NOVEL ULTRA-SENSITIVE DIGITAL PCR ASSAYS FOR SCREENING AND DETECTION OF RARE MISSENSE MUTATIONS IN (PROTO-)ONCOGENES

by

AREZOO (ROZA) BIDSHAHRI

B.Sc. (Hons.), The University of British Columbia, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Biomedical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2017

Abstract

Somatic mutations can lead to cancer, often by altering the activity of kinases within signaling pathways that control cell growth and proliferation. Targeted cancer therapeutics are designed and used to regulate these aberrant signaling pathways in cases where somatic mutations within kinase genes predict a positive patient response to those treatments. For example, the V600E mutation in *BRAF*, the gene coding for the BRAF serine threonine kinase, predicts the effectiveness of vemurafenib in treating metastatic melanoma, while the mutational status of codons G12/G13 in the *KRAS* gene predicts likely colorectal cancer patient response to the monoclonal antibody (mAb) cetuximab.[1-3] However, FDA approved assays currently used to detect missense mutations in *BRAF* V600 and *KRAS* G12/G13 are not capable of detecting clinically actionable mutations at mutational frequencies low enough to permit their robust application to early disease detection or minimal residual disease monitoring. Moreover, detection of all clinically actionable missense mutations is not certain or generally achieved, in part due to limitations to assay specificities and the inability to unequivocally discriminate missense mutations from synonymous germline sequence variations.

This thesis addresses that limitation through the development and validation of a novel platform for creating highly sensitive assays against all possible missense mutations in an oncogenic hotspot codon or adjacent set of hotspot codons that ameliorates the known limitations to current FDA-approved assays. The platform is designed to enable development of assays against all possible missense mutations in oncogenic hotspots and, if required, unequivocally differentiate them from synonymous germline alleles. It utilizes droplet digital PCR (ddPCR) technology and chimeric wild-type specific LNA/DNA probes to create a novel "WT-negative" screening paradigm. The platform is applied to the creation of two new assays of potential clinical use in cancer diagnostics and theranostics. The first provides a reliable and sensitive screening and detection of all known clinically actionable mutations in *BRAF* V600, and the second achieves the same for *KRAS* G12/G13. Both assays show complete diagnostic accuracy when applied to formalin-fixed paraffin-embedded (FFPE) tumor specimens from metastatic colorectal cancer patients deficient for Mut L homologue-1.

Lay Summary

The aim of this thesis work was to develop a novel diagnostic test that would detect all possible mutations at a cancer hotspot gene. This type of cancer test can be utilized prior to treating patients to determine their response to targeted cancer therapies.

As an example, melanoma patients carrying a mutation in codon 600 of the *BRAF* gene are responsive to *BRAF* inhibitors, such as Vemurafenib. Unfortunately, current clinical *BRAF* tests are only able to identify patients with the most common *BRAF* V600 mutations, such as *BRAF* V600E or V600K, while patients with a more rare mutation will not be identified using current tests and will not receive proper treatment.

This thesis work has developed a sensitive, rapid, and cheap screening test that correctly stratifies patients into wild-type from those carrying any possible mutation. This novel paradigm will ensure that all patients receive the appropriate treatment for them.

Preface

A version of Chapter 3 has been published in the ***Journal of Molecular Diagnostics***:
Bidshahri R, *et al.* (2015) Quantitative Detection and Resolution of *BRAF* V600 Status in Colorectal Cancer Using Droplet Digital PCR and a Novel Wild-Type Negative Assay.

I performed all of the research, with insights provided by Drs. Curtis Hughesman and Charles Haynes, and in collaboration with the pathology team at the Canadian Immunohistochemistry Quality Control (cIQc) who provided clinical FFPE-stabilized colorectal tumor samples. Kelly McNeil provided training on the extraction of DNA from the FFPE cores at the Department of Genetics and Molecular Diagnostics at BC Cancer Agency. I also co-developed novel software for rigorous statistical analysis of data from the digital PCR assay in collaboration with Dean Attali, Dr. Haynes and Dr. Jenny Bryan. In addition, I drafted the initial manuscript, with further contributions to it made by Dr. Charles Haynes. Valuable input from Dr. Aly Karsan was also received prior to submission for publication.

A version of Chapter 4 from this thesis has been submitted to a journal and is under review as:
Bidshahri R, *et al.* (2017) Quantitative analysis of *KRAS* G12/G13 status in colorectal cancer using a novel wild-type negative assay that unequivocally differentiates missense and synonymous alleles.

I performed all of the research, with insights provided by Drs. Curtis Hughesman and Charles Haynes, and tumor specimens used for assay validation provided by cIQc. In addition, I drafted the initial manuscript, with further contributions to it made by Dr. Charles Haynes.

Table of Contents

List of Tables

List of Figures

List of Symbols

ΔH	Change in enthalpy; kcal mol^{-1}
ΔS	Change in entropy; cal mol^{-1} K^{-1}
ΔG	Change in Gibb's free energy change; kcal mol^{-1}
ΔC_p	Change in heat capacity; cal mol^{-1} K^{-1}
K	Equilibrium constant
C_T	Total strand concentration; M
T_a	Annealing temperature; °C
T_m	Melting temperature; °C
T_{ref}	Reference temperature defined as 53 °C
θ	Fraction of total droplets having a positive end-point fluorescence

List of Abbreviations

A	Adenine
APC	Adenomatous polyposis coli
ARMS	Amplification refractory mutation system
AS	Allele specific
Bp	Base pair
BRAF	v-*raf* murine sarcoma viral oncogene homolog B1
C	Cytosine
CAST-PCR	Competitive allele specific Taq-Man® PCR
Cf-DNA	Cell free-DNA
CIMP	CpG island methylation phenotype
cIQc	Canadian Immunohistochemistry Quality Control
COSMIC	Catalogue of Somatic Mutations in Cancer
CPD	Copies per droplet
C_q	Quantification cycle
CRC	Colorectal carcinoma
CREB	Clinical ethics research board
Dab	Dabcyl
dNTP	Deoxynucleotide triphosphate
dPCR	Digital polymerase chain reaction
EGFR	Epidermal growth factor receptor
ERK	Extracellular signal-regulated kinase
FAM	6-carboxyfluorescein
FFPE	Formalin fixed paraffin embedded
FP	Forward primer
G	Guanine
gDNA	Genomic deoxyribonucleic acid
GDP	Guanosine diphosphate
GTP	Guanosine triphosphate

HEX	hexachloro-fluorescein
HRM	High resolution melt
IABkFQ	Iowa black fluorescence quencher
IHC	Immunohistochemistry
KRAS	Kirsten rat sarcoma
LNA	Locked nucleic acid
LOB	Limit of blank
LOD	Limit of detection
LOH	Loss of heterozygosity
LOQ	Limit of quantitation
LS	Lynch syndrome
mAb	Monoclonal antibody
MAPK	Mitogen activated protein kinase
MF	Mutant frequency
MGB	Minor groove binder
MLH1	Mut L homologue
MMR	Mismatch repair
MSI	Microsatellite instability
MT	Mutant
NCCN	National comprehensive cancer network
NGS	Next generation sequencing
NNT	Nearest neighbor thermodynamic
nt	Nucleotide
NTC	No template control
PCR	Polymerase chain reaction
qPCR	Quantitative real-time polymerase chain reaction
RAF	Rapidly accelerated fibrosarcoma
RP	Reverse primer
RTK	Receptor tyrosine kinase
SD	Standard deviation
SPM	Single point mutation

SSA	Sessile serrated adenoma
T	Thymine
TMA	Tumor microaaray
TP53	Tumor suppressor protein 53
UVM	UV monitored melt
VCH	Vancouver Coastal Health
WT	Wild-type

Acknowledgements

First and foremost I want to thank my supervisor Dr. Charles Haynes. It has been an honor to be one of his PhD students. I would have not been the scientist I am today without him. He thought me how to take one-step at a time, how to listen, and how to crystalize my ideas. He made me aware of my strengths and weaknesses, and ensured I work hard to strengthen both. He supported me beyond his role as a supervisor, and helped me as I transitioned from his lab to industry. I began working prior to completing my degree, and his tremendous support allowed me to complete my PhD thesis. Chip, without you I would have not made it so far. I will always be grateful.

Further thanks to the Haynes lab members, Dr. Louise Creagh for her endless support. I owe particular thanks to Dr. Curtis Hughesman for providing me knowledge about the thermodynamics of LNA probe design, which made the wild-type negative assay possible. I also want to thank Kareem Fakhfakh for his support in conducting ultraviolet melt analysis of the probes and for the stimulating discussions that made problem solving more enjoyable. Dr. Louise Lund and Dr. Eric Ouellet thank you for your moral support and friendship during all my PhD years. Dr. Jenny Bryan, thanks for exposing me to the world of R programming and Dean Attali, I really enjoyed collaborating with you and creating the ddPCR data analysis software tool pack.

My sincere thanks goes to Dr. Jennifer Won and everyone at cIQc for providing me the colorectal cancer FFPE specimen that allowed me to apply the wild-type negative assay to real cancer patient sample. Also thank you Kelly McNeil and Dr. Aly Karsan at the Department of Genetics and Molecular Diagnostics at BC Cancer Agency for their feedback on my assays.

I also want to thank Dr. Andre Marziali, my co-supervisor, for his support through tough times and for giving me a chance to experience being part of a start-up company at Boreal Genomics. He exposed me to the area of cancer diagnostics and he had a great role in shaping the topic of my PhD thesis. Through Boreal, I also learned how to conduct research in an industry setting. Further thanks for Dr. Karen Cheung for her time and feedback as a committee member. I would also like to thank the MITACS accelerate program for making such an internship possible.

Special thanks to my mom for inspiring me to pursue higher education, my dad for teaching me to be persistent and always follow my passion, and my brothers Roozbeh and Ramtin who always reminded me to push hard.

Last but not least, thank you to my husband Amir Ekhterai Sanai. Without you, I would have not started my PhD and without you, I would have not completed it. Our marriage and my PhD commenced at the same time and you were my companion in both.

Dedication

I dedicate this thesis to

To my family, for all the sacrifices they made to ensure I receive the best education possible.

To Amir, for being by my side in every step of my doctorate degree.

To Myrna, my daughter, for being with me as I conducted my final experiments. I hope this will one day inspire you.

I love you all dearly.

Chapter 1: Introduction

1.1 Thesis Motivation and Overview

In the United States (U.S.), one in four deaths is due to cancer, making it the second-leading cause of mortality.[4] Intense efforts within healthcare and scientific communities are therefore being directed toward advancing both the diagnosis and treatment of cancer, and within the past quarter century progress toward understanding and treating cancers has served to decrease overall incidences and improve patient survival.[5] But the costs of cancer diagnosis, treatment, and residual disease monitoring are rising at rates exceeding inflation. The National Cancer Institute (NCI) has projected the total annual cost for cancer-related care to rise to $173 billion by 2020, which represents a 39% increase from documented costs in 2010.[6] The U.S. views such increases as unsustainable, yet still seeks to improve quality of patient care and survival. It is hoped that progress in addressing these opposing challenges might be realized through a personalized approach to risk monitoring, diagnosis and treatment of cancer.[7] High-throughput (HT) sequencing of specific cancers is yielding an ever-improving understanding of the molecular events that initiate and accelerate malignant transformation and cancer progression. Those events often involve somatic (acquired) genetic mutations, copy number variations, and chromosomal translocations that drive the oncogenesis.[8] Many of the somatic mutations affect kinase activity. Targeted cancer therapeutics are therefore often designed to regulate signaling pathways whose activity has been altered by missense (i.e. nonsynonymous) mutations in pathway-associated kinases or downstream effectors, though modulation of molecular events impacting DNA metabolism and repair is receiving significant attention as well.[9]

This in turn is driving technological advances that improve the ability of private and hospital-associated cancer genomics clinics to define the mutation and copy-number status of key genetic biomarkers in a robust and cost-effective manner. For the case of missense mutation detection, Sanger sequencing may be considered a benchmark technology,[10] as it can provide a detailed assessment of somatic variations within an oncogene or defined set of oncogenes,[11] but at a throughput and cost that challenge the sustained operation of any testing clinic. Alternative mutation-analysis technologies have therefore been developed to accurately stratify cancer

patients in a more rapid and affordable manner. For detection and quantification of somatic mutations, they include restriction-enzyme based analyses,[12] antibody-based histochemical analyses,[13, 14] pyro-sequencing and other next-generation sequencing technologies,[15] high resolution melt (HRM) analyses,[16, 17] and PCR-based methods utilizing allele-specific (AS) primers and/or probes.[10, 18, 19] These powerful methods are not only being applied to cancer diagnosis and disease staging, they have become integral to evaluating cancer risk and treatment response, setting of proper courses of therapy, and post-treatment monitoring of patients.[20]

Due to its unique combination of sensitivity, specificity, speed, and low risk of contamination,[21] AS-PCR is now the platform technology most widely employed in clinics to detect and quantify somatic mutations in specific genes and genetic biomarkers. Although the detection of mutant frequencies below 0.1% has been achieved for certain targets, AS-PCR assays are generally capable of detecting a point mutation in genomic DNA (gDNA) purified from a tissue specimen when the frequency of that mutation relative to the background abundance of the parent germline allele equals or exceeds 1%.[10, 22, 23] 30 ng sample for AS-PCR analysis drawn from a standard tissue biopsy typically yields *ca.* 10,000 to 15,000 amplifiable copies of the gene of interest; that sample would therefore need to carry at least 100 copies of the mutation to permit its detection. This level of sensitivity is at times adequate for defining the proper initial course of therapy for cancer patients,[22-24] but has either precluded or limited the use of AS-PCR methods in detecting early-stage pre-malignant cancers or in post-treatment minimal-residual-disease monitoring.

Selectively amplifying and detecting a single copy of a mutant allele in a sample containing a high background of the parent germline allele is inherently difficult, challenged in part by cross-priming and/or probe cross-hybridization events that generate false positives.[19] Chemical modification of primers and probes, including through incorporation of locked nucleic acids (LNAs) – an RNA analog that contains a methylene bridge between the 2'O and 4'C of the ribose sugar – has been shown to improve their performance when used in PCR assays against rare somatic mutations. LNAs improve base-pair stability and mismatch discrimination, allowing for shorter and more specific primer and probe designs.[25-27] Significant improvements to the specificity and sensitivity of PCR assays against cancer-related somatic mutations may also be realized through the use of digital PCR (dPCR) formats[28, 29] that partition individual mutant

alleles into individual sub-nL droplets to permit their high-fidelity amplification by minimizing or eliminating cross-priming and cross-hybridization events.

This thesis leverages these advances to create a new diagnostic platform that utilizes dPCR and novel LNA-modified probes to reliably detect missense mutations present at frequencies of 0.05% or better. A molecular thermodynamic model capable of predicting the thermal stabilities of complementary and mismatched duplexes containing LNA substitutions[30-33] is described and used to design the probes needed to create two novel ultra-sensitive dPCR assays for detecting all cancer-relevant missense mutations within the "hotspot" codon V600 of the *BRAF* proto-oncogene, and codons G12/G13 of the *KRAS* oncogene, respectively. Each of these assays uses dPCR to segregate mutant alleles from wild-type (germline) forms of the gene to permit their accurate detection and quantification, and to thereby enable rapid and robust stratification of cancer patients, most notably colorectal cancer and metastatic melanoma patients. Finally, to facilitate the adoption of digital PCR in cancer clinics (where it is not currently used), novel software for rigorous statistical analysis of data from dPCR assays against driver mutations is described and shown to provide an objective means to make clinically reliable calls on the genomic variations specific to a particular patient's cancer. This fills a crucial application gap, as software for analyzing dPCR data generated from patient specimens was previously not available.

1.2 Somatic Mutations and Personalized Cancer Care

Personalized medicine has as a founding principle the molecular testing of clinically validated biomarkers that are of prognostic or predictive (theranostic) value.[5, 34-36] These markers may be of many forms, but are often identified within genomic DNA, transcripts and proteins of patients. Prognostic biomarkers correlate with clinical outcomes, including survival rates, independent of or following treatment. Predictive biomarkers, which are the primary focus of this thesis, drive clinical decisions by assessing the benefit of a possible treatment option to a specific patient. They are therefore used to forecast the effect of a drug on a disease, while prognostic biomarkers are used to define and stratify the disease present in the patient. Biomarker-based disease detection, patient stratification and therapeutic selection can thereby serve to reduce unnecessary or improper treatment, and decrease morbidity.

Biomarker discoveries and associated molecular diagnostics technologies are rapidly advancing as the molecular mechanisms that transform a normal cell to a malignant state become better understood.[37, 38] In oncology, those transformations are often associated with somatic missense mutations in kinases that deregulate signaling pathways, most notably the mitogen activated protein kinase (MAPK) pathway.[39] Indeed, many of the most useful predictive biomarkers in clinical molecular oncology identify mutations within that pathway, as exemplified by the V600E mutation in *BRAF* that predicts the effectiveness of vemurafenib in treating metastatic melanoma, and the mutational status of *KRAS* G12/G13 that predicts likely colorectal cancer patient response to the monoclonal antibody (mAb) cetuximab.[1-3] Understanding the MAPK pathway and the missense mutations that can alter its activity and pathology is therefore essential to establishing improved biomarker-based disease treatment and monitoring, as well as patient stratification.

1.3 The MAPK Pathway and the Role of RAS/RAF Mutations in Cancer

The MAPK/ERK (also known as the RAS-RAF-MEK-ERK) pathway (**Figure 1.1**) is the best studied of the six known mammalian MAPK pathways, and is deregulated in one-third of all human cancers. It is comprised of a series of evolutionarily conserved enzymes (mostly kinases) that link extracellular signals to regulatory transcription factors that control important cellular processes such as growth, proliferation, differentiation, migration, and apoptosis.[40] In healthy cells, MAPK stimulation starts with binding of an external ligand (e.g. a member of the epidermal growth factor family) to receptor tyrosine kinases (RTKs), including epidermal growth factor receptors (EGFRs). Upon stimulation, RTKs interact to form receptor dimers, which acts to change the conformation of the cytoplasmic domain of those RTKs to reveal a latent tyrosine kinase activity that stimulates RAS (a GTPase) to substitute its guanosine-5'-diphosphate (GDP) for a guanosine-5'-triphosphate (GTP). One major effector of RAS is the RAF family of serine/threonine kinases. RAFs signal through phosphorylation and activation of a downstream kinase, the mitogen-activated protein kinase kinase,[41] which subsequently phosphorylates and activates ERK, a member of the mitogen-activated protein kinase (MAPK) family.[42] Active ERK then phosphorylates further downstream effectors, including fos, MITF, myc and Ap-1.[43, 44]

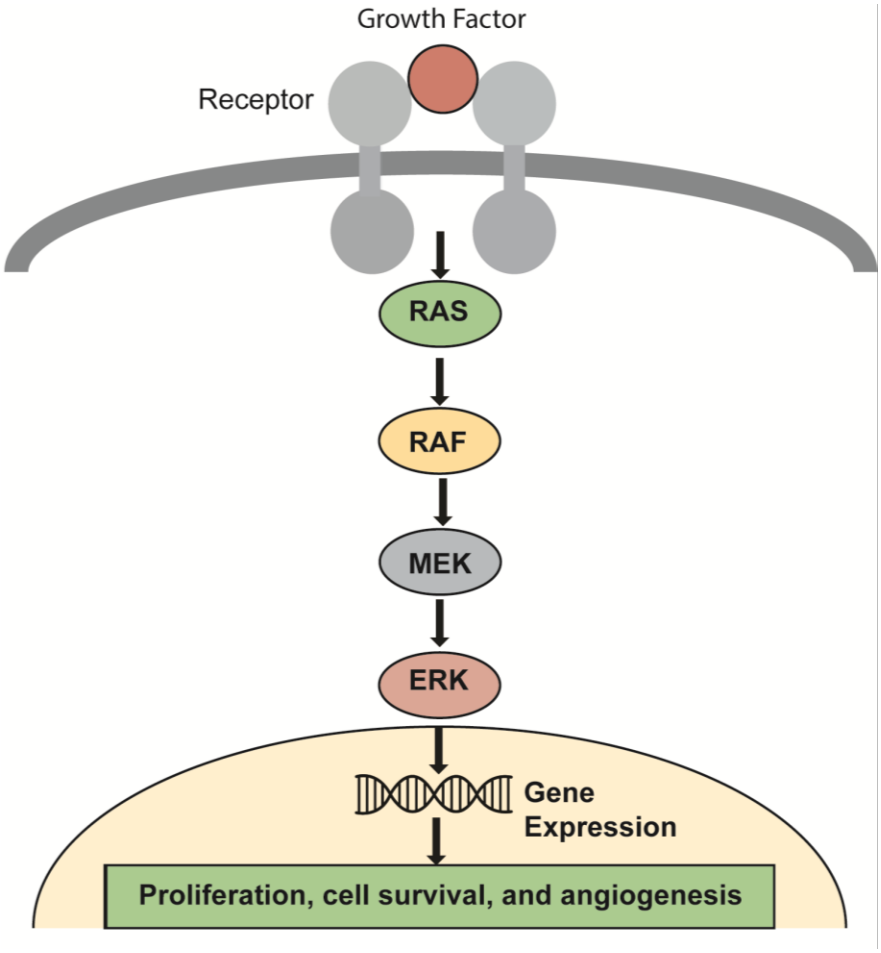

Figure 1.1 Mammalian MAPK/ERK Pathway. Adapted from Davies et al.[45]
Specific somatic mutations, often called driver mutations, alter the activities of kinases within the MAPK pathway in ways that result in the constitutive activation of the pathway. Most of these driver mutations affect kinases acting early within the pathway, such as RTKs, RAS, and RAF. The high frequency of activating mutations around the RAS-RAF axis suggests that it is a regulatory gateway within the MAPK/ERK pathway.[40]

Effective treatment of colorectal, melanoma and other MAPK-associated cancers has therefore been realized through development of therapeutics, both small molecules and biologics, that specifically target and inhibit aberrant activities in upstream activators and kinases of the MAPK signaling pathway, most notably RAS and RAF members. Predictive biomarkers based on those specific mutations have likewise been shown to enable selection of effective treatment regimens for MAPK-associated cancers.[46]

1.3.1 Mutations in the RAF Gene Family and Cancer

The first *RAF* (rapidly accelerated fibrosarcoma) gene was identified in 1983 as the murine retroviral oncogene *v-RAF*, a homolog of the human *CRAF* gene.[47] A year later, an avian homolog (*v-mil*) was identified.[48] Together, v-raf and v-mil (in this thesis, genes are written in italics, while gene products are not) were the first onco-proteins found to have serine/threonine kinase activity, an activity subsequently found to be shared by all RAF proteins. In mammals, three RAF isoforms have been identified, each originating from a distinct gene: *ARAF*, *BRAF* or *CRAF*. The murine sarcoma viral oncogene homolog B1 gene *BRAF*, for example, encodes a serine/threonine kinase that can be activated by the Kirsten rat sarcoma (KRAS) protein as the top-level element of the RAF/MEK/ERK (MAPK) kinase cascade. MAPK signaling controls proliferation, differentiation, and other aspects of cellular activity through phosphorylation of different ERK substrates, including transcription factors and cytoskeletal components. Crystal structures of RAF proteins suggest the valine at position 600 is required for BRAF to maintain an inactive conformation in the absence of interaction with KRAS.[49, 50] Amino acid substitutions at this position permit BRAF activation independent of its dimerization with either RAF1 or itself, which is normally required for activation. Missense mutation of *BRAF* V600 thereby permits MEK binding and phosphorylation, leading to BRAF-mediated signal transduction.

Sanger and next-generation sequencing have detected missense mutations in *BRAF* in approximately 8% of all human cancers.[45] Activating mutations in *BRAF* are present in approximately 40% to 60% of advanced melanomas,[45, 51] in 40% to 80% of papillary thyroid cancers (PTC),[52] and approximately 10-15% of all CRCs.[53] The most frequent *BRAF* V600 mutation is a substitution at the second position of the codon (GTG>GAG; c.1799 T>A) which results in an amino acid change in the gene product from a valine (V) to a glutamic acid (E) (p.V600E). Initial sequencing studies suggested that the p. V600E mutation was dominant, accounting for more than 90% of all *BRAF* mutations.[54] However, follow-up studies have shown that the p. V600E genotype is not as prevalent as reported in that early work, as the p. V600K and p. V600R mutations occur at a higher frequency than first reported.[55-57] Moreover, other rare BRAF V600 mutations have been identified, including p. V600E2 and p. V600D/M/G/A/L (**Table 1.1**),[19, 50] as well as rare mutations at exon 11 (the P-loop) of the kinase domain of BRAF.[39, 58]

Table 1.1 Frequencies of known missense mutations in the BRAF V600 codon. Frequencies of known missense mutations in the BRAF V600 codon. Values reported taken from.[19, 22, 56]

***BRAF* Mutation**	**Nucleotide Change**	**Frequency**
p.V600**E**	c.1799T>A or c.1799_1800TG>AA	79.0 – 84.0%
p.V600**K**	c.1798_1799GT>AA	8.0 – 12.4%
p.V600**R**	c.1798_1799GT>AG	2.2 – 5.0%
p.V600**D**	c.1799_1800TG>AT	0.3 – 1.3%
p.V600**M**	c.1798G>A	0.3 – 4.0%
p.V600**G**	c.1799T>G	1.3%
p.V600**A**	c.1799T>C	N.A
p.V600**L**	c.1798G>C or c.1798G>T	N.A

1.3.2 *BRAF* Mutations as a Predictive Biomarker in Targeted Cancer Therapy

Given the importance of *BRAF* V600 missense mutations in cancer development and progression, intense effort has been directed towards developing molecular therapeutics targeting these deregulated signals. In 2011, the first BRAF inhibitor (BRAFi) PLX4032 (vemurafenib; Roche, Basel, Switzerland) received approval from the U.S. Food and Drug Administration (FDA) for the treatment of *BRAF* V600E positive metastatic or unresectable melanomas,[1, 2] a form of skin cancer that originates in melanocytes, the pigment producing cells in the basal layer of human skin epidermis.[59] Incidence rates for melanoma are increasing, with over 76,000 new diagnoses and nearly 10,000 deaths reported in the U.S. in 2014.[4] Although it represents less than 5% of all incidences of skin cancer per annum, melanoma accounts for 65-80% of all deaths from skin-related malignancies.[60] In 2013, a second BRAFi, trafinlar, also known as dabrafenib,(GlaxoSmithKline; GSK, UK) received FDA approval for the same indication.[61, 62] In addition, trametinib (GSK), a MEK1/MEK2 inhibitor (MEKi), is approved for treatment of metastatic melanomas positive for either *BRAF* V600E or V600K.[54, 63] Either of these small molecule drugs binds the active state of the kinase domain to selectively inhibit the proliferation of cells with unregulated BRAF activity. Both received approval for treating the two most common *BRAF* V600 mutations, *BRAF V600E* (79% to 84% of all V600 missense mutations) and *V600K* (8% to 12%). However, in-vitro and preclinical data indicate that BRAF and MEK inhibitors can be effective in treating patients with a more rare mutation in codon 600 of the

BRAF gene,[1, 44, 64] including patients carrying a V600R (2% to 5%), V600D (0.3% to 1.3%), or V600M (0.3% to 4%) mutation.[57, 62, 65-67] In contrast, evidence of overexpression and activation of the MAPK pathways in the absence of activating mutations in *BRAF* V600 generally precludes treatment with BRAFis, as they are capable of accelerating growth in tumors harboring a wild-type[68] sequence at the V600 codon.[69, 70] Finally, resistance to vemurafenib or dabrafenib often occurs within 6 to 12 months of treatment, necessitating careful disease monitoring. When relapse is observed, the FDA has approved treatment of *BRAF* V600E positive patients with a combination of dabrafenib and the MEKi trametinib.[54]

Missense V600 mutations are also observed in approximately 10% to 15% of CRC patients, and at a higher frequency (50% to 60%) in Mut L homologue 1 (*MLH1*) deficient CRCs.[71] The clinical practice for colorectal cancer guidelines from the National Comprehensive Cancer Network therefore recommend *BRAF* V600 screening of metastatic CRC patients, as patients positive for *BRAF* V600E may not respond to anti-EGFR monoclonal antibodies such as cetuximab and panitumumab.

1.3.3 Mutations in the *RAS* Gene Family and Cancer

Somatic mutations in *RAS* (rat sarcoma) genes were the first disease-specific genetic alterations identified in human cancers.[72, 73] That initial RAS research, which dates from the 1960's, identified a murine virus induced sarcoma in new-born rodents and found that mutation of cellular RAS proteins (21 kDa) was a major oncogenic driver.[74, 75] In 1982, the first two human *RAS* oncogenes, *HRAS* and *KRAS*, were identified in human cancer cell lines.[76, 77] A year later, a third *RAS* gene (*NRAS*) was discovered in human neuroblastoma cells.[78] Mutations in these three *RAS* isoforms are found in ~ 30% of all human cancers,[72] with the highest frequency found in pancreatic cancers (69 - 95% of all patients),[79] colon cancer (40 - 45%),[53, 80, 81] malignant melanomas (15 - 20%)[82-84], and lung adenocarcinomas (16 - 20%).[85, 86]

The germline sequence of each *RAS* isoform is highly conserved. KRAS is a GTPase that serves as a central relay for signals originating at receptor tyrosine kinases, including the EGFRs within the intestinal epithelium and other human tissues. Those receptor tyrosine kinases control KRAS activation through guanine nucleotide exchange factors (GEFs), which can activate KRAS by

stimulating the release of guanosine diphosphate (GDP) to permit binding of guanosine triphosphate (GTP). Missense mutations within *RAS*, which in *KRAS* are most often observed within so-called "mutational hot-spot" codons 12, 13 and 61 (but can occur in codon 18, 117 or 146 as well),[87] impair intrinsic *RAS* GTPase activity, resulting in enhanced GTP binding and locking of the *RAS* protein into its active form. Constitutive activation of the MAPK pathway is then observed.[88, 89] Among the *RAS* isoforms, activating missense mutations are found most commonly in *KRAS* (85%), less commonly in *NRAS* (12%), and rarely in *HRAS* (3%).[90] In *KRAS*, single amino acid substitutions are generally observed, typically through mutation of codon 12 or 13; mutation of codon 61, 117 or 146 occurs at a much lower frequency. **Table 1.2** lists of all known missense mutations in *KRAS* codons 12 and 13.[50]

Table 1.2 Frequencies of clinically relevant missense mutations in KRAS. Values reported taken from.[50]

***KRAS* Mutation**	**Nucleotide Change**	**Frequency**
p.G12**D**	c.35G>A	35.0%
p.G12**V**	c.35G>T	23.8%
p.G13**D**	c.38G>A	13.1%
p.G12**C**	c.34G>T	11.9%
p.G12**A**	c.35G>C	5.7%
p.G12**R**	c.34G>C	3.2%
p.G13**C**	c.37G>T	0.9%
p.G13**S**	c.37G>A	0.2%
p.G13**R**	c.37G>C	0.2%
p.G13**A**	c.38G>C	0.1%
p.G13**V**	c.38G>T	0.1%
G12-13	Complex	0.4%

BRAF and *KRAS* missense mutations are often observed in colorectal cancers, with mutation of one generally thought to be mutually exclusive of the other.[91] Current dogma explains this by positing that missense mutation of *BRAF* and *KRAS* is either functionally redundant with respect to cancer pathogenesis, so that the second mutation (i.e. co-mutation of *BRAF* and *KRAS*) provides no selective advantage to the tumor, or that co-mutation of *BRAF* and *KRAS* is a tumor-lethal genotype.

1.3.4 Role of *KRAS* Mutations as a Predictive Biomarker in Targeted Cancer Therapy

CRC is the third most common cancer and the second leading cause of cancer-related death in the US.[92] In the past decade, survival of mCRC patients has approximately doubled.[87] This significant improvement is mainly due to the introduction of new combinational therapies and novel targeted therapies, such as anti-EGFR monoclonal antibodies(mAbs).[3] In 2004, the FDA approved the treatment of mCRC patients with erbitux (cetuximab; Bristol-Myers Squibb, US). Erbitux is a mAb targeting EGFR, an RTK upstream of RAS, that prevents activation of the receptor and kinases downstream.[93, 94] In 2007, a second anti-EGFR mAb, vectibex (panitumumab; Amgen, US), was approved for the treatment of mCRC.[95] Regrettably, initial clinical data showed that response rates of mCRC patients to either anti-EGFR therapy varied significantly, and that both targeted therapies increased treatment cost and toxicity. These factors spurred further studies that led to the discovery of biomarkers predictive of mCRC patient response to anti-EGFR therapies.[3] Missense mutation of *KRAS* was thereby identified as a negative predictive biomarker,[96] leading the American Society of Clinical Oncology[80] to recommend that all patients with mCRC have *KRAS* within their tumor tested for the most common G12/G13 mutations (**Table 1.2**); only mCRC patients with WT *KRAS* are then eligible to receive anti-EGFR therapy.[97] *KRAS* mutation testing in mCRC tumors is now mandatory in the USA, Europe, and Japan.[97, 98] The FDA has approved targeted therapy with panitumumab and cetuximab in mCRC patients negative for all *KRAS* G12/G13 mutations,[97, 99] while the European Society for Medical Oncology recommends establishing that the tumor is WT, without need to determine the specific missense mutation.[100]

More recently, activating mutations in other *KRAS* codons have also been suggested as negative predictive biomarkers for anti-EGFR therapy. These include rare mutations observed in *KRAS* codon 61, and in codons 117 and 146, which in CRC patients occur at a frequency of 4% (codon 61) and 6% (codons 117 and 146 combined), respectively **(Figure 1.2)**.[87, 101] Furthermore, missense mutations in *NRAS* codon 61 (4% of all CRC patients), or in *NRAS* codons 12 and 13 (3% of all CRC patients) have also been identified as negative predictive biomarkers for anti-EGFR therapy.[102]

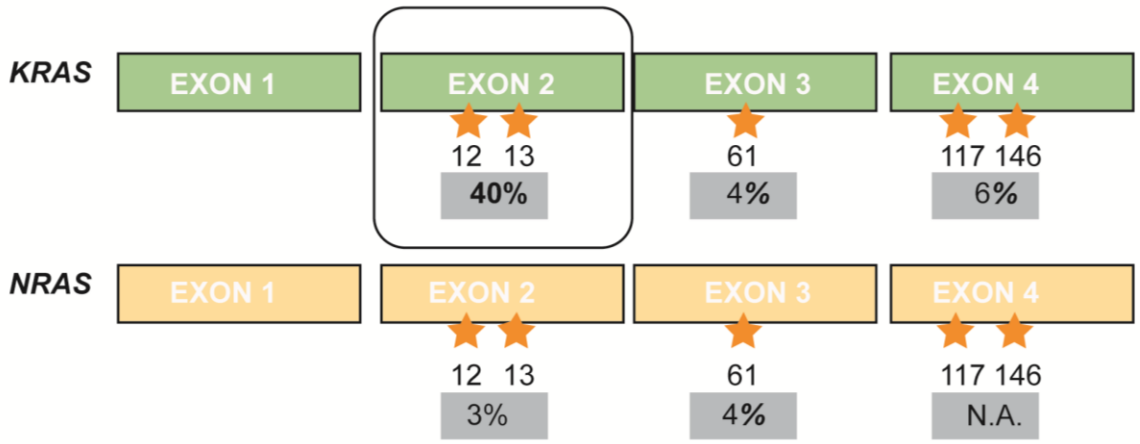

Figure 1.2 Maps of the *KRAS* and *NRAS* genes showing the mutational hot-spot codons. For each gene, the frequency of missense mutations within each hot-spot relative to the total frequency of all mutations in that gene is provided.[87]

1.4 Current Clinical Methods for Detecting Missense Mutations

1.4.1 Clinical Detection of *BRAF* Gene Mutations

BRAF V600 mutation status is routinely tested in cancer genomics laboratories to permit clinicians to select an appropriate treatment for the patient and thereby improve patient outcomes. Currently there are two FDA approved diagnostic tests against *BRAF* V600 missense mutations used by clinics to identify metastatic melanoma patients eligible for treatment with BRAF or MEK inhibitors.[103] These two assays include the cobas 4800 *BRAF V600* mutation test (Roche) and the THxID *BRAF* assay (BioMérieux). Both are real-time (RT) PCR (qPCR) tests. The cobas test was developed as a companion diagnostic to vemurafenib, and was first used to select patients eligible for inclusion in vemurafenib phase II and phase III clinical trials. It is now used to select *BRAF* V600E positive metastatic melanoma patients eligible for treatment with vemurafenib. The cobas test utilizes a common forward and reverse primer set that amplifies a 116 base pair[11] fragment of human chromosome 7q34 bearing that portion (exon 15) of the *BRAF* gene containing *BRAF* codon 600. The cobas test is specifically designed to detect a c.1799T>A mutation in the BRAF gene that results in a valine to glutamic acid substitution (V600E). Two dual-labeled hydrolysis probes are used for this purpose. The first is designed to

be specific to the *BRAF* WT sequence[68] at and adjacent to codon V600, and the second is designed to be specific to the mutant *BRAF* V600E sequence (**Figure 1.3**).

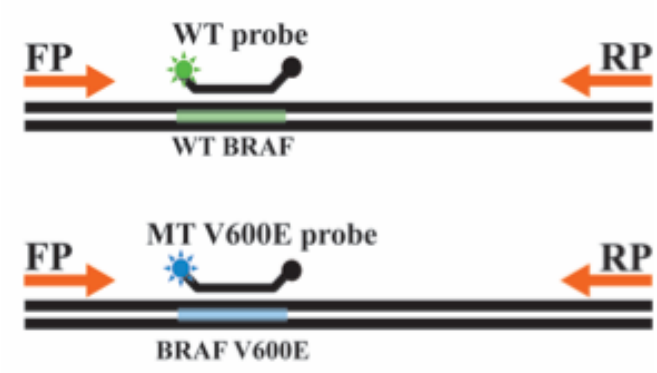

Figure 1.3 Schema showing the components of the cobas 4800 BRAF V600E mutation test.

Though designed to detect the V600E (c.1799T>A) missense mutation, the second probe shows varying degrees of cross reactivity with some but not all of the other known *BRAF* V600 missense mutations; those include detection of *BRAF* V600D (c.1799_1800TG>AT) when present at greater than 10% mutant frequency (MF), V600K (c.1798_1799GT>AA) at > 35% MF, and V600E2 (c.1799_1800TG>AA) at > 65% MF.[54] In principal, this is a desirable artifact of the assay as there is growing evidence metastatic melanomas bearing one of these less common missense mutations respond positively to treatment with vemurafenib.[54] However, given the relatively poor sensitivity of the assay to these mutations, particularly to V600E2, the clinical utility of these cross-reaction derived signals is marginal.

When applied clinically, the cobas assay is therefore effectively limited to detection of the *BRAF* V600E1 missense mutation, offering an analytical sensitivity of approximately 5-10% MF. A typical sample of genomic DNA (~ 30 ng) isolated from a tumor specimen will contain at least 10,000 total copies of *BRAF*. At least 500 copies of a *BRAF* V600E1 allele must therefore be present for reliable detection of the mutation. When those conditions are met, the test provides a call on *BRAF* V600 status, but does not quantify the mutant frequency present in a specimen. Moreover, as noted above, the assay does not detect all known missense mutations at *BRAF* V600. This is a significant limitation, as the committee for medicinal products for human use (CHMP) of the European Medicines Agency has recommended all melanoma patients carrying

MT *BRAF* V600 be eligible for treatment with a BRAF inhibitor. That agency has therefore recently ruled the cobas test insufficient as a companion diagnostic to vemurafenib because it does not capture the known clinical benefit to the broader population of melanoma patients carrying a V600 missense mutation.[104]

The second test, the THxID *BRAF* assay, has been approved for determining the eligibility of melanoma patients for treatment with tafinlar or mekinist. This assay is a qPCR test, but unlike the cobas assay it utilizes two AS-forward primers (instead of AS-probes) to achieve allele-specific detection. The first primer is designed to specifically amplify the WT *BRAF* V600 allele, while the other is designed to amplify both *BRAF* V600E1 and *BRAF* V600K mutant alleles, but has been found to show cross-reactivity to V600E2 and V600D. Its analytical sensitivity is between 5 – 10% MF for the two target mutant alleles, and is poorer for V600E2 and V600D. As with the cobas assay that predated it, MT calls by the THxID *BRAF* assay are qualitative and the assay does not detect all known *BRAF* V600 missense mutations.

Motivated by the limitations of the two FDA-approved assays, other methods to detect mutations in *BRAF* V600 have been developed, but none has proven sufficiently robust[105] or comprehensive to establish its clinical use in melanoma testing. Most are qPCR assays[18, 22-24, 56, 106, 107] employing either an allele-specific (AS) hydrolysis probe or AS primers, and may therefore be viewed as modified forms of the cobas and THxID *BRAF* assays. The emphasis on qPCR-based methods is likely due to the faster sample processing times and lower frequency of invalid results this approach generally offers over either mutation reagent analyses (*e.g.* ABI BRAF test[108]) or sequencing-based approaches. Nevertheless, Sanger and next-generation sequencing-based methods have been proposed,[19, 106, 109-112] as have various high-resolution melt (HRM) analysis methods.[113-116] In general, they can be classified into those that detect a select subset of V600 mutations with improved analytical specificity relative to that provided by the two FDA approved kits, and those with the capacity to detect all V600 mutations, but at an analytic specificity that is generally worse than 5%. An example of the first is provided by recent qPCR-based methods, such as CAST-PCR,[18] which is sensitive down to a mutant allele frequency of 2% but currently only applicable to the detection of V600E1 and V600K. Methodologies based on HRM analyses have likewise only been applied to the detection of a

subset of *BRAF* V600 mutations. They offer the advantage of simplicity and relatively low costs, but are generally more qualitative in nature due in part to a variability in results depending on the instrument and protocol used. In contrast, sequencing-based methods such as Sanger sequencing and next-generation pyrosequencing can detect all known and clinically relevant *BRAF* V600 mutations. However, the current sensitivity of these methods is such that a MF within the tumor of greater than 25% (Sanger) or 15-20% (pyrosequencing) is typically required.[109] Furthermore, sequencing methods require a relatively long turnaround time, and are labor intensive and expensive. As a result, they are not yet generally applied in cancer clinics.

Given that accurate diagnosis of *BRAF* MT-positive melanoma is critical to defining proper course of therapy, these facts indicate a need for the development of a next-generation assay that detects all known missense mutations within the *BRAF* V600 codon, and that does so at an improved analytical sensitivity that might permit reliable assignment of mutation status in patients where the melanoma has either not spread to the lymph nodes or not metastasized to more distant organs.

1.4.2 Clinical Detection of the *KRAS* Gene Mutations

KRAS G12/G13 (i.e. codons 12 and 13) mutational status is routinely tested in cancer genomics laboratories to permit clinicians to determine the eligibility of CRC patients for treatment with cetuximab and panitumumab. Currently there are two FDA-approved diagnostic assays for this purpose: the therascreen *KRAS* RGQ PCR kit (Qiagen) and the cobas *KRAS* mutation test (Roche).

