input quantities. Graphical representation of variance component analysis results displayed in **Figure 5.3-18 (upper panel)** indicate that preparation effects contributed little to the overall variance. Linear mixed model analysis, taking into consideration the full 6 – 96 GE methylated DNA input range and controlling for the variation due to the differences in these methylated DNA inputs, provides evidence that, overall, preparation did not contribute (**Figure 5.3-18, lower panel**). This suggests that there was no overall difference imparted upon similar DNA samples by HinP1I digestion and bisulphite treatment when measured by MethPlex ddPCR. Furthermore, when the contributions of the fixed effect (nominal methylated DNA input), random effect (preparation and MethPlex) and residual (predominantly PCR measurement error) to the total modelled variance were analysed, ddPCR random effects were quite insubstantial as a proportion of the total experimental variance (**Figure 5.3-18, right-hand panel**). In contrast, random effects were relatively large in both qMSP instances (*RASSF1* qMSP: 0.373, *WT1* qMSP: 0.291, *RASSF1* ddPCR: 0.086, *WT1* ddPCR: 0.07). Moreover, nominal methylated DNA input explained greater than 75% of the experimental variance when samples were analysed by ddPCR , whereas qMSP measurement explained around 50% ($R^2_{GLMM(m)}$ (Johnson, 2014; Nakagawa and Schielzeth, 2013) *RASSF1* qMSP = 0.541, *WT1* qMSP = 0.503) (**Figure 5.3-18, right-hand panel**).

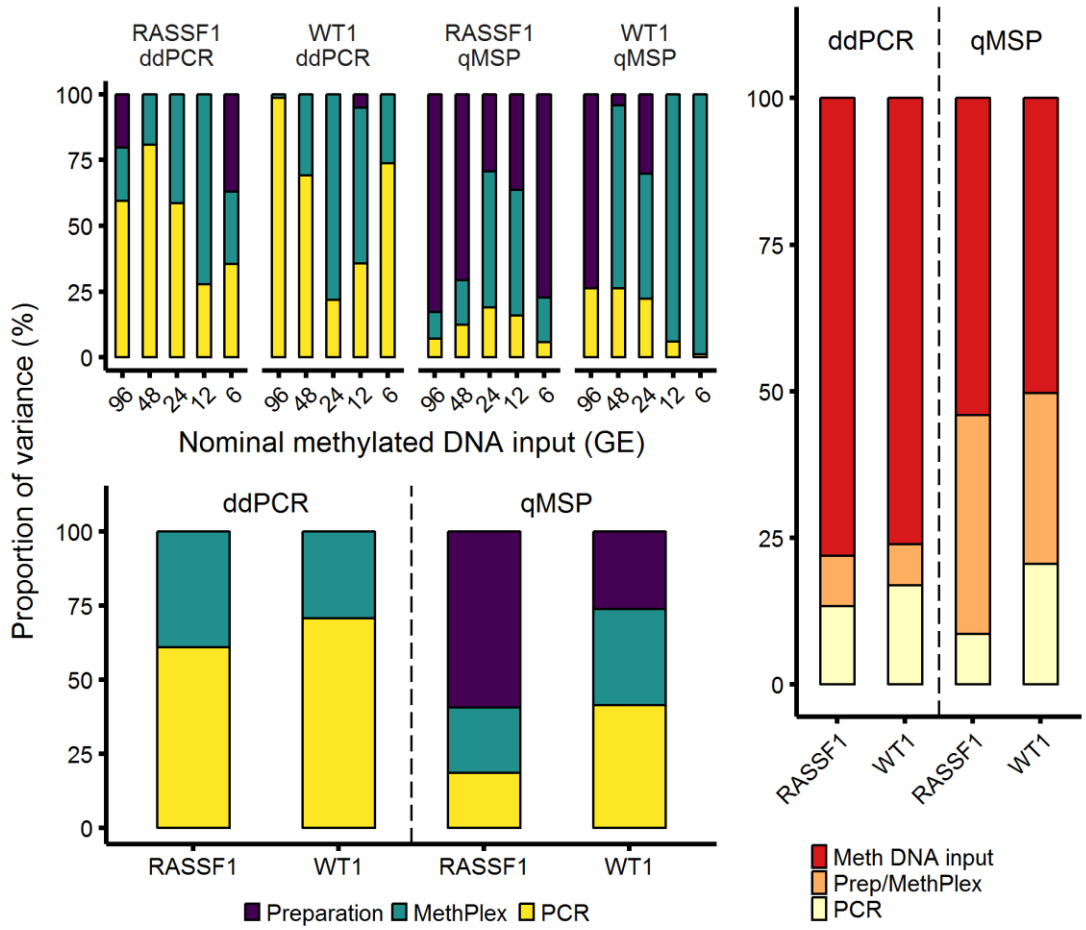


Figure 5.3-18 Variance component analysis of MethPlex methylated DNA quantification data
The **upper panel** shows the contribution of the different levels of the experimental nested hierarchy to the total variance at the different nominal methylated DNA input quantities. Results were acquired using nested ANOVA implemented in the VCA R package (Schuetzenmeister, 2017) and are shown in **Appendix, Table 8.6**. The four MethPlex DNA assay/ detection method combinations are displayed above. The **lower panel** shows these same contributions modelled across the methylated DNA input range, through analysis by linear mixed effects regression and assessment of intra-class correlation (ICC or repeatability) coefficients, controlling for the variation due to the methylated DNA input, using the rptR package (Stoffel et al., 2017). The **right-hand panel** shows the contributions of the methylated DNA input (fixed effect), preparation/MethPlex (random effects) and PCR measurement error (residual variance) to the total experimental variance.

5.3.2.6 MethPlex ddPCR models provide accurate and precise prediction of methylated DNA abundance but MethPlex qMSP models perform poorly

As in **5.3.1.6.1**, I made predictions using MethPlex enrichment models, predicting marginal measured relative methylated DNA abundance at specified input quantities for the average MethPlex enrichment sample within the average preparation. This

provided predicted average relative abundance estimates for the possible population of preparation and MethPlex enrichment processes. Sampling distributions for given predictions were again established by bootstrapping and associated 95% confidence intervals were constructed.

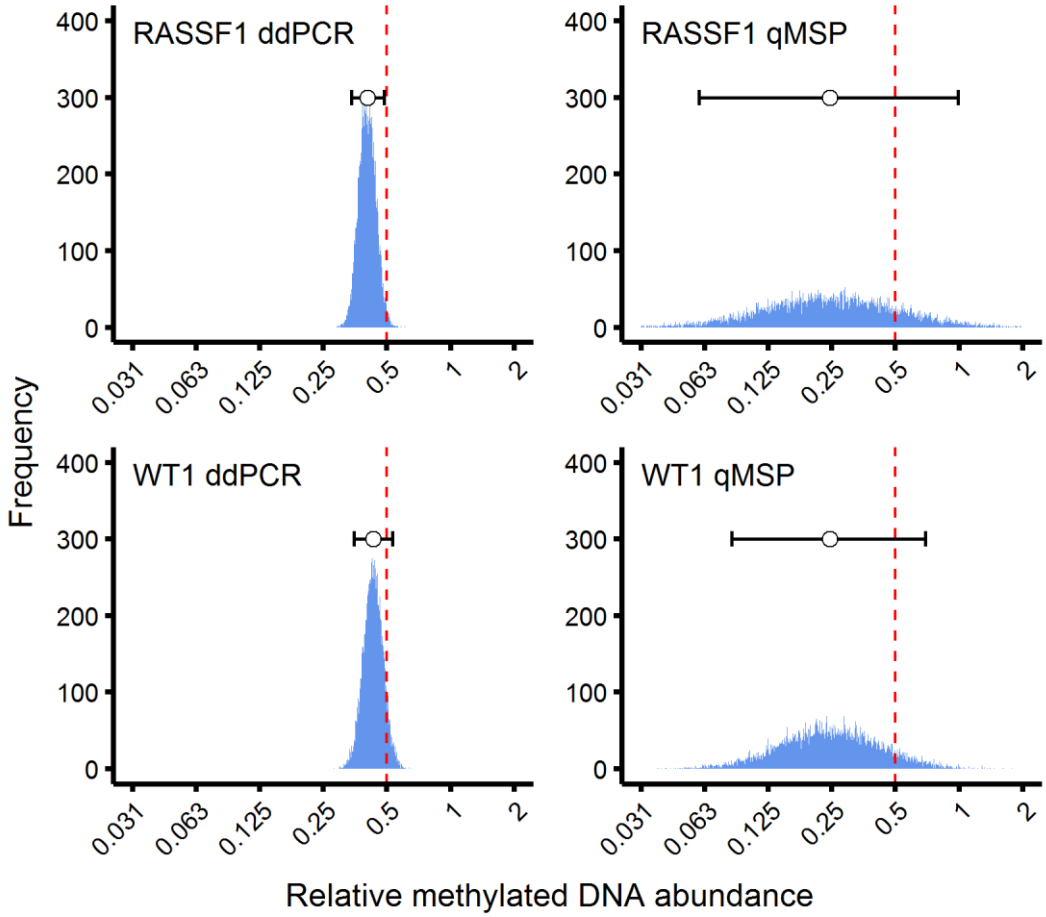


Figure 5.3-19 Histograms of bootstrap estimates for prediction at 48 GE nominal methylated DNA MethPlex pre-amplification input
Sampling distributions from 10,000 bootstrap model simulations presented for each gene promoter methylation assay and measurement method. Model output has been back-transformed into the relative measurement scale. If models and measurement assays are accurate distributions should peak at 0.5 relative methylated DNA abundance, indicated by the red dashed line. Point estimate of predictions are shown as empty circles and error bars represent 95% confidence intervals.

Predictions at x = 1 for the centred models correspond to a nominal methylated DNA input of 48 GE or, alternatively stated, expected relative methylated DNA abundance of 0.5. MethPlex ddPCR prediction estimates at this input, back-transformed into the

measurement scale, are more similar to 0.5 than qMSP estimates (**Figure 5.3-19** and **Table 5.3-15**), underestimates relative methylated DNA abundance by 50%. **Figure 5.3-19** illuminates us to the fact that MethPlex qMSP predictions are very uncertain, indicated by the very wide, flat sampling distributions achieved in qMSP model resampling. Underestimation of relative abundance and very high levels of uncertainty were present across the input range for MethPlex qMSP models, whereas MethPlex ddPCR predictions proved to be precise and accurate (**Table 5.3-15** and **Appendix, Figure 8.8**).

Table 5.3-15 MethPlex enrichment linear mixed model predictions and their 95% bootstrap confidence intervals

Gene	Methylated DNA input (GE)	Expected relative abundance[1]	ddPCR		qMSP	
			Model predicted relative abundance[1]	(95% CI)[1]	Model predicted relative abundance[1]	(95% CI)[1]
RASSF1	96	1	0.85	(0.68, 1.08)	0.83	(0.2, 3.51)
	48	0.5	0.41	(0.34, 0.49)	0.25	(0.06, 0.99)
	24	0.25	0.19	(0.16, 0.23)	0.07	(0.02, 0.31)
	12	0.125	0.09	(0.07, 0.12)	0.02	(0.01, 0.09)
	6	0.0625	0.19	(0.03, 0.06)	0.01	(0, 0.03)
WT1	96	1	0.85	(0.67, 1.07)	0.86	(0.27, 2.65)
	48	0.5	0.43	(0.35, 0.53)	0.24	(0.08, 0.69)
	24	0.25	0.22	(0.18, 0.27)	0.07	(0.02, 0.22)
	12	0.125	0.11	(0.09, 0.14)	0.02	(0.01, 0.07)
	6	0.0625	0.06	(0.05, 0.07)	0.01	(0, 0.03)

[1]Model-based predictions and confidence intervals were on the $\log_2$ scale and are here presented back-transformed into the measurement scale

5.4 Discussion

DNA methylation is a potential biomarker for the early detection of cancer (Belinsky et al., 1998; Fleischhacker and Schmidt, 2007). For this potential to be realised, however, the accuracy of appropriate methods needs to be established (Redshaw et al., 2014). In this chapter, I have presented a comprehensive technical analysis of ddPCR and qMSP performance in the detection of methylated DNA, diluted in an appropriate unmethylated DNA background, at a high total DNA input and in combination with MethPlex target enrichment at low DNA input. This evaluation was performed for two genes of interest.

Since digital PCR technologies have been recently developed, relevant DNA methylation detection literature is somewhat limited, making it difficult to put these results in context. Analytical performance characteristics are often not reported, are inadequate or are relegated to supplementary sections where they are generally presented in an unclear fashion, even using inappropriate assessment techniques (Barault et al., 2015; Garrigou et al., 2016; Hata et al., 2017; Hayashi et al., 2015; Liu et al., 2017b; Uehiro et al., 2016; Yu et al., 2015). In general, this oversight is worrisome, as inadequate consideration of such factors could discredit fundamental conclusions (Bustin, 2017, 2014; Vynck et al., 2017) .

It is interesting that *RASSF1* methylation signal was detected in normal control PBMC DNA by ddPCR but not qMSP analysis of high total DNA input (**Figure 5.3-1**). WGA technical unmethylated control DNA, which is truly unmethylated since it is enzymatically amplified in the absence of 5methyl-dCTP, did not show any amplification for either measurement method. This indicates that these assays have

good specificity, provided by a combination of the specificity of the TaqMan assay design and pre-digestion with the methylation-sensitive endonuclease HinP1I. The most likely explanation for the observed difference between detection methods, therefore, is that the PBMC pool population has an underlying low level of *RASSF1* methylation that is detected by ddPCR but not qMSP. This indicates that ddPCR is more sensitive than qMSP in the detection of *RASSF1* DNA methylation. The low background *RASSF1* methylation is not unexpected as PBMC control DNA was purified from many (>96) blood bank PBMC samples without positive cancer diagnosis. Furthermore, *WT1* methylation was detected by both techniques at high total DNA input.

RASSF1 and *WT1* methylation was detected at high input (20,000 GE) in normal PBMC DNA by ddPCR but not with low input MethPlex ddPCR (**Figure 5.3-1** and **Figure 5.3-10**). If the PBMC pool possessed a low degree of background methylation, say 0.0125%, then PBMC PCR reactions would contain on average 2.5 methylated DNA copies. There is variation in the copy number between wells, due to random sampling, and they, of course, cannot contain exactly 2.5 methylated DNA copies, some wells receiving more, some fewer, with probabilities dictated by the Poisson distribution. The probability of a reaction containing one or more methylated DNA copies would be greater than 92%, and DNA methylation would be readily detected. For the MethPlex low DNA input case (3000 GE), however, methylated DNA would be present on average at 0.38 copies per MethPlex enrichment reaction and the probability that a pre-amplification reaction contained one or more methylated DNA copies would be only 31%. MethPlex *WT1* qMSP reactions were methylation positive in two out of nine MethPlex enrichment samples, from different preparations (33%

occurrence for each of the two preparations). Taking the above together suggests that positive signals in PBMC DNA samples were genuine methylated DNA background, albeit at a very low incidence, which was effectively "diluted out" in low total input MethPlex assays. This therefore confirms the analytical specificity of these PCR-based techniques.

The above thought experiment may have a disadvantage: in exchange for high MethPlex ddPCR specificity there could potentially be an accompanying decrease in analytical sensitivity. MethPlex ddPCR LOD95 values of below 4 GE, however, suggest otherwise. In fact, the theoretical limit of detection for qPCR assays in general is three target molecules per reaction volume, and in practice is higher because of technical imprecision (Ståhlberg and Kubista, 2014). Furthermore, ddPCR limits of detection (**Table 5.3-2**, $\mathbf{X_D}$) for the 66 ng total DNA input study (i.e. without pre-amplification) were at least as low as those for methylated *EVL* and *NTRK3* ddPCR analysis of methylated DNA diluted in a water background (Yu et al., 2015). Moreover, MethPlex limits of detection were considerably lower, both for ddPCR and qMSP detection.

A novel minor differentially methylated allele population enrichment technique, involving double-strand specific nuclease digestion of bisulphite DNA-oligonucleotide probe duplexes (MS-NaME), was recently described (Liu et al., 2017b). MS-NaME followed by ddPCR analysis was able to increase methylated RARβ2 (*RARB*) representation from 0.1% to 20%, thus enabling detection (Liu et al., 2017b). This relative measure cannot be directly compared to the data from the current study; however, assuming a mass of 3.3 pg per haploid human genome, MS-NaME

inputs can be calculated, and 0.1% would equate to 6 GE, comparable to measurable MethPlex enrichment inputs. Unfortunately, this study lacks assessment of accuracy and precision.

Both ddPCR and MethPlex ddPCR, at high and low total DNA inputs respectively, demonstrated superior precision in the measurement of methylated *RASSF1* and *WT1* (**5.3.1.2** and **5.3.2.2**). Analysis of *RASSF1* methylation by ddPCR reduced the PCR level coefficient of variation across the measured methylation range by 34% for high total DNA inputs, imitating clinical assays performed on tissue samples. MethPlex enrichment of low DNA inputs, simulating cfDNA extracted from plasma specimens, reduced the same measure of precision by 50%. A number of studies have similarly established high digital PCR sensitivity and precision when compared to qPCR (Hindson et al., 2011, 2013; Whale et al., 2012). This contradicts data generated in the measurement of cytomegalovirus DNA (Hayden et al., 2013) and methylated DNA (Redshaw et al., 2014), with both of these studies attributing poorer precision to digital PCR when compared to qPCR measurement. Both studies, however, analysed a lower quantity of DNA in digital PCR analyses, as acknowledged by both investigating groups (Hayden et al., 2013; Redshaw et al., 2014).

Both qMSP and its MethPlex variant suffered from considerable biases when assessed directly and through modelling (**5.3.1.3**, **5.3.1.6**, **5.3.2.3** and **5.3.2.6**). It is most striking that predicted biases from qMSP models approached 100% over-estimation at the 160 GE nominal methylated DNA input level **(5.3.1.6.1**). Relative biases of similar magnitude were observed from direct estimation (**Appendix, Table 8.2)**

The apparent superior precision and trueness that ddPCR conferred on these assays could be explained through the partitioning into thousands of nanolitre-sized reactors potentially limiting the effect of PCR inhibitors. Furthermore, the fact that ddPCR is an end-point technique, and thus PCR efficiency is of less importance, could also provide tolerance to inhibitors (Hindson et al., 2013) and better enable quantification of highly fragmented bisulphite-treated DNA.

Variance component analysis of MethPlex enrichment data, both using classical nested ANOVA methodology and linear mixed models, indicated that preparation did not contribute substantively to the overall ddPCR variance. This is in agreement with likelihood ratio tests of preparation random intercepts (**Table 5.3-13**) which infer that preparation random intercepts were not significant in the case of ddPCR measurement. Significant qMSP preparation random effects (**Table 5.3-13**) and substantial preparation variance components (**Figure 5.3-18**) suggest differences in preparation groups despite ddPCR and qMSP being performed on the same samples. Considering superior ddPCR precision and trueness, and no difference between preparations observed in the high DNA input study, it is more likely that there was no true heterogeneity in preparations and differences seen in qMSP data were a manifestation of their higher uncertainty.

It is clear that there was substantial variation at the MethPlex level. This is unsurprising as multiplex reactions often exhibit amplification bias (Polz and Cavanaugh, 1998) and this is unlikely to be consistent from sample to sample due to the complex nature of the reaction kinetics involved. In future analysis of clinical samples, replication at this level may improve precision of acquired data. However,

since the quantity of methylated DNA in plasma samples is likely to be limiting, a more favourable approach would be to retain all of the bisulphite-treated DNA in a single MethPlex pre-amplification reaction and increase the likelihood of successful enrichment. In effect, the division into multiple MethPlex reactions reduces the number of methylated DNA copies per sample and so increases experimental noise (Tichopad et al., 2009).

The ability of ddPCR, and its combination with MethPlex enrichment at limiting input quantity, to resolve twofold differences in methylated DNA input could enable reliable and accurate quantification, representing a key improvement over the traditional real-time PCR method.

Mixed model analysis of data with only three random effects levels ("Preparation" effects) may well attract criticism: at least 5-6 random effects levels are generally recommended (Bolker et al., 2009). Between-cluster variances with fewer levels are likely to be poorly estimated. However, our interest is not in the individual preparation effects but how they vary and how any this variation in clusters applies to the population of possible preparations, in addition to controlling for non-independence of data resulting from hierarchical experimental design. Aggregating data would sacrifice important and interesting information relevant to assay performance and not including preparation predictors would potentially lead to incorrect inferences due to underestimation of regression coefficient standard errors. Thus, from a philosophical point of view, the designation of preparation as a random effect is justified. Furthermore, in both the high and low total DNA input cases, model (marginal) intercept and slope are identical if preparation is modelled as a

fixed or random effect. Calculated variances are very similar, but most importantly the inferences are the same. Statistical data tables illustrating this are included in **Appendix, Table 8.7** and **Appendix, Table 8.8**.

It is anticipated that this comprehensive technical study of various aspects of assay performance, including analytical specificity and sensitivity, precision, trueness, linearity and resolution, and an analysis of the influence of different workflow components, incorporating methods and procedures from ISO technical standards, forms a logical framework for the assessment and validation of potential biomarker assays at the analytical level, prior to pre-clinical study using patient specimens. In using such a framework, assays are more likely to perform satisfactorily in pre-clinical study and require little to no modification in order to meet stringent criteria demanded in formal clinical validation.

Chapter 6

DNA methylation panel ddPCR screening in blood plasma samples

6.1 Introduction

Conclusive lung cancer diagnosis typically requires histological examination of invasively obtained tissue biopsy tissue. Initial diagnostic work-up routinely includes less invasive bronchoscopy and cytological diagnosis. In Chapter 3, I evaluated the discriminatory accuracy of methylated DNA biomarkers detected in bronchoalveolar lavage specimens obtained during this procedure. While this is moderately invasive, it is not without risk to the patient. There is a clear need for genuinely non-invasive diagnostic procedures in the clinic. In addition, bronchial lavage is not a standardised process i.e. the bronchoscopist will sample the areas looking more suspicious rather than following a standard routine. Therefore, bronchial washings differ between patients regarding the representation of the bronchial tree. In contrast, blood is a readily accessible clinical specimen, highly standardised and presents an alternative, promising source of biomarkers (Crowley et al., 2013). Indeed, circulating tumour DNA (ctDNA) offers much potential and has been detected in patients with different cancer types (Schwarzenbach et al., 2011).

Detection of *SEPT9* methylation in plasma, marketed as the Epi proColon test, has recently been approved by the US FDA for colorectal cancer screening. This biomarker was previously validated in prospective trials (Church et al., 2014; Potter et al., 2014). *SHOX2* methylation assays were initially developed and validated for the diagnostic evaluation of bronchial aspirates (Dietrich, 2011; Schmidt et al., 2010). This marker has subsequently been assessed in plasma, with lower sensitivity (Kneip et al., 2011). However, recently, a combination of *SHOX2* and *PTGER4* methylation demonstrated a sensitivity of 67% and (fixed) specificity of 90% when validated using

plasma samples from lung cancer patients and individuals with and without benign lung disease (Weiss et al., 2017).

The utility of ddPCR analysis of plasma samples in copy number variation analysis (Gevensleben et al., 2013) and tumour genotyping (Oxnard et al., 2014) in cancer, and the detection in of nucleic acids in other diseases and disorders has been successfully demonstrated (Beck et al., 2013; Persaud et al., 2013; Strain et al., 2013).

Screening of this set of retrospective case-control plasma samples was intended to validate the use of MethPlex enrichment and methylation-specific ddPCR in the detection of DNA methylation in clinical plasma samples. It was performed to assess the ability of cfDNA concentration and individual methylation markers to discriminate between lung cancer cases and controls. Statistical modelling was planned to identify optimal discriminatory algorithms utilizing combinations of markers, and model validation by internal cross-validation was proposed.

6.2 Materials and methods

6.2.1 Liverpool Lung Project patients and samples

The plasma screening set included 180 individuals in total (60 lung cancer cases and 120 age/sex/storage duration frequency matched). All patients were recruited to the Liverpool Lung Project through the Liverpool Heart and Chest Hospital and were referred to the Liverpool Lung Cancer Unit because of clinical suspicion of lung cancer. As such, control subjects were not healthy volunteers, but received diagnosis for various diseases including, but not limited to, bronchitis, chronic obstructive pulmonary disorder (COPD), emphysema and tuberculosis. The LLP study protocol has been approved by the Liverpool Research Ethics Committee and all participants provided informed consent in accordance with the Declaration of Helsinki.

Blood plasma specimens were stored in at -80°C. DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen) according to the modified protocol in **2.2.2**.

6.2.2 Exploratory univariate analysis

Subjects' epidemiological, clinical and methylation characteristics were described by case-control status separately. Descriptive statistics were obtained and compared by using Chi-square test or Fischer's exact test for categorical variables and t-tests for normally distributed variables. The Mann-Whitney U test or Kruskal-Wallis test was used for continuous variables where the normality assumption failed. Epidemiological and clinical characteristics are presented in **Table 6.2-1**, plasma cfDNA yield and concentration statistics are displayed in **Table 6.3-1** and DNA methylation statistics are shown in **Table 6.3-2**. **Figure 6.2-1** provides additional evidence that there was no significant difference in the age, storage duration, smoking duration or smoking packyears distributions between case and control groups.

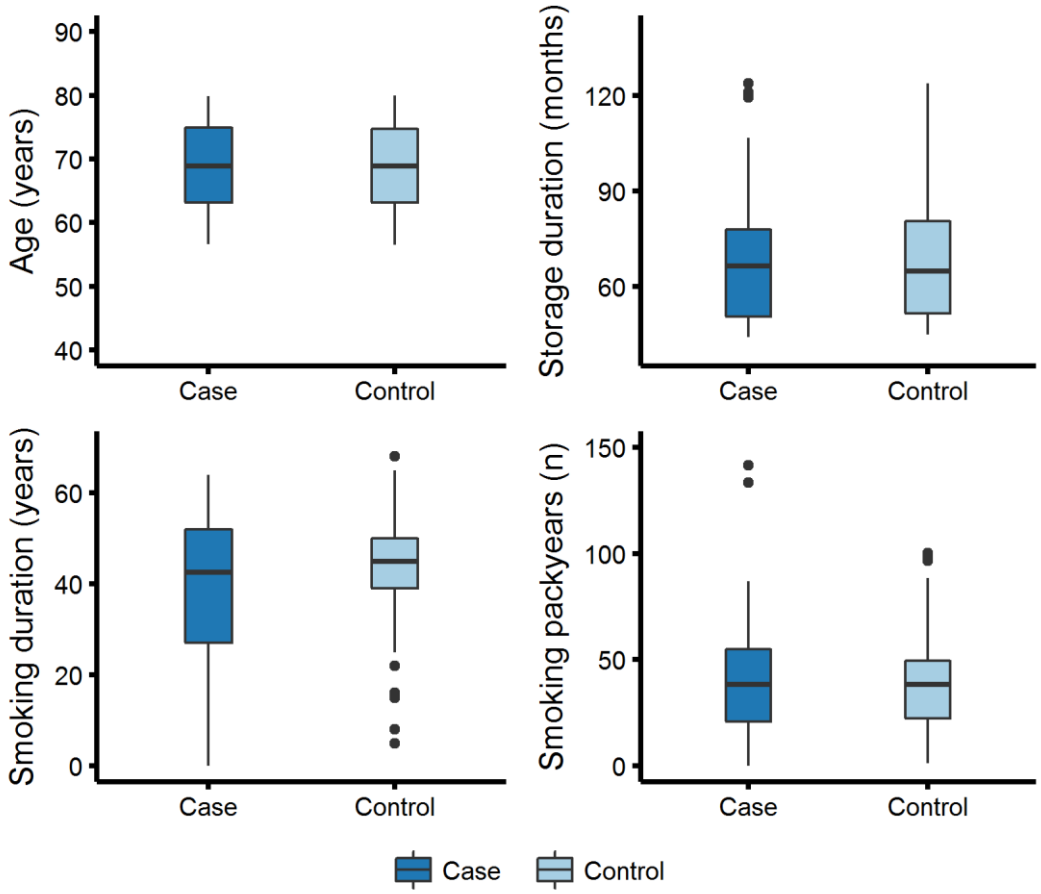


Figure 6.2-1 Box plots of epidemiological subject characteristics
Box limits represent first and third quartiles and the central line shows the median. Whiskers denote 1.5 x interquartile range and outliers beyond the limits of the whiskers appear as individual points. Note that there was no statistically significant difference between the distribution of cases and controls (P >0.05, Mann-Whitney U test).

