A NOVEL PLATFORM FOR CREATING DIGITAL PCR ASSAYS TO DETECT GENETIC TRANSLOCATIONS AND ITS APPLICATION TO THE INITIAL DIAGNOSIS OF CANCER

by

Helen Louise Lund

M.Eng., 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

(Genome Science and Technology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2016

Abstract

Chromosomal translocations can cause cancer, often through the formation of fusion genes that code for an unnatural tyrosine kinase that promotes constitutive activation of a signaling pathway controlling cell proliferation and differentiation. For example, the diagnostic hallmark of chronic myelogenous leukemia (CML) is an oncogene fusion formed from a reciprocal translocation (t(9;22)(q34.1;q11.2)) between chromosomes 9 and 22 that results in an altered chromosome 22q known as the Philadelphia chromosome. Approximately 95% of all CML patients harbor the gene fusion, *BCR-ABL*, which is formed via a double stranded break (DSB) within both the Abelson oncogene 1 (*ABL*) on chromosome 9q, which codes for a non-receptor tyrosine kinase (ABL), and the breakpoint cluster region gene (*BCR*) on chromosome 22q. *BCR-ABL* encodes a constitutively active tyrosine kinase BCR-ABL responsible for the uncontrolled proliferation associated with chronic myelogenous leukemia. The identification of these translocation events and/or associated fusion genes in clinical samples is critical to ensure the appropriate treatment for patients where the drug and related course of therapy target an activated fusion kinase. Clinical detection of complex chromosomal rearrangements is often conducted using fluorescent *in situ* hybridization (FISH). The FISH analysis, though effective, offers relatively poor sensitivity while being expensive, time-consuming and technically challenging to perform.

Here we have developed and validated a new general platform for creating assays against complex chromosomal rearrangements, including both reciprocal and non-reciprocal translocations. It utilizes droplet digital PCR (ddPCR) technology in lieu of FISH to quantify the rearrangement of proto-oncogenes that undergo rearrangement as part of the translocation event. The platform is applied to the creation of two new assays of potential clinical use in cancer diagnostics or theranostics. The first provides a reliable and sensitive measure of DSBs within the major breakpoint region of *BCR* (M-*BCR*), permitting initial diagnosis of CML through unequivocal detection of the *BCR-ABL* fusion gene to a frequency of 0.25%. The second provides for the highly sensitive detection of DSBs in the anaplastic lymphoma kinase (*ALK*) gene that result in a non-reciprocal (inversion) translocation (inv(2)(p21;p23)) associated with an *ALK*-positive non-small cell lung cancer (NSCLC).

Preface

A version of Chapter 2 has been accepted for publication in *Analytical and Bioanalytical Chemistry* as: Lund, Louise H et al., (2015), Initial Diagnosis of Chronic Myelogenous Leukemia Based on Quantification of *BCR* Status Using Droplet Digital PCR.

I performed all of the research with insights provided by Drs. Curtis Hughesman and Charles Haynes, and in collaboration with BC Cancer Agency cytogenetics lab personnel, who performed all of the cytogenetic testing. In addition, I drafted the initial manuscript, with further contributions to it made by Dr. Charles Haynes. Valuable input from Dr Leonard Foster and Dr Aly Karsan was also received prior to submission for publication.

A version of Chapter 3 from this thesis will be submitted for publication as Lund, Louise H et al., Initial Diagnosis of *ALK*-positive NSCLC Based on Quantification of *ALK* Status Using Droplet Digital PCR.

I performed all of the research with insights provided by Drs. Curtis Hughesman and Charles Haynes, and in collaboration with BC Cancer Agency cytogenetics lab personnel, who performed all of the cytogenetic testing. In addition, I drafted the initial manuscript, with further contributions to it made by Dr. Charles Haynes with valuable input from Dr Leonard Foster.

Table of Contents

List of Tables

List of Figures

List of Symbols

2-D	two-dimensional
3IABkFQ	3’ Iowa Black® FQ
9q	chromosome 9 long arm q
22q	chromosome 22 long arm q
a	adenine
A^+	Alexa Fluor® 488
b	biological double stranded break – translocation
$\bar{b}$	no biological double stranded break – translocation
bp	base pair
c	cytosine
ca	circa
Cq	quantitation cycle
$C_{template}$	average concentration of template
C_T	total strand concentration
e	exon
F^+	6-carboxyfluorescein
F1174C	amino acid substitution position 1174 –phenylalanine to cysteine
FP	forward primer
g	guanine
g	gravitational force
G1269A	amino acid substitution position 1269 –glycine to alanine
G1202R	amino acid substitution position 1202R –glycine to arginine
h	hours
H	hydrogen
H^+	hexachloro-fluorescein
i	intron
inv(2)(p21;p23)	chromosome 2 inversion translocation on the short arm p region (2) band (1) and region (2) band (3)
I1151T/N/S	amino acid substitution position 1171 –isoleucine to threonine/asparagine/serine

I1171T	amino acid substitution position 1171 –isoleucine to threonine
θ	fraction of total droplets having a positive end-point fluorescence
kbp	kilobase pair
KW2449	multi kinase inhibitor
l	loss of Hex signal – disruption
$\bar{l}$	no loss of Hex signal – disruption
L1196M	amino acid substitution position 1196 –leucine to methionine
mAU/ml	milliactivity unit per millilitre
Mbp	mega base pair
mg	milligram
mg/ml	milligrams per millilitre
min	minutes
mL	millilitre
mm	millimeters
mM	millimolar
n	cycle number
n	copies of template
ng	nanograms
nL	nanolitre
nm	nanometer
nM	nanomolar
OH	hydroxide
p230	230 kilo dalton protein
pL	picolitre
p(n)	probability that a given droplet contains n copies of template
RP	reverse primer
σ	standard deviation
s	mechanical double stranded break – shear
$\bar{s}$	no mechanical double stranded break – shear
S1206Y	amino acid substitution position 1206 –serine to tyrosine
t	thymine

T315I	amino acid substitution position 315 – threonine to isoleucine
t(9;22)(q34.1;q11.2)	chromosome 9 and chromosome 22 translocation breaks in region (3) band (4) sub-band (1) on long arm q of chromosome 9 and region (1) band (1) sub-band (2) of long arm q of chromosome 22
Ta	annealing temperature
λ_{em}	emission wavelength, m
λ_{ex}	excitation wavelength, m
µL	microliter
µM	micromolar
μm^2	micrometer
μm^2	micrometer squared
$V_{droplet}$	average droplet volume
ZEN™	IDT internal quencher

List of Abbreviations

ABL	abelson oncogene 1 protein
ABL	abelson oncogene 1
AKT	protein kinase B
Alexa	Alexa Fluor® 488
ALK	anaplastic lymphoma kinase protein
ALK	anaplastic lymphoma kinase gene
ALK-EML	anaplastic lymphoma kinase – echinoderm microtubule associated protein – like 4 fusion protein
ALK-EML4	anaplastic lymphoma kinase – echinoderm microtubule associated – like 4 fusion gene
ALK – positive NSCLC	anaplastic lymphoma kinase – positive non-small cell lung cancer
ALL	acute lymphoblastic leukemia
AML	acute myeloid leukemia
ASS	argininosaccinate synthase 1
ATP	adenosine triphosphate
BCL2	B-cell lymphoma 2 protein
BCL6	B-cell lymphoma 6 protein
BCR	breakpoint cluster region protein
BCR	breakpoint cluster region gene
BCR-ABL	breakpoint cluster region - abelson oncogene 1 fusion protein
BCR-ABL	breakpoint cluster region - abelson oncogene 1 fusion gene
BHQ1	black hole quencher® 1
BRAF	B-Raf proto-oncogeneserine/threonine kinase gene
CAP	College of American Pathologists
CBC	complete blood count
cdPCR	chip digital polymerase chain reaction
CI	confidence interval
cKIT	tyrosine protein kinase KIT
CML	chronic myelogenous leukemia
CPD	copies per droplet

CV	coefficient of variance
DAPI	4’,6’-diamidino-2-phenylindole
DDR	deoxyribonucleic acid damage response
D-FISH	dual-labeled fluorescent *in-situ* hybridization
ddPCR	droplet digital polymerase chain reaction
dPCR	digital polymerse chain reaction
dUTP	deoxyuridine-triphosphatase
DSB	double stranded break
EGFR	epidermal growth factor receptor protein
EGFR	epidermal growth factor receptor gene
ELN	European Leukemia NET
EML4	echinoderm microtubule associated protein – like 4
EML4	echinoderm microtubule associated – like 4 gene
FAM	6-carboxyfluorescein
FDA	US Food and Drug Administration
FFPE	formalin-fixed paraffin-embedded
FISH	fluorescent *in-situ* hybridization
FU	fluorescence units
GAP	GTPase-activating protein
gDNA	genomic deoxyribonucleic acid
GLI	GLI family zinc finger protein
HEX	hexachloro-fluorescein
Hh	hedgehog protein
HRR	homologous recombination repair
HSCs	hematopoietic stem cells
HSP 90	heat shock protein 90
IHC	immunohistochemistry
KIF5B	kinesin family member 5B
KRAS	kirsten rat sarcoma viral oncogene homolog
LOB	limit of blank
LOD	limit of detection

LOH	loss of heterozygosity
MAP	mitogen-activated protein
M-*BCR*	major breakpoint region on the breakpoint cluster region gene
m-*BCR*	minor breakpoint region on the breakpoint cluster region gene
MCI-1	mantel cell lymphoma 1 protein
MMEJ	microhomology-mediated end-joining
MRD	minimal residual disease
mRNA	messenger ribonucleic acid
mTOR	mechanistic target of rapamycin
NCCN	National Comprehensive Cancer Network
ND	NanoDrop
NGS	next generation sequencing
NHEJ	non-homologous end-joining
NK	natural killer
No.	number
NSCLC	non-small cell lung cancer
NTC	no template control
PCR	polymerase chain reaction
PhC	Philadelphia chromosome
PI3K	phosphoinositide 3-kinase
PML	promyelocytic leukemia protein
PPZa	protein phosphatase
PTEN	phosphatase and tensin homolog
qPCR	quantitative polymerase chain reaction
RNA	ribonucleic acid
RPMI	Roswell Park memorial institute media
RT	reverse transcriptase
RT-qPCR	reverse transcriptase quantitative polymerase chain reaction
SCLC	small cell lung cancer
SD	standard deviation
SH2	SRC homology 2

SMO	smoothened G protein-coupled receptor
SRC	SRC proto-oncogene, non receptor tyrosine kinase
TE	tris ethylene-diamine-tetra acetic acid
TFG	*TRK*-fused gene
TKI	tyrosine kinase inhibitor
μ-BCR	micro breakpoint region on the breakpoint cluster region gene
WNT/β catenin	wingless integration β catenin

Acknowledgements

I would like to acknowledge my supervisors Dr Charles Haynes and Dr Leonard Foster. Thank you Chip for taking me in and helping me develop my skills as a scientist and researcher. Thanks to Leonard for your continued and relentless support through my whole PhD journey. Thanks also to Dr Carl Hansen who has pushed me the whole journey to be a better scientist and Dr Catherine Poh for her valuable input into this thesis.

I couldn't have got this far without the support of the Haynes lab members Roza Bidshari, Kareem Fakhfakh, Dr Eric Ouellet and Dr Curtis Hughesman and also the Foster Lab members particularly Amanda Van Haga and Jenny Moon – Thanks.

A special thanks to Dr Aly Karsan and his team, Shahira Clemens, Ryan Pettersson and Kimberley Hocken for all of the cytogenetic analysis. I would also like to thank Kelly McNeil for his expert contributions along the way.

Dedication

To my family, Alan, Charlotte, Henry and Arwen!

Chapter 1: Introduction

1.1 Thesis Overview

Roukos describes translocations as

"Beautifully complex, yet threateningly dangerous"[1].

Translocations are genetic abnormalities that occur due to chromosomal rearrangement [2]. They are classified into those that are either reciprocal or non-reciprocal. Examples of non-reciprocal translocations include breakage within a chromosome that results in either the loss of genetic material or, if a pair of double stranded breaks (DSBs) occurs simultaneously within a given chromosome, inversion and recombination of the released genomic fragment back into the same chromosome. In contrast, a reciprocal translocation results in the exchange of genomic material between two chromosomes (Figure 1-1) [1].

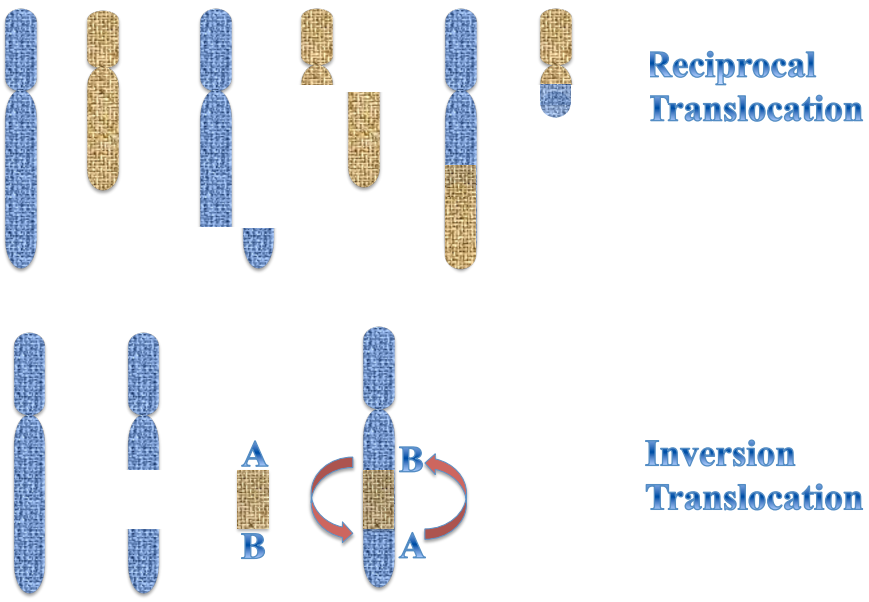


Figure 1-1. Schema for the formation of reciprocal and inversion translocations.
Either class of translocation is capable of forming fusion genes.

Translocations are known drivers of oncogenesis, and collectively they are attributed to 20% of all cancer-related mortalities [3]. The mechanisms resulting in formation of

translocations are not fully understood, though recent research into DSBs and DNA damage response (DDR) has shed some light on this very complex phenomena [1].

The most studied of all translocations is a reciprocal t(9;22)(q34;q11) translocation created through DSBs within the breakpoint cluster region (*BCR*) gene on chromosome 22, and c-*ABL* oncogene 1, a non-receptor tyrosine kinase (*ABL*) on chromosome 9 (Figure 1-2). The t(9;22)(q34;q11) translocation creates a *BCR-ABL* gene fusion [4] within the resultant aberrant chromosome 22, commonly known as the Philadelphia chromosome (PhC) [5], which was first observed in 1960 by Nowell and Hungerford [6]. The chimeric *BCR-ABL* gene encodes a non-natural tyrosine kinase, BCR-ABL, which is highly active and deregulates proliferation of PhC-positive cells of myeloid lineage, the oncogenic signature of chronic myelogenous leukemia (CML) [7]. Various forms of the BCR-ABL kinase exist and are dependent on the specific breakpoints within both the *BCR* gene on chromosome 22 and the *ABL1* gene on chromosome 9 [8].

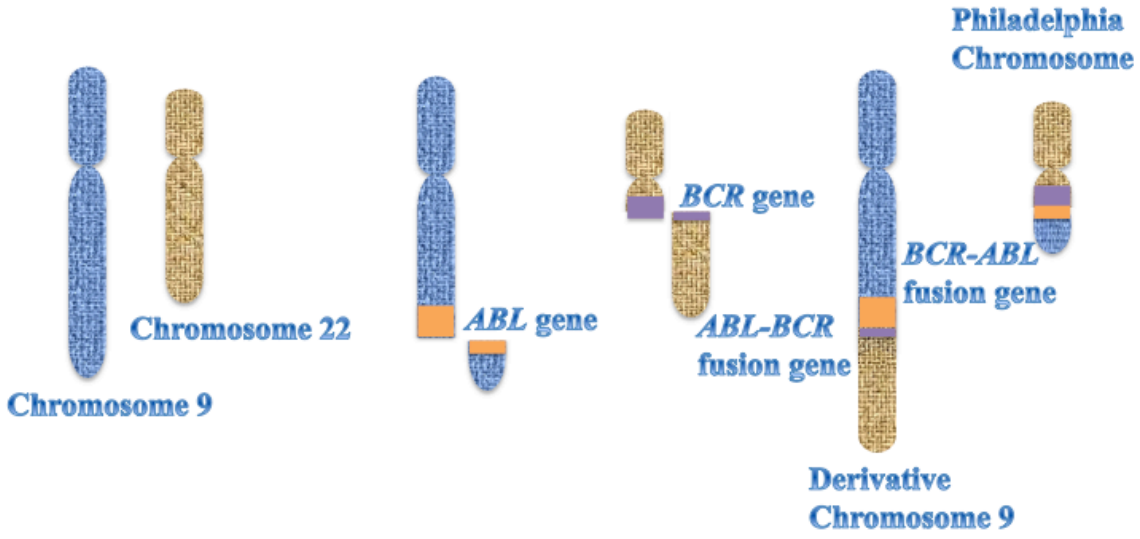


Figure 1-2. The Philadelphia chromosome, a reciprocal translocation between chromosome 22 and chromosome 9 resulting in formation of the *BCR*-ABL fusion gene.

CML is the most common myeloproliferative neoplasm, developing in immature blood cells in the bone marrow and affecting approximately 6000 patients in the USA annually [9]. Diagnosis and monitoring of this life-threatening cancer are based on detection of *BCR-ABL* and quantification of its frequency within the patient's granulocyte population. With its genetic

origins described by Rowley in 1973, the PhC was the first translocation to be associated with cancer [4]. Now proven to be causative of CML (approximately 95% of all CML patients are PhC positive in their cancerous (malignant) granulocytes), *BCR-ABL* and the PhC have since become a model for other cancers where translocations occur [10-12].

Non-small-cell lung cancer (NSCLC) is one such example. NSCLC patients often carry a non-reciprocal translocation on human chromosome 2 (Figure 1-3) in which DSBs within the echinoderm microtubule associated protein like 4 (*EML4*) gene and the anaplastic lymphoma kinase (*ALK*) gene allow an inversion of genetic material that forms the non-natural gene fusion *EML4-ALK* [13]. The chimeric *EML4-ALK* gene encodes a highly active tyrosine kinase, EML4-ALK, that is associated with greater than 5% of all NSCLCs. Lung cancer is the largest cause of oncological mortalities in the World, with approximately 1.8 million patients diagnosed each year; 85 to 90% of those patients are afflicted with NSCLC [14].

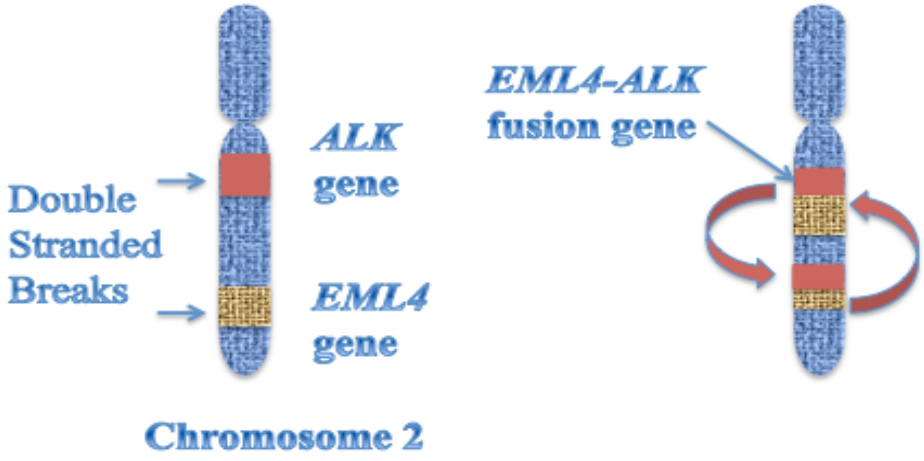


Figure 1-3. Schema of the inversion translocation on chromosome 2 that forms the *EML4-ALK* fusion gene.
The blue arrows indicate the regions of the double stranded breaks in the *ALK* and *EML4* genes. The red arrows illustrate the inversion process.

The detection of either the reciprocal translocation resulting in formation of *BCR-ABL* or the inversion translocation resulting in formation of *EML4-ALK* is currently achieved in clinics through direct visualization via a fluorescent *in-situ* hybridization (FISH) assay. For the *BCR-*

ABL fusion gene, the FISH assay utilizes two probes – one specific to the *BCR* gene and the other to the *ABL* gene – with fluorescence co-localization identifying *BCR-ABL*. Conversely, in FISH-based detection of *ALK*-positive NSCLC, the assay employs two probes against the *ALK* gene, with separation of signal indicating a translocation event. Known weaknesses of FISH-based translocation assays such as the *ALK*-positive NSCLC assay include a general inability to reliably detect the translocation when less than 15% of the assayed cell population carries it [15].

This thesis reports on a new platform that permits cost-effective and sensitive assays against either reciprocal or inversion translocation events. The platform exploits the unique advantages of droplet digital PCR (ddPCR) when applied to the detection of rare chromosomal rearrangements. The general applicability of this method is demonstrated through its use to create two new assays. The first enables detection of DSBs within the major breakpoint region (M-*BCR)* of the *BCR* gene. Breakage of *BCR* within its M-*BCR* is generally observed (> 98% of the time) in the *BCR-ABL* reciprocal translocation. The second assay detects the non-reciprocal *EML4-ALK* inversion translocation by assaying for DSBs within *ALK*. Both assays leverage the unique capacity of the polymerase chain reaction (PCR), when conducted in a ddPCR format, to detect and quantify complex translocations by isolating individual copies of target gene(s) or gene fragments into sub-nanoliter droplets (nL) [16]. End-point signals recorded for each droplet following ddPCR processing of the entire ensemble of droplets are used to count those containing either no template, amplified copies of DSB-generated and thus translocation-associated gene fragments, or amplified copies of the germline form of the target gene. A key benefit of this platform is the ability to directly visualize a translocation event in the raw output data. In addition, it also permits accurate quantification of the translocation frequency through application of a novel data-analysis method that utilizes Poisson statistics and data from multiplexed control reactions. Initial distributions of templates are defined among available droplets and loss of template integrity due to chromosomal shearing events during genomic DNA purification is quantified. The resulting ddPCR assays have the advantage of being far simpler and cheaper than the corresponding FISH assays, while providing greater reliability and a significantly lower limit of detection. As a result, the assays may find use in cancer genetics testing laboratories.