The therascreen assay, which received FDA approval in 2014, is a qPCR test used to identify patients carrying any one of the seven most common G12/G13 missense mutations in exon 2 of the *KRAS* gene.[117] CRC patients positive for a G12/G13 missense mutation are ineligible for anti-EFGR mAb treatment as a course of therapy. The therascreen *KRAS* RGQ PCR Kit contains reagents for eight separate real-time amplification reactions; seven mutation-specific reactions to amplify and detect missense mutations in codons 12 and 13 of the *KRAS* oncogene, and one control reaction that amplifies and detects a highly conserved region of exon 4 of *KRAS*. Each mutation-specific reaction makes use of an amplification refractory mutation system[91] type AS-

PCR scheme to selectively amplify mutant *KRAS* alleles within genomic DNA also containing WT *KRAS*.

The ARMS technique for detecting known point mutations was pioneered by Newton et al.[68] It is based on the principle of allele-specific priming of the PCR process, and it therefore relies on the ability to effectively design a primer that only extends when its 3' terminal nucleotide is complementary to the paired template base. A typical ARMS test for a somatic point mutation consists of a pair of amplifications in the same reaction mixture containing genomic DNA as substrate. One amplification product results from extension of the specific ARMS primer and its paired primer; that amplification product is therefore (expected to be) observed only when the mutation is present in the genomic DNA sample. The second amplification product is generated from the extension of a primer pair acting on a highly conserved sequence, typically within the target gene of interest but away from the mutational hot-spot being queried. The generation of this reference amplicon indicates the reaction mixture and thermal cycler are working properly, and may be used in some assays to verify the presence of the target gene within the specimen.[68]

The method has proven effective in detecting and quantifying mutations where the ARMS allele-specific primer forms a pronounced mismatch (i.e. a thermodynamically highly destabilizing mismatched base pair) with the germline allele. These include C–T, C–C, G–A and T–T mismatches, for which priming efficiencies of zero or below 5% can be achieved. But the method is far less effective when the mismatch formed is either moderately (e.g. A–A) or weakly destabilizing (e.g. G-T).[118] One may then try to introduce a secondary mismatch into the ARMS primer, often at the 3'-1 or 3'-2 position, with the hope of making extension of that primer more allele specific. But this strategy is empirical and, and in part for that reason, its effectiveness has proven highly variable. As a result, ARMS-PCR has not found widespread use in highly multi-plexed assays designed to detect a panel of possible mutations within an oncogenic hot-spot.

Due in part to these general weaknesses with ARMS technology, the therascreen *KRAS* RGQ PCR Kit has limitations. Several of the seven ARMS primers are cross-reactive with non-target alleles. For example, the ARMS primer for G12A also amplifies G12C, G12S, and G12V. This compromises both the specificity of the assay and its analytical sensitivity. Detection of G12/G13 mutations, including G12R and G13D, therefore requires a mutation frequency of 6.4% or

greater, making the assay unsuitable for early disease detection or residual disease monitoring. Moreover, the assay is limited in the number of G12/G13 mutations it can detect. In particular, the assay precludes detection of known but rare G12/G13 mutations, which collectively comprise approximately 2% of all CRC patients. This is unfortunate, as patients harboring any one of those mutations should not receive anti-EGFR therapy.

Finally, due in part to the method of detection used, the assay requires ~50 ng of DNA as the input, which is generally challenging to obtain from sparse needle core biopsy specimens from liver or lymph node metastases. The therascreen *KRAS* RGQ PCR kit utilizes a Scorpion®-type bifunctional hairpin probe that is attached to the 5' end of the ARMS primer to monitor amplification. After extension of the primer, the sequence specific probe can bind to its complement within newly formed amplicons, causing the fluorophore and quencher on the probe to become sufficiently separated to generate a detectable fluorescence signal.[119, 120] Though this detection strategy is advantageous in terms of reducing the number of reactions required, changes in relative fluorescence generated by Scorpion-type probes per amplicon produced are generally lower than those achieved using standard Taqman-type probes.

The second assay is the recently approved cobas *KRAS* Mutation Test kit, which can detect 19 different mutations known to occur within *KRAS* codons G12/G13 and Q61.[121] The cobas kit uses primers that amplify an 85 base-pair sequence within *KRAS* exon 2 containing codons 12 and 13 and a 75 base-pair sequence within exon 3 containing codon 61. In contrast to the therascreen test, detection of mutant alleles by the cobas *KRAS* mutation test is achieved using a set of allele-specific probes targeting either codons G12/G13, mutations in which are queried in one reaction chamber, or codon Q61, which is analyzed in a separate reaction. The fluorescence generated by hydrolysis of each uniquely labeled probe is recorded, and after amplification, each amplicon generated is subjected to a thermal melt analysis in which the temperature is ramped from 40°C to 95°C (a so-called "TaqMelt") and the melting transition of the amplicon(s) recorded. An amplicon fully complementary to the probe melts at a higher temperature than do amplicons with one or more mismatches, permitting mutation calls to be made. The cobas test is thereby more comprehensive in analyzing clinically actionable *KRAS* mutations when compared to the therascreen assay. This in turn provides evidence that assays employing allele-specific

probes may hold advantages over those employing allele-specific primers in highly multiplexed assays intended to detect a complex set of missense mutations within one or more codons of an oncogene. But limitations to the cobas assay are known. Probe specificity challenges limit the analytical sensitivity of the assay to between 5% and 10% mutation frequency,[121, 122] and though the assay is *generally* reliable in making mutant calls above that threshold, it does not quantify mutant frequencies. Moreover, it can generate false positives. For example, when analyzing 25 formalin-fixed, paraffin-embedded lung cancer samples of different sizes and tumor percentage using a range of available PCR and sequencing based approaches, Hinrichs et al.[123] reported cases where results from the cobas assay were at odds with those obtained from other reliable orthogonal methods. In each case, a false mutant-positive call was made, including for a sample that contained an *EGFR* mutation, making the presence of a *KRAS* mutation highly unlikely.[124]

Finally, the cobas test requires 100 ng total DNA, which can be obtained by sacrificing an entire 5 μm section, but which clearly represents a non-optimal use of needle core biopsy specimens from liver or lymph node metastases since it precludes application of DNA recovered from the sample to other informative tests.[125]

1.5 Thesis Objectives

The discovery of "driver mutations", named as such because they drive cancer progression, is transforming the practice of clinical oncology in important and positive ways. With respect to colorectal cancer, genetic testing can help show if an individual has a high risk of acquiring the disease due to inherited maladies such as Lynch syndrome (also known as hereditary non-polyposis colorectal cancer). Standard care of individuals newly diagnosed with CRC therefore includes testing for Lynch Syndrome in tumor biopsies to assess genetic instability associated with impaired DNA mismatch repair. Tumor evaluation also generally includes immunohistochemistry testing for the expression of the MMR proteins associated with either Lynch Syndrome or microsatellite instability (MSI), *BRAF* testing, and *MLH1* hyper-methylation analyses.

Genetic testing of CRC patients has likewise become a standard tool used by clinicians to determine proper courses of therapy and to monitor disease during and post-treatment.[54] In

particular, *KRAS* G12/G13 testing is now required as a prognostic biomarker for determining the eligibility of CRC patients for treatment with penitumumab or cetuximab.

The FDA approved assays currently used by clinics to detect missense mutations in *BRAF* V600 and *KRAS* G12/G13 are not capable of detecting clinically actionable mutations at mutational frequencies low enough to permit their robust application to early disease detection or minimal residual disease monitoring. Moreover, detection of all clinically actionable missense mutations is not certain, in part due to limitations to assay specificities and the inability to unequivocally discriminate missense mutations from synonymous germline sequence variations.

Thus, a need exists for robust and cost-effective assays capable of detecting the complete set of clinically informative somatic point mutations within an oncogenic hotspot. **The central objective of my thesis is to develop a novel platform for creating highly sensitive assays against all possible missense mutations in an oncogenic hotspot codon or adjacent set of hotspot codons that ameliorates the known limitations to current FDA-approved assays. The platform is designed to enable development of assays against all possible missense mutations in oncogenic hotspots and, if required, unequivocally differentiate them from synonymous germline alleles. That platform is first described in chapter 2 of this thesis, and then applied to the development of two clinically relevant assays: the first permits screening and detection of all known clinically actionable mutations in *BRAF* V600, and the second achieves the same for *KRAS* G12/G13. The clinical performance of each assay is assessed through its application to FFPE tumor specimens from a de-identified cohort of metastatic colorectal cancer patients**.

The platform is designed based on the hypothesis that significant improvements to the sensitivity and specificity of missense mutational analyses can be realized by collectively leveraging the template-partitioning capabilities of droplet digital PCR (ddPCR), the ability to use locked nucleic acid (LNA) substitutions to modulate the thermal stabilities of duplexes formed between single-strand template DNA and an allele-specific probe, and the development of a molecular thermodynamic model described in this thesis that enables the number and pattern of LNAs in

each such probe to be optimized such that the probe hybridizes only to its target allele at the annealing and amplification temperature of the ddPCR assay.

Specific technical objectives associated with development, assessment and ultimately clinical adoption of the platform and the two assays derived from it include:

- Co-defining and utilizing a nearest-neighbor type molecular thermodynamic model to design *in silico* LNA-substituted dual-labeled hydrolysis probes that are highly selective to their target allele
- Conceptualizing and creating a platform for creating a novel class of assays, termed "wild-type negative assays", that use ddPCR to unequivocally discriminate germline alleles from clinically actionable alleles bearing a missense mutation (or mutations) in one or a set of adjacent hot-spot codons within an oncogene
- Co-creating a software tool and associated graphical user interface that clinicians can use to analyze output data from wild-type negative assays to detect if a clinically actionable mutation is present in the specimen, to quantify the frequency of any such mutation, and to enable the clinician to make a sound decision on therapeutic course of action based on the results of the assay

Chapter 2 reviews relevant applications of ddPCR to the detection of mutations in genomic DNA, and then presents the concept of the wild-type negative assay and its critical operational features. One such feature is the need to design dual-labeled hydrolysis probes that are completely selective to their target allele, and the features of the model developed to design those probes are described.

Chapter 3 describes the development and validation of a wild-type negative ddPCR-based diagnostic assay against *BRAF* V600 mutations and its application to gDNA purified from FFPE mCRC tumor specimens.

Chapter 4 describes the development and testing of a wild-type negative assay for detecting G12/13 missense mutations in *KRAS* and its application to gDNA purified from FFPE mCRC tumor specimens.

Chapter 2: Digital PCR and the Proposed Wild-Type Negative Assay

2.1 Digital PCR: Basic Principles and Prior Applications to Mutation Detection

First described in 1999 by Vogelstein and Kinzler,[126] digital PCR (dPCR) is a method of absolute nucleic acid quantification.[127] Its use has grown in recent years through the development of microfluidic systems and emulsion chemistries to simplify and automate the process.[29, 128] A number of dPCR instruments are now available commercially. These include microfluidic chip-based digital PCR (cdPCR) machines[129] and droplet digital (ddPCR) equipment.[28] The latter method, specifically the Bio-Rad QX100 ddPCR instrument, is employed in this work. ddPCR is based on limited partitioning of individual molecules of template DNA into thousands of independent isolated sub-nL droplets; typically, the dilution and partitioning are designed so that each such droplet contains zero, one or a small number of copies of template. In the BioRad ddPCR system, sample DNA is solubilized in an appropriate PCR mastermix solution and then partitioned into droplets by emulsification of the aqueous reaction mixture within a thermostable oil. Massively parallel PCR amplification is performed on the ensemble of droplets using an appropriate reporting agent, usually a TaqMan™-type dual-labelled hydrolysis probe, to record the end-point fluorescence of each droplet. That end point fluorescence within each droplet is quantified by serially reading each droplet in a reader that operates using concepts similar to that of a flow cytometer.

Due to the sub-nL size of each droplet, conducting a suitable number of amplification cycles within a droplet containing a single copy of template DNA will result in saturation of the end-point fluorescence signal. A clear distinction between template-positive and template-negative droplets is therefore possible. As a result, the number of template-positive droplets may be combined with Poisson statistics to determine with good accuracy the abundance of the target sequence in the original sample. Digital PCR therefore offers advantages over traditional qPCR quantification of nucleic acids, most notably its capacity for absolute quantification without need for an external reference or calibration curve. Moreover, as it relies on an end-point fluorescence measurement, ddPCR results are in general far less sensitive to the presence of amplification inhibitors within the sample matrix.[130]

Using the end-point fluorescence data for all read droplets, the average copies of template per droplet (*CPD*) is determined using the relation:[131, 132]

$$CPD = -ln(1-\theta) \qquad 2.1$$

where θ is the fraction of the total droplets read having a positive end-point fluorescence (above a prescribed threshold value). From this *CPD*, Poisson statistics may then be used to compute the probability $p(n)$ that a given droplet initially contains n copies of template

$$p(n) = \frac{(CPD)^n\, e^{-CPD}}{n!} \qquad 2.2$$

Together, equations 2.1 and 2.2 thereby allow the total copies of the target sequence in the initial sample to be quantified, with the average concentration of template $c_{template}$ (copies/µL) in the initial sample given by

$$c_{template} = \frac{CPD}{V_{droplet}} = \frac{-ln(1-\theta)}{V_{droplet}} \qquad 2.3$$

where $V_{droplet}$ is the average droplet volume.

Multiplexing of a ddPCR experiment to quantify two different template sequences is possible. In the Bio-Rad QX100 or QX200 ddPCR system, the emulsification reaction typically partitions initial copies of each template among *ca.* 20,000 aqueous nL-sized droplets. End-point fluorescence signal(s) in each droplet is read at a high speed by a two-color droplet detector in which the excitation and emission filters are usually selected to detect FAM (fluorescein) and HEX dye fluorescence amplitudes. A 2-D plot may then be generated from the data in which the FAM end-point fluorescence amplitude generated in each droplet is plotted against the HEX end-point fluorescence amplitude such as shown in **Figure 2.1**. In a ddPCR experiment where two different targets are amplified (probe 1 – FAM-labeled; probe 2 – HEX-labeled), each read

droplet will typically appear in the plot in one of 4 unique droplet clusters: empty droplets containing neither template (FAM^-/HEX^- cluster), droplets containing template 1 (FAM^+/HEX^- cluster), droplets containing template 2 (FAM^-/HEX^+ cluster), and droplets containing both templates (FAM^+/HEX^+ cluster). The ability to completely segregate clusters in the 2-D display allows accurate enumeration of cluster populations and robust statistical analyses of the resulting data.[54]

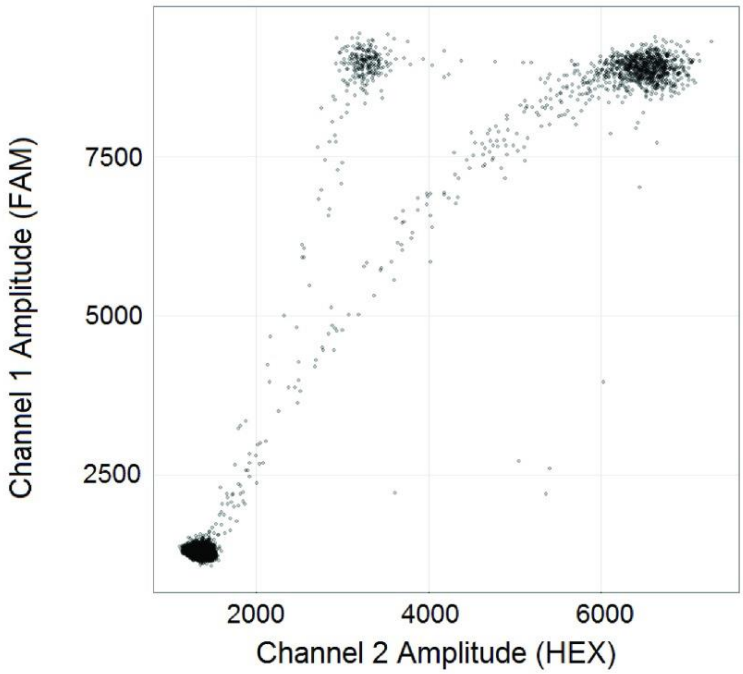

Figure 2.1 Illustration of typical output data from a multiplexed two-channel ddPCR experiment.

Compared to qPCR, dPCR has many advantages that could allow it to be used in clinics to perform diagnostic, prognostic, and predictive tests for disease.[133-137] In particular, ddPCR has been shown to offer improved precision that allows finer fold change measurements[138] and more sensitive detection of rare genetic events, including mutations.[28] Moreover, when compared to current next-generation sequencing (NGS) platforms, digital PCR is far more sensitive and able to operate on much lower quantities of DNA and template.[128, 134] It is also complementary to NGS in important respects. For example, ddPCR may be used to verify NGS sequencing results.[139]

But ddPCR has yet to be adopted in the clinic. Some of this is simply due to the newness of the technology, as clinically amenable ddPCR instruments have only been available for a few years. As a result, reliability/reproducibility, dynamic range, and cost questions remain.[140]

Furthermore, robust ddPCR-specific sample preparation protocols must be established and validated, as ddPCR is as susceptible as qPCR to upstream errors associated with sampling and DNA extraction.[141] Yet it is clear that ddPCR offers unique opportunities for highly sensitive molecular genetic analysis in cancer. Indeed, it has been applied to the detection of specific somatic missense mutations,[29, 126, 142-145] allelic imbalances,[146, 147] and loss of heterozygosity (LOH) in clinical specimens.[148] For example, Sanmamed et al[149] recently reported a ddPCR assay capable of specifically detecting the BRAF V600E missense mutation in circulating DNA to mutant frequencies of 0.005%. Other ddPCR assays have been reported against BRAF V600E, as well as against KRAS G12D or G12V mutations.[126, 133, 143-145, 149, 150] However, the multiplexing of dPCR to permit the collective detection of all possible somatic mutations at an oncogenic hotspot in a single well has received little attention to date. Taking as an example the detection of all known missense mutations in *BRAF* V600, in principal this could be achieved by conducting allele-specific amplifications for all known mutations in parallel. This concept is challenged by the need to establish a way to uniquely identify and quantify each mutation using a 2-color detection system, and a way to avoid a false positive call for one possible mutation due to cross-reaction of the probes and/or primers intended to detect a different mutation.

However, many clinical tests only require one to collectively differentiate all missense mutant alleles from all synonymous germline (wild-type) alleles, without need to determine the precise mutation when present. The multiplexing problem is then simplified somewhat, as one can then simply conduct two reactions in parallel: one that amplifies all copies of the gene of interest, regardless if they are mutated or not, and second that selectively amplifies either the entire collection of mutant alleles or all copies of the germline allele.[68] The challenges associated with either approach are effectively the same, with avoidance of the mutant-specific reaction showing cross-reactivity with WT alleles, or *vice versa*, being the most critical. This thesis focuses on assays that employ a second reaction that selectively amplifies only copies of the target gene that are wild-type within the oncogenic hotspot of interest.

2.2 The Wild-Type Negative Assay Concept

CRC patients are tested in the USA, as well as in Europe and Japan, for missense mutations within *KRAS*. In the USA, CRC patients who are *KRAS* G12/G13 mutation negative are eligible for treatment with an anti-EGFR mAb, while in Europe they are eligible if all *KRAS* within the tumor is WT. For *KRAS*, as well as for other oncogenic biomarkers including *BRAF*, *PIK3CA* (missense mutations preclude anti-HER2 treatment of breast cancer), etc., clinicians therefore require a means to sub-classify cancer patients in terms of the absence or presence of a missense mutation in the (proto-)oncogenic biomarker of interest.

This thesis is concerned with the development and assessment of a new class of diagnostic tests that are specifically designed to meet that clinical need. The fundamental elements of this novel class of assays, identified here as "Wild-Type Negative Assays", are illustrated in **Figure 2.2**.

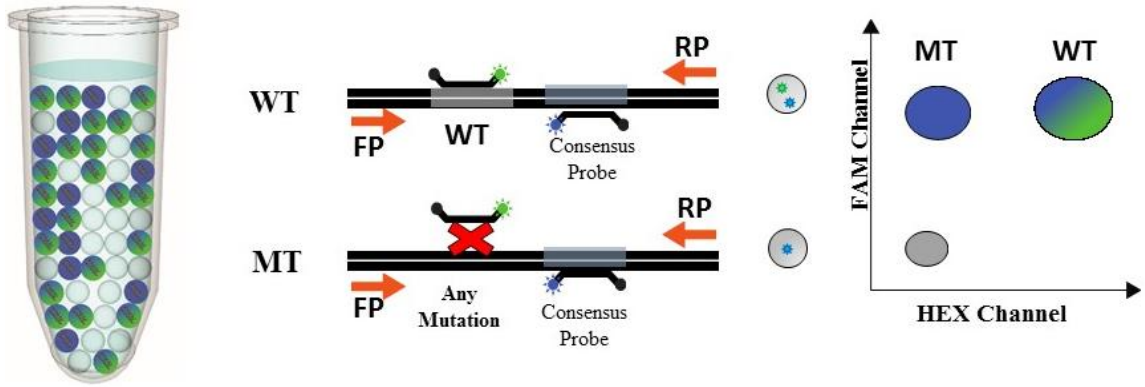

Figure 2.2 Basic elements and features of a wild-type negative mutational screening assay. Individual copies of target sequence to be analyzed are partitioned by limiting dilution into an excess number of sub-nL droplets and then independently amplified using a forward (FP) and reverse (RP) primer. In every droplet containing a single copy of a WT target sequence, that amplification will generate both a FAM (blue) and HEX (green) end-point fluorescence signal due to hydrolysis of both the FAM-labeled consensus probe and the HEX-labeled probe against the WT sequence within the hot-spot region of variability. A FAM signal, but no end-point HEX signal, is generated in droplets containing a copy of MT template due to the lack of WT-specific probe hybridization to the MT allele. All droplets containing a copy of the target may therefore be quantified by counting droplets displaying an end-point FAM signal, and the fraction of those that contain a copy of WT template then determined by counting the droplets having an end-point HEX signal.

In a wild-type negative assay, a fragment of the target gene is selected which contains the oncogenic hotspot of interest, a short germline sequence (10 – 20 bp in length) that has been shown to be highly conserved even in advanced stages of oncogenesis, and a set of priming sites that are predicted to offer efficient and specific amplification of the fragment. Two different dual-labeled hydrolysis probes are then prepared. The first, end-labeled with a unique reporter dye (e.g. FAM in **Figure 2.2**), is designed to target the highly conserved germline sequence within the fragment. The second, labeled with a different and spectrally distinct reporter dye (e.g., HEX), is designed to target the germline sequence of the proto-oncogene across the hotspot codon or pair of adjacent hotspot codons. The example provided in **Figure 2.2** therefore assumes the WT gene in comprised of only one known germline allele across the hotspot region. But this needn't be the case, as the assay can be multiplexed through creation of a set of probes, each one specific to a particular synonymous germline allele. These reaction components are mixed with DNA isolated from the tumor or tissue (e.g. blood) specimen, and then subjected to a limiting dilution that partitions those reactants and copies of the target gene into a large set (15,000 or more) of stable independent droplets. They are then subjected to ddPCR to amplify each copy of template present in each droplet. Amplified droplets are then individually read in both fluorescence channels to complete the ddPCR assay workflow as shown in **Figure 2.3**.

To fix ideas, consider first those droplets into which was partitioned one copy of the gene, with that copy being WT. Amplification of that copy will therefore produce a droplet displaying strong end-point fluorescence amplitudes from both released FAM and released HEX. In an output plot in which the FAM end-point fluorescence generated in each droplet is plotted against the HEX end-point fluorescence, all droplets of this type will therefore cluster in the FAM^+/HEX^+ region (**Figure 2.2**). In contrast, a droplet containing only a copy of the gene fragment that harbors a missense mutation within the oncogenic hotspot of interest will record an end-point fluorescence (FAM). A FAM signal results from hydrolysis of the probe against the highly conserved sequence, but not from the probe against the oncogenic hotspot, provided of course that the probe does not cross-react with any mutant alleles. Droplets of this type, which cluster in the FAM^+/HEX^- region of the output plot, are deemed wild-type negative. Lack of

fluorescence signal from the probe targeting the WT sequence within the oncogenic hotspot may therefore be used to detect the presence of a missense within the specimen. Assays of this type are therefore designated as "wild-type negative".

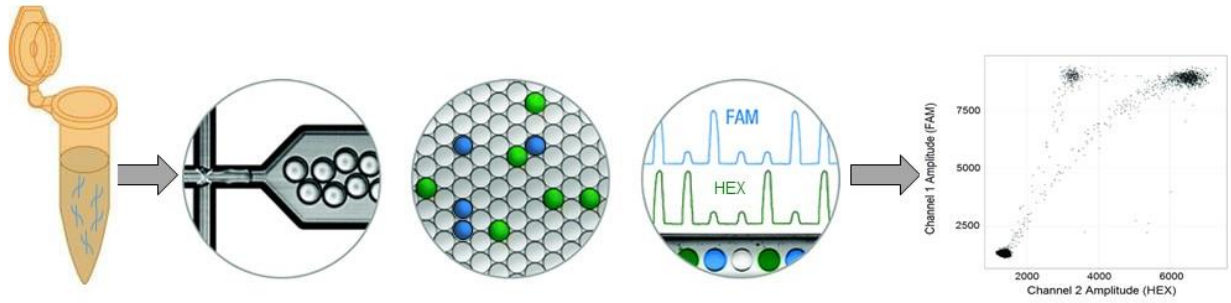

Figure 2.3 Basic elements of the workflow for a ddPCR-based wild-type negative assay. Figure adapted from Hindson *et al*[28] Template and PCR reagents are first mixed to form an aqueous reaction solution. A water-in-oil emulsion is then formed to partition the reaction solution into a large number of independent sub-nL droplets. That collection is then subjected to ddPCR to amplify each copy of template present in each droplet. Amplified droplets are individually read in both fluorescence channels, and a 2D plot of the set of end-point fluorescence amplitudes for each read droplet is constructed.

For each fluorescence amplitude channel, the total number of read droplets C and the fraction of those droplets displaying a significant fluorescence signal (above background) θ are recorded. Equations 2.1 and 2.2 are then used first to estimate the total copies of the gene $n_{template}$ present in the sample, and then again to estimate the total copies of that gene that are wild-type negative and therefore carry a mutation (n_{mutant}). The ratio $n_{mutant}/n_{template}$ x 100% gives the mutant frequency MF.

The uncertainty in the MF value can be determined from the corresponding uncertainty in θ, which is computed as

$$\theta \pm z_{CI}\sqrt{\frac{\theta(1-\theta)}{C}} \qquad 2.4$$

where z_{CI} is the confidence interval chosen (= 0.95 for a 95% confidence interval).

2.3 Critical Features of the Assay Design and Operation

Based on the design of the wild-type negative assay, its performance and clinical adoption are expected to depend on a number of important factors. Key among them are the following:

- Effective design of primers to provide robust and efficient amplification of the target gene fragment
- Design of a probe or set of probes targeting the oncogenic hotspot of interest, each of which is highly selective for its target WT allele, showing no cross-reactivity to any mutant allele
- Creation of a computational algorithm that objectively computes the MF value from the raw output of the ddPCR-based assay (i.e., the 2-D fluorescence data plot)

The principles and strategies used to address these challenges are described in the following sub-sections.

2.3.1 Effective Primer Design for ddPCR-Based Assays

Creation of a highly sensitive and specific ddPCR assay against one or more missense mutant alleles requires very careful consideration to the locations and characteristics of the amplicons, the probes, and the primers. Of particular importance for primers is the need to ensure the primer is specific to the target allele, and provides for efficient amplification of a single copy of that target sequence isolated within a nL-sized droplet through proper engineering of its sequence and its melting temperature relative to the annealing/extension temperature used for amplification. While numerous reports and studies provide sound guidelines for designing efficient PCR primers used in standard qPCR experiments, there is far less literature describing effective strategies for designing primers used in ddPCR experiments.

My thesis work, as well as research conducted by others in the laboratory, has served to establish useful empirical guidelines for designing efficient and specific primers for amplifying templates partitioned among nL-scale droplets by limiting dilution. In particular, though many of the

standard strategies used to design primer for standard bulk PCR may be applied with good effect to the design of primers for ddPCR amplifications, additional criteria and the augmentation of standard strategies can serve to improve the performance of those primers. Due to the unique features of ddPCR, the demands on primers used in ddPCR amplifications are in certain ways lower than those on primers used in standard bulk PCR. The limiting dilution used to partition individual copies of template into isolated droplets serves to also dilute those components/contaminants within the sample matrix that might inhibit or interfere with amplification of the target within each droplet; the likelihood of efficient amplification is therefore increased. Moreover, qPCR depends on accurate measurement of changes in relative fluorescence intensities with cycle number. Those changes are sensitive to changes in amplification efficiencies, which as noted can be influenced by inhibitory analytes within the sample.[151] In contrast, ddPCR depends on the detection of one (or if the reaction is multiplexed, several) end-point fluorescence value(s) for each droplet in the reaction. The precise value of each end-point fluorescence (EPF) amplitude is not relevant, as the experiment only requires counting of those droplets displaying an EPF amplitude above a defined threshold value. But primer design for ddPCR experiments is more challenging in other respects. In particular, due to the need to create a stable emulsion within which those droplets containing the aqueous reaction phase (containing the sample and the PCR reagents) are partitioned, the ddPCR system places constraints on the possible composition of that aqueous phase. In addition, each droplet will in general contain either one or no copies of the target sequence, but all droplets will nevertheless contain a substantial amount of total DNA (human genomic DNA in this work). In those droplets containing no copies of the target sequence, which might include more than half of all of the droplets formed, it is essential to avoid primers that cross-react with sequences within the background DNA, or with themselves, in ways that might lead to a false positive EPF signal.

These challenges were therefore considered, along with empirical evidence I and my coworkers in the laboratory have collected, to establish the following general strategy for designing primers for amplification of a target sequence by ddPCR:

- Primer and Amplicon GC content: ddPCR amplification efficiency and specificity generally decline if either the amplicon or either of the two primers (forward and reverse)

carry a large GC content, particularly if the GC content of either exceeds 60%. Whenever possible, the priming sites should therefore be positioned so as to generate an amplicon having a GC content of 50% ± ca. 8%. Primer sequences should likewise have a GC content near 50%, and ideally should not contain more 3 consecutive G residues. Satisfying all of these criteria generally allows the resulting ddPCR amplification to be performed at the preferred annealing temperature T_a of 60 °C.

- Primer Length, Sequence and Melting Temperature: Standard PCR primers are usually 18 to 30 bases in length. When a primer is used in a ddPCR amplification, length is not of paramount importance; instead, the most important considerations for effective primer design are 1) selection of a priming site to achieve good hybridization specificity at T_a, 2) the engineering of an optimal primer – template melting temperature ($T_{m,primer}$), and 3) the setting of each primer sequence to avoid cross-reactivity with non-target regions of genomic DNA, primer self-complementarity, and the formation of any stable secondary structures. There are often several regions of homologous sequence within a genome. Improper identification of sites for primer annealing is therefore a potential source of error in ddPCR and its avoidance can be challenging. However, several tools are available to aid selection of optimal priming sites to avoid cross reactivity. These include PrimerSelect, Primer Express, Primer Premier, OLIGO software series, and Primer3.[152-154] In general, intron and intergenic sequences are preferred as primer-binding sites for amplification of unique target sequences.

  Powerful software is also available designing primer sequences offering a desired $T_{m,primer}$. These include the OligoAnalyzer Tool provided on-line by IDT (https://www.idtdna.com/calc/analyzer), but other equally effective programs are available (e.g. dnaMATE, OligoCalc, OligoEvaluator). Note that it is important to apply these tools to prediction of $T_{m,primer}$ at the reaction conditions used in a ddPCR experiment. Aqueous reaction solutions in ddPCR usually contain 50 mM K^+, 3 mM Mg^{2+}, and about 0.8 mM dNTPs, but modest changes to those standard reaction conditions are possible. From the perspective of amplification efficiency, primer – template stabilities are often increased by including G and C nucleotides within the 3'–5

end of the primer. But primers with a high GC content at their 3' end also increase the likelihood of active hybridization with homologous non-target sequences. Limiting the GC content of the 3′ end of the primer (generally to no more than 3 G or C nucleotides) is therefore advised, as it minimizes the risk of false priming.

Finally, tools available to screen putative primer designs for self-complementarity and potential to form unwanted secondary structures include Primer3Plus,[155] which was primarily used in this work.

Primers used in standard bulk PCR are typically designed so that $T_{m,primer}$ lies 2 °C to 5 °C above a T_a, in part because when T_a = 60 °C this $T_{m,primer} - T_a$ difference generally results in optimal extension rates by polymerases used in PCR. In ddPCR, however, avoiding primer cross-reactivity, which increases with increasing $T_{m,primer} - T_a$, outweighs optimization of amplification efficiency (because of the use of end-point as opposed to temporal fluorescence measurements). $T_{m,primer}$ should therefore be set to avoid cross reactivity and false positives. While in some cases this criteria can be met using the standard guideline defined above, in others it may require selection of a lower $T_{m,primer}$ value, often equalling or within a °C of T_a. In either case, the $T_{m,primer}$ value of the forward and reverse primers should be the same (± 1 °C), as this improves the likelihood that they exhibit similar amplification efficiency and that neither cross reacts.

- **Setting the annealing temperature T_a and $T_{m,primer}$:** As noted above, T_a should be no more than 5 °C below $T_{m,primer}$, and must often have a value very close to $T_{m,primer}$ so as to avoid unwanted cross reactions leading to false positives. The thermal gradient feature available in current ddPCR instruments allows for facile screening of assay performance at different T_a values, and exploiting that capability is often critical to achieving desired assay performance. But the value of T_a, and thus the target $T_{m,primer}$ used for primer design, can change depending on the GC content of the amplicon and associated priming sites. In particular, efficient amplification of template sequences having high GC content can often require an increase in T_a, with $T_{m,primer}$ increased in kind so that a small $T_{m,primer} - T_a$ value is maintained.

2.3.2 Locked Nucleic Acids and Their Use in Allele-Specific Probes

The use of allele-specific probes to detect somatic point mutations in tumors or premalignant tissues generally relies on the fact that the stability of a duplex formed between two single-stranded (ss) oligonucleotides is sensitive to the presence of any mismatched base pairs within that duplex. As a result (**Figure 2.4**), a duplex formed between a short ssDNA probe and its perfect match (PM) will in general exhibit a higher T_m (here denoted $T_{m,PM}$) compared to that ($T_{m,MM}$) for a duplex between the same probe and a different allele with which it forms at least one base-pair mismatch (MM). Pure-DNA dual-labeled hydrolysis probes are generally 18 or more bases in length. The fractional contribution of any base within a probe, whether paired to its complementary base or not, to the overall thermal stability of a duplex is therefore relatively small (e.g. ~ 1/20 or 5%). This makes it difficult to design a pure-DNA probe where ΔT_m ($=T_{m,PM} - T_{m,MM}$) is sufficiently large so as to achieve selective hybridization of the probe to its target allele at the annealing (and extension) temperature, T_a, of a PCR-based assay designed to detect that allele. This is particularly true for mismatched base pairs known to be only weakly destabilizing to a duplex. It is known that certain mismatched base pairs, most notably G–A and G–T mismatches, weakly destabilize dsDNA, while C–A, T–T, A–A and C–T mismatches are much more strongly destabilizing.[118]

A number of methods and chemistries have therefore been used to increase ΔT_m in cases where standard DNA probes have proven ineffective. These include the use of pure DNA reagents, such as minor-groove binders (MGBs) that serve to alter hybridization thermodynamics[156] or unlabeled oligonucleotides designed to block hybridization of the AS probe to non-target alleles,[157] and the use of various synthetic analogues of DNA or RNA nucleotides that are designed to form more stable base pairs with their complementary deoxy-ribonucleotide.[158, 159] By far the best studied and most widely used of these analogues is the so-called Locked Nucleic Acid, or LNA[15, 25, 26]. As shown in **Figure 2.5**, LNAs contain a methylene bridge that connects the 2'-oxygen and 4'-carbon of a nucleotide's pentose sugar and "locks" that sugar into a C3'-*endo* configuration.[160]Any base within ssDNA can be substituted with the corresponding LNA base, with each such LNA substitution decreasing the entropy of the oligonucleotide due to the

lower number of degrees of configurational freedom available to the pentose sugar when locked.[31-33] Hughesman et al.[161] have shown that hybridization of an LNA-substituted oligonucleotide with its complementary ssDNA therefore occurs with both a lower loss of entropy and an enhancement of favorable base stacking interactions, making the resulting duplex more stable in proportion to the number of locked bases.[47] This enables the use of significantly shorter oligonucleotides as probes, which when combined with experimental evidence that LNAs are in general less energetically tolerant of forming mismatched base pairs,[27] can in turn result in larger differences in ΔT_m.

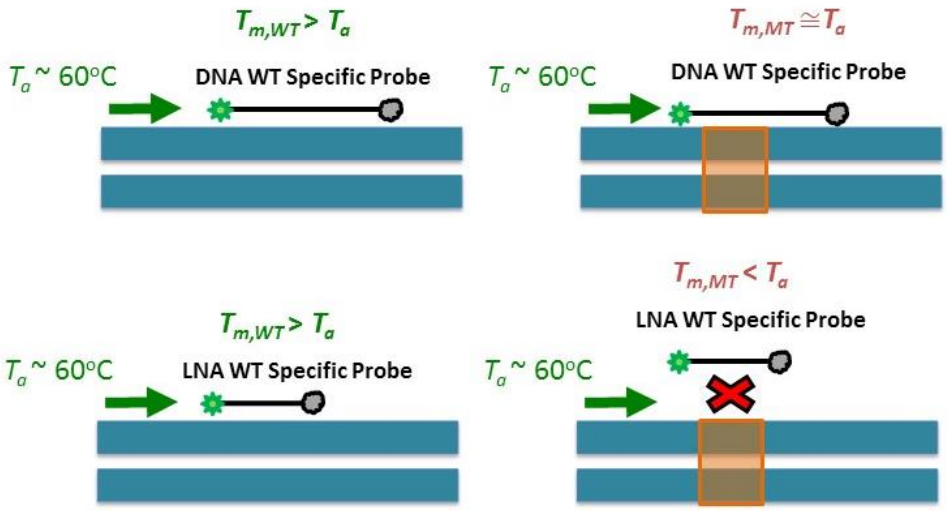

Figure 2.4 Thermal stabilities of pure-DNA and LNA-substituted dual-labeled hydrolysis probes and the dependence of allele-specific assay performance on them. Probes bearing LNA substitutions can display large ΔT_m values when compared to those generally realized using standard pure-DNA probes. As a result, false positives can be eliminated in cases where the corresponding pure-DNA probe cross reacts with non-target alleles.

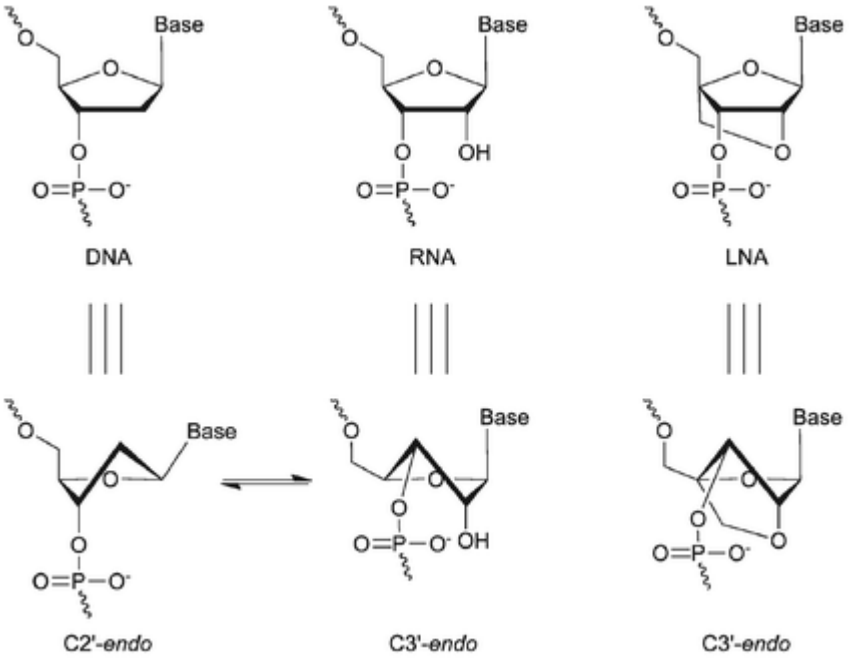

Figure 2.5 Chemical structures of DNA, RNA, and LNA nucleotides. Figure adapted from Campbell and Wengel [54]

AS probes substituted with one or more LNAs have been used in qPCR-based assays to improve specificity and lower the detection limit for target alleles, with LNA-substituted probes against a single specific mutation (e.g. *BRAF* V600E) being by far the most common example.[159, 162] As an example, Denys et al.[163] lowered the detection limit of a qPCR assay against *JAK2* V617F, a driver mutation in myeloproliferative neoplasms, by more than an order of magnitude by replacing a pure-DNA probe with a shorter LNA-substituted probe. LNAs have likewise been introduced into other diagnostic platforms to improve their selectivity, including DNA microarrays[164] and transcript (mRNA) expression profiling panels.[158] However, at present, the LNA substitution sites within the probes and other reagents used in these technologies are selected using either trial-and-error methods or empirical concepts (rules of thumb) drawn from those trail-and-error efforts. The platform described in this thesis, which is expected to enable rapid design of effective ddPCR-based assays capable of highly sensitive detection of somatic mutations, therefore requires a more rigorous, objective method to accurately predict $T_{m,PM}$ and $T_{m,MM}$ values for allele-specific probes as a function of the number and pattern of LNA substitutions.

2.3.3 Nearest-Neighbor Molecular Thermodynamic Models for *in silico* Design of Allele-Specific Probes

DNA (deoxyribonucleic acid) is essential to life and all areas of medical science. As a result, extensive research on the chemistry and properties of DNA has been conducted over the past century. The stability of chromosomal duplex DNA (dsDNA) is sufficient to preserve one's genetic code at physiologic conditions, yet portions of a chromosome can be made to dissociate into single stranded DNA (ssDNA) to permit, among other things, the transcription of genes. In addition to being essential for message and protein synthesis, the ability of dsDNA to dissociate into its component single strands is also exploited in many powerful techniques and technologies used in molecular biology and in clinical laboratories. Of specific relevance to this thesis, hybridization of oligonucleotide probes to ssDNA is used to identify specific sequences that are diagnostic of disease. Likewise, ssDNA primers are used in a wide range of applications, including to initiate complementary strand synthesis for sequencing or PCR-based amplification.