Table 6.2-1 Epidemiological and clinical characteristics separated by case-control status for the plasma screening sample set

Numbers in the main body of the table represent frequencies. Numbers in parentheses represent percentages.

	Plasma DNA methylation detection set (N = 180)	
Subject characteristics	**Case (n = 60)**	**Control (n = 120)**
Age group		
<60	3 (5.0)	6 (5.0)
60 – 79	55 (91.7)	111 (92.5)
80+	2 (3.3)	3 (2.5)
Age		
Mean ± SD	69.1 ± 6.49	69.1 ± 6.41
Median ± IQR	69.0 ± 12	68.9 ± 11.3
Gender		
Male	29 (48.3)	58 (48.3)
Female	31 (51.7)	62 (51.7)
Smoking status[1]		
Current	23 (38.3)	43 (35.8)
Former	28 (46.7)	76 (63.3)
Never	9 (15.0)	1 (0.8)
Smoking duration		
Mean ± SD	36.7 ± 19.75	43.2 ± 12.66
Median ± IQR	42.5 ± 25	45 ± 12
Smoking pack years		
Mean ± SD	38.1 ± 29.24	41.7 ± 24.27
Median ± IQR	38.4 ± 34.5	39.4 ± 24.5
Specimen storage duration (years)		
<5	28 (46.7)	56 (46.7)
5+	32 (53.3)	64 (53.3)
Histological diagnosis		
Adenocarcinoma	36 (60)	
Squamous cell carcinoma	24 (40)	
Stage (pT)		
1	31 (51.7)	
2	19 (31.7)	
3	9 (15)	
4	1 (1.7)	
Nodal status (pN)		
0	40 (66.7)	
1	11 (18.3)	
2	9 (15.0)	

[1] Statistically significant ($P < 0.05$)

6.2.3 MethPlex ddPCR plasma sample DNA methylation screening

The workflow for MethPlex ddPCR DNA methylation screening is shown in **Figure 6.2-2**. Relevant sections describing different workflow components are indicated in the figure and following text. Briefly, plasma samples were thawed at room temperature. Sonicated salmon sperm DNA (500ng) and methylated pUC19 DNA (10^6 copies) were added to 1.8 ml aliquots of blood plasma. DNA was extracted according to the modified protocol described in (**2.2.2**). DNA was eluted in a total volume of 110 µl 0.22 µm-filtered Tris-HCl, pH 8.0, 0.025% Tween-20. 10 µl plasma cfDNA was aliquoted into PCR strips for real-time PCR genomic DNA quantification (**2.3.3**, **4.2.3**, **0**). The remaining volume of plasma cfDNA was transferred to PCR strips, desiccated and resuspended in 20 µl Tris-HCl, pH 8.0, 0.025% Tween-20 (0.22 µm-filtered). cfDNA samples were digested using methylation-sensitive endonuclease HinP1I (**2.6**) and bisulphite treated (**2.7**). Bisulphite-treated DNA (20 µl) was enriched for methylated targets of interest by MethPlex pre-amplification (**2.11.1**) and PCR products were purified (**2.11.2.2**). MethPlex enrichment products were diluted 1/8 (**6.3.3**) or 1/2 (**6.3.4**) with ddH_2O and analysed by methylation-specific ddPCR.

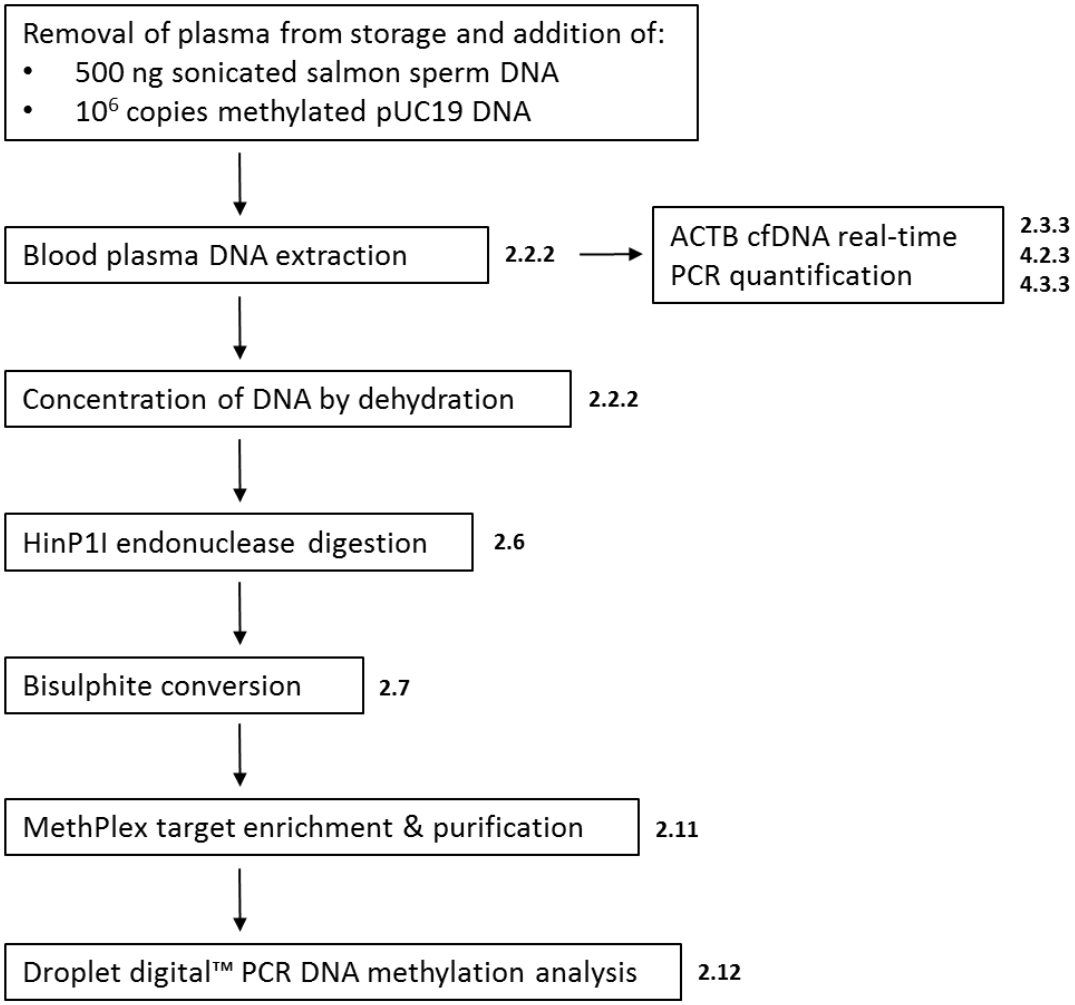


Figure 6.2-2 Outline of MethPlex ddPCR workflow
Numbers in bold to the right of boxes indicate chapter sections relevant to individual workflow components.

6.3 Results

6.3.1 *ACTB* cfDNA real-time PCR quantification assay displayed reliable and robust performance characteristics

DNA concentration was measured by qPCR using the in-house developed ACTB cfDNA real-time PCR assay (**2.3.3**). Samples were randomized across five 96-well qPCR plates. The assay demonstrated excellent precision (within-replicate Cq SD (pooled across plates and concentrations) = 0.166) and very good PCR efficiency (mean% = 95.9, 95% CI [93.2, 98.6]). It also produced an excellent average R^2 value of 0.998 (95% CI [0.997, 0.999]) (**Figure 6.3-1**). Amplification was detected in 178/180 plasma DNA samples tested, with two control subject samples providing no amplification signal (**Table 6.3-1**). No amplification was detected in non-template controls.

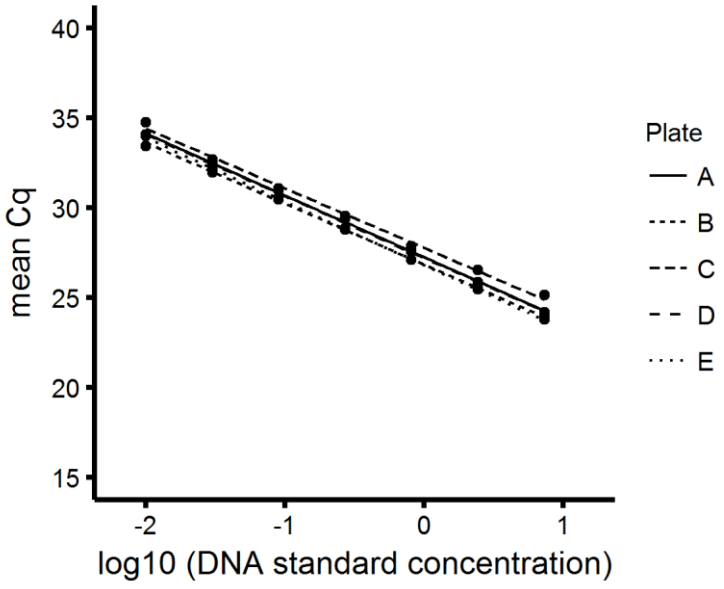


Figure 6.3-1 *ACTB* cfDNA real-time PCR quantification assay performance
Standard curves for all five cfDNA quantification plates are displayed. Final qPCR DNA concentrations ranged from 0.01 ng/µl to 7.29 ng/µl corresponding to total cfDNA yields of 1 – 729 ng.

6.3.2 Plasma samples yielded very low quantities of cfDNA

The absolute yields and concentrations of cfDNA recovered from subject plasma samples were lower than expected. *ACTB* qPCR analysis revealed that 73/180 samples demonstrated concentration values below the lowest PBMC DNA concentration standard. This means that these 73 samples yielded less than 1 ng amplifiable cfDNA overall. Nevertheless, these samples did provide clear amplification signals indicating the presence of amplifiable DNA, albeit at low levels. Yields and concentrations were calculated for these samples but should be regarded as somewhat tentative due to being outside the range of the standard curve. Extensive summary statistics of absolute cfDNA yields and concentrations (ng cfDNA/ml plasma) are shown in **Table 6.3-1**. The overall median yield from 1.8 ml plasma was 1.3 ng.

Table 6.3-1 Summary statistics of plasma cfDNA yields and concentrations for all samples and separated by disease status

Sample summary statistics (ng or ng/ml)	**Yield (ng)**	**Concentration (ng/ml plasma)**
	All samples (N = 180)	**All samples (N = 180)**
< 1	73 (40.6)	121 (67.2)
1 – 5	100 (55.6)	57 (31.7)
5 – 10	5 (2.8)	0 (0)
> 10	2 (1.1)	2 (1.1)
Mean ± SD	2.3 ± 7.77	1.3 ± 4.32
Median ± IQR	1.3 ± 1.45	0.7 ± 0.80
Range	0 – 88.98	0 – 49.4
Samples separated by status		
	Case (n = 60)	**Control (n = 120)**	**Case (n = 60)**	**Control (n = 120)**
< 1	17 (28.3)	56 (46.7)	34 (56.7)	87 (72.5)
1 – 5	37 (61.7)	63 (52.5)	24 (40.0)	33 (27.5)
5 – 10	4 (6.7)	1 (0.8)	0 (0)	0 (0)
> 10	2 (3.3)	0 (0)	2 (3.3)	0 (0)
Mean ± SD	4.3 ± 13.32	1.3 ± 1.08	2.4 ± 7.35	0.7 ± 0.60
Median ± IQR	1.4 ± 2.14	1.2 ± 1.43	0.8 ± 1.19	0.65 ± 0.79
Range	0.01 – 88.98	0 – 6.97	0.01 – 49.43	0 – 3.87

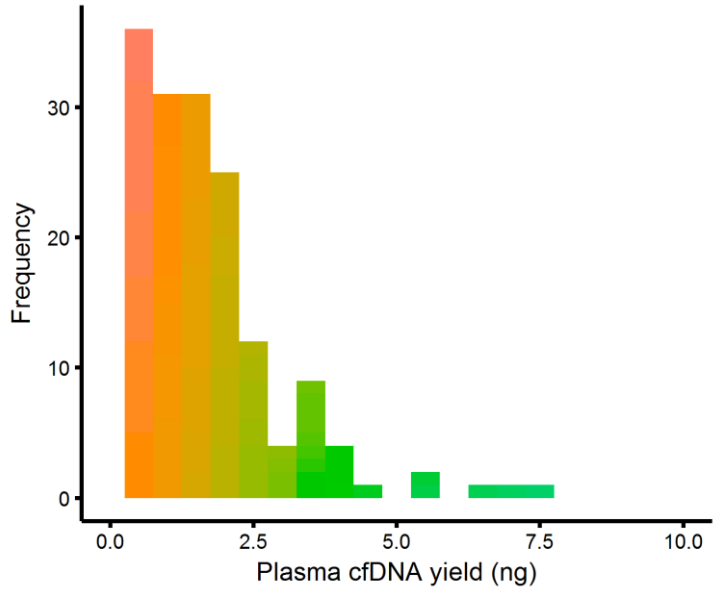


Figure 6.3-2 Distribution of plasma cfDNA yield for ddPCR methylation study
Histogram showing frequency of cfDNA yield based on concentration measured using *ACTB* cfDNA quantification assay (N = 178). Note that the x-axis is limited at a yield of 10 ng: two samples yielded higher cfDNA quantities (56.1 ng and 89.0 ng) but the figure axis was limited for better visualisation of the yield distribution

The distribution of DNA extraction yield for the 180 subject retrospective case-control blood plasma set is depicted in **Figure 6.3-2**. Inspection of this figure and formal testing clearly indicates that the data are not normally distributed ($P < 0.0001$, Shapiro-Wilk test). Boxplots of case and control yields do not display any obvious differences between the two groups, the clearest distinction being two very high yield lung cancer case outliers (**Figure 6.3-3, left panel**). Transformation of the data also revealed low yield outliers in the control group (**Figure 6.3-3, right panel**).

The case median yield (1.4 ng (IQR = 2.1)) was higher than that for the control group (1.2 ng (IQR = 1.4)) and the distributions of the two groups differed significantly (Mann–Whitney $U = 4522$, $n_1 = 60$, $n_2 = 120$, $P = 0.005$ two-tailed).

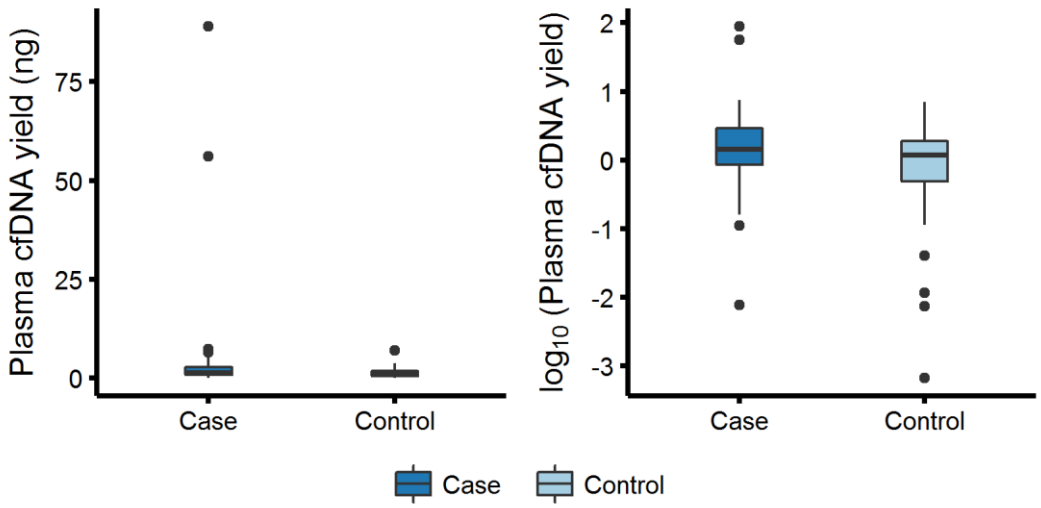


Figure 6.3-3 Boxplots of case and control yields
The **left hand panel** indicates measured cfDNA yields obtained using the ACTB cfDNA quantification assay. The **right hand panel** shows the log transformation of the same data for better interrogation of the underlying distributions.

More than two-thirds of cases and controls combined presented with concentrations of less than 1 ng cfDNA/ml plasma. Only two case samples yielded greater than 10 ng cfDNA overall (**Table 6.3-1**).

Yields were not significantly different between plasma samples stored for greater than or less than five years prior to analysis (P = 0.997, Mann-Whitney U test). No correlation was observed between cfDNA yield and storage time prior to analysis ($r_s(178) = -0.01$, P = 0.89, Spearman correlation).

There was no significant difference between the distributions of cfDNA concentration for groups with different pathological pT stage (H = 5.021, df = 3, P = 0.170, Kruskal-Wallis test), pathological pN stage (H = 2.749, df = 2, P = 0.253, Kruskal-Wallis test) or clinical stage (H = 6.028, df = 3, P = 0.110, Kruskal-Wallis test).

6.3.3 Initial screening of MethPlex ddPCR plasma samples

MethPlex ddPCR analysis of lung cancer case and control plasma samples for methylation of the *CDKN2A* promoter indicated very low frequency of methylation: only three out 60 cases (5%) were methylation positive. Methylation positive controls gave positive results and spike-in methylated pUC19 DNA, functioning as a process control, also produced positive, but variable, amplification signals (median ± IQR = 326.0 ± 354.2). The expectation was for a far higher proportion of methylation positive results from clinical samples. Bronchoalveolar lavage methylation screening (Error! Reference source not found.) and the earlier Nikolaidis tudy (Nikolaidis et al., 2012) displayed 16.2% and 25% *CDKN2A* methylation detection rates, respectively. Furthermore, the proportion of lung cancer patient plasma samples displaying promoter methylation of the same gene in published studies was greater than 20% (Belinsky et al., 2005; Zhang et al., 2011).

6.3.4 Increased ddPCR reaction inputs provided for increased methylation detection but DNA methylation frequencies were disappointingly low

I decided to continue screening samples, but, in order to conserve potentially wasted resources, the number of controls tested in subsequent ddPCR analysis was halved, resulting in a screening set of 60 cases and 60 controls. The absolute quantity of ddPCR input was also increased fourfold by changing the purified MethPlex enrichment product dilution factor from eight to two. A repeated screen of samples for *CDKN2A* methylation showed increased signal in previously positive samples and produced a further two positive cases. No non-lung cancer controls displayed *CDKN2A* positivity.

Plasma screening set univariate analysis statistics for all markers are presented in **Table 6.3-2**. All markers showed low frequency of methylation detection in plasma samples and therefore displayed poor sensitivity. With the exception of *MT1G*, *RASSF1* and *ABCB1*, no methylation was detected in controls samples. AUC statistics were all below 0.60, also indicating that individual markers were poor classifiers of disease status.

Table 6.3-2 Univariate statistics for DNA methylation biomarkers in blood plasma samples

	Positives			Model-based classification[1]		
Gene promoter	**Case (n=60)**	**Control (n=60)**	**χ^2 *P* value**	**Se/Sp**	**Accuracy (%)**	**AUC (95%CI)**
WT1	7	0	0.013	11.7/100	55.8	0.56 (0.52, 0.60)
MT1G	15	9	0.254	25/85	55.0	0.55 (0.48, 0.62)
RASSF1	9	4	0.239	15/93.3	54.2	0.54 (0.49, 0.60)
CDKN2A	5	0	0.119	8.3/100	53.3	0.53 (0.50, 57)
TERT	2	0	0.496	3.3/100	51.7	0.52 (0.49, 0.51)
F2R	2	0	0.496	3.3/100	51.7	0.52 (0.49, 0.54)
SHOX2	1	0	1	1.7/100	50.8	0.51 (0.49, 0.52)
ABCB1	2	4	-	-	-	-

[1] Disease class prediction based on predicted $Pr(D) \geq 0.5$

6.3.5 Multiple marker combination models did not discriminate disease status adequately

A logistic regression model including predictors for the four markers with the highest frequency of methylation detection in lung cancer case plasma samples (*WT1*, *MT1G*, *RASSF1* and *CDKN2A*) was assessed. This model displayed a sensitivity and specificity of 46.7% and 70.0%, respectively. Diagnostic accuracy was 58.3% and the ROC AUC was 0.60. The addition of cfDNA concentration (ng

cfDNA/ml plasma), assessed by *ACTB* cfDNA qPCR quantification assay, increased specificity to 85.0%. The diagnostic accuracy also improved (65.8%), as did the overall ability of the test to discriminate between individuals with lung cancer and those without (AUC = 0.68). Note that these are still modest performance characteristics.

6.3.6 Positive methylation detection was conclusive when present

Although frequency of methylation detection was low across the screening set, samples that displayed a positive signal for DNA methylation did so unambiguously. A representative example for the *SHOX2* assay is shown in **Figure 6.3-4**. Case 59_C produced an unquestionably positive methylation signal when analysed by MethPlex ddPCR, with 758 positively identified ddPCR droplets, equivalent to 26.5 copies/µl ddPCR reaction concentration. Control 59_A and the PBMC DNA methylation negative control produced no methylated *SHOX2* FAM-positive droplets. Furthermore, DNA methylation controls also exhibited clear ddPCR positive signal and there was a distinctive decline in number of positive droplets as the methylation percentage reduced from 1% (ddPCR concentration: 187.0 copies/µl ddPCR) to 0.1% (ddPCR concentration: 8.8 copies/µl).

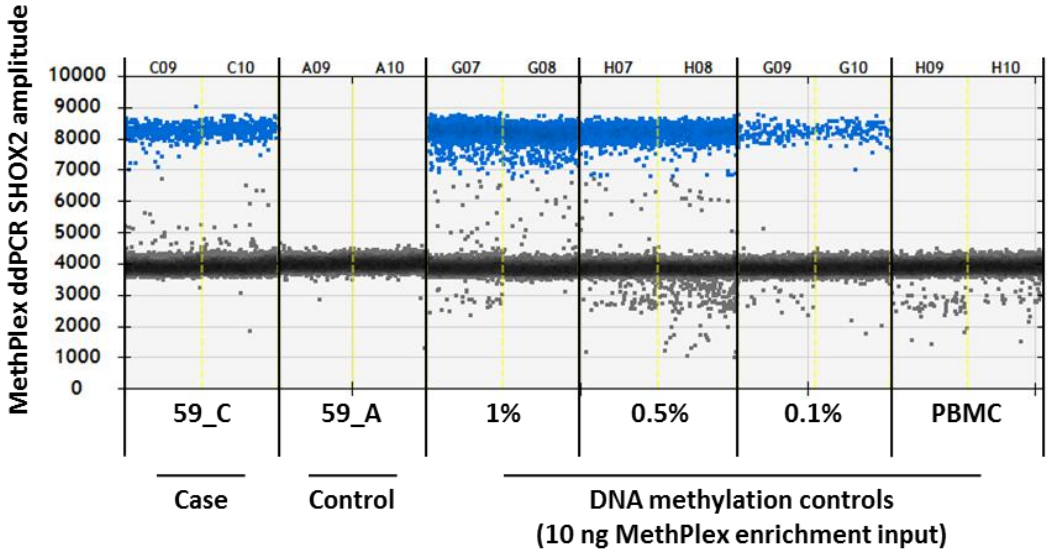


Figure 6.3-4 Representative modified QuantaSoft ddPCR 1D output for DNA methylation controls, and plasma cfDNA case and control samples
SHOX2 positive (FAM) positive droplets are indicated in blue, and negative droplets appear as dark grey. Each segment between solid vertical black lines contains data points from two wells (>32,500 individual ddPCR reactions).

6.3.7 Methylated DNA controls performed satisfactorily in plasma cfDNA screening

All methylated DNA positive controls in all runs displayed positive ddPCR detection, even at a Nominal MethPlex enrichment methylated DNA input of 3 GE. *CDKN2A*, *MT1G* and *SHOX2* assays also exhibited amplification in PBMC DNA wells. Assays had previously been tested with technical methylation negative WGA DNA controls, displaying no amplification, demonstrating their specificity. Signal present in PBMC DNA wells for the aforementioned assays is therefore most likely indicative of low levels of methylation in the untreated PBMC DNA control pool. The PBMC DNA measured concentration was consistently and significantly higher than that for 3 GE nominal methylated DNA MethPlex reaction input ($P = 0.029$, Mann-Whitney U test). This is symptomatic of the stochastic nature of amplification at very low levels.

All non-template controls did not produce any amplification signal. Plate-to-plate reproducibility was also assessed and overall inter-run coefficients of variation each DNA methylation assay were between 18.3% and 28.4%.

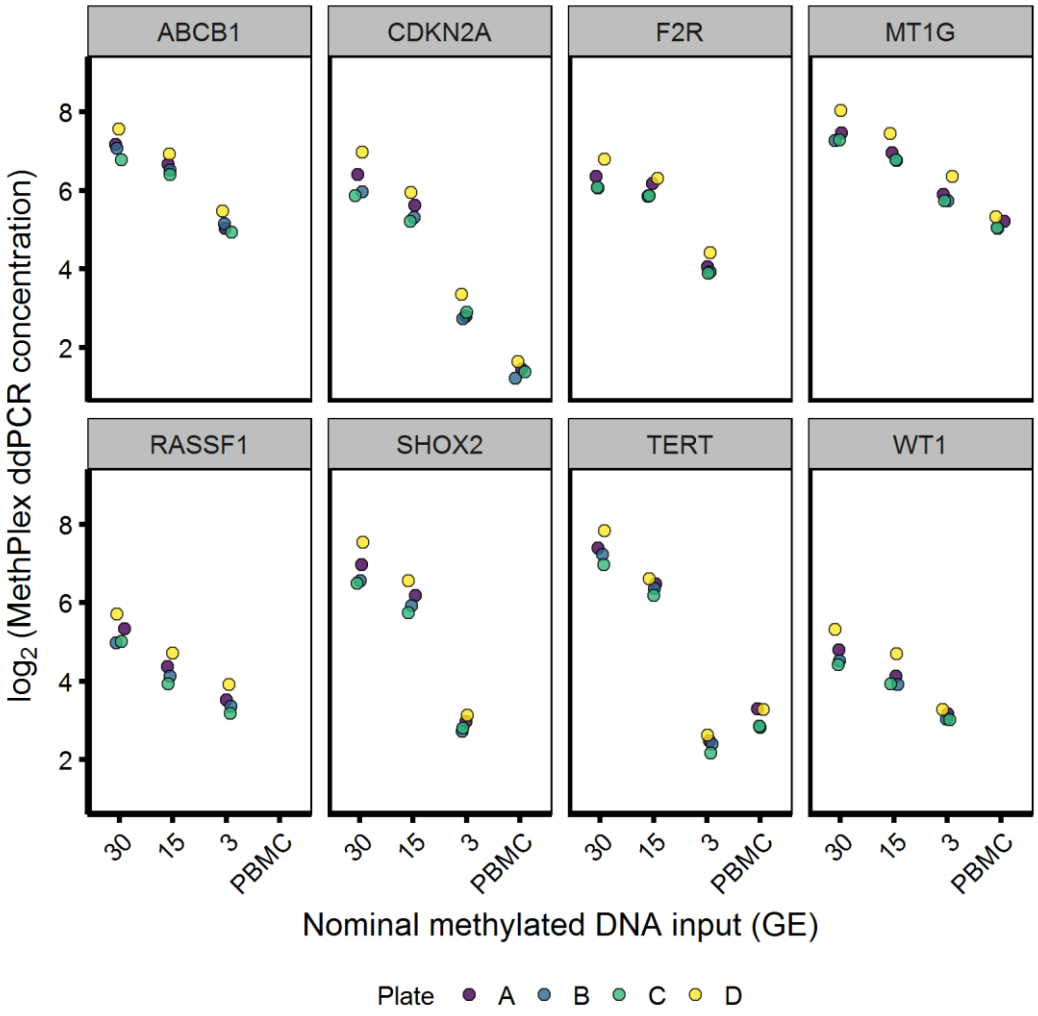


Figure 6.3-5 DNA methylation control performance in cfDNA plasma screening
MethPlex ddPCR concentrations were log_2 transformed for better visualisation. Plotted points are values are derived from the merging of duplicate wells as a single meta-well containing > 28,000 droplets. 30, 15 and 3 GE nominal methylated DNA input correspond to 1%, 0.5% and 0.1% methylated DNA proportions. Note that the x-axis is categorical and methylation controls were not equally spread therefore a linear trend is not necessarily expected to be observed.