1.2 Translocations and their Mechanism of Formation

Translocations are most often associated with hematological cancers such as CML, but are increasingly being identified in solid tumors to the extent that they have now been detected in almost every cancer type [17]. A consequence of certain translocations, most notably reciprocal and inversion translocations, is the formation of a fusion gene that encodes an unnatural protein conferring unique activity or attributes to the mutated cell. Many of these are highly active transcription factors or kinases capable of disrupting or deregulating basic cellular processes, including expansion, differentiation and anti-apoptosis. More than 350 unique gene fusions involving 337 different oncogenes and proto-oncogenes have been identified to date in neoplasia [3]. Similarly, 109 and 82 unique gene fusions encoding active proteins have been identified in acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL), respectively [17].

Translocations can occur in stem cells and progenitor cells during DNA replication, and these rare events are closely linked to cancer risk and progression. Tomasetti et al., [18] have shown, for instance, that a correlation exists between number of stem cell divisions and cancer risk within the associated organ(s). It is estimated that there are up to 10 DSBs per dividing cell per day [19]. These DSB's can occur through a variety of mechanisms, including replication errors, exogenous stresses such as ionizing radiation or chemotherapeutic agents, or chromosomal breaks induced during activation of the adaptive immune system [1, 20]. Translocations may form when one or more of those DSBs prompt the cell to activate the DNA damage response (DDR), a complex repair system that targets DNA repair factors at the damaged sites and triggers cell signaling pathways to pause the cell cycle and initiate repair [21]. The DDR machinery includes homologous recombination repair (HRR), which occurs in the S phase of the cell cycle, and the non-homologous end-joining (NHEJ) pathway, which is active throughout the cell cycle. The NHEJ pathway, including the associated microhomology-mediated end-joining (MMEJ) process, is implicated in many translocations because it is a more error-prone method of DNA repair. This is partly because the NHEJ pathway will often add or delete nucleotides before rejoining the broken DNA ends [22], and the resulting microhomology associated with these events is often observed next to DSBs that are known drivers of cancer progression [23]. In contrast, the HRR utilizes a roughly 100 bp section of the sister chromatid as

a template to rebuild the broken DNA with no nucleotide additions or deletions. As a result, HRR-mediated DNA repair is far less error prone [24].

There is evidence that the likelihood of a translocation event is influenced by the time that transpires between the DSB(s) and activation of the DDR [1]. Longer times increase the chances of chromosomal fragment separation and rearrangement within the nucleus by diffusion. Orthwein et al., [25] for instance, found that the DDR mechanism is inhibited during mitosis to prevent M-phase telomeres from fusing, and that the delay in DSB repair caused by that inhibition contributes to translocation formation. Following a DSB, the affected chromatin adopts a more open structure to allow the repair loci to be established at the damaged site. In mammalian cells, this chromatin opening is thought to initiate DDR [26] and enable the DSBs to have limited local motion measured in terms of a mean squared displacement of approximately 1 μm^2 per hour [27].

Roukos has reported that 80% of all chromosomal DSBs occur within 2.5 μm of each other within a given nucleus [28]. The combined mobility and proximity of the DSBs make unnatural recombination events possible [29], with close spatial arrangement of two different chromosomes increasing the possibility of a translocation [1], as has been visualized by techniques such as FISH showing that proximal chromosomes undergo translocations at higher frequency [30, 31]. Indeed, it is the close nuclear proximity of chromosomes 22 and 9 that facilitates the formation of the PhC and the associated *BCR-ABL* fusion gene [32, 33].

Finally, translocations can arise when replication of different chromosomal regions or gene pairs occurs at the same time, and early replication timing predisposes a chromosome to a translocation event [34, 35]. In mammalian cells, the *BCR* and *ABL* genes have been shown to be early replicators and to replicate at the same time, explaining in part their propensity to undergo reciprocal translocation [36].

1.3 The Philadelphia Chromosome, the *BCR-ABL* Fusion Gene and CML

Though the pathology of CML and the role of the PhC in the disease are not the focus of this thesis, some knowledge in this area is useful to understanding the design and intent of the assays and associated platform described in this thesis. There are a number of different neoplasm disorders, including AML [37], essential thrombocythemia [38], multiple myeloma [39] and B

cell type lymphoma [40], for which a subset of the patient population harbors the PhC. For example, 20 – 30% of adults [41] and 2 – 3% of children [42] afflicted with ALL carry the PhC. For 30 – 50% of that PhC-positive patient population, the DSB occurs within the M-*BCR*, with the remaining population typically having a DSB within the minor breakpoint region (m-*BCR*) [43].

In healthy bone marrow, hematopoietic stem cells (HSCs) may differentiate into myeloid progenitor cells or lymphoid progenitor cells. Lymphoid-lineage committed cells produced by lymphopoiesis include lymphoblasts, B and T cells, natural killer (NK) cells, and dendritic cells. Many of these lymphocytes are short-lived, and maintenance of the immune system requires continual HSC self-renewal and differentiation. Myeloid progenitor cells, in turn, give rise to megakaryocytes, which produce platelets, as well as erythrocytes, granulocytes and macrophages. Each of these mature blood cells can also be identified through appropriate cell-specific markers [44]. Cell enrichment based on these markers has been used to show that PhC-positive CML patients generally express the *BCR-ABL* fusion gene within their leukocytes, most notably their granulocyte population, and that those malignant granulocytes generally do not reach full maturity before entering the blood stream (50).

In 98% of the PhC-positive CML patient population, the *BCR* gene on chromosome 22 is cleaved within its M-*BCR* [45], which spans *BCR* exons 12 to 16 and is approximately 5.8 kbp in length [46, 47]. The DSB does not occur at a particular base pair, but rather can occur at any position within this region. The DSB in the remaining *ca.* 2% of PhC-positive CML patients generally occurs in one of two other regions. Between 1% and 2% of CML patients harbor a PhC formed by a DSB within the m-*BCR*, a 54.4 kbp long "minor–breakpoint" section of the *BCR* gene lying between alternative exons 2' and 2 [48]. Less than 1% have a breakpoint within the micro breakpoint region (μ-*BCR*), a 1 kbp segment between exons 21 and 22 of *BCR* [49].

The position of the DSB within the *ABL* gene on chromosome 9 is less defined, generally occurring within a 150 kbp region that spans from the 5' terminus of the *ABL* gene to a position within intron a2 [50]. The reciprocal translocation process can produce a variety of different *BCR-ABL* gene fusion constructs depending on where the breakpoints occur in *BCR* and *ABL*. In CML, by far the most common construct results from breakpoints within the M-*BCR* and between exons a1 and a2 of *ABL*. These patients transcribe *BCR-ABL* mRNA that produces a BCR-ABL 'p210" chimeric protein (210 kDa) displaying the non-natural tyrosine kinase activity

causative of CML. However, an alternative p190 BCR-ABL kinase (190 kDa), more closely associated with ALL, may be expressed as a result of breakpoints occurring within the m-*BCR* and within intron a2 of the *ABL* gene [47]. A very small percentage of leukemias translate a p230 BCR-ABL kinase as a result of DBSs within the μ-*BCR* and intron a1 of *ABL*. Each BCR-ABL fusion protein produced from these three different transcript constructs exhibits deregulated tyrosine kinase activity and the CML phenotype in mouse models [51]. Small pathological differences are reported. For example, there is some indication of generally lower levels of p230 kinase, leading to slower disease progression. That unique pathology is identified as neutrophilic CML [52].

Collectively, there is compelling evidence that the malignant transformation causative of CML is due to the BCR-ABL chimeric protein's unnatural tyrosine kinase activity, irrespective of the mRNA from which it was translated. In particular, the BCR-ABL tyrosine kinase is fully active independent of signaling from upstream effectors, and its presence can result in aberrant expression of genes that enable malignant cell proliferation [53], as well as deregulation of anti-apoptotic and metabolic pathways [54]. Elements encoded within exon 1 of the *BCR* gene facilitate the constitutive activation of the latent kinase activity within the ABL components of BCR-ABL [55, 56]. Exon 1 of *BCR* also encodes a capacity to bind the SH2 domain of ABL, as well as SH2 domains present in other non-receptor tyrosine kinases (e.g., SRC), in a phospho-tyrosine-independent manner [57]. This is important, as the oncogenicity of BCR-ABL requires binding to SH2 domains on ABL, other SH2-containing kinases, or both; deletion of BCR sequences mediating SH2-binding render BCR-ABL non-transforming [58]. The BCR elements of BCR-ABL generally bind SH2-domains through phosphorylated tyrosine residues [59]. In particular, phosphorylation of tyrosine 177 of BCR-ABL is reported to be essential for activation of downstream effectors such as RAS [60].

The classic PhC is created through a balanced reciprocal translocation. A corresponding *ABL-BCR* fusion gene is consequently formed on derivative chromosome 9 [61, 62]. As a result, the GTPase-activating protein (GAP) activity encoded within the *BCR* gene is transferred to the *ABL-BCR* gene product. That finding has motivated research into the function of the ABL-BCR fusion protein and its role in CML oncogenesis [63]. It was originally thought that the ABL-BCR fusion protein might exert a regulatory action on BCR-ABL kinase activity, due in part to the observed poor prognosis of patients showing loss of derivative chromosome 9 (and thus *ABL-*

BCR), but this has recently been disproved; when treated with the tyrosine kinase inhibitor (TKI) imatinib, patients showing loss of derivative chromosome 9 have equal response to that of the general CML patient population [63].

1.3.1 Stages of CML

CML develops slowly, over months to years, with patients often asymptomatic or only mildly symptomatic early in the cancer pathogenesis [64]. Its progression is clinically defined in three stages: chronic phase, accelerated phase, and blast phase. In the chronic phase, less than 10% of all cells in the blood and bone marrow are blasts, which are abnormal immature leukocytes/granulocytes. CML often progresses slowly in the chronic phase, which can last several years. CML patients in the chronic phase usually respond well to treatment with TKIs, which act to decrease or eliminate populations of cells harboring a PhC; TKI treatment thereby slows or prevents disease progression to the accelerated or blast phase. In the accelerated phase, 10 – 20% of cells in the blood and bone marrow are blasts, and relative to their chronic-phase states those abnormal cells display more genomic damage and accelerated mitosis. Chromosomal mutations within and in addition to the PhC are often observed, and symptoms include fever, poor appetite and weight loss. Treatment with TKIs becomes less effective, with disease reoccurrence often observed within 2 years. Allo-stem cell transplantation is a secondary treatment option for CML patients in this phase [65].

CML in the accelerated phase can quickly progress into the blast phase, with the more severe symptoms of high fever, malaise and an enlarged spleen collectively called blast crisis. Blast crisis symptoms are consistent with those of an acute leukemia [66] – more than 30% of cells in blood and bone marrow are blasts, and those blast cells often have spread to tissues and organs outside the bone marrow and circulatory system. Leukocyte and platelet counts become highly abnormal, and bleeding and infections may occur. TKI treatment response is limited to a few months, and stem cell transplant therapy is less effective than observed in the accelerated phase. Expression of *BCR-ABL* no longer drives disease progression, as other oncogenic chromosomal and molecular alterations occur at the onset of and during blast phase, and these additional genome imbalances, which include losses on chromosomes 1, 5 and 9, and significant gains on chromosomes 1, 8, 9, 16 and 22, accelerate the cancer [67, 68].

1.3.2 Further Variations within the Philadelphia Chromosome and Derivative Chromosome 9

Deletions within derivative chromosome 9 have been observed in 15.7% of all CML patients [69-72]. About 12% of patients with a classic PhC carry losses within derivative chromosome 9, [72-76] whereas nearly 40% of patients with a variant *BCR-ABL* translocation (10% of the patient population) show such deletions [72-77]. The size of the deletions vary considerably between patients, with 260 kbp to 41.8 Mbp losses often observed on the 5′ end of *ABL* and 230 kbp to 12.9 Mbp losses on the 3′ end of *BCR*. These deletions generally occur at the onset of the disease, but can arise during disease progression.

Huntly et al., [78] analyzed 193 CML patients and found that 34% of patients having a loss in derivative chromosome 9 lacked the *ABL-BCR* transcript. In testing 71 CML patients with a loss in derivative chromosome 9, Albano et al., [79] found that 66% had deletions in both the 5′*ABL* and 3′*BCR* gene regions, while an additional 18% and 16% had a loss in only the 5′*ABL* or the 3′*BCR* gene region, respectively. Deletions in derivative chromosome 9 were initially though to be an indicator of poor prognosis during treatment with interferon α, but this became a controversial hypothesis once treatment with imatinib was established [75].

Variations in chromosome 22 as a result of or following the translocation event have also been observed, such that *ca.* 10% of all CML patients are classified as having a variant PhC rather than a classic PhC. These patients may also display rearrangements within chromosome 9 and/or other chromosomes, but still harbor the active *BCR-ABL* fusion gene causative of CML [17, 80]. Commonly observed variations in chromosome 22 include deletions [81], as well as additional scrambling of elements of the *BCR* and/or *ABL* genes, with 11% of the variant patient population harboring micro-insertions of *ABL* genetic information within the *BCR* gene or *vice versa* [82]. Finally, elements of the 3′ end of *BCR* have also been shown to transfer, in some cases, to a chromosome other than chromosome 9 [83].

There is no indication that either a variant PhC or a loss of derivative chromosome 9 has any impact on prognosis of patients treated with TKIs [75, 84]; for example, 73% of a cohort of patients carrying a variant PhC showed no significant change in response to imatinib therapy relative to a classic PhC control group [85]. The pathology of CML characterized by a variant PhC and losses in chromosome 9 is still controversial. It also remains unclear whether the

translocation occurs in a single step, as in the formation of the PhC, or via a multistep process [84]. Marzochi et al., [84], in analyzing 30 patients with a variant PhC, found that the majority, 80%, formed their variant translocation in a one-step process involving 3 chromosomes; 6% of the patients instead underwent a multistep process in which a classic PhC was first formed and then sequentially modified. But irrespective of whether a patient carries a classic or variant PhC [75, 84, 85], the *BCR-ABL* fusion gene is generally expressed, and its detection is a sufficient diagnosis of CML.

1.3.3 Tyrosine Kinase Inhibitors and Treatment of CML

The first TKI against BCR-ABL was developed by Novartis (formerly Ciba-Geigy). It is a small-molecule aromatic compound originally named GCP57148B, but now more commonly identified as either imatinib or imatinib mesylate [86]. Imatinib inhibits BCR-ABL tyrosine kinase activity by binding to the kinase domain. At sub-micromolar concentrations in blood, imatinib prevents BCR-ABL from phosphorylating activating tyrosines on downstream effectors [87], inducing apoptosis and blocking transmission of proliferative signals to the nucleus [88]. Based on highly successful clinical trails [89], imatinib (trade name Gleevec/Glivec) was approved by the US Food and Drug Administration (FDA) in 2002 for treatment of CML [87, 90].

Hughes et al., found that the overall survival rate for chronic CML patients on daily imatinib therapy (400 mg standard dose) is about 88% [91]. Of the cohort of CML patients within that large study, 14% progressed to later stages of the disease, while imatinib therapy was discontinued in 5% due to serious side effects [91]. Early treatment of patients in the chronic phase of CML has been shown more effective than treatment during later stages of the disease [92]. An eight year follow-up study of patients treated early in CML pathogenesis found that 83% had a complete cytogenetic response (see Figure 1-5), with an event free survival rate of 81% [93]. An early cytogenetic response is an indicator of a positive long-term outcome, while lack of complete cytogenetic response in 12 months of treatment correlates with a poor prognosis, as the probability of disease progression or development of imatinib resistance increases to 38% [91]. Overall, imatinib treatment results in an eight year survival rate of greater than 90% for patients treated in chronic phase CML [94].

As resistance to imatinib treatment is observed, the European LeukemiaNet (ELN) and National Comprehensive Cancer Network have adopted a shared set of recommendations for treatment regimens in this instance [95]. Imatinib resistance is a time-dependent diagnosis that monitors reduced response after initiating therapy. Resistance then manifests as an increase of leukemic load at any time during therapy. For CML patients found to be resistant or intolerant to imatinib, treatment with a 2nd-line TKI is initiated. FDA-approved second line therapies include dasatinib (Sprycel) and nilotinib (Tasigna) [96]. Both have therapeutic efficacies comparable to imatinib [97], though patients in general experience more serious side effects. Dasatinib is more promiscuous, blocking many kinases in addition to BCR-ABL [98, 99] and exhibiting tolerance to a range of common somatic mutations in *BCR-ABL*.

For approximately 20% to 25% of CML patients who have developed resistance to a TKI, higher doses may be applied, with 30% to 50% of patients showing positive response [100]. In some instances the resistance is associated with a T315I (a threonine to isoleucine gate-keeper type) somatic point mutation in the kinase domain of BCR-ABL [101]. The T315I mutation alters the structure of the ATP-binding pocket and thereby eliminates binding of first- and second-generation TKIs [94]. Ponatinib, a multi-targeted 3rd line kinase inhibitor, has been approved by the FDA to treat CML patients who develop the T315I mutation [102]; it is effective, though patients often experience very serious side effects [102]. Other FDA-approved treatments are also available in this case, including omacetaxine, which is not a TKI, and bosutinib, which is [103]. Omacetaxin's mechanism of therapeutic action is unclear, but it is thought to target the BCL2, MCI-1 and HSP90 pathways [104].

Though TKIs against BCR-ABL are effective, they are not a cure [66]. Patients in complete molecular remission after TKI treatment may still harbor BCR-ABL expressing leukemic cells [89, 105, 106]. Current research is working to complement BCR-ABL kinase inhibition with therapies that target downstream effectors to eliminate leukemic stem cells. For example, Hedgehog (Hh) signaling is activated via up-regulation of the smoothened (SMO) G protein-coupled receptor by BCR-ABL tyrosine kinase, and inhibition of SMO has been shown to reduce leukemic stem cells [107, 108]. SMO activates the GLI family of transcription factors, which in turn activate Hh [109, 110]. Novartis is currently developing a SMO receptor antagonist as a potential treatment option for CML patients.

Similarly, the P13K/AKT/mTOR pathway is activated by BCR-ABL. Rapamycin-based inhibition of mTOR signaling, including in combination with the anti-tumor agent celecoxib, is also being pursued as a next-generation CML treatment strategy [111], with clinical trails now underway [112]. In addition, the aurora kinase family regulates mitosis and their over expression has been associated with many cancers [113]. Aurora kinase and ABL kinase co-inhibitors currently being developed for CML treatment include tozasertib, danusertib and KW2449 [114]. Other potential options for targeting downstream BCR-ABL effectors include the SRC kinases [115], protein phosphatase 2 (PP2a) [116], B-cell lymphoma 6 protein (BCL6) [117], promyelocytic leukemia protein (PML) [117, 118], PTEN [119], and Wnt/β-catenin [120]. The goal is to achieve remission [121] without the requirement for further treatment [122, 123].

1.4 *ALK*-positive Non-Small Cell Lung Cancer

Approximately 1.5 million mortalities worldwide are attributed to lung cancer each year [124]. There are 2 major forms of lung cancer – non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC accounts for 85% to 90% of all lung cancers [125] and can be further classified into 3 major histological subtypes – adenocarcinoma, squamous-cell carcinoma and large cell lung cancer [126]. Approximately 25% of all lung cancers are in patients who have never smoked and are diagnosed with the adenocarcinoma NSCLC sub-type [127].

Soda et al., [13, 79] discovered that a subset of NSCLC patients harbor rearrangements in the anaplastic lymphoma kinase (*ALK*) gene that encodes a transmembrane receptor tyrosine kinase ALK thought to be involved in development. Constitutive activation of ALK has been shown to occur in a number of different ways. In NSCLC, it most often occurs through a paracentric (non-reciprocal) translocation (inv(2)(p21;p23)) on chromosome 2 in which the gene fragment encoding an N terminal region of the echinoderm microtubule associated protein-like 4 protein (EML4) is inverted and fused to a fragment of the 3′ end of *ALK* to create an *EML4-ALK* fusion gene encoding an unnatural tyrosine kinase EML4-ALK (79). This "*ALK*-positive NSCLC" subtype comprises 2.6% to 6.7% of the total NSCLC population, equating to approximately 40,000 to 100,000 patients diagnosed worldwide each year [13, 14, 128-131]. Further analysis has revealed that the *EML4-ALK* translocation tends to occur in younger male adenocarcinoma NSCLC patients who have never smoked or are only light smokers [15, 128,

132], though there have been a few documented cases of heavy smokers who are *ALK* positive [15]. There is likewise some indication of heightened prevalence of *EML4-ALK* translocation in the Asian community, where it most often affects younger females [128]. However, recent studies on a Western population indicate the subset of NSCLC patients who are *ALK* positive is ca. 5.6%, consistent with the levels seen in Asia [133].

The normal function of ALK is not well defined [134], but it is believed to be an insulin receptor with a potential role in ligand binding within the nervous system [135]. The *ALK* gene is currently implicated in 16 cancer-related fusions, 3 of which have been observed in solid tumors (*ALK* fusions to *EML4*, *TFG* and *KIF5B* [3, 136, 137]. The *EML4-ALK* gene fusion has also been identified in breast and colorectal cancer [137].