The successful design and application of these ssDNA reagents typically requires knowledge of the T_m of the duplex they are expected to form with their target sequence, and how that T_m depends on probe or duplex length, sequence and concentration. Other solution variables, including salt concentration, pH, and added metal ions or organic solvents, are also known to affect duplex stability.[165, 166] Models and tools to understand and predict the melting properties of dsDNA, particularly short (< 25 bp) complementary dsDNA (e.g. a probe – template duplex), must account for these sensitivities, and their development has been an intense area of research for more than 50 years.[33, 167, 168] While many different types of models have arisen from that collective effort, a certain class, the so-called Nearest-Neighbor Thermodynamic (NNT) models, have proven highly accurate in predicting T_m values and are without question the most widely used.[118, 169-171]

Over the past half-century, an extensive T_m and melting thermodynamics database has been compiled for short B-form dsDNA across a wide range of duplex sequences and lengths, as well as solvent compositions. Collectively, these data show that short dsDNA melting thermodynamics depend not only on duplex length and the number of A–T and G–C base-pairs

(i.e., the pyrimidines, cytosine (C) and thymine (T), and the purines, adenine (A) and guanine (G)), but also on sequence. Spectroscopic studies have shown that the sequence dependence arises, at least in large part, from the greater stability of the G–C base pair and contributions from base-stacking interactions within the B-form duplex.[172] A successful model of dsDNA thermal denaturation must therefore account for both base-pairing and base-stacking interactions. Among the simplest models capable of this is one that assumes that base-pairing and base-stacking contributions can be captured at the nearest-neighbor level. As originally proposed by Gray and Tinoco, Jr., NNT models assume that hydrogen bonds formed between the m^{th} base pair are sensitive to structural and electronic perturbations caused by the neighboring $(m + 1)^{th}$ base pair, and that the energy of stacking interactions between the m^{th} and $(m + 1)^{th}$ base pairs depend only on the types of base pairs in those positions of the duplex.[173] All longer-range contributions are therefore ignored. ΔH_{DNA}, the enthalpy change for the denaturation reaction, may then be computed as a simple summation of the energy ΔH_j^{init} required to initiate denaturation through the dissociation of terminal base pairs (terminal base pairs have unique (typically weaker) energetics due to the fact that they are unbounded on one side), and the energy of denaturation ΔH_{NN_i} for each nearest-neighbor (NN) base-pair *i* within the set of base-pair doublets comprising the duplex:

$$\Delta H_{DNA} = \sum_{j=1}^{2} m_j \Delta H_j^{init} + \sum_{i=1}^{10} n_i \Delta H_{NN_i} \qquad 2.5$$

In equation 2.5, ΔH_{DNA} is the enthalpy change for duplex denaturation at T_m, *j* counts the possible terminal base-pairs (A–T or C–G), m_j is the number (0, 1 or 2) of type *j* terminal base pairs in the duplex, *i* counts the 10 energetically unique Watson-Crick nearest neighbor base pairs, and n_i is the number of each nearest neighbor base pair of type *i* in the duplex. Here it is important to note that Watson-Crick base-pairing requirements reduce the 16 (i.e., 4^2) total nearest neighbor base pairs (doublets) to 10 *energetically unique* nearest neighbors within complementary dsDNA. This is because the two strands are antiparallel; the doublet $N_{3'+m}N_{3'+(m+1)}/N_{5'+m}N_{5'+(m+1)}$ is therefore equivalent to the doublet $N_{5'+m}N_{5'+(m+1)}/N_{3'+m}N_{3'+(m+1)}$.

Entropy is computed under the same assumptions. The total entropy change accompanying the helix-to-coil transition (ΔS_{DNA}) is partitioned into a sum of nearest-neighbor contributions, so that

$$\Delta S_{DNA} = \Delta S^{sym} + \sum_{j=1}^{2} m_j \Delta S_j^{init} + \sum_{i=1}^{10} n_i \Delta S_{NN_i} \qquad 2.6$$

In equation 2.6, ΔS_j^{init} accounts for the unique entropy change of the terminal base pairs; ΔS^{sym}, which is applied only to self-complementary sequences, accounts for the fact that a bimolecular complex formed from self-complementary strands has a rotational symmetry that is not present in a duplex formed from non-self-complementary strands.

In standard NNT models, all terms on the right-hand-side of equations 2.5 and 2.6 are temperature independent. This includes the most widely used NNT model, the so-called "unified" nearest-neighbor model of Santa Lucia Jr. and coworkers,[171, 174] which computes ΔH_{DNA} and ΔS_{DNA} as temperature-independent values by invoking the assumption that ΔC_p, the heat capacity change for the denaturation reaction, is zero. The "unified" NNT model of Santa Lucia Jr. et al. therefore predicts T_m values for a short (<30 bp) complementary B-form duplex using the thermodynamic relation:

$$T_m = \frac{\Delta H_{DNA}}{\Delta S_{DNA} - R\ ln(K)} \qquad 2.7$$

where K is the concentration-dependent equilibrium constant for the denaturation reaction, R is the ideal gas constant (1.987 cal mol^{-1} K^{-1}), and ΔH_{DNA} and ΔS_{DNA} are given by equations 2.5 and 2.6, respectively. K can be computed based on the concentrations of the single strands and knowledge of any duplex symmetry. For a non-self-complementary strand (e.g. 5'-aaaaaaaa-3' cannot form a duplex with itself), K is given by $C_T/4$ when the strands are added in equal concentration. Here, C_T is the total strand concentration. If one strand is added in a greater concentration, $K = C_A - C_B/2$, where C_A and C_B are the more and less concentrated strands

respectively. And finally, in the case of a self-complementary duplex (i.e. $dsDNA \Leftrightarrow 2\ ssDNA_1$), K is equal to C_T. [54]

Recently, Hughesman et al.[54] used differential scanning calorimetry to show that duplex denaturation is accompanied by a positive heat capacity change per base pair, ΔC_p^{bp} of 42 ± 16 cal mol^{-1} K^{-1} bp^{-1}. They found that T_m is therefore more accurately predicted by the thermodynamic relation

$$T_m = \frac{\Delta H^o_{DNA}(T_{ref}) + \Delta C_p(T_m - T_{ref})}{\Delta S^o_{DNA} + \Delta C_p ln(T_m/T_{ref}) - R\ ln(K)} \qquad 2.8$$

where ΔH^o_{DNA} and ΔS^o_{DNA} are now the change in enthalpy and entropy, respectively, for duplex dissociation *at a specific reference state* temperature of 53 °C. They are computed as:

$$\Delta H^o_{DNA} = \sum_{j=1}^{2} m_j \Delta H_j^{init} + \sum_{i=1}^{10} n_i \Delta H^o_{NN_i} \qquad 2.9$$

$$\Delta S^o_{DNA} = \Delta S^{sym} + \sum_{j=1}^{2} m_j \Delta S_j^{init} + \sum_{i=1}^{10} n_i \Delta S^o_{NN_i} \qquad 2.10$$

The standard state nearest-neighbor $\Delta H^o_{NN_i}$ and $\Delta S^o_{NN_i}$ parameters used in equations 2.9 and 2.10 to compute ΔH^o_{DNA} and ΔS^o_{DNA} are provided in **Table 2.1**, along with the required values for ΔH_j^{init}. The contribution to the denaturation entropy arising from sequence symmetry (i.e. self-complementarity), ΔS^{sym}, is also given. In equations 2.9 and 2.10, n_i indexes the total number of times each Watson-Crick nearest neighbor base pair of type i is present in the duplex, and m_j is the number of each terminal base pair of type j.

Prediction of T_m by equation 2.8 requires a value for the heat capacity change, ΔC_p, for the helix-to-coil (dissociation) transition. It is given by $n_{bp}\Delta C_p^{bp}$, where n_{bp} is the total number of base-pairs in the duplex and ΔC_p^{bp} is 42 cal mol^{-1} K^{-1} bp^{-1}.[32]

Applicable to the melting of short complementary dsDNA (< ~ 25 bp), the model of Hughesman et al. is quite accurate, showing a mean error and the standard deviation of -0.2 ± 1.4 °C when applied to a set of 1258 different duplexes with known T_m.[33] Compared to the unified NNT model it is significantly better at predicting T_m values for duplexes melting near or above 65 °C, where the assumption of zero ΔC_p is especially poor. However, it cannot predict T_m values for duplexes in which one of the two strands contains LNAs. Nor can it predict the T_m for a duplex containing both LNAs in one strand and a mismatched base pair which may or may not include an LNA.

Table 2.1 Nearest-neighbor enthalpy and entropy parameters for denaturation of complementary base pair doublets in 1 M NaCl. Parameters taken from[168] with change in sign made to conform to the model developed and presented in this work.

Nearest Neighbor Base Pairs	ΔH_{NN_i} (kcal/mol)	ΔS_{NN_i} (cal/(mol K)
AA/TT	7.9	22.2
AT/TA	7.2	20.4
TA/AT	7.2	21.3
CA/GT	8.5	22.7
GT/CA	8.4	22.4
CT/GA	7.8	21.0
GA/CT	8.2	22.2
CG/GC	10.6	27.2
GC/CG	9.8	24.4
GG/CC	8.0	19.9
Initiation with terminal G–C bp	- 0.1	2.8
Initiation with terminal A–T bp	- 2.3	- 4.1
Chain symmetry correction	0	1.4

A model providing those predictive capabilities was desired in support of the wild-type negative assay development platform central to this thesis work. The resulting new model,[161] which is described in detail in Appendix A, accurately predicts $T_{m,PM}$ and $T_{m,MM}$ values for allele-specific probes as a function of the number and pattern of LNA substitutions. Importantly, it can provide those predictions at the solution conditions used in ddPCR-based wild-type negative assays. The model was co-developed and co-published[161] with others in the laboratory, most notably Dr. Curtis Hughesman, and the associated model parameters, along with example calculations in which the model is used to predict T_m and ΔT_m values for one of the allele-specific probes designed and used in this work, are provided in Appendix A.

2.3.4 Automated Analysis of Assay Output Data to Quantify Mutant Frequencies

As described in section 2.2, output data from a ddPCR-based wild-type negative experiment is most conveniently displayed in the form of a two-dimensional scatter plot (**Figure 2.2**), in which the FAM and HEX fluorescence amplitudes recorded for each droplet are plotted against each other in Cartesian coordinates. In a well-conceived ddPCR experiment designed to detect, discriminate and quantify two (or possibly more) different alleles, the read droplets following ddPCR amplification will segregate into unique groups (clusters) that may include FAM^-/HEX^- (i.e., double negative or empty), FAM^+/HEX^-, FAM^-/HEX^+, and/or FAM^+/HEX^+ droplet clusters. At sufficiently high CPDs, positive/double-positive, double-positive/positive and double-positive/ double-positive clusters may also be observed. In addition, some droplets may record a more ambiguous set of FAM and HEX fluorescent signals that fall between the distinct positive and negative populations defined above. Such droplets are termed “rain” and are generally observed between all clusters.

One must therefore properly assign droplets into groups/clusters based on their fluorescence signals to effectively use ddPCR output data to detect mutations and accurately quantify their frequency. Regrettably, no software was available to objectively conduct those assignments for a wild-type negative assay. Indeed, overall there is a paucity of algorithms and associated software available for analyzing digital PCR data of any type, with the few tools available not applicable to the types of assays described in this work.[175, 176] In collaboration with Dean Attali, a software engineer, I therefore co-developed a tool for automated unbiased assignment of read droplets into

defined clusters. Droplets in each cluster can thereby be reliably counted to detect mutations and accurately quantify mutant frequencies.

The software tool, entitled *ddpcr*, is written in the R programming language[177] and can be used to explore, visualize, and analyze two-channel ddPCR data. The R language was used because it is open-source and cross-platform, allowing anyone to freely use it on any operating system. R is a popular language in the field of computational biology, and is the main data analysis language for many biological and health scientists. To improve access and ease of use, the *ddpcr* software has been implemented as an interactive web resource using Shiny, through which one can apply it to ddPCR datasets using a simple point-and-click interface.[178] The full R script comprising the *ddpcr* algorithm and the web-based Shiny application can be accessed through the link provided in Appendix B, which also provides a detailed description of the algorithm. Key statistical concepts and methods used in the *ddpcr* algorithm to treat the raw datasets include Gaussian distributions and kernel density estimates of distributions. Pertinent basic information relevant to each of those concepts and methods is provided below.

2.3.4.1 Gaussian Distributions and Two-Component Gaussian Mixture Models

As illustrated in **Figure 2.1**, in a typical multiplexed 2-channel ddPCR experiment the set of fluorescence amplitudes for all read droplets recorded in either channel (FAM or HEX) will take on a range of values, with the value for most droplets centered around either a low amplitude or a high amplitude mean. To fix ideas, let us assume for the moment that the cluster of droplets near the low-amplitude mean (e.g. the FAM^- cluster) is composed of droplets whose fluorescence amplitudes are distributed normally about that mean μ. The probability p_X that a given droplet within that population has a fluorescence amplitude X is then described by the Gaussian distribution

$$p_X(X) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(x-\mu)^2}{(2\sigma^2)}\right] \qquad 2.11$$

where σ is the standard deviation. For the case of ddPCR datasets, the fluorescence amplitude is not a continuous function, but rather a set of distinct X_i values from the N droplets within the

cluster. The maximum likelihood estimate for the value of μ is therefore determined by taking the log of equation 2.11 (which gives the log-likelihood), differentiating that with respect to μ, and then finding the maximum

$$\frac{d}{d\mu}\ln\left(p_{X_i}(X_i)\right) = \sum_{i=1}^{N}\frac{1}{\sigma^2}(X_i - \mu) = 0 \qquad 2.12$$

which gives

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad 2.13$$

with the standard deviation of the distribution given by

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)} \qquad 2.14$$

However, as shown in **Figure 2.1**, the complete set of X_i values in a given channel generally partition into two (or more) distinct clusters, each of which can be well-modeled by a Gaussian distribution. Gaussian mixture models are useful for describing such data, with a two-component Gaussian mixture model being appropriate when the X_i data cluster around the means of two different Gaussian distributions.

More generally, a mixture model may be used to describe a set of *K* component distributions that combine into a mixture distribution $f(X_i)$ according to the relation

$$f(X_i) = \sum_{k=1}^{K} \propto_k f_k(X_i) \qquad 2.15$$

where $f_k(X_i)$ is component distribution k, and α_k is its weight within the mixture model such that $\sum_{k=1}^{K} \alpha_k = 1$. Each f_k can represent a distribution of any type, but if $K = 2$ and both f_K are normal distributions (and therefore each f_k is described by the probability density distribution given in equation 2.11) equation 2.15 becomes a two-component Gaussian mixture model of the form

$$p_X(X_i|\mu, \sigma, \propto) = \sum_{k=1}^{K} \propto_k p_k(X_i|\mu_k, \sigma_k) \qquad 2.16$$

The 2-component model therefore requires a total of 6 parameters (μ_1, μ_2, σ_1, σ_2, α_1, and α_2) whose values must be determined through model regression to the data set. Initial estimates for those parameters may be obtained by subjecting the dataset to k-means analysis, where

$$\mu_k = \frac{\sum_{i=1}^{N_k} X_{i,k}}{N_k} \qquad 2.17$$

$$\sigma_k = \sqrt{\frac{\sum_{i=1}^{N_k} (X_{i,k} - \mu_k)}{N_k}} \qquad 2.18$$

and

$$\alpha_k = \frac{N_k}{N} \qquad 2.19$$

Here, N_k is the set (number) of data points in the k^{th} component of the distribution. A number of publicly available algorithms based on the Expectation Maximization method and written in R script are available and may be used to then optimize the fit of equation 2.16 through refinement of the parameter estimates.[179]

2.3.4.2 Kernel Density Estimation of a Distribution of Discrete Data

Computational analysis with the *ddpcr* algorithm of the population of read droplets in terms of the set of fluorescence amplitudes recorded for each is facilitated within certain steps by describing distributions (i.e. $f(X_i)$ in equation 2.15) comprised of discrete fluorescence amplitude values X_i with a smoothed function $\hat{f}(X)$. The main advantage of this is the ability to create a sufficiently smooth density estimation that global features of the distribution (that might otherwise be obscured by the coarseness of the discrete data set) become more apparent and quantifiable. One means of achieving this transformation is the kernel density estimation method, which computes the smoothed distribution function $\hat{f}(X)$, known as the kernel density estimate, from $f(X_i)$ through the relation

$$\hat{f}(X) = \frac{1}{N\lambda}\sum_{i=1}^{N} f\left(\frac{X - X_i}{\lambda}\right) \qquad 2.20$$

where λ is the chosen bandwidth, an adjustable parameter whose value must be set properly. Too large a λ value will over-smooth the density estimate, while too small a value will yield a course "choppy" estimate. Choosing an appropriate bandwidth is therefore important, and useful methods for setting λ are available. One such relation, which was used in this work to gain a good initial estimate, is given by

$$\lambda = \frac{1.06}{N^{1/5}}\,\sigma \qquad 2.21$$

Values both above and below that initial λ value are then screened to see which works best in representing a particular dataset.

Chapter 3: Quantitative Detection and Resolution of *BRAF* V600 Status in Colorectal Cancer Using Droplet Digital PCR and a Novel Wild-Type Negative Assay

3.1 Background

Somatic variations within (proto-)oncogenes and key signaling pathways are screened in cancer testing laboratories to refine disease diagnoses and enable targeted approaches to therapy.[180-182] Detection of somatic point mutations (SPMs) in exons 19 and 21 of the gene encoding epidermal growth factor receptor (EGFR), for example, is used to establish the therapeutic value of EGFR tyrosine kinase inhibitors in treating non-small-cell lung cancer,[183] while SPMs within codons 12 and 13 of *KRAS* are theranostic of response to anti-EGFR antibody based treatment of colorectal cancer (CRC).[184] For these and many other clinically actionable genes, one of the various forms of allele-specific (AS) or mutation-specific (MS) PCR is often used to detect somatic variants within a single codon or adjacent codons.[185-189] Alternatively, an immunohistochemical (IHC) staining assay may be used to detect the resulting amino acid substitution(s) within the gene product.[190] In either case, the assays are generally designed to detect a specific SPM or a limited set of SPMs, even in instances where a larger number of mutations in the target oncogene are known to occur and to be of clinical significance. Examples of assays designed to detect all known mutant (MT) alleles within an oncogene are rare. They generally are qPCR assays utilizing nested AS-primers, significant multiplexing and a relatively large number of reaction wells.[191] The assays are therefore relatively complex in structure.

Comprehensive somatic variant analysis may also be achieved by next generation sequencing (NGS), which is expected to find increasing use in clinical testing of patients as the sensitivity of the method improves, and the cost and throughput of the technology become more manageable to healthcare providers.[192-194] Several bench-top NGS instruments suitable for targeted molecular diagnostics are now available for clinical use. As well, large repositories and panels of primers have been assembled (*e.g.*, AmpliSeq™, TruSeq™) to enable preparation of amplicon libraries targeting oncogenic hot-spot regions that are frequently mutated.[195] In response to these growing capabilities, governments, healthcare providers and payers are actively working toward defining

when NGS may be used in clinical tests, with current updates for reimbursement through the Protecting Access to Medicare Act (see Genetic Tests for Cancer Diagnosis. May 1, 2013; http://www.cms.gov/medicare-coverage-database/) indicating that health insurers in the United States recommend declining reimbursements for sequencing-based somatic profiling in the absence of evidence that 1) clinically actionable genomic alterations are (likely) present in the specimen, 2) the quality and quantity of the genomic DNA (gDNA) recovered from the specimen are sufficiently high, particularly if it is a formalin-fixed paraffin-embedded (FFPE) sample, and 3) the frequency of the mutation(s) is great enough to permit reliable sequence analysis. Current targeted re-sequencing methods generally offer a mutant frequency (MF) detection limit (LOD) of *ca.* 10%, 5% at best, but projected advances in NGS and associated bioinformatics suggest this LOD can ultimately be reduced to levels approaching 1% MF.[196, 197] Though clearly an improvement, the higher depths of sequencing required to achieve this performance will likely decrease NGS throughput. The creation of sensitive and robust AS-probe based droplet-digital PCR (ddPCR) assays that clinics may employ as a simple and inexpensive stand-alone method to define proper courses of therapy through comprehensive profiling of somatic variations in oncogene(s) present in either FFPE specimens or circulating tumor DNA at a MF of 0.1% or higher could therefore be of considerable clinical value. Assays offering such sensitivity would provide a means of rapidly and cheaply identifying patients carrying a prognostic mutation, and for whom acquisition of further sequencing data is justified.

The activating V600E mutation in exon 15 of the v-raf murine sarcoma viral oncogene homolog B1 gene (*BRAF*) on human chromosome 7q34 is present in approximately 40 – 60% of advanced melanomas,[45, 51] as well as in 40 – 80% of papillary thyroid cancers (PTCs) and ~50% of CRCs exhibiting Mut L homologue-1 (MLH1)-associated microsatellite instability (MSI).[52, 53] *BRAF* encodes the serine-threonine protein kinase BRAF associated with the mitogen-activated protein kinase pathway regulating cell expansion, differentiation and apoptosis. FDA-approved therapeutics for *BRAF* V600E positive metastatic or unresectable melanomas that lack evidence of activating mutations in downstream effectors include the mutant BRAF inhibitors PLX4032 (vemurafenib (Roche)) and Tafinlar (dabrafenib (GlaxoSmithKline)).[1, 198] Either small molecule binds the active state of the kinase domain to selectively inhibit the proliferation of cells with unregulated BRAF activity. Resistance to vemurafenib or dabrafenib often occurs within 6 to 12

months, necessitating careful disease monitoring.[199] When relapse is observed, the FDA has approved treatment of *BRAF* V600E positive patients using a combination of dabrafenib and the MEK-inhibitor trametinib; that regimen is also approved for treating *BRAF* V600K positive metastatic or unresectable melanomas.[200] In addition to V600E (79–84% of all V600 coding mutations) and V600K (8–12%), V600R (2–5%), V600M (0.3–4%), V600D (0.3–1.3%), V600G (~1.3%), V600A (< 1%) and V600L (< 1%) mutations are observed, and there is evidence that V600 mutation status correlates with metastasis-free survival and may be associated with other distinct oncological features of melanoma.[19, 56, 201] Importantly, there is increasing evidence that BRAF and MEK inhibitors can be effective in treating late-stage melanoma patients harboring any non-synonymous *BRAF* V600 mutation.[31]

National Comprehensive Cancer Network (NCCN) clinical practice guidelines for colorectal cancer (http://www.nccn.org/professionals/physician_gls/f_guidelines.asp) also recommend *BRAF* mutational testing of MLH1-deficient CRC patients, as the anti-tumor activity of anti-EGFR-antibodies (cetuximab, panitumumab) may be suppressed or lost in *BRAF* V600E positive patients.[202] Moreover, the *BRAF* V600E mutation correlates significantly with adverse pathological features and distinct clinical characteristics of CRC, including altered differentiation mucinous histology, MSI, and CpG (i.e., C-phosphate-G) island methylator phenotype.[203, 204] As with melanoma, other *BRAF* V600 mutations have been observed in CRC, with the total collective incidence of non-synonymous V600 mutations estimated at 10 – 15%, including a ~ 50 – 60% frequency in MLH1-deficient CRCs.[205]

BRAF status is also being considered as a prognostic biomarker for papillary thyroid cancer, as disease progression and reoccurrence correlate significantly with *BRAF* V600E.[206] Correct identification of V600 status is therefore integral to understanding and treating cancers in which BRAF activity plays an important role. A clinical pipeline that permits rapid, accurate and sensitive detection of *BRAF* V600E, as well as tumors harboring a rare non-synonymous V600 mutation at a frequency suitable for NGS analysis, could serve to improve clinical and pathological staging to achieve better management of *BRAF*-related cancers.

Toward that goal, a diagnostic technology is developed and presented here that leverages the unique capabilities of droplet digital PCR (ddPCR) to rapidly and effectively discriminate patients carrying wild-type *BRAF* from those carrying a *BRAF* V600E mutation or any other V600 mutation. The nearest-neighbor type molecular thermodynamic model described in Chapter 2 and Appendix A[32, 161] is used to design a locked nucleic acid (LNA) substituted dual-labeled hydrolysis probe against *BRAF* V600E1 (Val600Glu c.1799*t*>*a*) to quantitatively detect that clinically relevant somatic variation. Those models, as well as a corresponding model for un-substituted DNA, are also used to optimize two dual-labeled hydrolysis probes against WT *BRAF* over the hot-spot oncogenic region within codons 598 to 603: one in a standard pure-DNA format, and the other a LNA/DNA chimera.[33] Either of these WT-specific probes is expected to show sufficiently low cross-reactivity to any known V600 mutation that MT and WT alleles can be unequivocally discriminated. Data are presented showing that this essential capability is not provided by the standard pure-DNA probe, but is fully realized by the LNA-substituted probe when applied in a ddPCR format. A second LNA-substituted probe targeting a highly conserved 16-nt sequence within *BRAF* is used to quantify the total number of amplifiable *BRAF* templates within the gDNA sample, permitting the MF to be quantified with high accuracy to a limit of detection (LOD) of 0.05% while also defining the total quantity of amplifiable DNA present in the sample. The method offers advantages over either AS/MS-PCR or IHC assays through its sensitivity to all *BRAF* V600 mutations; it likewise holds advantages over high-resolution melt analysis, particularly when applied to genomic DNA recovered from FFPE samples, by not only detecting the presence of a mutation, but also quantifying the mutant frequency and the mass of high-quality amplifiable DNA present in the sample: two parameters essential to properly assessing the potential success of a subsequent NGS run and the depth of sequencing that would be required.

The performance of this novel ddPCR *BRAF* V600 status assay is demonstrated through successive application first to a set of plasmid-DNA standards, each presenting a specific V600 mutant allele, over a range of mutant frequencies, and then to sets of reference cell lines and FFPE tissues. Finally, clinical utility is assessed by comparing results from the new ddPCR *BRAF* V600 status assay to those from a validated *BRAF* V600E IHC assay when applied to FFPE tumor specimens from 41 MSI-positive CRC patients.

3.2 Materials and Methods

3.2.1 Oligonucleotides

All primers, pure-DNA and LNA-substituted dual-labeled hydrolysis probes, and WT and MT *BRAF* alleles were purchased from IDT, Inc. (Coralville, IA). Probes were HPLC purified, while primers and templates were purified by desalting. Purified primers, probes, and templates were resuspended to 100 µM in TE (10 mM Tris, pH 8.0, 0.1 mM EDTA) buffer and stored at –20 °C prior to use.

3.2.2 Primers and Probes Design

Forward (FP) and reverse (RP) primers were designed using Primer3.[155] Primers used to amplify a 165 bp fragment spanning across the V600 codon were designed within the *BRAF* intron/exon 15 boundary. All primers were analyzed by primer-BLAST to find any sequence similarities within the human genome database.

Table 3.1 Sequence and Concentrations of Primers and Dual-Labeled Hydrolysis Probes Used in the ddPCR-based BRAF V600 Status Assay. LNAs shown in bold.

Primer/ Probe	**Sequence**	**Conc.** (µM)
Forward Primer	5'-CTACTGTTTTCCTTTACTTACTACACCTCAGA-3'	0.9
Reverse Primer	5'-AGCCTCAATTCTTACCATCCA-3'	0.9
LNA V600E1 Probe	5'-(6-FAM)/AGATTT**CTC**T**GT**AGC/(BHQ1)-3'	0.25
DNA WT V600 Probe	5'-(6-FAM)/CATCGAGATTTCACTGTAGCTAGACC/(BHQ1)-3'	0.25
LNA WT V600 Probe	5'-(6-FAM)/CGAGATTT**CACTGT**A/(BHQ1)-3'	0.25
Consensus *BRAF* Probe	5'-(HEX)/TCCCATCAG/ZEN/TTTGAACAGTTGTCTGG/(IABkFQ)-3'	0.25

6-FAM = 6- carboxyfluorescein; HEX = hexachloro-fluoroscein; BHQ1 = Black Hole Quencher 1; ZEN = the ZEN internal quencher; IABkFQ = Iowa Black fluorescence quencher; the consensus probe is labeled with HEX in well 1 (as shown), but with 6-FAM in well 2 of the assay.

Each dual-labeled hydrolysis probe (**Table 3.1**) was likewise engineered to minimize PCR artifacts and to selectively hybridize to a specific sequence within *BRAF* at ddPCR reaction conditions. The consensus probe, a 16 nucleotide (nt) FAM or HEX-labeled/BHQ1-quenched probe spanning *BRAF* codons 607 to 612, was designed using the NNT melting thermodynamics model to i) define the combinations of LNA substitutions needed to limit probe length to within a highly conserved region of exon 15, and ii) to achieve a melting temperature (T_m) of *ca.* 66 – 67 °C when duplexed to WT *BRAF*. As detailed in Chapter 2 and Appendix A, that model accounts for the effects of PCR solution conditions and the addition of a fluorescent reporter dye (HEX or FAM) and quencher (e.g., BHQ1 – Black-Hole Quencher 1) on probe-template melting thermodynamics. Potential probe-derived PCR artifacts were identified and avoided using primer-Blast software.[207] The LNA *BRAF* V600 WT-specific probe and the LNA *BRAF* V600E1 specific probe spanning *BRAF* hot-spot codons 598 to 603 were designed in a similar manner. There, however, LNA-DNA mismatches must be accounted for in the model to identify probe sequences that minimize probe cross-reactivity to non-target alleles. Finally, a 26 nt pure-DNA version of the WT-specific probe was designed as a benchmark for evaluating the performance improvements conferred by LNA substitutions. Melting thermodynamics required for WT-specific DNA probe design were predicted using a corresponding melting thermodynamic model developed by Hughesman *et al.*[33] for un-substituted DNA duplexes. Model predicted T_m values for probes were verified experimentally by UV melt spectroscopy according to standard methods.

3.2.3 Tumor and Reference Samples, and DNA Extraction Protocols

Formalin fixed paraffin-embedded (FFPE) tissues from a cohort of metastatic CRC (mCRC) patients were obtained from Lion's Gate Hospital (North Vancouver, Canada) with approval from the UBC Clinical Ethics Research Board (CREB) number H14-00577. The colorectal carcinoma cases were selected from a pool of MLH1-deficient tumors, identified by IHC testing as part of the population-based Vancouver Coastal Health [172] Lynch-syndrome screening program. Three non-MLH1 deficient cases (samples 3, 5 and 20) were also included. Specimens were derived from 10% neutral buffered formalin fixed resections in all but one case (sample 37), which was a formalin fixed malignant ascites fluid specimen. The original diagnosis was confirmed by a pathologist (Dr. Robert Wolber, Lions Gate Hospital, North Vancouver, BC, Canada), and an optimal tumor block containing invasive carcinoma was selected for sampling.

Two cores of each carcinoma were taken for construction of tumor microarray (TMA) blocks, and two additional cores were taken for the ddPCR studies. Four FFPE standards, each harboring either WT *BRAF* or a verified frequency of a particular V600 mutant allele (0.8% V600K MF, 1.4% V600E, or 50% V600R), were purchased from Horizon Discovery (HDx) Ltd. (Cambridge, UK). DNA from FFPE reference standards and CRC FFPE cores was purified using the QIAamp DNA FFPE tissue kit (Qiagen, Inc.; Santa Clarita, CA) under a modified protocol for the xylene-assisted paraffin removal step that serves to reduce shearing of genomic DNA during that process. Xylene (0.8 mL) was added to FFPE cores (two) collected in 1.5 mL micro-centrifuge tube) and the mixture equilibrated by rotational mixing for 10 minutes. The sample was then centrifuged at 14,000 rpm for 1 min to collect tissue. The waste xylene was removed and the process repeated twice more. Three successive washes with 0.8 mL of 100%, 100% and 70% ethanol, respectively, were then completed using the same procedure. Finally, the sample was placed in a Vacufuge™ (eppendorf, Hamburg, Germany) for 5 minutes at medium heat to dry, and the residue carefully removed for further processing according to the standard QIAamp DNA FFPE protocol.

Purified gDNA from the SW480 cell line (WT *BRAF*) and the YUMAC cell line (homozygous for *BRAF* V600K) was kindly provided by the BC Cancer Agency. pIDTSMART-AMP plasmids, each containing a 280 bp gene fragment (exon 15) of either WT *BRAF* or a *BRAF* V600 MT allele, were purchased from IDT.

DNA concentration was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific; Waltham, MA). As plasmids exhibit supercoiling, pIDTSMART plasmid DNA (1.5 µg) was linearized with ClaI restriction enzyme as per the manufacturer's instructions (Life Technologies; Carlsbad, CA) prior to concentration determination and use in assays.

3.2.4 ddPCR Assay Workflow and Characterization

Digital PCR mastermix solutions (20 µL) were prepared from 2X dUTP-free ddPCR supermix for probes (BioRad, Inc.; Hercules, CA), 900 nM each of FP and RP, 250 nM of the consensus probe, 250 nM of either the LNA *BRAF* V600E1 probe (well 1) or the LNA *BRAF* V600 WT-specific probe or pure-DNA *BRAF* V600 WT-specific probe (well 2), and an amount of purified

plasmid or gDNA containing ~10000 copies of *BRAF* (gDNA purified from clinical colorectal FFPE samples sometimes contained fewer copies (~ 5000 – 8000) of amplifiable *BRAF*). No template control (NTC) samples were prepared in IDTE buffer. Emulsified sub-nL reaction droplets were created by introducing a 20 µl sample into a well of an eight-channel disposable DG8 droplet generator cartridge and adding 60 µl of droplet generation oil (BioRad). 40 µl of the resulting droplet emulsion (~13000 to 15000 droplets) generated on the BioRad QX-100 Droplet Generator were transferred by multichannel p100 pipette to an Eppendorf Twin.tec semi-skirted 96-well PCR plate, which was then heat-sealed with foil sheets. The droplet emulsions were thermally cycled in the CFX96™ thermocycler using the following protocol: denaturation at 95 °C for 10 min, followed by 50 cycles of 94 °C for 30 s and 60 °C for 1 min (ramp rate set to 2.5 °C/s). The end-point fluorescence of each thermally cycled droplet was measured using the QX100 Droplet Reader (BioRad). After cycling, a final 10 minute hold at 98 °C is applied to deactivate the enzyme and stabilize the droplets. The raw ddPCR data was collected and visualized using the QuantaSoft v1.2 program (BioRad). MF values were then determined from droplet counts either through operator-based assignment of MT and WT data clusters (hereafter referred to as the manual method) or using an automated data analysis algorithm described below.

Serial dilutions (1000 – 1 copies/µl) of the WT *BRAF* isolated from plasmid DNA, as well as each of 10 MT *BRAF* alleles, were prepared in IDTE buffer [158], with the concentration of *BRAF* template (copies/ul) in each dilution quantified on a BioRad QX100 ddPCR instrument. For dynamic-range measurements, 1000 – 1 copies/µl of BRAF V600 MT plasmid-derived DNA was prepared in a high fixed background (10,000 copies/µl) of WT *BRAF* plasmid DNA. Serial dilutions (1000 – 1 copies/µl) of *BRAF* V600K gDNA from YUMAC cells, as well as *BRAF*^{V600K} (YUMAC cells) in WT *BRAF* (SW480 cells), were prepared in the same manner.

3.2.5 ddPCR Raw Data Analysis Algorithm

The droplet event data from a ddPCR-based *BRAF* WT-negative assay are exported from QuantaSoft V1.3.2 into custom software I co-developed that analyzes the data to assign each read droplet as either empty, “rain” (droplets having a signal lying within an indeterminate region between a pair of distinct positive and negative droplet clusters), MT-positive, or WT-positive.

The assigned droplets are then used to detect MT alleles and compute the MF in each well. The software was built in the R programming language[177, 179] and is available on GitHub at https://github.com/daattali/ddpcr, with a detailed description of its features and computational elements provided in our publication.[178] An illustration of the data analysis process and a brief overview of the core computational steps of the algorithm are provided in **Figure 3.1**.

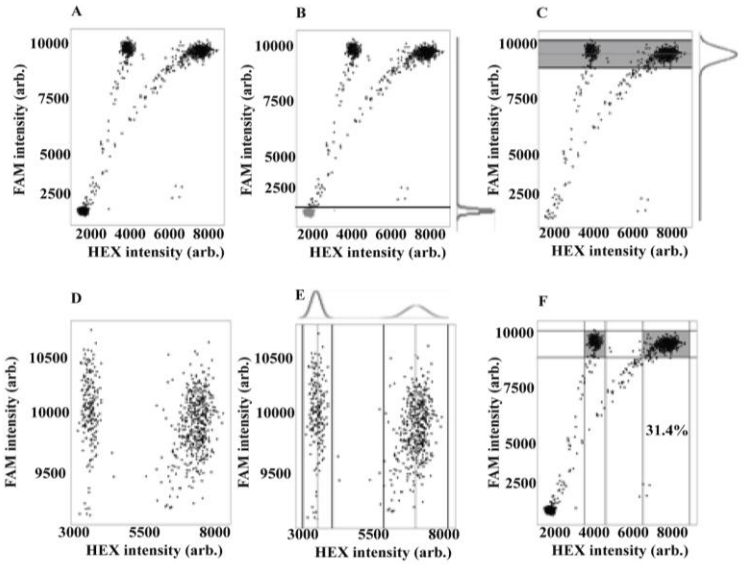

Figure 3.1 Automated Data Analysis Overview for the ddPCR-Based BRAF Status Assay. Representative output data for the WT negative portion (well 2) of the assay are visualized in a 2D scatter plot. (A) Empty droplets (marked as grey dots) are identified by fitting two normal distributions to the signal from the FAM fluorescence amplitude channel, one of low mean intensity (B) and the other of high mean intensity (C). All droplets (B) below a threshold FAM intensity (horizontal black line) set at the mean of the lower-intensity distribution plus 7 standard deviations (SDs) are deemed empty and removed from the analysis. All filled droplets are then identified (C) as lying within the mean ± 3 SDs of the upper distribution (grey band). Droplets displaying FAM intensities outside this range (often referred to as ddPCR "rain") are taken as being poor in signal quality and excluded from the analysis (D). *BRAF* V600 WT-negative and WT-positive droplets, respectively, are identified by fitting two normal distributions to the signal from the HEX fluorescence amplitude channel (E), with the relevant populations bounded by the mean ± 3 SDs in each case. The distribution having the lower mean HEX fluorescence amplitude contains the population of WT-negative droplets. The use of this automated algorithm, the full details of which are presented in Appendix B, to analyze the raw ddPCR data in (A) to determine the population of WT-negative (left grey box) and WT-positive (right grey box) droplets, gives a mutant frequency value (calculated as the ratio of WT-negative droplets to total droplets) of 195/(195 + 426)*100% = 31.4%.

3.2.6 BRAF V600E Mutation-Specific Antibody Staining

The colorectal carcinoma TMA was cut at 4-micron thick sections for IHC studies. Deparaffinized sections were stained on a Ventana (Tucson, AZ) Benchmark XT automated immuno-stainer using a 32 min CC1 heat induced antigen retrieval protocol. Incubation with VE-1 mouse monoclonal primary antibody (Spring Biosciences, Pleasanton, CA) directed against BRAF V600E protein was for 16 minutes at 37 °C. Antibody detection was done by the Ventana Optiview system using DAB as a chromogen. Previous experience with this antibody demonstrated that homogeneous cytoplasmic staining of carcinoma cells indicates the presence of BRAF V600E protein. Nonspecific nuclear staining, readily distinguished from cytoplasmic staining, was encountered in both benign and malignant colorectal epithelium in some cases. Blinded IHC scoring was performed by two pathologists, and a histotechnologist over a multi-headed microscope. TMA cases scored as indeterminate were retested by IHC on the whole section from which the corresponding TMA core was taken, and then rescored.

3.3 Results

3.3.1 Assay Design and the Unique Advantages of ddPCR

Our assay of *BRAF* V600 status is a two-well test utilizing ddPCR for target amplification and detection. In both wells, gDNA purified from an FFPE tissue specimen is loaded to a copies-per-drop (CPD) of 0.2(±0.05), ensuring that at the start of the PCR most droplets contain either 0 or 1 copy of a *BRAF* gene. Approximately 10000 copies of *BRAF* are therefore analyzed per test. Well 1 (**Figure 3.2A**) also contains ddPCR supermix for probes without dUTP (BioRad), as well as primers and two dual-hydrolysis probes whose sequences and labeling chemistries are reported in **Table 3.1**. The resulting amplicons (WT and MT) are 165 bp in length, spanning most of *BRAF* exon 5 including codons 598 to 603. During amplification, fluorescence generated from hydrolysis of the *BRAF* consensus probe confirms the presence of a *BRAF* gene (WT or MT) in the droplet, allowing the total amplifiable copies of *BRAF* in the specimen to be quantified by HEX-positive droplet counts and Poisson statistics. Well 1 also contains a *BRAF* V600E1 MT-specific probe that is model-designed (see below) to contain a pattern of LNA-substitutions that ensures cross-reaction of the probe to WT *BRAF* or other V600 MT alleles is sufficiently low at ddPCR conditions to avoid overlap of *BRAF* V600E1 and non-$V600E_1$ data

fields. Droplets clustered in the HEX-positive/FAM-positive quadrant therefore unambiguously quantify *BRAF* V600E1 alleles, while those displaying a HEX-positive/FAM-negative signal quantify all WT plus non-V600E$_1$ alleles, from which the frequency of *BRAF* V600E1 within the specimen may be quantified. Well 1 therefore is a standard AS/MS-PCR assay that targets and quantifies a MT allele (*BRAF* V600E1) within a high background of WT allele. It is conducted here in a ddPCR format, but other studies have shown that it can also be effectively performed in a standard qPCR format using either an AS-primer or an AS-probe, often in the presence blocking agents that minimize cross-reactivity.[24]

In the second well (**Figure 3.2B**), our novel "WT-negative" screen is applied to the gDNA sample to collectively detect and quantify all *BRAF* V600 mutations. The *BRAF* consensus probe, now FAM labeled, functions as described above. The second probe is designed to selectively hybridize only WT *BRAF* and to thereby discriminate WT *BRAF* (FAM-positive/HEX-positive signal) from any *BRAF* allele carrying a V600 mutation (FAM-positive/HEX-negative signal). In both wells, droplets recording a HEX-negative/FAM-negative signal contain no *BRAF*. Digital PCR offers capabilities essential to effective execution of the WT-negative screen. To fix ideas, consider a gDNA specimen harboring 10,000 total copies of *BRAF*, 10% of which are a MT *BRAF* allele and the remainder WT *BRAF*. In the WT-negative screen, the probe targets the WT allele. If the screen were conducted in a qPCR format, the decrease in copies of WT *BRAF* from 10,000 (no MT alleles) to 9,000 (10% MT frequency) would, in theory, result in a C_q (quantitation cycle) change of *ca*. 0.1, which cannot be measured with statistical significance (ΔC_q errors in AS-PCR assays are typically *ca*. ± 0.3). Conversely, when the WT-negative assay is conducted in a ddPCR format, templates are partitioned into individual droplets, enabling droplets containing WT *BRAF* to be fully differentiated from those containing MT *BRAF* through their end-point HEX intensities. Very small populations of MT *BRAF* may thereby be reliably detected and quantified.