6.4 Discussion

In this chapter, I have clearly demonstrated that MethPlex ddPCR analysis can detect DNA methylation in plasma samples. Frequencies of methylation detection for all markers in clinical specimens were lower than anticipated, resulting in poor discrimination between cases and controls as individual markers and combined as a discriminatory model. However, it should be noted that 27/60 cases gave a positive DNA methylation signal for one or more markers when assessed by MethPlex ddPCR. It is likely that modest marker sensitivity was dictated by the low cfDNA yields achieved in extraction of cfDNA from blood plasma. It has previously been demonstrated that DNA yields from plasma samples subject to long-term storage can decrease by an average of 30% per year (Sozzi et al., 2005). The use of samples that had been in frozen storage for at least three years was necessary in order to provide sufficient subject follow-up. In fact, 90% of samples had a storage duration of greater than four years. Furthermore, microcentrifuge tubes with untreated tube walls, as used in this study, have been shown to negatively influence PCR results at low concentration (Ellison et al., 2006; Teo et al., 2002), presumably due to adsorption of DNA molecules to untreated plastics.

In addition to the plasma DNA yield, the source of the extracted DNA and the proportion originating from malignant cells and/or the tumour microenvironment are also of vital importance (Thierry et al., 2016) . The structural properties of circulating DNA are still under investigation and the most frequently communicated circulating DNA fragment length resembles the size of mono- or oligo-nucleosomes (Fleischhacker and Schmidt, 2007; Holdenrieder et al., 2001; Jahr et al., 2001; Thierry et al., 2016). Experiments in colorectal cancer patients and animal xenograft models

somewhat elegantly demonstrated that ctDNA molecules in plasma are more fragmented than cfDNA molecules, being primarily less than 147 bp in length (Mouliere et al., 2013; Thierry et al., 2010). This was later confirmed independently by next-generation sequencing of hepatocellular carcinoma plasma samples (Jiang et al., 2015). DNA extraction kits assessed for use in this project were evaluated using high integrity, high molecular weight PBMC DNA spiked into plasma samples and as such did not assess the recovery of these short cfDNA fragments. Recent studies focussing on DNA methylation in plasma (Barault et al., 2015; Garrigou et al., 2016; Uehiro et al., 2016), or indeed ctDNA in general (Bettegowda et al., 2014; Oxnard et al., 2014; Sacher et al., 2016), used the QIAamp Circulating Nucleic Acid kit, which is presumed to specifically enrich for shorter DNA fragments, or other extraction methods developed specifically for the extraction of DNA from plasma (Kneip et al., 2011; Potter et al., 2014; Weiss et al., 2017). However, other studies have identified large sized DNA fragments, potentially originating from necrotic cells (Gormally et al., 2007; Jahr et al., 2001; Wang et al., 2003). It is possible, therefore, that the extraction technique used in this study failed to isolate the apparently more abundant small cfDNA fragments originating from apoptotic cells or from phagocytosis of hypoxia-induced necrotic tumour/tumour microenvironment cells by macrophages (Diehl et al., 2005), but retained cfDNA released from necrotic cells or actively secreted within exosomes (Kahlert et al., 2014; Thakur et al., 2014). The two cases with high cfDNA concentrations and the small proportion of plasma samples in which DNA methylation was detected could quite conceivably contain high molecular weight DNA of such origin.

The *ACTB* cfDNA quantification assay performed well and reproducibly presented excellent analytical characteristics, enabling reliable DNA quantification as low as 0.01 ng/µl (PCR reaction concentration) equivalent to a total yield of 1 ng amplifiable cfDNA. The strong analytical performance of DNA methylation controls was promising. The observation of amplification in all samples with 3 GE nominal methylated DNA MethPlex reaction input highlights the undoubted analytical sensitivity of the developed assays. This suggests that a lack of positive droplets in clinical sample wells was genuinely indicative of methylation negative samples, either because of absence of methylation of target gene promoters, insufficient ctDNA in patient circulation or inadequate extraction of cfDNA.

This part of the study had a number of weaknesses related to the nucleic acid isolation and the quality control process. Assessment of cfDNA extraction methods should have been carried out using DNA of appropriate length, obtained through sonication or enzymatic digestion, or more appropriately, isolated nucleosome-associated DNA. Clinical plasma with known high cfDNA yield would be the most relevant material to use for this purpose but is likely to be required for other studies. Secondly, the methylated pUC19 workflow control was 2686 bp in length; a control of shorter length would have been more relevant. It was also added directly to plasma samples prior to lysis. Degradation by DNases could account for the variability seen in ddPCR analysis of this control. The methylated pUC19 vector should have been added post-lysis or introduced in a protected form, for example, encapsulated in a protein coat (Brown and Pasloske, 2001). Assessment of methylated pUC19 recovery in a multiplex assay, along with *ACTB*, at the cfDNA quantification stage may also have been informative of the effective recovery in DNA extraction. Methylated DNA controls were

introduced to the workflow at the MethPlex enrichment stage. Additional analytical methylation controls should have been included in the workflow at the DNA extraction stage.

In conclusion, this chapter establishes that MethPlex enrichment methylation-specific ddPCR can detect methylated DNA in plasma samples, and this was more common in lung cancer cases than controls. However, due to a number of experimental limitations, primarily the choice of DNA extraction methodology, the methylation detection frequency was lower than anticipated and the data acquired did not perform well in the classification of lung cancer and controls, and could not adequately discriminate disease, either as single markers or as a panel. DNA methylation analytical controls were measured with reasonable precision, considering the exceptionally low reaction inputs, and were reliably detected. It is expected that the DNA methylation detection assays assessed in this study will perform well in future study if a more appropriate extraction technique is used.

Chapter 7

General discussion

7.1 Study justification

The clinical need for the early detection of lung cancer is unequivocal. Improved patient survival in the event of early detection has previously been demonstrated (McPhail et al., 2015). Data generated by the randomized US National Lung Cancer Screening Trial (NLST) established that low-dose computed tomography (LDCT) screening can reduce lung cancer mortality by 20% (National Lung Screening Trial Research Team, 2011). However, the increased survival comes at a cost of a high false-positive rate leading to patient complications caused by additional diagnostic procedures (Bach et al., 2012). The assistance of non-invasive or minimally invasive lung cancer biomarkers could potentially improve the specificity of CT screening. Furthermore, biomarker detection in body fluids may detect disease earlier than it can be identified using imaging. The present work in developing DNA methylation panels and assessing new technologies to facilitate the reliable detection of DNA methylation in plasma is therefore justified.

7.2 Cancer Research UK Diagnostic Biomarker Roadmap

As noted in chapter 4, a considerable number of academic cancer biomarker studies have been published yet only a small number of marker assays have been validated for clinical use. Many academic studies do not involve separate discovery and validation phases, are frequently underpowered and regularly lack appropriate assay optimisation (Liloglou and Field, 2010; Liloglou et al., 2014; Sandoval et al., 2013b). Cancer Research UK provides funding for biomarker assay development, validation and qualification to projects that have evident potential for clinical implementation under the Biomarker Projects Awards scheme. The strength of study design, including

statistical design and rationale, is one of the key Clinical Research Committee assessment criteria. This funding stream is intended to support appropriate biomarker studies only and result in projects that do not suffer from the above listed limitations, and therefore have an increased likelihood of providing fit-for-purpose biomarker assays that can be transferred to routine clinical use. A 'Diagnostic Biomarker Roadmap' is provided by CRUK as a supporting resource, describing a recommended model development process from initial biomarker discovery, assay development, retrospective comparison with current gold standard diagnostic practice and prospective clinical qualification (**Figure 7.2-1**). The work contributing to this thesis was executed in accordance with these guidelines, with the bronchoalveolar lavage screening corresponding to the BIDD (Biomarkers and Imaging Discovery and Development) BM (Biomarker) Discovery – Stage 2 ("Study the relationship between BM and gold standard diagnostic practice using clinical samples collected retrospectively"). Moderately invasive specimens such as BAL would not be fit-for-purpose in potential screening or diagnostic programmes yet methylation biomarkers displayed excellent diagnostic performance. This indicated the feasibility of lung cancer diagnosis using DNA methylation biomarkers in body fluids and the potential for utility of BAL methylation assays in assisting diagnosis. The biomarker assay development process proceeded, therefore, with the development of ddPCR assays, corresponding to BIDD Assay Development - Stage 1 ("Development of an accurate and reproducible assay to measure BM"). The screening of case-control plasma samples with MethPlex ddPCR assays was equivalent to BIDD BM Discovery – Stage 2 in the CRUK scheme **Figure 7.2-1**.

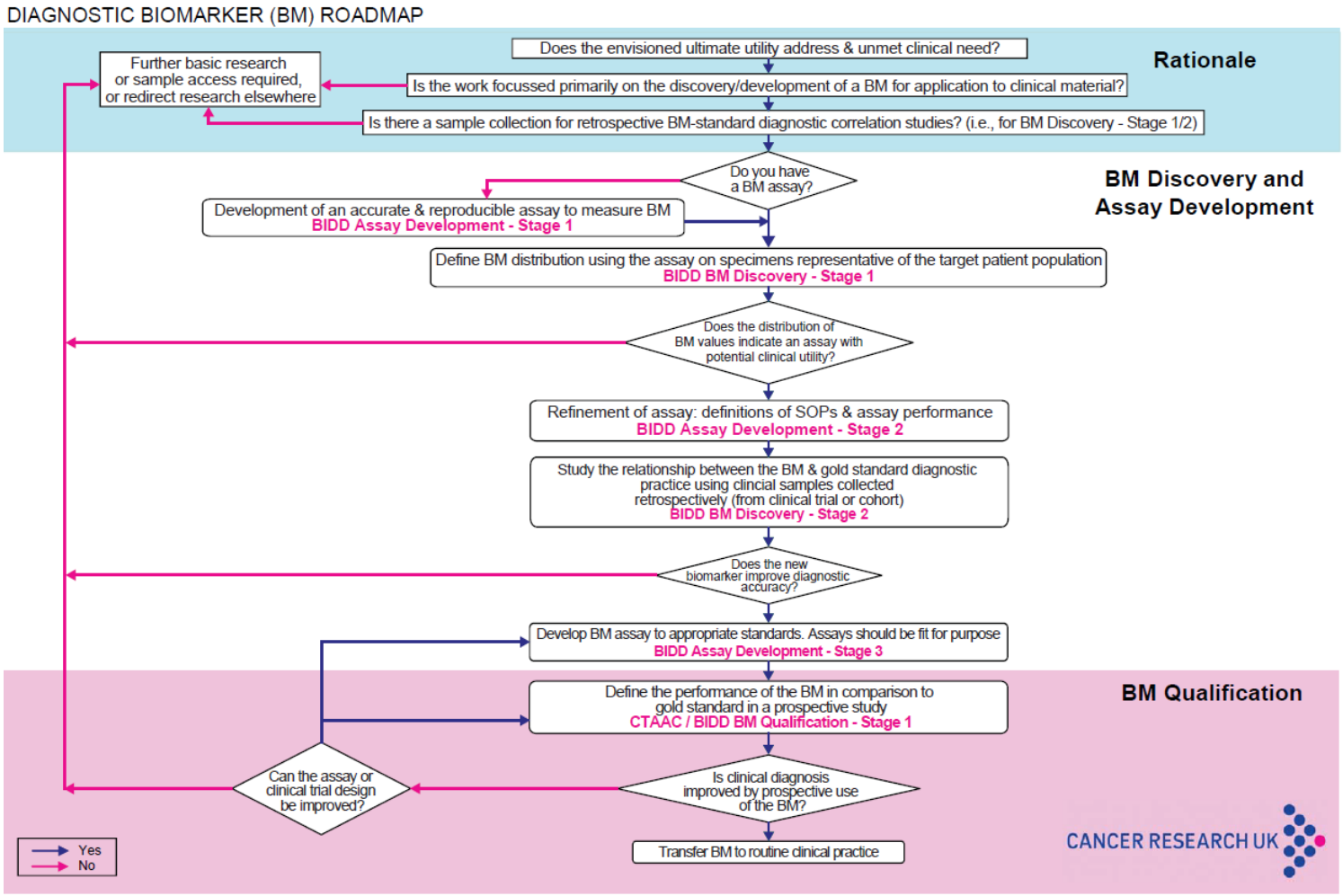


Figure 7.2-1 Cancer Research UK Diagnostic Biomarkers Roadmap

7.3 Point-of-care testing

The MethPlex ddPCR methodology that I have developed and described in this thesis displays high sensitivity in the detection of methylated DNA. However, this technique is complicated, time-consuming, labour-intensive and requires extensive technical expertise. Clinicians are better able to make timely and correct decisions regarding patient care when pertinent information is immediately available through point-of-care testing (POCT), potentially improving patient survival through early intervention (Louie et al., 2000). The development of automated real-time PCR systems for clinical molecular diagnostic detection of common cancer hot-spot mutations is encouraging (Janku et al., 2015, 2016). Furthermore, the extensive development of DNA methylation signature detection into integrated lab-on-a-chip devices may be desirable for routine clinical diagnosis (Kalofonou et al., 2012; Toumazou et al., 2013). Use of such compact, inexpensive devices would enable rapid diagnosis by individuals with minimal training, reducing public health costs and lessening patient anxiety.

According to the GLOBOCAN series of the International Agency for Research on Cancer 2012, c. 58% of all lung cancers occurred in developing countries (Ferlay et al., 2015; Wong et al., 2017). Lung cancer incidence and mortality is increasing in several low- and middle- income countries due to increased tobacco usage (Sankaranarayanan et al., 2011; Torre et al., 2016). The low cancer diagnosis rate in the African continent and presentation at advanced stage also leads to lower survival rate and increased mortality (Gaafar, 2017). In addition to any desired attempt to halt the tobacco epidemic, there is therefore an urgent need for better and earlier lung cancer diagnosis in these developing countries. Point-of-care tests have potential to address this demand in resource-limited settings where sophisticated instrumentation

may not be available. Furthermore, since the electricity supply may not be reliable or even present in some environments, and patient loss to follow-up may be of concern, rapid diagnostic result acquisition with low-power or self-powered devices may also be imperative (Rogers-Broadway and Karteris, 2015).

7.4 Future work

7.4.1 Revised cfDNA extraction methodology

The MethPlex ddPCR method developed and described in this thesis sensitively and specifically detected DNA methylation in technical analytical testing (**Chapter 5**), methylation positive controls run in parallel with clinical specimens (**Chapter 6**) and in a limited number of patient plasma samples (**Chapter 6**). Low input positive controls tested in plasma screening runs indicated that MethPlex ddPCR consistently detected 3 GE of methylated target DNA. It is therefore probable that methylated DNA was not present in cfDNA extracts where no signal was observed, either because the interrogated gene promoters were unmethylated or cfDNA extraction was inadequate. Low cfDNA yields measured by *ACTB* cfDNA quantification assay (**6.3.2**) add support to the latter interpretation. As indicated in **Chapter 6**, recent ctDNA studies (Bettegowda et al., 2014; Oxnard et al., 2014; Sacher et al., 2016) and plasma-based DNA methylation studies (Barault et al., 2015; Garrigou et al., 2016; Uehiro et al., 2016) used extraction methods presumed to specifically enrich for shorter DNA fragments present in the blood (Mouliere et al., 2011, 2013; Thierry et al., 2010, 2016). Recovery of cfDNA from plasma specimens with known high cfDNA yield or plasma samples spiked with short, fragmented DNA using the QIAamp Circulating Nucleic Acid kit should be assessed. Assuming satisfactory

DNA recovery, this kit should be used to extract cfDNA from clinical plasma specimens for a further MethPlex ddPCR retrospective case-control study.

7.4.2 DNA methylation biomarker discovery in plasma

The initial discovery phase(s) of potential DNA methylation biomarker identification prior to this study, and also in the previous Nikolaidis study (Nikolaidis et al., 2012), were implemented using lung cancer and normal lung tissue. Analysis of singular, localised biopsy samples may not correctly reveal the patient tumour DNA methylation profile because of intra-tumour heterogeneity (Gerlinger et al., 2012; McGranahan and Swanton, 2017). Liquid biopsy, by contrast, takes a blood-based ctDNA sample derived from diverse tumour regions with a profile likely representative of overall disease (De Mattos-Arruda et al., 2014; Jamal-Hanjani et al., 2016). Therefore, plasma could be a more informative test material when used directly during the initial discovery phase.

The utility of liquid biopsy and ctDNA analysis by next-generation sequencing as an alternative to traditional tissue biopsy in metastatic disease has been demonstrated (Lebofsky et al., 2015; Page et al., 2017; Shaw et al., 2017). High-throughput sequencing to identify multiple mutations in ctDNA of cancer patients and in longitudinally monitoring mutational profiles in high-burden disease has also been described (Forshew et al., 2012; Murtaza et al., 2013). Both targeted and genome-wide bisulphite sequencing approaches have recently been used to successfully map the tissue origin of plasma DNA through the identification of tissue-specific methylation patterns (Lehmann-Werman et al., 2016; Sun et al., 2015). Therefore,

high-throughput bisulphite sequencing presents a viable alternative approach for potential biomarker discovery.

Initial biomarker discovery prior to this study was also restricted to CpG islands within promoter regions of protein-encoding genes. Given the emerging importance of miRNA (Bediaga et al., 2013; Seol et al., 2014; Sozzi et al., 2014) and lncRNA (Ling et al., 2015; Schmitt and Chang, 2016) in cancer, and the differential methylation of regulatory enhancer and super-enhancer regions observed in the comparison of human cancer and normal cells (Heyn et al., 2016; Vidal et al., 2017), the interrogation of the wider non-coding genome may also be important.

Considering the above points, I propose that a future discovery phase should be implemented comprising whole-genome bisulphite sequencing of plasma samples. MethPlex enrichment and ddPCR assays targeting the most promising potential biomarkers should subsequently be developed, validated and used in patient plasma screening.

7.4.3 Exosomes as a potential methylated DNA pool

Exosomes are small (30 to 150 nm) extracellular vesicles (EVs) released by exocytosis and are of multivesicular endosomal origin (Kalluri, 2016; Reclusa et al., 2017), containing a diverse assortment of biomolecules including mRNA, lncRNA, miRNA, lipids and proteins, protected from degradation by a lipid bilayer (Reclusa et al., 2016; Vanni et al., 2017). Exosomal miRNAs have demonstrated diagnostic and prognostic utility in lung cancer (Cazzoli et al., 2013; Jin et al., 2017; Liu et al., 2017a). The presence of double-stranded DNA (dsDNA) in exosomes derived from cancer cell

lines and in pancreatic cancer patient serum has been described (Kahlert et al., 2014; Thakur et al., 2014). There is also limited evidence indicating that the degree of exosomal DNA methylation is comparable to that observed in gDNA (Thakur et al., 2014). Therefore, isolated exosomes may provide an alternative source of tumour-derived methylated DNA that should be investigated further and potentially screened by MethPlex ddPCR.

7.5 Concluding remarks

Screening of the expanded methylation DNA panel in bronchoalveolar lavage specimens was successful, improving on the diagnostic accuracy of the Nikolaidis study (Nikolaidis et al., 2012). Although the newly identified markers were not selected in the optimal model, diagnostic algorithms including these novel biomarkers showed very strong performance characteristics. In effect, the final choice of markers within the screened DNA methylation panel seemed unimportant as the combinations of three or more markers attained very similar sensitivity and specificity in the validation set. The multiplexing of assays may have contributed to the improved discriminatory performance of DNA methylation markers, indicated by improved attributes for the GN model used in the previous study (Nikolaidis et al., 2012).

BAL specimens are obtained during invasive bronchoscopy on suspicion of lung cancer. The application of minimally invasive blood-based testing for lung cancer diagnosis and screening would be preferable. However, a large proportion of patients with early stage cancer have been observed to have low abundance or undetectable levels of ctDNA (Bettegowda et al., 2014; Cohen et al., 2017; Wang et al., 2015). ddPCR has displayed highly precise quantification of nucleic acids (Pinheiro et al.,

2012) and presented greater precision compared to qPCR (Hindson et al., 2013; Sedlak et al., 2014; Strain et al., 2013). The next objective, therefore, was to develop and optimise ddPCR-based DNA methylation assays and associated workflows for utilisation with blood plasma samples, leveraging the potentially improved analytical performance of this platform. Eight DNA methylation assays were developed and extensively optimised. A robust cfDNA qPCR quantification assay was also established and revealed prohibitively low total DNA yields in patient plasma. This lead to the development of a targeted methylation-specific pre-amplification strategy, MethPlex enrichment, which was also meticulously optimised and used in conjunction with ddPCR analysis.

The next objective was to compare the newly developed ddPCR assays with legacy qMSP analysis of DNA methylation. ddPCR was revealed to be more precise and more sensitive than qMSP, exhibiting limits of detection of approximately 3 GE methylated DNA input when used in combination with MethPlex enrichment. ddPCR could also resolve twofold differences in methylated DNA input at moderately high (66ng) total DNA input and at low (10ng) total DNA inputs. As part of the process of comprehensively assessing assay performance, a framework for evaluating assays before use in pre-clinical studies was also developed.

The final objective was to screen the successfully developed MethPlex ddPCR assays in a retrospective lung cancer case-control plasma study. Diagnostic accuracy of markers, individually and combined as marker panels, was inadequate. This was probably as a result of very low cfDNA yields, potentially arising from diminished extraction of cfDNA of short fragment length (Underhill et al., 2016). Nevertheless,

DNA methylation assays clearly and unambiguously detected DNA methylation more frequently in lung cancer cases than controls. On a technical level, DNA methylation controls were reliably and consistently detected, even at 3 GE methylated DNA MethPlex reaction input. Despite the fact that the objective of validating these markers in plasma was not met, the technical performance of ddPCR DNA methylation assays was highly promising.

In conclusion, I have demonstrated that a three gene DNA methylation biomarker panel assessed by qMSP could assist in the diagnosis of lung cancer using bronchoalveolar lavage samples, in particular in cases occult to bronchoscopic diagnosis. I have determined that MethPlex ddPCR is a suitable method for the analysis of low abundance methylated DNA and displays greater sensitivity, higher precision, lower bias and better resolution than legacy qMSP assays. Diagnostic performance in minimally invasive plasma samples was severely limited by inadequate cfDNA yields. It is expected that biomarker assays will demonstrate improved performance if a more suitable cfDNA extraction technique is used in the future.

Chapter 8

Appendix

Blood plasma specimen SOP

Standard Operating Procedure Title:

SEPARATION OF WHOLE BLOOD INTO PLASMA AND LYMPHOCYTE SAMPLES

SOP Number:	LLP.21.02	**Effective Date:**
Version Number & Date:	V2 06/01/12	**Review Date:**

Superseded Version Number & Date (if applicable): V1. 09.01.2009,

Controlled Copy Number:

BACKGROUND

Blood is one type of biological specimen sample collected from Liverpool Lung Project (LLP) Hospital Cohort and Population Cohort participants at a number of clinics across Merseyside. At most, a total of 3 tubes of blood are collected from the participant and, once delivered to the Cancer Research Centre (CRC), individual blood samples are either frozen whole or processed within the laboratory to produce plasma and lymphocyte samples.

PURPOSE

To correctly separate whole blood in to plasma and lymphocyte samples.

SCOPE

1. WHO?

This procedure applies to any Laboratory Staff Member that performs this assay for the LLP.

2. WHEN?

This procedure should be applied when LLP blood samples collected from Population Cohort or Hospital Cohort participants are ready for processing at CRC.

3. HOW?

3.1. Reagents

3.1.1. Lysis buffer.

3.1.2. Haz Tab tablets – Haz Tab solution prepared as follows: Dissolve 1 tablet in approximately 2L of tap water.

3.1.3. Virkon.

3.1.4.

3.2. Equipment

3.2.1. 50ml Falcon Tubes.

3.2.2. 150mm Disposable Glass Pasteur Pipettes.

3.2.3. 3ml Disposable Plastic Pasteur Pipettes.

3.2.4. 2ml Microtube Vials.

3.2.5. 5ml Scintillation Vials.

3.2.6. Vacuum Pump.

3.2.7. Barcode Labels.

3.2.8. Microfuge.

3.2.9. Type II Flow Cabinet.

3.2.10. Centrifuge.

3.2.11. FTA classic cards.

3.2.12. Multi-Barrier pouches.

3.2.13. Desiccant tablets.

3.3. Method

Three tubes of blood is the amount collected from a Population Cohort participant attending a clinic for specimen collection. The following procedure is planned on the assumption of that being the number of blood samples available to work with. Adjustments are included for a total of 2 blood samples; if there is only 1 sample of whole blood it will be used for inoculation of the FTA card (step 3.3.4) and then stored in a -80^{o}C freezer in Freezer Room S16 without further processing.

To be carried out in the Type II Flow Cabinet in Laboratory F27 of CRC

FOR EACH SET OF PARTICIPANT BLOOD SAMPLES:

3.3.1. Each tube of blood is labelled by affixing a barcode label and also writing the relevant lab number on the side of the tube.

3.3.2. A 50ml Falcon tube is labelled by writing the relevant lab number on the lid and on the side of the tube and then placed in a rack.

3.3.3. 3x5ml vials for plasma storage and 3x2ml vials for lymphocyte storage are labelled by affixing barcode labels and also by writing the relevant lab number on the vial caps, and then placed in a rack.

3.3.4. Using a 3ml plastic Pasteur pipette and the required amount of one tube of blood, 4 circles of an FTA classic card (labelled with participant's name, D.O.B, barcode label and today's date) are inoculated and left to dry overnight in an incubator at room temperature. *At this and each subsequent step, when used, the Pasteur pipette is used once and then placed in virkon for 24hrs before disposal.*

3.3.5. Combined contents of the other 2 tubes of blood (or the remaining contents of each tube, if the total no. is only 2) are transferred to its labelled 50ml Falcon tube. The Falcon tube is then centrifuged for 10 minutes at 3000 rpm.