Though rarely observed in other types of NSCLC, the *EML4-ALK* fusion gene has been shown to be a key driver of NSCLC adenocarcinoma pathogenesis [138] and is often observed in lung fibroblasts [126], in which the gene product then activates signaling pathways such as MAP kinase, PI3K and STAT, deregulating cell proliferation and anti-apoptotic behavior [128, 139].

1.4.1 Current Treatments for *ALK*-positive NSCLC

Crizotinib (xALKori; Pfizer Inc.) [140, 141] is a TKI against ALK and MET kinases that received accelerated FDA approval for treatment of advanced stage *ALK*-positive NSCLC in 2011. The approval was linked to a companion diagnostic test, the Vysis *ALK* break-apart FISH assay (Abbott, US). Crizotinib was later approved for therapeutic use in Canada (2012). Though a companion assay for *ALK*-positive NSCLC was not specified in that instance, Health Canada currently requires clinical evidence of the inv(2)(p21;p23) translocation (or a rearrangement in *ALK*) prior to crizotinib treatment [142]. Just over half (55%) of all *ALK*-positive NSCLC patients respond positively to crizotinib, exhibiting a progression free survival at 6 months of 72% [15] and a 2-year overall survival rate of 54% [141]. Crizotinib is currently the preferred first line therapy. When compared to standard chemotherapeutic treatment, which is characterized by a one year survival rate of 33% [143] and a 10% positive response rate [15], crizotinib offers a much improved treatment option.

NSCLC patients on crizotinib will eventually develop resistance to treatment, usually within 10-12 months [144, 145]. Crizotinib associates with the ATP binding domain of the

EML4-ALK tyrosine kinase [128] to inhibit its activity. However, somatic mutations can occur within the *ALK* portion of *EML4-ALK*, much the same way as in *BCR-ABL*, which prevent the TKI from binding. L1196M has been identified as a gate-keeper mutation in *EML4-ALK*; it alters the ATP binding site to create steric hindrance that prevents the TKI from binding [145-147].

A number of second generation TKIs have been developed to try and combat this resistance, including ceritinib (zykadia; Novartis Inc.) [148] and alectinib (RO5425802/CH5424802; Chugai-Roche Inc. [149, 150]). These differ from crizotinib through their higher selectivity and potency against the ALK domain of the EML4-ALK fusion kinase. In 2014, ceritinib received accelerated FDA approval [151] for treatment of *ALK*-positive NSCLC in cases where there is (re)progression of the disease or an intolerance of crizotinib [152, 151]. Ceritinib has recently shown an overall response rate of 58%, with a progression free survival of 7 months in patients previously treated with crizotinib [153]. On average, progression-free survival of *ALK*-positive NSCLC patients on crizotinib is 8.2 months, with sequential treatment with ceritinib furthering that to a total of 17.4 months [149].

Patients carrying a L1196M, G1269A, I1171T, S1206Y or C1156Y mutation in *EML4-ALK* respond to ceritinib treatment, while those with a G1202R or F1174C point mutation show some resistance [153]. Furthermore, patients harboring an I1151T/N/S somatic point mutation are resistant to alectinib but sensitive to ceritinib [154]. Ceritinib has the added advantage that it can cross the blood brain barrier and is thereby able to act at the most common site of NSCLC-related metastasis [150].

Other treatment strategies are in development. An inhibitor to heat shock protein HSP90 (a chaperone necessary for folding and stabilization of tyrosine kinases) is currently in stage II clinical trials and shows promise for treatment of *ALK*-positive NSCLC [155]. In addition, BCR-ABL and EML4-ALK tyrosine kinases share common downstream signaling mediators that may undergo alterations to provide an alternative mechanism for cancer relapse in the presence of the targeted TKI [156]. One third of patients who develop resistance to crizotinib exhibit hyperactive epidermal growth factor receptor (EGFR) and cKIT activated MAP kinase pathways [146]. In this patient population, a 44% increase in EGFR phosphorylation within tumor biopsies collected at the time of resistance was observed (relative to levels prior to crizotinib treatment (negative control)) [145]. Combining TKI therapy with other agents, particularly monoclonal antibodies,

targeting EGFR and cKIT activated pathways is also under investigation as a treatment for *ALK*-positive NSCLC [157].

Finally, the kirsten rat sarcoma viral oncogene homolog (*KRAS*) is mutated in 15% of NSCLC patients [158] and may serve as an alternative driver of oncogenesis. The *EML4-ALK* gene fusion has been found to be mutually exclusive of *KRAS* and *EGFR* mutations [15, 159]. As a result, *ALK*-positive NSCLC patients rarely harbor a *KRAS* or *EGFR* mutation [132, 160]. It is clinically important to identify patients that harbor the *EML4-ALK* gene, as TKI treatment against mutations in *EGFR* or *KRAS* is ineffective for *ALK*-positive NSCLC [132], while ceritinib is effective in these cases.

1.5 Current Methods for Detecting Translocations

1.5.1 Clinical Detection of the *BCR-ABL* Fusion Gene

Preliminary diagnosis of CML is usually indirect and based primarily on a complete blood count (CBC) that measures the number and quality of leukocytes, erythrocytes and platelets. CML is suspected when granulocytes within the leukocyte population are irregular and immature. Low platelet counts, known as thrombocytopenia, low neutrophil counts, or low red blood cell counts are also suggestive of leukemia. Various blood chemistry, clotting, immunohistochemical and cyto-chemical tests are used to evaluate leukemia in conjunction with CBC data. Fatigue, weight loss, joint pain, and/or an enlarged spleen are also indirect indicators of CML. During patient treatment, regular CBC tests are conducted to monitor treatment outcome and disease relapse.

Unequivocal diagnosis of CML (as well as specific types of ALL) is generally provided through direct or indirect detection of the *BCR-ABL* fusion gene. Several diagnostic platforms are available to detect the *BCR-ABL* fusion. Until recently, conventional cytogenic analysis of metaphases (dividing cells) in an aspirated bone-marrow sample was the method of choice [161]. That method is based on using Giemsa staining of metaphase chromosomal spreads to gauge the percentage of PhC-positive cells. However, it suffers from several known limitations, including high costs, applicability only to dividing cells, limited sensitivity and unacceptable failure rates arising primarily from lack of metaphases. National Comprehensive Cancer Network (NCCN)

guidelines require evaluation of a minimum of 20 metaphases, setting the detection limit at no better than 5%, while requiring the patient to undergo an invasive and painful bone marrow aspiration procedure [162].

Conventional karyotyping methods have largely been replaced by molecular tests of the translocation, fusion gene, fusion transcript or fusion-gene product. For initial diagnosis of CML, by far the most widely employed of these are fluorescent *in-situ* hybridization (FISH) assays (Figure 1-4). FISH is a molecular cytogenetic technique that allows genes to be analyzed at the chromosomal level. The Vysis LS1 *BCR-ABL* dual-color, dual-fusion translocation (D-FISH) assay (Abbott Molecular) uses two long oligonucleotide probes. The first is 650 kbp in length, labeled with a unique fluorescent reporter (spectrum orange: λ_{ex} = 559 nm, λ_{em} = 588 nm (red)) and specific to a region of chromosome 9q34 spanning from just before the 5′ end of the *ASS* gene to just after the 3′end of the *ABL* gene. The second targets a region of the *BCR* gene near its 5′ end that spans the m-*BCR* and most of the M-*BCR*; it is 300 kbp in length and is labeled with a different reporter (spectrum green: λ_{ex} = 497 nm, λ_{em} = 524 nm (green)). The *BCR-ABL* D-FISH assay typically analyses 100 to 500 nuclei isolated from either bone marrow or peripheral blood specimens and reports on the fusion patterns observed. The *BCR-ABL* fusion gene within each imaged granulocyte may be identified by the co-localization of the reporters to create a yellow fluorescence signal [163].

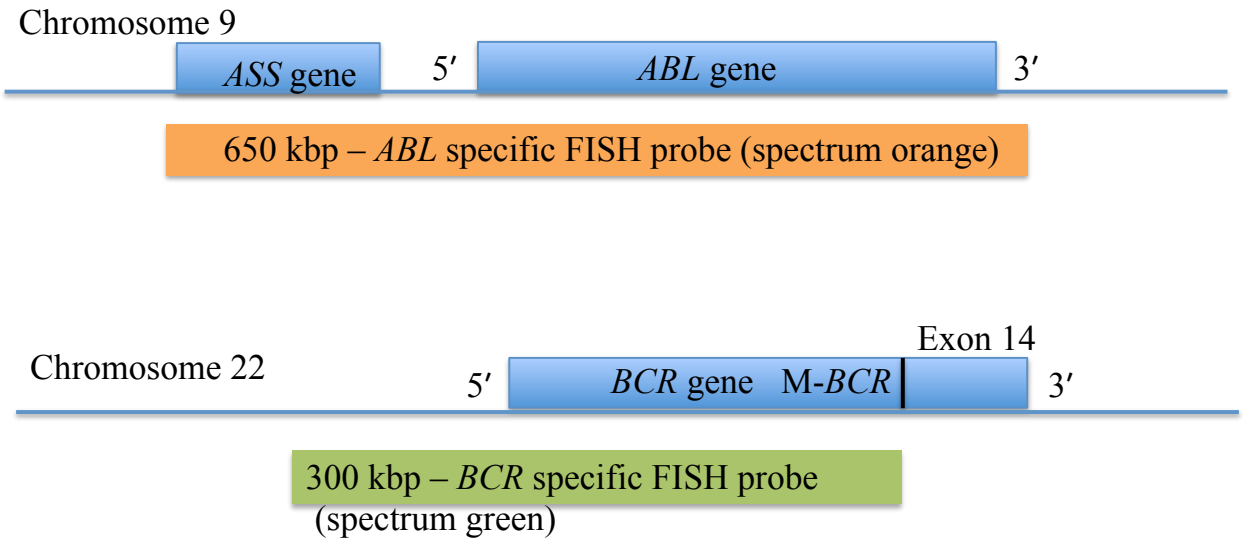


Figure 1-4. Basic structure and targets of the *BCR-ABL* FISH assay.
Black vertical line on Chromosome 22 indicates that the 300 kbp *BCR* specific FISH probe interrogates only a portion of the M-*BCR*.

D-FISH can be performed on metaphase cells, where it is reasonably accurate. It is less precise when applied to interphase (non-dividing) cells because *BCR* and *ABL* specific probe signals coincidentally overlap to give a "false fusion" co-localization signal (i.e., false positive) in about 4 - 5% of normal nuclei [164]. The dual-probe dual-color *BCR-ABL* D-FISH assay is in general unable to distinguish a normal chromosome 9 from a derivative chromosome 9, so the information content of the assay is effectively restricted to analysis of rearrangements in chromosome 22.

To reduce false positive readings, a tri-color FISH probe kit may be utilized, that incorporates into the traditional 2 probe *BCR-ABL* FISH assay, a third 270 bp probe labeled with a unique reporter that allows detection of both the PhC and derivative chromosome 9 in PhC-positive cells [165]. This improves analysis of interphase cells and is reported to reduce false positives to 2 ± 0.3% [166].

False negatives may also be an issue. Chase et al., [164] analyzed the D-FISH assay to find that the false negative rate is influenced both by the nucleus size and the position of the breakpoint on the *ABL* gene. The lowest false negative rate (ca. 2%) was achieved for nuclei

between 300 and 500 nm in nominal diameter and when the breakpoint on the *ABL* gene is closer to exon a2 than the 5′ end [164].

The advantages of FISH over conventional karyotyping include its applicability to both bone marrow and more easily accessed peripheral blood specimens, its ability to analyze both interphase and metaphase cells, and its ability to identify variant translocations. The most advanced *BCR-ABL* FISH assays have a limit of detection of 2%. The limitations of these assays are well documented and represent a serious clinical challenge. Fluorescence signals arising from FISH staining of nuclei can be incorrectly scored in many PhC-positive nuclei (false-negative cells) because scoring of *BCR-ABL* fusion signals is subjective and generally requires a highly trained and experienced technician [167]. The assay requires costly reagents [168] and is time consuming [169], requiring about 48 hours to complete. Furthermore, it is unable to detect micro-deletions of less than approximately 190 kbp [170]. These technical challenges limit the clinical utility of FISH to identify and quantify *BCR-ABL* and PhC-positive cells.

Following diagnosis, CML patients undergoing treatment with a targeted first-line (e.g., imatinib) or second-line (e.g., dasatinib) TKI and are typically monitored using an appropriate molecular test to gauge the efficacy of the therapy and to monitor minimal residual disease (MRD) during (and following) treatment. *BCR-ABL* levels at initial diagnosis are used as a reference. Poor response to treatment may thereby be identified and used clinically to guide decisions pertaining to adoption of alternative therapeutics or treatment regimens. The high false positive rates and corresponding high detection limit (typically 2% – 5%) observed in *BCR-ABL* FISH assays make that method unsuitable for MRD monitoring, as the ability to observe a complete molecular response in *BCR-ABL* fusion gene frequencies ($< 10^{-4}$) is typically required. Instead MRD monitoring is usually performed by reverse-transcription quantitative PCR (RT-qPCR) analysis of CML-patient specific *BCR-ABL* mRNA every 3 to 6 months. Patient response is then measured on an internationally adopted scale (Figure 1-5) that correlates leukemic cell load to log reduction in *BCR-ABL* transcripts, with 10^{12} leukemia cells equating to a 100% *BCR-ABL* to control gene (either *BCR* or *ABL*) ratio [171]. A complete cytogenetic response is a 2-log normalized reduction, while a major molecular response is a 3-log reduction in *BCR-ABL* transcripts (i.e., $\leq$ 0.1% *BCR-ABL* frequency) [91], which is known to correlate with good progression-free survival.

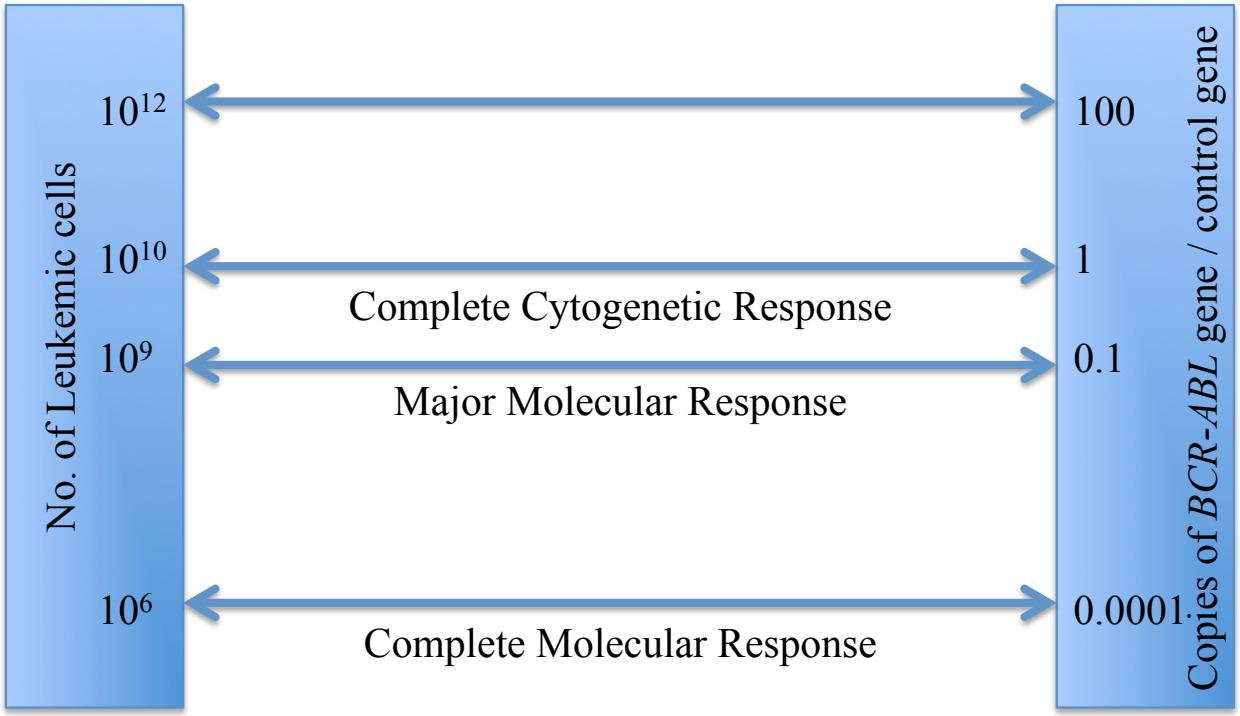


Figure 1-5. International scale for the monitoring of Minimal Residual Disease in CML patients.
Scale is based on a reference of 10^{12} Leukemic cells equating to 100 copies of *BCR-ABL*/(control gene).

Reported detection limit (LOD) values for RT-qPCR *BCR-ABL* assays are as low as 0.0001% *BCR-ABL* mRNA, or about 10^6 leukemic cells [172]. The RT-qPCR *BCR-ABL* assay is multistep, technically challenging and difficult to score. In surveys conducted by the College of American Pathologists on identical samples [173] significant variations in *BCR-ABL* to control ratio values were reported, with assay performance dependent on specimen collection and mRNA isolation methods, internal control selection, reverse transcription efficiency, and standard curve creation. LOD is likewise dependent on the quality of the mRNA sample and the efficiency of the RT step, as well as the intrinsic detection limit of the quantitative PCR step [174].

Relative to the FISH assays used for initial CML diagnosis, the significantly better sensitivity of RT-qPCR assays used for MRD monitoring is directly attributable to the fact that the latter uses the specific *BCR-ABL* transcript sequence of the patient as a basis for template and

signal amplification. In contrast, the amplification-free FISH assay is designed for general application to all patients, and requires no prior knowledge of the patient's *BCR* or *ABL* gene sequence.

Weaknesses to RT-qPCR based monitoring of PhC-positive leukemic cells are well documented [175, 176]. RNA is far less stable than DNA due to the 2′-OH group of a ribonucleotide being more reactive that the 2′-H of a deoxyribonucleotide [177]. Results from RT-qPCR depend on per cell *BCR-ABL* transcript expression levels which are variable and do not provide a direct measure of the number of leukemic cells. Normalized *BCR-ABL* transcript abundances have been shown to correlate relatively poorly and non-linearly with number of leukemic cells. But perhaps most importantly, the significant variability in RT-qPCR assay results has made it impossible to establish standards across all laboratories [89], and guidelines for acceptable levels of reproducibility and sensitivity are lacking [90].

The RT-qPCR assay may be applied to mRNA isolates from either peripheral blood or bone marrow. Differences in the levels of leukemic cells measured in these two different specimens tend to be reasonably small, permitting less-invasive peripheral blood samples to be used to monitor chronic-phase CML patients undergoing treatment or post-treatment. In the advanced phases of the disease, the use of bone marrow is preferred [90]. Disease relapse in a patient is generally indicated by a greater than 2-fold increase in *BCR-ABL* levels above minimum levels during treatment [90]. When a relapse is confirmed, *BCR-ABL* mutation analysis is undertaken [178].

1.5.2 Clinical Detection of the *ALK* Biological DSB

Due to the established clinical benefit in treating *ALK*-positive NSCLC patients with crizotinib, a number of different methods have been developed to detect *ALK* rearrangement or constitutive ALK (EML4-ALK) activation. These include qPCR analysis of the ALK kinase domain [179], RT-qPCR of the *EML4-ALK* fusion transcript [147, 180], and immunohistochemical (IHC) staining of the chimeric EML4-ALK kinase [130, 140, 181]. However, the only FDA-approved testing method is a "break-apart" FISH assay (Vysis *ALK* break-apart FISH assay; Abbott Molecular Laboratories). That assay (Figure 1-6) dominates clinical use despite its complexity and the associated need for highly trained technicians.

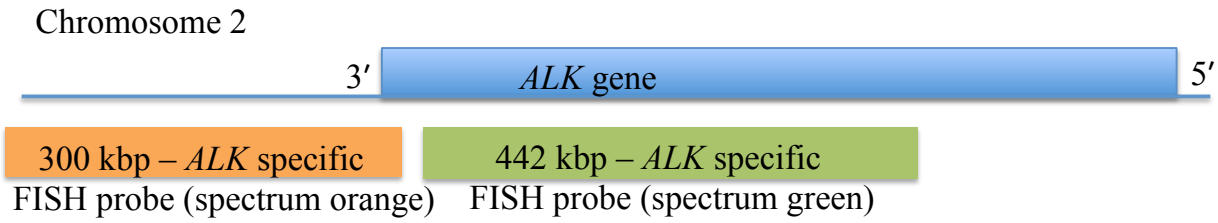


Figure 1-6. Basic structure and targets of the *ALK* break-apart FISH assay.

Testing for *ALK* rearrangements requires biopsy of potentially tumorous lung tissue, and the acquisition of those specimens is difficult and painful. To reduce costs and ease patient discomfort [147], surgical resection of the tumor is generally avoided, particularly in advance stages of NSCLC, in favor of a fine needle biopsy, a trans-bronchial biopsy, bronchial washing or pleural fluid collection. Much of that small formalin-fixed paraffin-embedded (FFPE) sample is typically devoted to pathological testing, so only minimal tumorous tissue may be available for molecular testing of *ALK* [142]. Furthermore, to satisfy Medicare insurance requirements in the US, testing for oncogenic *EGFR* mutations is usually done before *EML4-ALK* testing or any broader sequencing studies [182]. FISH testing of *ALK* rearrangements is based on specific binding of two long uniquely labeled probes to sequences to the 5′ side and the 3′ side, respectively, of the region within exon 20 of the *ALK* gene in which the *ALK* breakpoint is most commonly observed in *ALK*-positive NSCLC. Co-localization of the red reporter on one probe with the green reporter on the other to produce a yellow ("fused") signal indicates a normal *ALK* gene state within an imaged cell nucleus. Separation (break apart) of red and green signals (by at least 2 signal diameters) indicates an *ALK* rearrangement, and thus the presence of tumor nuclei, within the tissue specimen. Loss of a green signal in nuclei displaying a red signal is also observed. This result is presumed to be positive evidence of a rearrangement, in this case resulting in loss of the binding site for the probe specific to the 5′ side of the breakpoint region. Because both cell imaging and signal interpretation are quite difficult in the assay, a FISH result is considered positive for *ALK* rearrangement only when greater than 15% of the nuclei within a sample have delocalized *ALK* 5′ and *ALK* 3′ probe signals, loss of 5′ probe signal, or some combination thereof [15]. Lost signals, non-specific probe hybridization and background noise

all serve to further complicate this assay [183]. For example, loss of red signal in nuclei displaying a green signal is not counted as positive evidence of *ALK* rearrangement, due in part to the possibility the result might indicate loss of *ALK* kinase activity.