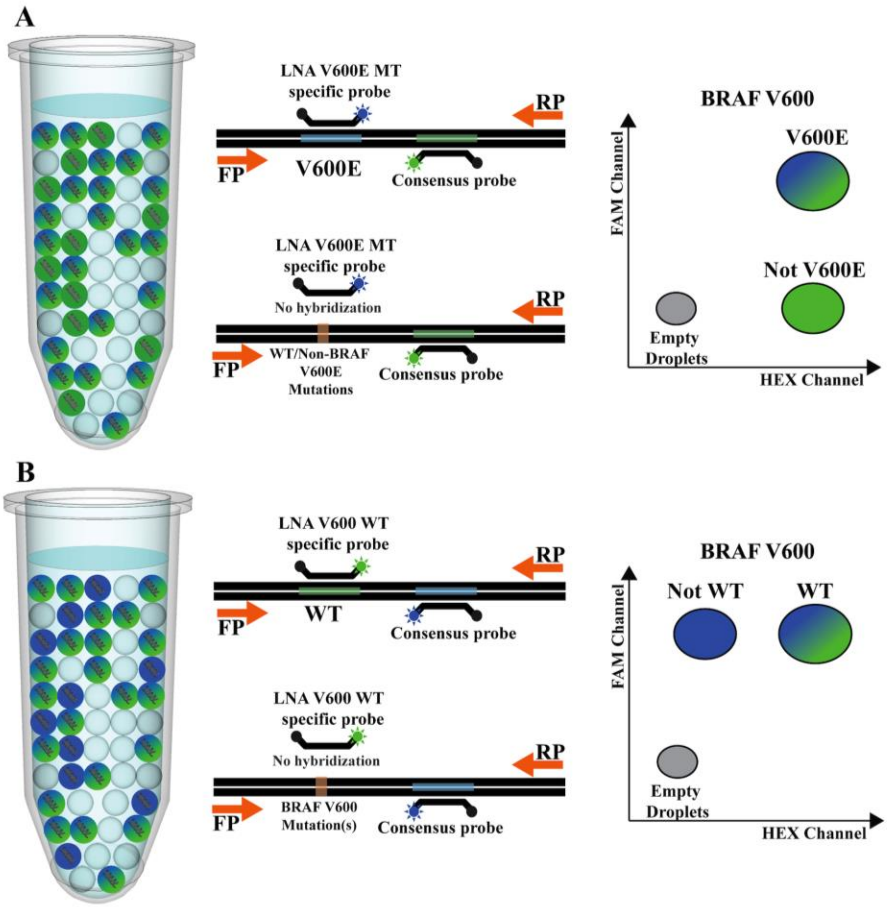

Figure 3.2 Schematic of the ddPCR-Based BRAF V600 Status Assay for Identifying WT, V600E1 and Less Common BRAF V600 Mutant Alleles in CRC Tumor Specimens. Water-in-oil emulsion droplets are generated to contain ddPCR master-mix and a small aliquot of genomic DNA containing, on average, 0.2 to 0.4 copies of *BRAF* per droplet (CPD). The two-well assay utilizes forward and reverse primers that amplify a 165-bp fragment of the *BRAF* gene spanning codon V600. Two dual-labeled hydrolysis probes spanning *BRAF* codons 598 to 603 are employed. The first, used in well 1 (A), is a FAM-labeled LNA-substituted probe designed to selectively hybridize and unequivocally detect *BRAF* V600 E1 by showing no cross-reactivity to WT *BRAF* or any other clinically relevant *BRAF* alleles. The second, employed in well 2 (B), is a HEX-labeled probe designed to hybridize only to WT *BRAF* and thereby distinguish WT *BRAF* V600 from all *BRAF* V600 MT alleles. A "consensus" probe that binds to a highly conserved sequence within the *BRAF* amplification fragment is also added to each reaction well. Using well 2 (B) as an illustration of assay mechanics, when a *BRAF* V600 WT allele (upper duplex of Figure B) is amplified, end-point fluorescence signals from the WT-specific probe (green) and the consensus probe (blue) are both detected. Amplification of any *BRAF* V600 MT allele (lower duplex of Figure B) results in the generation of an end-point signal from the consensus probe only. In the resulting ddPCR 2D plot (output data), droplets containing amplicons of a V600 WT allele will cluster in the FAM-positive/HEX-negative quadrant (top left). Droplets containing no *BRAF* will cluster in the bottom left (FAM-negative/HEX-negative) quadrant.

Together, the combined ddPCR output from the two wells permits unequivocal stratification of tumors bearing WT *BRAF* from those carrying either *BRAF* V600E1 or a rare V600 mutation, as well providing values for MF and the total abundance of amplifiable *BRAF*.

3.3.2 Engineering Allele Specificity into Probes Used in ddPCR Assays

In addition to the unique capabilities offered by ddPCR, successful execution of the two-well *BRAF* V600 status assay described in **Figure 3.2** requires the two probes against the *BRAF* V600 region to be sufficiently specific to their target allele that cross-reactivity at PCR conditions does not diminish assay performance. Designing a real-time probe to unequivocally discriminate a target allele from an ensemble of alleles that may differ from the target sequence by as little as a single base is, in part, a problem rooted in chemical equilibria. Taking design of the probe against WT *BRAF* V600 (**Figure 3.2B**) as an example, it requires creating sufficient difference ($\Delta T_{m,(WT-MT_i)}$) between $T_{m,WT}$, the melting temperature (T_m) of the perfectly matched duplex formed by the probe and its WT target, and T_{m,MT_i}, the T_m for the mismatched duplex formed by the probe and any mutant allele *i*. In particular, $T_{m,WT}$ must exceed T_a, the annealing temperature of the PCR assay, such that the duplex formed between the probe and WT template is sufficiently stable that creation of each WT amplicon results in hydrolysis of the bound probe and reporter release. T_{m,MT_i} must then be low enough that concurrent amplification of any MT allele *i* at T_a results in a negligible signal from the WT-specific probe over the course of the ddPCR run. If melting heat capacity effects, which are small,[37] are ignored, the fraction α of MT template *i* to which a WT-specific probe is hybridized at T_a can be estimated from the thermodynamic relation

$$ln\left(\frac{C_T(1-\alpha)^2}{2\alpha}\right) = -\frac{\Delta G_{MT_i}(T_a)}{RT_a} \qquad 3.1$$

where $\Delta G_{MT_i}(T_a)$ is the Gibbs energy change for the melting transition at T_a and C_T is the total strand concentration, probe plus allele. Equation 3.1 shows that the value of α is a metric of the cross-reactivity of the probe; moreover, through its dependence on $\Delta G_{MT_i}(T_a)$, α also provides an indication of the stability of the duplex formed between the probe and mutant template *i*.

Previous qPCR studies suggest that, at least qualitatively, probe efficiency (fraction of amplification events that result in probe hydrolysis) decreases with decreasing stability of the probe:template duplex.[46] This is confirmed in **Figure 3.3**, which shows that the dependence of probe efficiency on duplex stability causes the end-point fluorescence in a qPCR assay to decrease rapidly and nonlinearly with decreasing α. As a result, total elimination of probe cross-reactivity (*i.e.* $\alpha = 0$) is not required to discriminate WT and MT alleles in either a qPCR or ddPCR assay. For example, in the ddPCR assays described here, we find that an $\alpha < ca.$ 0.3 at T_a is sufficient to eliminate a false WT end-point fluorescence signal arising from cross-hybridization of the WT-specific probe.

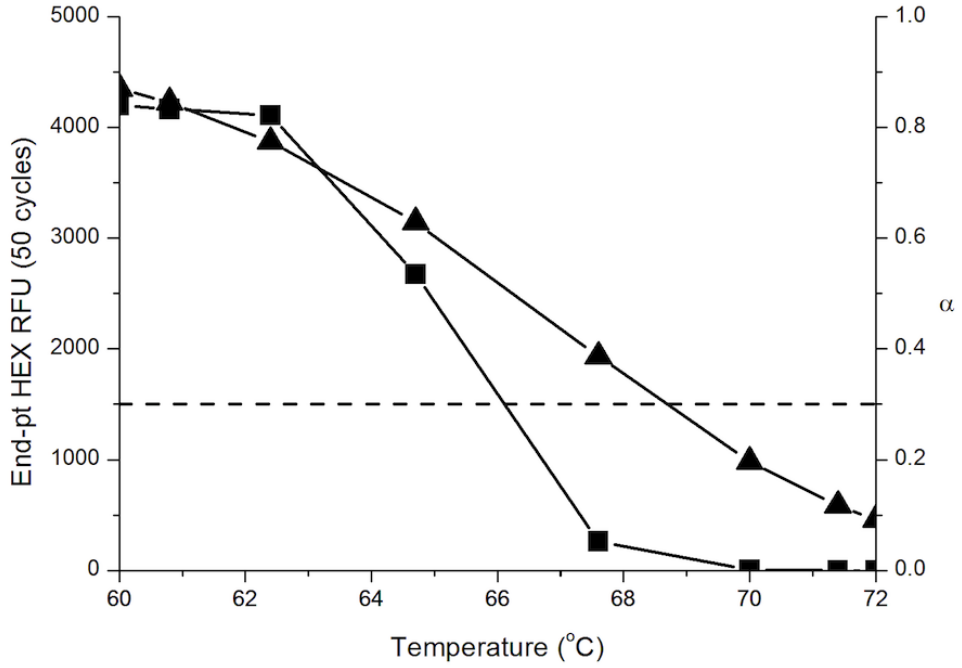

Figure 3.3 General relation between the end-point fluorescence (filled squares; left y axis) and the fraction α of template bearing a dual-labeled hydrolysis probe (filled triangles; right y axis) as a function of the PCR annealing temperature Ta. Data reported for end-point fluorescence monitoring of qPCR amplification of WT *BRAF* (plasmid; C_T = 0.5 µM) using the LNA *BRAF* V600 WT specific probe (Table 3.1). Data point represent mean values for triplicate runs at each T_a, with the size of the data points shown commensurate with the standard deviation for that data point having the largest experimental error.

The lack of appreciable end-point fluorescence at $\alpha < 0.3$ may be related to findings of Holland *et al.*,[208] who reported that during polymerization *Taq* displaces the first two or so bases it encounters before cleaving at that site. Thus, probe hydrolysis by *Taq* requires lifting the 5'-end of the probe off of the template. This destabilizes the probe:template duplex such that, when $\alpha <$ 0.3, intact probe release occurs in lieu of fluorescent signal generation, which only occurs if the probe remains duplexed with template through cleavage of the labeled 5'-base.

Based on these concepts, a standard pure-DNA dual-hydrolysis probe against WT *BRAF* V600 (**Table 3.1**) was designed *in silico* using the models of Hughesman *et al.*[33] and SantaLucia, Jr.[168] to melt at a $T_{m,WT}$ of *ca.* 66 – 68 °C at ddPCR solution conditions, while also offering the best possible discrimination ($\Delta T_{m,(WT-MT_i)}$) against all clinically relevant SPMs in codon V600. For that probe, both model-predicted and experimental $\Delta T_{m,(WT-MT_i)}$ values are small (generally ≤ 7 °C) for all *BRAF* MT alleles differing from WT *BRAF* by a single SPM within codon V600 (**Table 3.2**), reflecting the relatively low proportional contribution of the mismatch to the thermal stability of the cross-hybridization duplex. The consequences of this are reflected in **Figure 3.4A**, which overlays ddPCR output data for the WT-negative assay (**Figure 3.2B**) sequentially applied to WT *BRAF* and to a representative set of clinically relevant V600 coding mutations presented in the form of purified linearized plasmid DNA. Application of the WT-negative screen against a WT *BRAF* sample generates a tight cluster of droplets displaying a strong FAM-positive/HEX-positive signal. For MT alleles carrying two SPMs in codon V600 (e.g. V600E2, V600K), $\Delta T_{m,(WT-MT_i)}$ is sufficiently large that an appreciable HEX-positive signal resulting from cross-hybridization of the WT-specific probe is avoided; a tight cluster of FAM-positive/HEX-negative droplets showing no overlap with the WT *BRAF* cluster is therefore observed. However, for alleles having a single missense point mutation in codon V600, the MT-positive cluster tends to carry a significant HEX-positive signal; overlap of WT and MT data clusters is therefore observed for several V600 mutants (G, A, and most-significantly E1), negating the ability to unequivocally discriminate WT and MT V600 alleles. For those cross-reactive MT alleles (**Figure 3.4A**), $\Delta T_{m,(WT-MT_i)}$ is low (< 6 °C) and α is > 0.3.

Table 3.2 Experimental and model predicted predicted $\Delta T_{m,WT-MT_i}$ **values for mismatched duplexes formed between a BRAF V600 MT allele and either a pure-DNA or LNA-substituted BRAF V600 WT specific probe (Table 3.1).**
Also reported are prevalence data for *BRAF* V600 missense mutations in melanoma[19, 56] and α values for V600 MT alleles bearing a single point mutation.

BRAF* V600 Mutation *i		**Mismatch**	**Prevalence**	**DNA WT *BRAF* V600 Probe**			**LNA WT *BRAF* V600 Probe**		
				$\Delta T_{m,WT-MT_i}$ (K)		α	$\Delta T_{m,WT-MT_i}$ (K)		α
				Pred[¶]	Expt		Pred	Expt	
V600E1	c.1799 T>A	A/A	79 – 84%	4.3	4.2	0.85	11.8	11.0	< 0.01
V600E2	c.1799_1800 TG>AA	CA/AA	-[£]	-[§]	8.5	-	-	23.9	-
V600K	c.1798_1799 GT>AA	AC/AA	8 – 12.4%	-	7.4	-	-	23.8	-
V600R	c.1798_1799 GT>AG	AC/AG	2.2 – 5%	-	6.6	-	-	18.5	-
V600M	c.1798 G>A	C/A	0.14 – 4%	4.6	6.2	0.69	12.0	14.1	< 0.01
V600D	c.1799_1800 TG>AT	CA/AT	0.3 – 1.3%	-	8.6	-	-	24.7	-
V600G	c.1799 T>G	A/G	0 – 1.3%	1.9	1.4	0.89	7.7	7.6	0.21
V600L1	c.1798 G>C	C/C	0 – 0.29%	7.6	7.4	0.27	16.9	17.8	< 0.01
V600L2	c.1798 G>T	C/T	NA*	5.1	6.6	0.63	12.6	15.4	< 0.01
V600A	c.1799 T>C	A/C	NA	4.9	4.6	0.66	12.0	11.6	< 0.01

[£] The *BRAF* V600E1 and V600E2 missense mutations collectively occur at a prevalence of 79 – 84%, with the V600E1 mutation representing most of that prevalence; * NA – no data available; [¶] Experimental (Expt) $\Delta T_{m,WT-MT_i}$ values were measured by UV melt spectroscopy at a C_T of 2 µM (50 mM K^+; 3 mM Mg^{2+}; pH 7) and then corrected to the ddPCR C_T of 0.5 µM, and predicted (Pred) $\Delta T_{m,WT-MT_i}$ values were computed using the NNT model described in Chapter 2 and Appendix A; [§] that model does not permit prediction of T_m values for duplexes containing two adjacent mismatched base pairs.

Alternative strategies for achieving suitable $\Delta T_{m,(WT-MT_i)}$ and α values for each possible V600 coding mutation include chemically modifying the V600 WT-specific probe through either addition of a 3' terminal minor groove binder (MGB) ligand[209] or by substituting nucleotides within the probe with their corresponding locked nucleic acid (LNA).[210] We pursued the latter approach. **Table 3.1** reports an LNA-substituted probe against WT *BRAF* V600 designed in the same *in silico* manner to achieve a $\Delta T_{m,(WT-MT_i)}$ > 7 °C and an α < 0.3 for all known V600 mutations (**Table 3.2**).

Use of the LNA/DNA chimeric probe in the WT-negative screen then results in complete segregation of the WT data cluster from all MT data clusters (**Figure 3.4B**), permitting unequivocal detection of a V600 mutation and determination of the mutant frequency through MT and WT droplet counts. Design of the *BRAF* V600E1 allele-specific probe used in well 1 of the ddPCR assay followed the same strategy and yielded comparable performance results.

3.3.3 Assay Limit of Detection on Plasmid, Cell-Line and FFPE Standards

Serial dilutions ($n \geq 8$ for each dilution) of *BRAF* MT plasmid DNA into *BRAF* WT plasmid DNA to MT frequencies down to 0.01% were used to define the LOD in each reaction well. Due to the greater challenge of achieving the probe specificity required in the WT-negative screen, the LOD recorded in well 2 defined the overall LOD for the two-well assay. For serial dilutions of V600K into WT *BRAF* plasmid DNA, a LOD of ≤ 0.05% MF was recorded (**Figure 3.5A**). An equivalent or better LOD was recorded for all other V600 mutations. As a result, when applied to gDNA samples of high quality (minimal chemically or mechanically (*e.g.* shear) induced degradation), the two-well assay can be used to quantify MFs down to 0.1% to a very high degree of accuracy (**Table 3.3**).

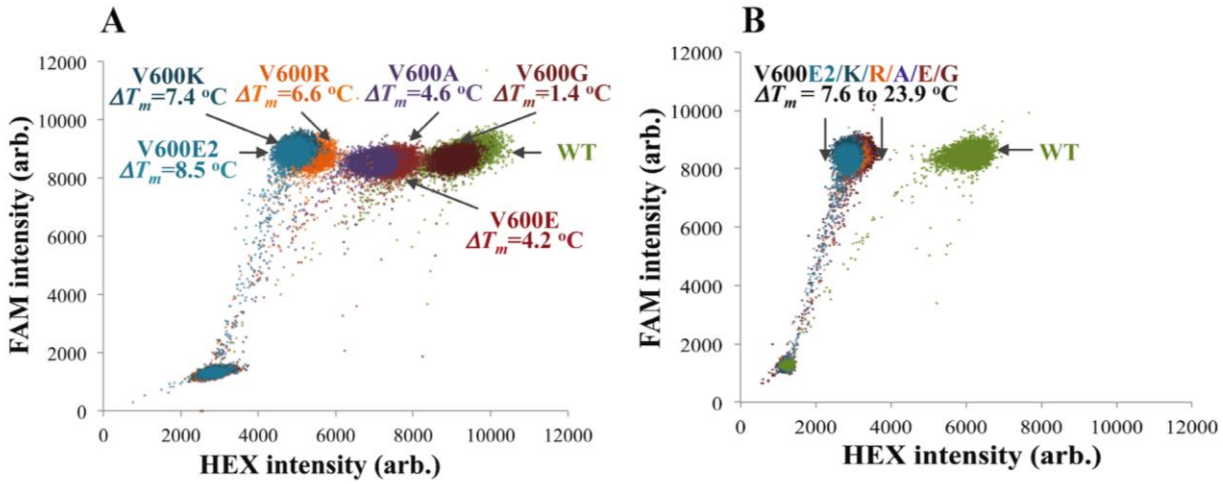

Figure 3.4 Influence of ΔTm value on the performance of the WT-negative component of the ddPCR-based BRAF V600 status assay. Overlay of 2D data output for the *BRAF* V600 WT-negative component of the assay independently applied to samples of WT *BRAF* and each of six clinically relevant MT *BRAF* V600 plasmid DNA samples. (A) pure-DNA *BRAF* V600 WT-specific probe: data clusters for assay applied against MT alleles merge with the corresponding data cluster for the WT allele as ΔT_m decreases to values less than 7 °C, as seen for *BRAF* V600E, V600A and V600G. (B) LNA *BRAF* V600 WT-specific probe: a ΔT_m greater than 7 °C is achieved against every mutant allele, resulting in complete segregation of the data cluster for WT *BRAF* from that for every clinically relevant V600 missense mutation.

Table 3.3 Mean mutant frequency (MF) and standard deviation (n = 8) values measured using the ddPCR-based BRAF V600 status assay. Samples having 0.10% MF were created from MT and WT plasmid DNA. Data for clinically relevant *BRAF* V600 missense mutations are shown.

***BRAF* Allele**	**Expected MF**	**Measure MF**
V600E	0.10 %	0.10 (± 0.05) %
V600K	0.10 %	0.14 (± 0.06) %
V600R	0.10 %	0.11 (± 0.04) %
V600D	0.10 %	0.10 (± 0.08) %
V600M	0.10 %	0.09 (± 0.04) %
V600G	0.10 %	0.11 (± 0.05) %
V600A	0.10 %	0.09 (± 0.05) %
V600L	0.10 %	0.10 (± 0.06) %

MF values were computed by exporting raw ddPCR data (e.g., **Figure 3.4**) into the custom-designed data analysis software described in Chapter 2 (section 2.3.4) and Appendix B. For the *BRAF* status assay described here, that algorithm accurately identifies and isolates unique data clusters (*e.g.*, MT droplet cluster, WT droplet cluster) using a Gaussian mixture model[51] to set cluster borders in both the HEX and FAM dimensions to either side of the mean of the cluster (see **Figure 3.1**). Droplets falling outside those limits [181] are eliminated prior to MF calculation. The data analysis algorithm is fully automated, and results recorded by it, including computed standard deviations and confidence intervals agree quantitatively with manually computed results across the full dynamic range of the assay (0.05 – 100% MF).

As illustrated for *BRAF* V600K (**Figure 3.5B**), the basic metrics of assay performance (LOD, limit of blank (LOB), confidence intervals, dynamic range) remain unchanged when applied to serial dilutions of gDNA isolated from the YUMAC cell line (homozygous for *BRAF* V600K) into that from the SW480 cell line (wild-type *BRAF*).

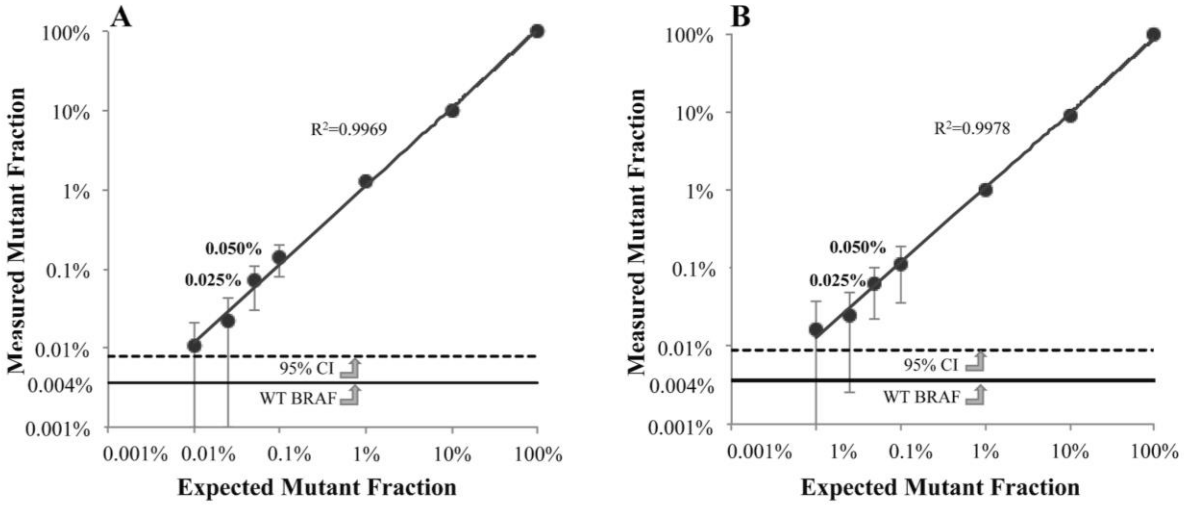

Figure 3.5 Analytical sensitivity (LOD) of the WT-negative component of the ddPCR-based *BRAF* V600 status assay when applied to (A) V600K plasmid and (B) cell-line derived DNA standards.

In both figures, measured mutant frequency (MF) and standard deviation values (n = 24 for 0.01% - 0.1% data, and n = 4 for MF > 0.1% data) are plotted versus expected MF for serial

dilutions down to 0.01% MF. The dotted line is the limit of blank of the WT-negative component of the assay. Significant linear correlation ($R^2 > 0.996$; $p < 0.0001$) between the measured and expected MF is observed down to 0.05% MF. Replicates ($n = 24$) of *BRAF* V600 WT DNA from the specified source (plasmid or cell line) were used to define mean false-positives ("WT *BRAF*" labeled solid horizontal line) and standard deviations, from which the 95% confidence interval was determined and used to define the LOB – 0.008% MF for both systems. Statistically significant MF values can be obtained using the assay to an LOD of 0.05%.

V600 MT frequencies detected in gDNA isolated from FFPE standards (**Figure 3.6**) carry larger uncertainties. The quality (and often the amplifiable quantity) of target template is generally lower in gDNA prepared from FFPE specimens, leading to an array of possible sequence and PCR artifacts.[52] The heterogeneous cell population comprising tissue samples will also display larger genomic sequence diversity. Replicate ($n = 24$) ddPCR *BRAF* V600 status assay runs on gDNA from the 1.4% *BRAF* V600E FFPE reference standard, as well as gDNA from that standard serially diluted into gDNA from the 100% *BRAF* WT V600 FFPE reference (1%, 0.8%, 0.6%, and 0.4% MT frequencies) were therefore conducted to estimate a LOD for the assay when working on gDNA recovered from FFPE samples as an input. For those replicates, statistically significant V600 WT-negative calls could be made for MT frequencies of 0.8 % and higher. Evidence of MT template (*i.e.*, a distinct MT cluster) could be observed at MT frequencies between 0.8% and 0.4% (see **Figure 3.6B**, for example), but the standard deviation in MT template counts was too large to discriminate, for instance, 0.6% and 0.4% MFs. An MF of 0.8% was therefore assigned as the sensitivity of the ddPCR-based *BRAF* V600 status assay when applied to FFPE tissue specimens.

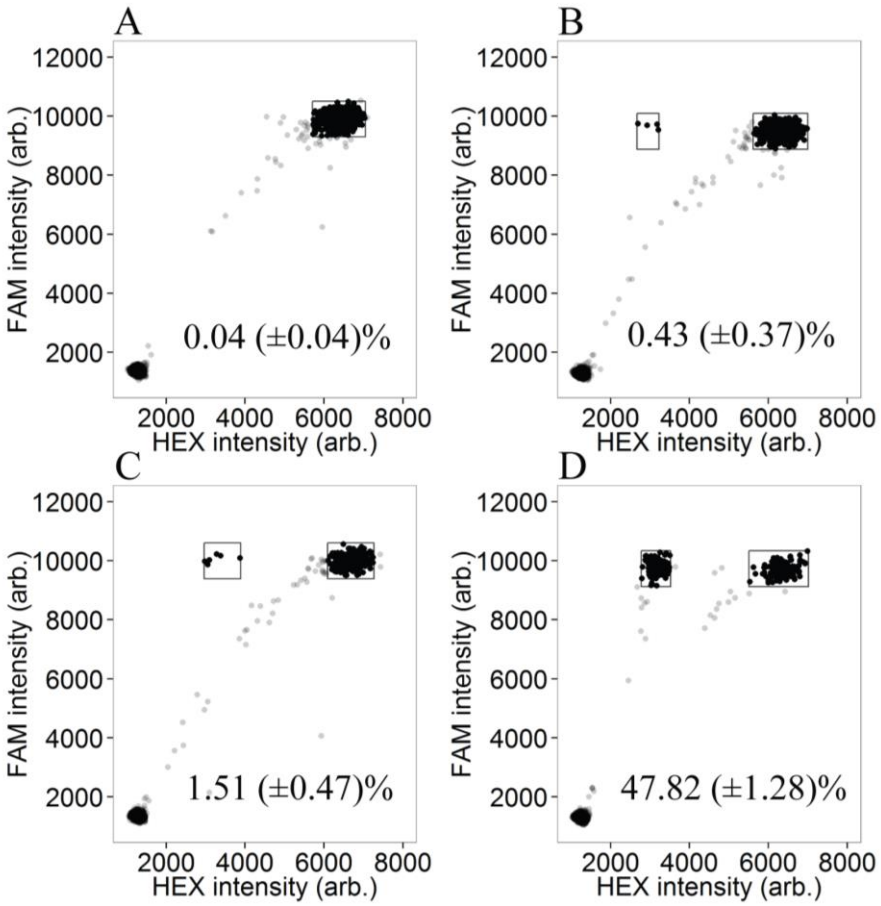

Figure 3.6 BRAF V600 WT-negative screen (well 2) output when applied to HDx Reference FFPE standards. Automated analysis of the data yields mean mutant frequencies of (A) 0.04 (± 0.04) % for a 100 % *BRAF* WT V600 FFPE reference standard (*n* = 24), (B) 0.43 (± 0.37) % for a 0.4 % *BRAF* V600E FFPE reference standard, (C) 1.51 (± 0.47) % for a 1.4 % *BRAF* V600E FFPE reference standard, and (D) 47.82 (± 1.28) % for a 50% *BRAF* V600E FFPE reference standard. Values are expressed as mean ± SEM (*n* = 2). FAM = 6-carboxyfluorescein; HEX = hexachloro-fluorescein; FFPE = formalin-fixed paraffin-embedded; WT = wild-type.

3.3.4 Assay Validation on Clinical Colorectal Cancer Tumor Samples

Colorectal carcinoma cases were selected from a pool of MLH1-deficient tumors identified by IHC testing as part of the population-based Vancouver Coastal Health [172] Lynch-syndrome screening program. Three non-MLH1 deficient cases (specimens 3, 5 and 20) were also included. Genomic DNA specimens purified from the cohort of *n* = 41 colorectal cancer FFPE-stabilized tumor samples were subjected to the ddPCR assay, and complete BRAF V600 status results for each specimen are reported in **Table 3.4** along with corresponding V600E calls made

using the orthogonal gold-standard mutation-specific IHC VE1 assay. The IHC VE1 staining results, examples of which are reported in **Figure 3.7**, are expressed as staining negative (N) or positive (P) for V600E. Results of the ddPCR *BRAF* V600 status assay are expressed as N or P for V600E1 (well 1), and as WT or WT-negative (well 2), with the mutant frequency calculated as the percentage of WT-negative V600 alleles over total *BRAF* alleles.

Table 3.4 Analysis of genomic DNA recovered from patient FFPE tumor specimens by the VE1 IHC (V600E) staining assay and the ddPCR-based BRAF V600 status assay.

Sample	IHC-VE1 Staining Assay (V600E)	dPCR *BRAF* V600 Status Assay		
		Well 1 (V600E)	Well 2	BRAF V600 Mutant Frequency
1	N	N	WT	0.08 (±0.08) %
2	P	P	WT Negative	45.05 (±1.35) %
3	N	N	WT	0.04 (±0.04) %
4	P	P	WT Negative	30.00 (±1.10) %
5	N	N	WT	0.12 (±0.08) %
6	P	P	WT Negative	14.45 (±0.35) %
7	P	P	WT Negative	28.40 (±2.30) %
8	N	N	WT	0.25 (±0.09) %
9	P	P	WT Negative	22.66 (±0.35) %
10	N	N	WT	0.054 (±0.054) %
11	N	N	WT	0.047 (±0.047) %
12	P	P	WT Negative	44.25 (±1.48) %
13	N	N	WT	0.19 (±0.19) %
14	P	P	WT Negative	35.55 (±0.85) %
15	N	N	WT	0.24 (±0.07) %
16	P	P	WT Negative	35.95 (±0.95) %
17	P	P	WT Negative	29.37 (±3.24) %
18	P	P	WT Negative	30.90 (±2.90) %
19	P	P	WT Negative	32.40 (±2.13) %
20	N	N	WT	0.17 (±0.17) %
21	P	P	WT Negative	28.50 (±0.40) %
22	N	N	WT	0.20 (±0.20) %
23	P	P	WT Negative	31.90 (±2.12) %
24	**N**	**N**	**WT Negative**	**57.40 (±1.85) %**

Sample	IHC-VE1 Staining Assay (V600E)	dPCR *BRAF* V600 Status Assay		
		Well 1 (V600E)	Well 2	BRAF V600 Mutant Frequency
25	P	P	WT Negative	25.00 (±1.90) %
26	N	N	WT	0.32 (±0.11) %
27	P	P	WT Negative	34.80 (±0.60) %
28	P	P	WT Negative	30.75 (±0.65) %
29	P	P	WT Negative	40.10 (±0.40) %
30	P	P	WT Negative	31.20 (±0.20) %
31	N	N	WT	0.31 (±0.08) %
32	P	P	WT Negative	36.05 (±0.15) %
33	P	P	WT Negative	22.4 (±0.25) %
34	P	P	WT Negative	31.95 (±1.95) %
35	N	N	WT	0.00 (±0.00) %
36	P	P	WT Negative	12.10 (±0.56) %
37	P	P	WT Negative	18.30 (±3.30) %
38	N	N	WT	0.39 (±0.34) %
39	P	P	WT Negative	48.95 (±12.25) %
40	N	N	WT	0.35 (±0.32) %
41	N	N	WT	0.33 (±0.09) %

As summarized in **Table 3.5**, 17 of 41 CRC samples tested negative for *BRAF* V600E by VE1 staining, with a typical negative V600E staining result shown in **Figure 3.7A**. All 24 MLH1-deficient tumors testing positive for *BRAF* V600E by VE1 staining (see, for example **Figure 3.7B**) tested both positive for *BRAF* V600E1 (well 1) and WT-negative (well 2) in the ddPCR assay, confirming the accuracy of both components of the assay against the clinically actionable V600E mutation. Thus, VE1 IHC staining data and ddPCR *BRAF* V600 status assay results were in concordance for all clinical samples except sample number 24 (bold), which stained negative for V600E in the VE1 IHC assay (**Figure 3.7C**). That sample likewise tested negative for *BRAF* V600E1 in well 1 of our dPCR assay, but tested WT-negative in well 2, with a recorded MF of 57.4 (±1.9) %. Sanger sequencing results for the amplified template (**Figure 3.7D**) confirm that the patient carries the rare *BRAF* V600R mutation, highlighting the comprehensive ability of the ddPCR *BRAF* V600 status assay to both detect clinically relevant V600 mutations and, in the

case of rare V600 mutations, provide quantitative evidence that the MF is sufficient to permit unequivocal MT-sequence verification.

Table 3.5 Summary of BRAF V600 status calls made using the ddPCR-based assay and the gold-standard IHC VE1 assay.

IHC VE1 Assay		**ddPCR-based *BRAF* V600 Status Assay**		
V600E Negative	V600E positive	WT	V600E1	WT negative
17/41	24/41	16/41	24/41	25/41

The remaining n = 16 specimens testing negative for *BRAF* V600E by VE1 staining tested WT-positive in the ddPCR assay based on a mean MF (and error) computed from duplicate runs falling below the sensitivity of the assay when applied to FFPE specimens.

3.4 Discussion

The growing library of biomarkers prognostic of cancer risk and progression includes somatic mutations in (proto-)oncogenes that are thought to drive malignant cell proliferation or prolong their survival. Those markers are enabling individualize genetic analyses of tumors, which is in turn helping to transform oncology toward the effective design and use of therapeutics against molecular targets and associated signaling-pathway aberrations that are specific to a patient's cancer. Relative to conventional (chemo)-therapeutic treatments, targeted molecular therapies generally offer fewer side effects, can often be administered as a patient-friendly oral dosage, and can prove effective against certain tumor types for which standard therapies offer little to no benefit. The development of panels of biomarkers offering increasingly detailed pharmaco-genetic analyses of patient-specific cancers also points to more comprehensive mutational analyses being utilized in cancer genetics testing laboratories to facilitate clinical decisions.[211] In the case of an oncogene that is susceptible to either common or more rare somatic mutations in a single codon or adjacent set of codons, this includes assays able to detect the complete set of relevant mutant alleles in a manner that not only identifies clinically-actionable mutations, but also alerts the clinician of a rare mutation that might necessitate more aggressive clinical monitoring by NGS or a personalized course of treatment.

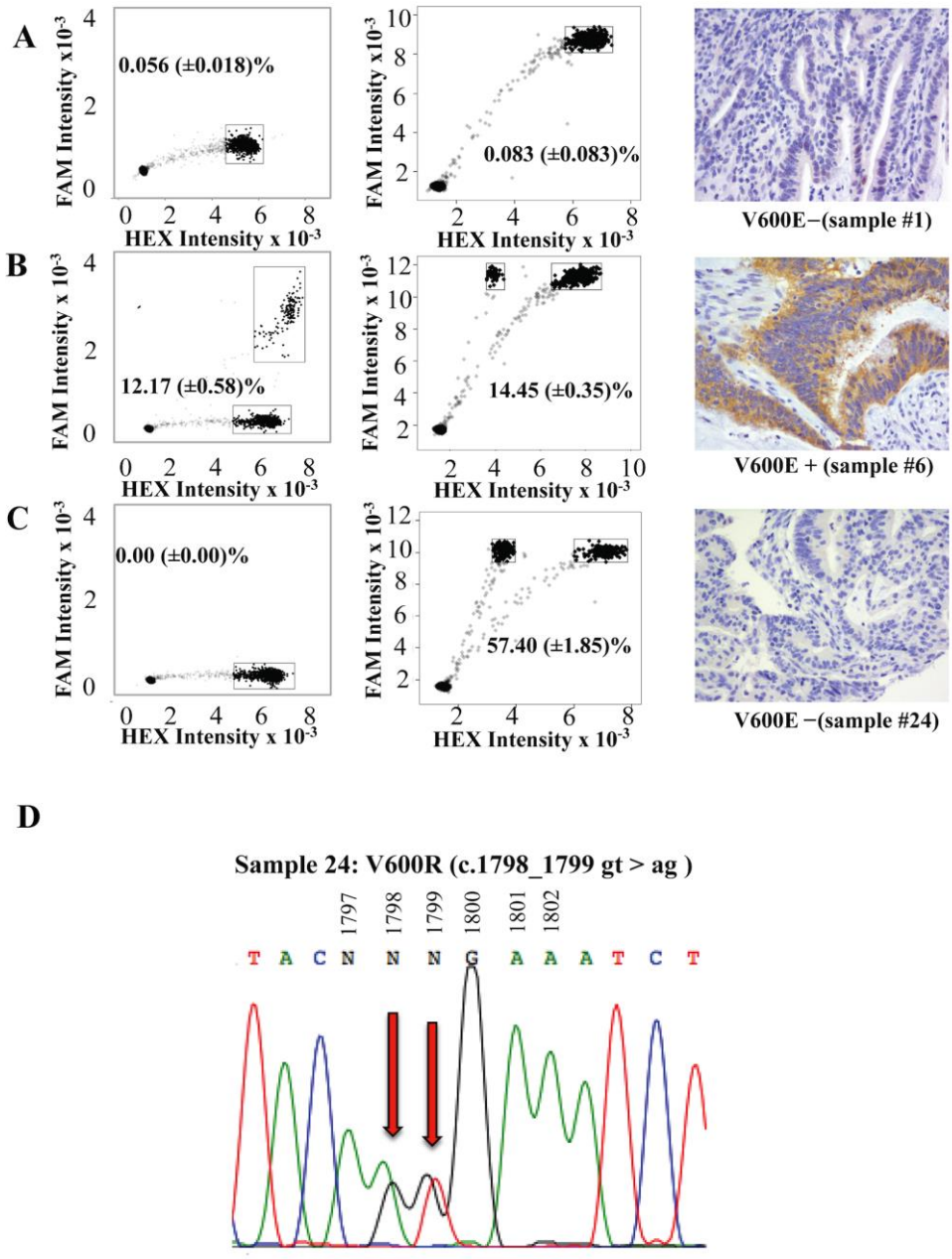

Figure 3.7 Comparison of ddPCR-based BRAF V600 status assay results to those of the VE1 IHC assay for representative MT and WT BRAF tumor samples. (A) Sample #1 (see Table 3.4) – testing V600E negative by the VE1 IHC assay, and WT-positive by the ddPCR-based assay; (B) Sample #6 – testing V600E positive by the VE1 IHC assay, and V600E1-positive/WT-negative by the ddPCR-based assay; (C) Sample #24 – testing V600E negative by the VE1 IHC assay, and V600E1-negative/WT-negative by the ddPCR-based assay; (D) Sanger sequence of gDNA of sample #24 shows a *BRAF* V600R missense mutation (red arrow indicate mutated nucleotides) in concordance with the ddPCR-based *BRAF* V600 status assay results.

For clinical testing of *BRAF* V600 mutations associated with colorectal cancer, we have shown here how such an assay can be realized by using ddPCR to i) unambiguously detect *BRAF*^{V600E1} and thereby identify MLH1-deficient CRC patients unsuitable for anti-EGFR-antibody therapy, and ii) further segregate MLH1-deficient tumors into those bearing WT V600 and therefore eligible for treatment with cetuximab or panitumumab, and those harboring a rare V600 mutation, for which a more comprehensive NGS-based screen of CRC biomarkers may be justified. Differentiation of WT V600 from MT V600 is achieved using a WT-negative screening strategy, and the assay can be successfully applied to as little as 8 ng gDNA at a total cost of goods of *ca.* $8 US per sample. Interestingly, the concept of the WT-negative screen has only been described once before – in the landmark paper of Vogelstein and Kinzler[126] that is best known for providing the first description and demonstration of digital PCR. It has not been demonstrated on clinical samples. To our knowledge, the work reported here therefore provides the first evidence that a WT-negative screen can be designed for and successfully applied to mutational analyses of tumor specimens; it therefore adds to a number of powerful new applications of digital PCR to screening of rare alleles.[145, 212] Through use of a model-designed LNA/DNA chimeric probe against WT V600, we achieve a sensitivity of 0.8% MF when the assay is applied to gDNA isolated from FFPE specimens. That LOD is dictated by the poorer quality and generally lower quantity of amplifiable *BRAF* template (often 5000 – 8000 copies) that could (and typically can) be extracted from the colorectal FFPE specimens, but nevertheless exceeds that projected for the coming generation of clinical NGS instrumentation (≥ 1% mutant frequency) and is at least 3.5-fold better than provided by the best NGS technology currently available to clinics. In particular, specific calls on all possible mutations in the *BRAF* V600 codon and proximal codons has been achieved by deep pyrosequencing down to MF values ≥ 3%.[213] The sensitivity and the mutational coverage of our assay also exceed what can be achieved by conventional IHC testing.

We note that comparable or better sensitivities can be realized using allele-specific PCR assays against a particular MT *BRAF* allele. For example, for plasmid DNA samples, sensitivities as low as 0.001% have been reported in assays against *BRAF* V600E utilizing allele-specific qPCR,[214] wild-type blocking PCR,[215] E-ice-COLD-PCR,[216] or ddPCR.[149] The capabilities of those assays, however, differ considerably from that reported here, as they are not designed to or capable of detecting all *BRAF* V600 mutations. Moreover, they generally do not operate on (or

at least have not been applied to) gDNA recovered from FFPE specimens, for which (as demonstrated in this work) the assay LOD is largely determined by the nature and quantity of amplifiable template available per test.