3.3.6. The Falcon tube is carefully removed from the centrifuge and returned to the Flow Cabinet in the rack.

3.3.7. Plasma (upper layer) is carefully transferred from the Falcon tube to the 3 pre-labelled 5ml vials using a 3ml plastic Pasteur pipette (approximately 2.5ml plasma per vial) and caps are then attached.

3.3.8. The contents of the Falcon tube is then topped up to 45ml with lysis buffer and gently mixed by inverting tube until sediments have dissolved.

3.3.9. The Falcon tube is then centrifuged for a further 10 minutes at 3000rpm.

3.3.10. The Falcon tube is carefully removed from the centrifuge and returned to the Flow Cabinet in rack.

3.3.11. Using a disposable glass Pasteur pipette and vacuum pump, supernatant is aspirated and discarded in to a 5L flask containing Haz tab solution, leaving a cell pellet at the bottom of the tube. The glass pipette is then placed in virkon for 24hrs before discarding.

3.3.12. The cell pellet is thoroughly resuspended in approximately 7.5ml of lysis buffer, using a 3ml plastic Pasteur pipette to gently mix.

3.3.13. The cell suspension is aliquoted equally in to the 3 pre-labelled 2ml vials using a 3ml plastic Pasteur pipette, the labelled caps are attached and the vials are spun in the microfuge for 2 minutes at 12000rpm.

3.3.14. The vials are removed from the microfuge and returned to the Flow Cabinet in the rack. Supernatant is poured off in to virkon, leaving the lymphocyte cell pellets at the

bottom of each vial, and caps are re-attached to the vials (the virkon solution is left for 24hrs before discarding).

3.3.15. Lymphocyte containing and plasma containing vials are stored in a -80°C freezer in Storage Room S03, and any remaining tube of whole blood is stored in a -80°C freezer in Freezer Room S16.

3.3.16. The following day the FTA card is placed with a desiccant tablet in a multi-barrier pouch, the pouch is then sealed and attached to the back of the Specimen Form, which is then stored in the box in Storage Room S03.

ROLES AND RESPONSIBILITIES

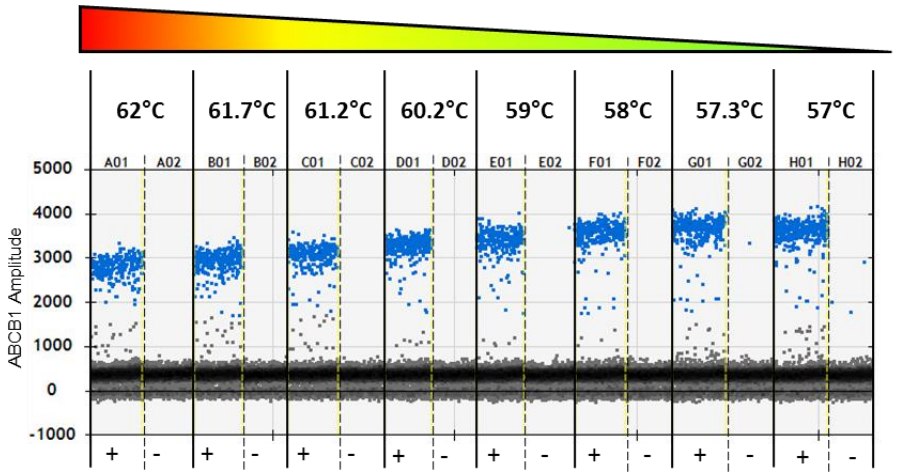
Laboratory Staff Member

- Correct labelling of vials and Falcon tube.
- Using the correct number of blood samples for the separation procedure, keeping one sample of whole blood for freezing if the total number is 3.

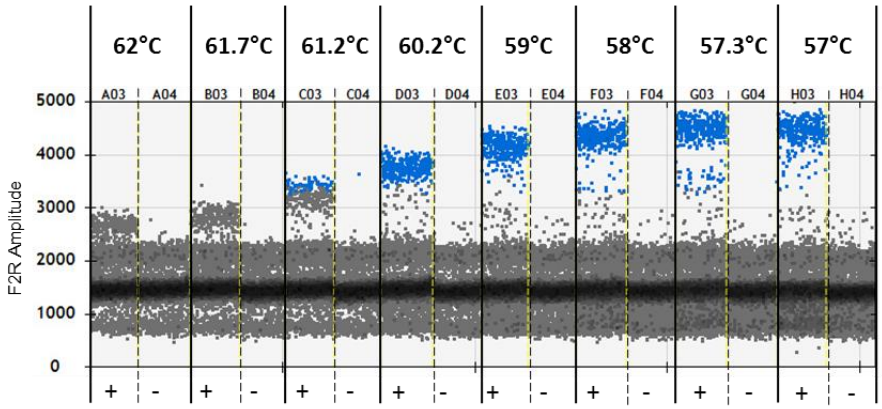
OTHER RELATED PROCEDURES

SOP LLP.19. 'Unpacking of LLP Blood, Sputum, Buccal & Bronchial Lavage Specimens with Distribution of Paperwork for Entry of Specimen Details in to LIMS' (Current Version)

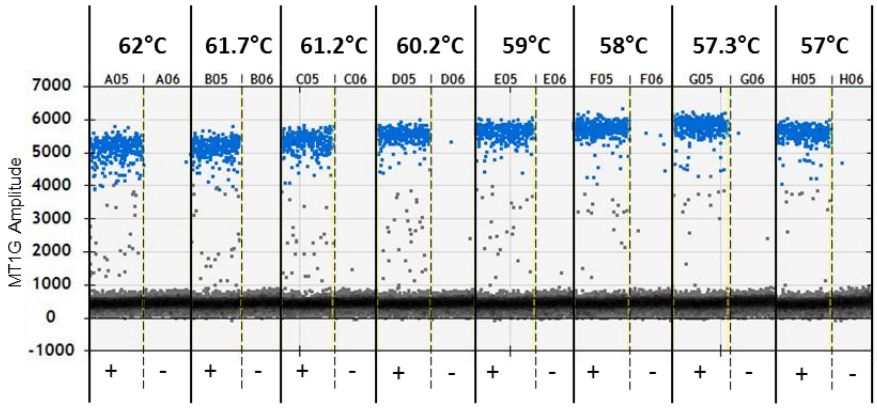
ABCB1



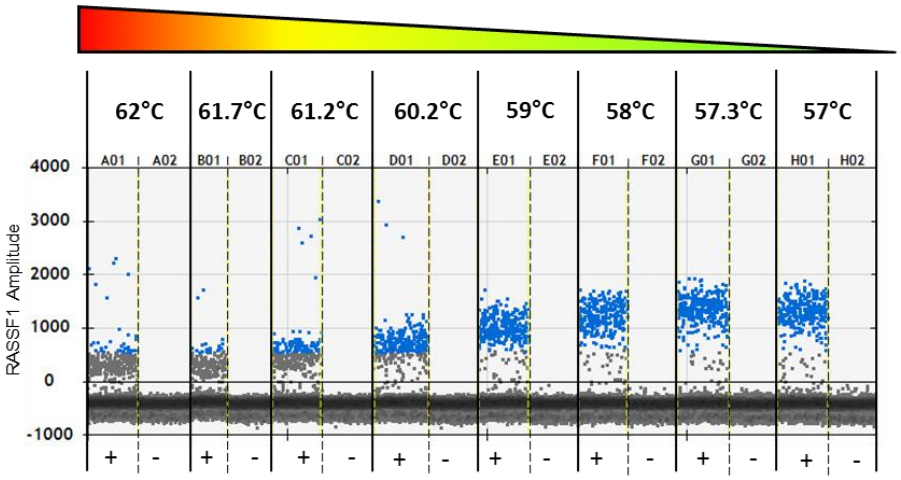
F2R



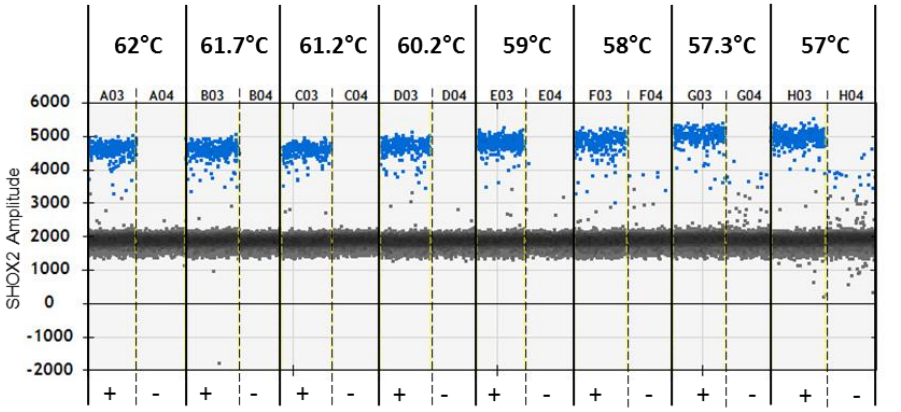
MT1G



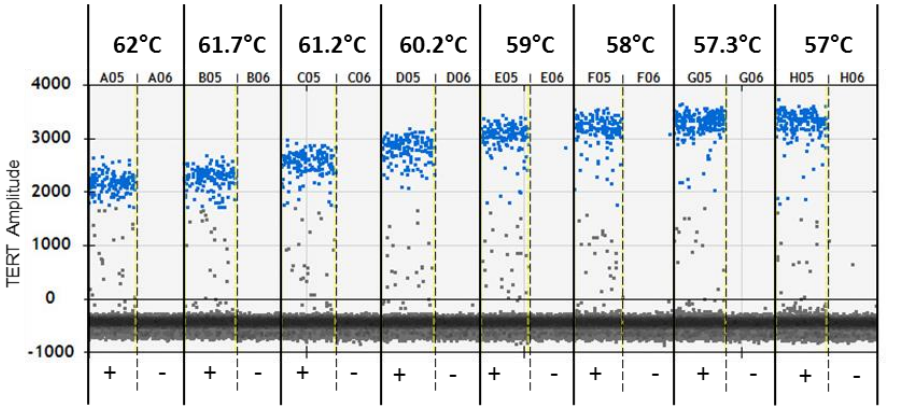
RASSF1



SHOX2



TERT



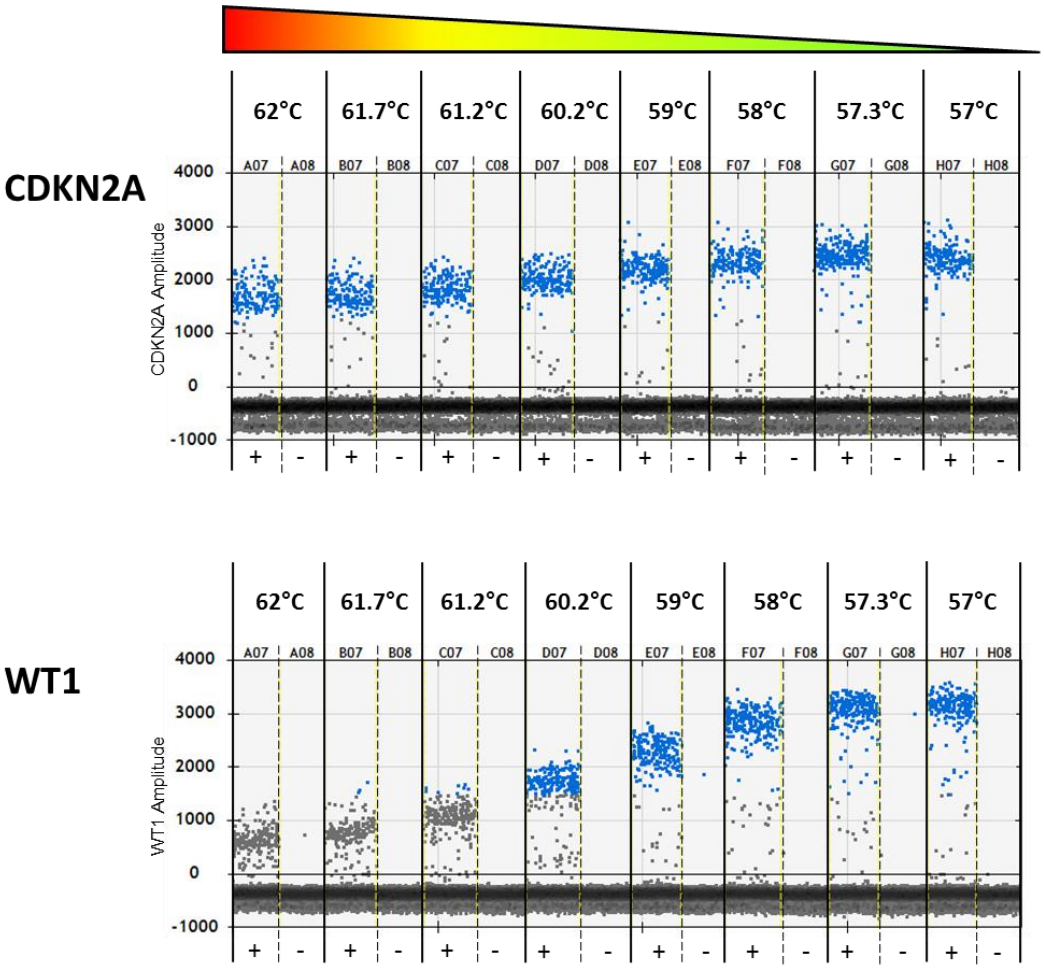
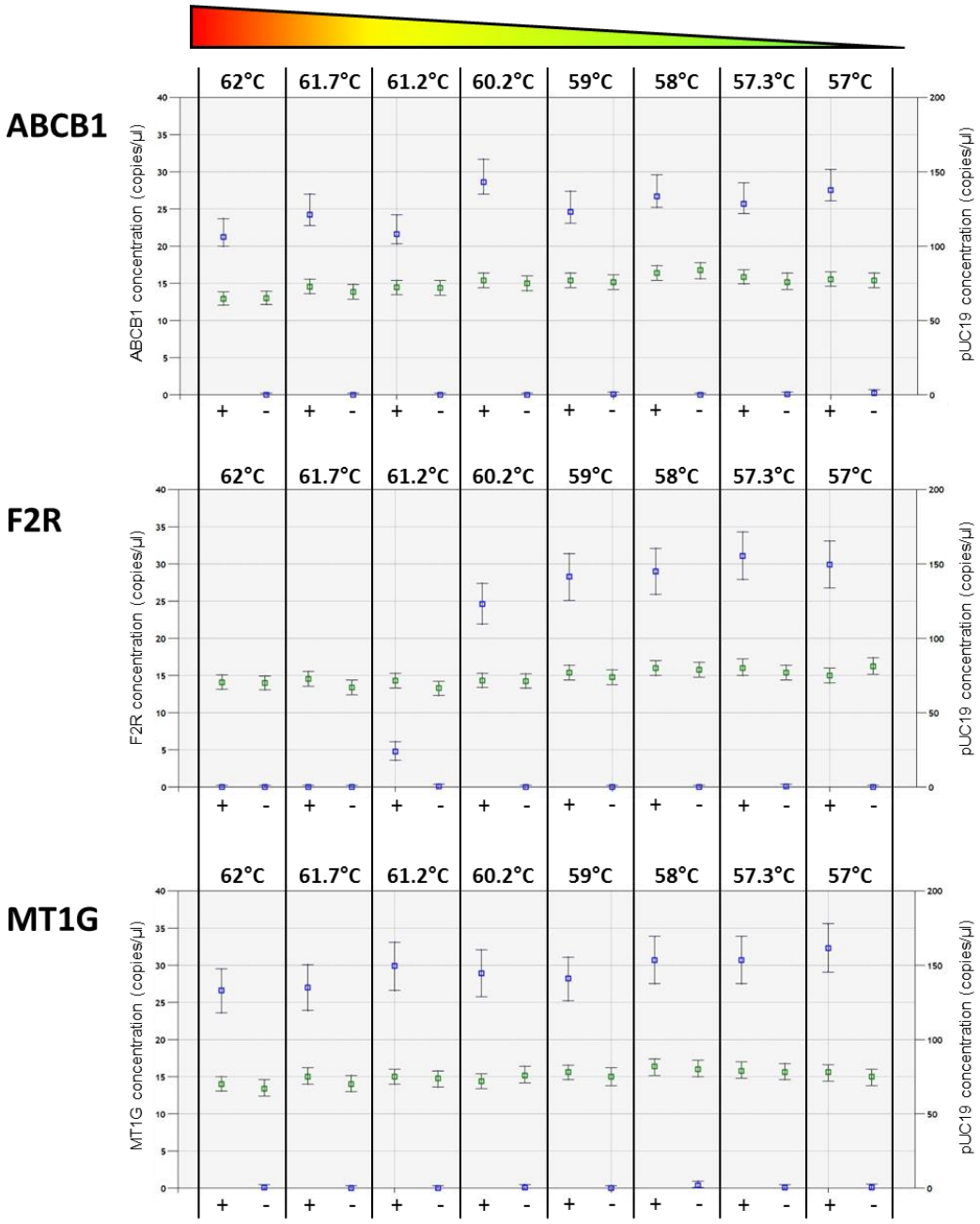
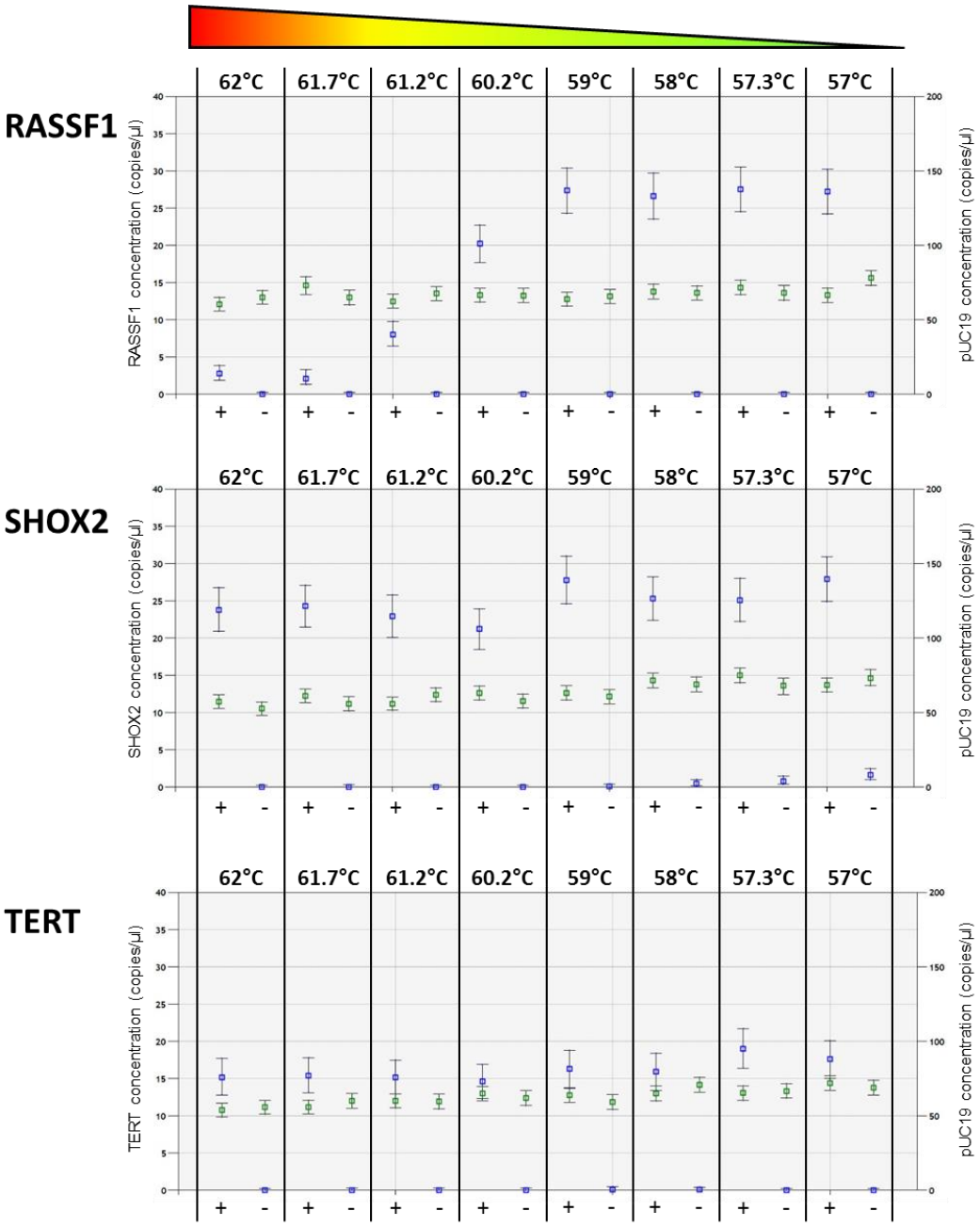


Figure 8.1 Modified QuantaSoft ddPCR 1D output for annealing temperature optimisation of all target assays in duplex with pUC19 control

ddPCR fluorescence amplitude and positive/negative droplet identification for all methylated DNA targets. Target positive (FAM) positive droplets are indicated in blue and negative droplets appear as dark grey. Each segment between solid vertical black lines contains data points for >10, 000 individual ddPCR partitions.


ABCB1
F2R
MT1G
62°C
61.7°C
61.2°C
60.2°C
59°C
58°C
57.3°C
57°C
ABCB1 concentration (copies/µl)
F2R concentration (copies/µl)
MT1G concentration (copies/µl)
pUC19 concentration (copies/µl)
+
-


RASSF1
SHOX2
TERT
62°C
61.7°C
61.2°C
60.2°C
59°C
58°C
57.3°C
57°C
RASSF1 concentration (copies/µl)
SHOX2 concentration (copies/µl)
TERT concentration (copies/µl)
pUC19 concentration (copies/µl)
+
-

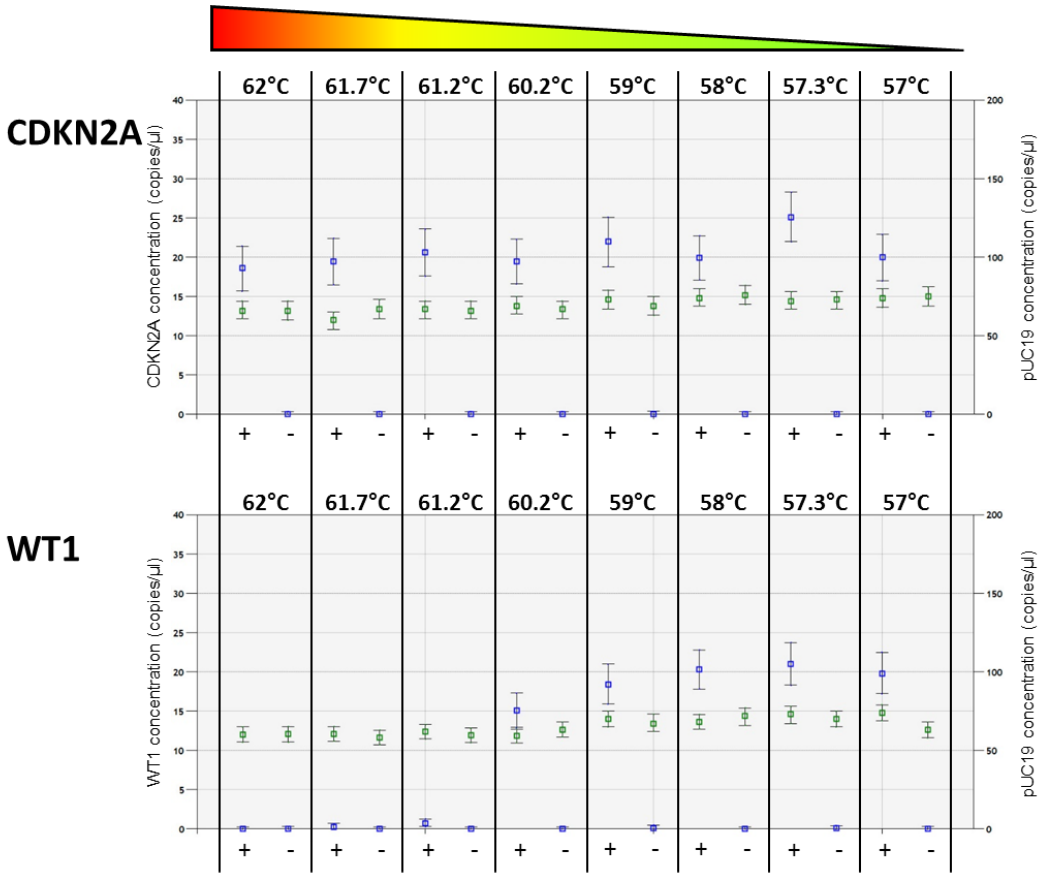
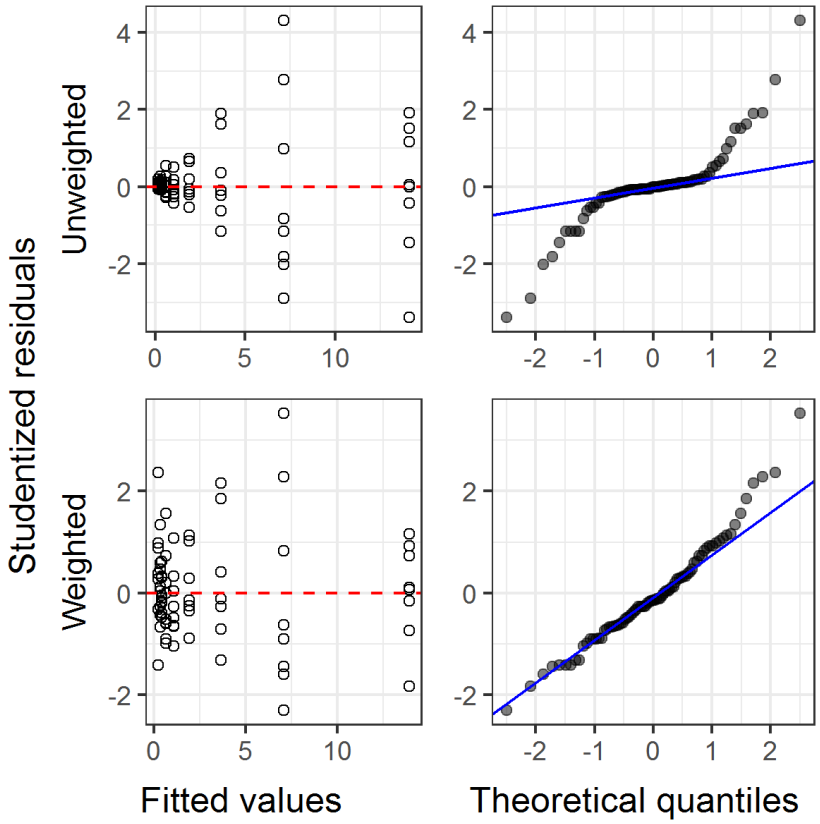


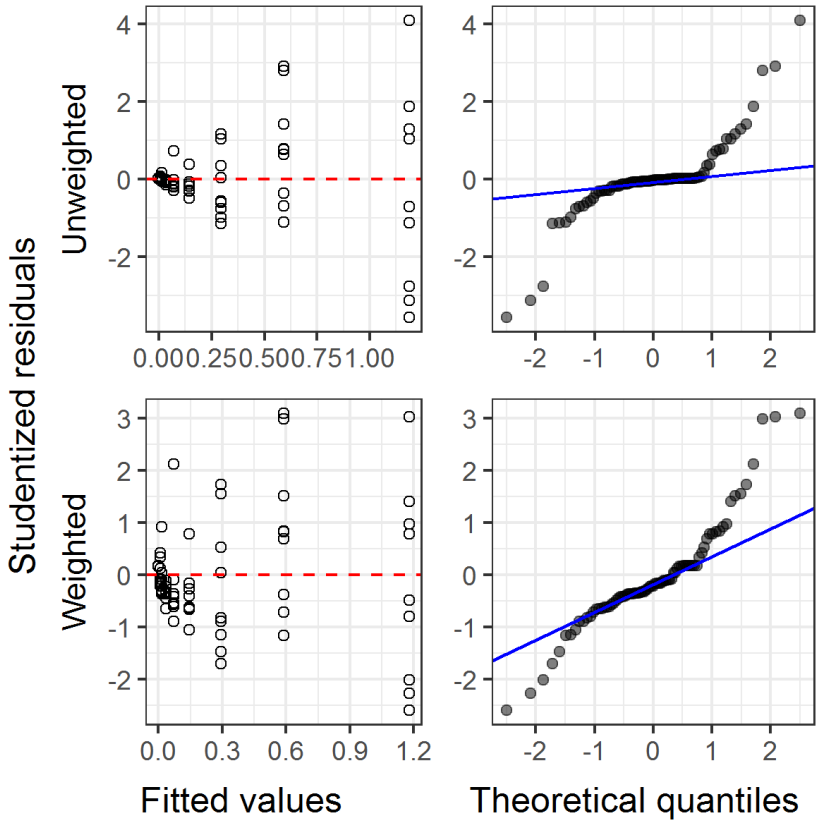
Figure 8.2 Modified QuantaSoft ddPCR concentration output for annealing temperature optimisation of all target assays in duplex with pUC19 control

QuantaSoft concentration output. Blue points correspond to methylated target DNA concentration, green points correspond to pUC19 concentration. Error bars represent 95% Poisson confidence interval.