Improved methods for assaying rearrangements in *ALK* are clearly needed and some have been developed. Among the most advanced is the IHC test for *ALK*-positive NSCLC developed by Roche, which was approved in Europe in 2012 but is not approved in either the US or Canada. Published results for the IHC assay are of similar quality to those for the break-apart FISH assay [184], though significant differences in the results from the two assays have been reported [140, 181]. Those differences cannot all be attributed to the technical and interpretive issues [140] associated with the break-apart FISH assay, as quality control of the ALK antibodies used in the IHC assay remains an issue [130], and difficulties in achieving acceptable levels of staining have also been reported.

Similarly, the qPCR and RT-qPCR methods reported to date have not proven sufficiently robust [147] or comprehensive [180] to establish their clinical use in *ALK*-positive NSCLC testing.

As accurate diagnosis of *ALK*-positive NSCLC is critical to defining proper course of therapy (182), there remains a need for a reliable method to detect *ALK* rearrangements, most notably the *EML4-ALK* fusion gene, at low frequencies. The development of that capability would enable early disease diagnosis and an improvement in survival rates [185].

1.6 Thesis Objectives

Chromosomal translocations are major genomic events known to correlate with cancer risk and progression [1]. Identifying the presence and frequency of certain translocations has been shown to allow for a targeted approach to treatment of associated cancers that improves patient outcomes and quality of life [186]. While the FISH assays currently used clinically provide a means of detecting cancer-associated translocations, including the reciprocal translocation forming the *BCR-ABL* fusion gene and the inversion translocation associated with *ALK*-positive NSCLC, they generally do so with a detection limit of no better than about 2%. The utility of FISH assays as a means of early cancer detection is therefore limited, as the

confidence one has in the assay output diminishes near the detection limit, while the high cost and long processing times of FISH strain the budget and workflow of cancer genetics testing laboratories.

The central objective of my thesis is to develop a new general method for creating higher sensitivity assays against cancer-associated translocations that ameliorates limitations in current assays. The method is expected to be applicable to development of assays against either a reciprocal or non-reciprocal (inversion) translocation. This will be demonstrated through use of the platform to develop assays that permit reliable initial diagnoses of CML (reciprocal translocation) and *ALK*-positive NSCLC (inversion translocation) early in the cancer pathogenesis. The platform will leverage the many known advantages of detecting and quantifying genomic alterations/mutations by ddPCR, which to date has not been applied to the direct detection and analysis of translocation events at the genomic DNA (gDNA) level. It will operate on a purified gDNA specimen derived from relevant tissue drawn from the patient, and will require high-quality representations of the gene(s) or chromosomal region(s) to be interrogated.

Specific technical objectives associated with development and validation of the platform that will be addressed include:

- Tailoring of current protocols for purifying chromosomal DNA from human tissues to minimize shear-associated loss of the genomic information to be interrogated.
- Applying ddPCR principles and technology to develop a low-cost, time-efficient and sensitive assay to detect fragmentation events in the *BCR* gene indicative of the *BCR-ABL* fusion gene and CML.
- Further exploiting ddPCR to develop an improved assay to detect fragmentation events within *ALK* indicative of *ALK*-positive NSCLC.

It is anticipated that the assays developed using the platform will permit quantitative detection of translocations at frequencies ($\geq$ 1%) relevant to initial cancer diagnoses.

Chapter 2 describes the refined chromosomal DNA purification method, as well as the development and validation of a ddPCR-based diagnostic assay for initial detection of CML.

Chapter 3 describes the development and testing of a new ddPCR-based diagnostic assay for *ALK*- positive NSCLC.

1.7 Purifying Genomic DNA from Tissue Specimens and Cell Lines

As exemplified by the *BCR-ABL* fusion gene, the DSBs associated with a translocation event are often not defined at a specific base pair, but rather occur anywhere within a large span of one or more of the participating genes. This fact, when coupled with the distributed intron/exon structures of genes, means that PCR-based assays against translocations must generally operate on relatively long fragments of chromosomal DNA, often more than 100 kbp in length. This is potentially problematic, as duplex gDNA is susceptible to shear, particularly during the mechanically intensive steps used to extract it from tissue or cultures cells. The probability of an unwanted mechanical break within the gene(s) to be interrogated depends on the nature of the shear events associated with sample processing and, as shear–mediated breakage of DNA is a stochastic process, fragment length [187].

A variety of methods and commercial kits are available to purify gDNA (e.g., Qiagen's Gene Read DNA FFPE kit**,** and QIAamp DNA FFPE Tissue Kit**,** Promega's Wizard Genomic DNA Purification Kit). Most of these follow the same basic sequence of processing steps – tissue collection, cell disruption/lysis, removal of proteins and other contaminants (gDNA purification), gDNA recovery, and gDNA re-solubilization. The precise procedure used in each step varies. For the gDNA purification step, for example, salt or alcohol-induced precipitation (Promega's Wizard Genomic DNA Purification Kit), liquid-liquid extraction (Life Technologies DNAzol), solid-phase extraction (Life Technologies Pure link genomic DNA kit, Qiagen Gene Read DNA FFPE kit), chromatography (most often anion-exchange chromatography) (DNeasy Blood and Tissue Extraction Kit), and various combinations of these operations have been proposed and used.

In this thesis work, as a minor objective, comparative analysis and tailoring of methods applied in each processing step will be used to establish a protocol for purifying gDNA from cultured cells, blood tissue, and formalin-fixed paraffin-embedded specimens of a quality suitable for ddPCR-based assaying of genomic events indicative of reciprocal and inversion translocations.

1.8 Digital PCR

1.8.1 Basic Principles and Data Analysis Methods

The ability to detect specific sequences of genomic or plasmid DNA (e.g., genes) and quantify their abundances has become an essential capability within the modern toolkit of molecular biology and clinical science. It permits absolute quantification of, for example, pathogen or viral adulterants in foodstuffs and drinking water [188, 189], as well as the detection of gene copy number variations or rare point mutations that enable genetic analyses of behavioral traits or disease [16, 190, 191].

The real-time polymerase chain reaction (real-time PCR or qPCR) method is most commonly used for this purpose, and a large number of commercial qPCR instruments are available, each offering certain performance features while operating on the same basic principles. In particular, they all link the amplification of a target DNA sequence (the "template") to the generation of a fluorescence signal that can be detected in real time over the course of a set number of PCR cycles. In a well-designed amplification reaction, the number of copies (amplicons) of the template increases at (nearly) 2^n, where n is the cycle number, with fluorescence intensity increasing in proportion to the amount of amplified DNA. Monitoring the change in fluorescence intensity with time (the amplification curve) permits the efficiency of the amplification reaction to be estimated. From that, the abundance of template in the original sample may be estimated.

The PCR cycle in which fluorescence is first detected with statistical certainty is termed the quantitation cycle (C_q) and is the central parameter of a qPCR experiment aimed at quantifying template abundance in the initial sample – a lower C_q value means a higher initial template concentration. Precise quantification typically involves comparing the C_q value for the sample either to the C_q value for a reference sample (internal control, leading to a calculation of ΔC_q, from which the relative abundance of template may be determined) or to a standard curve of known amounts of the target (permitting absolute quantification of copy number). The C_q value is typically also compared to that obtained when the target is not present (e.g., the background signal), in part to enable determination of detection limits. Uncertainties in C_q values vary, but are typically of order ± 0.2 – 0.5 cycles (standard deviation). This includes the uncertainty in the

C_q value for the template-free qPCR control. As a result of these uncertainties and other factors, the qPCR method, though tremendously useful, exhibits limitations, most notably when used to quantify small variations in copy number or, in certain cases, to detect target alleles or genomic alterations present at low abundance within a sample. For example, the qPCR-based Cobas™ 4800 BRAF V600 mutation assay is capable of detecting the V600E somatic point mutation in the *BRAF* oncogene, a known driver of metastatic melanoma and colorectal cancer, to mutational frequencies no lower that 5% [192].

First described by Vogelstein and Kinzler [193], who used the method to identify rare mutations in the *KRAS* gene within gDNA isolated from stool samples of patients with colorectal cancer, digital PCR (dPCR) overcomes many of the limitations of qPCR by re-imagining the manner by which the template is quantified. In a dPCR experiment, a gDNA sample is partitioned into a large set of elements, such that each element contains no, one or at most a few copies of the template (e.g., allele or allele fragment) to be amplified and thereby detected. The distribution of template copies among the elements is then reasonably well described by Poisson statistics. A PCR reaction is simultaneously conducted on the entire ensemble of partitioned templates. The fluorescence intensity of each element (e.g., droplet) is then recorded at the end of the PCR, typically comprised of 40 to 50 cycles. Thus, while qPCR requires accurate measurement of fluorescence intensity in real time over the entire time-course of the amplification reaction, dPCR is fashioned so as to require only the end-point fluorescence amplitude of each droplet. The total number of droplets with positive end-point fluorescence amplitude is then recorded. From that information a number of useful analyses based on Poisson statistics may be completed, starting with calculation of the average copies (of template) per droplet (*CPD*):

$$CPD = -ln(1 - \theta) \qquad 1.1$$

where θ is the fraction of total droplets having a positive end-point fluorescence (above a prescribed threshold value). From this *CPD*, the Poisson distribution gives the probability $p(n)$ that a given droplet will initially contain n copies of template

$$p(n) = \frac{(CPD)^n \, e^{-CPD}}{n!} \qquad 1.2$$

Equation 1.2 thereby allows the total copies of the target sequence in the initial sample to be estimated. The average concentration of template $c_{template}$ (copies/μL) in the initial sample may be computed as well through the relation

$$c_{template} = \frac{CPD}{V_{droplet}} = \frac{-ln(1-\theta)}{V_{droplet}} \qquad 1.3$$

where $V_{droplet}$ is the average droplet volume (μL).

A number of digital PCR instruments are now available commercially. Bio-Rad Inc. and RainDance Inc. market droplet digital (ddPCR) equipment, while Fluidigm Inc. and Life Technologies Inc. market microfluidic chip-based digital PCR (cdPCR) machines [194]. All of these instruments partition template copies into individual elements to permit their enumeration by end-point fluorescence detection (the two cdPCR machines are designed to also collect real-time data if required) [195, 196]. The manner by which the partitioning is conducted varies among instruments. In the case of the Bio-Rad QX100 ddPCR instrument used in this work, an emulsification reaction is used to partition initial copies of template among *ca.* 20,000 aqueous sub nL-sized droplets, each then holding zero, one or a few template copies as well as the necessary reagents for the amplification reaction.

Through this partitioning process, individual templates present at very low levels in the original sample are isolated and enriched, as each droplet in the emulsion represents an independent nL-scale PCR capable of amplifying a single template molecule to concentrations easily detected by end-point fluorescence. For that detection process, droplets are focused and read single file by a flow cytometry based detector to realize highly accurate enumeration of amplicon-positive and negative droplets, and θ, at high speed. Thus, in comparison to standard qPCR, ddPCR takes advantage of the ability to digitally count every copy of template to enable

accurate detection of either very small changes in copy number [16] or very small absolute numbers of template in the initial sample [197-199]. Both applications are made possible in part by the fact that isolation of a small quantity of gDNA in each droplet promotes favorable primer-template interactions and efficient amplification [200]. In quantifying template copy numbers by detection of end-point fluorescence above a threshold value, dPCR data quality is far less sensitive than qPCR results to any matrix effects that serve to inhibit amplification [201, 196]. Uncertainties in C_q values and their contribution to corresponding uncertainties in copy number estimates are avoided [195]. Besides, unlike qPCR, ddPCR does not in principle require a calibration curve or internal control to quantify copy numbers absolutely [202].

As with qPCR, combinations of dual labelled probes, where each carries a unique fluorescent reporter group coupled to an effective quenching agent, may be used to monitor amplification of multiple templates in the same reaction. The ddPCR data from an experiment in which two different targets are PCR amplified can be viewed in a 2-D plot in which the end-point fluorescence generated in each droplet by amplification-associated hydrolysis of probe 1 (a FAM-labeled probe for instance) is plotted against that generated by hydrolysis of the second probe (e.g., a HEX labeled probe). In a well-designed ddPCR experiment, each droplet will appear in one of 4 unique clusters: 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 [203].

When compared to current next-generation sequencing (NGS) platforms, digital PCR (dPCR) is far more sensitive and able to operate on much lower quantities of DNA and template [204, 205]. In addition, dPCR may be used to verify NGS sequencing results [206], indicating that dPCR is often complimentary to as opposed to competitive with NGS [207].

However, dPCR has limitations. In general, qPCR has a larger dynamic range [208], although the megapixel dPCR technology pioneered by Heyries *et. al*., [198] largely addresses this deficiency. Specialized mixtures (generally more expensive dPCR master-mixes) of PCR reagents are needed, and the design of a robust dPCR assay or experiment generally requires

more thought than does a typical qPCR experiment. Finally, understanding and accounting for the inherent failings and sources of bias in dPCR is essential for a successful assay [209].

1.8.2 Applications of Digital PCR Analysis in Cancer Diagnostics

The many unique attributes of dPCR have enabled its rapid development as a platform for analyses of oncogenes and other cancer biomarkers [210-212]. Through its ability to detect mutant alleles at frequencies as low as 1 copy in a background of 100,000 copies of the germline allele [195, 204, 213], dPCR has been used to detect rare somatic point mutations and more complex variations within one or more codons of a (proto)-oncogene [193, 214, 215]. Variations in allele copy numbers [195, 216-218] and/or loss of heterozygosity (LOH) [210] have also been quantified by dPCR. As a result, dPCR assays have found clinical use in the detection of, for example, breast cancer metastases [219], ocular infections [220], and fetal DNA in maternal plasma [221]. The T315I mutation in *ABL*, the presence of which results in ineffective treatment of CML with imatinib, has also been detected by RT-cdPCR [222].

The application of dPCR to the detection of translocation events has been minimal, with the only published example being that of Shuga et al., [223], who developed an assay against the t(14;18) translocation associated with follicular lymphoma. That method did not use dPCR to detect the translocation within gDNA. Rather, a digital PCR reaction was used to generate sufficient amplicon material to enable gel purification and deep NGS of the translocation. The direct application of dPCR to the detection of reciprocal or inversion translocations and the quantification of their frequency has never been demonstrated.

Chapter 2: Initial Diagnosis of Chronic Myelogenous Leukemia Based on Quantification of M-*BCR* Status Using Droplet Digital PCR

Formed from a reciprocal translocation t(9;22)(q34;q11) of genetic material between the long arms of human chromosomes 9 and 22, the constitutively active breakpoint cluster region (BCR) Abelson 1 (ABL) tyrosine kinase BCR-ABL is known to be causative of chronic myelogenous leukemia (CML). In 98% of CML patients harboring the t(9;22)(q34;q11) translocation, known as the Philadelphia chromosome, the chimeric *BCR-ABL* oncogene is created through cleavage of the *BCR* gene within its major breakpoint region (M-*BCR*), and breakage of the *ABL* gene within a 100 kbp region downstream of exon 2a. Clinical detection of the fused *BCR-ABL* oncogene currently relies on direct visualization by FISH, a relatively tedious assay that typically offers a detection limit of *ca.* 2%. This chapter describes a general method for using droplet digital PCR (ddPCR) technology to reliably detect translocation events. It is applied to the development of a new assay to measure M-*BCR* status and the presence of *BCR-ABL*. When applied to cell-line models of CML, the assay accurately quantifies *BCR-ABL* frequency to a detection limit of 0.25%. It consequently offers improved specificity relative to FISH, and may allow identification of variant translocation patterns, including derivative chromosome 9 deletions.

2.1 Introduction

Chronic myelogenous leukemia (CML), the most common chronic myeloproliferative neoplasm, affects approximately 6000 patients in the USA [9], with 140,000 newly diagnosed cases worldwide each year [224]. Clinically, CML pathology is described in three distinct phases. In the *chronic* phase, which generally lasts several (~ 6 to 8) years, myeloid progenitor cells expand and differentiate in an apparently normal fashion, with a gradual increase in immature and mature myeloid elements observed in bone marrow and blood. The leukemia may then progress through an *accelerated* phase (increased blasts and basophils) or directly to an acute phase, known as the *blast* phase or blast crisis, in which greatly increased proliferation of abnormal white blood cells, including 20% or more blasts in the bone marrow or blood, is observed along with loss or inhibition of tumor suppressor genes and activation of pathways

influencing myeloid differentiation [44]. The diagnostic hallmark of CML is an oncogene fusion formed from a reciprocal translocation (t(9:22)(q34;q11)) between chromosomes 9 and 22 that results in an altered chromosome 22q known as the Philadelphia chromosome (PhC). Approximately 95% of all CML patients harbor the gene fusion, *BCR-ABL* [10-12], which is formed via a double stranded break (DSB) within both the Abelson oncogene 1 (*ABL*) on chromosome 9q, which codes for a non-receptor tyrosine kinase (ABL), and the breakpoint cluster region gene (*BCR*) on chromosome 22q [50]. The DSB in *BCR* usually occurs (~ 98% of the time) somewhere within a region bounded by exons e12 and e16, the so-called major breakpoint region (M-*BCR*) [45], while that in *ABL* generally occurs somewhere within exon a2, though a breakpoint in other regions of *ABL* can occur [75]. The DNA repair mechanism then joins the exchanged chromosomal sections to form a *BCR-ABL* chimeric oncogene comprised of the cleaved 5′ fragment of *BCR* fused to the 3′ fragment of *ABL* (A derivative chromosome 9 that encodes a reciprocal *ABL-BCR* fusion gene within 9q is typically also formed [78]). *BCR-ABL* encodes a constitutively active tyrosine kinase BCR-ABL responsible for the uncontrolled proliferation associated with CML [10, 57].

The work-up for individuals exhibiting symptoms consistent with the chronic phase of CML begins with histo-pathologic analyses (complete blood count, peripheral blood smear) to determine counts and morphologies of white blood cells (WBCs) and platelets. When elevated WBC counts (e.g., ≥ 20,000 per μL), blasts, and possibly abnormal platelet counts are observed, a bone marrow specimen may be collected and used to test for the *BCR-ABL* oncogene and other irregularities within the myeloid lineage. Initial clinical detection of *BCR-ABL* is generally achieved using FISH [225, 226]. In its standard (S) format, the cytogenetic S-FISH (hereafter FISH) assay utilizes two long, uniquely labeled oligonucleotide probes that together span the most common chimeric-gene fusion points. Granulocyte nuclei staining with the probes thereby permits detection of *BCR-ABL* via the fused signal created from co-localization of the two reporters on a derivative chromosome 22q. In addition to its clinical application to initial diagnosis of CML, FISH and alternative dual-fused-signal forms (D-FISH) have been used to diagnose questionable cases where cytogenetics failed, investigate more complex PhC rearrangements, and detect cryptic *BCR-ABL* fusions [163]. However, drawbacks and limitations to applying FISH to *BCR-ABL* detection are well known and include a limit of detection (LOD) of ~ 2% (FISH, somewhat lower for D-FISH) [227], and relatively high costs due in part to the

time, labor [167] and high level of technical expertise required [228]. The development of a reliable alternative that offers improved performance at lower cost could therefore be of significant clinical value.

Alternative genotyping technologies include next generation sequencing (NGS) [229, 230], microarray-based screening [231, 232], immunohistochemical methods [233], and various modalities of the polymerase chain reaction (PCR) [234, 235]. Each has proven effective in detecting and quantifying oncogenic biomarkers, including somatic point and codon-hotspot mutations [236-238], copy number variations [239-241], and loss of heterozygosity events [242, 243]. Applications of these platforms to the detection of reciprocal translocations within genomic DNA (gDNA) are considerably more limited. In particular, though demonstrations of assaying balanced chromosomal rearrangements by NGS or microarrays have been reported [244, 245], neither approach has found clinical use, likely due, at least in part, to the current costs and turnaround times of those methods relative to FISH.

In patients for whom a *BCR-ABL* gene fusion has been detected, quantitative RT-qPCR may then be used to monitor breakpoint-specific *BCR-ABL* mRNA during treatment with TKIs as a surrogate metric of MRD and disease relapse [97]. As the specific BCR-ABL construct produced by a CML patient undergoing treatment is generally not known, a limited panel of forward and reverse primers is screened to identify a pair capable of amplifying the patient-specific *BCR-ABL* cDNA. Calibration standards, though problematic, may then be used in conjunction with RT-qPCR to estimate *BCR-ABL* abundance through comparison to the mRNA level of a selected control gene [246]. Recently, a method for instead quantifying *BCR-ABL* fusion transcripts by digital PCR has been described [247], offering the advantage of providing absolute copy number values, which can serve to make use of standard curves potentially unnecessary.