Finally, some discussion is warranted as to how the ddPCR-based assay described may be extended for application to testing of *BRAF* V600 status in advanced melanomas. The *BRAF* V600E mutation has been accepted as a biomarker predictive of melanoma patient response to the mutant BRAF inhibitors vemurafenib and dabrafenib, while a combination of dabrafenib and trametinib has been generally accepted for treatment of V600K-positive metastatic and unresectable melanomas. Multiplexing of the ddPCR assay could permit specific detection of V600E and V600K mutations in well 1. This would require model-based design of a probe against V600K that exhibits no cross-reactivity to WT V600 or any other relevant V600 MT (analogous to the V600E1-specifc probe described in this work). FAM labeling of the V600E-specific probe and, say, Alexa Fluor 488 labeling of the corresponding V600K-specific probe may then be used to unambiguously detect and quantify either mutation in the resulting 2D ddPCR data plot for well 1. Through its ability to detect rare *BRAF* V600 mutations, the WT-negative assay conducted in well 2 could then serve to inform the clinician of the need for enhanced patient monitoring or a more aggressive course of treatment. In this regard, the ability to detect all clinically relevant V600 mutations at frequencies lower than currently achieved by Sanger sequencing or NGS could ameliorate the dangers of a negative sequencing result arising not from absence or remission of disease, but rather from a MF below the detection limit of the instrument (~ 5%). This concern is heightened by growing evidence that minor sub-clones positive for a V600 coding mutation have altered immune responses. Moreover, though they typically provide significant clinical response over several months, BRAF inhibitors can alter immune inflammatory mechanisms associated with those aberrant cells. Knowledge of the V600 mutation and the associated immunogenicity of the tumor-associated cells bearing that mutation may therefore provide a sound basis for designing combinations of BRAF inhibitors and immuno-therapeutics that improve progression-free survival by strengthening immune responses or counteracting immune escape mechanisms.[217]

Chapter 4: Quantitative analysis of *KRAS* G12/G13 status in colorectal cancer using a novel wild-type negative assay that unequivocally differentiates missense and synonymous alleles

4.1 Introduction

Nearly 1.4 million new cases of colorectal carcinoma (CRC) are diagnosed each year worldwide, making it the third most common cancer and fourth most common cause of cancer-related deaths.[218] Current established targeted therapies for metastatic CRC (mCRC) include the anti-epidermal growth factor receptor (EGFR) antibodies (mAbs) panitumumab (Vectibix; Amgen Inc.) and cetuximab (Erbitux; Bristol-Myers Squibb Inc.). Nonsynonymous somatic point (missense) mutations in the Kirsten Ras (*KRAS*) oncogene are theranostic of mCRC patient response to treatment with anti-EGFR mAbs, with patients carrying any one of a set of known missense mutations in *KRAS* being nonresponsive or poorly responsive. *KRAS* is located on the short arm of chromosome 12 and encodes the GTPase KRAS that signals downstream of EGFR in the mitogen-activated protein kinase (MAPK; RAF/MAPK) and PI3K signaling pathways regulating cell expansion, differentiation, and apoptosis. Missense mutations in *KRAS* are observed in 30-50% of CRC tumors,[80] with codons 12 and 13 being two hot spots that account for approximately 95% of all observed *KRAS* mutations.[219, 220] Both of these codons code for glycine in wild-type[68] human *KRAS*. Mutation of either one or of both of the first two bases in either codon is often observed in mCRC and leads to an amino acid substitution in *KRAS* that results in activation of the MAPK pathway without need for ligand binding to EGFR. Resistance to anti-EGFR mAb treatment is then observed.[3, 202, 221, 222]

KRAS mutational testing of mCRC patients is now mandatory in the US, Europe and Japan,[97, 98] with panitumumab or cetuximab therapy approved in the US for *KRAS* G12/G13 mutation-negative mCRC[97, 99] and in Europe if *KRAS* within the tumor is WT.[100] In either jurisdiction, evaluation of *KRAS* status across codons 12 and 13 is therefore required. Sequencing can be and is (albeit rarely) used clinically for this purpose, but is known to suffer from a lack of sensitivity. This creates potential for misinterpretation of results leading to lack of response to anti-EGRF therapy in patients with a *KRAS* mutation frequency (MF) below ~ 10%.[223]

Clinical testing of *KRAS* mutational status is therefore most often conducted using a real-time PCR (qPCR) based assay, and two FDA-approved test kits are available for this purpose: the TheraScreen® *KRAS* mutation kit (DxS-Qiagen) and the cobas® *KRAS* mutation test (Roche Molecular Systems). The TheraScreen assay detects the seven most frequent somatic point mutations (SPMs) observed in codons 12 and 13 of *KRAS*. The analytical sensitivity (LOD) to each of these mutant (MT) alleles is approximately 5%, depending on the DNA quality. However, other missense mutations in codons G12/G13 promoting constitutive activation of *KRAS* are known, and the TheraScreen assay does not identify the ~2% of mCRC patients carrying one of those mutations. The cobas *KRAS* assay is more comprehensive in that it can detect those less common mutations. However, the assay is likewise limited by an LOD of ~5%, and results provided by it are qualitative (i.e. the test does not quantify MF) above that detection limit.[223] Other methods have been reported,[16, 17, 29, 224, 225] but none to our knowledge have received regulatory approval or significant clinical use.

More sensitive and quantitative *KRAS* detection methods capable of detecting and quantifying all clinically actionable missense mutations within codons G12/G13 would therefore be of considerable clinical value, as anti-EGFR therapy could be avoided in CRC patients harboring a *KRAS* mutation at a frequency below 5%. As shown in this thesis, droplet digital PCR (ddPCR) can be used to inexpensively and quantitatively profile somatic variations and their frequencies in oncogenes present within formalin-fixed paraffin-embedded (FFPE) tissue specimens, as well as in fresh cellular tissues isolated from blood, or circulating tumor DNA .[133, 149, 226-232] The novel ddPCR-based *BRAF* V600 status assay (Chapter 3) detects all known missense mutations within the V600 codon of *BRAF* to enable clinics to reliably and rapidly determine *BRAF* gene status in metastatic melanoma and mCRC patients.[233] When applied to a cohort of mCRC patients positive for microsatellite instability, that assay reported 100% clinical accuracy to an LOD well below 1% MF, offering the ability to detect and quantify the frequency of the most prevalent V600 mutations (V600E and V600K) as well as all known rare *BRAF* V600 mutations, such as V600R.

Those advances are leveraged here to create a multiplexed ddPCR assay to discriminate patients carrying WT *KRAS* from those carrying a missense mutation within codons G12/G13 or within codons 14 to 17 of *KRAS*. The nearest-neighbor type (NNT) molecular thermodynamic model (Chapter 2 and Appendix A) is used to design a highly specific locked nucleic acid (LNA) substituted dual-labeled hydrolysis probe against each known WT *KRAS* allele over the hot-spot (proto-)oncogenic region comprising codons 12 to 14. In the *BRAF* WT-negative assay (Chapter 3), a single LNA-substituted WT-specific probe designed using these models was shown to unequivocally discriminate between WT *BRAF* V600 and all known nonsynonymous *BRAF* V600 alleles.[233] Due to the specificity of the probe used, ddPCR assays of this type were therefore collectively named "wild-type negative assays".

Compared to *BRAF*, creation of a WT-negative assay for *KRAS* is complicated significantly by the fact that seven synonymous *KRAS* alleles are observed across codons 12 and 13, while *BRAF* is characterized by a single WT sequence within and adjacent to its V600 codon. Creation of a WT-negative test of *KRAS* status therefore requires significant multiplexing to enable unequivocal discrimination of all known *KRAS* WT alleles from all known G12/G13 missense mutations. Such capability has not yet been demonstrated in a ddPCR assay, and collectively requires the set of dual-labelled WT-*KRAS*-specific hydrolysis probes one utilizes to show no cross-reactivity to any known missense mutation within codons 12 and 13. To achieve a duplex stability that enables its efficient hydrolysis, each such WT-specific probe must extend beyond codons G12/G13 and into *KRAS* codon 14. The assay therefore must include a means to determine if a lack of WT-specific probe binding is due to a known (but rare) SPM within codon V14 (three V14 mutations have been reported in the COSMIC (*Catalogue of Somatic Mutations in Cancer*) database [50] at a total MF of 0.07%). An LNA-substituted probe targeting codons 14 – 17 is employed in the assay for this purpose. Finally, a dual-labeled "consensus" probe targeting a highly conserved 23-nt sequence within *KRAS* is used to detect the total number of amplifiable *KRAS* templates within the genomic DNA (gDNA) sample, permitting the MF to be quantified for plasmid and cell-line DNA samples with high accuracy to an LOD of 0.025%, while also defining the total amount of amplifiable DNA present in the sample. The method therefore offers advantages over current qPCR-based tests of *KRAS* status through a combination of its improved sensitivity, its ability to not only

detect the presence of all clinically actionable *KRAS* G12/G13 missense mutations but to also accurately quantify the MF, and, particularly when applied to gDNA recovered from FFPE samples, its ability to quantify the total copies of amplifiable *KRAS* present in the sample.

The performance of this novel ddPCR-based *KRAS* G12/G13 status assay is first demonstrated through application to a set of plasmid-DNA standards, each presenting a germline or specific G12/G13 mutant allele, and to gDNA recovered from mutant *KRAS* cell-line and FFPE standards. The assay is then applied in combination with the previously validated *BRAF* V600 status assay[233] to clinical FFPE tumor specimens from 87 mCRC patients positive for high microsatellite instability (MSI-H). CRCs often develop as a "classical" adenoma to carcinoma progression in which *KRAS* mutation may be observed. However, approximately 20 – 30% of CRCs arise via an alternative pathway, most often one of the serrated pathways characterized by progression of micro-vesicular hyperplastic polyps through sessile serrated adenoma (SSA) to full SSA dysplasia, and then ultimately to carcinoma. CRCs arising through the serrated pathway are often characterized by MSI-H resulting from epigenetic inactivation of DNA repair genes, most notably the Mut L homologue-1 (*MLH1*) gene. CpG island methylation phenotype (CIMP) and *BRAF* V600 mutation are also but not always observed in serrated-pathway associated CRCs and the status of each may be used to further sub-classify the cancer.[234, 235] Co-mutation of both *BRAF* V600 and *KRAS* G12/G13 is a rare event, observed before in only 3 patients[236], whose pathology has not been well described. However, through application of the *KRAS* status assay described here, two additional MLH1-deficient mCRC patients carry tumors harboring a missense mutation in both *BRAF* V600 and *KRAS* G12/G13 are identified. In one of those patients, the frequency of the *KRAS* MT is low (~ 2%), below the sensitivity offered by either the TheraScreen or cobas assay, providing evidence of the utility of the new assay as well as pathological insights into the potential origin of this rare event.

4.2 Materials and Methods

4.2.1 Oligonucleotides

All primers, pure-DNA and LNA-substituted dual-labeled hydrolysis probes, and WT and MT *BRAF* alleles were purchased from IDT, Inc. (Coralville, IA). Probes were HPLC purified, while primers and templates were purified by desalting. Purified primers, probes and templates were resuspended to 100 µM in IDTE (10 mM Tris, pH 8.0, 0.1 mM EDTA) buffer and stored at -20 °C prior to use.

4.2.2 Primers and Probes Design

Forward (FP) and reverse (RP) primers (**Table 4.1**) were designed using Primer3 software. All primers were designed by first applying primer-BLAST software to identify and avoid sequences that might show cross-reactivity to non-target sequences within the human genome. Each dual-labeled hydrolysis probe (**Table 4.1**) was likewise engineered to minimize PCR artifacts and to ensure selective hybridization to a specific sequence within *KRAS* at ddPCR reaction conditions. The consensus probe, a 23 nucleotide (nt) HEX and Iowa black fluorescence quencher (IABkFQ) labeled pure-DNA probe spanning elements of *KRAS* intron/exon 2 (includes exon 2 codons 1 to 3), was designed using the NNT model of Hughesman *et al.* [161] to achieve a melting temperature (T_m) of *ca.* 68 °C when duplexed to WT *KRAS*. That model for pure-DNA probes accounts for the effects of PCR solution conditions and the addition of a fluorescent reporter dye (HEX or FAM) and quencher (IABkFQ, BHQ, or Dab) on probe-template melting thermodynamics. Potential probe-derived PCR artifacts were identified and avoided using primer-Blast software. Each LNA-substituted *KRAS* G12/G13 WT-specific probe (denoted WTP1 to WTP7), as well as the LNA-substituted probe against WT *KRAS* codons 14 to 17, were designed using the NNT model described in this thesis (Chapter 2; Appendix A) to define the combination and pattern of LNA substitutions needed to limit probe length to the target codons while achieving a suitable T_m (~ 67 to 72 °C) for use in the assay. For each known germline allele, a set of putative probe sequences was thereby designed, and the LNA-DNA mismatch capabilities of the model then used to identify among them that probe sequence that minimizes cross-reactivity to non-target alleles. Model predicted

T_m values for probes were verified experimentally by UV melt spectroscopy according to standard methods.[237]

Table 4.1 Sequences and concentrations of primers and dual-labeled hydrolysis probes used in the ddPCR-based KRAS G12/G13 screening assay.

Primer/Probe	Sequence	Conc. (nM)
Forward Primer	5'-ATTTGATAGTGTATTAACCTTATGTGTGAC-3'	900
Reverse Primer	5'-ACCTCTATTGTTGGATCATATTCG-3'	900
WTP1 (39G, 36A)	5'-(6-FAM)/A**C**G**C**C**A**C**C**/(Dab)-3'	250
WTP2 (39G, 36T)	5'-(6-FAM)/A**C****G****C**C**T**C**C**/(Dab)-3'	250
WTP3 (39A, 36A)	5'-(6-FAM)/A**C****A****C**C**A**C**C**/(Dab)-3'	250
WTP4 (39T, 36A)	5'-(6-FAM)/A**C****T****C**C**A**C**C**/(Dab)-3'	250
WTP5 (39C, 36A)	5'-(6-FAM)/A**C**C**C**C**A**C**C**/(Dab)-3'	150
WTP6 (39G, 36C)	5'-(6-FAM)/A**C**G**C**C**C**C**C**/(Dab)-3'	100
WTP7 (39G, 36G)	5'-(6-FAM)/A**C**G**C**C**G**C**C**/(Dab)-3'	100
WTP14-17	5'-(HEX)/CT**C**T**T**G**C**C**T**A**C**G/(BHQ_1)-3'	150
Consensus Probe	5'-(HEX)/AAGGCCTGC/ZEN/TGAAAATGACTGAA/IABkFQ)-3'	150

Locked nucleic acids are in **bold**. Bases that vary between WG-specific probes are underlined. Nomenclature: WTP = wild-type specific probe; 6-FAM = 6-carboxy-fluorescein; Dab = dabcyl; HEX = hexachloro-fluorescein; BHQ_1 = Black Hole Quencher® 1; ZEN = ZEN internal quencher; IABkFQ = Iowa Black fluorescence quencher.

4.2.3 Tumor and Reference Samples, and DNA Extraction Protocols

FFPE tissue specimens from a cohort of mCRC patients were obtained from Lion's Gate Hospital (North Vancouver, Canada) with approval from the UBC Clinical Research Ethics Board (CREB; certificate number H14-00577) and patient consent following the Helsinki protocol. The CRC MSI+ cases were selected from a pool of MLH1-deficient tumors, identified by immune-histochemical (IHC) testing as part of the population-based Vancouver Coastal Health [172] Lynch-syndrome screening program. Three non-MLH1-deficient cases (samples 3, 5, and 20) were also included. Specimens were derived from 10% neutral buffered formalin-fixed resections in all but one case (#37), which was a formalin-fixed malignant ascites fluid specimen. The original diagnosis was confirmed by a pathologist (Dr. Robert Wolber, Lion's Gate Hospital, North Vancouver, BC, Canada), and an optimal tumor block containing invasive carcinoma was selected for sampling. Two cores of each carcinoma were

taken for construction of tumor microarray (TMA) blocks, and two additional cores were taken for the ddPCR studies reported.

Eight FFPE standards, each harboring either a WT *KRAS* allele or a ~5% MF of one of the prevalent G12/G13 mutants, were purchased from Horizon Discovery Ltd. (Cambridge, UK). gDNA specimens from the FFPE reference standards or from mCRC FFPE cores were purified using the QIAamp DNA FFPE tissue kit (Qiagen, Inc.; Santa Clarita, CA) under a modified protocol for the xylene-assisted paraffin removal step that serves to reduce DNA shearing. In that modified method, xylene (0.8 mL) was added to FFPE cores (two) collected in a 1.5 mL micro-centrifuge tube and the mixture equilibrated by rotational mixing for 10 min. The sample was then centrifuged at 14,000 rpm for 1 min to collect tissue. Waste xylene was removed and the process repeated twice more. Three successive washes with 0.8 mL of 100%, 100% and 70% ethanol, respectively, were then completed using the same procedure. Finally, the sample was placed in a Vacufuge™ (eppendorf, Hamburg, Germany) for 5 min under medium heat to dry, and the residue carefully removed for further processing according to the standard QIAamp DNA FFPE protocol.

Purified gDNA specimens from cell lines HT-29 (homozygous for the germline *KRAS* allele WT1; labeled WT1 to reflect that it is the most prevalent WT allele (the sequence of WT1 from codon 11 through 17 is GGTGGCGTAGGCAAGAGT), SW480 (homozygous for *KRAS* G12V), SW460 (homozygous for *KRAS* G12A*)*, SW116 (heterozygous for *KRAS* G12A), MIA (homozygous for *KRAS* G12C), PL45 (heterozygous for *KRAS* G12D), A549 (homozygous for *KRAS* G12S), H1355 (heterozygous for *KRAS* G13C), and LOVO (heterozygous for *KRAS* G13D) were kindly provided by the BC Cancer Agency. pIDTSMART-AMP plasmids, each containing a 270 bp gene fragment (intron/exon 2) of either WT1 *KRAS* or a *KRAS* G12/13 MT allele, were purchased from IDT.

The concentration of DNA within purified samples was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific; Waltham, MA). As plasmids exhibit supercoiling, pIDTSMART plasmid DNA (1.5 µg) was linearized with ClaI restriction enzyme as per the

manufacturer's instructions (Life Technologies; Carlsbad, CA) prior to concentration determination and use in the assays.

4.2.4 ddPCR Assay Workflow and Characterization

Digital PCR mastermix solutions (20 µL) were prepared from 2X dUTP-free ddPCR supermix for probes (BioRad, Inc.; Hercules, CA), 900 nM each of FP and RP, 250 nM of the consensus probe, 100 nM to 250 nM (**Table 4.1**) of each of the LNA-substituted *KRAS* G12/G13 WT-specific probes, and typically an amount of purified plasmid or gDNA containing ~10000 copies of *KRAS*; gDNA purified from clinical colorectal FFPE samples sometimes contained roughly 2-fold fewer (but still ≥ 5000) copies of amplifiable *KRAS*. No template control (NTC) samples were prepared in IDTE buffer. Emulsified sub-nL reaction droplets were created by introducing a 20 µl sample into a well of an 8-channel disposable DG8 droplet generator cartridge and adding 60 µl of droplet generation oil (BioRad). 40 µl of the resulting droplet emulsion (~13000 to 15000 droplets) generated on the BioRad QX-100 Droplet Generator were transferred by multichannel p100 pipette to an Eppendorf Twin.tec semi-skirted 96-well PCR plate, which was then heat-sealed with foil sheets. The droplet emulsions were thermally cycled in the CFX96™ thermocycler using the following protocol: denaturation at 95 °C for 10 min, followed by 50 cycles of 94 °C for 30 s and 60 °C for 1 min (ramp rate set to 2.5 °C/s). The end-point fluorescence of each thermally cycled droplet was measured using the QX100 Droplet Reader (BioRad). After cycling, a final 10 min hold at 98 °C was applied to deactivate the enzyme and stabilize the droplets. The raw ddPCR data was collected and visualized using the QuantaSoft v1.3.2 program (BioRad Inc.). Droplet assignments into MT and WT data clusters and subsequent counts, from which MF values were computed, were then made by exporting the data from QuantaSoft into the custom software (and associated graphical user interface) described in Chapter 2 and Appendix B [178].

Serial dilutions (1000 – 1 copies/µl) of WT1 *KRAS* isolated from plasmid DNA, as well as each of 6 other known WT *KRAS* alleles and each of 13 clinically relevant MT G12/G13 *KRAS* alleles, were prepared in IDTE buffer [158], with the concentration of *KRAS* template (copies/µl) in each dilution quantified on a BioRad QX100 ddPCR instrument. For dynamic-range measurements, serial samples containing 1000 – 1 copies/µl of *KRAS* G12D or G12S plasmid

DNA were prepared in a high fixed background (10,000 copies/µl) of WT1 *KRAS* plasmid DNA. Samples of *KRAS* G12V gDNA from SW480 cells, as well as *KRAS* G12V (SW480 cells) combined with WT *KRAS* (HT-29 cells), were prepared in the same manner.

4.2.5 ddPCR Raw Data Analysis Algorithm

Droplet event data exported from QuantaSoft V1.3.2 into the custom software[178] were automatically analyzed by the algorithm to assign each read droplet as either empty, "rain" (droplets having a signal lying within an indeterminate region between a pair of distinct positive and negative droplet clusters), MT-positive, or WT-positive. The droplets assigned to each category were then counted and used to detect MT alleles and compute the MF in each well. The software is built in the R programming language and is available on GitHub at https://github.com/daattali/ddpcr and a detailed description of its features and computational elements is provided in Appendix B. An illustration of the data analysis process as applied to data from the ddPCR-based *KRAS* status assay, and a brief overview of the core computational steps of the algorithm, are provided in **Figure 4.1**.

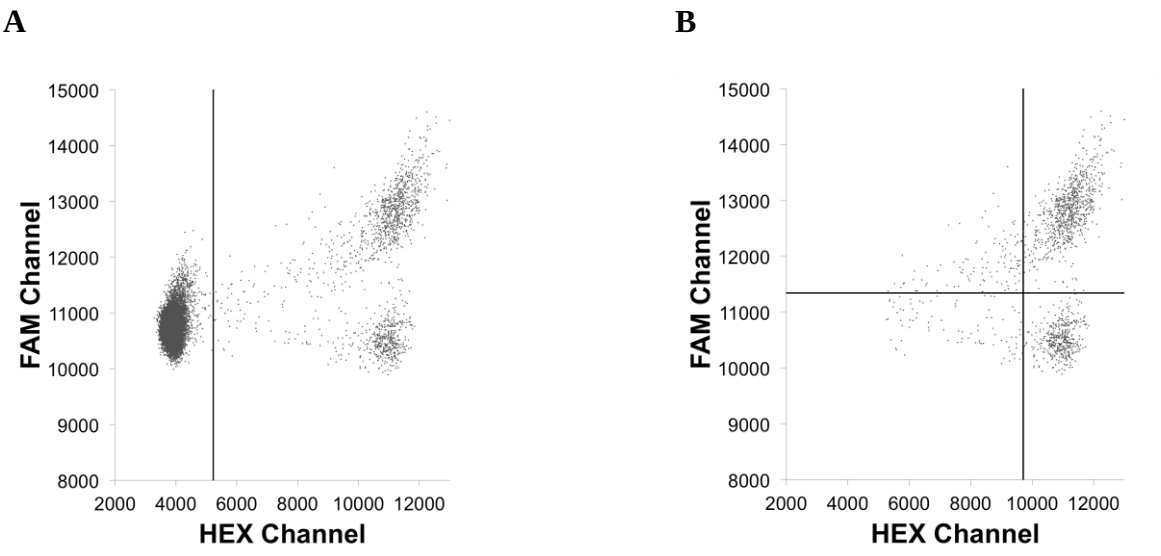

Figure 4.1 KRAS G12/G13 status assay automated data analysis overview. Representative output data for the KRAS WT-negative assay is visualized in a two-dimensional scatter plot. A) Empty droplets are identified by fitting a two-component Gaussian mixture model to the HEX channel, generating one distribution of low mean intensity (representing the cluster left of the vertical line) and another of high mean intensity. All droplets in the low mean density distribution are deemed empty and removed from analysis. B) All filled droplets are then identified as those lying within ± 3 standard deviations (SDs) from the mean of the high mean density distribution (grey band). Droplets displaying HEX intensities outside this range (often referred to as digital PCR rain) are taken as being poor in signal quality and are excluded from the analysis. KRAS G12/G13 WT and MT droplets, respectively, are then identified by fitting two normal distributions to the FAM channel signal, with the relevant populations bounded by ±3 SDs in each case. The distribution having the lower mean FAM intensity contains the population of MT-positive droplets, from which the mutant frequency is calculated as the ratio of the MT droplets to total filled droplets (in this sample: 439/(439+837)x100 = 34% MF). Nomenclature: FAM, 6-carboxyfluorescein; HEX, hexachloro-fluorescein; WT, wild-type; MT, mutant.

4.3 Results

4.3.1 Assay Design and Readout

The ddPCR-based *KRAS* G12/G13 status assay, a schema for which is provided in **Figure 4.2**, is a single-well assay utilizing ddPCR for amplification and detection of missense mutations within codons 12 and 13. In each well, gDNA purified from a FFPE tissue specimen is loaded to a copies-per-drop (CPD) of 0.2 (±0.05), ensuring at the start of the ddPCR that most droplets contain either 0 or 1 copy of a *KRAS* allele and that ~10,000 amplifiable copies of *KRAS* are analyzed per test. Each well contains ddPCR supermix for probes without dUTP (BioRad), as well as the required set of primers. A total of 9 dual-labeled hydrolysis probes, whose sequences and labeling chemistries are reported in Table 4.1 are also included in each well. The resulting *KRAS* amplicons (both WT and MT) are 195 bp in length and span a portion of intron 2 and all of *KRAS* exon 2, including the oncogenic region of interest (codons 12 to 14). During amplification, fluorescence created from hydrolysis of the HEX-labeled *KRAS* consensus probe confirms the presence of a *KRAS* allele (WT or MT) within a droplet, allowing the total amplifiable copies of *KRAS* in the specimen to be quantified by HEX-positive droplet counts and Poisson statistics. Fluorescence generated from any one of the seven FAM-labeled LNA-substituted probes that collectively target all known germline *KRAS* G12/G13 alleles indicates that the copy of *KRAS* present in a given read droplet is WT across codons 12 to 14. Finally, each droplet contains a HEX-labeled probe targeting germline *KRAS* across codons 14 – 17. As noted above, the COSMIC database identifies a dominant WT sequence across codons 14 – 17, permitting this probe to confirm not only the lack of a mutations in codon 14, but to also discriminate between codon 14 mutations and mutations present within codons 15-17. Thus droplets within the FAM-positive/double-HEX positive[1] cluster quantify WT *KRAS* templates, whereas droplets within the FAM-negative/double HEX-

[1] For convenience, we have named the signal arising from amplification of germline *KRAS* as "double"-HEX positive to reflect the fact that those templates are fully complementary to both the HEX-labeled consensus probe and the HEX-labeled probe against WT *KRAS* across codons 14 – 17. However, for a WT *KRAS* allele, the latter probe must compete with a FAM-labeled probe against codons 12 – 14. As a result, the end-point HEX signal recorded is reduced somewhat.

positive cluster quantify missense mutations within codon 12 or 13. In the rare case of a droplet containing a *KRAS* allele bearing a somatic mutation in codon 14, an end-point signal will be generated from the HEX-labeled consensus probe only, and these droplets therefore form a unique FAM-negative/single-HEX positive cluster, while droplets containing a *KRAS* allele bearing a mutation in codons 15 – 17 form a FAM-positive/single-HEX-positive cluster resulting from an end-point signal from one of the FAM-labeled WT specific probes and from the HEX-labeled consensus probe. Finally, droplets recording neither a HEX nor a FAM end-point signal contain no copies of *KRAS*. This novel ddPCR-based WT-negative screening assay thereby permits unequivocal differentiation of tumors bearing germline *KRAS* across codons 12 – 14 from those carrying a missense mutation within codons G12/G13. The assay also quantifies MF and the total abundance of amplifiable *KRAS* through signal from hydrolysis of the HEX-labeled *KRAS* consensus probe.

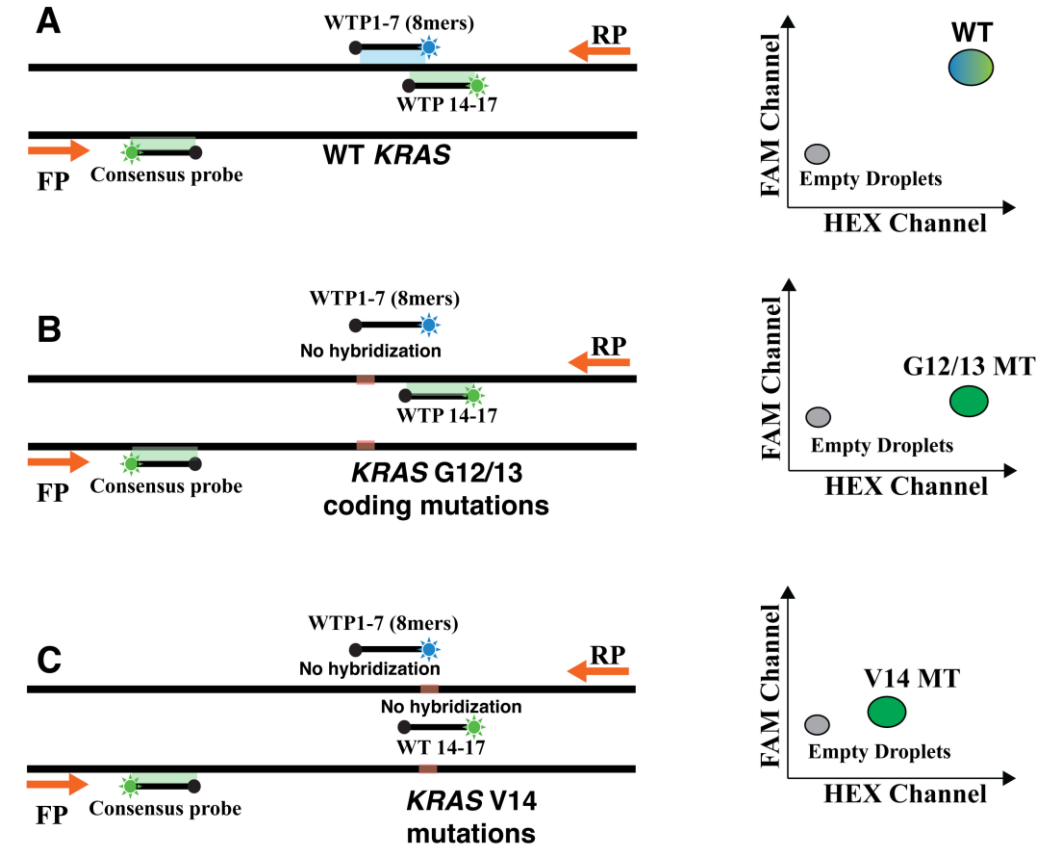

Figure 4.2 Schematic of key features, components and expected output of the WT-negative KRAS screening assay. This single well ddPCR assay uses forward and reverse primers that amplify a 195-bp fragment of the *KRAS* gene that spans exon 2 and codons G12/G13. The assay utilizes nine dual-labeled hydrolysis probes. Seven of those probes are FAM-labeled WT-specific LNA-substituted probes (spanning codons 12 – 14). Each is designed to selectively hybridize to a synonymous *KRAS* G12/G13 allele and thereby collectively distinguish WT *KRAS* from all *KRAS* G12/G13 missense mutations. The other two probes, which are HEX labeled, include an LNA-substituted probe against WT *KRAS* within codons 14 – 17 and a "consensus" probe that binds a highly conserved sequence within the *KRAS* amplification template. A) When a WT *KRAS* G12/G13 allele is amplified, end-point fluorescence signals from the FAM-labeled probe against that WT-allele (blue), the HEX-labeled WTP14-17 probe (green), and HEX-labeled consensus probe (green) are all detected to create a distinct WT-allele cluster of droplets, along with a population of empty droplets, in the 2D output [28]. B) Amplification of any *KRAS* G12/13 MT allele results in the generation of an end-point signal only from the consensus probe (green) and the WTP14-17 probe (green). C) In the rare case where there is a mutation in codon 14, an end-point signal is recorded from the consensus probe only, while a mutation in codons 15 – 17 results in no end-point signals from the WT-specific probe (blue) and the consensus probe (green). In this manner, the assay uniquely and unequivocally detects all clinically actionable *KRAS* G12/G13 missense mutations by differentiating them from WT *KRAS* alleles as well as less common mutations within *KRAS* codons 14 – 17.

4.3.2 Engineering Allele Specificity into Probes Used in ddPCR Assays

In addition to the unique capabilities offered by ddPCR, successful execution of the *KRAS* G12/G13 status assay described in **Figure 4.2** requires the seven probes against the *KRAS* G12/G13 region to be sufficiently short and specific to their target WT allele that cross-reactivity at PCR conditions does not diminish assay performance. Data presented in Chapter 3 shows that designing a WT-specific probe to unequivocally discriminate a WT allele from an ensemble of possible MT alleles that may differ from the target WT sequence by as little as a single base generally requires creating a difference ($\Delta T_{m,(WT-MT_i)}$) of ≥ 7 °C between $T_{m,WT}$, the melting temperature (T_m) of the perfectly matched duplex formed by the probe and its target WT allele, and T_{m,MT_i}, the T_m for the mismatched duplex formed by the probe and any mutant allele i.[233] In addition, $T_{m,WT}$ must exceed T_a, the annealing temperature of the PCR assay, such that the duplex formed between the probe and its WT template is sufficiently stable that the creation of each WT amplicon results in hydrolysis of a bound probe and release of the reporter dye. T_{m,MT_i} must then be low enough that concurrent amplification of any MT allele i at T_a results in a negligible signal from the WT-specific probe over the course of the ddPCR run.

Locked nucleic acid (LNA) substituted WT-specific probes were designed for this purpose using the NNT model[161] described in Appendix B that permits *in silico* selection of the number and pattern of LNA substitutions needed to achieve a suitable $\Delta T_{m,(WT-MT_i)}$. Melting thermodynamics (including T_m data) collected by the UV-melt spectroscopy method[237] confirmed that each WT-specific probe so designed had a $\Delta T_{m,(WT-MT_i)}$ > 7 °C for all known G12/G13 mutations (**Table 4.2**).

Table 4.2 Melting temperature Tm data collected by UV-melt spectroscopy for each WT-specific probe duplexed either to its complementary (PM – perfect match) germline sequence or to a G12/G13 mutant allele.

	T_m (°C)						
Allele	**WTP1**	**WTP2**	**WTP3**	**WTP4**	**WTP5**	**WTP6**	**WTP7**
PM	**71.7**	**67.8**	**68.4**	**69.4**	**69.6**	**71.7**	**70.8**
G12A	54.8	52.4	54.2	57.1	52.9	53.7	52.1
G12C	57.1	56.5	55.5	55.2	51.9	52.7	53.3
G12D	54.6	51.5	54.2	59.8	55.7	51.6	50.7
G12R	57.9	57.2	54.6	56.1	57.3	49.2	51.9
G12S	56.1	53.6	52.5	54.1	54.9	51.4	50.7
G12V	50.1	53.7	56.5	54.3	54.0	52.5	52.8
G13D	50.3	56.4	58.4	59.5	54.8	55.8	60.0

Use of these LNA/DNA chimeric probes in the *KRAS* WT-negative screening assay thereby results in the segregation and differentiation of the WT data cluster from all MT data clusters (**Figure 4.3**), permitting detection of a G12/G13 missense mutation and accurate quantification of the mutant frequency through MT and total *KRAS*-positive droplet counts.

4.3.3 Assay Application to Plasmid, Cell-Line and FFPE Standards

Serial dilutions ($n \geq 4$ for each dilution) of MT *KRAS* plasmid DNA into WT1 *KRAS* plasmid DNA to MT frequencies down to 0.01% were used to define the LOD for each G12/G13 mutation. For *KRAS* G12D, an LOD of ≤ 0.025% MF was recorded (**Figure 4.4A**). An equivalent LOD was recorded for all other known G12/G13 missense mutations (e.g., **Figure 4.4B**). The assay was also applied to gDNA isolated from nine cell-lines harboring various *KRAS* G12/G13 mutations, and we again observed significant linear correlation ($R^2 \geq 0.998$; $P < 0.05$) between the expected MF and that determined using the ddPCR assay, with the basic metrics of assay performance (LOD, limit of blank (LOB), confidence intervals, dynamic range) remaining unchanged from those recorded for plasmid DNA.

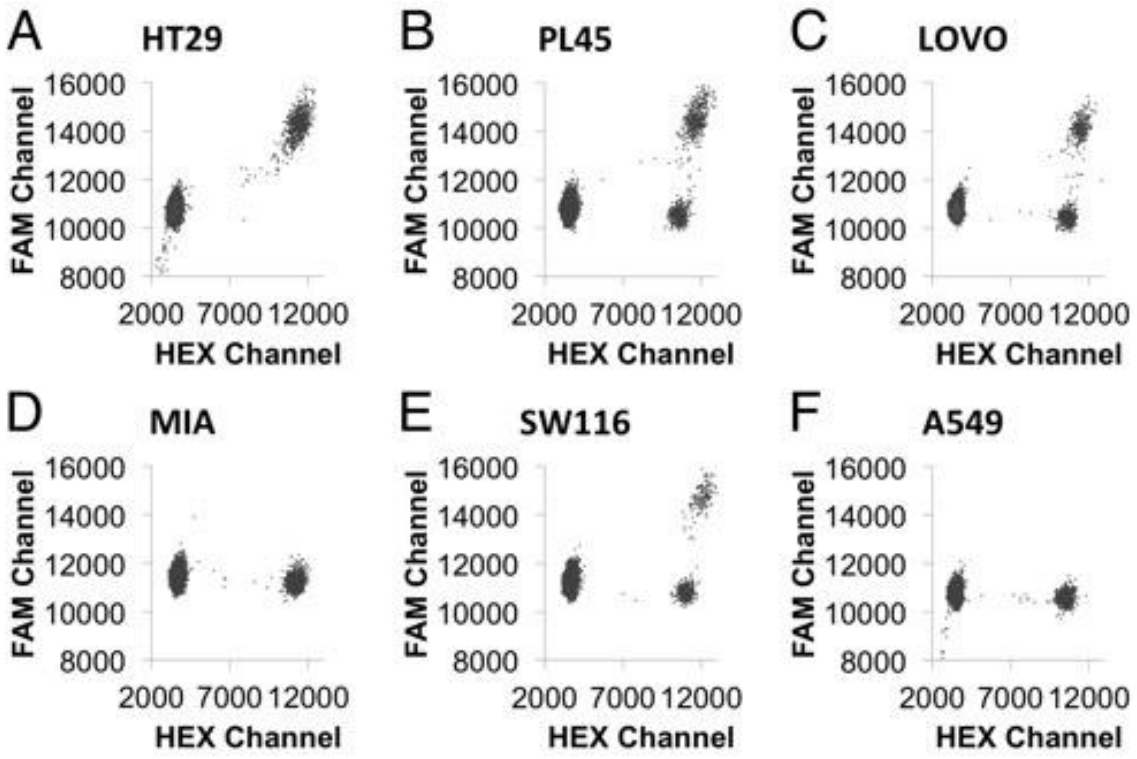

Figure 4.3 Representative output from the WT-negative KRAS screening assay when applied to gDNA isolated from various cell lines. A) HT29 cells (WT *KRAS*); B) PL45 cells (heterozygous for *KRAS* G12D); C) LOVO cells (heterozygous for *KRAS* G13D); D) MIA cells (homozygous for *KRAS* G12C); E) SW116 cells (heterozygous for *KRAS* G12A); and F) A549 cells (homozygous for *KRAS* G12S). Nomenclature: FAM, 6-carboxyfluorescein; HEX, hexachloro-fluorescein; WT, wild type.

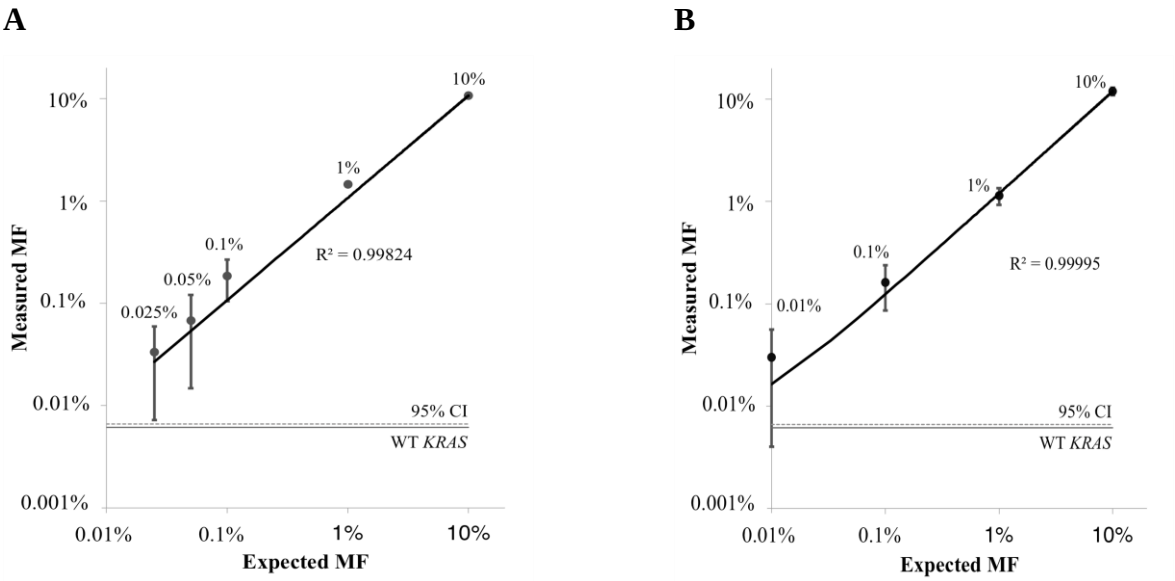

Figure 4.4 Analytical sensitivity (LOD) of the KRAS screening assay when applied to plasmid DNA. A) *KRAS* G12D into WT1 *KRAS*, B) G12S into WT1. Measured mutant frequency (MF) and standard deviation values are plotted versus expected MF for serial dilutions down to 0.01% MF. The dotted line is the limit of blank of the *KRAS* screening assay. Significant linear correlation ($R^2 \geq 0.998$; $p < 0.05$) between the measured and expected MF is observed down to the analytical detection limit (LOD) of 0.025% MF for G12D. Replicates ($n = 24$) of WT *KRAS* plasmid were used to define the mean false positive (WT *KRAS*, solid horizontal line) and SDs, from which the 95% confidence interval (CI; dashed horizontal line) was determined and used to define the LOB (0.007%). Nomenclature: LOD, limit of detection; LOB, limit of blank; MF, mutant frequency; WT, wild-type.

Finally, the ddPCR *KRAS* status assay was applied to gDNA prepared from seven *KRAS* FFPE reference standards, each harboring 5% MF in codons G12/G13. Once again, statistically significant G12/G13 missense mutation calls could be made (**Table 4.3**).