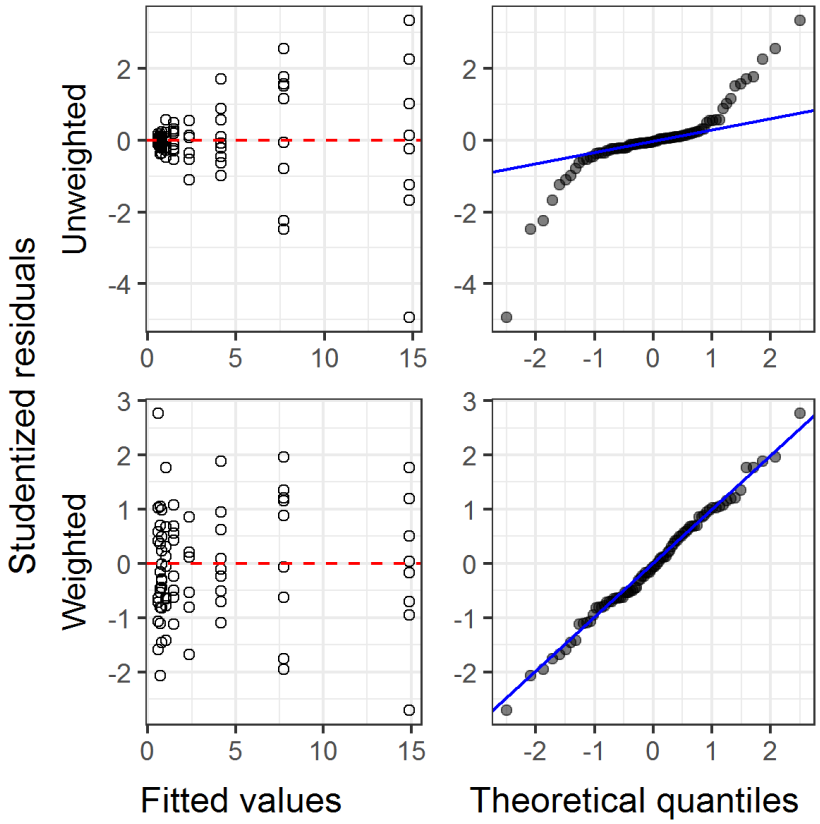
RASSF1 ddPCR



RASSF1 qMSP



WT1 ddPCR



WT1 qMSP

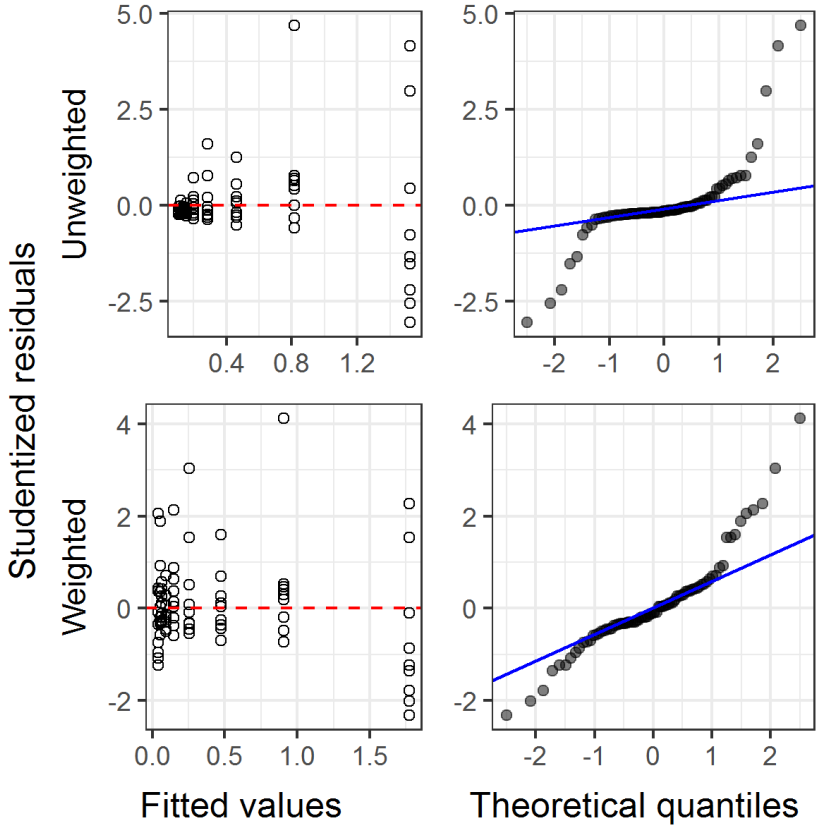


Figure 8.3 High total DNA input unweighted and weighted regression diagnostics

Studentized residuals are plotted against model fitted values in the left-hand panels and are indicative of scedasticity (constancy of variance). Distribution of residuals is represented by Q-Q plots in right-hand panels with points lying along the plotted line indicating a normal distribution. Upper panels show diagnostics for unweighted linear regression, while lower panels display those for weighted least squares regression.

Figure 8.4 High total DNA input RASSF1 ddPCR and qMSP linear regression diagnostic plots

Studentized residuals are plotted against model fitted values in the left-hand panels and are indicative of scedasticity. Distribution of residuals is represented by Q-Q plots in right-hand panels with points lying along the plotted line indicating a normal distribution. Upper panels show diagnostics for modelled ranges > 20 GE, while lower panels display those for ranges > LOQ.

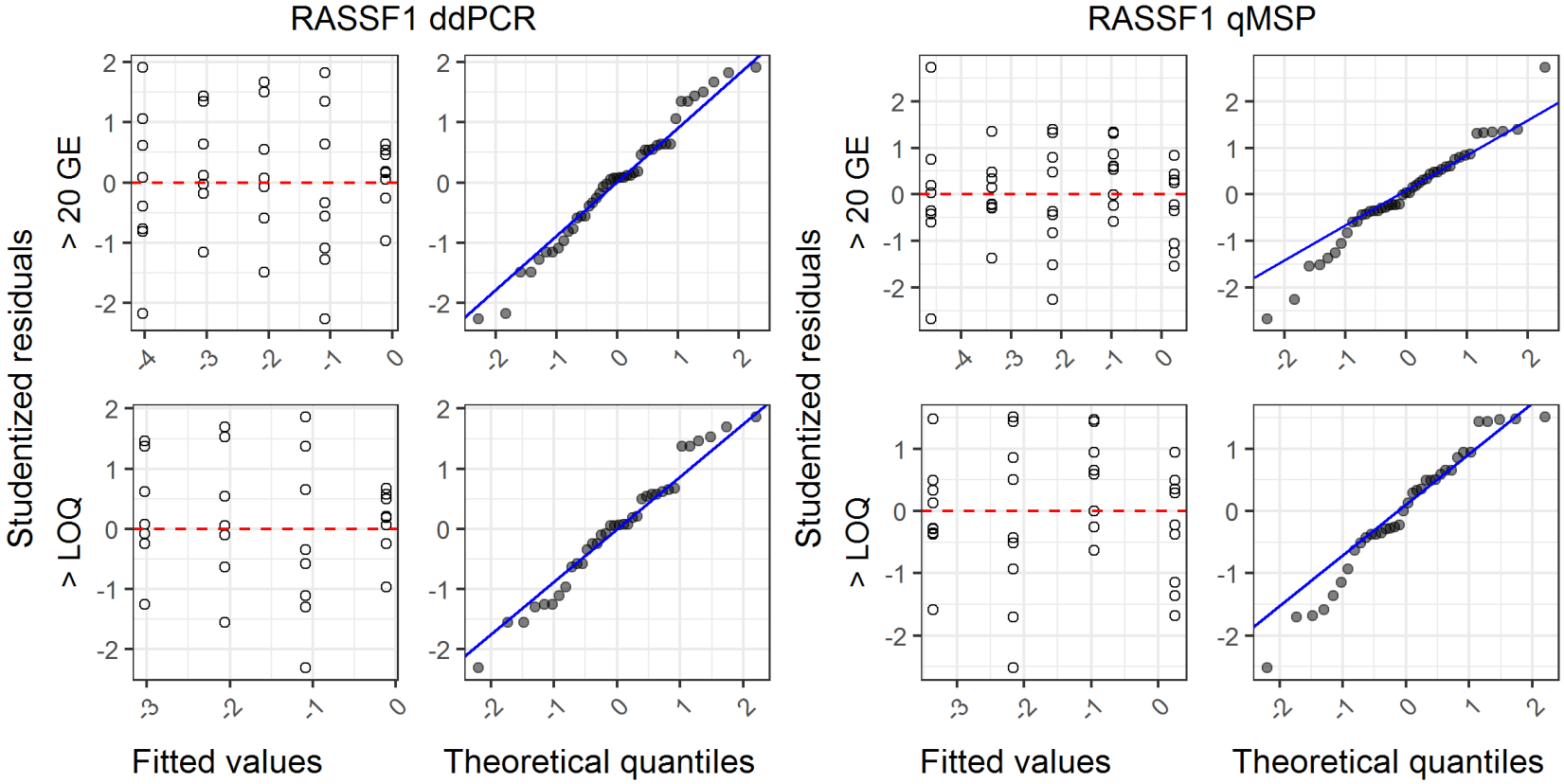


Figure 8.5 High total DNA input WT1 ddPCR and qMSP linear regression diagnostic plots

Studentized residuals are plotted against model fitted values in the left-hand panels and are indicative of scedasticity. Distribution of residuals is represented by Q-Q plots in right-hand panels with points lying along the plotted line indicating a normal distribution. Upper panels show diagnostics for modelled ranges > 20 GE, while lower panels display those for ranges > LOQ.

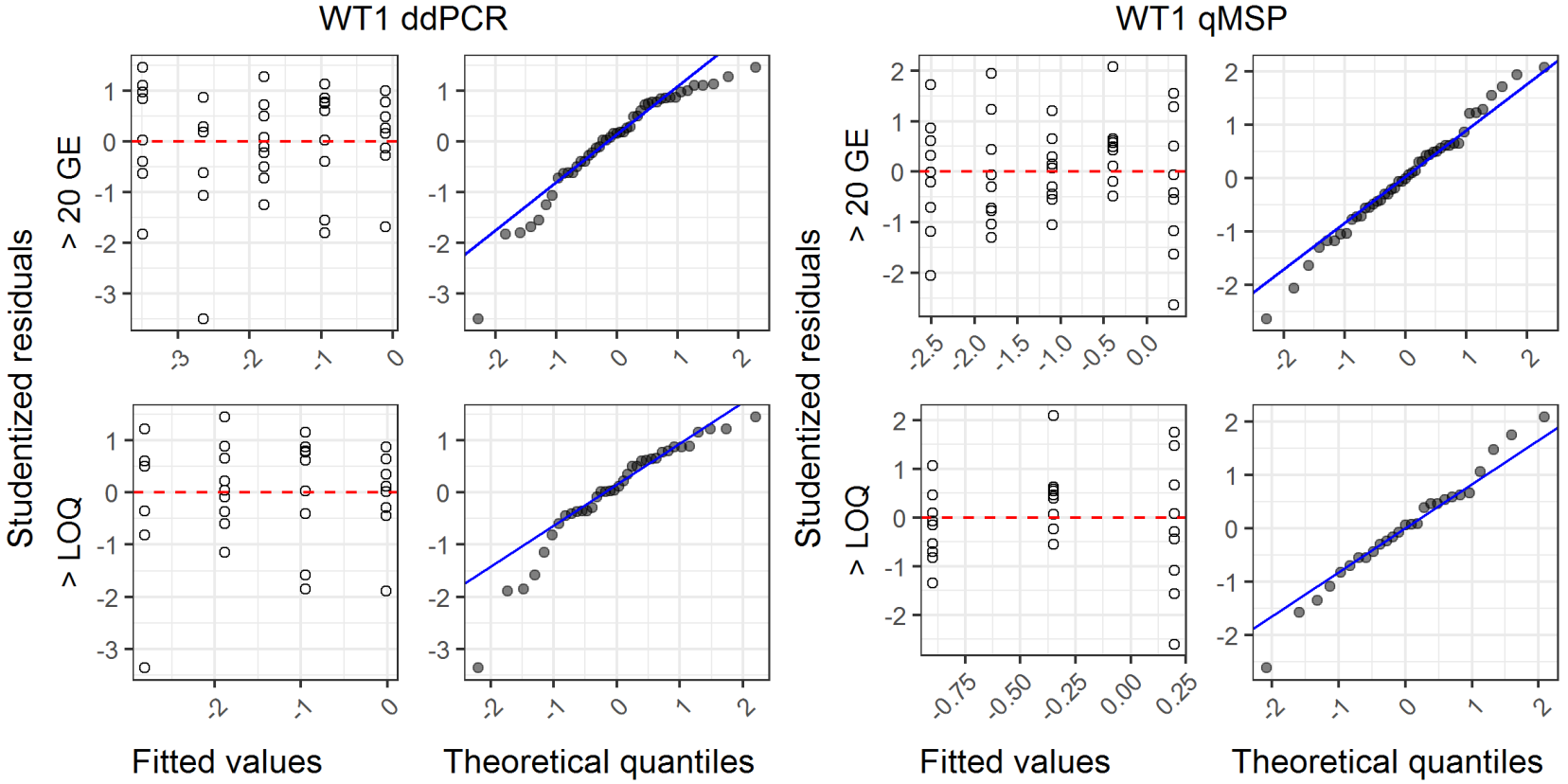
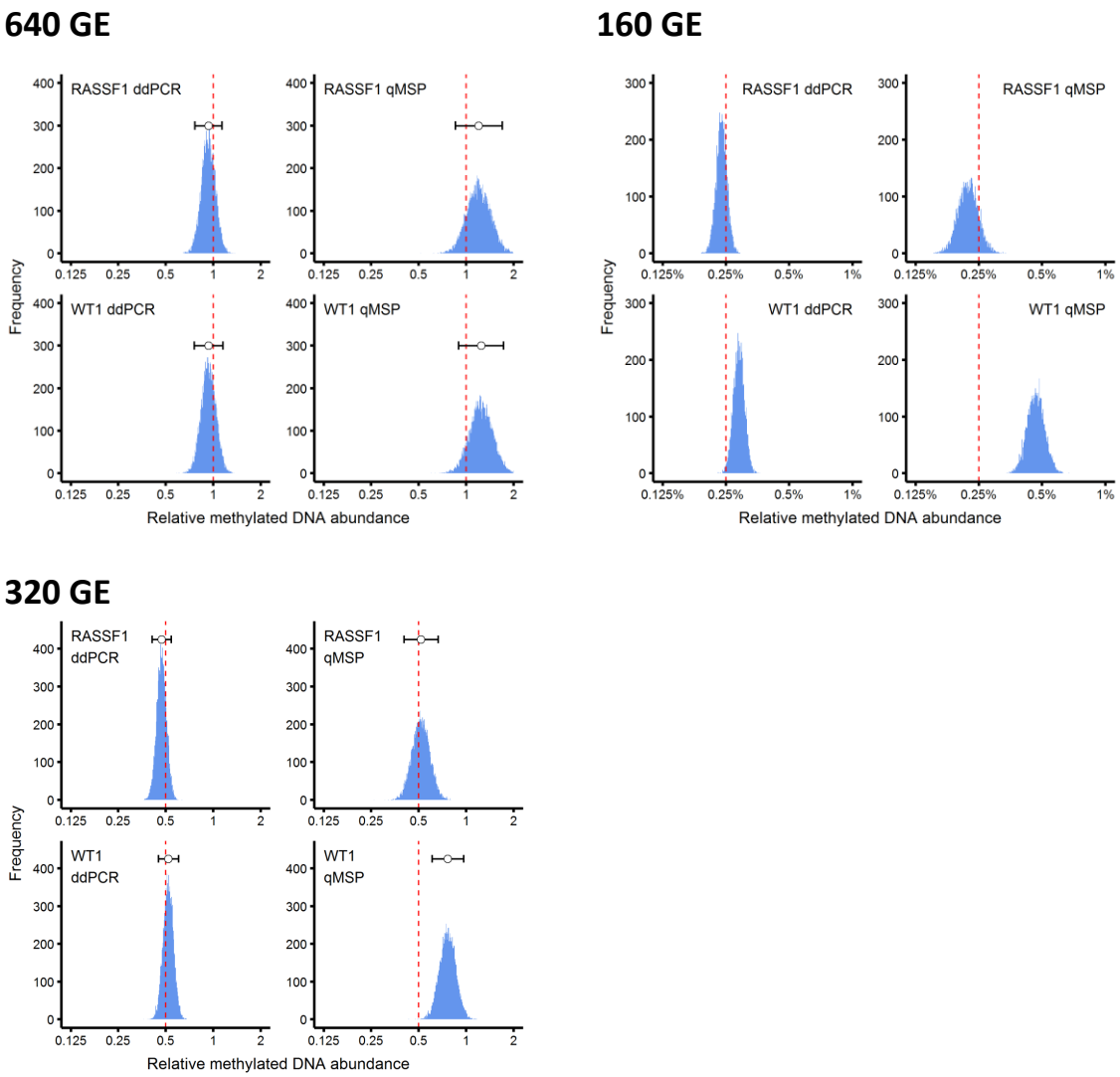


Figure 8.6 Representative histograms of bootstrap estimates for mixed model predictions at 640 GE, 320 GE, and 160 GE nominal methylated DNA input



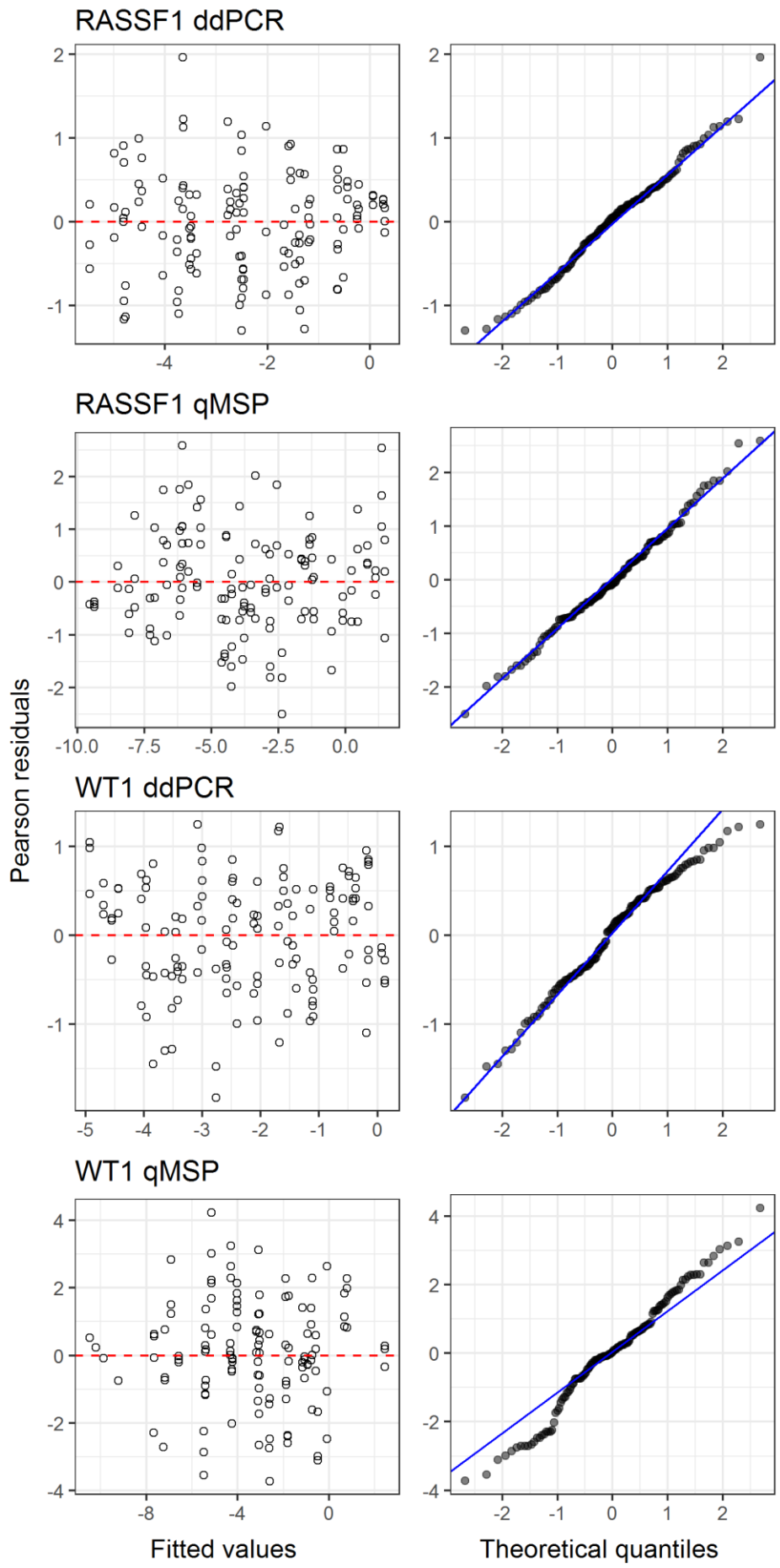
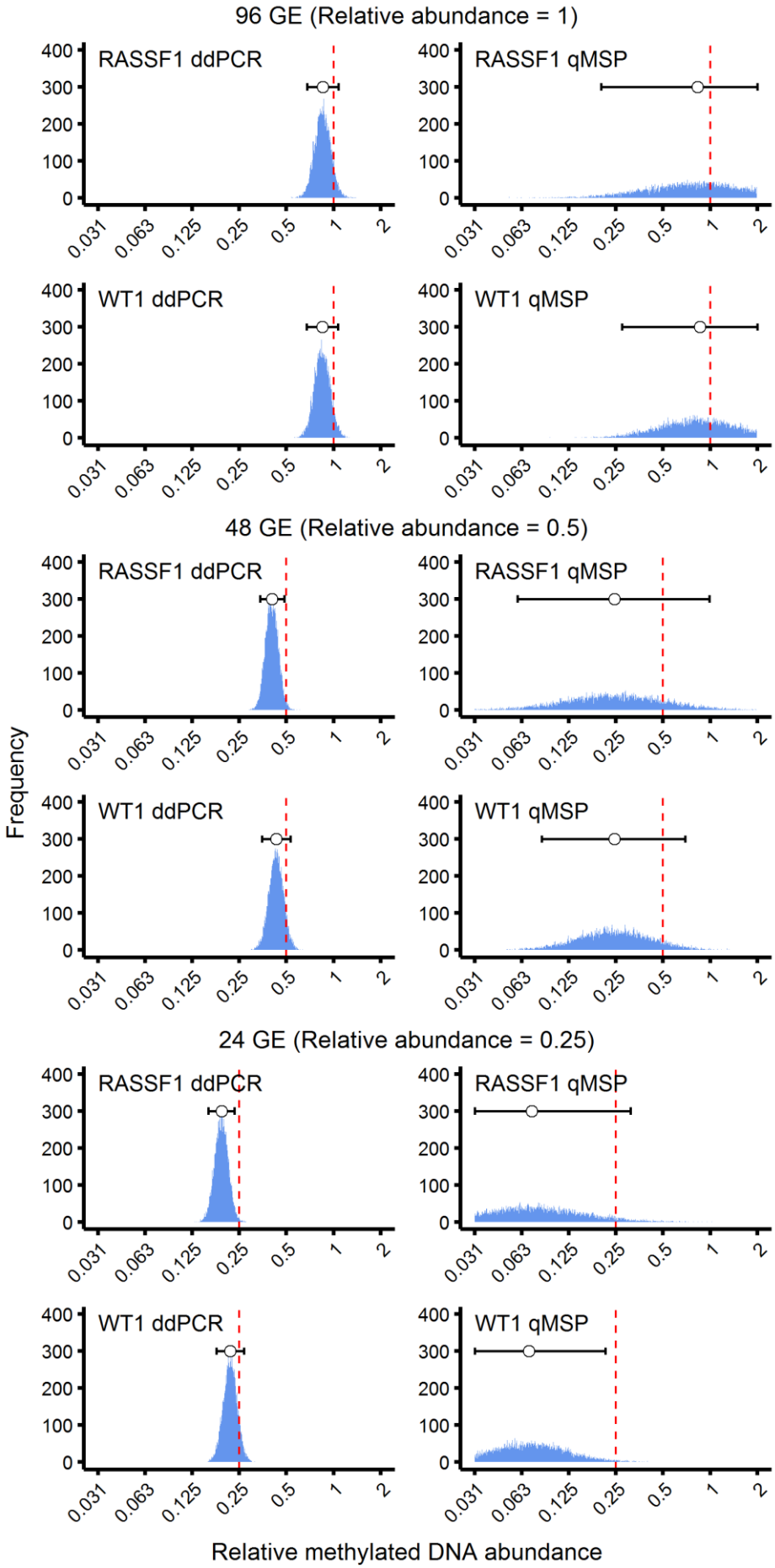

RASSF1 ddPCR
RASSF1 qMSP
WT1 ddPCR
WT1 qMSP
Pearson residuals
Fitted values
Theoretical quantiles

Figure 8.7 Low total DNA input bias assessment regression diagnostics

Studentized residuals are plotted against model fitted values in the left-hand panels and are indicative of scedasticity (constancy of variance). Distribution of residuals is represented by Q-Q plots in right-hand panels with points lying along the plotted line indicating a normal distribution.