Here, a method is presented that enables ddPCR technology to be used effectively to detect and quantify chromosomal rearrangements associated with a translocation event. That method is applied to the problem of reciprocal translocation detection through the development of a new assay that may be used in lieu of FISH to provide a reliable and sensitive measure of M-*BCR* status within gDNA isolated from various cultured cells either positive or negative for *BCR-ABL*. The assay exploits the fact that DSBs in *BCR* are associated with formation of both a PhC

and *BCR-ABL* in over 90% of all CML cases [248]. Most of the remaining CML patients harbor an alternative translocation, typically still comprised of 9q34 and 22q11, but containing one or more additional elements from other chromosomal regions. These patients generally also test positive for *BCR-ABL*, though rare *BCR* fusions to other chromosomes and gene fragments have been reported [249-251]. In approximately 98% all *BCR-ABL* positive CML cases, the DSB occurs within the M-*BCR*. M-*BCR* status correlates with *BCR-ABL* in the vast majority of CML cases.

The general method described here is based on amplifying three unique short (100 to 150 bp) sequences within a target gene, in this case *BCR*, with the presence of amplicons generated from a given sequence detected using a uniquely labeled hydrolysis probe. Two of those sequences flank the 5′ and 3′ borders of the M-*BCR*, respectively. Co-localization within a droplet of end-point fluorescent signals generated by amplification of these two sequences thereby allows one to detect and enumerate those droplets harboring an intact M-*BCR*. In the platform validation and assay development work described here, the droplets are loaded to an average copies-per-droplet (CPD) of *ca.* 0.2, so that only *ca.* 0.1% of all droplets contain more than two copies. Virtually all droplets contain either one or no copies of intact M-*BCR*.

Given its length (~ 7 kbp), the M-*BCR* can fragment mechanically (e.g., shear-induced fragmentation) during the required gDNA isolation process. A novel probabilistic model integral to the platform is presented to quantify the copies of M-*BCR* lost due to non-biological fragmentation. When combined with Poisson statistics and knowledge of the total copies loaded, that model permits the frequency of biological disruption of the M-*BCR* in various *BCR-ABL* positive cell lines to be quantified accurately and sensitively. Comparison to FISH results for the same cell lines is used to show that the measured frequency of M-*BCR* disruption quantitatively matches the *BCR-ABL* frequency across the full dynamic range. The digital assay can thereby be used to accurately identify and quantify *BCR-ABL* to a detection limit of 0.25% at a CPD of 0.2 – 0.3. Relative to FISH, it offers a significant improvement in sensitivity in a lower cost, more time-efficient procedure. It may therefore find clinical use in initial diagnoses of CML, and possibly MRD monitoring, through detection of biological DSBs in *BCR* isolated from either bone marrow or blood cells.

By utilizing the unique advantages of digital PCR [16, 197, 210, 211], most notably template partitioning and lack of need for external (calibration) standards, the platform and associated assay add to the growing list of applications of digital amplification to cancer diagnostics, which include detection of oncogene copy number variations, rare mutations in liquid biopsy and transplantation, alterations to genome structure (e.g., loss of heterozygosity), somatic mosaicism, and single nucleotide polymorphisms and related markers of inherited disease or disease risk [195, 214-218, 252]. Relative to those more active areas of digital PCR technology development, the use of digital amplification to detect balanced chromosomal rearrangements is limited, likely due in part to the added complexity of analyzing these large genomic alterations. The only significant demonstration is that of Shuga et al*.,* [223], who developed an elegant platform that combines multiple rounds of hemi-nested PCR and NGS to enable detection of individual copies of the t(14;18) balanced translocation associated with follicular lymphoma. As it requires custom-built microfluidics, multiple rounds of PCR using hemi-nested sets of primers, as well as amplicon sequencing capabilities, that technology is well positioned to become a powerful research tool, but is likely too costly and complex for clinical adoption at this time. Moreover, in that work dPCR is used only for template amplification; detection of *BCR-ABL* is achieved by sequencing. This work shows for the first time that ddPCR performed on a standard commercial droplet digital PCR instrument (the Bio-Rad QX100 system; Hercules, CA) can be used to detect and quantify reciprocal translocation events in a reliable and cost effective manner suitable for clinical adoption.

2.2 Materials and Methods

2.2.1 Oligonucleotides

All primers and dual-labeled hydrolysis probes were purchased from IDT, Inc. Probes were HPLC purified, while primers were purified by desalting. Purified primers and probes were re-suspended to 100 μM in IDTE (10 mM Tris, pH 8.0, 0.1 mM EDTA) buffer and stored at -20 °C prior to use.

2.2.2 Cell Lines

KU812 and MEG01 cell lines were purchased from ATCC, while K562 and HL60 cell lines were provided by the British Columbia Cancer Agency. The *BCR-ABL* positive KU812 cell line[1] was established from the peripheral blood of a 38-year-old CML patient in early blast crisis [253]. Through passaging, it has altered karyotype to present two distinct clones, with the more abundant (~ 80%) containing two Philadelphia chromosomes and the other presenting one classic PhC and one acrocentric marker containing the *BCR-ABL* oncogene fusion at elevated copy number [254, 255]. The K562 cell line, originally derived from a pleural effusion of a 53-year-old CML patient in blast crisis, has been a widely used model in CML research [256, 257]. The highly passaged K562 cell line assayed here[2] is hyper-diploid, displays two clones during culturing, and is highly mutated in terms of further chromosomal arrangements [254, 258]. The K562 used is a PhC-negative/*BCR-ABL* oncogene positive cell line that does not harbor the classic PhC, has four copies of chromosome 22 per nuclei, two being marker chromosomes containing multiple repeats of *BCR-ABL* [259]. Despite the substantial chromosomal rearrangements noted, the breakpoint sequence appears to be conserved in the majority (if not all) *BCR-ABL* fusions in K562, which has permitted its use as a blast crisis model [260, 261]. The MEG01 cell line[3] was established from bone marrow of a 55 year old patient in blast crisis. It is positive for *BCR-ABL* [262] through a classic PhC carrying two copies of *BCR-ABL*, and an acrocentric chromosome containing one *BCR-ABL* copy [67, 260]. The PhC negative and *BCR-ABL* negative HL60 cell line served as a negative control.

---

[1] KU812 (60,XYY,-2,-3,add(4)(p11),-5,+6,-7,+8,-9,t(9;22)(q34;q11.2)x2,-10,i(11)(q10),-12,-16, del(17)(q11.2q24),i(17)(q10),-18,+19,-20,-22.ish del(17)(TP53+),add(4) (p11)(wcp4-)(18)/60,sl,+19)

[2] K562 (67~70,XX,-X,add(2)(q3?3),-3,+5,add(5)(q11.2),ins(6;?)(p21;?),+7,der(7)t(7;7) (p1?1.1;q22), -9,del(9)(p13), der(9)t(9;9)(p1?3;q22), der(10)t(3;10)(p21;q2?3),-13, add(13)(p11.2),-14,+17,der(17) t(10;17)(q11.2;q11.2)der(10)t(3;10),der(17)t(17;20) (p11.2;p11.2),+19,-20,?der(21)t(1;21) (q21;p11.1) -22,+4mar.ish add(2) (BCR+,ABL1+,BCR con ABL1x1), add(5)(D5S23+,EGR1-), der(9)t(9;9) (ABL1++),der(17)t(17;20) (TP53-), der(17)t(10;17)der(10)t(3;10)(p53+),mar1(BCR+,ABL1+, BCR con ABL1++++), mar2(BCR+,ABL1+,BCR con ABL1++++)[cp7])

[3] MEG01 (97~100<4n>,XX,-Y,-Y,-1,der(1)t(1;2)(p21;q21), der(2;14)(q10;q10)x2,+6,+6,-7,+8,+8, add(8)(p13),-9,t(9;22)(q34;q11.2)x3,-10,-10,-10,add(10)(p12),+11,del(11)(q13q23)x2,-12,-13,-13, add(13)(q34),+17,+19,+19,+19,+19,+19, add(19)(p13.1)x2,add(19)(q13.1),+21,+21,-22,+der(22)t(9;22), +9mar[cp8])

All cell lines were cultured in HyClone RPMI 1640 media (GE Healthcare), with 10% fetal bovine serum, 1% glutamine and 1% penicillin/streptomycin (all from Invitrogen Canada).

2.2.3 Genomic DNA Purification

The assay described in this work interrogates two adjacent 7 kbp regions of the *BCR* gene; the limit of detection (LOD) can be diminished, in part, by excessive mechanical fragmentation of those regions during genomic DNA (gDNA) purification from a cell culture or tissue specimen. Various extraction protocols were evaluated, including the Qiagen Gene-read DNA FFPE kit, which utilizes a column extraction method, and the Qiagen MagAttract HMW DNA kit that exploits the use of magnetic beads. None proved suitable, as on average greater than 60% of M-*BCR* copies in gDNA recovered from *BCR-ABL* cell lines were lost to non-biologic fragmentation (shear) during purification (the method used to determine % non-biologic fragmentation is described in detail in the Results section of this chapter).

A new gDNA extraction method was empirically developed that uses precipitation, minimal pipetting and no vortexing to minimize shear-induced fragmentation. In addition, all centrifugation steps following DNA extraction are performed at 2000 g. The method thereby delivers M-*BCR* with less mechanical fragmentation, but at a potentially lower total DNA yield than can be achieved using more mechanically aggressive commercial kits. Due to the gentler low-g/precipitation approach used, typically only 15% to at most 30% of either of the 7 kbp regions interrogated is lost via non-biologic fragmentation.

This method (Figure 2-1) begins by collecting cells (approximately 5 x 10^6) cultured in 20 mL of total media by centrifugation (250 g) to a final concentrate volume of 50 µL. The cell concentrate is then washed 6X with phosphate buffered saline, and re-concentrated to 50 µL by strong centrifugation (12,000 g for 10 secs) and supernatant removal. The gDNA is extracted by re-suspending the washed concentrate in 600 µL of nuclei lysis solution (Promega), to which proteinase K (20 µL; > 600 mAU/mL solution) and RNase A (2 µL; 100 mg/mL solution) are added (Qiagen). The mixture is incubated overnight at 56 °C, after which 200 µL of protein precipitation solution (Promega) is added. That mixture is then incubated on ice for 10 min to facilitate precipitation of proteins, which are pelleted by centrifugation (2000 g). The gDNA

supernatant is decanted and recovered, to which is added 600 µL of 100% iso-propanol (Sigma) to precipitate and recover the purified gDNA, as well as 2 µL of glycogen (10 µg/ml) to mark the location of the DNA pellet. The gDNA is recovered, washed with 600 µL of 70% ethanol, and then air-dried. The purified gDNA is then re-suspended in 50 µL of DNA rehydration solution (Promega) and stored at 4 °C prior to use to avoid freeze/thaw steps that might damage gDNA.

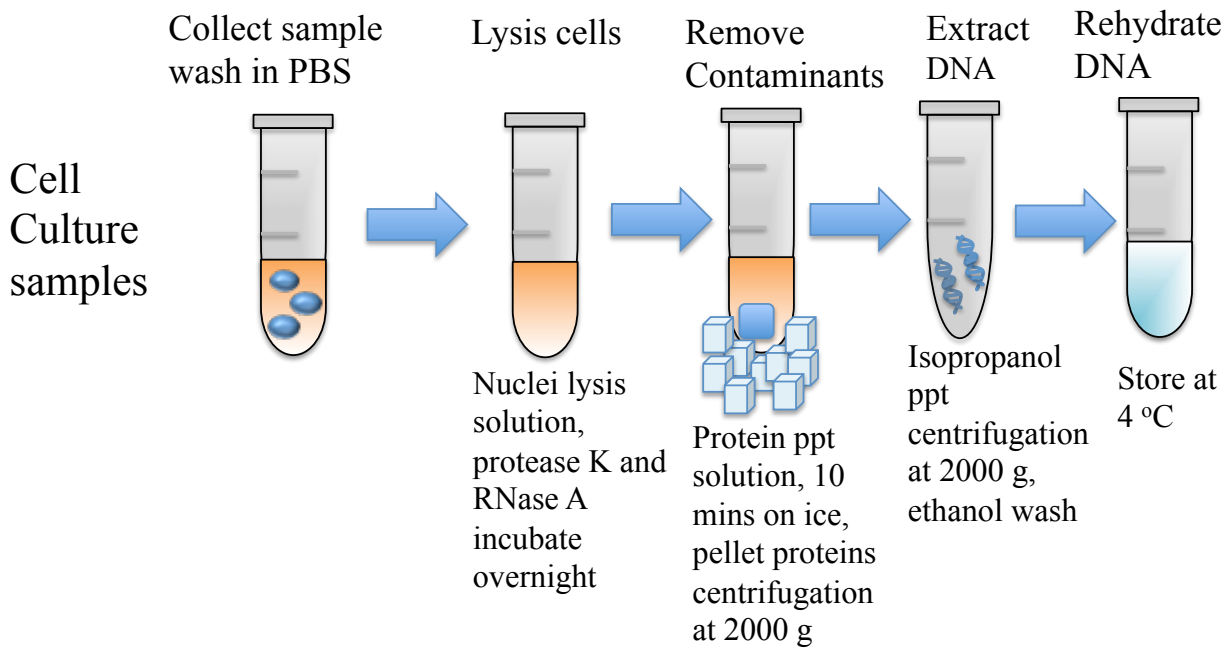


Figure 2-1. Genomic DNA extraction method.
A two-step precipitation and extraction method was developed to purify gDNA from cell lines in order to minimize any mechanical disruption of the gDNA. There was no vortexing, centrifugation steps were minimized and conducted at low g, and minimal pipetting.

2.2.4 Primer and Probe Design

The human *BCR* sequence was taken from the NCBI database (sequence NG_009244.1; http://www.ncbi.nlm.nih.gov/). Forward (FP) and reverse (RP) primers were designed using Primer3 software (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). All primers were then analyzed by primer-BLAST to find any sequence similarities within the human genome database (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Possible common (> 1% minor allele frequency) single nucleotide polymorphisms that might diminish primer (or probe) performance were analyzed using the human genome database browser (http://genome.ucsc.edu/cgi-bin/hgGateway), while the self-complementarity of various possible primer and probe pairs was scored using the Exiqon OligoAnalyzer tool (https://www.exiqon.com/oligo-tools). When annealed to their fully complementary template, primers and probes were designed to melt at 60 – 63°C and 66 – 70°C, respectively, under PCR conditions (50 mM K^+, 3 mM Mg^{2+} and a total strand concentration (C_T) of either 250 or 900 nM) (http://biophysics.idtdna.com/). An annealing temperature (T_a) gradient study was then used to optimize T_a (60 °C) for the assay.

2.2.5 Droplet Digital PCR (ddPCR) Assay Workflow

Samples (20 µL) for ddPCR analysis were prepared from 2X dUTP-free ddPCR™ Supermix for probes (Bio-Rad), 900 nM of each FP and RP, 250 nM of the FAM and the HEX dual-labeled hydrolysis probe, 325 nM of the Alexa dual-labeled probe, and an amount of purified gDNA, typically 6 – 10 ng, where DNA concentrations were measured using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific). Addition of gDNA was omitted for no template control (NTC) reactions. The concentration of the Alexa probe was increased to a value 1.5x that of the FAM probe to ensure segregation of the cluster outputs in the ddPCR 2D output graph.

The 20 µL ddPCR mixture was loaded into the sample well of the DG8 eight-channel disposable droplet generator cartridge (Bio-Rad); 60 µL of droplet generation oil was then loaded into the oil well for each channel. The cartridge was processed in a QX-100 Droplet Generator (Bio-Rad) to generate the droplet emulsion with a total CPD of 0.2 – 0.3. A 40 µL sample of the

emulsion containing approximately 12,000 – 15,000 readable droplets was transferred by multichannel p100 pipette to an Eppendorf Twin.tec semi-skirted 96-well PCR plate, which was then heat-sealed with foil sheets (Bio-Rad). The droplet emulsion samples were amplified in a CFX96™ thermocycler according to the following protocol: 95 °C for 10 min, followed by 50 cycles of 94 °C for 30 sec and 60 °C (T_a) for 1 min. Finally, the droplets were stabilized by heating to 98 °C for 10 min. End-point fluorescence readings within each droplet were measured using the QX100 Droplet Reader (Bio-Rad), and raw ddPCR data collected and visualized using the QuantaSoft v1.2 program (Bio-Rad).

For dynamic-range and limit of detection evaluation, serial dilutions of *BCR-ABL* positive gDNA isolated from KU812 cells into gDNA from *BCR-ABL* negative HL60 cells (total *BCR* held constant at *ca.* 2000 copies) were prepared in IDTE buffer (IDT), with the total concentration of *BCR* template in each dilution quantified on a Bio-Rad QX100 ddPCR instrument.

2.2.6 *BCR-ABL* FISH Assay

Cells were cultured as described for 24 h, and then prepared for fixation by arresting during division using colcemid. Aliquots of ~ 3000 - 5000 arrested cells in 10 mL RPMI 1640 culture medium were centrifuged (250 g) for 10 min and the supernatant removed. The cells were re-suspended in 10 mL of hypotonic solution (37 °C) and incubated for 25 min. Excess hypotonic solution was then removed by centrifugation (250 g for 10 min) and the cells fixed in up to 10 mL of a 3:1 methanol to acetic acid solution. That fixing process was repeated three times, with the resulting fixed cells then stored at 4°C until use. When required for FISH, samples were centrifuged and supernatant removed to leave ca. 0.75 mL. The cells were re-suspended in fixative to a final volume of *ca.* 1 mL to achieve an opaque suspension. Aliquots of 8 µL were spotted onto glass microscope slides (Leica Surgipath Snowcoat Precleaned 1 x 3 x 1mm). FISH was then performed using the Vysis LS1 *BCR-ABL* dual-color, extra signal translocation probe (Abbott Molecular) according to the manufacturer's instructions. Interphase nuclei (200) were analyzed using a Zeiss Axio-imager Z2 epifluorescence microscope with a triple-band pass filter and DAPI counter stain added to aid visualization of nuclei.

2.2.7 Statistics

Basic metrics of assay performance (LOD, limit of blank (LOB), confidence interval (CI), coefficient of variation (CV), dynamic range) were defined by computing means and standard deviations (SDs) from replicates. Statistical errors and the significance of recorded frequencies of biological disruption of *BCR* were determined using a paired Student's *t*-test. The error associated with the CPD ($= -ln(empty\ droplets/total\ droplets)$) was estimate using the method described in Dube et al., [190].

2.3 Results

2.3.1 Platform Concept

Oncogenic chromosomal translocations often comprise of gene rearrangements that reorganize the sequences and change the locations of proto-oncogenes, which may result in fusion genes coding active unnatural gene products. Such translocations are complex biological processes, but irrespective of whether they are reciprocal or non-reciprocal, they include two essential steps. First, DSBs on the same or different chromosomes occur simultaneously or nearly simultaneously at the two loci. Second, ends created by the opposing DSBs ligate together. The key concept behind the platform developed in this thesis work, applies ddPCR technology to the detection and quantification of the rearrangement of a proto-oncogene that is an integral part of the translocation event (Figure 2-2). Typically, the location of the DSB within the oncogene is ill-defined, occurring somewhere within a region of the gene, referred to as the breakpoint region, this may span several introns and exons, and thus several kbp. Amplification of the entire breakpoint region, at least with a single pair of forward and reverse primers, is therefore not possible. In theory, this problem can be addressed by a large set of nested primers covering the entire breakpoint region, and this concept has recently been explored in a qPCR format [175]. However, that approach has obvious drawbacks, including the required massive multiplexing (across many plates) of the qPCR reaction, and the lack of assurance that a suitable combination of primers will be discovered.

The alternative method described in this thesis defines templates, each comprised of highly conserved sequences, at the two boundaries of the breakpoint region (Figure 2-2). The unique attributes of ddPCR are then exploited by partitioning individual copies of the target proto-oncogene, either fully intact or fragmented due to a DSB, into isolated droplets. Droplets containing the intact gene display end-point signals from both amplicons, while those that carry a fragment will display only the end-point signal. As demonstrated below, the method allows enumeration of the frequency of DSBs in the gDNA sample.

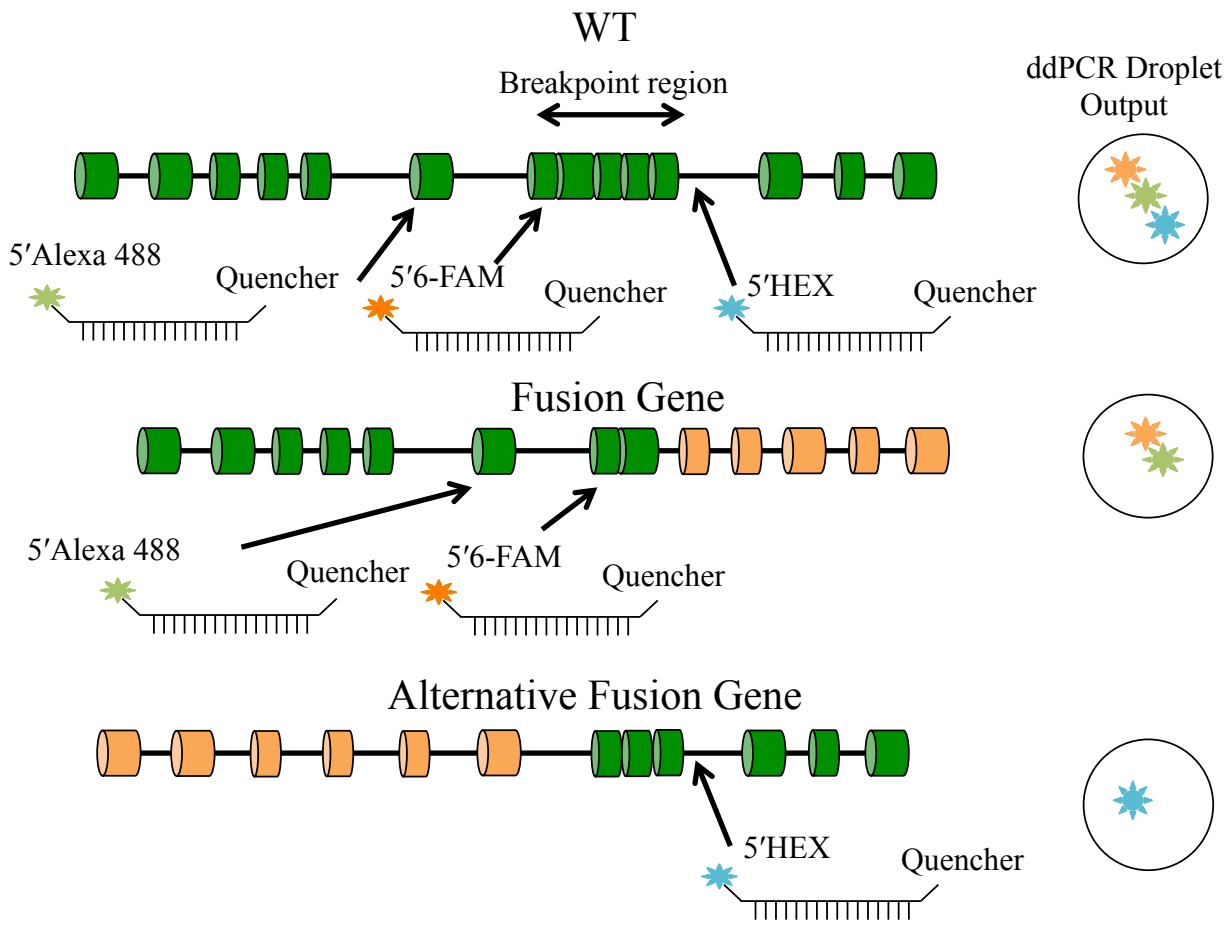


Figure 2-2. Triple probe platform concept.
Detailed platform concept showing 3 uniquely labeled probes used to identify a biological and a non-biologic DSB. The ddPCR 2D output contains 3 fluorescent probes indicative of intact gDNA. An Alexa 488 and a FAM signal with a separated HEX signal is indicative of a biological DSB.