Table 4.3 Application of the WT-negative KRAS screening assay to gDNA from 5% MF FFPE standards.Measured mean MF and SD (*n* = 2) values are reported.

***KRAS* Allele**	**Expected %MF**	**Measured %MF (Mean ± (SD))**
WT	0	0.00 (± 0.00)
G12D	5	4.59 (± 0.07)
G12A	5	6.23 (± 0.87)
G12C	5	7.18 (± 1.25)
G12S	5	5.54 (± 0.61)
G12R	5	3.48 (± 1.03)
G12V	5	4.17 (± 0.84)
G13D	5	7.21 (± 1.03)

DNA was extracted from FFPE standards bearing either WT *KRAS* or a 5% *KRAS* G12/G13 MF as indicated. Nomenclature: ddPCR, droplet-digital PCR; MF, mutant frequency.

4.3.4 Application to FFPE-stabilized MLH1-deficient CRC tumor cores

Colorectal carcinoma cases were selected from a pool of MLH1-deficient tumors identified by IHC testing as part of the population-based VCH Lynch-syndrome screening program. Three non-MLH1 deficient cases (samples 3, 5, 20) were also included.

Duplicate gDNA specimens purified from the cohort of $n = 87$ tumor samples were subjected to our ddPCR-based assay and representative output data are shown in **Figure 4.5**.

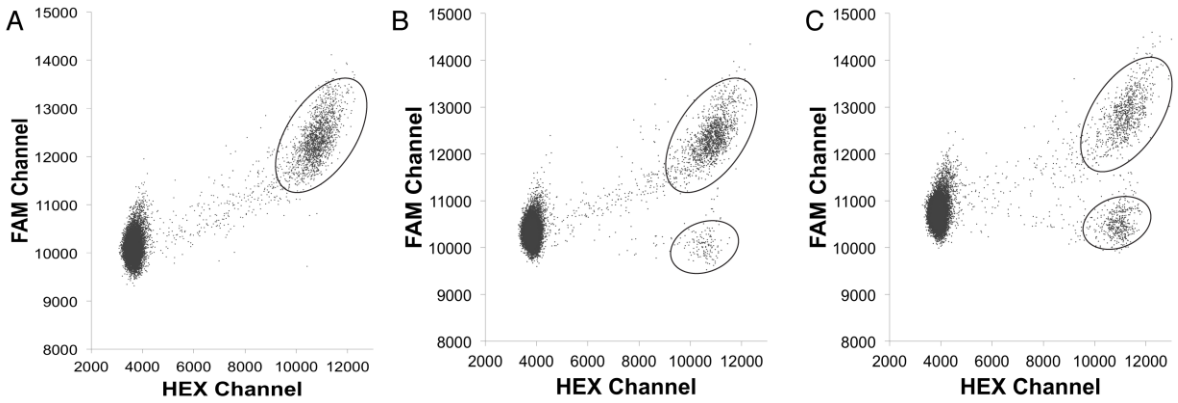

Figure 4.5 WT-negative KRAS G12/G13 screening assay data for representative clinical CRC tumor specimens. A) Representative G12/G13 WT sample (#1) (Table 4.4); B) Representative G12/G13 MT-positive sample (#20) displaying a low MF (10.95 (±1.63) %); C) Representative G12/13 MT-positive sample (#5) displaying a high MF (52.84 (±0.96) %). Nomenclature: CRC, colorectal cancer; MLH1, Mut L homologue; ddPCR, droplet digital PCR; WT, wild-type; MT, mutant.

The measured *KRAS* status of each specimen is reported in **Table 4.4**. 11 of the 87 CRC samples tested positive for a G12/G13 missense mutation, while the remaining 76 samples were WT *KRAS* positive. The *BRAF V600* status of each specimen was also determined (**Table 4.4**) using the *BRAF* WT-negative ddPCR assay we previously reported.[233] Mutational testing of each specimen was also conducted independently by the Canadian Immunohistochemistry Quality Control (cIQc) agency. Test results obtained from these orthogonal approaches were in agreement.

Of the 87 specimens tested using the pair of ddPCR-based assays, 54 were MT *BRAF* V600 positive/MT *KRAS* G12/G13 negative, a phenotype often observed in MLH1-deficient/MSI-H CRCs. However, 2 specimens (samples 4 and 37) were positive for a missense mutation in both *BRAF* V600 and *KRAS* G12/G13. Importantly, for sample 4 the ddPCR assay records a *KRAS* G12/G13 missense mutation having a MF of 2.3%, which falls below the detection limit of either the TheraScreen or cobas *KRAS* assay.

Table 4.4 Analysis of FFPE tumor specimens from a cohort of 87 MLH1-deficient colorectal cancer patients using the WT-negative KRAS screening assay and a previously reported ddPCR-based BRAF V600 screening assay.[233]Results for either assay are expressed as WT or MT positive, with the mutant frequency (MF) and standard error (in parenthesis) reported for each *KRAS* MT-positive sample. FFPE processing of tissue samples serves to degrade the quality and quantity of isolated amplifiable DNA. Droplet frequencies and standard errors recorded for the 76 *KRAS* WT-positive specimens identified below were therefore used to compute the limit of quantitation (LOQ) of the *KRAS* screening assay when applied to clinical specimens, which was found to be 0.38 (± 0.15) %. Sample 4 and sample 37 were positive for a missense mutation in both *KRAS* G12/G13 and *BRAF* V600; a *BRAF* V600 MF of 30.00 (±1.10) % and 18.30 (±3.30) %, were recorded for samples 4 and 37, respectively.

Patient Sample	***KRAS* G12/13 Call**	***KRAS* G12/G13 MF**	***BRAF* V600 Call**
1	WT	-	WT
2	WT	-	**MT**
3	WT	-	WT
4	**MT**	**2.26 (±0.72) %**	**MT**
5	**MT**	**52.84 (±0.96) %**	WT
6	WT	-	**MT**
7	WT	-	**MT**
8	WT	-	WT
9	WT	-	**MT**
10	WT	-	WT
11	WT	-	WT
12	WT	-	**MT**
13	WT	-	WT
14	WT	-	**MT**
15	WT	-	WT
16	WT	-	**MT**
17	WT	-	**MT**
18	WT	-	**MT**
19	WT	-	**MT**
20	**MT**	**10.95 (±1.63) %**	WT
21	WT	-	**MT**
22	WT	-	WT
23	WT	-	**MT**
24	WT	-	MT
25	WT	-	**MT**
26	WT	-	WT
27	WT	-	**MT**
28	WT	-	**MT**
29	WT	-	**MT**
30	WT	-	**MT**

Patient Sample	***KRAS* G12/13 Call**	***KRAS* G12/G13 MF**	***BRAF* V600 Call**
31	WT	-	WT
32	WT	-	**MT**
33	WT	-	**MT**
34	WT	-	**MT**
35	**MT**	**58.43 (±7.46) %**	WT
36	WT	-	**MT**
37	**MT**	**7.48 (±0.21) %**	**MT**
38	WT	-	WT
39	WT	-	**MT**
40	WT	-	WT
41	WT	-	WT
42	WT	-	**MT**
43	WT	-	WT
44	WT	-	**MT**
45	WT	-	**MT**
46	**MT**	**67.67 (±2.28)** %	WT
47	WT	-	**MT**
48	WT	-	**MT**
49	WT	-	WT
50	WT	-	**MT**
51	**MT**	**39.02 (±0.39)** %	WT
52	WT	-	WT
53	WT	-	MT
54	**MT**	**49.67 (±1.36)** %	WT
55	WT	-	WT
56	WT	-	**MT**
57	WT	-	WT
58	WT	-	WT
59	WT	-	WT
60	WT	-	**MT**
61	**MT**	**29.54 (±1.68)** %	WT
62	WT	-	**MT**
63	WT	-	WT
64	WT	-	**MT**
65	WT	-	**MT**
66	WT	-	**MT**
67	WT	-	WT
68	WT	-	**MT**
69	WT	-	**MT**
70	WT	-	**MT**
71	WT	-	**MT**
72	**MT**	**43.23 (±0.33)** %	WT

Patient Sample	***KRAS* G12/13 Call**	***KRAS* G12/G13 MF**	***BRAF* V600 Call**
73	WT	-	**MT**
74	WT	-	**MT**
75	WT	-	WT
76	WT	-	**MT**
77	WT	-	WT
78	**MT**	**10.75 (±0.24)** %	WT
79	WT	-	**MT**
80	WT	-	**MT**
81	WT	-	**MT**
82	WT	-	**MT**
83	WT	-	**MT**
84	WT	-	**MT**
85	WT	-	**MT**
86	WT	-	**MT**
87	WT	-	**MT**

4.4 Discussion

In CRC patients, missense mutation of *KRAS* is a clinically actionable biomaker of negative response to anti-EGFR mAb therapy. As delineated in **Table 4.5**, a single base missense mutation in either codon 12 or 13 is the most common MT *KRAS* in CRC, with a c.35G > T (p.Gly12Val) or c.35G > A (p.Gly12Asp) mutation observed with highest frequency in codon 12, else a c.38G > A (p.Gly13Asp) mutation in codon 13.

Table 4.5 Frequency of clinically relevant KRAS G12/G13 missense mutations and WT alleles different from WT1.

***KRAS* Mutation**	**Nucleotide Change**	**Frequency**
p.G12D	c.35G>A	35.0%
p.G12V	c.35G>T	23.8%
p.G13D	c.38G>A	13.1%
p.G12C	c.34G>T	11.9%
p.G12A	c.35G>C	5.7%
p.G12R	c.34G>C	3.2%
p.G13C	c.37G>T	0.9%
p.G13S	c.37G>A	0.2%
p.G13R	c.37G>C	0.2%
p.G13A	c.38G>C	0.1%
p.G13V	c.38G>T	0.1%
G12-13	Complex	0.4%
G12G/G13G	Synonymous	0.1%

Other single-base missense mutations have been observed at low frequency in both codons, as well as in *KRAS* codons 61, 117, and 146.[238] Rare instances of mutation of more than one base in codons 12/13 have likewise been reported, as have deletion and insertion mutations.[11] Current dogma is that mutation of *BRAF* V600 and *KRAS* G12/G13 tend toward mutually exclusive events in CRC, presumably due to their co-involvement in regulation of MAPK/ERK signaling;[202, 236, 239] but our results, when combined with a previous study reporting three additional cases of CRC patients carrying co-mutations,[236] challenge this assertion by suggesting missense mutations in both of these oncogenic hot spots can be observed with reasonable frequency.

Methods widely employed by clinics and oncogenetics labs to test CRC patient populations for *KRAS* mutations include Sanger sequencing and various next-generation sequencing technologies, as well as various modalities of qPCR. The methods differ in the mutations they can identify and in their sensitivities in detecting those mutations, though in general they cannot detect a MF below ca. 5%.[121, 240] Thus, though widely used, sequencing and qPCR strategies typically do not offer sufficient analytical sensitivity to detect mutations in biopsy specimens with low tumor content, including those isolated from surgical margins. Some improvements in MT-detection sensitivity can be achieved through careful microdissection of a tissue specimen to enrich the tumor cell density prior to gDNA extraction, but that preparation procedure is time-consuming and typically improves sensitivity by a factor of less than 2.

The WT-negative *KRAS* screening assay described in this work addresses this limitation and associated unmet clinical need by providing an inexpensive (cost of goods of ~ $8 US per specimen) and rapid (96 samples can be assayed in 6 hrs) means for clinics to reliably and quantitatively detect missense mutations in codons G12/G13 of *KRAS*. The ddPCR-based assay offers an analytical sensitivity of 0.025% MF when applied to MT standards.

When then applied in conjunction with the ddPCR-based *BRAF* V600 status assay (Chapter 3) to FFPE tumor biopsies from 87 mCRC patients deficient in the *MLH1* DNA repair gene, two patients were identified as carrying a missense mutation both in *KRAS* G12/G13 and in *BRAF* V600. Importantly, for one of those patients (#4) the *KRAS* MT frequency was low (2.26 (±0.72) %), falling below the detection limit of either sequencing or FDA-approved qPCR assays of *KRAS* G12/G13 status (at best 5%). Thus, the rare observance to date of *KRAS* and *BRAF* co-mutation may be due in part to the limited ability of clinics to detect the pathology.

CRC is thought to arise predominantly (but not exclusively) through one of two genetic pathways. In the classical pathway,[81] mutation of proto-oncogenes within the MAPK and PI3K signaling pathways and/or within various tumor suppressor genes serve to deregulate control of cell proliferation and apoptosis.[241] Common driver mutations activating this pathway include ones within *KRAS*, *NRAS*, *APC* (adenomatous polyposis coli) – a multi-functional tumor

suppressor gene – and *TP53* – a gene encoding tumor suppressor protein 53 (p53) involved in regulating the cell cycle. Pathogenic missense variants of *KRAS* and *NRAS* in particular, as well as of *APC* or *TP53*, can encode mutant versions of the gene product that no longer regulate cell growth and division properly.

The alternative serrated pathway to CRC is characterized in part by impairment of DNA mismatch repair (MMR) functions that often results from MSI-H. Lynch syndrome (LS), an autosomal dominant condition associated with inherited polymorphisms in genes within the MMR pathway, is the most common hereditary predisposition to CRC. Roughly 3% of all CRC malignancies and ~15% of MSI-H CRCs progress from LS via somatic-mutation induced failure of the MMR pathway, often through epigenetic silencing of *MLH1* by hyper-methylation of the promoter controlling transcription. Oncogenic *BRAF* V600 mutations are associated with MSI and are observed in ca. 60% MSI-H tumors, but in only 5 – 10% of microsatellite stable (MSS) tumors. MT *BRAF* V600 is predictive of poor prognosis and increased mortality in MSS patients, while survival statistics for serrated-pathway CRC patients displaying the (MT *BRAF* V600/MSI-H)-positive subtype are significantly better.

When combined with the previous study by Sahin *et al.*,[236] the finding of CRC patients carrying a missense mutation in both *KRAS* G12/G13 and *BRAF* V600 suggests that overlap between or co-existence of the two primary CRC pathologies, at least with respect to genetic signatures, may occur at a low but statistically significant frequency. Precisely how this co-mutation arises remains unclear. There is, however, evidence that co-mutation of two MAPK-associated oncogenes within the same clone is generally lethal.[242] The observance of both *BRAF* V600 and *KRAS* G12/G13 missense mutations may therefore reflect a clonally heterogeneous tumor, which is consistent with the considerable difference in KRAS and BRAF MFs observed in each patient.

As noted, one patient found in this study to carry missense mutations in both *KRAS* and *BRAF* exhibited a *KRAS* MF < 5%, and thus would not have been classified as MT-*KRAS*/MT-*BRAF* positive by other tests, including the cobas platform which provides for testing of both *BRAF* and *KRAS* mutations. This makes clear the potential clinical value of the ddPCR-based *KRAS*

status test reported here, which quantifies MF down to a detection limit of ~1%. Moreover, it points to the potential need to more comprehensively sub-classify CRCs in terms of a larger panel of tumor characteristics that might include MMR gene status, chromosomal instability phenotype, *KRAS*, *BRAF*, *APC* and *FBXW7* (an F-box ubiquitin ligase that is thought to antagonize cancer development) mutational status, and CIMP status. While it is known that *BRAF* V600 mutation promotes hyper-methylation of the *MLH1* gene promotor, the impact of co-mutation of KRAS on that process and on activation of the MAPK pathway is not understood.

Chapter 5: Future Work

Future work based on the concepts and findings presented in this thesis includes both an existing plan for continued clinical adoption of the ddPCR-based *BRAF* and *KRAS* wild-type negative tests I developed, and exploration of ideas for new wild-type negative tests and areas for their application.

By working in collaboration with clinicians at Lions Gate Hospital (N. Vancouver) and the Canadian Immunohistochemistry Quality Control Centre (Vancouver), I was able to test and validate both assays on genomic DNA samples from a cohort of colorectal cancer patients with known pathologies. On each sample tested, the assays were judged clinically accurate and informative. But the number of patients tested was relatively small – less than 100 in each case. As a goal for each assay is to realize its widespread clinical adoption and use, further clinical validation will be required, ideally by one or more independent clinical testing laboratories with expertise in digital pathology and ddPCR. This is particularly relevant to the goal of introducing the assays into genetic testing clinics in the US, as legislation approved as part of the Affordable Care Act now mandates FDA approval of genetic tests for either inherited or acquired diseases. Independent testing results are required by the FDA as part of the filing and approval process for new genetic tests for human disease or therapeutic eligibility.

Specifically, clinical validation at the standard imposed by the FDA would require a third party CLIA laboratory to conduct retrospective and prospective double-blinded studies comparing *BRAF* V600 or *KRAS* G12/13 WT negative assays results to patient outcomes and/or therapeutic responses. While designed to define the patient population to which each wild-type negative assay can be applied, these studies will also serve to refine the limit of quantitation (LOQ) of each test; that is, the minimum mutant frequency at which a clinical action based on the presence of that mutation can be taken without risking harm to the patient.

Further refinements to and clinical adoption of the data-analysis tools (and graphical user interface) presented and used in this thesis work would also prove valuable in enhancing clinical adoption and use of wild-type negative assays. The automated data analysis method presented in

Appendix B and used throughout the thesis is capable of accurately identifying a limited set of well-defined clusters within the FAM+ and HEX+FAM+ quadrants. The R code could in principle be extended to enable automated identification of clusters in other quadrants as well. That extended data analysis tool pack would then be applicable to not only the wild-type negative assays described in this work, but more generally to any multiplexed ddPCR assay in which the output data display as a set of identifiable droplet clusters.

The quality and accuracy of genetic testing are known to depend on the quality and quantity of the DNA (or RNA) purified from the specimen and used. While this issue has been exhaustively studied for competing testing technologies, most notably qPCR assays and next-generation sequencing tests, very little is known about how DNA quality recovered from patient specimens, particularly the method used to recover DNA from FFPE samples, impacts genetic assays conducted using ddPCR. Incomplete deparaffinization of the sample, as well as cross-linking of the DNA during formalin fixation, are known to impact gene amplification kinetics in traditional PCR.[243] It therefore stands to reason that these issues might also impact template amplification within a droplet format. In this thesis, genomic DNA was isolated from FFPE specimens using a commercial kit (QIAamp FFPE tissue kit; Qiagen) that is widely used in clinics. The results obtained and reported in chapters 3 and 4 show that detection limits are an order of magnitude lower than current FDA-approved *BRAF* and *KRAS* tests can be achieved when the ddPCR tests I developed are applied to gDNA recovered from FFPE samples using this commercial kit. Nevertheless, more attention certainly needs to be given to understanding the impact of gDNA quality on results obtained by a ddPCR-based WT-negative test, and to establishing best clinical practices with respect to the DNA purification protocol(s) used.

While instrumentation development was not a focus of my thesis, it is important to note that the FDA and cancer testing clinics are only just now truly considering approving and adopting genetic tests conducted by digital PCR. The robustness and reliability of ddPCR instrumentation in a clinical setting therefore remain uncertain, but a general concern that has been raised with regard to droplet digital ddPCR instruments is the variability in the number of "readable" droplets created for a given sample. Efforts to reduce that variation and to realize other instrument improvements would be of clear value.

Finally, an R program and associated graphical user interface that permits automated solution of the molecular thermodynamic model used in this thesis to design LNA-substituted dual-labeled hydrolysis probes does not currently exist. Its creation and availability as a web-based tool could be of great value, as it would enable clinicians and other users outside of our laboratory to rapidly design truly allele-specific probes in silico.

Ideas for new applications of the platform presented in this thesis include creation of WT-negative assays against other oncogenic biomarkers comprised of one or more missense mutations that either drive cancer progression or determine the eligibility of a patient for a given targeted therapy. Important clinically relevant examples include *PIK3CA* and *NRAS*.[80, 87] The *PIK3CA* gene on chromosome 3q26.3 codes for a 124 kDa size protein, PIK3Ca, a heterodimeric lipid kinase. Somatic missense mutations in the *PIK3CA* gene are thought to drive many human cancer types, including colorectal, breast, brain, liver, stomach and lung cancers. Each serves to increase the kinase activity of the mutated PIK3CA, contributing to deregulation of cellular proliferation and transformation. Oncogenic hotspot codons in *PIK3CA* include E542, E545 and H1047, and missense mutations in each are implicated in cancer progression.

Approximately 20% of all metastatic melanoma patients harbor a missense mutation in *NRAS*, a gene on human chromosome 1 coding for the NRas protein, a GTPase that converts GTP into GDP. It is the second most common type of mutation in melanoma after *BRAF* mutations, with mutational hotspots present at codons 12, 13 and 61. *NRAS* and *BRAF* mutations are thought to be mutually exclusive, but the incidence of *NRAS* mutations among metastatic melanoma patients wild-type for *BRAF* is almost 40%. Melanoma patients positive for an *NRAS* mutation show a higher mitotic rate and therefore generally exhibit a much poorer prognosis. Moreover, there is evidence that *NRAS* mutations may affect the clinical outcome of melanoma patients treated with immune therapies. In the context of this thesis work, it is important to note that treatment of colorectal cancer with anti-EGFR mAbs has proven most effective in patients wild-type for not only *BRAF* and *KRAS*, but for *NRAS* and PIK3CA as well.[80, 87] The concept of the multiplexed WT-negative assay might therefore be extended to more fully stratify CRC patient populations.

A final and particularly exciting possible application of the platform described in this thesis is its use in creating assays for detecting oncogenic biomarkers within the free circulating DNA (cf-DNA) population in a patient's blood. Significant amounts of circulating cf-DNA are present in the plasma of cancer patients, and the concept of "liquid biopsies" is now very attractive as it allows for non-invasive monitoring of cancer patients post-treatment. Thus a potentially impactful advance, at least from the perspective of a clinical need, would be the development of WT negative ddPCR assays capable of amplifying fragments shorter that < 160 bp and therefore interrogating the short cf-DNA fragments typically found in plasma.[244]

References

[1] Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M, Hogg D, Lorigan P, Lebbe C, Jouary T, Schadendorf D, Ribas A, O'Day SJ, Sosman JA, Kirkwood JM, Eggermont AM, Dreno B, Nolop K, Li J, Nelson B, Hou J, Lee RJ, Flaherty KT, McArthur GA: Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 2011, 364:2507-16.
[2] Sosman JA, Kim KB, Schuchter L, Gonzalez R, Pavlick AC, Weber JS, McArthur GA, Hutson TE, Moschos SJ, Flaherty KT, Hersey P, Kefford R, Lawrence D, Puzanov I, Lewis KD, Amaravadi RK, Chmielowski B, Lawrence HJ, Shyr Y, Ye F, Li J, Nolop KB, Lee RJ, Joe AK, Ribas A: Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. N Engl J Med 2012, 366:707-14.
[3] Lièvre A, Bachet J-B, Le Corre D, Boige V, Landi B, Emile J-F, Côté J-F, Tomasic G, Penna C, Ducreux M, Rougier P, Penault-Llorca F, Laurent-Puig P: KRAS Mutation Status Is Predictive of Response to Cetuximab Therapy in Colorectal Cancer. Cancer Research 2006, 66:3992-5.
[4] Siegel R, Ma J, Zou Z, Jemal A: Cancer statistics, 2014. CA: A Cancer Journal for Clinicians 2014, 64:9-29.
[5] Kalia M: Personalized oncology: recent advances and future challenges. Metabolism 2013, 62 Suppl 1:S11-4.
[6] Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML: Projections of the cost of cancer care in the United States: 2010-2020. J Natl Cancer Inst 2011, 103:117-28.
[7] Schilsky RL: Personalized medicine in oncology: the future is now. Nat Rev Drug Discov 2010, 9:363-6.
[8] Schilsky RL: Personalizing cancer care: American Society of Clinical Oncology presidential address 2009. J Clin Oncol 2009, 27:3725-30.
[9] Griewank KG, Scolyer RA, Thompson JF, Flaherty KT, Schadendorf D, Murali R: Genetic alterations and personalized medicine in melanoma: progress and future prospects. J Natl Cancer Inst 2014, 106:djt435.
[10] Richter A, Grieu F, Carrello A, Amanuel B, Namdarian K, Rynska A, Lucas A, Michael V, Bell A, Fox SB, Hewitt CA, Do H, McArthur GA, Wong SQ, Dobrovic A, Iacopetta B: A multisite blinded study for the detection of BRAF mutations in formalin-fixed, paraffin-embedded malignant melanoma. Sci Rep 2013, 3:1659.
[11] Muzny DM BM, Chang K, Dinh HH, Drummond JA, Fowler G, Kovar CL, Lewis LR, Morgan MB, Newsham IF, Reid JG, Santibanez J, Shinbrot E, Trevino LR, Wu YQ, Wang M, Gunaratne P, Donehower LA, Creighton CJ, Wheeler DA, Gibbs RA, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, Stojanov P, McKenna A, Lander ES, Gabriel S, Getz G, Ding L, Fulton RS, Koboldt DC, Wylie T, Walker J, Dooling DJ, Fulton L, Delehaunty KD, Fronick CC, Demeter R, Mardis ER, Wilson RK, Chu A, Chun HJ, Mungall AJ, Pleasance E, Robertson A, Stoll D, Balasundaram M, Birol I, Butterfield YS, Chuah E, Coope RJ, Dhalla N, Guin R, Hirst C, Hirst M, Holt RA, Lee D, Li HI, Mayo M, Moore RA, Schein JE, Slobodan JR, Tam A, Thiessen N, Varhol R, Zeng T, Zhao Y, Jones SJ, Marra MA, Bass AJ, Ramos AH, Saksena G, Cherniack AD, Schumacher SE, Tabak B, Carter SL, Pho NH, Nguyen H, Onofrio RC, Crenshaw A, Ardlie K, Beroukhim R, Winckler W, Getz G, Meyerson M, Protopopov A, Zhang J, Hadjipanayis A, Lee E, Xi R, Yang L, Ren X, Zhang H, Sathiamoorthy N, Shukla S, Chen PC, Haseley P, Xiao Y, Lee S, Seidman J, Chin L, Park PJ, Kucherlapati R, Auman JT,

Hoadley KA, Du Y, Wilkerson MD, Shi Y, Liquori C, Meng S, Li L, Turman YJ, Topal MD, Tan D, Waring S, Buda E, Walsh J, Jones CD, Mieczkowski PA, Singh D, Wu J, Gulabani A, Dolina P, Bodenheimer T, Hoyle AP, Simons JV, Soloway M, Mose LE, Jefferys SR, Balu S, O'Connor BD, Prins JF, Chiang DY, Hayes D, Perou CM, Hinoue T, Weisenberger DJ, Maglinte DT, Pan F, Berman BP, Van Den Berg DJ, Shen H, Triche T Jr, Baylin SB, Laird PW, Getz G, Noble M, Voet D, Saksena G, Gehlenborg N, DiCara D, Zhang J, Zhang H, Wu CJ, Liu SY, Shukla S, Lawrence MS, Zhou L, Sivachenko A, Lin P, Stojanov P, Jing R, Park RW, Nazaire MD, Robinson J, Thorvaldsdottir H, Mesirov J, Park PJ, Chin L, Thorsson V, Reynolds SM, Bernard B, Kreisberg R, Lin J, Iype L, Bressler R, Erkkilä T, Gundapuneni M, Liu Y, Norberg A, Robinson T, Yang D, Zhang W, Shmulevich I, de Ronde JJ, Schultz N, Cerami E, Ciriello G, Goldberg AP, Gross B, Jacobsen A, Gao J, Kaczkowski B, Sinha R, Aksoy B, Antipin Y, Reva B, Shen R, Taylor BS, Chan TA, Ladanyi M, Sander C, Akbani R, Zhang N, Broom BM, Casasent T, Unruh A, Wakefield C, Hamilton SR, Cason R, Baggerly KA, Weinstein JN, Haussler D, Benz CC, Stuart JM, Benz SC, Sanborn J, Vaske CJ, Zhu J, Szeto C, Scott GK, Yau C, Ng S, Goldstein T, Ellrott K, Collisson E, Cozen AE, Zerbino D, Wilks C, Craft B, Spellman P, Penny R, Shelton T, Hatfield M, Morris S, Yena P, Shelton C, Sherman M, Paulauskis J, Gastier-Foster JM, Bowen J, Ramirez NC, Black A, Pyatt R, Wise L, White P, Bertagnolli M, Brown J, Chan TA, Chu GC, Czerwinski C, Denstman F, Dhir R, Dörner A, Fuchs CS, Guillem JG, Iacocca M, Juhl H, Kaufman A, Kohl B 3rd, Van Le X, Mariano MC, Medina EN, Meyers M, Nash GM, Paty PB, Petrelli N, Rabeno B, Richards WG, Solit D, Swanson P, Temple L, Tepper JE, Thorp R, Vakiani E, Weiser MR, Willis JE, Witkin G, Zeng Z, Zinner MJ, Zornig C, Jensen MA, Sfeir R, Kahn AB, Chu AL, Kothiyal P, Wang Z, Snyder EE, Pontius J, Pihl TD, Ayala B, Backus M, Walton J, Whitmore J, Baboud J, Berton DL, Nicholls MC, Srinivasan D, Raman R, Girshik S, Kigonya PA, Alonso S, Sanbhadti RN, Barletta SP, Greene JM, Pot DA, Shaw KR, Dillon LA, Buetow K, Davidsen T, Demchok JA, Eley G, Ferguson M, Fielding P, Schaefer C, Sheth M, Yang L, Guyer MS, Ozenberger BA, Palchik JD, Peterson J, Sofia HJ, Thomson E.: Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012, 487:330-7.

[12] Parry JM, Shamsher M, Skibinski DO: Restriction site mutation analysis, a proposed methodology for the detection and study of DNA base changes following mutagen exposure. Mutagenesis 1990, 5:209-12.

[13] Kuan SF, Navina S, Cressman KL, Pai RK: Immunohistochemical detection of BRAF V600E mutant protein using the VE1 antibody in colorectal carcinoma is highly concordant with molecular testing but requires rigorous antibody optimization. Hum Pathol 2014, 45:464-72.

[14] Long GV, Wilmott JS, Capper D, Preusser M, Zhang YE, Thompson JF, Kefford RF, von Deimling A, Scolyer RA: Immunohistochemistry is highly sensitive and specific for the detection of V600E BRAF mutation in melanoma. Am J Surg Pathol 2013, 37:61-5.

[15] Choleva Y, Kongsbak L, Moller S, Mouritzen P, Nielsen AT, Pfundheller HM: Single nucleotide polymorphism genotyping using locked nucleic acid (LNA[TM]). Expert Review of Molecular Diagnostics 2003, 3:27+.

[16] Krypuy M, Newnham GM, Thomas DM, Conron M, Dobrovic A: High resolution melting analysis for the rapid and sensitive detection of mutations in clinical samples: KRAS codon 12 and 13 mutations in non-small cell lung cancer. BMC Cancer 2006, 6:295.

[17] Simi L, Pratesi N, Vignoli M, Sestini R, Cianchi F, Valanzano R, Nobili S, Mini E, Pazzagli M, Orlando C: High-resolution melting analysis for rapid detection of KRAS, BRAF, and PIK3CA gene mutations in colorectal cancer. Am J Clin Pathol 2008, 130:247-53.

[18] Didelot A, Le Corre D, Luscan A, Cazes A, Pallier K, Emile J-F, Laurent-Puig P, Blons H: Competitive allele specific TaqMan PCR for KRAS, BRAF and EGFR mutation detection in clinical formalin fixed paraffin embedded samples. Experimental and Molecular Pathology 2012, 92:275-80.
[19] Heinzerling L, Kuhnapfel S, Meckbach D, Baiter M, Kaempgen E, Keikavoussi P, Schuler G, Agaimy A, Bauer J, Hartmann A, Kiesewetter F, Schneider-Stock R: Rare BRAF mutations in melanoma patients: implications for molecular testing in clinical practice. Br J Cancer 2013, 108:2164-71.
[20] Huang M, Shen A, Ding J, Geng M: Molecularly targeted cancer therapy: some lessons from the past decade. Trends Pharmacol Sci 2014, 35:41-50.
[21] Navarro E, Serrano-Heras G, Castano MJ, Solera J: Real-time PCR detection chemistry. Clin Chim Acta 2015, 439C:231-50.
[22] Anderson S, Bloom KJ, Vallera DU, Rueschoff J, Meldrum C, Schilling R, Kovach B, Lee JR-J, Ochoa P, Langland R, Halait H, Lawrence HJ, Dugan MC: Multisite Analytic Performance Studies of a Real-Time Polymerase Chain Reaction Assay for the Detection of BRAF V600E Mutations in Formalin-Fixed, Paraffin-Embedded Tissue Specimens of Malignant Melanoma. Archives of Pathology & Laboratory Medicine 2012, 136:1385-91.
[23] Curry JL, Torres-Cabala CA, Tetzlaff MT, Bowman C, Prieto VG: Molecular platforms utilized to detect BRAF V600E mutation in melanoma. Semin Cutan Med Surg 2012, 31:267-73.
[24] Halait H, Demartin K, Shah S, Soviero S, Langland R, Cheng S, Hillman G, Wu L, Lawrence HJ: Analytical performance of a real-time PCR-based assay for V600 mutations in the BRAF gene, used as the companion diagnostic test for the novel BRAF inhibitor vemurafenib in metastatic melanoma. Diagn Mol Pathol 2012, 21:1-8.
[25] Vester B, Wengel J: LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. Biochemistry 2004, 43:13233-41.
[26] Ugozzoli LA, Latorra D, Puckett R, Arar K, Hamby K: Real-time genotyping with oligonucleotide probes containing locked nucleic acids. Anal Biochem 2004, 324:143-52.
[27] You Y, Moreira BG, Behlke MA, Owczarzy R: Design of LNA probes that improve mismatch discrimination. Nucleic Acids Research 2006, 34:e60.
[28] Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC, Kitano TK, Hodel MR, Petersen JF, Wyatt PW, Steenblock ER, Shah PH, Bousse LJ, Troup CB, Mellen JC, Wittmann DK, Erndt NG, Cauley TH, Koehler RT, So AP, Dube S, Rose KA, Montesclaros L, Wang S, Stumbo DP, Hodges SP, Romine S, Milanovich FP, White HE, Regan JF, Karlin-Neumann GA, Hindson CM, Saxonov S, Colston BW: High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal Chem 2011, 83:8604-10.
[29] Pekin D, Skhiri Y, Baret J-C, Le Corre D, Mazutis L, Ben Salem C, Millot F, El Harrak A, Hutchison JB, Larson JW, Link DR, Laurent-Puig P, Griffiths AD, Taly V: Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. Lab on a Chip 2011, 11:2156-66.
[30] Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW: Cancer genome landscapes. Science 2013, 339:1546-58.
[31] Hughesman CB, Turner RFB, Haynes C: Stability and Mismatch Discrimination of DNA duplexes containing 2,6-Diaminopurine and 2-Thiothymidine Locked Nucleic Acid bases. Nucleic Acids Symposium Series 2008, 52:245-6.

[32] Hughesman CB, Turner RF, Haynes CA: Role of the heat capacity change in understanding and modeling melting thermodynamics of complementary duplexes containing standard and nucleobase-modified LNA. Biochemistry 2011, 50:5354-68.
[33] Hughesman CB, Turner RF, Haynes C: Correcting for heat capacity and 5'-TA type terminal nearest neighbors improves prediction of DNA melting temperatures using nearest-neighbor thermodynamic models. Biochemistry 2011, 50:2642-9.
[34] Reed GH, Kent JO, Wittwer CT: High-resolution DNA melting analysis for simple and efficient molecular diagnostics. Pharmacogenomics 2007, 8:597-608.
[35] Nalejska E, Maczynska E, Lewandowska MA: Prognostic and predictive biomarkers: tools in personalized oncology. Mol Diagn Ther 2014, 18:273-84.
[36] Kalia M: Biomarkers for personalized oncology: recent advances and future challenges. Metabolism 2015, 64:S16-21.
[37] Bailey AM, Mao Y, Zeng J, Holla V, Johnson A, Brusco L, Chen K, Mendelsohn J, Routbort MJ, Mills GB, Meric-Bernstam F: Implementation of biomarker-driven cancer therapy: existing tools and remaining gaps. Discov Med 2014, 17:101-14.
[38] Sawyers CL, van 't Veer LJ: Reliable and effective diagnostics are keys to accelerating personalized cancer medicine and transforming cancer care: a policy statement from the american association for cancer research. Clin Cancer Res 2014, 20:4978-81.
[39] Roberts PJ, Der CJ: Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. Oncogene 2007, 26:3291-310.
[40] Dhillon AS, Hagan S, Rath O, Kolch W: MAP kinase signalling pathways in cancer. Oncogene 0000, 26:3279-90.
[41] Abdel-Wahab O, Klimek VM, Gaskell AA, Viale A, Cheng D, Kim E, Rampal R, Bluth M, Harding JJ, Callahan MK, Merghoub T, Berger MF, Solit DB, Rosen N, Levine RL, Chapman PB: Efficacy of intermittent combined RAF and MEK inhibition in a patient with concurrent BRAF- and NRAS-mutant malignancies. Cancer Discov 2014, 4:538-45.
[42] Ney JT, Froehner S, Roesler A, Buettner R, Merkelbach-Bruse S: High-Resolution Melting Analysis as a Sensitive Prescreening Diagnostic Tool to Detect KRAS, BRAF, PIK3CA, and AKT1 Mutations in Formalin-Fixed, Paraffin-Embedded Tissues. Archives of Pathology & Laboratory Medicine 2012, 136:983-92.
[43] Lemmon MA, Schlessinger J: Cell signaling by receptor tyrosine kinases. Cell 2010, 141:1117-34.
[44] Yang H, Higgins B, Kolinsky K, Packman K, Go Z, Iyer R, Kolis S, Zhao S, Lee R, Grippo JF, Schostack K, Simcox ME, Heimbrook D, Bollag G, Su F: RG7204 (PLX4032), a selective BRAFV600E inhibitor, displays potent antitumor activity in preclinical melanoma models. Cancer Res 2010, 70:5518-27.
[45] Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins GJ, Bigner DD, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho JW, Leung SY, Yuen ST, Weber BL, Seigler HF, Darrow TL, Paterson H, Marais R, Marshall CJ, Wooster R, Stratton MR, Futreal PA: Mutations of the BRAF gene in human cancer. Nature 2002, 417:949-54.

[46] Santarpia L, Lippman SL, El-Naggar AK: Targeting the Mitogen-Activated Protein Kinase RAS-RAF Signaling Pathway in Cancer Therapy. Expert opinion on therapeutic targets 2012, 16:103-19.
[47] Rapp UR, Goldsborough MD, Mark GE, Bonner TI, Groffen J, Reynolds FH, Jr., Stephenson JR: Structure and biological activity of v-raf, a unique oncogene transduced by a retrovirus. Proc Natl Acad Sci U S A 1983, 80:4218-22.
[48] Sutrave P, Bonner TI, Rapp UR, Jansen HW, Patschinsky T, Bister K: Nucleotide sequence of avian retroviral oncogene v-mil: homologue of murine retroviral oncogene v-raf. Nature 1984, 309:85-8.
[49] Pratilas C, Xing F, Solit D: TARGETING ONCOGENIC BRAF IN HUMAN CANCER. Current topics in microbiology and immunology 2012, 355:83-98.
[50] Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, Teague JW, Stratton MR, McDermott U, Campbell PJ: COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Research 2015, 43:D805-D11.
[51] Curtin JA, Fridlyand J, Kageshita T, Patel HN, Busam KJ, Kutzner H, Cho K-H, Aiba S, Bröcker E-B, LeBoit PE, Pinkel D, Bastian BC: Distinct Sets of Genetic Alterations in Melanoma. New England Journal of Medicine 2005, 353:2135-47.
[52] Yip L, Nikiforova MN, Carty SE, Yim JH, Stang MT, Tublin MJ, Lebeau SO, Hodak SP, Ogilvie JB, Nikiforov YE: Optimizing surgical treatment of papillary thyroid carcinoma associated with BRAF mutation. Surgery 2009, 146:1215-23.
[53] Vaughn CP, Zobell SD, Furtado LV, Baker CL, Samowitz WS: Frequency of KRAS, BRAF, and NRAS mutations in colorectal cancer. Genes Chromosomes Cancer 2011, 50:307-12.
[54] Flaherty KT, Robert C, Hersey P, Nathan P, Garbe C, Milhem M, Demidov LV, Hassel JC, Rutkowski P, Mohr P, Dummer R, Trefzer U, Larkin JMG, Utikal J, Dreno B, Nyakas M, Middleton MR, Becker JC, Casey M, Sherman LJ, Wu FS, Ouellet D, Martin A-M, Patel K, Schadendorf D: Improved Survival with MEK Inhibition in BRAF-Mutated Melanoma. New England Journal of Medicine 2012, 367:107-14.
[55] Rubinstein JC, Sznol M, Pavlick AC, Ariyan S, Cheng E, Bacchiocchi A, Kluger HM, Narayan D, Halaban R: Incidence of the V600K mutation among melanoma patients with BRAF mutations, and potential therapeutic response to the specific BRAF inhibitor PLX4032. J Transl Med 2010, 8:67.
[56] Lovly CM, Dahlman KB, Fohn LE, Su Z, Dias-Santagata D, Hicks DJ, Hucks D, Berry E, Terry C, Duke M, Su Y, Sobolik-Delmaire T, Richmond A, Kelley MC, Vnencak-Jones CL, Iafrate AJ, Sosman J, Pao W: Routine Multiplex Mutational Profiling of Melanomas Enables Enrollment in Genotype-Driven Therapeutic Trials. PLoS ONE 2012, 7:e35309.
[57] Klein O, Clements A, Menzies AM, O'Toole S, Kefford RF, Long GV: BRAF inhibitor activity in V600R metastatic melanoma. European Journal of Cancer 2013, 49:1073-9.
[58] Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, Jones CM, Marshall CJ, Springer CJ, Barford D, Marais R: Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell 2004, 116:855-67.
[59] Gray-Schopfer V, Wellbrock C, Marais R: Melanoma biology and new targeted therapy. Nature 2007, 445:851-7.
[60] Cummins DL, Cummins JM, Pantle H, Silverman MA, Leonard AL, Chanmugam A: Cutaneous malignant melanoma. Mayo Clin Proc 2006, 81:500-7.