96 GE (Relative abundance = 1)
RASSF1 ddPCR
RASSF1 qMSP
WT1 ddPCR
WT1 qMSP
48 GE (Relative abundance = 0.5)
RASSF1 ddPCR
RASSF1 qMSP
WT1 ddPCR
WT1 qMSP
24 GE (Relative abundance = 0.25)
RASSF1 ddPCR
RASSF1 qMSP
WT1 ddPCR
WT1 qMSP
Frequency
0
100
200
300
400
0.031
0.063
0.125
0.25
0.5
1
2
Relative methylated DNA abundance

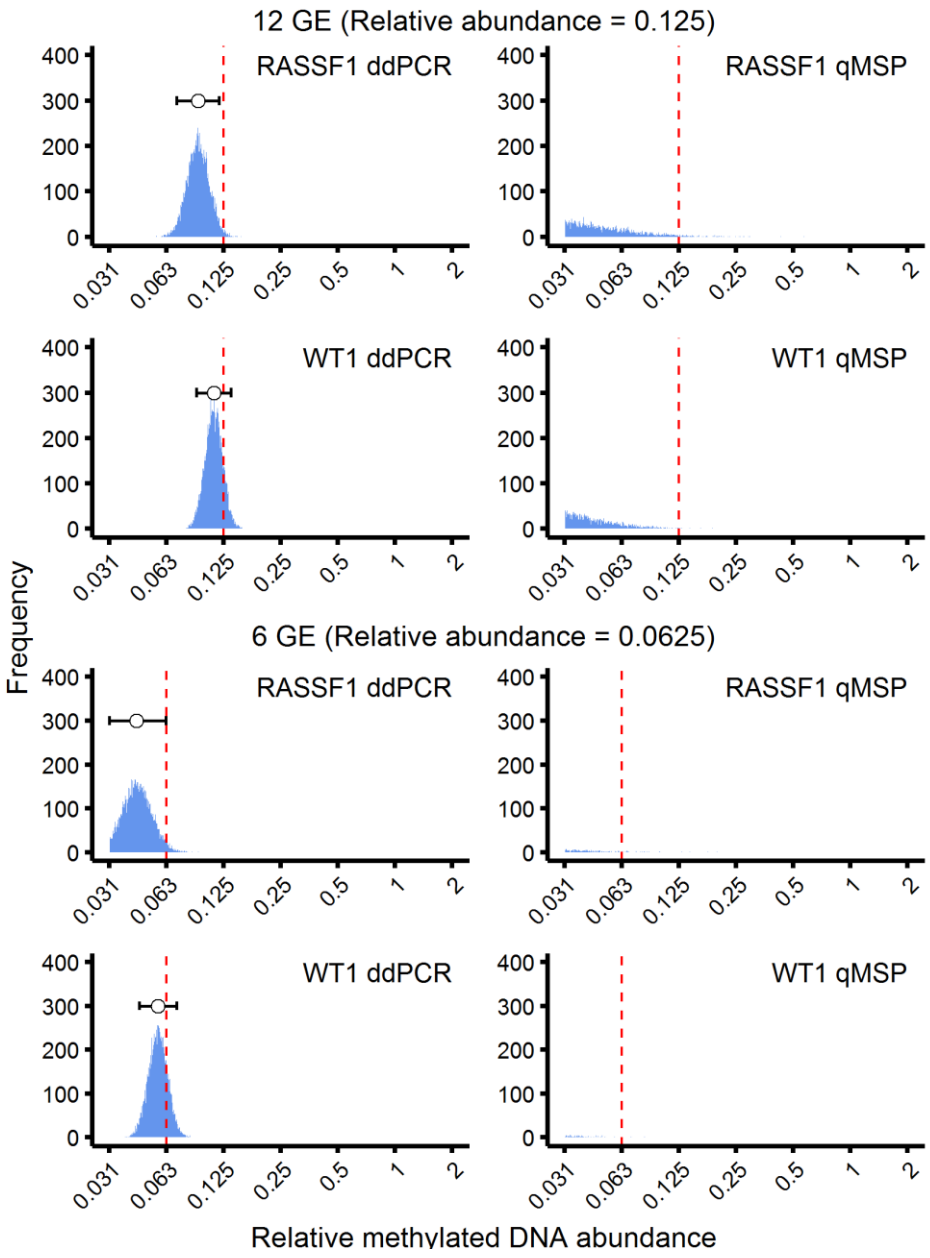


Figure 8.8 Histograms of bootstrap estimates for mixed model predictions for all experimental nominal methylated DNA inputs in the 6 – 96 GE range

Table 8.1 Statistical table for Dunnett's multiple comparison test for high total DNA input RASSF1/WT1 ddPCR/qMSP assays

All comparison are to 66ng PBMC DNA baseline

		ddPCR				qMSP			
Gene	Comparison (Methylated DNA input) [GE]	Estimate	Standard error	Dunnett t statistic	Adjusted P value	Estimate	Standard error	Dunnett t statistic	Adjusted P value
RASSF1	5 - PBMC	0.119	0.071	1.672	0.537	0.007	0.002	3.278	0.008
	10 - PBMC	0.206	0.060	3.407	0.005	0.012	0.005	2.415	0.119
	20 - PBMC	0.367	0.141	2.599	0.071	0.018	0.003	6.447	9.1×10^{-10}
	40 - PBMC	0.710	0.147	4.825	1.1×10^{-5}	0.058	0.022	2.687	0.056
	80 - PBMC	1.686	0.232	7.274	2.8×10^{-12}	0.105	0.017	6.005	1.5×10^{-8}
	160 - PBMC	3.508	0.549	6.390	1.3×10^{-9}	0.259	0.060	4.285	1.5×10^{-4}
	320 - PBMC	6.608	1.200	5.508	2.9×10^{-7}	0.758	0.100	7.594	2.5×10^{-13}
	640 - PBMC	14.030	0.826	16.986	$< 2 \times 10^{-16}$	1.118	0.176	6.337	1.9×10^{-9}
WT1	5 - PBMC	-0.009	0.109	-0.082	1.000	0.029	0.019	1.478	0.690
	10 - PBMC	0.129	0.143	0.903	0.972	0.034	0.013	2.561	0.080
	20 - PBMC	0.383	0.193	1.983	0.335	0.053	0.018	2.916	0.028
	40 - PBMC	0.952	0.267	3.565	0.005	0.171	0.048	3.570	0.003
	80 - PBMC	1.601	0.373	4.290	4.4×10^{-4}	0.312	0.095	3.267	0.009
	160 - PBMC	3.706	0.525	7.064	6.0×10^{-9}	0.474	0.079	6.002	1.5×10^{-8}
	320 - PBMC	7.639	0.740	10.328	2.2×10^{-14}	1.111	0.225	4.939	6.3×10^{-6}
	640 - PBMC	13.894	1.044	13.304	$< 2 \times 10^{-16}$	1.306	0.351	3.718	0.002

Table 8.2 Table of precision and trueness estimates for high total DNA input methylated DNA measurement

Asterisks indicate significant differences in repeatability variance (s_r^2) at the 95% confidence level

			ddPCR						qMSP					
				Repeatability		Intermediate precision				Repeatability		Intermediate precision		
Gene	Nominal methylated DNA input (GE)	Expected relative abundance	Mean relative abundance	SD (s_r)	CV (%)	SD ($s_{i(prep)}$)	CV (%)	Relative bias (%)	Mean relative abundance	SD (s_r)	CV (%)	SD ($s_{i(prep)}$)	CV (%)	Relative bias (%)
RASSF1	640	1.00	1.00	0.13*	13.37	0.18	18.50	0.00	1.12	0.39*	34.89	0.57	50.81	11.80
	320	0.50	0.48	0.29	60.71	0.29	60.71	-4.13	0.76	0.34	44.31	0.34	44.31	51.68
	160	0.25	0.26	0.13	49.24	0.13	49.24	4.75	0.26	0.17	65.93	0.18	71.28	3.59
	80	0.13	0.13	0.06	41.09	0.06	41.09	7.25	0.10	0.05	50.48	0.05	50.48	-16.34
	40	0.06	0.07	0.03	51.13	0.03	51.13	5.00	0.06	0.07	120.90	0.07	120.90	-7.42
WT1	640	1.00	1.00	0.30*	30.10	0.30	30.10	0.00	1.34	1.19*	88.90	1.19	88.90	34.18
	320	0.50	0.57	0.23*	40.18	0.23	40.18	13.72	1.15	0.68*	59.09	0.68	59.09	129.49
	160	0.25	0.30	0.11*	36.06	0.11	36.06	18.93	0.51	0.24*	46.55	0.24	46.55	104.13
	80	0.13	0.15	0.07*	42.93	0.07	42.93	21.75	0.35	0.25*	70.68	0.30	85.38	178.42
	40	0.06	0.11	0.04*	34.89	0.04	37.00	71.89	0.21	0.13*	61.54	0.15	70.21	230.75

Table 8.3 Table of statistics resulting from Tukey's honest significant difference test for ddPCR and qMSP analysis of methylated DNA in the input range 40 – 640 GE

		ddPCR				qMSP			
Gene	Comparison (Methylated DNA input) [GE]	Estimate	95% CI	Tukey's q statistic	Adjusted P value	Estimate	95% CI	Tukey's q statistic	Adjusted P value
RASSF1	80 - 40	1.08	(0.33, 1.83)	3.91	0.001	1.24	(0.05, 2.43)	2.84	0.036
	160 - 40	2.00	(1.25, 2.75)	7.28	1.4×10^{-12}	2.32	(1.13, 3.51)	5.32	8.0×10^{-7}
	320 - 40	2.83	(2.08, 3.58)	10.29	$<2 \times 10^{-16}$	4.13	(2.94, 5.32)	9.49	$<2 \times 10^{-16}$
	640 - 40	4.04	(3.28, 4.79)	14.67	$<2 \times 10^{-16}$	4.64	(3.45, 5.83)	10.65	$<2 \times 10^{-16}$
	160 - 80	0.93	(0.17, 1.68)	3.36	0.007	1.08	(-0.11, 2.27)	2.47	0.097
	320 - 80	1.75	(1.00, 2.50)	6.37	7.3×10^{-10}	2.89	(1.71, 4.08)	6.64	1.2×10^{-10}
	640 - 80	2.96	(2.21, 3.71)	10.76	$<2 \times 10^{-16}$	3.40	(2.21, 4.59)	7.80	3.2×10^{-14}
	320 - 160	0.83	(0.08, 1.58)	3.01	0.022	1.82	(0.63, 3.00)	4.17	2.8×10^{-4}
	640 - 160	2.03	(1.28, 2.78)	7.39	6.4×10^{-13}	2.32	(1.13, 3.51)	5.33	1.3×10^{-6}
	640 - 320	1.21	(0.45, 1.96)	4.38	1.2×10^{-4}	0.51	(-0.68, 1.69)	1.16	0.774
WT1	80 - 40	0.46	(-0.32, 1.23)	1.61	0.493	0.71	(-0.49, 1.91)	1.62	0.486
	160 - 40	1.51	(0.73, 2.29)	5.31	1.0×10^{-6}	1.48	(0.28, 2.68)	3.36	0.007
	320 - 40	2.40	(1.63, 3.18)	8.44	1.1×10^{-16}	2.62	(1.42, 3.82)	5.95	1.3×10^{-8}
	640 - 40	3.27	(2.49, 4.05)	11.50	$<2 \times 10^{-16}$	2.58	(1.37, 3.78)	5.85	3.5×10^{-8}
	160 - 80	1.05	(0.28, 1.83)	3.71	0.002	0.77	(-0.43, 1.97)	1.74	0.407
	320 - 80	1.94	(1.17, 2.72)	6.84	3.78×10^{-11}	1.91	(0.70, 3.11)	4.33	1.5×10^{-4}
	640 - 80	2.81	(2.04, 3.59)	9.89	$<2 \times 10^{-16}$	1.86	(0.66, 3.06)	4.23	2.6×10^{-4}
	320 - 160	0.89	(0.12, 1.67)	3.13	0.015	1.14	(-0.06, 2.34)	2.58	0.073
	640 - 160	1.76	(0.98, 2.54)	6.19	9.7×10^{-9}	1.10	(-0.11, 2.30)	2.49	0.094
	640 - 320	0.87	(0.09, 1.64)	3.06	0.019	-0.04	(-1.24, 1.16)	-0.10	1

Table 8.4 Likelihood ratio test results of significance of MethPlex random effects and Preparation random effects in LOD95 modelling

		MethPlex random effects			Preparation random effects		
Gene	**Method**	**df**	**χ^2**	***P* value**	**df**	**χ^2**	***P* value**
RASSF1	ddPCR	(0, 1)	3.41	0.032	(0, 1)	0	1
	qMSP	(0, 1)	14.21	8.2×10^{-5}	(0, 1)	7.7×10^{-8}	0.50
WT1	ddPCR	(0, 1)	12.05	2.5×10^{-4}	(0, 1)	0	1
	qMSP	(0, 1)	26.40	1.4×10^{-7}	(0, 1)	3.7×10^{-6}	0.5

Table 8.5 Table of precision estimates for 10ng total DNA input methylated DNA MethPlex enrichment quantification

Asterisks indicate significant differences in repeatability variance (s_r^2) at the 95% confidence level

			ddPCR							qMSP						
				PCR		MethPlex		Preparation			PCR		MethPlex		Preparation	
Gene	Nominal methylated DNA input(GE)	Expected relative abundance	Mean relative abundance	SD (s_r)	CV%	SD ($s_{i(MethPlex)}$)	CV%	SD ($s_{i(Prep)}$)	CV%	Mean relative abundance	SD (s_r)	CV%	SD ($s_{i(MethPlex)}$)	CV%	SD ($s_{i(Prep)}$)	CV%
RASSF1	96	1.00	1.00	0.23*	23.26	0.29	29.05	0.33	33.02	2.33	1.82	78.05	2.88	123.64	3.34	143.50
	48	0.50	0.39	0.17	43.65	0.20	51.99	0.20	51.99	0.34	0.17	49.69	0.19	55.54	0.29	83.55
	24	0.25	0.22	0.11	47.99	0.17	73.87	0.17	73.87	0.13	0.06*	46.03	0.24	186.48	0.25	188.76
	12	0.13	0.12	0.06	46.31	0.09	75.06	0.09	75.06	0.07	0.08	123.27	0.12	183.64	0.13	197.51
	6	0.06	0.05	0.02	38.12	0.03	49.53	0.03	59.22	0.01	0.01	90.26	0.01	98.51	0.02	171.32
WT1	96	1.00	1.00	0.33*	32.91	0.33	32.91	0.34	33.52	2.33	1.64	70.48	1.64	70.48	3.03	130.30
	48	0.50	0.44	0.16*	36.00	0.22	49.01	0.22	49.22	0.37	0.26	69.76	0.43	114.89	0.45	119.75
	24	0.25	0.26	0.09*	34.69	0.18	68.68	0.18	68.68	0.21	0.23	108.72	0.25	122.05	0.27	128.69
	12	0.13	0.12	0.05	40.48	0.08	62.48	0.08	62.48	0.03	0.03*	82.97	0.07	193.22	0.07	196.82
	6	0.06	0.07	0.02*	30.58	0.02	35.26	0.02	35.26	0.05	0.04	77.40	0.11	214.70	0.12	224.08

Table 8.6 MethPlex low total DNA input methylated DNA analysis variance component table

Variance components were calculated by nested ANOVA using the VCA R package (Schuetzenmeister, 2017). Equivalent repeatability (ICC) calculations of hierarchical decomposition of variances using a mixed model approach in the rptR package (Stoffel et al., 2017) were similar.

Gene	Method	Variance component	Estimated contribution to total variance (%)				
			Nominal methylated DNA input (GE)				
			96	48	24	12	6
RASSF1	ddPCR	Preparation	20.3	0	0	0	37.1
		MethPlex	20.2	19.2	41.5	72.3	27.5
		PCR	59.4	80.8	58.5	27.7	35.4
	qMSP	Preparation	82.8	70.7	29.3	36.4	77.2
		MethPlex	10.2	16.9	51.7	47.7	16.9
		PCR	7	12.4	19	16	5.9
WT1	ddPCR	Preparation	0	0	0	5.2	0
		MethPlex	1.5	30.9	78.2	59.0	26.3
		PCR	98.5	69.1	21.8	35.8	73.7
	qMSP	Preparation	73.8	4.3	30.2	0	0.2
		MethPlex	0	69.5	47.6	93.9	98.6
		PCR	26.2	26.3	22.2	6.1	1.2

Table 8.7 Equivalence of preparation as a fixed effect and random effect in the 66 ng total DNA ddPCR input case

		Preparation variance		Residual variance		Model intercept		Intercept standard error		Preparation effect *P* value	
Gene	**Method**	**Random**	**Fixed**	**Random**	**Fixed**	**Random**	**Fixed**	**Random**	**Fixed**	**Random[1]**	**Fixed[2]**
RASSF1	ddPCR	0.000	0.004	0.323	0.336	-2.069	-2.069	0.085	0.086	1	0.822
	qMSP	0.017	0.074	0.856	0.856	-2.173	-2.173	0.157	0.138	0.5	0.283
WT1	ddPCR	0.000	0.03	0.355	0.370	-1.797	-1.797	0.089	0.091	1	0.871
	qMSP	0.000	0.048	0.874	0.881	-1.098	-1.098	0.139	0.140	1	0.447

[1] Likelihood Ratio Test, corrected P value (Self and Liang, 1987; Stram and Lee, 1994)
[2] F-test

Table 8.8 Equivalence of preparation as a fixed effect and random effect in the 10 ng total DNA MethPlex enrichment input case

		Preparation variance		Residual variance		Model intercept		Intercept standard error		Preparation effect *P* value	
Gene	**Method**	**Random**	**Fixed**	**Random**	**Fixed**	**Random**	**Fixed**	**Random**	**Fixed**	**Random[1]**	**Fixed[2]**
RASSF1	ddPCR	0.000	0.070	0.397	0.397	-2.377	-2.377	0.134	0.157	1	0.457
	qMSP	3.159	3.418	0.990	0.990	-3.790	-3.790	0.861	0.253	0.002	0.002
WT1	ddPCR	0.000	0.061	0.428	0.428	-2.196	-2.196	0.142	0.126	1	0.472
	qMSP	1.711	1.852	2.712	2.712	-3.850	-3.850	0.691	0.317	0.007	0.012

[1] Likelihood Ratio Test, corrected *P* value (Self and Liang, 1987; Stram and Lee, 1994)
[2] Kenward-Roger F-test

Chapter 9

References

Alnaes, G., Ronneberg, J., Kristensen, V., and Tost, J. (2015). Heterogeneous DNA Methylation Patterns in the GSTP1 Promoter Lead to Discordant Results between Assay Technologies and Impede Its Implementation as Epigenetic Biomarkers in Breast Cancer. Genes *6*, 878–900.

Amatu, A., Sartore-Bianchi, A., Moutinho, C., Belotti, A., Bencardino, K., Chirico, G., Cassingena, A., Rusconi, F., Esposito, A., Nichelatti, M., et al. (2013). Promoter CpG island hypermethylation of the DNA repair enzyme MGMT predicts clinical response to dacarbazine in a phase II study for metastatic colorectal cancer. Clin. Cancer Res. *19*, 2265–2272.

Asada, K., Nakajima, T., Shimazu, T., Yamamichi, N., Maekita, T., Yokoi, C., Oda, I., Ando, T., Yoshida, T., Nanjo, S., et al. (2015). Demonstration of the usefulness of epigenetic cancer risk prediction by a multicentre prospective cohort study. Gut *64*, 388–396.

Bach, P.B., Mirkin, J.N., Oliver, T.K., Azzoli, C.G., Berry, D.A., Brawley, O.W., Byers, T., Colditz, G.A., Gould, M.K., Jett, J.R., et al. (2012). Benefits and harms of CT screening for lung cancer: a systematic review. JAMA *307*, 2418–2429.

Baird, D.M. (2010). Variation at the TERT locus and predisposition for cancer. Expert Rev. Mol. Med. *12*, e16.

Barault, L., Amatu, A., Bleeker, F.E., Moutinho, C., Falcomatà, C., Fiano, V., Cassingena, A., Siravegna, G., Milione, M., Cassoni, P., et al. (2015). Digital PCR quantification of *MGMT* methylation refines prediction of clinical benefit from alkylating agents in glioblastoma and metastatic colorectal cancer. Ann. Oncol. *26*, 1994–1999.

Bashtrykov, P., Jankevicius, G., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2014). The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. J. Biol. Chem. *289*, 4106–4115.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. *67*.

Baylin, S.B., and Jones, P.A. (2011). A decade of exploring the cancer epigenome — biological and translational implications. Nat. Rev. Cancer *11*, 726–734.

Beck, J., Bierau, S., Balzer, S., Andag, R., Kanzow, P., Schmitz, J., Gaedcke, J., Moerer, O., Slotta, J.E., Walson, P., et al. (2013). Digital droplet PCR for rapid quantification of donor DNA in the circulation of transplant recipients as a potential universal biomarker of graft injury. Clin. Chem. *59*, 1732–1741.

Bediaga, N.G., Davies, M.P.A., Acha-Sagredo, A., Hyde, R., Raji, O.Y., Page, R., Walshaw, M., Gosney, J., Alfirevic, A., Field, J.K., et al. (2013). A microRNA-based prediction algorithm for diagnosis of non-small lung cell carcinoma in minimal biopsy material. Br. J. Cancer *109*, 2404–2411.

Belinsky, S.A., Nikula, K.J., Palmisano, W.A., Michels, R., Saccomanno, G., Gabrielson, E., Baylin, S.B., and Herman, J.G. (1998). Aberrant methylation of p16INK4a is an early event in lung cancer and a potential biomarker for early diagnosis. Proc. Natl. Acad. Sci. *95*, 11891–11896.

Belinsky, S.A., Klinge, D.M., Dekker, J.D., Smith, M.W., Bocklage, T.J., Gilliland, F.D., Crowell, R.E., Karp, D.D., Stidley, C.A., and Picchi, M.A. (2005). Gene Promoter

Methylation in Plasma and Sputum Increases with Lung Cancer Risk. Clin. Cancer Res. *11*, 6505–6511.

Bettegowda, C., Sausen, M., Leary, R.J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B.R., Wang, H., Luber, B., Alani, R.M., et al. (2014). Detection of circulating tumor DNA in early-and late-stage human malignancies. Sci. Transl. Med. *6*, 224ra24–224ra24.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91–99.

Blakely, C.M., and Bivona, T.G. (2012). Resiliency of Lung Cancers to EGFR Inhibitor Treatment Unveiled, Offering Opportunities to Divide and Conquer EGFR Inhibitor Resistance: Figure 1. Cancer Discov. *2*, 872–875.

Bland, J.M. (2006). How should I calculate a within-subject coefficient of variation?

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., and White, J.-S.S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol. Evol. *24*, 127–135.

Brown, D., and Pasloske, B.L. (2001). Ribonuclease-Resistant RNA Controls and Standards. In Methods in Enzymology, A.W. Nicholson, ed. (Academic Press), pp. 648–654.

Bundo, M., Sunaga, F., Ueda, J., Kasai, K., Kato, T., and Iwamoto, K. (2012). A systematic evaluation of whole genome amplification of bisulfite-modified DNA. Clin. Epigenetics *4*, 22.

Burmester, T., Ebner, B., Weich, B., and Hankeln, T. (2002). Cytoglobin: A Novel Globin Type Ubiquitously Expressed inVertebrate Tissues. Mol. Biol. Evol. *19*, 416–421.

Bustin, S. (2017). The continuing problem of poor transparency of reporting and use of inappropriate methods for RT-qPCR. Biomol. Detect. Quantif. *12*, 7–9.

Bustin, S.A. (2014). The reproducibility of biomedical research: Sleepers awake! Biomol. Detect. Quantif. *2*, 35–42.

Bustin, S., and Huggett, J. (2017). qPCR primer design revisited. Biomol. Detect. Quantif. *14*, 19–28.

Calcagno, V. (2013). glmulti: Model selection and multimodel inference made easy.

Calle, M.L., Urrea, V., Malats, N., and Van Steen, K. (2010). mbmdr: an R package for exploring gene-gene interactions associated with binary or quantitative traits. Bioinforma. Oxf. Engl. *26*, 2198–2199.

Cancer Research UK (2015a). Cancer incidence for common cancers.

Cancer Research UK (2015b). Cancer survival for common cancers.

Cancer Research UK (2015c). Cancer mortality for common cancers.

Cancer Research UK (2015d). Lung cancer diagnosis and treatment statistics.

Candiloro, I.L.M., Mikeska, T., and Dobrovic, A. (2011). Closed-Tube PCR Methods for Locus-Specific DNA Methylation Analysis. In Epigenetics Protocols, T.O. Tollefsbol, ed. (Totowa, NJ: Humana Press), pp. 55–71.

Cardarella, S., Ogino, A., Nishino, M., Butaney, M., Shen, J., Lydon, C., Yeap, B.Y., Sholl, L.M., Johnson, B.E., and Jänne, P.A. (2013). Clinical, pathologic, and biologic features associated with BRAF mutations in non-small cell lung cancer. Clin. Cancer Res. *19*, 4532–4540.

Carmeliet, P., and Jain, R.K. (2011). Molecular mechanisms and clinical applications of angiogenesis. Nature *473*, 298–307.

Castonguay, A., Lin, D., Stoner, G.D., Radok, P., Furuya, K., Hecht, S.S., Schut, H.A., and Klaunig, J.E. (1983). Comparative carcinogenicity in A/J mice and metabolism by cultured mouse peripheral lung of N'-nitrosonornicotine, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone, and their analogues. Cancer Res. *43*, 1223–1229.

Cazzoli, R., Buttitta, F., Di Nicola, M., Malatesta, S., Marchetti, A., Rom, W.N., and Pass, H.I. (2013). microRNAs derived from circulating exosomes as noninvasive biomarkers for screening and diagnosing lung cancer. J. Thorac. Oncol. *8*, 1156–1162.

Chen, Y., Breeze, C.E., Zhen, S., Beck, S., and Teschendorff, A.E. (2016). Tissue-independent and tissue-specific patterns of DNA methylation alteration in cancer. Epigenetics Chromatin *9*.

Church, T.R., Wandell, M., Lofton-Day, C., Mongin, S.J., Burger, M., Payne, S.R., Castaños-Vélez, E., Blumenstein, B.A., Rösch, T., Osborn, N., et al. (2014). Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut *63*, 317–325.

Clark, S.J., Statham, A., Stirzaker, C., Molloy, P.L., and Frommer, M. (2006). DNA methylation: Bisulphite modification and analysis. Nat. Protoc. *1*, 2353–2364.

Claus, R., Wilop, S., Hielscher, T., Sonnet, M., Dahl, E., Galm, O., Jost, E., and Plass, C. (2012). A systematic comparison of quantitative high-resolution DNA methylation analysis and methylation-specific PCR. Epigenetics *7*, 772–780.

Clement-Jones, M., Schiller, S., Rao, E., Blaschke, R.J., Zuniga, A., Zeller, R., Robson, S.C., Binder, G., Glass, I., Strachan, T., et al. (2000). The short stature homeobox gene SHOX is involved in skeletal abnormalities in Turner syndrome. Hum. Mol. Genet. *9*, 695–702.

Cohen, J.D., Javed, A.A., Thoburn, C., Wong, F., Tie, J., Gibbs, P., Schmidt, C.M., Yip-Schneider, M.T., Allen, P.J., Schattner, M., et al. (2017). Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. Proc. Natl. Acad. Sci. *114*, 10202–10207.