2.3.2 Assay Design

Here, the method described in section 2.3.1 is used to design an assay that utilizes ddPCR to partition and amplify three specific segments of *BCR* (Figure 2-3) proto-oncogene to quantify the fraction of total M-*BCR* harboring a DSB. The first two reactions are designed to amplify short (100 to 150 bp) segments, 7 kbp apart, lying immediately to the 5′ (within *BCR* intron 11) and 3′ (within intron 16) sides of the M-*BCR*. Poisson statistics, in combination with counts of droplets in which the end-point fluorescence signals from the two released dyes (FAM and HEX) co-localize, can then be used to quantify intact M-*BCR*. As noted above, segregation of FAM and HEX signals between different droplets typically then reflects biological disruption (a DSB) within the M-*BCR*, but may in some instances instead reflect non-biologic (e.g., shear-induced) disruption during sample processing [263, 264]. These two M-*BCR* disruption mechanisms – biological and non-biologic – are differentiated in the assay through amplification of a third short segment of *BCR* (control region, detected with an Alexa-488 labeled probe) lying the same distance from (7 kbp; within intron 8), but now to the 5′ side of the FAM-detected sequence immediately upstream of the M-*BCR*. Biological DSBs within this second 7 kbp stretch of *BCR* have not been reported [50]. Thus, as shear-induced breaks within gDNA are known to arise stochastically [187] at a length-dependent frequency [265], quantification of segregated Alexa and FAM signals may be used as a surrogate to estimate the frequency of non-biologic disruptions within the M-*BCR*. The resulting data sets can then be analyzed using a probabilistic model described below, so as to leverage the unique ability of ddPCR to partition signals from the three equidistantly spaced amplification reactions to enumerate states of the M-*BCR* arising from either non-biologic or biologic fragmentation, respectively.

Optimized primers and the dual-labeled probe used in each amplification reaction are reported in Table 2-1, with their positions within *BCR* and relative to the M-*BCR* shown in Figure 2-3.

Table 2-1. Primer and probe sequences used in each amplification reaction comprising the ddPCR based M-*BCR* status assay.

Amplicon Template	Reagent	Sequence
117 bp segment of intron 8	FP	5′-aggagcatagccagcatcat-3′
	RP	5′-aaatcatcatccccacagga-3′
	Probe[£]	5′-(Alexa-488)- tcggctcattcccaaaggaa -(BHQ1)-3′
133 bp segment – 236 bp upstream of exon 12	FP	5′-attgaatgcaggaggtcagg-3′
	RP	5′-acaccatctctcacccgaac-3′
	Probe[£]	5′-(6-FAM)-ctgccagcatcacaccctga-(BHQ1)-3′
142 bp segment – 214 bp downstream of exon 16	FP	5′-ggctctgaaacatccatcgt-3′
	RP	5′-cagctgcaaaaccaagttga-3′
	Probe[£]	5′-(HEX)- aagatgcaggctgtcctggc -(BHQ1)-3′

[£] Alexa-488 (Alexa Fluor® 488), 6-FAM (6-carboxyfluorescein), HEX (hexachloro-fluorescein), and BHQ1 (Black Hole Quencher® 1)

2.3.3 Detecting M-*BCR* Status in *BCR-ABL* Positive and Negative Cell Lines

Raw data from application of the platform-derived ddPCR-based M-*BCR*-status assay to gDNA isolated from the KU812 (*BCR-ABL* positive) or HL60 (*BCR-ABL* negative) cell line are shown as two-dimensional (2D) output plots in Figure 2-4. A negative (empty) droplet cluster is observed in the lower left quadrant, along with seven distinct non-overlapping clusters positive for at least one of the reporting fluorophores. In the *BCR-ABL* positive KU812 cell line, loss of normal chromosome 22q11 through a concomitant double-stranded break in the M-*BCR* results in a sparsely populated $Alexa^{+}FAM^{+}HEX^{+}$ cluster, indicating intact M-*BCR* is present in very low abundance (Figure 2-4A), and thus that repeated passaging of this cell line has resulted in genomic heterogeneities. Dense $Alexa^{+}FAM^{+}$ and HEX^{+} clusters are also observed due to a combination of biological (i.e., DSB) and mechanical disruption of the M-*BCR*, with the lower

density of the Alexa^{+}FAM^{+} cluster (715 droplets) relative to the HEX^{+} cluster (1272 droplets) providing a coarse indication of the level of non-biologic disruption.

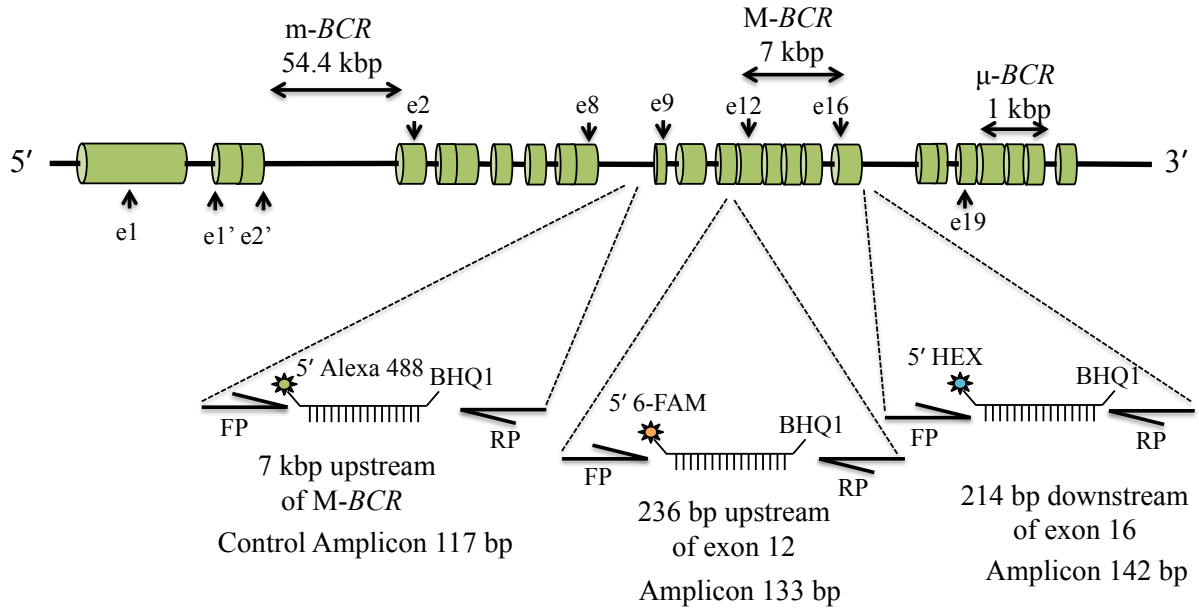


Figure 2-3. M-*BCR* status assay reactions schema showing amplification templates used to detect biological cleavage and mechanical fragmentation within the M-*BCR*.
The intron (i)/exon (e) structure of the human breakpoint cluster region (*BCR*) gene on chromosome 22q (NCBI Reference Sequence: NG_009244.1) is shown. The assay amplifies three equidistantly spaced templates: the first (133 bp, within i11, 236 bp upstream of e12, and detected using a 5′6-FAM labeled dual-hydrolysis probe) is positioned on the immediate 5′ side of the M-*BCR* (major breakpoint region – 7 kbp sequence spanning e12 through e16); the second (142 bp, within i16, 214 bp downstream of e16, detected using a 5′HEX labeled probe) lies to the immediate 3′ side of the M-*BCR*; the last, a control template (117 bp, within i8, 7 kbp to the 5′ side of the first (6-FAM) template, detected using a 5′Alexa Fluor 488 labeled probe), is used to quantify the degree of mechanical fragmentation of the M-*BCR* in the purified gDNA sample. The fraction of the total copies of *BCR* showing fragmentation between templates 3 (i8) and 1 (i11), detected in the assay by segregation of Alexa and FAM signals among droplets, is used as a surrogate to quantify the frequency of shear events within the M-*BCR*. The minor (m-*BCR*) and micro (μ-*BCR*) breakpoint regions of *BCR* are not interrogated within this assay.

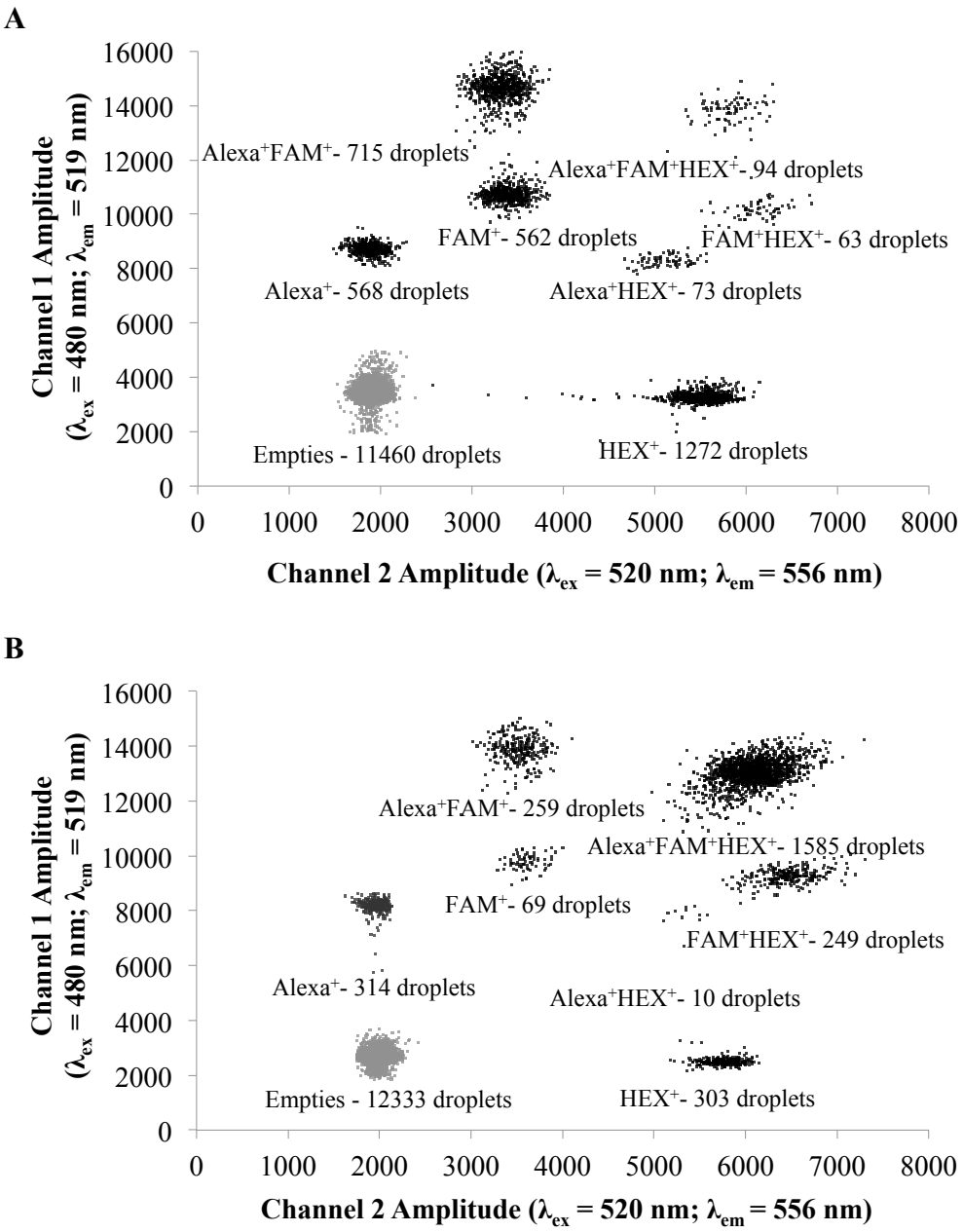


Figure 2-4. Raw 2D data output from the M-*BCR* status assay applied to the (A) *BCR-ABL* positive KU812 and (B) *BCR-ABL* negative HL60 cell lines.
The primers and probes listed in Table 2-1 were applied. For the *BCR-ABL* positive KU812 cell line (A), total read droplets = 14807 and CPD = 0.256. The sparsely populated $Alexa^{+}FAM^{+}HEX^{+}$ cluster (94 droplets), and dense $Alexa^{+}FAM^{+}$ and HEX^{+} clusters indicate disruption of the M-*BCR* in most copies of *BCR*. For the *BCR-ABL* negative HL60 cell line (B), total read droplets = 15123 and CPD = 0.206. The highly populated (1585 droplets) $Alexa^{+}FAM^{+}HEX^{+}$ cluster indicates *BCR* in the sample is mainly fully intact M-*BCR* (Figure 2-3), with the more sparse populations recorded in the remaining clusters due to random mechanical fragmentation (shearing) of M-*BCR* during sample processing.

In contrast, raw 2D output data for the *BCR-ABL* negative HL60 cell line (Figure 2-4B) show a densely populated $Alexa^{+}FAM^{+}HEX^{+}$ cluster, as well as sparse $Alexa^{+}FAM^{+}$ and HEX^{+} clusters. The sum of droplet counts (562 droplets) in the $Alexa^{+}FAM^{+}$ cluster and HEX^{+} cluster now equals that of the $FAM^{+}HEX^{+}$ cluster and $Alexa^{+}$ cluster (563 droplets) due to the fact that either sum reflects only non-biologic disruption of the M-*BCR* in this case. Thus, without the aid of any data processing, direct visual interpretation of the raw 2D data output from the assay provides clear evidence of disruption of *BCR* due to a biological DSB within the M-*BCR*.

2.3.4 Model-Based Quantification of M-*BCR* Status

While M-*BCR* status can be qualitatively defined by comparing droplet counts in the primary ($Alexa^{+}FAM^{+}HEX^{+}$, $Alexa^{+}FAM^{+}$ and HEX^{+}) clusters of the raw 2D output plot, the added information content within the distinct minor droplet clusters ($Alexa^{+}$, FAM^{+}, $FAM^{+}HEX^{+}$, and $Alexa^{+}HEX^{+}$) provides a means to reliably quantify the fraction (frequency) of total copies in the sample that have been disrupted through a biological DSB. For example, droplets in the FAM^{+} cluster arise from a combination of a mechanical break in the 7 kbp control region upstream of the M-*BCR* and either a biologic or non-biologic break within the M-*BCR*; counts of $Alexa^{+}HEX^{+}$ droplets, which can only arise through random co-partitioning of the Alexa-probed sequence and the HEX-probed sequence (Figure 2-1), provide a checkpoint of Poisson statistics, as the frequency of $Alexa^{+}$ and HEX^{+} signals appearing in the same droplet should satisfy the Poisson distribution for those two independent signals. Relative to more traditional real-time PCR (qPCR) or FISH assays, ddPCR-based analysis of M-*BCR* status, and by extension of other complex genomic rearrangements, can provide a significantly richer description of the various states of the assayed target through its partitioning capabilities.

A probabilistic model was derived that uses the full 2D output data set to accurately quantify M-*BCR* status by first defining all possible states of the target, and then connecting those states to the observed droplet counts within each cluster. Figure 2-5 reports a diagram of all possible states of the M-*BCR* and the adjoining (7 kbp) control region, and the sets of droplet clusters that can be populated by each of those states. The "tree" of possible states begins with segregation of *BCR* copies into those that have (b) or have not ($\bar{b}$) undergone a biological DSB

within the M-*BCR*. For each of these two primary branches, all states arising from shear-induced (*s*) non-biologic disruption within either the M-*BCR*, the 7 kbp control region flanking the M-*BCR*, or both (*ss*) are then considered. As exemplified by the K562 cell line, blast crisis can result in loss of derivative chromosome 9 and the 3′ fragment of *BCR* downstream of the M-*BCR* that contributes to *ABL-BCR*. Loss of HEX^+ droplets due to loss (*l*) or disruption of that fragment (Figure 2-3 may thereby arise (recorded by FISH as a loss of derivative chromosome 9 [75]) and is also considered in the model. Finally, more than one copy of M-*BCR* can partition into a single droplet and be recorded as single event within the $Alexa^+FAM^+HEX^+$ cluster, particularly at CPD values above *ca.* 0.35. Poisson statistics and droplet counts are used in the model to quantify all hidden signals within each cluster (See Appendix A). At a CPD of approximately 0.2, which was employed in the development work for this assay to help define the minimum number of biologically disrupted M-*BCR* that can be reliably detected, the proportion of droplets containing three or more copies is minimal (~ 0.1%). In this case, droplets are analyzed with a model that considers no more than 2 copies per drop.

Total copies of *BCR* (taken as the average of all Alexa-positive droplets and all FAM-positive droplets) loaded may be combined with the diagram of possible states (Figure 2-5) to quantify M-*BCR* status. The contributions of each possible state of the M-*BCR* and adjoining control region to the set of normalized data clusters are given by

$$\frac{Alexa^+FAM^+HEX^+}{Total\ BCR} = (1-b)(1-s)^2 \tag{2.1}$$

$$\frac{FAM^+HEX^+}{Total\ BCR} = s(1-b)(1-s) \tag{2.2}$$

$$\frac{HEX^+}{Total\ BCR} = b(1-l) + s\,(1-b) \tag{2.3}$$

$$\frac{Alexa^+FAM^+}{Total\ BCR} = b(1-2s^2) + s(1-3b-s) \tag{2.4}$$

$$\frac{FAM^+}{Total\ BCR} = 2sb(1-s) + s^2 \tag{2.5}$$

$$\frac{Alexa^+}{Total\ BCR} = sb(1-sb) + s \tag{2.6}$$

where the states considered are those defined in Figure 2-5 and for example, $\bar{b}$ represents the fraction of total M-*BCR* copies that have not undergone a biological DSB, which is given by (1 –

b) in the model. In this assay, some states, such as $s\bar{s}\bar{b}$ and $\bar{s}s\bar{b}$, are indistinguishable and are treated as one state in the model to avoid double counting of total M-*BCR*.

As there are only 3 unknowns (*b*, *s*, and *l*), this system of equations is over-specified; we can solve this by creating a set of 3 equations by adding pairs of equations to give:

$$\frac{Alexa^{+}FAM^{+}HEX^{+}}{Total\ BCR} + \frac{FAM^{+}HEX^{+}}{Total\ BCR} = (1-b)(1-s) \tag{2.7}$$

$$\frac{HEX^{+}}{Total\ BCR} + \frac{Alexa^{+}FAM^{+}}{Total\ BCR} = b(2-l-2s^{2}) + s\,(2-4b-s) \tag{2.8}$$

$$\frac{FAM^{+}}{Total\ BCR} + \frac{Alexa^{+}}{Total\ BCR} = sb(3-2s-sb) + s(1+s) \tag{2.9}$$

The solution is independent of the pairings chosen, and only one such pairing is shown and used. All normalized values on the left-hand sides of equations 2.7 – 2.9 can be determined from the total number of droplets and the droplet counts for the indicated cluster. As an example, the data analysis method at the *n* = 2 level (i.e., no more than 2 copies per droplet considered) is applied below to the 2D output data shown in Figure 2-4A for gDNA purified from KU812 cells.

Droplet counts for each positive cluster are: Alexa^{+} cluster = 568 droplets, FAM^{+} cluster = 562 droplets, HEX^{+} cluster = 1272 droplets, Alexa^{+}FAM^{+} cluster = 715 droplets, FAM^{+}HEX^{+} cluster = 63 droplets, Alexa^{+}HEX^{+} cluster = 73 droplets, and Alexa^{+}FAM^{+}HEX^{+} cluster = 94 droplets. The total number of empty droplets = 11460 droplets, and the total number of read droplets = 14807.

The CPD (= $-ln(empty\ droplets/total\ droplets)$ = 0.256) is first computed and then combined with Poisson statistics to estimate the fraction of total read droplets containing *n* = 0, 1, 2, 3, … copies as $\frac{CPD^{n}exp^{-CPD}}{n!}$. For this sample, the fraction of empty droplets is 0.7740, *n* = 1 copy droplets is 0.1983, and *n* = 2 droplets is 0.0254.