[61] Hauschild A, Grob JJ, Demidov LV, Jouary T, Gutzmer R, Millward M, Rutkowski P, Blank CU, Miller WH, Jr., Kaempgen E, Martin-Algarra S, Karaszewska B, Mauch C, Chiarion-Sileni V, Martin AM, Swann S, Haney P, Mirakhur B, Guckert ME, Goodman V, Chapman PB: Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. Lancet 2012, 380:358-65.
[62] Falchook GS, Long GV, Kurzrock R, Kim KB, Arkenau TH, Brown MP, Hamid O, Infante JR, Millward M, Pavlick AC, O'Day SJ, Blackman SC, Curtis CM, Lebowitz P, Ma B, Ouellet D, Kefford RF: Dabrafenib in patients with melanoma, untreated brain metastases, and other solid tumours: a phase 1 dose-escalation trial. Lancet 2012, 379:1893-901.
[63] Falchook GS, Lewis KD, Infante JR, Gordon MS, Vogelzang NJ, DeMarini DJ, Sun P, Moy C, Szabo SA, Roadcap LT, Peddareddigari VG, Lebowitz PF, Le NT, Burris HA, 3rd, Messersmith WA, O'Dwyer PJ, Kim KB, Flaherty K, Bendell JC, Gonzalez R, Kurzrock R, Fecher LA: Activity of the oral MEK inhibitor trametinib in patients with advanced melanoma: a phase 1 dose-escalation trial. Lancet Oncol 2012, 13:782-9.
[64] Halaban R, Zhang W, Bacchiocchi A, Cheng E, Parisi F, Ariyan S, Krauthammer M, McCusker JP, Kluger Y, Sznol M: PLX4032, a selective BRAFV600E kinase inhibitor, activates the ERK pathway and enhances cell migration and proliferation of BRAFWT melanoma cells. Pigment Cell & Melanoma Research 2010, 23:190-200.
[65] Ascierto PA, Grimaldi AM, Curti B, Faries MB, Ferrone S, Flaherty K, Fox BA, Gajewski TF, Gershenwald JE, Gogas H, Grossmann K, Hauschild A, Hodi FS, Kefford R, Kirkwood JM, Leachmann S, Maio M, Marais R, Palmieri G, Morton DL, Ribas A, Stroncek DF, Stewart R, Wang E, Mozzillo N, Marincola FM: Future perspectives in melanoma research. Meeting report from the "Melanoma research: a bridge from Naples to the World. Napoli, December 5th-6th 2011". J Transl Med 2012, 10:83.
[66] Fisher R, Larkin J: Vemurafenib: a new treatment for BRAF-V600 mutated advanced melanoma. Cancer Manag Res 2012, 4:243-52.
[67] Ponti G, Pellacani G, Tomasi A, Gelsomino F, Spallanzani A, Depenni R, Al Jalbout S, Simi L, Garagnani L, Borsari S, Conti A, Ruini C, Fontana A, Luppi G: The somatic affairs of BRAF: tailored therapies for advanced malignant melanoma and orphan non-V600E (V600R-M) mutations. Journal of Clinical Pathology 2013, 66:441-5.
[68] Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF: Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Research 1989, 17:2503-16.
[69] Hatzivassiliou G, Song K, Yen I, Brandhuber BJ, Anderson DJ, Alvarado R, Ludlam MJ, Stokoe D, Gloor SL, Vigers G, Morales T, Aliagas I, Liu B, Sideris S, Hoeflich KP, Jaiswal BS, Seshagiri S, Koeppen H, Belvin M, Friedman LS, Malek S: RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. Nature 2010, 464:431-5.
[70] Poulikakos PI, Zhang C, Bollag G, Shokat KM, Rosen N: RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. Nature 2010, 464:427-30.
[71] Pietrantonio F, Petrelli F, Coinu A, Di Bartolomeo M, Borgonovo K, Maggi C, Cabiddu M, Iacovelli R, Bossi I, Lonati V, Ghilardi M, de Braud F, Barni S: Predictive role of BRAF mutations in patients with advanced colorectal cancer receiving cetuximab and panitumumab: A meta-analysis. European Journal of Cancer, 51:587-94.
[72] Fernández-Medarde A, Santos E: Ras in Cancer and Developmental Diseases. Genes & Cancer 2011, 2:344-58.

[73] Malumbres M, Barbacid M: RAS oncogenes: the first 30 years. Nat Rev Cancer 2003, 3:459-65.
[74] Harvey JJ: AN UNIDENTIFIED VIRUS WHICH CAUSES THE RAPID PRODUCTION OF TUMOURS IN MICE. Nature 1964, 204:1104-5.
[75] Kirsten WH, Mayer LA: Morphologic responses to a murine erythroblastosis virus. J Natl Cancer Inst 1967, 39:311-35.
[76] Parada LF, Tabin CJ, Shih C, Weinberg RA: Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus ras gene. Nature 1982, 297:474-8.
[77] Santos E, Tronick SR, Aaronson SA, Pulciani S, Barbacid M: T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB- and Harvey-MSV transforming genes. Nature 1982, 298:343-7.
[78] Hall A, Marshall CJ, Spurr NK, Weiss RA: Identification of transforming gene in two human sarcoma cell lines as a new member of the ras gene family located on chromosome 1. Nature 1983, 303:396-400.
[79] Almoguera C, Shibata D, Forrester K, Martin J, Arnheim N, Perucho M: Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. Cell, 53:549-54.
[80] De Roock W, Claes B, Bernasconi D, De Schutter J, Biesmans B, Fountzilas G, Kalogeras KT, Kotoula V, Papamichael D, Laurent-Puig P, Penault-Llorca F, Rougier P, Vincenzi B, Santini D, Tonini G, Cappuzzo F, Frattini M, Molinari F, Saletti P, De Dosso S, Martini M, Bardelli A, Siena S, Sartore-Bianchi A, Tabernero J, Macarulla T, Di Fiore F, Gangloff AO, Ciardiello F, Pfeiffer P, Qvortrup C, Hansen TP, Van Cutsem E, Piessevaux H, Lambrechts D, Delorenzi M, Tejpar S: Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. Lancet Oncol 2010, 11:753-62.
[81] Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Smits AMM, Bos JL: Genetic Alterations during Colorectal-Tumor Development. New England Journal of Medicine 1988, 319:525-32.
[82] Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, Nickerson E, Auclair D, Li L, Place C, Dicara D, Ramos AH, Lawrence MS, Cibulskis K, Sivachenko A, Voet D, Saksena G, Stransky N, Onofrio RC, Winckler W, Ardlie K, Wagle N, Wargo J, Chong K, Morton DL, Stemke-Hale K, Chen G, Noble M, Meyerson M, Ladbury JE, Davies MA, Gershenwald JE, Wagner SN, Hoon DS, Schadendorf D, Lander ES, Gabriel SB, Getz G, Garraway LA, Chin L: A landscape of driver mutations in melanoma. Cell 2012, 150:251-63.
[83] Jakob JA, Bassett RL, Jr., Ng CS, Curry JL, Joseph RW, Alvarado GC, Rohlfs ML, Richard J, Gershenwald JE, Kim KB, Lazar AJ, Hwu P, Davies MA: NRAS mutation status is an independent prognostic factor in metastatic melanoma. Cancer 2012, 118:4014-23.
[84] Fedorenko IV, Gibney GT, Smalley KS: NRAS mutant melanoma: biological behavior and future strategies for therapeutic management. Oncogene 2013, 32:3009-18.
[85] Bos JL: ras oncogenes in human cancer: a review. Cancer Res 1989, 49:4682-9.
[86] Mills NE, Fishman CL, Rom WN, Dubin N, Jacobson DR: Increased prevalence of K-ras oncogene mutations in lung adenocarcinoma. Cancer Res 1995, 55:1444-7.
[87] Douillard JY, Oliner KS, Siena S, Tabernero J, Burkes R, Barugel M, Humblet Y, Bodoky G, Cunningham D, Jassem J, Rivera F, Kocakova I, Ruff P, Blasinska-Morawiec M, Smakal M, Canon JL, Rother M, Williams R, Rong A, Wiezorek J, Sidhu R, Patterson SD: Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. N Engl J Med 2013, 369:1023-34.

[88] Downward J: Targeting RAS signalling pathways in cancer therapy. Nat Rev Cancer 2003, 3:11-22.
[89] Karnoub AE, Weinberg RA: Ras oncogenes: split personalities. Nature Reviews Molecular Cell Biology 2008, 9:517+.
[90] Cox AD, Der CJ: Ras history: The saga continues. Small GTPases 2010, 1:2-27.
[91] Thomas RK, Baker AC, Debiasi RM, Winckler W, Laframboise T, Lin WM, Wang M, Feng W, Zander T, MacConaill L, Lee JC, Nicoletti R, Hatton C, Goyette M, Girard L, Majmudar K, Ziaugra L, Wong KK, Gabriel S, Beroukhim R, Peyton M, Barretina J, Dutt A, Emery C, Greulich H, Shah K, Sasaki H, Gazdar A, Minna J, Armstrong SA, Mellinghoff IK, Hodi FS, Dranoff G, Mischel PS, Cloughesy TF, Nelson SF, Liau LM, Mertz K, Rubin MA, Moch H, Loda M, Catalona W, Fletcher J, Signoretti S, Kaye F, Anderson KC, Demetri GD, Dummer R, Wagner S, Herlyn M, Sellers WR, Meyerson M, Garraway LA: High-throughput oncogene mutation profiling in human cancer. Nat Genet 2007, 39:347-51.
[92] Siegel RL, Miller KD, Jemal A: Cancer statistics, 2015. CA: A Cancer Journal for Clinicians 2015, 65:5-29.
[93] Saltz LB, Meropol NJ, Loehrer PJ, Sr., Needle MN, Kopit J, Mayer RJ: Phase II trial of cetuximab in patients with refractory colorectal cancer that expresses the epidermal growth factor receptor. J Clin Oncol 2004, 22:1201-8.
[94] Cunningham D, Humblet Y, Siena S, Khayat D, Bleiberg H, Santoro A, Bets D, Mueser M, Harstrick A, Verslype C, Chau I, Van Cutsem E: Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. N Engl J Med 2004, 351:337-45.
[95] Messersmith WA, Hidalgo M: Panitumumab, a Monoclonal Anti–Epidermal Growth Factor Receptor Antibody in Colorectal Cancer: Another One or the One? Clinical Cancer Research 2007, 13:4664-6.
[96] Lievre A, Bachet JB, Boige V, Cayre A, Le Corre D, Buc E, Ychou M, Bouche O, Landi B, Louvet C, Andre T, Bibeau F, Diebold MD, Rougier P, Ducreux M, Tomasic G, Emile JF, Penault-Llorca F, Laurent-Puig P: KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. J Clin Oncol 2008, 26:374-9.
[97] Allegra CJ, Jessup JM, Somerfield MR, Hamilton SR, Hammond EH, Hayes DF, McAllister PK, Morton RF, Schilsky RL: American Society of Clinical Oncology provisional clinical opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. J Clin Oncol 2009, 27:2091-6.
[98] Ciardiello F, Tejpar S, Normanno N, Mercadante D, Teague T, Wohlschlegel B, Van Cutsem E: Uptake of KRAS mutation testing in patients with metastatic colorectal cancer in Europe, Latin America and Asia. Target Oncol 2011, 6:133-45.
[99] Engstrom PF, Arnoletti JP, Benson AB, 3rd, Chen YJ, Choti MA, Cooper HS, Covey A, Dilawari RA, Early DS, Enzinger PC, Fakih MG, Fleshman J, Jr., Fuchs C, Grem JL, Kiel K, Knol JA, Leong LA, Lin E, Mulcahy MF, Rao S, Ryan DP, Saltz L, Shibata D, Skibber JM, Sofocleous C, Thomas J, Venook AP, Willett C: NCCN Clinical Practice Guidelines in Oncology: colon cancer. J Natl Compr Canc Netw 2009, 7:778-831.
[100] Van Cutsem E, Nordlinger B, Cervantes A, Group ObotEGW: Advanced colorectal cancer: ESMO Clinical Practice Guidelines for treatment. Annals of Oncology 2010, 21:v93-v7.
[101] Janakiraman M, Vakiani E, Zeng Z, Pratilas CA, Taylor BS, Chitale D, Halilovic E, Wilson M, Huberman K, Ricarte Filho JC, Persaud Y, Levine DA, Fagin JA, Jhanwar SC, Mariadason JM, Lash A, Ladanyi M, Saltz LB, Heguy A, Paty PB, Solit DB: Genomic and

Biological Characterization of Exon 4 KRAS Mutations in Human Cancer. Cancer Research 2010, 70:5901-11.
[102] Sorich MJ, Wiese MD, Rowland A, Kichenadasse G, McKinnon RA, Karapetis CS: Extended RAS mutations and anti-EGFR monoclonal antibody survival benefit in metastatic colorectal cancer: a meta-analysis of randomized, controlled trials. Ann Oncol 2015, 26:13-21.
[103] Spagnolo F, Ghiorzo P, Orgiano L, Pastorino L, Picasso V, Tornari E, Ottaviano V, Queirolo P: BRAF-mutant melanoma: treatment approaches, resistance mechanisms, and diagnostic strategies. Onco Targets Ther 2015, 8:157-68.
[104] da Rocha Dias S, Salmonson T, van Zwieten-Boot B, Jonsson B, Marchetti S, Schellens JHM, Giuliani R, Pignatti F: The European Medicines Agency review of vemurafenib (Zelboraf®) for the treatment of adult patients with BRAF V600 mutation-positive unresectable or metastatic melanoma: Summary of the scientific assessment of the Committee for Medicinal Products for Human Use. European Journal of Cancer 2013, 49:1654-61.
[105] Gerber DE, Minna JD: ALK Inhibition for Non-Small Cell Lung Cancer: From Discovery to Therapy in Record Time. Cancer Cell 2010, 18:548-51.
[106] Curry JL, Torres-Cabala CA, Tetzlaff MT, Bowman C, Prieto VG: Molecular Platforms Utilized to Detect BRAF V600E Mutation in Melanoma. Seminars in Cutaneous Medicine and Surgery, 31:267-73.
[107] Pinzani P, Santucci C, Mancini I, Simi L, Salvianti F, Pratesi N, Massi D, De Giorgi V, Pazzagli M, Orlando C: BRAFV600E detection in melanoma is highly improved by COLD-PCR. Clinica Chimica Acta 2011, 412:901-5.
[108] Lopez-Rios F, Angulo B, Gomez B, Mair D, Martinez R, Conde E, Shieh F, Vaks J, Langland R, Lawrence HJ, de Castro DG: Comparison of testing methods for the detection of BRAF V600E mutations in malignant melanoma: pre-approval validation study of the companion diagnostic test for vemurafenib. PLoS One 2013, 8:e53733.
[109] Spittle C, Ward MR, Nathanson KL, Gimotty PA, Rappaport E, Brose MS, Medina A, Letrero R, Herlyn M, Edwards RH: Application of a BRAF Pyrosequencing Assay for Mutation Detection and Copy Number Analysis in Malignant Melanoma. The Journal of molecular diagnostics : JMD 2007, 9:464-71.
[110] Lade-Keller J, Romer KM, Guldberg P, Riber-Hansen R, Hansen LL, Steiniche T, Hager H, Kristensen LS: Evaluation of BRAF mutation testing methodologies in formalin-fixed, paraffin-embedded cutaneous melanomas. J Mol Diagn 2013, 15:70-80.
[111] Rechsteiner M, von Teichman A, Ruschoff JH, Fankhauser N, Pestalozzi B, Schraml P, Weber A, Wild P, Zimmermann D, Moch H: KRAS, BRAF, and TP53 deep sequencing for colorectal carcinoma patient diagnostics. J Mol Diagn 2013, 15:299-311.
[112] Siroy AE, Boland GM, Milton DR, Roszik J, Frankian S, Malke J, Haydu L, Prieto VG, Tetzlaff M, Ivan D, Wang WL, Torres-Cabala C, Curry J, Roy-Chowdhuri S, Broaddus R, Rashid A, Stewart J, Gershenwald JE, Amaria RN, Patel SP, Papadopoulos NE, Bedikian A, Hwu WJ, Hwu P, Diab A, Woodman SE, Aldape KD, Luthra R, Patel KP, Shaw KR, Mills GB, Mendelsohn J, Meric-Bernstam F, Kim KB, Routbort MJ, Lazar AJ, Davies MA: Beyond BRAF(V600): Clinical Mutation Panel Testing by Next-Generation Sequencing in Advanced Melanoma. J Invest Dermatol 2015, 135:508-15.
[113] Willmore-Payne C, Holden JA, Tripp S, Layfield LJ: Human malignant melanoma: detection of BRAF- and c-kit-activating mutations by high-resolution amplicon melting analysis. Hum Pathol 2005, 36:486-93.

[114] Pichler M, Balic M, Stadelmeyer E, Ausch C, Wild M, Guelly C, Bauernhofer T, Samonigg H, Hoefler G, Dandachi N: Evaluation of High-Resolution Melting Analysis as a Diagnostic Tool to Detect the BRAF V600E Mutation in Colorectal Tumors. The Journal of Molecular Diagnostics 2009, 11:140-7.
[115] Ney JT, Froehner S, Roesler A, Buettner R, Merkelbach-Bruse S: High-resolution melting analysis as a sensitive prescreening diagnostic tool to detect KRAS , BRAF , PIK3CA , and AKT1 mutations in formalin-fixed, paraffin-embedded tissues. Arch Pathol Lab Med 2012, 136:983-92.
[116] Chen D, Wang Y-Y, Chuai Z-R, Huang J-F, Wang Y-X, Liu K, Zhang L-Q, Yang Z, Shi D-C, Liu Q, Huang Q, Fu W-L: High-Resolution Melting Analysis for accurate detection of BRAF mutations: a systematic review and meta-analysis. Sci Rep 2014, 4.
[117] Tan C, Du X: KRAS mutation testing in metastatic colorectal cancer. World Journal of Gastroenterology : WJG 2012, 18:5171-80.
[118] Allawi HT, SantaLucia J: Thermodynamics of internal C.T mismatches in DNA. Nucleic Acids Research 1998, 26:2694-701.
[119] Whitcombe D, Theaker J, Guy SP, Brown T, Little S: Detection of PCR products using self-probing amplicons and fluorescence. Nat Biotechnol 1999, 17:804-7.
[120] Thelwell N, Millington S, Solinas A, Booth J, Brown T: Mode of action and application of Scorpion primers to mutation detection. Nucleic Acids Res 2000, 28:3752-61.
[121] Gonzalez de Castro D, Angulo B, Gomez B, Mair D, Martinez R, Suarez-Gauthier A, Shieh F, Velez M, Brophy VH, Lawrence HJ, Lopez-Rios F: A comparison of three methods for detecting KRAS mutations in formalin-fixed colorectal cancer specimens. Br J Cancer 2012, 107:345-51.
[122] Sharma A, Zhang G, Aslam S, Yu K, Chee M, Palma JF: Novel Approach for Clinical Validation of the cobas KRAS Mutation Test in Advanced Colorectal Cancer. Molecular Diagnosis & Therapy 2016, 20:231-40.
[123] Hinrichs JWJ, Marja van Blokland WT, Moons MJ, Radersma RD, Radersma-van Loon JH, de Voijs CMA, Rappel SB, Koudijs MJ, Besselink NJM, Willems SM, de Weger RA: Comparison of Next-Generation Sequencing and Mutation-Specific Platforms in Clinical Practice. American Journal of Clinical Pathology 2015, 143:573-8.
[124] Oh JE, An CH, Yoo NJ, Lee SH: Detection of low-level EGFR T790M mutation in lung cancer tissues. Apmis 2011, 119:403-11.
[125] Alqahtani QM, Crowley A, Rapp S, Cushman-Vokoun AM: QIAGEN Therascreen *KRAS* RGQ Assay, QIAGEN *KRAS* Pyro Assay, and Dideoxy Sequencing for Clinical Laboratory Analysis of *KRAS* Mutations in Tumor Specimens. Laboratory Medicine 2016, 47:30.
[126] Vogelstein B, Kinzler KW: Digital PCR. Proc Natl Acad Sci U S A 1999, 96:9236-41.
[127] Morley AA: Digital PCR: A brief history. Biomolecular Detection and Quantification 2014, 1:1-2.
[128] Sanders R, Huggett JF, Bushell CA, Cowen S, Scott DJ, Foy CA: Evaluation of Digital PCR for Absolute DNA Quantification. Analytical Chemistry 2011, 83:6474-84.
[129] Baker SG, Cappuccio A, Potter JD: Research on Early-Stage Carcinogenesis: Are We Approaching Paradigm Instability? Journal of Clinical Oncology 2010, 28:3215-8.
[130] Dingle TC, Sedlak RH, Cook L, Jerome KR: Tolerance of Droplet-Digital PCR vs Real-Time Quantitative PCR to Inhibitory Substances. Clin Chem 2013.

[131] Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, Vessella RL, Tewari M: Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat Methods 2013, 10:1003-5.
[132] Huggett JF, Whale A: Digital PCR as a novel technology and its potential implications for molecular diagnostics. Clin Chem 2013, 59:1691-3.
[133] Azuara D, Ginesta MM, Gausachs M, Rodriguez-Moranta F, Fabregat J, Busquets J, Pelaez N, Boadas J, Galter S, Moreno V, Costa J, de Oca J, Capella G: Nanofluidic digital PCR for KRAS mutation detection and quantification in gastrointestinal cancer. Clin Chem 2012, 58:1332-41.
[134] Pohl G, Shih Ie M: Principle and applications of digital PCR. Expert Rev Mol Diagn 2004, 4:41-7.
[135] White RA, 3rd, Blainey PC, Fan HC, Quake SR: Digital PCR provides sensitive and absolute calibration for high throughput sequencing. BMC Genomics 2009, 10:116.
[136] White RA, 3rd, Quake SR, Curr K: Digital PCR provides absolute quantitation of viral load for an occult RNA virus. J Virol Methods 2012, 179:45-50.
[137] Zhong Q, Bhattacharya S, Kotsopoulos S, Olson J, Taly V, Griffiths AD, Link DR, Larson JW: Multiplex digital PCR: breaking the one target per color barrier of quantitative PCR. Lab on a Chip 2011, 11:2167-74.
[138] Whale AS, Cowen S, Foy CA, Huggett JF: Methods for applying accurate digital PCR analysis on low copy DNA samples. PLoS One 2013, 8:e58177.
[139] Day E, Dear PH, McCaughan F: Digital PCR strategies in the development and analysis of molecular biomarkers for personalized medicine. Methods 2013, 59:101-7.
[140] Jones GM, Busby E, Garson JA, Grant PR, Nastouli E, Devonshire AS, Whale AS: Digital PCR dynamic range is approaching that of real-time quantitative PCR. Biomolecular Detection and Quantification 2016, 10:31-3.
[141] Huggett JF, Cowen S, Foy CA: Considerations for Digital PCR as an Accurate Molecular Diagnostic Tool. Clin Chem 2015, 61:79-88.
[142] Heyries KA, Tropini C, Vaninsberghe M, Doolin C, Petriv OI, Singhal A, Leung K, Hughesman CB, Hansen CL: Megapixel digital PCR. Nat Methods 2011, 8:649-51.
[143] Taly V, Pekin D, Benhaim L, Kotsopoulos SK, Le Corre D, Li X, Atochin I, Link DR, Griffiths AD, Pallier K, Blons H, Bouche O, Landi B, Hutchison JB, Laurent-Puig P: Multiplex picodroplet digital PCR to detect KRAS mutations in circulating DNA from the plasma of colorectal cancer patients. Clin Chem 2013, 59:1722-31.
[144] Reid AL, Freeman JB, Millward M, Ziman M, Gray ES: Detection of BRAF-V600E and V600K in melanoma circulating tumour cells by droplet digital PCR. Clin Biochem 2015, 48:999-1002.
[145] Lamy PJ, Castan F, Lozano N, Montelion C, Audran P, Bibeau F, Roques S, Montels F, Laberenne AC: Next-Generation Genotyping by Digital PCR to Detect and Quantify the BRAF V600E Mutation in Melanoma Biopsies. J Mol Diagn 2015, 17:366-73.
[146] Shih IM, Zhou W, Goodman SN, Lengauer C, Kinzler KW, Vogelstein B: Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. Cancer Res 2001, 61:818-22.
[147] Zhou W, Goodman SN, Galizia G, Lieto E, Ferraraccio F, Pignatelli C, Purdie CA, Piris J, Morris R, Harrison DJ, Paty PB, Culliford A, Romans KE, Montgomery EA, Choti MA, Kinzler KW, Vogelstein B: Counting alleles to predict recurrence of early-stage colorectal cancers. The Lancet 2002, 359:219-25.

[148] Cochran RL, Cravero K, Chu D, Erlanger B, Toro PV, Beaver JA, Zabransky DJ, Wong HY, Cidado J, Croessmann S, Parsons HA, Kim M, Wheelan SJ, Argani P, Park BH: Analysis of BRCA2 loss of heterozygosity in tumor tissue using droplet digital polymerase chain reaction. Hum Pathol 2014, 45:1546-50.
[149] Sanmamed MF, Fernandez-Landazuri S, Rodriguez C, Zarate R, Lozano MD, Zubiri L, Perez-Gracia JL, Martin-Algarra S, Gonzalez A: Quantitative cell-free circulating BRAFV600E mutation analysis by use of droplet digital PCR in the follow-up of patients with melanoma being treated with BRAF inhibitors. Clin Chem 2015, 61:297-304.
[150] Laurent-Puig P, Pekin D, Normand C, S KK, Nizard P, Perez Toralla K, Rowell R, Olson J, Srinivasan P, Le Corre D, Hor T, El Harrak Z, Li X, Link D, Bouche O, Emile JF, Landi B, Boige V, Hutchison B, Taly V: Clinical relevance of KRAS-mutated sub-clones detected with picodroplet digital PCR in advanced colorectal cancer treated with anti-EGFR therapy. Clin Cancer Res 2014.
[151] Gallup JM, Ackermann MR: Addressing fluorogenic real-time qPCR inhibition using the novel custom Excel file system 'FocusField2-6GallupqPCRSet-upTool-001' to attain consistently high fidelity qPCR reactions. Biological Procedures Online 2006, 8:87-152.
[152] Plasterer TN: PRIMERSELECT. Primer and probe design. Methods Mol Biol 1997, 70:291-302.
[153] Rychlik W: OLIGO 7 primer analysis software. Methods Mol Biol 2007, 402:35-60.
[154] Rozen S, Skaletsky H: Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 2000, 132.
[155] Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG: Primer3—new capabilities and interfaces. Nucleic Acids Research 2012, 40:e115-e.
[156] Kutyavin IV, Afonina IA, Mills A, Gorn VV, Lukhtanov EA, Belousov ES, Singer MJ, Walburger DK, Lokhov SG, Gall AA, Dempcy R, Reed MW, Meyer RB, Hedgpeth J: 3′-Minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. Nucleic Acids Research 2000, 28:655-61.
[157] Dominguez PL, Kolodney MS: Wild-type blocking polymerase chain reaction for detection of single nucleotide minority mutations from clinical specimens. Oncogene 2005, 24:6830-4.
[158] Castoldi M, Schmidt S, Benes V, Noerholm M, Kulozik AE, Hentze MW, Muckenthaler MU: A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). Rna 2006, 12:913-20.
[159] Morandi L, de Biase D, Visani M, Cesari V, De Maglio G, Pizzolitto S, Pession A, Tallini G: Allele specific locked nucleic acid quantitative PCR (ASLNAqPCR): an accurate and cost-effective assay to diagnose and quantify KRAS and BRAF mutation. PLoS One 2012, 7:e36084.
[160] Koshkin AA, Singh SK, Nielsen P, Rajwanshi VK, Kumar R, Meldgaard M, Olsen CE, Wengel J: LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. Tetrahedron 1998, 54:3607-30.
[161] Hughesman C, Fakhfakh K, Bidshahri R, Lund HL, Haynes C: A New General Model for Predicting Melting Thermodynamics of Complementary and Mismatched B-Form Duplexes Containing Locked Nucleic Acids: Application to Probe Design for Digital PCR Detection of Somatic Mutations. Biochemistry 2015.

[162] Pinzani P, Salvianti F, Cascella R, Massi D, De Giorgi V, Pazzagli M, Orlando C: Allele specific Taqman-based real-time PCR assay to quantify circulating BRAFV600E mutated DNA in plasma of melanoma patients. Clin Chim Acta 2010, 411:1319-24.
[163] Denys B, El Housni H, Nollet F, Verhasselt B, Philippé J: A Real-Time Polymerase Chain Reaction Assay for Rapid, Sensitive, and Specific Quantification of the JAK2V617F Mutation Using a Locked Nucleic Acid-Modified Oligonucleotide. The Journal of Molecular Diagnostics, 12:512-9.
[164] Tolstrup N, Nielsen PS, Kolberg JG, Frankel AM, Vissing H, Kauppinen S: OligoDesign: Optimal design of LNA (locked nucleic acid) oligonucleotide capture probes for gene expression profiling. Nucleic Acids Res 2003, 31:3758-62.
[165] Schildkraut C: Dependence of the melting temperature of DNA on salt concentration. Biopolymers 1965, 3:195-208.
[166] Melchior WB, Jr., Von Hippel PH: Alteration of the relative stability of dA-dT and dG-dC base pairs in DNA. Proc Natl Acad Sci U S A 1973, 70:298-302.
[167] Breslauer KJ, Frank R, Blocker H, Marky LA: Predicting DNA duplex stability from the base sequence. Proc Natl Acad Sci U S A 1986, 83:3746-50.
[168] SantaLucia J: A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proceedings of the National Academy of Sciences 1998, 95:1460-5.
[169] Allawi HT, SantaLucia J, Jr.: Thermodynamics and NMR of internal G.T mismatches in DNA. Biochemistry 1997, 36:10581-94.
[170] Peyret N, Seneviratne PA, Allawi HT, SantaLucia J, Jr.: Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. Biochemistry 1999, 38:3468-77.
[171] SantaLucia J, Jr., Allawi HT, Seneviratne PA: Improved nearest-neighbor parameters for predicting DNA duplex stability. Biochemistry 1996, 35:3555-62.
[172] Yakovchuk P, Protozanova E, Frank-Kamenetskii MD: Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Research 2006, 34:564-74.
[173] Borer PN, Dengler B, Tinoco I, Jr., Uhlenbeck OC: Stability of ribonucleic acid double-stranded helices. J Mol Biol 1974, 86:843-53.
[174] SantaLucia J, Jr.: A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci U S A 1998, 95:1460-5.
[175] Jones M, Williams J, Gärtner K, Phillips R, Hurst J, Frater J: Low copy target detection by Droplet Digital PCR through application of a novel open access bioinformatic pipeline, ‘definetherain’. Journal of Virological Methods 2014, 202:46-53.
[176] Trypsteen W, Vynck M, De Neve J, Bonczkowski P, Kiselinova M, Malatinkova E, Vervisch K, Thas O, Vandekerckhove L, De Spiegelaere W: ddpcRquant: threshold determination for single channel droplet digital PCR experiments. Anal Bioanal Chem 2015, 407:5827-34.
[177] R Core Team: A Language and Environment for Statistical Computing. . Vienna, Austria: R Foundation for Statistical Computing., 2016.
[178] Attali D, Bidshahri R, Haynes C, Bryan J: ddpcr: an R package and web application for analysis of droplet digital PCR data [version 1; referees: 1 approved], 2016.
[179] Benaglia T CD, Hunter DR, Young D: Mixtools: an R package for analyzing finite mixture models. Journal of Statistical Software 2009, 32:1-29.

[180] Tomasetti C, Marchionni L, Nowak MA, Parmigiani G, Vogelstein B: Only three driver gene mutations are required for the development of lung and colorectal cancers. Proc Natl Acad Sci U S A 2015, 112:118-23.
[181] Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale A-L, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjord JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinsk M, Jager N, Jones DTW, Jones D, Knappskog S, Kool M, Lakhani SR, Lopez-Otin C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt ANJ, Valdes-Mas R, van Buuren MM, van /'t Veer L, Vincent-Salomon A, Waddell N, Yates LR, Australian Pancreatic Cancer Genome I, Consortium IBC, Consortium IM-S, PedBrain I, Zucman-Rossi J, Andrew Futreal P, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR: Signatures of mutational processes in human cancer. Nature 2013, 500:415-21.
[182] Nikiforov YE, Nikiforova MN: Molecular genetics and diagnosis of thyroid cancer. Nat Rev Endocrinol 2011, 7:569-80.
[183] Gazdar AF: Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. Oncogene 2009, 28:S24-S31.
[184] Fakih MG: Metastatic Colorectal Cancer: Current State and Future Directions. Journal of Clinical Oncology 2015, 33:1809-24.
[185] Bolton L, Reiman A, Lucas K, Timms J, Cree IA: KRAS Mutation Analysis by PCR: A Comparison of Two Methods. PLOS ONE 2015, 10:e0115672.
[186] Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S, Vassiliou GS, Bench AJ, Boyd EM, Curtin N, Scott MA, Erber WN, Green AR: Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet 2005, 365:1054-61.
[187] Marchetti A, Martella C, Felicioni L, Barassi F, Salvatore S, Chella A, Camplese PP, Iarussi T, Mucilli F, Mezzetti A, Cuccurullo F, Sacco R, Buttitta F: EGFR mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. J Clin Oncol 2005, 23:857-65.
[188] Loughrey MB, Waring PM, Tan A, Trivett M, Kovalenko S, Beshay V, Young M-A, McArthur G, Boussioutas A, Dobrovic A: Incorporation of somatic BRAF mutation testing into an algorithm for the investigation of hereditary non-polyposis colorectal cancer. Familial Cancer 2007, 6:301-10.
[189] Sensi M, Nicolini G, Petti C, Bersani I, Lozupone F, Molla A, Vegetti C, Nonaka D, Mortarini R, Parmiani G, Fais S, Anichini A: Mutually exclusive NRASQ61R and BRAFV600E mutations at the single-cell level in the same human melanoma. Oncogene 2006, 25:3357-64.
[190] Lasota J, Kowalik A, Wasag B, Wang ZF, Felisiak-Golabek A, Coates T, Kopczynski J, Gozdz S, Miettinen M: Detection of the BRAF V600E mutation in colon carcinoma: critical evaluation of the imunohistochemical approach. Am J Surg Pathol 2014, 38:1235-41.
[191] Chang Y-S, Yeh K-T, Hsu NC, Lin S-H, Chang T-J, Chang J-G: Detection of N-, H-, and KRAS codons 12, 13, and 61 mutations with universal RAS primer multiplex PCR and N-, H-, and KRAS-specific primer extension. Clinical Biochemistry 2010, 43:296-301.
[192] Hovelson DH, McDaniel AS, Cani AK, Johnson B, Rhodes K, Williams PD, Bandla S, Bien G, Choppa P, Hyland F, Gottimukkala R, Liu G, Manivannan M, Schageman J, Ballesteros-

Villagrana E, Grasso CS, Quist MJ, Yadati V, Amin A, Siddiqui J, Betz BL, Knudsen KE, Cooney KA, Feng FY, Roh MH, Nelson PS, Liu C-J, Beer DG, Wyngaard P, Chinnaiyan AM, Sadis S, Rhodes DR, Tomlins SA: Development and Validation of a Scalable Next-Generation Sequencing System for Assessing Relevant Somatic Variants in Solid Tumors. Neoplasia 2015, 17:385-99.
[193] Dumur CI: Available resources and challenges for the clinical annotation of somatic variations. Cancer Cytopathology 2014, 122:730-6.
[194] Xuan J, Yu Y, Qing T, Guo L, Shi L: Next-generation sequencing in the clinic: promises and challenges. Cancer Lett 2013, 340:284-95.
[195] Tsongalis GJ, Peterson JD, de Abreu FB, Tunkey CD, Gallagher TL, Strausbaugh LD, Wells WA, Amos CI: Routine use of the Ion Torrent AmpliSeq Cancer Hotspot Panel for identification of clinically actionable somatic mutations. Clin Chem Lab Med 2014, 52:707-14.
[196] Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B: Detection and quantification of rare mutations with massively parallel sequencing. Proceedings of the National Academy of Sciences 2011, 108:9530-5.
[197] Lou DI, Hussmann JA, McBee RM, Acevedo A, Andino R, Press WH, Sawyer SL: High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. Proceedings of the National Academy of Sciences 2013, 110:19872-7.
[198] Flaherty KT, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, O'Dwyer PJ, Lee RJ, Grippo JF, Nolop K, Chapman PB: Inhibition of mutated, activated BRAF in metastatic melanoma. N Engl J Med 2010, 363:809-19.
[199] Ravnan MC, Matalka MS: Vemurafenib in Patients With BRAF V600E Mutational Positive Advanced Melanoma. Clinical Therapeutics 2012, 34:1474-86.
[200] Luke JJ, Hodi FS: Ipilimumab, vemurafenib, dabrafenib, and trametinib: synergistic competitors in the clinical management of BRAF mutant malignant melanoma. Oncologist 2013, 18:717-25.
[201] Rubinstein J, Sznol M, Pavlick A, Ariyan S, Cheng E, Bacchiocchi A, Kluger H, Narayan D, Halaban R: Incidence of the V600K mutation among melanoma patients with BRAF mutations, and potential therapeutic response to the specific BRAF inhibitor PLX4032. Journal of Translational Medicine 2010, 8:67.
[202] Di Nicolantonio F, Martini M, Molinari F, Sartore-Bianchi A, Arena S, Saletti P, De Dosso S, Mazzucchelli L, Frattini M, Siena S, Bardelli A: Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. J Clin Oncol 2008, 26:5705-12.
[203] Minoo P, Moyer MP, Jass JR: Role of BRAF-V600E in the serrated pathway of colorectal tumourigenesis. J Pathol 2007, 212:124-33.
[204] Prahallad A, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, Beijersbergen RL, Bardelli A, Bernards R: Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. Nature 2012, 483:100-3.
[205] Pietrantonio F, Petrelli F, Coinu A, Di Bartolomeo M, Borgonovo K, Maggi C, Cabiddu M, Iacovelli R, Bossi I, Lonati V, Ghilardi M, de Braud F, Barni S: Predictive role of BRAF mutations in patients with advanced colorectal cancer receiving cetuximab and panitumumab: A meta-analysis. European Journal of Cancer 2015, 51:587-94.
[206] Tufano RP, Teixeira GV, Bishop J, Carson KA, Xing M: BRAF mutation in papillary thyroid cancer and its value in tailoring initial treatment: a systematic review and meta-analysis. Medicine (Baltimore) 2012, 91:274-86.

[207] Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL: Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 2012, 13:134.
[208] Holland PM, Abramson RD, Watson R, Gelfand DH: Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. Proceedings of the National Academy of Sciences 1991, 88:7276-80.
[209] Itabashi T, Maesawa C, Uchiyama M, Higuchi T, Masuda T: Quantitative detection of mutant alleles of the K-ras gene with minor groove binder-conjugated fluorogenic DNA probes. Int J Oncol 2004, 24:687-96.
[210] Denys B, El Housni H, Nollet F, Verhasselt B, Philippé J: A Real-Time Polymerase Chain Reaction Assay for Rapid, Sensitive, and Specific Quantification of the JAK2V617F Mutation Using a Locked Nucleic Acid-Modified Oligonucleotide. The Journal of Molecular Diagnostics 2010, 12:512-9.
[211] Kim RY, Xu H, Myllykangas S, Ji H: Genetic-based biomarkers and next-generation sequencing: the future of personalized care in colorectal cancer. Personalized Medicine 2011, 8:331-45.
[212] Castellanos-Rizaldos E, Paweletz C, Song C, Oxnard GR, Mamon H, Janne PA, Makrigiorgos GM: Enhanced ratio of signals enables digital mutation scanning for rare allele detection. J Mol Diagn 2015, 17:284-92.
[213] Skorokhod A, Helmbold P, Brors B, Schirmacher P, Enk A, Penzel R: Automated Universal BRAF State Detection within the Activation Segment in Skin Metastases by Pyrosequencing-Based Assay U-BRAFV600. PLOS ONE 2013, 8:e59221.
[214] Szankasi P, Reading NS, Vaughn CP, Prchal JT, Bahler DW, Kelley TW: A Quantitative Allele-Specific PCR Test for the BRAF V600E Mutation Using a Single Heterozygous Control Plasmid for Quantitation: A Model for qPCR Testing without Standard Curves. The Journal of Molecular Diagnostics 2013, 15:248-54.
[215] Chen D, Huang J-F, Xia H, Duan G-J, Chuai Z-R, Yang Z, Fu W-L, Huang Q: High-sensitivity PCR method for detecting BRAF V600Emutations in metastatic colorectal cancer using LNA/DNA chimeras to block wild-type alleles. Analytical and Bioanalytical Chemistry 2014, 406:2477-87.
[216] How-Kit A, Lebbé C, Bousard A, Daunay A, Mazaleyrat N, Daviaud C, Mourah S, Tost J: Ultrasensitive detection and identification of BRAF V600 mutations in fresh frozen, FFPE, and plasma samples of melanoma patients by E-ice-COLD-PCR. Analytical and Bioanalytical Chemistry 2014, 406:5513-20.
[217] Ilieva KM, Correa I, Josephs DH, Karagiannis P, Egbuniwe IU, Cafferkey MJ, Spicer JF, Harries M, Nestle FO, Lacy KE, Karagiannis SN: Effects of BRAF mutations and BRAF inhibition on immune responses to melanoma. Mol Cancer Ther 2014, 13:2769-83.
[218] Haggar FA, Boushey RP: Colorectal Cancer Epidemiology: Incidence, Mortality, Survival, and Risk Factors. Clinics in Colon and Rectal Surgery 2009, 22:191-7.
[219] Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D: Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer 2009, 9:489-99.
[220] De Roock W, De Vriendt V, Normanno N, Ciardiello F, Tejpar S: KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer. Lancet Oncol 2011, 12:594-603.