Cooper, D.N., Taggart, M.H., and Bird, A.P. (1983). Unmethylated domains in vertebrate DNA. Nucleic Acids Res. *11*, 647–658.

Coughlin, S.R. (2000). Thrombin signalling and protease-activated receptors. Nature *407*, 258–264.

Crowley, E., Di Nicolantonio, F., Loupakis, F., and Bardelli, A. (2013). Liquid biopsy: monitoring cancer-genetics in the blood. Nat. Rev. Clin. Oncol. *10*, 472–484.

Dammann, R., Li, C., Yoon, J.-H., Chin, P.L., Bates, S., and Pfeifer, G.P. (2000). Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. Nat. Genet. *25*, 315–319.

Daskalos, A., Nikolaidis, G., Xinarianos, G., Savvari, P., Cassidy, A., Zakopoulou, R., Kotsinas, A., Gorgoulis, V., Field, J.K., and Liloglou, T. (2009). Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. Int. J. Cancer *124*, 81–87.

Daskalos, A., Oleksiewicz, U., Filia, A., Nikolaidis, G., Xinarianos, G., Gosney, J.R., Malliri, A., Field, J.K., and Liloglou, T. (2011). UHRF1-mediated tumor suppressor gene inactivation in nonsmall cell lung cancer. Cancer *117*, 1027–1037.

Davidson, M.R., Gazdar, A.F., and Clarke, B.E. (2013). The pivotal role of pathology in the management of lung cancer. J. Thorac. Dis. *5*, S463.

Davies, K.D., Le, A.T., Theodoro, M.F., Skokan, M.C., Aisner, D.L., Berge, E.M., Terracciano, L.M., Cappuzzo, F., Incarbone, M., Roncalli, M., et al. (2012). Identifying and targeting ROS1 gene fusions in non-small cell lung cancer. Clin. Cancer Res. *18*, 4570–4579.

Day, E., Dear, P.H., and McCaughan, F. (2013). Digital PCR strategies in the development and analysis of molecular biomarkers for personalized medicine. Methods *59*, 101–107.

De Mattos-Arruda, L., Weigelt, B., Cortes, J., Won, H.H., Ng, C.K.Y., Nuciforo, P., Bidard, F.-C., Aura, C., Saura, C., Peg, V., et al. (2014). Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. Ann. Oncol. *25*, 1729–1735.

Dela Cruz, C.S., Tanoue, L.T., and Matthay, R.A. (2011). Lung cancer: epidemiology, etiology, and prevention. Clin. Chest Med. *32*, 605–644.

DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics *44*, 837–845.

Devonshire, A.S., Whale, A.S., Gutteridge, A., Jones, G., Cowen, S., Foy, C.A., and Huggett, J.F. (2014). Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. Anal. Bioanal. Chem. *406*, 6499–6512.

Diehl, F., Li, M., Dressman, D., He, Y., Shen, D., Szabo, S., Diaz, L.A., Goodman, S.N., David, K.A., and Juhl, H. (2005). Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc. Natl. Acad. Sci. *102*, 16368–16373.

Dietrich, D. (2011). Performance evaluation of the DNA methylation biomarker SHOX2 for the aid in diagnosis of lung cancer based on the analysis of bronchial aspirates. Int. J. Oncol.

Dingle, T.C., Sedlak, R.H., Cook, L., and Jerome, K.R. (2013). Tolerance of droplet-digital PCR vs real-time quantitative PCR to inhibitory substances. Clin. Chem. *59*, 1670–1672.

Dubois, F., Keller, M., Calvayrac, O., Soncin, F., Hoa, L., Hergovich, A., Parrini, M.-C., Mazières, J., Vaisse-Lesteven, M., Camonis, J., et al. (2016). RASSF1A Suppresses the Invasion and Metastatic Potential of Human Non-Small Cell Lung Cancer Cells by

Inhibiting YAP Activation through the GEF-H1/RhoB Pathway. Cancer Res. *76*, 1627–1640.

Eads, C.A., Danenberg, K.D., Kawakami, K., Saltz, L.B., Blake, C., Shibata, D., Danenberg, P.V., and Laird, P.W. (2000). MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res. *28*, E32.

Ellison, S.L., English, C.A., Burns, M.J., and Keer, J.T. (2006). Routes to improving the reliability of low level DNA analysis using real-time PCR. BMC Biotechnol. *6*, 1.

Engelman, J.A., Zejnullahu, K., Mitsudomi, T., Song, Y., Hyland, C., Park, J.O., Lindeman, N., Gale, C.-M., Zhao, X., Christensen, J., et al. (2007). MET Amplification Leads to Gefitinib Resistance in Lung Cancer by Activating ERBB3 Signaling. Science *316*, 1039–1043.

Feinberg, A.P., Ohlsson, R., and Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. Nat. Rev. Genet. *7*, 21–33.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer *136*, E359–E386.

Field, J.K., and Raji, O.Y. (2010). The potential for using risk models in future lung cancer screening trials. F1000 Med. Rep. *2*.

Field, J.K., Smith, D.L., Duffy, S., and Cassidy, A. (2005). The Liverpool Lung Project research protocol. Int. J. Oncol. *27*, 1633–1645.

Fleischhacker, M., and Schmidt, B. (2007). Circulating nucleic acids (CNAs) and cancer—A survey. Biochim. Biophys. Acta BBA - Rev. Cancer *1775*, 181–232.

Fontanini, G., Vignati, S., Basolo, F., Bevilacqua, G., Lucchi, M., Mussi, A., Angeletti, C.A., Ciardiello, F., De Laurentiis, M., and De Placido, S. (1997). Angiogenesis as a Prognostic Indicator of Survival in Non-Small-Cell Lung Carcinoma: a Prospective Study. JNCI J. Natl. Cancer Inst. *89*, 881–886.

Forshew, T., Murtaza, M., Parkinson, C., Gale, D., Tsui, D.W.Y., Kaper, F., Dawson, S.-J., Piskorz, A.M., Jimenez-Linan, M., Bentley, D., et al. (2012). Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA. Sci. Transl. Med. *4*, 136ra68-136ra68.

Fraga, M.F., and Esteller, M. (2002). DNA methylation: a profile of methods and applications. Biotechniques *33*, 632–649.

de Fraipont, F., Moro-Sibilot, D., Michelland, S., Brambilla, E., Brambilla, C., and Favrot, M.C. (2005). Promoter methylation of genes in bronchial lavages: A marker for early diagnosis of primary and relapsing non-small cell lung cancer? Lung Cancer *50*, 199–209.

Fruh, M., De Ruysscher, D., Popat, S., Crino, L., Peters, S., Felip, E., and on behalf of the ESMO Guidelines Working Group (2013). Small-cell lung cancer (SCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. *24*, vi99–vi105.

Fu, J., Lv, H., Guan, H., Ma, X., Ji, M., He, N., Shi, B., and Hou, P. (2013). Metallothionein 1G functions as a tumor suppressor in thyroid cancer through modulating the PI3K/Akt signaling pathway. BMC Cancer *13*, 462.

Gaafar, R. (2017). SC17.05 Lung Cancer in Africa: Challenges and Perspectives. J. Thorac. Oncol. *12*, S115–S116.

Gadgeel, S.M. (2016). Personalized Therapy of Non-small Cell Lung Cancer (NSCLC). In Lung Cancer and Personalized Medicine: Novel Therapies and Clinical Management, (Springer, Cham), pp. 203–222.

Gaertner, J., Weingärtner, V., Lange, S., Hausner, E., Gerhardus, A., Simon, S.T., Voltz, R., Becker, G., and Schmacke, N. (2015). The Role of End-of-Life Issues in the Design and Reporting of Cancer Clinical Trials: A Structured Literature Review. PLOS ONE *10*, e0136640.

Garrigou, S., Perkins, G., Garlan, F., Normand, C., Didelot, A., Le Corre, D., Peyvandi, S., Mulot, C., Niarra, R., Aucouturier, P., et al. (2016). A Study of Hypermethylated Circulating Tumor DNA as a Universal Colorectal Cancer Biomarker. Clin. Chem. *62*, 1129–1139.

Gaspar, L.E., McNamara, E.J., Gay, E.G., Putnam, J.B., Crawford, J., Herbst, R.S., and Bonner, J.A. (2012). Small-cell lung cancer: prognostic factors and changing treatment over 15 years. Clin. Lung Cancer *13*, 115–122.

Gelman, A., and Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. (Cambridge ; Cambridge University Press, 2007.).

Gelman, A., and Su, Y.-S. (2016). arm: Data Analysis Using Regression and Multilevel/Hierarchical Models.

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., et al. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. N. Engl. J. Med. *366*, 883–892.

Gevensleben, H., Garcia-Murillas, I., Graeser, M.K., Schiavon, G., Osin, P., Parton, M., Smith, I.E., Ashworth, A., and Turner, N.C. (2013). Noninvasive detection of HER2 amplification with plasma DNA digital PCR. Clin. Cancer Res. *19*, 3276–3284.

Ghosh, S., Yates, A.J., Frühwald, M.C., Miecznikowski, J.C., Plass, C., and Smiraglia, D. (2010). Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. Epigenetics *5*, 527–538.

Goldstraw, P., Chansky, K., Crowley, J., Rami-Porta, R., Asamura, H., Eberhardt, W.E.E., Nicholson, A.G., Groome, P., Mitchell, A., Bolejack, V., et al. (2016). The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. J. Thorac. Oncol. *11*, 39–51.

Gormally, E., Caboux, E., Vineis, P., and Hainaut, P. (2007). Circulating free DNA in plasma or serum as biomarker of carcinogenesis: Practical aspects and biological significance. Mutat. Res. *635*, 105–117.

Goss, G., Tsai, C.-M., Shepherd, F.A., Bazhenova, L., Lee, J.S., Chang, G.-C., Crino, L., Satouchi, M., Chu, Q., Hida, T., et al. (2016). Osimertinib for pretreated EGFR Thr790Met-positive advanced non-small-cell lung cancer (AURA2): a multicentre, open-label, single-arm, phase 2 study. Lancet Oncol. *17*, 1643–1652.

Grunau, C., Clark, S.J., and Rosenthal, A. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res. *29*, e65–e65.

Haasbeek, C.J.A., Slotman, B.J., and Senan, S. (2009). Radiotherapy for lung cancer: Clinical impact of recent technical advances. Lung Cancer *64*, 1–8.

Haasbeek, C.J.A., Palma, D., Visser, O., Lagerwaard, F.J., Slotman, B., and Senan, S. (2012). Early-stage lung cancer in elderly patients: A population-based study of changes in treatment patterns and survival in the Netherlands. Ann. Oncol. *23*, 2743–2747.

Hahn, L.W., Ritchie, M.D., and Moore, J.H. (2003). Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. Bioinforma. Oxf. Engl. *19*, 376–382.

Halekoh, U., and Højsgaard, S. (2014). A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models–the R package pbkrtest. J. Stat. Softw. *59*, 1–32.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674.

Hanna, N., Johnson, D., Temin, S., Baker, S., Brahmer, J., Ellis, P.M., Giaccone, G., Hesketh, P.J., Jaiyesimi, I., Leighl, N.B., et al. (2017). Systemic Therapy for Stage IV Non–Small-Cell Lung Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update. J. Clin. Oncol. *35*, 3484–3515.

Hara, E., Smith, R., Parry, D., Tahara, H., Stone, S., and Peters, G. (1996). Regulation of p16CDKN2 expression and its implications for cell immortalization and senescence. Mol. Cell. Biol. *16*, 859–867.

Hartig, F. (2017). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models.

Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Dev. Camb. Engl. *129*, 1983–1993.

Hata, T., Dal Molin, M., Hong, S.-M., Tamura, K., Suenaga, M., Yu, J., Sedogawa, H., Weiss, M.J., Wolfgang, C.L., Lennon, A.M., et al. (2017). Predicting the Grade of Dysplasia of Pancreatic Cystic Neoplasms Using Cyst Fluid DNA Methylation Markers. Clin. Cancer Res. *23*, 3935–3944.

Hayashi, M., Guerrero-Preston, R., Sidransky, D., and Koch, W.M. (2015). Paired Box 5 Methylation Detection by Droplet Digital PCR for Ultra-Sensitive Deep Surgical Margins Analysis of Head and Neck Squamous Cell Carcinoma. Cancer Prev. Res. (Phila. Pa.) *8*, 1017–1026.

Hayden, R.T., Gu, Z., Ingersoll, J., Abdul-Ali, D., Shi, L., Pounds, S., and Caliendo, A.M. (2013). Comparison of Droplet Digital PCR to Real-Time PCR for Quantitative Detection of Cytomegalovirus. J. Clin. Microbiol. *51*, 540–546.

Herman, J.G., Graff, J.R., Myöhänen, S., Nelkin, B.D., and Baylin, S.B. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc. Natl. Acad. Sci. *93*, 9821–9826.

Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J. Biol. Chem. *279*, 48350–48359.

Herth, F.J.F., Eberhardt, R., Vilmann, P., Krasnik, M., and Ernst, A. (2006). Real-time endobronchial ultrasound guided transbronchial needle aspiration for sampling mediastinal lymph nodes. Thorax *61*, 795–798.

Heuvers, M.E., Wisnivesky, J., Stricker, B.H., and Aerts, J.G. (2012). Generalizability of results from the National Lung Screening Trial. Eur. J. Epidemiol. *27*, 669–672.

Heyn, H., Vidal, E., Ferreira, H.J., Vizoso, M., Sayols, S., Gomez, A., Moran, S., Boque-Sastre, R., Guil, S., Martinez-Cardus, A., et al. (2016). Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. Genome Biol. *17*.

Hicklin, D.J., and Ellis, L.M. (2005). Role of the Vascular Endothelial Growth Factor Pathway in Tumor Growth and Angiogenesis. J. Clin. Oncol. *23*, 1011–1027.

Higgins, C.F. (2007). Multiple molecular mechanisms for multidrug resistance transporters. Nature *446*, 749–757.

Hindson, B.J., Ness, K.D., Masquelier, D.A., Belgrader, P., Heredia, N.J., Makarewicz, A.J., Bright, I.J., Lucero, M.Y., Hiddessen, A.L., Legler, T.C., et al. (2011). High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. Anal. Chem. *83*, 8604–8610.

Hindson, C.M., Chevillet, J.R., Briggs, H.A., Gallichotte, E.N., Ruf, I.K., Hindson, B.J., Vessella, R.L., and Tewari, M. (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat. Methods *10*, 1003–1005.

Holdenrieder, S., Stieber, P., Bodenmüller, H., Busch, M., Pawel, J., Schalhorn, A., Nagel, D., and Seidel, D. (2001). Circulating nucleosomes in serum. Ann. N. Y. Acad. Sci. *945*, 93–102.

Holmberg, L., Sandin, F., Bray, F., Richards, M., Spicer, J., Lambe, M., Klint, A., Peake, M., Strand, T.-E., Linklater, K., et al. (2010). National comparisons of lung cancer survival in England, Norway and Sweden 2001-2004: differences occur early in follow-up. Thorax *65*, 436–441.

Holmes, E.E., Jung, M., Meller, S., Leisse, A., Sailer, V., Zech, J., Mengdehl, M., Garbe, L.-A., Uhl, B., Kristiansen, G., et al. (2014). Performance Evaluation of Kits for Bisulfite-Conversion of DNA from Tissues, Cell Lines, FFPE Tissues, Aspirates, Lavages, Effusions, Plasma, Serum, and Urine. PLoS ONE *9*, e93933.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous Inference in General Parametric Models. Biom. J. *50*, 346–363.

How Kit, A., Nielsen, H.M., and Tost, J. (2012). DNA methylation based biomarkers: Practical considerations and applications. Biochimie *94*, 2314–2337.

Hubers, A.J., van der Drift, M.A., Prinsen, C.F.M., Witte, B.I., Wang, Y., Shivapurkar, N., Stastny, V., Bolijn, A.S., Hol, B.E.A., Feng, Z., et al. (2014). Methylation analysis in spontaneous sputum for lung cancer diagnosis. Lung Cancer *84*, 127–133.

Huggett, J.F., Foy, C.A., Benes, V., Emslie, K., Garson, J.A., Haynes, R., Hellemans, J., Kubista, M., Mueller, R.D., Nolan, T., et al. (2013). The Digital MIQE Guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments. Clin. Chem. *59*, 892–902.

Hulbert, A., Jusue-Torres, I., Stark, A., Chen, C., Rodgers, K., Lee, B., Griffin, C., Yang, A., Huang, P., Wrangle, J., et al. (2017). Early Detection of Lung Cancer Using DNA Promoter Hypermethylation in Plasma and Sputum. Clin. Cancer Res. *23*, 1998–2005.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. *41*, 178–186.

ISO (1994a). ISO 5725-2:1994(en), Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.

ISO (1994b). ISO 5725-3:1994(en), Accuracy (trueness and precision) of measurement methods and results — Part 3: Intermediate measures of the precision of a standard measurement method.

ISO (1994c). ISO 5725-6:1994(en), Accuracy (trueness and precision) of measurement methods and results — Part 6: Use in practice of accuracy values.

ISO (2000). ISO 11843-2:2000 - Capability of detection -- Part 2: Methodology in the linear calibration case.

Jahr, S., Hentze, H., Englisch, S., Hardt, D., Fackelmayer, F.O., Hesch, R.-D., and Knippers, R. (2001). DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res. *61*, 1659–1665.

Jamal-Hanjani, M., Wilson, G.A., Horswell, S., Mitter, R., Sakarya, O., Constantin, T., Salari, R., Kirkizlar, E., Sigurjonsson, S., Pelham, R., et al. (2016). Detection of ubiquitous and heterogeneous mutations in cell-free DNA from patients with early-stage non-small-cell lung cancer. Ann. Oncol. *27*, 862–867.

Janku, F., Claes, B., Huang, H.J., Falchook, G.S., Devogelaere, B., Kockx, M., Bempt, I.V., Reijans, M., Naing, A., Fu, S., et al. (2015). BRAF mutation testing with a rapid, fully integrated molecular diagnostics system. Oncotarget *6*, 26886–26894.

Janku, F., Huang, H.J., Claes, B., Falchook, G.S., Fu, S., Hong, D., Ramzanali, N.M., Nitti, G., Cabrilo, G., Tsimberidou, A.M., et al. (2016). BRAF Mutation Testing in Cell-Free DNA from the Plasma of Patients with Advanced Cancers Using a Rapid, Automated Molecular Diagnostics System. Mol. Cancer Ther. *15*, 1397–1404.

Jänne, P.A., Yang, J.C.-H., Kim, D.-W., Planchard, D., Ohe, Y., Ramalingam, S.S., Ahn, M.-J., Kim, S.-W., Su, W.-C., Horn, L., et al. (2015). AZD9291 in EGFR Inhibitor–Resistant Non–Small-Cell Lung Cancer. N. Engl. J. Med. *372*, 1689–1699.

Jiang, P., Chan, C.W., Chan, K.A., Cheng, S.H., Wong, J., Wong, V.W.-S., Wong, G.L., Chan, S.L., Mok, T.S., and Chan, H.L. (2015). Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc. Natl. Acad. Sci. *112*, E1317–E1325.

Jin, X., Chen, Y., Chen, H., Fei, S., Chen, D., Cai, X., Liu, L., Lin, B., Su, H., Zhao, L., et al. (2017). Evaluation of Tumor-Derived Exosomal miRNA as Potential Diagnostic Biomarkers for Early-Stage Non-Small Cell Lung Cancer Using Next-Generation Sequencing. Clin. Cancer Res. *23*, 5311–5319.

Johnson, P.C.D. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. Methods Ecol. Evol. *5*, 944–946.

Johnson, D.H., Fehrenbacher, L., Novotny, W.F., Herbst, R.S., Nemunaitis, J.J., Jablons, D.M., Langer, C.J., DeVore, R.F., Gaudreault, J., Damico, L.A., et al. (2004). Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. J. Clin. Oncol. *22*, 2184–2191.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. Cell *128*, 683–692.

Jones, P.A., and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. Nat. Rev. Genet. *10*, 805–811.

Jones, M., Williams, J., Gärtner, K., Phillips, R., Hurst, J., and Frater, J. (2014). Low copy target detection by Droplet Digital PCR through application of a novel open access bioinformatic pipeline, 'definetherain.' J. Virol. Methods *202*, 46–53.

Kacem, S., and Feil, R. (2009). Chromatin mechanisms in genomic imprinting. Mamm. Genome *20*, 544–556.

Kahlert, C., Melo, S.A., Protopopov, A., Tang, J., Seth, S., Koch, M., Zhang, J., Weitz, J., Chin, L., Futreal, A., et al. (2014). Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. J. Biol. Chem. *289*, 3869–3875.

Kalluri, R. (2016). The biology and function of exosomes in cancer. J. Clin. Invest. *126*, 1208–1215.

Kalofonou, M., Georgiou, P., Ou, C.-P., and Toumazou, C. (2012). An ISFET based translinear sensor for DNA methylation detection. Sens. Actuators B Chem. *161*, 156–162.

Kelly, T.K., Miranda, T.B., Liang, G., Berman, B.P., Lin, J.C., Tanay, A., and Jones, P.A. (2010). H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. Mol. Cell *39*, 901–911.

Kenward, M.G., and Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. Biometrics *53*, 983–997.

Khakwani, A., Rich, A.L., Tata, L.J., Powell, H.A., Stanley, R.A., Baldwin, D.R., and Hubbard, R.B. (2014). Small-Cell Lung Cancer in England: Trends in Survival and Chemotherapy Using the National Lung Cancer Audit. PLOS ONE *9*, e89426.

Kneip, C., Schmidt, B., Seegebarth, A., Weickmann, S., Fleischhacker, M., Liebenberg, V., Field, J.K., and Dietrich, D. (2011). SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. J. Thorac. Oncol. *6*, 1632–1638.

Kobayashi, S., Boggon, T.J., Dayaram, T., Jänne, P.A., Kocher, O., Meyerson, M., Johnson, B.E., Eck, M.J., Tenen, D.G., and Halmos, B. (2005). EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N. Engl. J. Med. *352*, 786–792.

Krumlauf, R. (1994). Hox genes in vertebrate development. Cell *78*, 191–201.

Kulis, M., and Esteller, M. (2010). DNA methylation and cancer. Adv. Genet. *70*, 27–56.

Lackey, A., and Donington, J.S. (2013). Surgical Management of Lung Cancer. Semin. Interv. Radiol. *30*, 133–140.

Laird, P.W. (2003). Early detection: The power and the promise of DNA methylation markers. Nat. Rev. Cancer *3*, 253–266.

Laird, N.M., and Ware, J.H. (1982). Random-effects models for longitudinal data. Biometrics 963–974.

Laird, P.W., and Jaenisch, R. (1996). The role of DNA methylation in cancer genetic and epigenetics. Annu. Rev. Genet. *30*, 441–464.

Lamprecht, B., Porsch, P., Pirich, C., and Studnicka, M. (2009). Electromagnetic navigation bronchoscopy in combination with PET-CT and rapid on-site cytopathologic examination for diagnosis of peripheral lung lesions. Lung *187*, 55–59.

Lander, E., Linton, L., Barren, B., and International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Langer, C.J., Besse, B., Gualberto, A., Brambilla, E., and Soria, J.-C. (2010). The evolving role of histology in the management of advanced non-small-cell lung cancer. J. Clin. Oncol. *28*, 5311–5320.

Lavagnini, I., and Magno, F. (2007). A statistical overview on univariate calibration, inverse regression, and detection limits: Application to gas chromatography/mass spectrometry technique. Mass Spectrom. Rev. *26*, 1–18.

Lebofsky, R., Decraene, C., Bernard, V., Kamal, M., Blin, A., Leroy, Q., Rio Frio, T., Pierron, G., Callens, C., Bieche, I., et al. (2015). Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types. Mol. Oncol. *9*, 783–790.

Lehmann-Werman, R., Neiman, D., Zemmour, H., Moss, J., Magenheim, J., Vaknin-Dembinsky, A., Rubertsson, S., Nellgård, B., Blennow, K., Zetterberg, H., et al. (2016). Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc. Natl. Acad. Sci. *113*, E1826–E1834.

Leng, S., Do, K., Yingling, C.M., Picchi, M.A., Wolf, H.J., Kennedy, T.C., Feser, W.J., Baron, A.E., Franklin, W.A., Brock, M.V., et al. (2012). Defining a Gene Promoter Methylation Signature in Sputum for Lung Cancer Risk Assessment. Clin. Cancer Res. *18*, 3387–3395.

Leontiou, C.A., Hadjidaniel, M.D., Mina, P., Antoniou, P., Ioannides, M., and Patsalis, P.C. (2015). Bisulfite Conversion of DNA: Performance Comparison of Different Kits and Methylation Quantitation of Epigenetic Biomarkers that Have the Potential to Be Used in Non-Invasive Prenatal Testing. PLOS ONE *10*, e0135058.

Lev Maor, G., Yearim, A., and Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. Trends Genet. TIG *31*, 274–280.

Liloglou, T., and Field, J.K. (2010). Detection of DNA Methylation Changes in Body Fluids. Adv. Genet. *71*, 177.

Liloglou, T., Bediaga, N.G., Brown, B.R.B., Field, J.K., and Davies, M.P.A. (2014). Epigenetic biomarkers in lung cancer. Cancer Lett. *342*, 200–212.

Ling, H., Vincent, K., Pichler, M., Fodde, R., Berindan-Neagoe, I., Slack, F.J., and Calin, G.A. (2015). Junk DNA and the long non-coding RNA twist in cancer genetics. Oncogene *34*, 5003–5011.

Liu, Q., Yu, Z., Yuan, S., Xie, W., Li, C., Hu, Z., Xiang, Y., Wu, N., Wu, L., Bai, L., et al. (2017a). Circulating exosomal microRNAs as prognostic biomarkers for non-small-cell lung cancer. Oncotarget *8*, 13048–13058.

Liu, Y., Lee, M.O., Wang, H.G., Li, Y., Hashimoto, Y., Klaus, M., Reed, J.C., and Zhang, X. (1996). Retinoic acid receptor beta mediates the growth-inhibitory effect of retinoic acid by promoting apoptosis in human breast cancer cells. Mol. Cell. Biol. *16*, 1138–1149.

Liu, Y., Song, C., Ladas, I., Fitarelli-Kiehl, M., and Makrigiorgos, G.M. (2017b). Methylation-sensitive enrichment of minor DNA alleles using a double-strand DNA-specific nuclease. Nucleic Acids Res. *45*, e39–e39.

Livak, K.J., Wills, Q.F., Tipping, A.J., Datta, K., Mittal, R., Goldson, A.J., Sexton, D.W., and Holmes, C.C. (2013). Methods for qPCR gene expression profiling applied to 1440 lymphoblastoid single cells. Methods *59*, 71–79.

Louie, A.V., Palma, D.A., Dahele, M., Rodrigues, G.B., and Senan, S. (2015). Management of early-stage non-small cell lung cancer using stereotactic ablative radiotherapy: Controversies, insights, and changing horizons. Radiother. Oncol. *114*, 138–147.

Louie, R.F., Tang, Z., Shelby, D.G., and Kost, G.J. (2000). Point-of-Care Testing: Millennium Technology for Critical Care. Lab. Med. *31*, 402–408.

Lynch, M.D., Smith, A.J.H., Gobbi, M.D., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L., et al. (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J. *31*, 317–329.