Using these values, we can calculate the fraction *f*2 of all filled droplets that contain 2 copies (0.0254/(0.1983+0.0254) = 0.1136) to enable determination of the number of droplets within each cluster containing 2 copies. Beginning with the Alexa^{+}FAM^{+}HEX^{+} cluster:

$Alexa^+FAM^+HEX^+$ droplets containing 2 copies = cluster count x $f2$ = 94 x 0.1136 = 11

All pairs of copies that would generate a combined $Alexa^+FAM^+HEX^+$ signal within a droplet can then be defined. Each copy may either be intact *BCR* (which on its own generates an $Alexa^+FAM^+HEX^+$ signal within a droplet) or a fragment of *BCR*. The 5 possible fragments of interest generate an $Alexa^+$ signal, a FAM^+ signal, a HEX^+ signal, an $Alexa^+FAM^+$ signal, or a FAM^+HEX^+ signal, respectively. Due to the 2X degeneracy of unlike pairs, there are 17 combinations that generate an $Alexa^+FAM^+HEX^+$ signal within a droplet (e.g., $Alexa^+$/ $Alexa^+FAM^+HEX^+$ and $Alexa^+FAM^+HEX^+$/$Alexa^+$ are two possible combinations, while the $Alexa^+FAM^+HEX^+$/$Alexa^+FAM^+HEX^+$ combination can only be formed in one way due to the two contributing signals being indistinguishable). From this, the total abundance of each unique signal within the $Alexa^+FAM^+HEX^+$ droplets containing two copies is computed as:

$Alexa^+$	= 11 x 4/17 = 3	FAM^+	= 11 x 2/17 = 1
HEX^+	= 11 x 4/17 = 3	$Alexa^+FAM^+$	= 11 x 6/17 = 4
FAM^+HEX^+	= 11 x 6/17 = 4	$Alexa^+FAM^+HEX^+$	= 11 x (10/17 + 2x1/17) = 8

This analysis is repeated for the set of 5 positive clusters displaying a signal for one of the 5 possible *BCR* fragments generated by a biologic double stranded break (DSB) and/or a shear event:

$Alexa^+$ cluster droplets containing 2 copies = 568 x 0.1136 = 65

$Alexa^+$ = 65 x (2x1/1) = 130

FAM^+ cluster droplets containing 2 copies = 562 x 0.1136 = 64

FAM^+ = 64 x (2x1/1) = 128

HEX^+ cluster droplets containing 2 copies = 1272 x 0.1136 = 144

HEX^+ = 144 x (2x1/1) = 288

$Alexa^+FAM^+$ cluster droplets containing 2 copies = 715 x 0.1136 = 81

$Alexa^+$	= 81 x 4/7 = 46	FAM^+	= 81 x 4/7 = 46
$Alexa^+FAM^+$	= 81 x (4/7 + 2x1/7) = 69		

FAM^+HEX^+ cluster droplets containing 2 copies = 63 x 0.1136 = 7

FAM^+ = 7 x 4/7 = 4 HEX^+ = 7 x 4/7 = 4

FAM^+HEX^+ = 7 x (4/7 + 2x1/7) = 6

From this information and the fact that the $Alexa^+HEX^+$ cluster (73 droplets) must be comprised of 73 pairs of $Alexa^+$ + HEX^+, the total abundance of each unique signal is then computed.

Total $Alexa^+FAM^+HEX^+$ = 94 – 11 + 8 = 91

Total $Alexa^+$ = 568 – 65 + 130 + 3 + 46 + 73 = 755

Total FAM^+ = 562 – 64 + 128 + 1 + 46 + 4 = 677

Total HEX^+ = 1272 – 144 + 288 + 3 + 4 + 73 = 1496

Total $Alexa^+FAM^+$ = 715 – 81 + 69 + 4 = 707

Total FAM^+HEX^+ = 63 – 7 + 6 + 4 = 66

Finally, the total abundance of all forms of the *BCR* gene in the sample is estimated as

Total of all $Alexa^+$-containing signals = 91 + 755 + 707 = 1553

Total of all FAM^+-containing signals = 91 + 677 + 707 + 66 = 1541

Total *BCR* gene = (1553 + 1541)/2 = 1547

These results are used to compute the required value on the left-hand side of equations 2.7 – 2.9. For equation 2.7, for example:

$$\frac{Alexa^+FAM^+HEX^+}{Total\ BCR} + \frac{FAM^+HEX^+}{Total\ BCR} = \frac{91}{1547} + \frac{66}{1547} = (1-b)(1-s)$$

Equations 2.7 to 2.9 may then be solved to determine values for *b* (fraction of total M-*BCR* copies that have undergone a biological DSB in the M-*BCR*), *s* (fraction of total copies that have undergone a shear event in the M-*BCR*), and *l* (fraction of total copies that exhibit a loss of HEX signal).

By taking the difference between the experimental values and these calculated model values we can use the least squares minimization approach, to achieve one value for the "Solver"

subroutine of Excel (Microsoft Corp.) to minimize. Excel Solver has proven reliable in all fields of biology [266] and utilizes non-linear function fitting via an iterative algorithm [267] based on the generalized reduced gradient (GRG) method. Values of regressed parameters were constrained by the requirement that *s*, *b* and *l* all be ≤ 1 and ≥ 0. (See Appendix A for representative results).

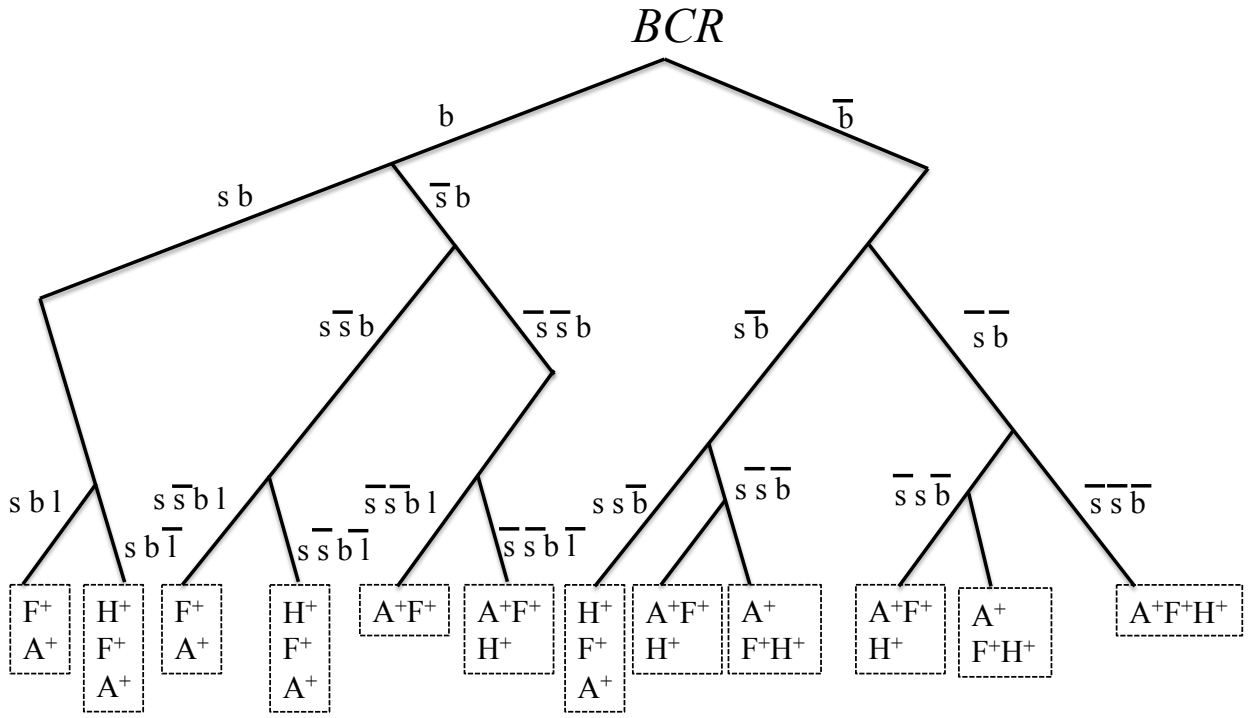


Figure 2-5. Tree diagram of all potential states of M-*BCR* within gDNA isolated and analyzed by the ddPCR based M-*BCR* status assay.
Analysis of all possible states of a copy of M-*BCR*, where *b* indicates a biological DSB within the M-*BCR*, *s* mechanical fragmentation of *BCR* within the M-*BCR* or 7 kbp control region, *ss* mechanical fragmentation within both the M-*BCR* and control region, and *l* loss of HEX^+ due to loss of the template immediately to the 3′ side of the M-*BCR*. A bar above a letter indicates that the event has not occurred. States are mapped to clusters within the ddPCR assay output data, with the set of clusters populated by each state shown in the connected hashed box, where A^+ = $Alexa^+$ cluster, F^+ = FAM^+ cluster, A^+F^+= $Alexa^+FAM^+$ cluster, F^+H^+ = FAM^+HEX^+ cluster, and $A^+F^+H^+$ = $Alexa^+FAM^+HEX^+$ cluster. Note that the $s\bar{s}\bar{b}$ and $\bar{s}s\bar{b}$ states are indistinguishable and therefore treated as one state in the model to avoid double counting of total M-*BCR*.

2.3.5 Operating Conditions and Algorithm for Accurate Cluster Assignments

Use of ddPCR and model equations 2.7 to 2.9 to reliably quantify *BCR* status requires accurate assignment of clusters, and those assignments are dependent on the ability to obtain identifiable, non-overlapping clusters within the 2D output. The raw ddPCR data reported in Figure 2-4 were collected using optimized reagents (primers/probes) and reaction conditions, and well-defined clusters are observed for both samples. Mean cluster intensities and associated standard deviations can then be computed and used to define cluster borders. In general, positioning of cluster borders can be confounded by excessive numbers of droplets falling along the vector connecting a pair of clusters – commonly known as ddPCR "rain". Proposed mechanisms of rain formation include differences in amplification efficiency resulting from sub-optimal primer/probe designs, template sequence variations (e.g., somatic point mutations), or non-uniform partitioning of the PCR Mastermix components [268]. The level of rain observed is related not only to the nature of the template(s) being analyzed, but also to the chosen reaction conditions, reagents, and thermal cycling profile. For the M-*BCR* status assay reported here, the degree of observed rain is strongly dependent on the thermal stabilities of the primers and probes, and the stabilities of both were optimized using the thermal gradient function of the ddPCR instrument. An example of un-optimized ddPCR assay operating conditions resulting in significant rain in the 2D output plot is shown in Figure 2-6.

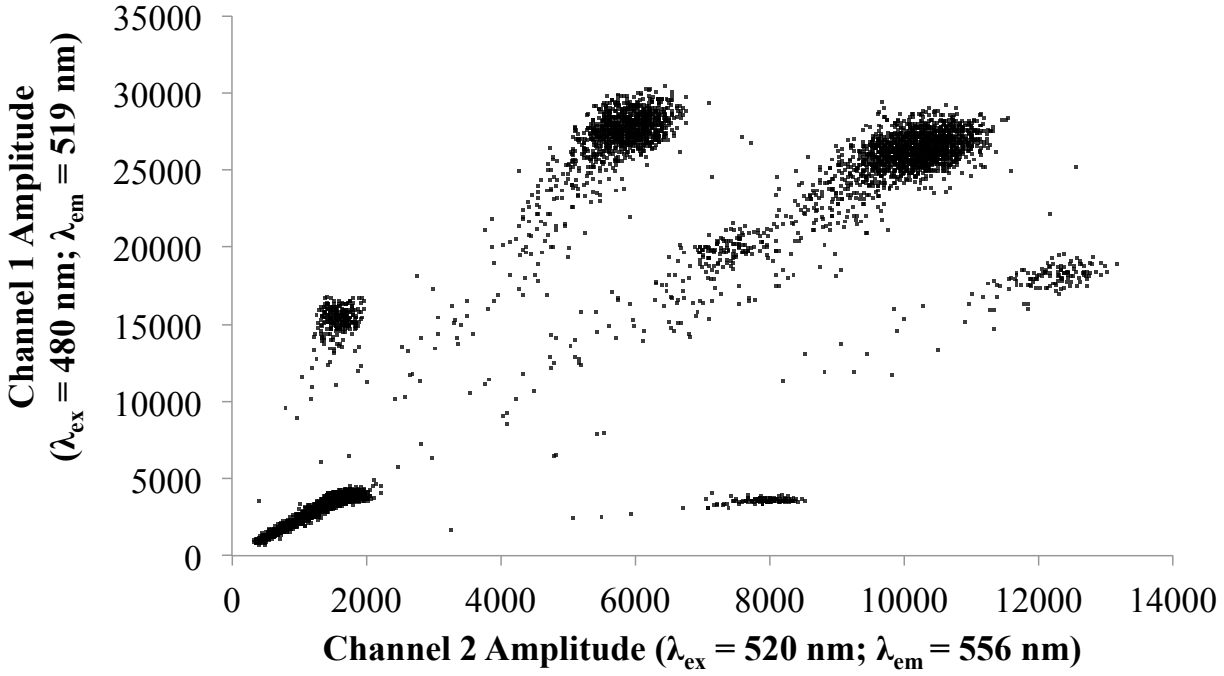


Figure 2-6. Example of excessive "rain" in the raw data from the ddPCR *BCR* status assay when operated at un-optimized conditions.
For the *BCR-ABL* positive KU812 cell line (CPD =0.249) "rain effect" pattern observed due to the non-uniform partitioning of the PCR Mastermix components

Under optimized ddPCR assay conditions (e.g., Figure 2-4), however, the rain effects are reduced to the point (typically < 0.1% of total droplet counts) that significant interference with clusters is avoided, allowing the droplet event data to be analyzed to assign each read droplet as either empty (negative), rain, or positive for at least one fluorophore. Though mild rain does skew data within a given positive cluster towards, for example, the empty droplet cluster, the bias is quite weak, allowing expected positive data clusters and their quality to be reliably defined by fitting a normal distribution first to the channel 1 (λ_{em} = 519 nm) signals for droplets in a chosen cluster. From that the mean and standard deviation (σ) are computed, with the mean ± 3σ then setting the cluster borders in the channel 1 dimension. That same analysis is repeated for channel 2 (λ_{em} = 556 nm) to define the remaining borders of each cluster. An illustration of the outcome of this data analysis process is provided in Figure 2-6

Droplet counts within each cluster may then be read and the total number of positive droplets and CPD computed. Total droplets successfully read (typically 12,000 to 15,000) must

exceed a threshold value set at 10,000, while the total CPD must exceed 0.2 to avoid excessive contributions to the total error from sub-sampling and partitioning uncertainties (see below); failure to satisfy either criteria constitutes a failed run. In addition, the standard deviation in the channel 1 signals of the empty (negative) population of droplets is computed and scored as a metric of the overall quality of data clusters. We find that a $\sigma \leq 200$ FU signifies formation of clusters within the 2D output that are sufficiently dense and segregated that their analysis with the model embodied in equations 2.7 – 2.9 is reliable.

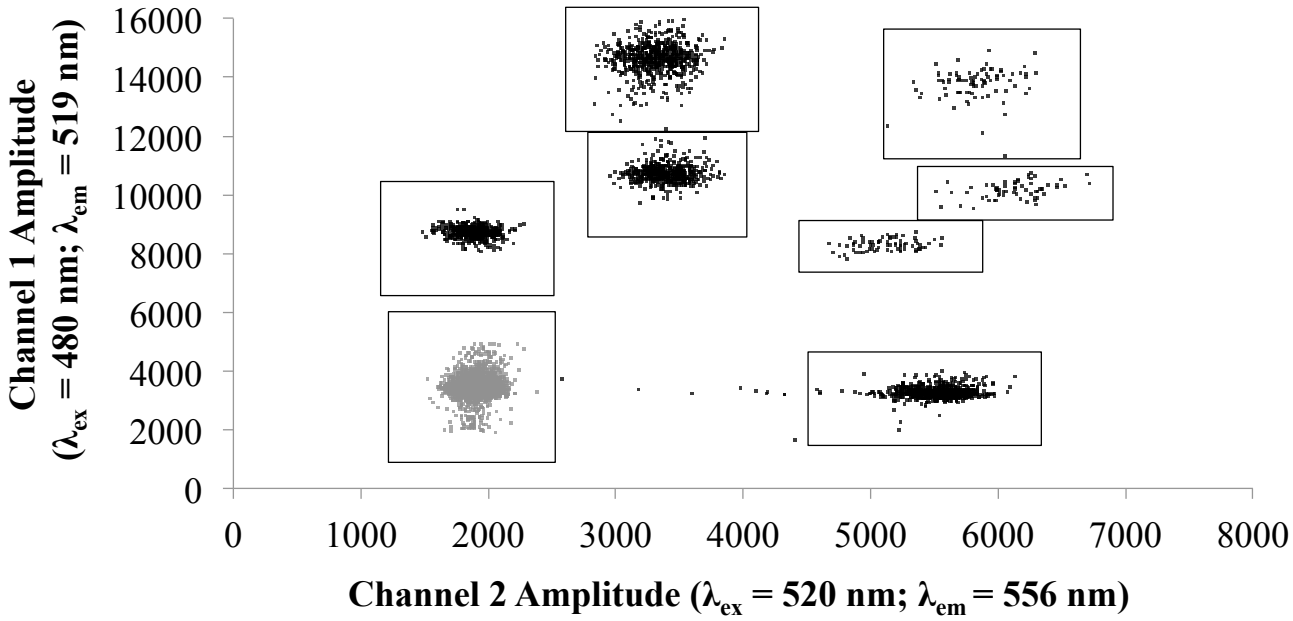


Figure 2-7. 2D output plots for the M-*BCR* status assay applied to gDNA from the KU816 cell line showing the result of application of the data processing algorithm for cluster assignment.

Cluster borders are first defined by fitting a normal distribution first to the channel 1 (λ_{em} = 519 nm) signals for droplets in a chosen cluster. From that the mean and standard deviation (σ) are computed, with the mean ± 3σ then setting the cluster borders in the channel 1 dimension. That same analysis is repeated for channel 2 (λ_{em} = 556 nm) to define the remaining borders of each cluster. Model fitting to the resulting cluster assignments yields the following results: total read droplets = 14807, CPD = 0.256, frequency of biological disruption of the M-*BCR* = 93.3%, frequency of mechanical disruption of the M-*BCR* = 20.9%. The data constitute an acceptable run based on our assay failure criteria.

2.3.6 M-*BCR* Status Assay Shows 1:1 Correspondence with *BCR-ABL* Frequencies

For each *BCR-ABL* positive cell line studied, results from the M-*BCR* status assay were compared to FISH. In each case, the frequency of biological disruption of the M-*BCR* recorded using the ddPCR assay matches the frequency of the *BCR-ABL* fusion oncogene recorded by FISH (Table 2-1). For instance, for the KU812 cell line the raw 2D output data shown in Figure 2-8 and associated data analysis process record a frequency of biologically disrupted M-*BCR* of 93.1%, while FISH records a corresponding *BCR-ABL* frequency of 97%. Note that the ddPCR

data for KU812 shown in Figure 2-8 is from an independent experiment, and may also be compared to those in Figure 2-4A (M-*BCR* DSB frequency = 93.3(±2.0)%) or to the average value from $n = 6$ replicates to assess assay reproducibility.

As noted above, a loss of HEX^+ droplets may be expected in the K562 cell line, and this is indeed observed in the 2D data output (Figure 2-9). But by explicitly accounting for it, the data analysis model embodied in equations 2.7 – 2.9 yields a biological M-*BCR* disruption frequency (92.1%) that again agrees with the *BCR-ABL* frequency (> 83%) reported by FISH. As expected based on K562 karyotype, the FISH data show a large number of *BCR-ABL* repeats on one chromosome, which made counting difficult due to signal overlap. The *BCR-ABL* frequency could only be estimated from FISH images for K562 samples.

The importance of the new gDNA extraction protocol described in section 2.2.3 to the accuracy and precision of the ddPCR *BCR* status assay, and the new translocation analysis platform described in this thesis, is made clear in Table 2-2, which reports assay results for various samples when commercial gDNA extraction kits were instead applied. Use of the magnetic-bead based gDNA extraction method results in *ca.* 60% loss of *BCR* template to shear (non-biologic disruption). As a result, for the KU812 cell line, a highly variable translocation frequency is recorded. More important, the value (10% - 18%) does not agree with the corresponding FISH result (Table 2-3). Results when the column extraction method is employed are even worse due to the very high levels of template lost to shear when that standard purification method is employed. This same problem is also observed for the MEG01 cell line.

Table 2-2. Results of the ddPCR *BCR* status assay applied to various samples when the new gDNA extraction protocol is replaced with a commercial gDNA extraction kit.
Kits tested include the MagAttract HMW DNA kit (magnetic beads) and the. Qiagen Gene-read DNA FFPE kit (column extraction) both from Qiagen

EXTRACTION METHOD and sample gDNA	CPD	Translocation % *b*	Shear % *s*	Loss in HEX % *l*
KU812 DNA magnetic beads	0.149	18.1	62.4	2.5
KU812 DNA magnetic beads	0.149	10.8	63.8	41.3
MEG01 DNA magnetic beads	0.216	83.4	58.4	51.7
MEG01 DNA magnetic beads	0.211	84.4	58.7	55.6
KU812 DNA column extraction	0.286	0.0	84.0	0.0
KU812 DNA column extraction	0.202	12.9	81.8	100.0
KU812 DNA column extraction	0.195	13.3	82.0	100.0
KU812 DNA column extraction	0.202	10.4	82.3	100.0
KU812 DNA column extraction	0.212	0.0	92.3	0.0
KU812 DNA column extraction	0.213	5.3	87.7	100.0

To evaluate if the observed disruption of the target sequence (Figure 2-3) to the immediate 3′ side of the M-*BCR* could be related to (partial) loss of derivative chromosome 9 in K562, the presence of or losses within the region of 22q spanning from the 3′ boundary of the M-*BCR* to 23 kbp downstream of *BCR* was interrogated using other short PCR targets in lieu of that sequence flanking the 3′ boundary of the M-*BCR*. For these additional reactions, the same reduction in amplification signal was observed, suggesting that loss of HEX^+ droplets may indeed reflect loss of or within derivative chromosome 9 [62]. Through the same mechanism [67, 262] a larger than expected loss of HEX^+ droplets is also recorded for the MEG01 cell line (Figure 2-10). Nevertheless, concordance of the M-*BCR* disruption frequency with the *BCR-ABL* oncogene frequency recorded by FISH is again observed, indicating that the ddPCR based M-*BCR* status assay provides an indirect but quantitative measure of *BCR-ABL* when present in either a classic or variant PhC. Representative FISH images for each cell line are also shown in Figures 2-8, 2-9 and 2-10.