[221] Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, Juan T, Sikorski R, Suggs S, Radinsky R, Patterson SD, Chang DD: Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. J Clin Oncol 2008, 26:1626-34.
[222] Di Fiore F, Blanchard F, Charbonnier F, Le Pessot F, Lamy A, Galais MP, Bastit L, Killian A, Sesboue R, Tuech JJ, Queuniet AM, Paillot B, Sabourin JC, Michot F, Michel P, Frebourg T: Clinical relevance of KRAS mutation detection in metastatic colorectal cancer treated by Cetuximab plus chemotherapy. Br J Cancer 2007, 96:1166-9.
[223] Chretien A-S, Harlé A, Meyer-Lefebvre M, Rouyer M, Husson M, Ramacci C, Harter V, Genin P, Leroux A, Merlin J-L: Optimization of routine KRAS mutation PCR-based testing procedure for rational individualized first-line-targeted therapy selection in metastatic colorectal cancer. Cancer Medicine 2013, 2:11-20.
[224] Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, Murphy KM: Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. J Mol Diagn 2010, 12:425-32.
[225] Dharmasiri U, Njoroge SK, Witek MA, Adebiyi MG, Kamande JW, Hupert ML, Barany F, Soper SA: High-throughput selection, enumeration, electrokinetic manipulation, and molecular profiling of low-abundance circulating tumor cells using a microfluidic system. Anal Chem 2011, 83:2301-9.
[226] Taly V, Pekin D, El Abed A, Laurent-Puig P: Detecting biomarkers with microdroplet technology. Trends Mol Med 2012, 18:405-16.
[227] Watanabe M, Kawaguchi T, Isa S, Ando M, Tamiya A, Kubo A, Saka H, Takeo S, Adachi H, Tagawa T, Kakegawa S, Yamashita M, Kataoka K, Ichinose Y, Takeuchi Y, Sakamoto K, Matsumura A, Koh Y: Ultra-Sensitive Detection of the Pretreatment EGFR T790M Mutation in Non-Small Cell Lung Cancer Patients with an EGFR-Activating Mutation Using Droplet Digital PCR. Clin Cancer Res 2015, 21:3552-60.
[228] Zhu G, Ye X, Dong Z, Lu YC, Sun Y, Liu Y, McCormack R, Gu Y, Liu X: Highly Sensitive Droplet Digital PCR Method for Detection of EGFR-Activating Mutations in Plasma Cell-Free DNA from Patients with Advanced Non-Small Cell Lung Cancer. J Mol Diagn 2015, 17:265-72.
[229] Denis JA, Patroni A, Guillerm E, Pepin D, Benali-Furet N, Wechsler J, Manceau G, Bernard M, Coulet F, Larsen AK, Karoui M, Lacorte JM: Droplet digital PCR of circulating tumor cells from colorectal cancer patients can predict KRAS mutations before surgery. Mol Oncol 2016, 10:1221-31.
[230] Hughesman CB, Lu XJ, Liu KY, Zhu Y, Poh CF, Haynes C: A Robust Protocol for Using Multiplexed Droplet Digital PCR to Quantify Somatic Copy Number Alterations in Clinical Tissue Specimens. PLoS One 2016, 11:e0161274.
[231] Lund HL, Hughesman CB, Fakhfakh K, McNeil K, Clemens S, Hocken K, Pettersson R, Karsan A, Foster LJ, Haynes C: Initial Diagnosis of ALK-Positive Non-Small-Cell Lung Cancer Based on Analysis of ALK Status Utilizing Droplet Digital PCR. Anal Chem 2016, 88:4879-85.
[232] Lund HL, Hughesman CB, McNeil K, Clemens S, Hocken K, Pettersson R, Karsan A, Foster LJ, Haynes C: Initial diagnosis of chronic myelogenous leukemia based on quantification of M-BCR status using droplet digital PCR. Anal Bioanal Chem 2016, 408:1079-94.
[233] Bidshahri R, Attali D, Fakhfakh K, McNeil K, Karsan A, Won JR, Wolber R, Bryan J, Hughesman C, Haynes C: Quantitative Detection and Resolution of BRAF V600 Status in Colorectal Cancer Using Droplet Digital PCR and a Novel Wild-Type Negative Assay. J Mol Diagn 2016, 18:190-204.

[234] Bettington M, Walker N, Clouston A, Brown I, Leggett B, Whitehall V: The serrated pathway to colorectal carcinoma: current concepts and challenges. Histopathology 2013, 62:367-86.
[235] Jass JR: Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology 2007, 50:113-30.
[236] Sahin IH, Kazmi SM, Yorio JT, Bhadkamkar NA, Kee BK, Garrett CR: Rare Though Not Mutually Exclusive: A Report of Three Cases of Concomitant KRAS and BRAF Mutation and a Review of the Literature. J Cancer 2013, 4:320-2.
[237] Fakhfakh K, Hughesman CB, Louise Creagh A, Kao V, Haynes C: Chapter Five - Calorimetric and Spectroscopic Analysis of the Thermal Stability of Short Duplex DNA-Containing Sugar and Base-Modified Nucleotides. Methods in Enzymology. Edited by Andrew LF. Academic Press, 2016. pp. 97-127.
[238] Dienstmann R, Tabernero J: Spectrum of Gene Mutations in Colorectal Cancer: Implications for Treatment. Cancer J 2016, 22:149-55.
[239] Yang S, Farraye FA, Mack C, Posnik O, O'Brien MJ: BRAF and KRAS Mutations in hyperplastic polyps and serrated adenomas of the colorectum: relationship to histology and CpG island methylation status. Am J Surg Pathol 2004, 28:1452-9.
[240] Chin EL, da Silva C, Hegde M: Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. BMC Genet 2013, 14:6.
[241] Clarke CN, Kopetz ES: BRAF mutant colorectal cancer as a distinct subset of colorectal cancer: clinical characteristics, clinical behavior, and response to targeted therapies. J Gastrointest Oncol 2015, 6:660-7.
[242] Unni AM, Lockwood WW, Zejnullahu K, Lee-Lin SQ, Varmus H: Evidence that synthetic lethality underlies the mutual exclusivity of oncogenic KRAS and EGFR mutations in lung adenocarcinoma. Elife 2015, 4:e06907.
[243] Do H, Dobrovic A: Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil-DNA glycosylase. Oncotarget 2012, 3:546-58.
[244] Mouliere F, Robert B, Arnau Peyrotte E, Del Rio M, Ychou M, Molina F, Gongora C, Thierry AR: High fragmentation characterizes tumour-derived circulating DNA. PLoS One 2011, 6:e23418.
[245] Moreira BG, You Y, Behlke MA, Owczarzy R: Effects of fluorescent dyes, quenchers, and dangling ends on DNA duplex stability. Biochem Biophys Res Commun 2005, 327:473-84.
[246] von Ahsen N, Wittwer CT, Schütz E: Oligonucleotide Melting Temperatures under PCR Conditions: Nearest-Neighbor Corrections for Mg²⁺, Deoxynucleotide Triphosphate, and Dimethyl Sulfoxide Concentrations with Comparison to Alternative Empirical Formulas. Clinical Chemistry 2001, 47:1956-61.

Appendices

Appendix A Predicting T_m Values for dsDNA Containing LNAs

A.1. A new NNT model to predict T_m values for complementary and mismatched dual-labeled hydrolysis probes containing LNAs

A version of Appendix A has been published in the journal *Biochemistry*: Hughesman C, Fakhfakh K, Bidshahri R, Lund HL, and Haynes C. A New General Model for Predicting Melting Thermodynamics of Complementary and Mismatched B-Form Duplexes Containing Locked Nucleic Acids: Application to Probe Design for Digital PCR Detection of Somatic Mutations. *Biochemistry* 54: 1338-1352 (2015).

While the NNT models described in Chapter 2 enable prediction of melting thermodynamics for short complementary dsDNA, I desired a model capable of predicting T_m values for complementary and mismatched dual-labeled hydrolysis probes containing LNAs. Regrettably, such a model did not exist. I therefore worked closely with two colleagues, Dr. Curtis Hughesman and Kareem Fakhfakh, to address that need. The result of that collaborative effort was a new model for predicting melting thermodynamics of complementary and mismatched B-form duplexes containing LNAs in one strand that can be applied to the design of allele-specific probes for digital PCR detection of missense mutations. The full details of that model can be found in the 2015 *Biochemistry* paper cited at the start of this Applendix. As this represents collaborative work completed as part of my thesis, I present here only the key features of this new model, and then provide in section A.2 an example of the model's use in designing one of the probes used in my WT-negative ddPCR assays.

The model, which extends the capabilities of the NNT model of Hughesman et al.[33] described in Chapter 2 (equations 2.8 to 2.10) for predicting melting thermodynamics of pure-DNA duplexes, permits prediction of the T_m of a short duplex bearing any pattern of LNA substitutions in one of the strands. It can be applied to a short LNA-modified oligonucleotide duplexed either to its perfect complement or to a template with which it forms a mismatched base pair. For a duplex

composed of non-self-complementary oligonucleotides, the T_m value is predicted from the sequence and chemistry of the strands through the relation:

$$T_m = \frac{\Delta H_{LNA}(T_m) + \Delta\Delta H_{f/q} + \Delta\Delta G_{DNA:MM} + \Delta\Delta G_{LNA:MM}}{\Delta S_{LNA}(T_m) + \Delta\Delta S_{f/q} + \Delta\Delta S_{salt} - R\ ln(K)} \quad \text{A.1}$$

where $\Delta H_{LNA}(T_m)$ and $\Delta S_{LNA}(T_m)$ are the enthalpy and entropy changes, respectively, for denaturation of the LNA containing duplex at T_m. $\Delta\Delta G_{DNA:MM}$ and $\Delta\Delta G_{LNA:MM}$ are the perturbations to duplex stability (in terms of the Gibbs energy) arising from any DNA:DNA or LNA:DNA mismatches, respectively. The effect on duplex stability of any terminal fluorescent reporter dyes and quenchers is captured through their perturbations to the melting enthalpy, $\Delta\Delta H_{f/q}$, and melting entropy, $\Delta\Delta S_{f/q}$, while $\Delta\Delta S_{salt}$ accounts for the dependence of duplex stability on salt composition (i.e. Mg^{2+}, Na^+, and other ion concentrations). The required $\Delta\Delta H_{f/q}$ and $\Delta\Delta S_{f/q}$ parameters are given in Moreira et al,[245] while $\Delta\Delta S_{salt}$ is computed using a modification to the method of von Ahsen et al.[246]

$$\Delta\Delta S_{salt} = 0.874\ n_{bp}\ log_{10}\left(\frac{\left[Na^+_{eq}\right]}{1000}\right) \quad \text{A.2}$$

$$\left[Na^+_{eq}\right] = [monovalent\ cations] + 120\sqrt{[Mg^{2+}] - [dNTPs]} \quad \text{A.3}$$

In equations A.2 and A.3, $\left[Na^+_{eq}\right]$ is the concentration in mM of sodium ion equivalents in the sample, $[Mg^{2+}]$ is the magnesium ion concentration (mM), and [dNTPs] is the mM concentration of deoxyribonucleotide triphosphates in the sample.

$\Delta H_{LNA}(T_m)$ and $\Delta S_{LNA}(T_m)$ are computed as:

$$\Delta H_{LNA}(T_m) = \Delta H^o_{DNA}\left(T_{ref}\right) + \Delta\Delta H^o_{LNA} + \Delta C_p\left(T_m - T_{ref}\right) \quad \text{A.4}$$

$$\Delta S_{LNA}(T_m) = \Delta S^o_{DNA}\left(T_{ref}\right) + \Delta\Delta S^o_{LNA} + \Delta C_p\ ln\left(T_m/T_{ref}\right) \quad \text{A.5}$$

where $\Delta H^o_{DNA}(T_{ref})$, $\Delta S^o_{DNA}(T_{ref})$, and ΔC_p are computed as described in Chapter 2. $\Delta\Delta H^o_{LNA}$ and $\Delta\Delta S^o_{LNA}$ are the incremental enthalpy and entropy changes, respectfully, to duplex stability provided by LNA substitutions, and are computed as:

$$\Delta\Delta H^o_{LNA} = \sum_{i=1}^{4} n_i \, \Delta\Delta H^o_i \qquad \text{A.6}$$

$$\Delta\Delta S^o_{LNA} = \sum_{i=1}^{4} n_i \, \Delta\Delta S^o_i \qquad \text{A.7}$$

In equations A.6 and A.7, n_i is the number of LNA substitutions of type *i*, and $\Delta\Delta H^o_i$ and $\Delta\Delta S^o_i$ are the incremental enthalpy and entropy parameters for each possible LNA–DNA base-pair *i*. based on calorimetry data, all $\Delta\Delta H^o_i$ parameters were found to be negligible in value, indicating that the stabilizing effect of locking a nucleotide is purely entropic in nature. The required $\Delta\Delta S^o_i$ parameters are provided in **Table A.1** along with values of $\Delta\Delta H_{f/q}$ and $\Delta\Delta S_{f/q}$ for the reporter-dye/quencher pairs used in this thesis work.

Equations A.1 to A.7 can be used to predict T_m values for fully complementary duplexes containing any number and pattern of LNAs in one strand with an accuracy of 0.0 ± 1.4 °C (mean error ± standard deviation).

The model can also predict T_m values for short duplexes bearing mismatched DNA–DNA and/or LNA–DNA base pairs. Perturbations to the Gibbs energy of the duplex resulting from DNA–DNA or LNA–DNA base-pair mismatches are accounted for in equation A.1 through $\Delta\Delta G_{DNA:MM}$ and $\Delta\Delta G_{LNA:MM}$, respectively, where:

$$\Delta\Delta G_{DNA:MM} = -\sum_{i=1}^{10} n_i \, \Delta G^o_{DNA_i} + \sum_{j=1}^{40} n_j \, \Delta G_{DNA:MM_i} \qquad \text{A.8}$$

$$\Delta\Delta G_{LNA:MM} = \sum_{i=1}^{12} n_i \, \Delta G_{LNA:MM_i} + \Delta\Delta G_{5'NN-LNA} + \Delta\Delta G_{3'NN-LNA} \qquad \text{A.9}$$

As indicated in equation A.8, $\Delta\Delta G_{DNA:MM}$ is computed by first subtracting the Gibbs energy of each complementary nearest neighbor base pair *i* lost in the formation of a mismatch; the Gibbs energy of the nearest neighbor doublets, *j*, containing the mismatch are then added through use of the $\Delta\Delta G_{DNA:MM_i}$ parameters reported in **Table A.2**. Participation in a mismatch of a more structurally rigid LNA nucleotide generally impacts duplex stability far more punitively. $\Delta\Delta G_{LNA:MM}$ accounts for this added perturbation to the transition energy change, with the $\Delta\Delta G_{LNA:MM_i}$ parameters (**Table A.3**) describing the energetic penalty of each possible LNA–DNA mismatch relative to the corresponding isosequential pure DNA–DNA mismatch. Finally, perturbations to the energy change for denaturation of a duplex where an LNA is immediately to the 5' side or 3' side of a mismatch are then accounted for through the two base independent parameters, $\Delta\Delta G_{5'NN-LNA}$ = – 0.17 kcal mol^{-1} and $\Delta\Delta G_{3'NN-LNA}$ = – 0.55 kcal mol^{-1}, respectively.

Table A.1 Incremental entropy parameters ($\Delta\Delta S_i^o$), as well as $\Delta\Delta H_{f/q}$ and $\Delta\Delta S_{f/q}$ values [taken from Moreira BG, et al. (2005) *Biochem Biophys Res Commun* 327: 473-484**] for the reporter-dye/quencher pairs used in this thesis work.** LNAs shown in **bold** type.

LNA–DNA base pair	$\Delta\Delta S_i^o$ (cal mol^{-1} K^{-1})	
A–T	- 2.3	
T–A	- 3.2	
G–C	- 2.5	
C–G	- 4.8	
Reporter dye or Quenching agent	$\Delta\Delta S_{f/q}$ (cal mol^{-1} K^{-1})	$\Delta\Delta H_{f/q}$ (kcal mol^{-1})
5'-FAM	- 12	- 4
5'-HEX	- 19	- 7
3'-BHQ1	9	4
3'=IABkFQ	- 42	- 14

Table A.2 Incremental nearest-neighbor energy parameter $\Delta G_{DNA:MM_i}$ (@ 37 °C) values (kcal mol^{-1}) for DNA:DNA base-pair mismatches next to complementary Watson-Crick base pairs in 1 M NaCl. Energies reported are for the denaturation reaction (dsDNA → ssDNA)

Nearest Neighbor Base Sequence	X	Y			
		A	C	G	T
GX/CY	A	- 0.17	- 0.81	0.25	PM
	C	- 0.47	- 0.79	PM	- 0.62
	T	0.52	PM	1.11	- 0.08
	G	PM	- 0.98	0.59	- 0.45
CX/GY	A	- 0.43	- 0.75	- 0.03	PM
	C	- 0.79	- 0.70	PM	- 0.62
	G	- 0.11	PM	0.11	0.47
	T	PM	- 0.40	0.32	0.12
AX/TY	A	- 0.61	- 0.88	- 0.14	PM
	C	- 0.77	- 1.33	PM	- 0.64
	G	- 0.02	PM	0.13	- 0.71
	T	PM	- 0.73	- 0.07	- 0.69
TX/AY	A	- 0.69	- 0.92	- 0.42	PM
	C	- 1.33	- 1.05	PM	- 0.97
	G	- 0.74	PM	- 0.44	- 0.43
	T	PM	- 0.75	- 0.34	- 0.67

Parameter values are reported without their associated errors; parameter errors can be found in the original references [see Santa Lucia, Jr. J and Hicks D (2004) *Annual Rev Biophys Biomol Structure* 33: 415-440; and references therein]

Table A.3 Incremental nearest neighbor energy parameter $\Delta G_{LNA:MM_i}$ values (kcal mol^{-1}) for LNA:DNA base pair mismatches next to complementary Watson-Crick base pairs in 1 M NaCl. Energies reported are for the denaturation reaction (dsDNA → ssDNA). LNA shown in **bold**.

Mismatched LNA:DNA Base Pair	$\Delta G_{LNA:MM_i}$ (kcal/mol)
A–A	- 0.50
A–G	- 0.63
G–A	- 1.18
G–G	- 0.82
C–C	- 0.59
C–T	- 0.44
T–C	- 0.35
T–T	- 0.32
A–C	- 0.11
G–T	0.38
C–A	- 0.17
T–G	- 0.28

A.2. Sample Model-Based Calculations for LNA-Substituted Probes

Below, the model described in Appendix A.1 is applied to the prediction of T_m values for the LNA-substituted *BRAF* WT-specific dual-labeled hydrolysis probe (see Chapter 3; the sequence for the probe is 5'-HEX-CGAGATTT**CACTGT**A-BHQ1-3'; see **Table 3.1**) when duplexed to either I) WT *BRAF* or II) *BRAF* V600E1.

I. T_m Prediction for Duplex Formed Between WT BRAF and the BRAF WT-specific probe

As this is a fully complementary duplex, $\Delta\Delta G_{DNA:MM}$ and $\Delta\Delta G_{LNA:MM}$ equal 0 in this calculation.

Step 1: Compute ΔH^o_{DNA} and ΔS^o_{DNA} using equations 2.9 and 2.10 and the parameters provided in **Table 2.1**:

$$\Delta H^o_{DNA} = \sum_{j=1}^{2} m_j \Delta H^{init}_j + \sum_{i=1}^{10} n_i \Delta H^o_{NN_i}$$

ΔH^o_{DNA} = [(-0.1) + (-2.3)] + [10.6 + 3*(8.2) + 2*(7.8) + 2*(7.2) + 2*(7.9) + 2*(8.5) + 2*(8.4)]

ΔH^o_{DNA} = 112.4 kcal mol^{-1}

$$\Delta S^o_{DNA} = \Delta S^{sym} + \sum_{j=1}^{2} m_j \Delta S^{init}_j + \sum_{i=1}^{10} n_i \Delta S^o_{NN_i}$$

ΔS^o_{DNA} = 0 + [2.8 + (-4.1)] + 302.1 = 310.8 cal mol^{-1} K^{-1}

Step 2: Set $\Delta\Delta H^o_{LNA}$ = 0 and compute $\Delta\Delta S^o_{LNA}$ using equation A.7 and the parameters in **Table A.1**

$$\Delta\Delta S^{o}_{LNA} = \sum_{i=1}^{4} n_i \, \Delta\Delta S^{o}_{i}$$

$\Delta\Delta S^{o}_{LNA}$ = – 20.8 cal mol^{-1} K^{-1}

Step 3: Compute ΔC_p

From chapter 2:

$\Delta C_p = n_{bp}\Delta C_p^{bp}$ = 15*(42 cal mol^{-1} K^{-1}) = 630 cal mol^{-1} K^{-1}

Step 4: Compute $\Delta H_{LNA}(T_m)$ in kcal mol^{-1} and $\Delta S_{LNA}(T_m)$ in cal mol^{-1} K^{-1} as a function of T_m using equations A.4 (with $\Delta\Delta H^{o}_{LNA}$ = 0) and A.5 and the values determined in Steps 1 to 3.

$\Delta H_{LNA}(T_m) = \Delta H^{o}_{DNA}(T_{ref}) + \Delta C_p(T_m - T_{ref})$ = 112.4 + 0.630*(T_m – 326.15)

$\Delta S_{LNA}(T_m) = \Delta S^{o}_{DNA}(T_{ref}) + \Delta\Delta S^{o}_{LNA} + \Delta C_p \, ln\,(T_m/T_{ref})$ = 290 + 630 *ln* (T_m/326.15)

Step 5: Compute $\Delta\Delta S_{salt}$ from equations A.2 and A.3 for the chosen PCR solution conditions [C_T = 0.25 µM; 50 mM K^+ and 3 mM Mg^{2+}; dNTP concentration is ignored in this calculation]. In the absence of the salt correction, the model predicts the *Tm* at standard thermodynamic solution conditions [1 M NaCl at pH 7]. Note that all concentrations in these equations are in mM units, with $\Delta\Delta S_{salt}$ then given in cal mol^{-1} K^{-1}.

$[Na^+_{eq}] = [monovalent\ cations] + 120\sqrt{[Mg^{2+}] - [dNTPs]}$ = 257.8 mM

$\Delta\Delta S_{salt} = 0.874\,(n_{bp} - 1)\,log_{10}\left(\frac{[Na^+_{eq}]}{1000}\right)$ = – 7.2 cal mol^{-1} K^{-1}

Step 6: Iteratively compute T_m using equation A.1 with $\Delta\Delta G_{DNA:MM}$ and $\Delta\Delta G_{LNA:MM}$ set equal to 0. Note that T_m appears on both sides of the equation. One must therefore solve for T_m using a root-finding algorithm, such as Newton's method or the bisection method, both of which are widely available, including on many hand-held programmable calculators. Calculating the T_m for the complementary duplex when the probe contains no attached fluorophore or quencher, so that the $\Delta\Delta H_{f/q}$ and $\Delta\Delta S_{f/q}$ terms in equation A.1 are ignored, gives

$$T_m = \frac{\Delta H_{LNA}(T_m)}{\Delta S_{LNA}(T_m) + \Delta\Delta S_{salt} - R\ ln(K)}$$

T_m = 340.75 K = 67.6 °C

For comparison, the experimentally determined (UV melt spectroscopy, with T_m value then corrected to a C_T = 0.25 μM) value of T_m at these conditions is 67.9 °C.

When the $\Delta\Delta H_{f/q}$ and $\Delta\Delta S_{f/q}$ terms correcting for the presence of the 5'-HEX fluorophore and 3' quencher, the predicted T_m is

$$T_m = \frac{\Delta H_{LNA}(T_m) + \Delta\Delta H_{f/q}}{\Delta S_{LNA}(T_m) + \Delta\Delta S_{f/q} + \Delta\Delta S_{salt} - R\ ln(K)}$$

T_m = 340.05 K = 66.9 °C

II. T_m Prediction for Duplex Formed Between BRAF V600E1 *and the BRAF WT-specific probe*

This duplex is the same as in example *I*, except for the replacement of a complementary A–T Watson-Crick base pair at the 5'–10 position with a mismatched A–A base pair (*BRAF* c.1799 T>A mutation). The calculations in example *I* must therefore be extended to account for that mismatch. As a result, $\Delta\Delta G_{DNA:MM}$ and $\Delta\Delta G_{LNA:MM}$ are no longer equal to 0 in this calculation.

The calculation follow that of Example *I* with corrections made to account for the mismatched base-pair.

Step 1: Compute $\Delta\Delta G_{DNA:MM}$ by identifying the two doublets (reading from the 5' end) altered by the mutation, and then applying equation A.8 and the parameters in **Table 2.1** and **Table A.2**

$$\Delta\Delta G_{DNA:MM} = -\sum_{i=1}^{10} n_i \, \Delta G^o_{DNA_i} + \sum_{j=1}^{40} n_j \, \Delta G_{DNA:MM_i}$$

$\Delta\Delta G_{DNA:MM}$ = - (1.45 + 1.44) + ((-0.43) + (-0.17)) = - 3.49 kcal mol^{-1}

Step 2: Define $\Delta\Delta G_{5'NN-LNA}$ and $\Delta\Delta G_{3'NN-LNA}$. In this probe, LNA substitutions are present on the base to the 5' side of the mismatch and on the base to the 3' side of the mismatch, so both parameters must be applied.

$\Delta\Delta G_{5'NN-LNA}$ = -0.17 kcal mol^{-1}

$\Delta\Delta G_{3'NN-LNA}$ = -0.55 kcal mol^{-1}

Step 3: Use those values and the parameter in **Table A.3** for an **A**–A mismatch to compute $\Delta\Delta G_{LNA:MM}$ using equation A.9

$$\Delta\Delta G_{LNA:MM} = \sum_{i=1}^{12} n_i \, \Delta G_{LNA:MM_i} + \Delta\Delta G_{5'NN-LNA} + \Delta\Delta G_{3'NN-LNA}$$

$\Delta\Delta G_{LNA:MM}$ = - 0.5 + ((-0.17) + (-0.55)) = - 1.22 kcal mol^{-1}

Step 4: Compute T_m as before but with energy corrections ($\Delta\Delta G_{DNA:MM}$ and $\Delta\Delta G_{LNA:MM}$) accounting for the mismatched base pair included.

$$T_m = \frac{\Delta H_{LNA}(T_m) + \Delta\Delta H_{f/q} + \Delta\Delta G_{DNA:MM} + \Delta\Delta G_{LNA:MM}}{\Delta S_{LNA}(T_m) + \Delta\Delta S_{f/q} + \Delta\Delta S_{salt} - R\ ln(K)}$$

T_m = 328.2 K = 55.1 °C

Appendix B Algorithm and Software Tool for Analyzing Data from a ddPCR WT-Negative Assay

Further information on the software tool described here in appendix B, including the complete R code and the Shiny web application, has been published in *F1000Research*:

Attali D, Bidshahri R, Haynes C and Bryan J. ddpcr: an R package and web application for analysis of droplet digital PCR data. *F1000Research* 2016, **5**: 1411 (doi: 10.12688/f1000research.9022.1)

B.1 Description of the *ddpcr* Algorithm

The algorithm for automated analysis of output data from a wild-type negative assay consists of five main data-processing steps. To provide context as to how each step is conducted, I describe here the application of the algorithm to the analysis of output data from the *BRAF* WT-negative assay described in chapter 3. The output from that assay is comprised of three clusters: a FAM^-/HEX^- cluster of empty droplets, a FAM^+/HEX^+ cluster (droplets containing WT *BRAF*), and a FAM^+/HEX^- cluster (WT negative droplets). The software automatically gates droplets into unique clusters using kernel density estimation and Gaussian mixture models applied to the droplet fluorescence amplitudes. The five main steps of the algorithm are as follows:

Step 1: Identify Failed Wells

Any wells that clearly failed the ddPCR run are first identified by applying four quality control metrics. The first check ensures that the total number of droplets in a well exceeds a specific threshold value of 5000 droplets. The instrument manufacturer (BioRad Inc.) claims that wells should have 20,000 readable droplets, but in practice we usually read 12,000 – 17,000 droplets per well. Lower values are observed on occasion, and any well with less than 5000 read droplets is considered a failure. The other three metrics evaluate the expected droplet clusters (empty, mutant, wildtype) and their quality by fitting a two-component Gaussian mixture model to the FAM signals for all read droplets. The mean center and standard deviation of each component are computed and recorded. In a typical well, the distribution with the lower center will capture the empty droplets, while the higher distribution captures all FAM+ (template-containing)

droplets. The second metric involves looking for acceptable segregation of empty droplets from the FAM+ cluster by measuring the difference between the centers of the two distributions and ensuring it is acceptably large. The third metric evaluates the relative droplet frequency within the lower (empty) population to ensure the fraction of empty droplets is above a default threshold of 0.3. Failure to meet this criterion indicates a lack of a defined empty cluster. Similarly, the fourth quality control check ensures the empty droplet frequency is below a default fraction of 0.99, as having too many empty droplets is a sign that there is not enough amplifiable templates in the well. Any well that does not meet all four criteria is deemed a failed run, and such wells are removed from further analysis.

Step 2: Identify outlier droplets

For each signal channel (FAM and HEX), the *1%* percent of droplets having the highest signal value are identified. An outlier threshold is then defined as $Q3 + k*IQR$ (where $Q3$ is the 3rd quartile, IQR is the interquartile range, and the percentile k is set as a default to 5 (*i.e.* 5%)). Any droplet in any well that has a FAM signal exceeding the FAM outlier threshold or a HEX value exceeding the HEX outlier threshold is considered an outlier droplet, and removed from further analysis.

Step 3: Identify and eliminate empty droplets

Droplets with very low fluorescent signal in both channels are considered empty and are removed from further analysis as these droplets do not contain any amplifiable template. The removal of empty droplets is beneficial for two reasons. First, it greatly reduces the dimension of the data, which consequently allows for faster computations on the remaining droplets. Secondly, removing the empty droplets also serves to eliminate any bias in data analysis that might occur due to the large number of empty droplets. No useful information is lost, as all the template-containing droplets are retained.

Empty droplets are identified by first fitting a two-component Gaussian mixture model to the FAM signals of all droplets in a well. Only the FAM signal is used because all empty droplets emit no FAM signal while all template-containing droplets emit FAM signal. Note that the same argument cannot be applied to the HEX signal because both empty droplets and mutant-positive

droplets are HEX^-, so while FAM can be used to identify empty droplets, HEX cannot. The Gaussian distribution with the lower mean is assumed to be modeling the empty droplets, and generally has a small standard deviation since the empty droplets tend to densely cluster. A FAM threshold for empty droplets is then calculated under the assumption that the FAM value of empty droplets can be roughly modeled by a normal distribution. Specifically, a threshold is defined as the $mean + k*\sigma$ (where σ is the standard deviation) of the lower (empty) distribution. A default percentile of $k = 7$ is used but can be adjusted as needed to improve model fit. Any droplets in the well with a FAM value lower than the threshold are deemed empty and not considered in further analysis. Figure B1 shows an illustration of this step.

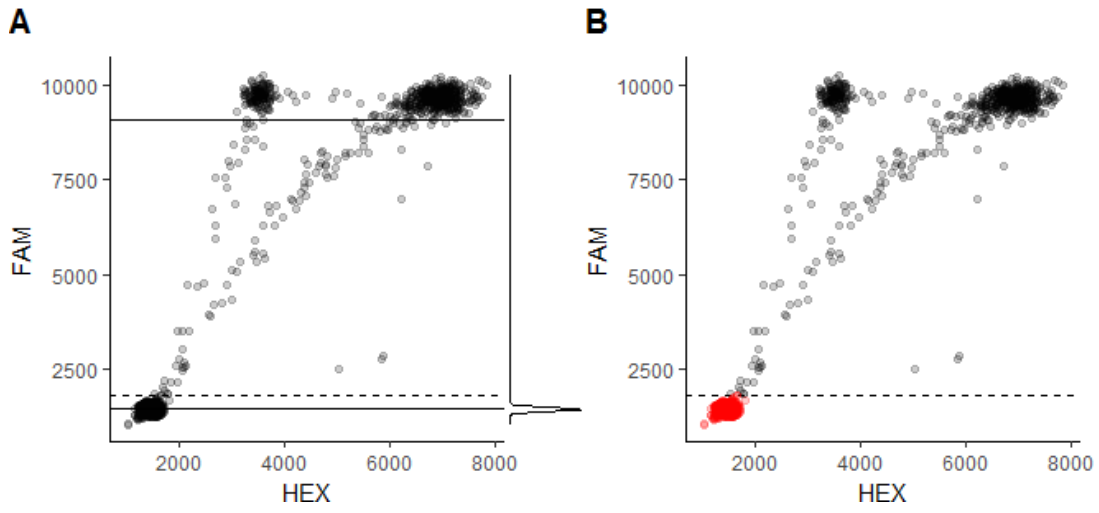

Figure B1. A) Typical output data from the ddPCR-based *BRAF* WT-negative assay and a marginal density plot of the FAM values for all FAM^- droplets. The lower solid line is the mean of the lower Gaussian distribution fitted to the FAM values, while the dotted line is the $mean + 7*\sigma$ for that lower distribution. B) Droplets falling below the $mean + 7*\sigma$ line (dashed line) for lower distribution are identified as empty droplets (marked in red) and discarded.

Step 4: Identify and gate template-positive droplets

The remaining droplets are automatically classified according to their FAM and HEX fluorescence amplitudes. For instance, non-empty droplets recorded in the *BRAF* WT-negative assay are classified as either *mutant*, *wild-type*, or *rain*. Rain droplets are droplets that are not empty, but that have fluorescence amplitudes that fall outside the boundaries of a defined

positive-droplet cluster. The exact cause of rain droplets is not well understood, but they are omitted from our analysis due to their ambiguity. The remaining droplets, which contain high quality amplifiable template, are defined as the *filled* droplet population; every filled droplet is then assigned to a cluster through use of a droplet gating algorithm that may be broken down into its sub-steps.

Sub-step 1: Identify rain droplets: Similar to the way empty droplets are identified in each well, rain droplets are identified using a FAM threshold. A two-component Gaussian mixture model is fitted to the FAM signals of all non-empty droplets in each well. The population of filled droplets have similar FAM fluorescence amplitudes, while the rain droplets have a wide range of FAM signals that are collectively of lower amplitude. Therefore, the Gaussian distribution with the higher mean captures the filled droplet population. A FAM threshold is calculated as the *mean* - $k*\sigma$ (k = default is 3) of the higher (filled) distribution. Any droplets below this value are assigned as rain and removed from further analysis.

Sub-step 2: Identify mutant versus wildtype droplets: In a *BRAF* WT-negative assay, all remaining droplets at this stage contain either a mutant or wildtype template. These droplets all have similar FAM values, while their HEX values differ and may be used to further discriminate and assign droplets. While employing a simple clustering method, such as a k-means, can in principle be used to cluster the two groups of droplets, that approach did not yield good results on many datasets. We therefore defined the two template-positive clusters by first computing the kernel density estimate of the distribution of HEX values. The local minimum of the density estimate is then used as the gate separating mutant-positive and wild-type positive droplets. When computing the kernel density estimate, the degree of smoothing is an important variable that can affect the densities of the resulting clusters. The algorithm therefore attempts to find the optimal smoothing bandwidth by iteratively increasing the smoothing parameter and using heuristics to assess the quality of the density estimations.

At an ideal smoothing bandwidth, the density estimator of the HEX values will have two local maxima with one local minimum between them. The two local maxima capture the centers of the mutant-positive and wild-type-positive clusters, while the local minimum defines the border that

separates those clusters. If the smoothing bandwidth is too high, the mutant cluster will not be identified, while too low of a bandwidth will falsely identify most droplets as mutant. An iterative process is used to find the optimal smoothing parameter by multiplying the default bandwidth by an increasingly larger value of *k*.

Initially, *k* is set to a specific value (the default (starting) *k_min* is 4) and the kernel density is estimated by adjusting the default smoothing bandwidth by *k*. If there is only one local maximum in the density curve, then all filled droplets are taken as having wild-type template. If two local maxima are identified, then a gate is created at the local minimum between the two maxima, with the droplets within the distribution of lower mean HEX signal assigned as the mutant cluster and droplets within the high-HEX distribution assigned as wild-type. If there are more than two local maxima, *k* is increased in small increments until a final value of *k_max* (*k_max* = default is 20). At each iteration, the number of local maxima is recorded until there are only two, at which point the gate is defined as described. If no values of *k* produce acceptable results, the gate is still defined using the left-most local minimum, and the well flagged as inaccurate in order for the user to review it.

Figure B2 provides an example of this process, including the identification of rain droplets, and the subsequent gating of the filled droplet populations.

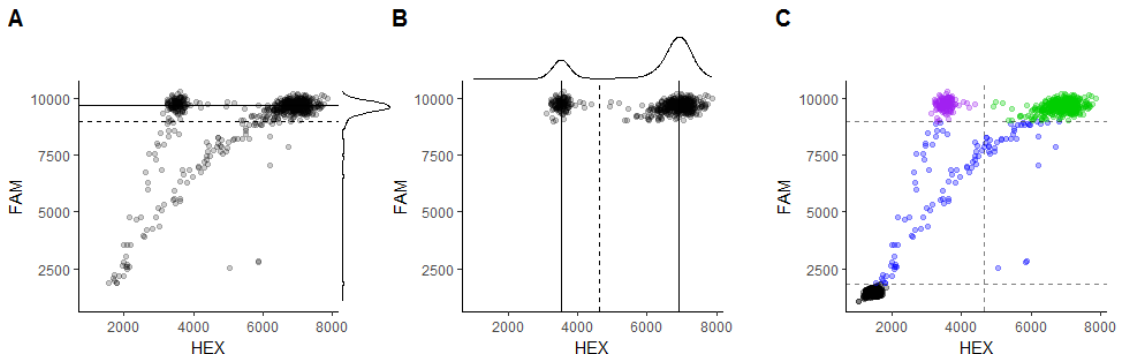

Figure B2. **Treatment of empty-droplet excluded raw ddPCR data to identify and remove rain droplets, and then gate the filled droplet populations.** A) Marginal density plot of FAM amplitude values for a data set in which the empty droplet population has been removed (see Figure B1 for that process). The solid line is the mean of the Gaussian distribution fitted to the high FAM amplitude droplet population, and the dotted line is the *mean - 3*σ* of that distribution. This dotted line defines the threshold, with droplets below it assigned as rain. B) Marginal density plot of HEX amplitude values for a data set in which the empty droplet population and the rain have been removed. The two vertical solid lines define the two local maxima in the density kernel, while the dotted lines is the value of the local minimum between the two maxima. The dotted line is used as a threshold to distinguish between the two filled droplet populations (mutant droplet population to the left and wildtype droplet population to the right). C) Complete droplet population with each droplet assigned, along with the thresholds used to determine those assignments. The outlier droplets are excluded from this visualization. The black droplets are classified as empty, blue droplets as rain, green droplets as containing WT *BRAF* template, and purple droplets as containing mutant *BRAF* V600.

Sub-step 3: Computation of the mutant frequency: Once all filled droplets have been assigned to a cluster, the *mutant frequency* (*MF*) within the sample is calculated as *MF* = (*number of mutant droplets / total number of filled droplets*) * *100%*. Using the *MF* value, it is then possible to classify each sample as harboring wildtype or mutant *BRAF*. In this algorithm, a sample classified as MT is defined as having a *MF* that is higher than *p*% (where *p*% is defined by the operator based on the limit of quantitation (*LOQ*) of the assay) by an amount defined by the confidence interval (*CI*). The binomial test is carried out for this task, with the null hypothesis being that the real mutant frequency is at most *p*%.

A well containing 500 filled droplets that include 7 mutant droplets would therefore be classified as containing a wildtype sample, as the *p*-value is higher than 0.01 despite a mutant frequency of 1.4% (*7 / 500 * 100*), while a well with 5000 filled droplets that includes 70 mutant droplets has

an identical *MF*, but is classified as a mutant sample because the *p*-value is then statistically significant (< 0.01).

Sub-step 4: Proper gating of a wildtype sample: Mutant samples display clearly defined clusters of mutant and wildtype droplets, making gating of clusters relatively straightforward. On the other hand, a wildtype sample may present small number of droplets with HEX signals significantly different from the cluster mean. Assignment of those droplets can be problematic. However, it is possible to leverage the data acquired from mutant-positive samples to accurately assign those droplets as containing either mutant or wildtype template. That required reference data is provided by results from (at least) $n = 4$ wells containing a mutant-positive sample.

Those data are collective used to define the average distribution of the mutant-positive clusters relative to that of the wildtype clusters. To conduct this analysis, a *mutant-to-wildtype ratio* is calculated for every mutant-positive sample by comparing the HEX value of the right-most mutant-positive droplet to the median HEX value of the wildtype-positive droplets (*mutant-to-wildtype ratio = max(mutant) / median(wildtype)*). After calculating this ratio for all available mutant-positive-samples, a "consensus" ratio is computed by choosing the *Q3* percentile. The median HEX value of filled droplets in each indeterminate sample well is then multiplied by the consensus *mutant-to-wildtype ratio*, and the resulting value used as the new border between mutant and wildtype droplets.

B.2 Implementation

As reported in chapter 3, the automated *ddpcr* software was applied to output data of the *BRAF* WT-negative assay when applied to FFPE specimens from a cohort of colorectal cancer (CRC) patients. Through its droplet gating algorithm, *ddpcr* accurately identified droplet clusters and the total number of filled droplets within each sample to provide the information needed to compute the frequency of mutated *BRAF* genes (Figure B3).

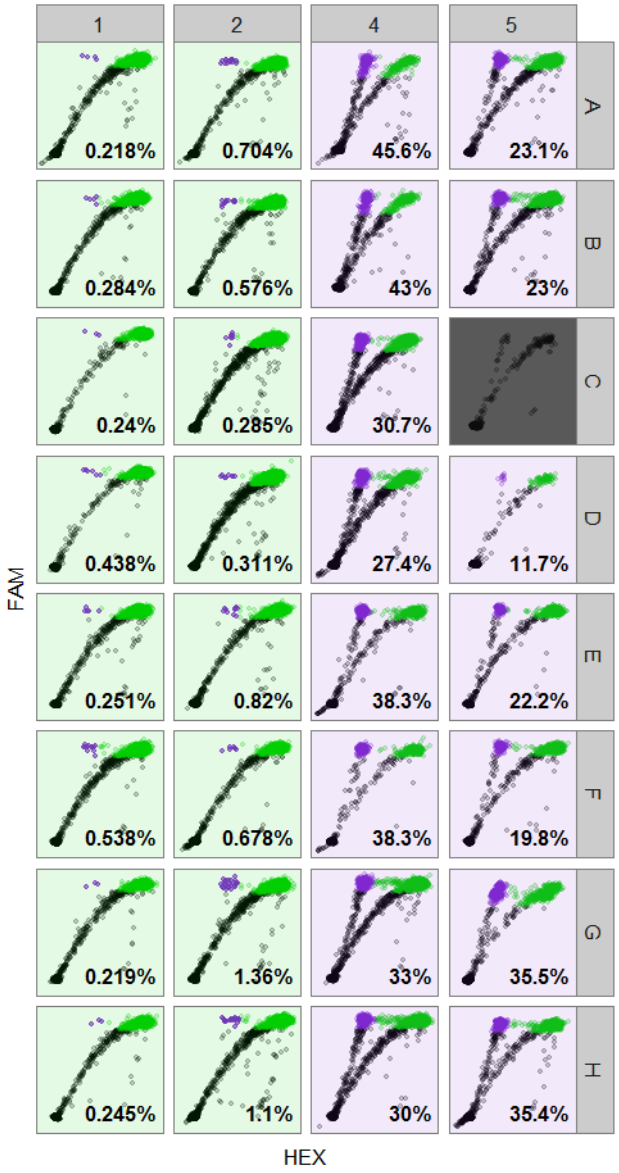

Figure B3. **Results from applying the ddpcr software to DNA samples isolated from a cohort of colorectal cancer (CRC) patients.** Wells highlighted in green are those containing samples classified as containing WT *BRAF*. Wells highlighted in purple are those containing samples classified as containing mutant *BRAF* V600, with the recorded mutation frequency shown. The grey well (C5) represents a failed run.

To assess the accuracy of results obtained from the *ddpcr* software, we compared *BRAF*-V600 mutation frequencies determined by it with results obtained from two independent methods: 1) manual analysis of each ddPCR data set by an experienced operator, and 2) a certified immunohistochemical staining assay against *BRAF* V600E. V600 mutation frequencies

computed from the automated *ddpcr* software were within 3% of those obtained by manual analysis of the ddPCR data by an experienced operator. Likewise, for all samples testing positive for either WT *BRAF* or *BRAF* V600E using *ddpcr*, the *BRAF*-V600 status corresponded to that provided by a certified pathologist using an immunohistochemical staining assay. We therefore obtained agreement between the pathologist's binary classification of *BRAF* status and that determined using *ddpcr*.