Mazières, J., Peters, S., Lepage, B., Cortot, A.B., Barlesi, F., Beau-Faller, M., Besse, B., Blons, H., Mansuet-Lupo, A., Urban, T., et al. (2013). Lung cancer that harbors an HER2 mutation: epidemiologic characteristics and therapeutic perspectives. J. Clin. Oncol. *31*, 1997–2003.

McElnay, P., and Lim, E. (2014). Adjuvant or neoadjuvant chemotherapy for NSCLC. J. Thorac. Dis. *6*, S224–S227.

McGranahan, N., and Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell *168*, 613–628.

McLeod, A.I., and Xu, C. (2017). bestglm: Best Subset GLM.

McPhail, S., Johnson, S., Greenberg, D., Peake, M., and Rous, B. (2015). Stage at diagnosis and early mortality from cancer in England. Br. J. Cancer *112*, S108–S115.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Merlo, A., Herman, J.G., Mao, L., Lee, D.J., Gabrielson, E., Burger, P.C., Baylin, S.B., and Sidransky, D. (1995). 5′ CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nat. Med. *1*, 686–692.

Millar, D., Christova, Y., and Holliger, P. (2015). A polymerase engineered for bisulfite sequencing. Nucleic Acids Res. *43*, e155–e155.

Mok, T.S., Wu, Y.-L., Thongprasert, S., Yang, C.-H., Chu, D.-T., Saijo, N., Sunpaweravong, P., Han, B., Margono, B., Ichinose, Y., et al. (2009). Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma. N. Engl. J. Med. *361*, 947–957.

Morgensztern, D., Campo, M.J., Dahlberg, S.E., Doebele, R.C., Garon, E., Gerber, D.E., Goldberg, S.B., Hammerman, P.S., Heist, R.S., Hensing, T., et al. (2015). Molecularly Targeted Therapies in Non–Small-Cell Lung Cancer Annual Update 2014. J. Thorac. Oncol. *10*, S1–S63.

Mouliere, F., Robert, B., Arnau Peyrotte, E., Del Rio, M., Ychou, M., Molina, F., Gongora, C., and Thierry, A.R. (2011). High Fragmentation Characterizes Tumour-Derived Circulating DNA. PLoS ONE *6*, e23418.

Mouliere, F., El Messaoudi, S., Gongora, C., Guedj, A.-S., Robert, B., Del Rio, M., Molina, F., Lamy, P.-J., Lopez-Crapez, E., Mathonnet, M., et al. (2013). Circulating Cell-Free DNA from Colorectal Cancer Patients May Reveal High KRAS or BRAF Mutation Load. Transl. Oncol. *6*, 319–328.

Mulshine, J.L., and van Klaveren, R.J. (2011). Lung cancer screening: what is the benefit and what do we do about it? Lung Cancer *71*, 247–248.

Murtaza, M., Dawson, S.-J., Tsui, D.W.Y., Gale, D., Forshew, T., Piskorz, A.M., Parkinson, C., Chin, S.-F., Kingsbury, Z., Wong, A.S.C., et al. (2013). Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature *497*, 108–112.

Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods Ecol. Evol. *4*, 133–142.

Narsule, C.K., Ebright, M.I., and Fernando, H.C. (2011). Sublobar versus lobar resection: current status. Cancer J. Sudbury Mass *17*, 23–27.

National Lung Screening Trial Research Team (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med *2011*, 395–409.

Nikolaidis, G., Raji, O.Y., Markopoulou, S., Gosney, J.R., Bryan, J., Warburton, C., Walshaw, M., Sheard, J., Field, J.K., and Liloglou, T. (2012). DNA Methylation Biomarkers Offer Improved Diagnostic Efficiency in Lung Cancer. Cancer Res. *72*, 5692–5701.

Novello, S., Barlesi, F., Califano, R., Cufer, T., Ekman, S., Levra, M.G., Kerr, K., Popat, S., Reck, M., Senan, S., et al. (2016). Metastatic non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Ann. Oncol. *27*, v1–v27.

Ohashi, K., Maruvka, Y.E., Michor, F., and Pao, W. (2013). Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor–Resistant Disease. J. Clin. Oncol. *31*, 1070–1080.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Ooki, A., Maleki, Z., Tsay, J.-C.J., Goparaju, C., Brait, M., Turaga, N., Nam, H.-S., Rom, W.N., Pass, H.I., Sidransky, D., et al. (2017). A Panel of Novel Detection and Prognostic Methylated DNA Markers in Primary Non–Small Cell Lung Cancer and Serum DNA. Clin. Cancer Res. *23*, 7141–7152.

Ortiz-Vega, S., Khokhlatchev, A., Nedwidek, M., Zhang, X., Dammann, R., Pfeifer, G.P., and Avruch, J. (2002). The putative tumor suppressor RASSF1A homodimerizes and heterodimerizes with the Ras-GTP binding protein Nore1. Oncogene *21*, 1381–1390.

Ostrow, K.L., Hoque, M.O., Loyo, M., Brait, M., Greenberg, A., Siegfried, J.M., Grandis, J.R., Gaither Davis, A., Bigbee, W.L., Rom, W., et al. (2010). Molecular Analysis of Plasma DNA for the Early Detection of Lung Cancer by Quantitative Methylation-Specific PCR. Clin. Cancer Res. *16*, 3463–3472.

Oxnard, G.R., Arcila, M.E., Chmielecki, J., Ladanyi, M., Miller, V.A., and Pao, W. (2011). New Strategies in Overcoming Acquired Resistance to Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors in Lung Cancer. Clin. Cancer Res. *17*, 5530–5537.

Oxnard, G.R., Paweletz, C.P., Kuang, Y., Mach, S.L., O'Connell, A., Messineo, M.M., Luke, J.J., Butaney, M., Kirschmeier, P., Jackman, D.M., et al. (2014). Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. Clin. Cancer Res. *20*, 1698–1705.

Page, K., Guttery, D.S., Fernandez-Garcia, D., Hills, A., Hastings, R.K., Luo, J., Goddard, K., Shahin, V., Woodley-Barker, L., Rosales, B.M., et al. (2017). Next Generation Sequencing of Circulating Cell-Free DNA for Evaluating Mutations and Gene Amplification in Metastatic Breast Cancer. Clin. Chem. *63*, 532–541.

Pao, W., Miller, V.A., Politi, K.A., Riely, G.J., Somwar, R., Zakowski, M.F., Kris, M.G., and Varmus, H. (2005). Acquired Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib Is Associated with a Second Mutation in the EGFR Kinase Domain. PLOS Med. *2*, e73.

Pepe, M.S., Etzioni, R., Feng, Z., Potter, J.D., Thompson, M.L., Thornquist, M., Winget, M., and Yasui, Y. (2001). Phases of biomarker development for early detection of cancer. J. Natl. Cancer Inst. *93*, 1054–1061.

Persaud, D., Gay, H., Ziemniak, C., Chen, Y.H., Piatak, M., Chun, T.-W., Strain, M., Richman, D., and Luzuriaga, K. (2013). Absence of detectable HIV-1 viremia after treatment cessation in an infant. N. Engl. J. Med. *369*, 1828–1835.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res. *29*, e45.

Pfeifer, G.P. (2010). Environmental exposures and mutational patterns of cancer genomes. Genome Med. *2*, 54.

Pinheiro, J.C., and Bates, D.M. (2000). Mixed-effects models in S and S-PLUS (New York, NY [u.a.]: Springer).

Pinheiro, L.B., Coleman, V.A., Hindson, C.M., Herrmann, J., Hindson, B.J., Bhat, S., and Emslie, K.R. (2012). Evaluation of a Droplet Digital Polymerase Chain Reaction Format for DNA Copy Number Quantification. Anal. Chem. *84*, 1003–1011.

Polz, M.F., and Cavanaugh, C.M. (1998). Bias in Template-to-Product Ratios in Multitemplate PCR. Appl. Environ. Microbiol. *64*, 3724–3730.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nat. Biotechnol. *28*, 1057–1068.

Postmus, P.E., Kerr, K.M., Oudkerk, M., Senan, S., Waller, D.A., Vansteenkiste, J., Escriu, C., Peters, S., and on behalf of the ESMO Guidelines Committee (2017). Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Ann. Oncol. *28*, iv1–iv21.

Potter, N.T., Hurban, P., White, M.N., Whitlock, K.D., Lofton-Day, C.E., Tetzner, R., Koenig, T., Quigley, N.B., and Weiss, G. (2014). Validation of a Real-Time PCR-Based Qualitative Assay for the Detection of Methylated SEPT9 DNA in Human Plasma. Clin. Chem. *60*, 1183–1191.

Pulverer, W., Wielscher, M., Panzer-Grümayer, R., Plessl, T., Kriegner, A., Vierlinger, K., and Weinhäusel, A. (2012). The stem cell signature of CHH/CHG methylation is not present in 271 cancer associated 5′UTR gene regions. Biochimie *94*, 2345–2352.

Quentien, M.H., Barlier, A., Franc, J.L., Pellegrini, I., Brue, T., and Enjalbert, A. (2006). Pituitary transcription factors: from congenital deficiencies to gene therapy. J. Neuroendocrinol. *18*, 633–642.

R Core Team (2017). R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna Austria.

Raizis, A.M., Schmitt, F., and Jost, J.P. (1995). A Bisulfite Method of 5-Methylcytosine Mapping That Minimizes Template Degradation. Anal. Biochem. *226*, 161–166.

Raynal, N.J.-M., Si, J., Taby, R.F., Gharibyan, V., Ahmed, S., Jelinek, J., Estecio, M.R.H., and Issa, J.-P.J. (2012). DNA Methylation Does Not Stably Lock Gene Expression but Instead Serves as a Molecular Mark for Gene Silencing Memory. Cancer Res. *72*, 1170–1181.

Reclusa, P., Sirera, R., Araujo, A., Giallombardo, M., Valentino, A., Sorber, L., Bazo, I.G., Pauwels, P., and Rolfo, C. (2016). Exosomes genetic cargo in lung cancer: a truly Pandora's box. Transl. Lung Cancer Res. *5*, 483–491.

Reclusa, P., Taverna, S., Pucci, M., Durendez, E., Calabuig, S., Manca, P., Serrano, M.J., Sober, L., Pauwels, P., Russo, A., et al. (2017). Exosomes as diagnostic and predictive biomarkers in lung cancer. J. Thorac. Dis. *9*, S1373–S1382.

Redshaw, N., Huggett, J.F., Taylor, M.S., Foy, C.A., and Devonshire, A.S. (2014). Quantification of epigenetic biomarkers: an evaluation of established and emerging methods for DNA methylation analysis. BMC Genomics *15*, 1174.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425–432.

Reik, W., and Lewis, A. (2005). Co-evolution of X-chromosome inactivation and imprinting in mammals. Nat. Rev. Genet. *6*, 403–410.

Rideout, W.M., Coetzee, G.A., Olumi, A.F., and Jones, P.A. (1990). 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. Science *249*, 1288–1290.

Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., et al. (2007). Global Survey of Phosphotyrosine Signaling Identifies Oncogenic Kinases in Lung Cancer. Cell *131*, 1190–1203.

Rivera, M.P., Mehta, A.C., and Wahidi, M.M. (2013). Establishing the diagnosis of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest *143*, e142S-e165S.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics *12*, 77.

Rogers-Broadway, K.-R., and Karteris, E. (2015). Amplification efficiency and thermal stability of qPCR instrumentation: Current landscape and future perspectives. Exp. Ther. Med. *10*, 1261–1264.

Rosell, R., Carcereny, E., Gervais, R., Vergnenegre, A., Massuti, B., Felip, E., Palmero, R., Garcia-Gomez, R., Pallares, C., Sanchez, J.M., et al. (2012). Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol. *13*, 239–246.

Sacher, A.G., Paweletz, C., Dahlberg, S.E., Alden, R.S., O'Connell, A., Feeney, N., Mach, S.L., Jänne, P.A., and Oxnard, G.R. (2016). Prospective validation of rapid plasma genotyping as a sensitive and specific tool for guiding lung cancer care. JAMA Oncol. *2*, 1014–1022.

Sandoval, J., Mendez-Gonzalez, J., Nadal, E., Chen, G., Carmona, F.J., Sayols, S., Moran, S., Heyn, H., Vizoso, M., and Gomez, A. (2013a). A prognostic DNA methylation signature for stage I non–small-cell lung cancer. J. Clin. Oncol. *31*, 4140–4147.

Sandoval, J., Peiró-Chova, L., Pallardó, F.V., and García-Giménez, J.L. (2013b). Epigenetic biomarkers in laboratory diagnostics: emerging approaches and opportunities. Expert Rev. Mol. Diagn. *13*, 457–471.

Sankaranarayanan, R., Swaminathan, R., Jayant, K., and Brenner, H. (2011). An overview of cancer survival in Africa, Asia, the Caribbean and Central America: the case for investment in cancer health services. IARC Sci. Publ. 257–291.

Schilling, E., and Rehli, M. (2007). Global, comparative analysis of tissue-specific promoter CpG methylation. Genomics *90*, 314–323.

Schmidt, B., Liebenberg, V., Dietrich, D., Schlegel, T., Kneip, C., Seegebarth, A., Flemming, N., Seemann, S., Distler, J., and Lewin, J. (2010). SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer based on bronchial aspirates. BMC Cancer *10*, 600.

Schmitt, A.M., and Chang, H.Y. (2016). Long Noncoding RNAs in Cancer Pathways. Cancer Cell *29*, 452–463.

Schuetzenmeister, A. (2017). VCA: Variance Component Analysis.

Schwarzenbach, H., Hoon, D.S.B., and Pantel, K. (2011). Cell-free nucleic acids as biomarkers in cancer patients. Nat. Rev. Cancer *11*, 426–437.

Sedlak, R.H., Cook, L., Cheng, A., Magaret, A., and Jerome, K.R. (2014). Clinical Utility of Droplet Digital PCR for Human Cytomegalovirus. J. Clin. Microbiol. *52*, 2844–2848.

Self, S.G., and Liang, K.-Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. J. Am. Stat. Assoc. *82*, 605–610.

Senan, S., Paul, M.A., and Lagerwaard, F.J. (2013). Treatment of early-stage lung cancer detected by screening: surgery or stereotactic ablative radiotherapy? Lancet Oncol. *14*, e270–e274.

Seol, H.S., Akiyama, Y., Shimada, S., Lee, H.J., Kim, T.I., Chun, S.M., Singh, S.R., and Jang, S.J. (2014). Epigenetic silencing of microRNA-373 to epithelial-mesenchymal transition in non-small cell lung cancer through IRAK2 and LAMP1 axes. Cancer Lett. *353*, 232–241.

Sharma, S., Kelly, T.K., and Jones, P.A. (2010). Epigenetics in cancer. Carcinogenesis *31*, 27–36.

Shaw, J.A., Guttery, D.S., Hills, A., Fernandez-Garcia, D., Page, K., Rosales, B.M., Goddard, K.S., Hastings, R.K., Luo, J., Ogle, O., et al. (2017). Mutation Analysis of Cell-Free DNA and Single Circulating Tumor Cells in Metastatic Breast Cancer Patients with High Circulating Tumor Cell Counts. Clin. Cancer Res. *23*, 88–96.

Soda, M., Choi, Y.L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S., Watanabe, H., Kurashina, K., Hatanaka, H., et al. (2007). Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. Nature *448*, 561–566.

Sozzi, G., Roz, L., Conte, D., Mariani, L., Andriani, F., Verderio, P., and Pastorino, U. (2005). Effects of Prolonged Storage of Whole Plasma or Isolated Plasma DNA on the Results of Circulating DNA Quantification Assays. JNCI J. Natl. Cancer Inst. *97*, 1848–1850.

Sozzi, G., Boeri, M., Rossi, M., Verri, C., Suatoni, P., Bravi, F., Roz, L., Conte, D., Grassi, M., Sverzellati, N., et al. (2014). Clinical Utility of a Plasma-Based miRNA Signature Classifier Within Computed Tomography Lung Cancer Screening: A Correlative MILD Trial Study. J. Clin. Oncol. *32*, 768–773.

Ståhlberg, A., and Kubista, M. (2014). The workflow of single-cell expression profiling using quantitative real-time PCR. Expert Rev. Mol. Diagn. *14*, 323–331.

Stoffel, M.A., Nakagawa, S., and Schielzeth, H. (2017). rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods Ecol. Evol. *8*, 1639–1644.

Strain, M.C., Lada, S.M., Luong, T., Rought, S.E., Gianella, S., Terry, V.H., Spina, C.A., Woelk, C.H., and Richman, D.D. (2013). Highly precise measurement of HIV DNA by droplet digital PCR. PloS One *8*, e55943.

Stram, D.O., and Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects model. Biometrics *50*, 1171–1177.

Sun, K., Jiang, P., Chan, K.C.A., Wong, J., Cheng, Y.K.Y., Liang, R.H.S., Chan, W., Ma, E.S.K., Chan, S.L., Cheng, S.H., et al. (2015). Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc. Natl. Acad. Sci. *112*, E5503–E5512.

Suzuki, T., Ohsumi, S., and Makino, K. (1994). Mechanistic studies on depurination and apurinic site chain breakage in oligodeoxyribonucleotides. Nucleic Acids Res. *22*, 4997–5003.

Tamkovich, S.N., Bryzgunova, O.E., Rykova, E.Y., Permyakova, V.I., Vlassov, V.V., and Laktionov, P.P. (2005). Circulating Nucleic Acids in Blood of Healthy Male and Female Donors. Clin. Chem. *51*, 1317–1319.

Tanay, A., O'Donnell, A.H., Damelin, M., and Bestor, T.H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. Proc. Natl. Acad. Sci. *104*, 5521–5526.

Tang, F. (2006). MicroRNA expression profiling of single whole embryonic stem cells. Nucleic Acids Res. *34*, e9–e9.

Tartarone, A., Giordano, P., Lerose, R., Rodriquenz, M.G., Conca, R., and Aieta, M. (2017). Progress and challenges in the treatment of small cell lung cancer. Med. Oncol. *34*, 110.

Taylor, S.C., Carbonneau, J., Shelton, D.N., and Boivin, G. (2015). Optimization of Droplet Digital PCR from RNA and DNA extracts with direct comparison to RT-qPCR: Clinical implications for quantification of Oseltamivir-resistant subpopulations. J. Virol. Methods *224*, 58–66.

Teo, I.A., Choi, J.W., Morlese, J., Taylor, G., and Shaunak, S. (2002). LightCycler qPCR optimisation for low copy number target DNA. J. Immunol. Methods *270*, 119–133.

Thakur, B.K., Zhang, H., Becker, A., Matei, I., Huang, Y., Costa-Silva, B., Zheng, Y., Hoshino, A., Brazier, H., and Xiang, J. (2014). Double-stranded DNA in exosomes: a novel biomarker in cancer detection. Cell Res. *24*, 766–769.

Thierry, A.R., Mouliere, F., Gongora, C., Ollier, J., Robert, B., Ychou, M., Del Rio, M., and Molina, F. (2010). Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. Nucleic Acids Res. *38*, 6159–6175.

Thierry, A.R., El Messaoudi, S., Gahan, P.B., Anker, P., and Stroun, M. (2016). Origins, structures, and functions of circulating DNA in oncology. Cancer Metastasis Rev. *35*, 347–376.

Tichopad, A., Kitchen, R., Riedmaier, I., Becker, C., Ståhlberg, A., and Kubista, M. (2009). Design and Optimization of Reverse-Transcription Quantitative PCR Experiments. Clin. Chem. *55*, 1816–1823.

Torre, L.A., Siegel, R.L., Ward, E.M., and Jemal, A. (2016). Global Cancer Incidence and Mortality Rates and Trends—An Update. Cancer Epidemiol. Prev. Biomark. *25*, 16–27.

Tost, J. (2016). Current and Emerging Technologies for the Analysis of the Genome-Wide and Locus-Specific DNA Methylation Patterns. In DNA Methyltransferases - Role and Function, A. Jeltsch, and R.Z. Jurkowska, eds. (Cham: Springer International Publishing), pp. 343–430.

Tost, J., and Gut, I.G. (2007). DNA methylation analysis by pyrosequencing. Nat. Protoc. *2*, 2265–2275.

Toumazou, C., Shepherd, L.M., Reed, S.C., Chen, G.I., Patel, A., Garner, D.M., Wang, C.-J.A., Ou, C.-P., Amin-Desai, K., Athanasiou, P., et al. (2013). Simultaneous DNA amplification and detection using a pH-sensing semiconductor system. Nat. Methods *10*, 641–646.

Travis, W.D., Brambilla, E., and Riely, G.J. (2013a). New Pathologic Classification of Lung Cancer: Relevance for Clinical Practice and Clinical Trials. J. Clin. Oncol. *31*, 992–1001.

Travis, W.D., Brambilla, E., Noguchi, M., Nicholson, A.G., Geisinger, K., Yatabe, Y., Ishikawa, Y., Wistuba, I., Flieder, D.B., Franklin, W., et al. (2013b). Diagnosis of lung cancer in small biopsies and cytology: implications of the 2011 International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification. Arch. Pathol. Lab. Med. *137*, 668–684.

Uehiro, N., Sato, F., Pu, F., Tanaka, S., Kawashima, M., Kawaguchi, K., Sugimoto, M., Saji, S., and Toi, M. (2016). Circulating cell-free DNA-based epigenetic assay can detect early breast cancer. Breast Cancer Res. *18*.

Underhill, H.R., Kitzman, J.O., Hellwig, S., Welker, N.C., Daza, R., Baker, D.N., Gligorich, K.M., Rostomily, R.C., Bronner, M.P., and Shendure, J. (2016). Fragment Length of Circulating Tumor DNA. PLOS Genet. *12*, e1006162.

Vanni, I., Alama, A., Grossi, F., Dal Bello, M.G., and Coco, S. (2017). Exosomes: a new horizon in lung cancer. Drug Discov. Today *22*, 927–936.

Venables, W.N., and Ripley, B.D. (2002). Modern Applied Statistics with S (New York: Springer).

Vidal, E., Sayols, S., Moran, S., Guillaumet-Adkins, A., Schroeder, M.P., Royo, R., Orozco, M., Gut, M., Gut, I., Lopez-Bigas, N., et al. (2017). A DNA methylation map of human cancer at single base-pair resolution. Oncogene *36*, 5648–5657.

Vuillemenot, B.R., Pulling, L.C., Palmisano, W.A., Hutt, J.A., and Belinsky, S.A. (2004). Carcinogen exposure differentially modulates RAR-β promoter hypermethylation, an early and frequent event in mouse lung carcinogenesis. Carcinogenesis *25*, 623–629.

Vynck, M., Vandesompele, J., and Thas, O. (2017). Quality control of digital PCR assays and platforms. Anal. Bioanal. Chem. *409*, 5919–5931.

Wain, H.M., Bruford, E.A., Lovering, R.C., Lush, M.J., Wright, M.W., and Povey, S. (2002). Guidelines for human gene nomenclature. Genomics *79*, 464–470.

Wakelee, H.A., Dahlberg, S.E., Keller, S.M., Tester, W.J., Gandara, D.R., Graziano, S.L., Adjei, A.A., Leighl, N.B., Butts, C.A., Aisner, S.C., et al. (2016). E1505: Adjuvant chemotherapy +/- bevacizumab for early stage NSCLC—Outcomes based on chemotherapy subsets. J. Clin. Oncol. *34*, 8507–8507.

Wang, B.G., Huang, H.-Y., Chen, Y.-C., Bristow, R.E., Kassauei, K., Cheng, C.-C., Roden, R., Sokoll, L.J., Chan, D.W., and Shih, I.-M. (2003). Increased plasma DNA integrity in cancer patients. Cancer Res. *63*, 3966–3968.

Wang, R., Hu, H., Pan, Y., Li, Y., Ye, T., Li, C., Luo, X., Wang, L., Li, H., Zhang, Y., et al. (2012). RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *30*, 4352–4359.

Wang, Y., Springer, S., Mulvey, C.L., Silliman, N., Schaefer, J., Sausen, M., James, N., Rettig, E.M., Guo, T., Pickering, C.R., et al. (2015). Detection of somatic mutations and HPV in the saliva and plasma of patients with head and neck squamous cell carcinomas. Sci. Transl. Med. *7*, 293ra104-293ra104.

Warton, K., and Samimi, G. (2015). Methylation of cell-free circulating DNA in the diagnosis of cancer. Front. Mol. Biosci. *2*.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. *39*, 457–466.

Weisenberger, D.J. (2005). Analysis of repetitive element DNA methylation by MethyLight. Nucleic Acids Res. *33*, 6823–6836.

Weiss, G., Schlegel, A., Kottwitz, D., König, T., and Tetzner, R. (2017). Validation of the SHOX2/PTGER4 DNA Methylation Marker Panel for Plasma-Based Discrimination between Patients with Malignant and Nonmalignant Lung Disease. J. Thorac. Oncol. *12*, 77–84.

Whale, A.S., Huggett, J.F., Cowen, S., Speirs, V., Shaw, J., Ellison, S., Foy, C.A., and Scott, D.J. (2012). Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. Nucleic Acids Res. *40*, e82–e82.

Wilm, B., and Muñoz-Chapuli, R. (2016). The Role of WT1 in Embryonic Development and Normal Organ Homeostasis. Methods Mol. Biol. Clifton NJ *1467*, 23–39.

Winham, S. (2012). MDR: Detect gene-gene interactions using multifactor dimensionality reduction.

Winham, S.J., and Motsinger-Reif, A.A. (2011). An R package implementation of multifactor dimensionality reduction. BioData Min. *4*, 24.

Wong, M.C.S., Lao, X.Q., Ho, K.-F., Goggins, W.B., and Tse, S.L.A. (2017). Incidence and mortality of lung cancer: global trends and association with socioeconomic status. Sci. Rep. *7*.

Wu, H., Caffo, B., Jaffee, H.A., Irizarry, R.A., and Feinberg, A.P. (2010). Redefining CpG islands using hidden Markov models. Biostat. Oxf. Engl. *11*, 499–514.

Yang, J.C.-H., Ahn, M.-J., Kim, D.-W., Ramalingam, S.S., Sequist, L.V., Su, W.-C., Kim, S.-W., Kim, J.-H., Planchard, D., Felip, E., et al. (2017). Osimertinib in Pretreated T790M-Positive Advanced Non-Small-Cell Lung Cancer: AURA Study Phase II Extension Component. J. Clin. Oncol. *35*, 1288–1296.

Yu, H.A., Arcila, M.E., Rekhtman, N., Sima, C.S., Zakowski, M.F., Pao, W., Kris, M.G., Miller, V.A., Ladanyi, M., and Riely, G.J. (2013). Analysis of Tumor Specimens at the Time

of Acquired Resistance to EGFR-TKI Therapy in 155 Patients with EGFR-Mutant Lung Cancers. Clin. Cancer Res. *19*, 2240–2247.

Yu, M., Carter, K.T., Makar, K.W., Vickers, K., Ulrich, C.M., Schoen, R.E., Brenner, D., Markowitz, S.D., and Grady, W.M. (2015). MethyLight droplet digital PCR for detection and absolute quantification of infrequently methylated alleles. Epigenetics *10*, 803–809.

Zappa, C., and Mousa, S.A. (2016). Non-small cell lung cancer: current treatment and future advances. Transl. Lung Cancer Res. *5*, 288–300.

Zhang, Y., Wang, R., Song, H., Huang, G., Yi, J., Zheng, Y., Wang, J., and Chen, L. (2011). Methylation of multiple genes as a candidate biomarker in non-small cell lung cancer. Cancer Lett. *303*, 21–28.

Zuur, A.F. (2009). Mixed effects models and extensions in ecology with R. [electronic book]. (New York ; Springer, 2009.).