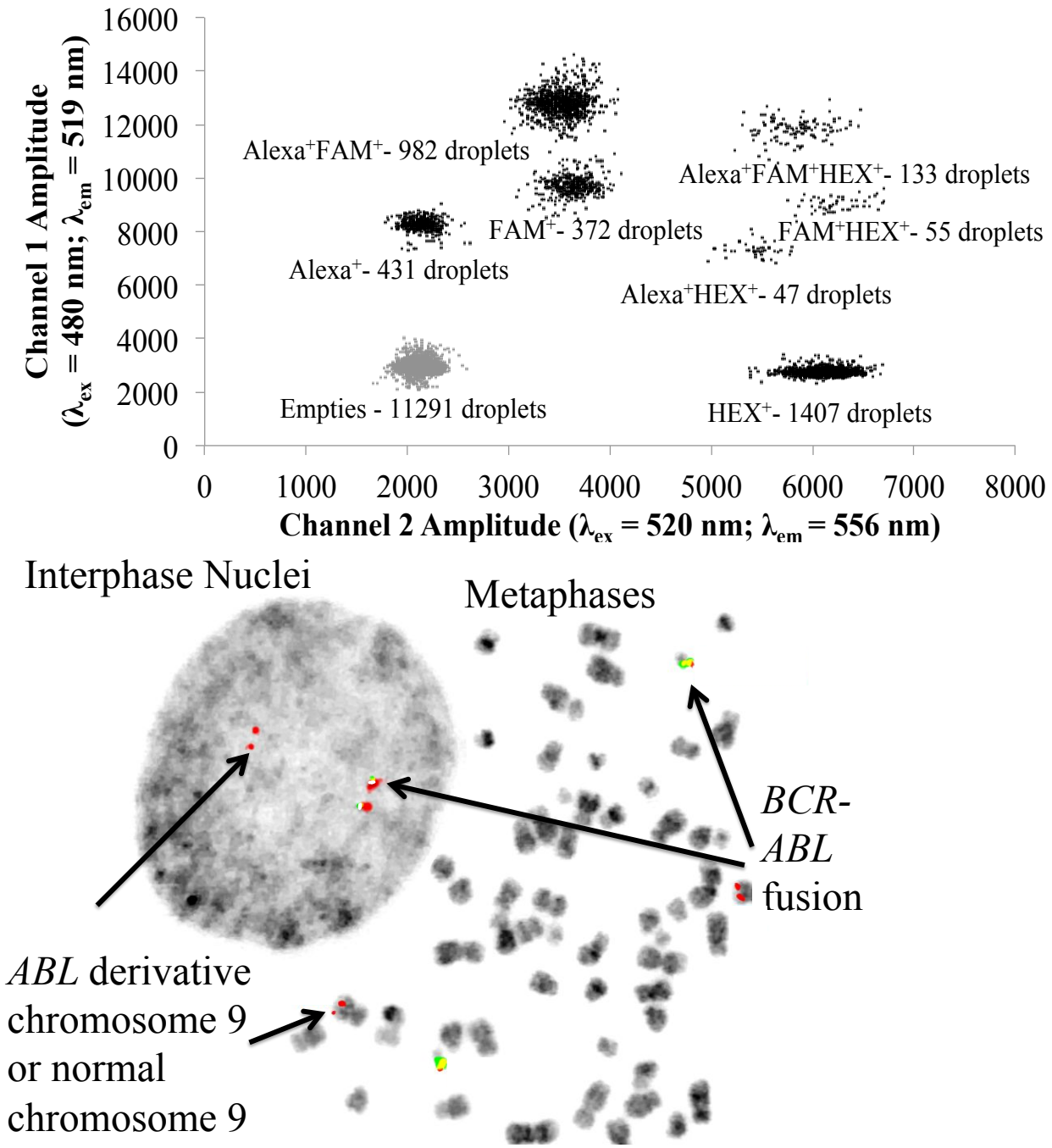


Figure 2-8. Representative M-*BCR* status assay and FISH results for cell line KU812.
The raw ddPCR data for the *BCR-ABL* positive cell line provides clear visual evidence of biological cleavage within the M-*BCR* that is consistent with the *BCR-ABL* fusion oncogene detected in the corresponding FISH images. Values for the frequency of biological disruption of the M-*BCR* recorded in the ddPCR assay, and for the *BCR-ABL* frequency recorded by FISH are reported in Table 2-3. The ddPCR assay provides additional information: KU812 (total read droplets = 14718, CPD = 0.265, frequency of mechanical disruption of the M-*BCR* = 20.9%).

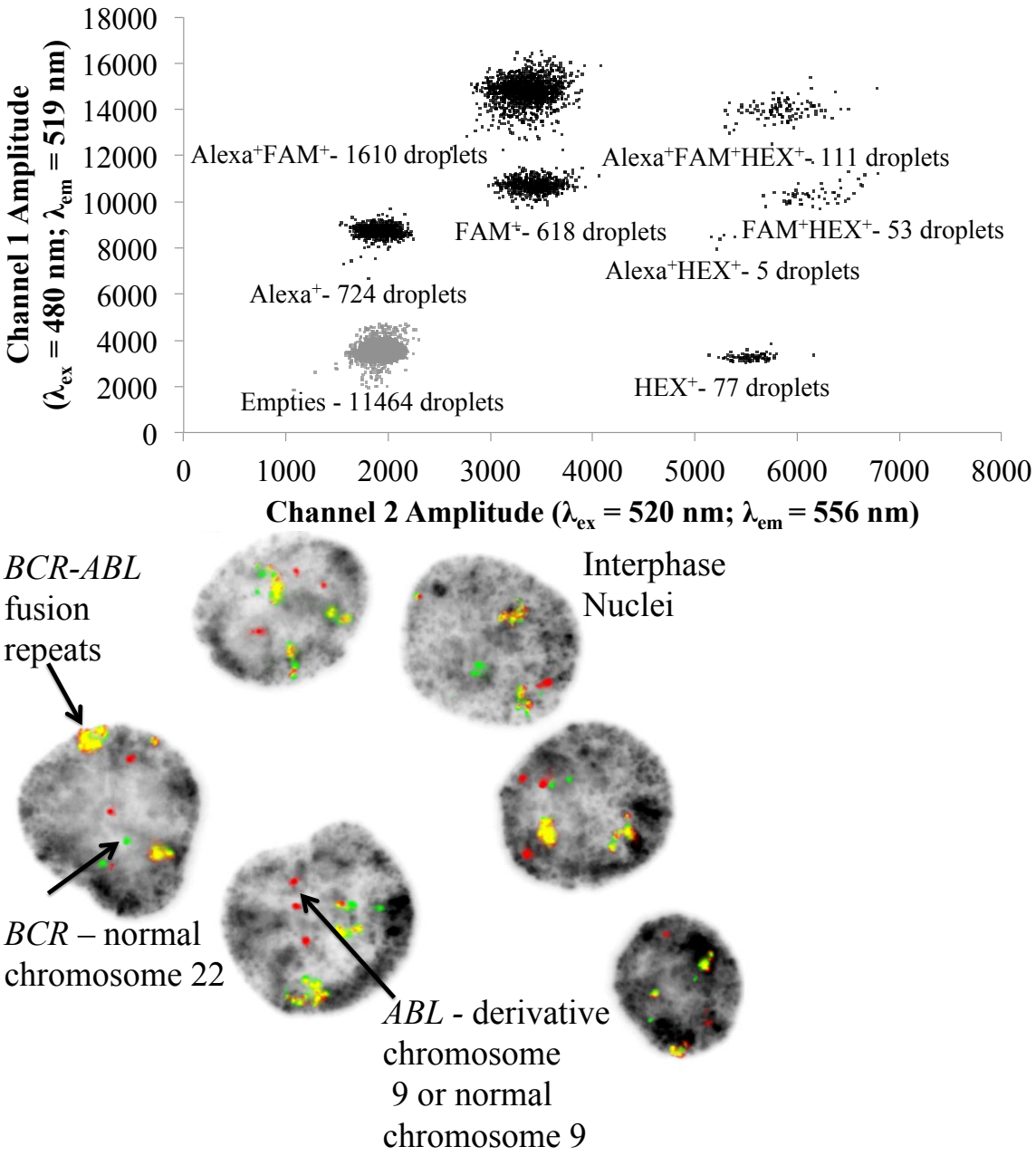
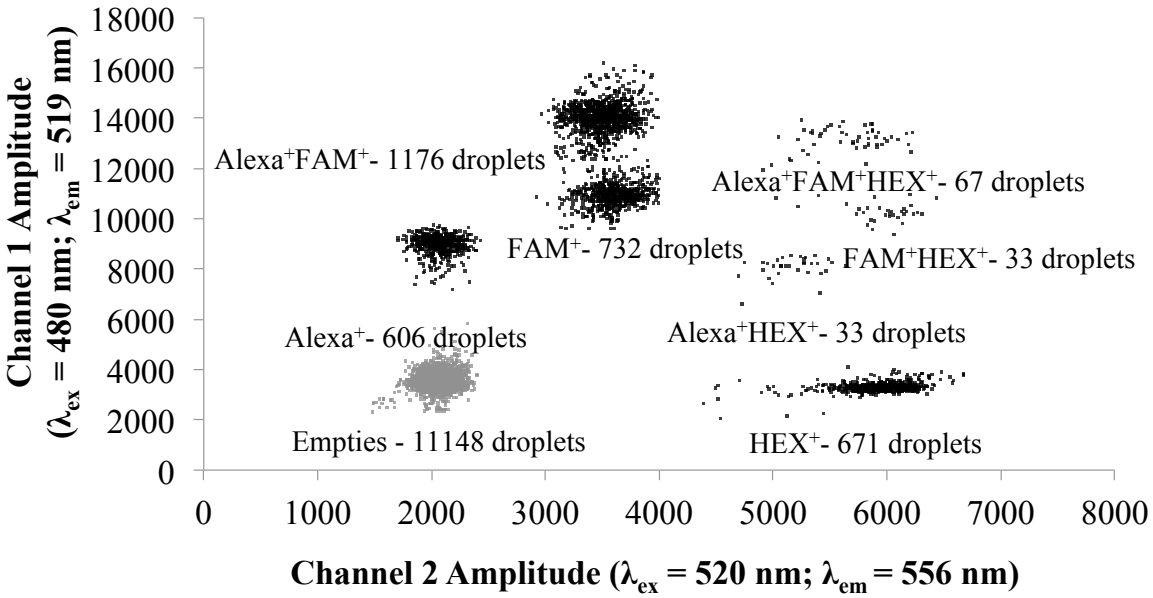


Figure 2-9. Representative M-*BCR* status assay and FISH results for cell line K562.
The raw ddPCR data for the *BCR-ABL* positive cell line provides clear visual evidence of biological cleavage within the M-*BCR* that is consistent with the *BCR-ABL* fusion oncogene detected in the corresponding FISH images. M-*BCR* status assay data for K562 show a loss of HEX^+ droplets, consistent with loss of derivative chromosome 9 in these models of blast crisis. Values for the frequency of biological disruption of the M-*BCR* recorded in the ddPCR assay, and for the *BCR-ABL* frequency recorded by FISH are reported in Table 2-3. The ddPCR assay provides additional information: K562 (total read droplets = 14662, CPD = 0.246, frequency of mechanical disruption of the M-*BCR* = 18.8%).



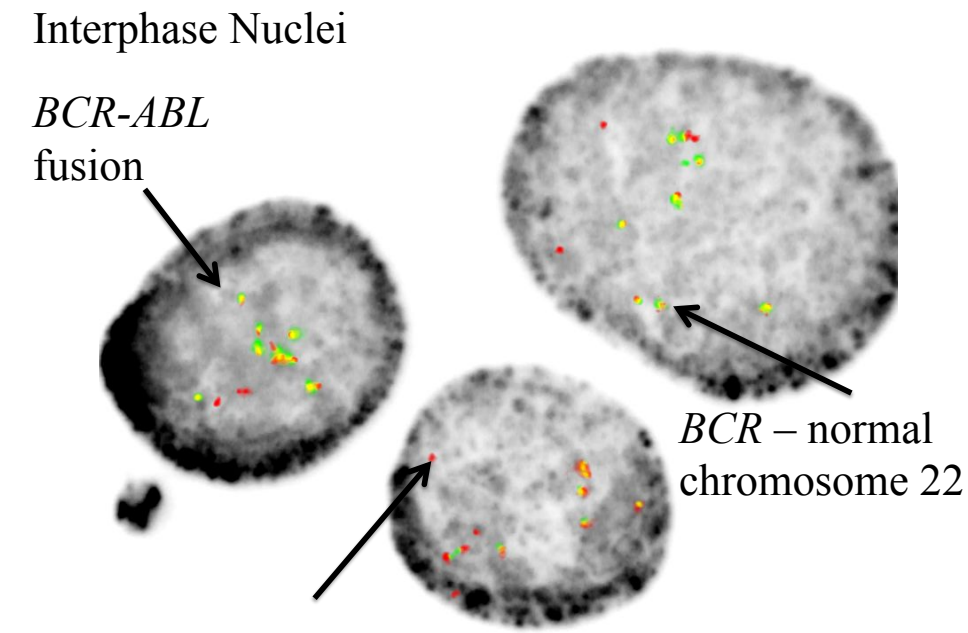


Figure 2-10. Representative M-*BCR* status assay and FISH results for cell line MEG01. The raw ddPCR data for each *BCR-ABL* positive cell line MEG01 provides clear visual evidence of biological cleavage within the M-*BCR* that is consistent with the *BCR-ABL* fusion oncogene detected in the corresponding FISH images. M-*BCR* status assay data for MEG01 show a loss of HEX^+ droplets, consistent with loss of derivative chromosome 9 in these models of blast crisis. Values for the frequency of biological disruption of the M-*BCR* recorded in the ddPCR assay, and for the *BCR-ABL* frequency recorded by FISH are reported in Table 2-3. The ddPCR assay provides additional information: MEG01 (total read droplets = 14466, CPD = 0.261, frequency of mechanical disruption of the M-*BCR* = 23.8%).

Table 2-3. Comparison of M-*BCR* status assay results to benchmark FISH data for cell lines K562, MEG01 and KU812.
FISH data (100 imaged nuclei) were analyzed by calculating translocation as the number of *BCR-ABL* fusions divided by the sum of the normal *BCR* transcripts and *BCR-ABL* fusions. The K562 data showed a large number of *BCR-ABL* repeats on one chromosome, making counting difficult due to signal overlap [259]. The translocation could only be estimated in this case (> 83%), but is consistent with the corresponding value (92.1% ± 1.9%) recorded by the ddPCR assay applied to 6 independent samples.

Cell line	No. nuclei	Fusion *(BCR-ABL)*	*ABL* normal or derivative	*BCR* normal	Total translocated	Total normal	FISH % translocation (*BCR-ABL*/ ((*BCR-ABL*)+*BCR*)	Assay % translocation
K562	**100**	>5	1	1	>500	100	**>83**	**92.1 ±1.9**
	80	3	3	0	240	0		
	6	3	3	0	18	0		
	1	3	2	1	3	1		
MEG01	1	3	4	0	3	0		
	11	3	3	2	33	22		
	1	2	2	1	2	1		
Sum	**100**				**299**	**24**	**94**	**94.1 ± 2.0**
	88	1	2	0	88	0		
	1	0	3	1	0	1		
	1	0	2	1	0	1		
KU812	3	1	3	0	3	0		
	3	1	1	0	3	0		
	3	1	2	1	3	3		
	1	0	2	1	0	1		
Sum	**100**				**97**	**6**	**97**	**93.3 ±2.0**
Negative	93	0	2	2	0	186		
Control	4	1	1	1	4	4		
HL60	1	0	2	1	0	1		
	2	0	3	2	0	4		
Sum	**100**				**4**	**195**	**≤ 2[£]**	**0.01 ±0.02**

£ - LOD for FISH provided by the Cancer Genetics Laboratory of the BC Cancer Agency.
∞ - LOD of the ddPCR M-*BCR* status assay based on independent replicates (n = 24) of gDNA isolated from the *BCR-ABL* negative HL60 cell line, and standard deviation of the mean values on n = 4 replicates of gDNA isolated from the indicated *BCR-ABL* positive cell line.

2.3.7 Assay Limit of Detection when applied to gDNA from Cell Lines

Serial dilutions (n = 4 for each dilution) of KU812 in HL60 gDNA down to KU812-derived M-*BCR* frequencies of 0.25% were used to define the LOD of the assay (Figure 2-9). The assay was first applied to gDNA (n = 24) from HL60 alone (negative control), from which the Limit of Blank (LOB (= 0.05%)), taken as the mean + 95% CI of the set of M-*BCR* disruption frequencies, was determined. Measured M-*BCR* disruption frequencies correlate linearly with expected *BCR-ABL* frequencies down to a recorded LOD of 0.25% based on the LOB and a paired Student's *t*-test. In these experiments, the CPD was set between 0.2 and 0.3, and the reported LOB and LOD apply to those conditions. Though indirect, the assay can therefore reliably identify a *BCR-ABL* positive sample through the detection of as few as *ca.* 3 copies of *BCR* showing a biological disruption within the M-*BCR*.

In comparison, the LOD of FISH is 2% when calculated, as was done here, from the mean + 3 standard deviations of the expected signal pattern seen in 100 nuclei from 10 different constitutionally normal individuals (1000 nuclei total).

2.4 Discussion

Clinical interest in applying digital technologies to analysis of molecular mutations diagnostic or theranostic of cancer is building [269] due to the recognized advantages of the platform [270]. Though currently no ddPCR-based molecular diagnostic test has been adopted for CLIA-certified clinical use, several are in development or under-going clinical validation, including but not limited to assays against mutations in epidermal growth factor receptor in non-small-cell lung cancer [271], cancer-associated viruses [209, 272], and HER2 in breast cancer specimens [211].

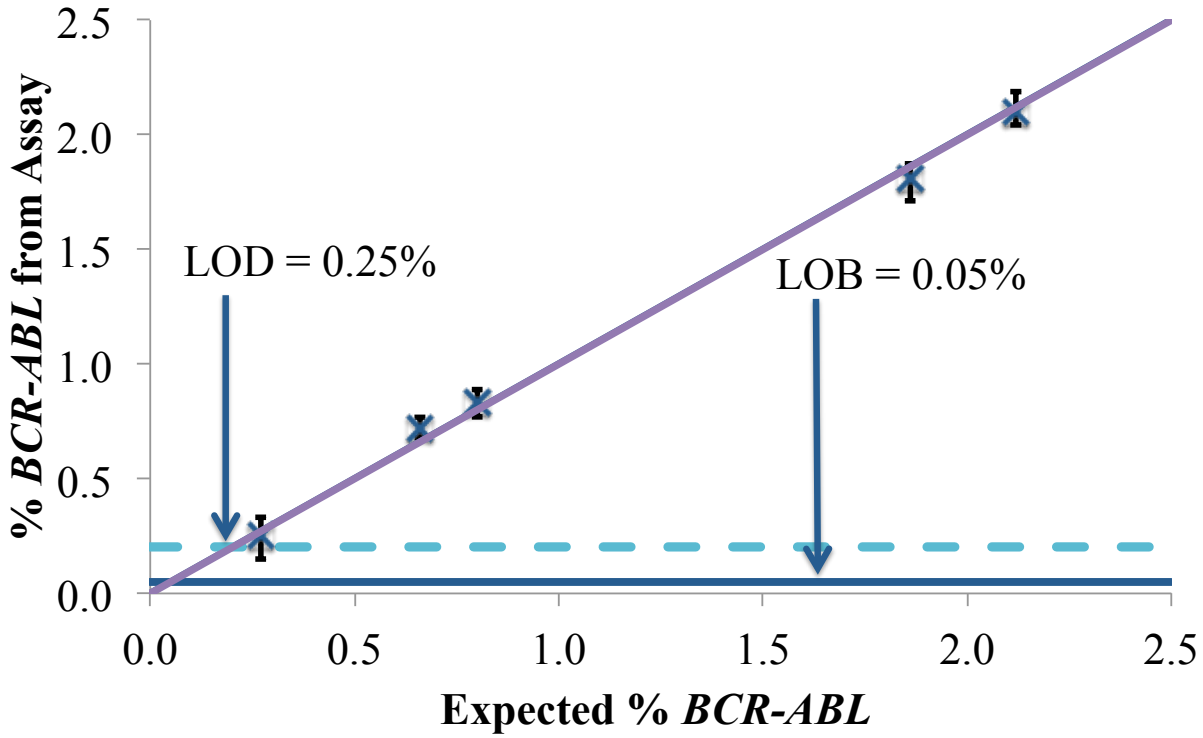


Figure 2-11. Accuracy, precision and limit of detection of the M-*BCR* status assay. Measured *BCR-ABL* frequencies and standard deviations ($n \geq 4$) are plotted versus expected *BCR-ABL* frequencies for serial dilutions of KU812 in HL60 gDNA down to a frequency of 0.25%. Significant linear correlation ($R^2 \geq 0.9995$; $P < 0.0001$) between the measured and expected *BCR-ABL* frequencies is observed down to 0.25%. Replicates ($n = 24$) of gDNA from the *BCR-ABL* negative HL60 cell line were used to define the mean and standard deviation of false positives, from which the 95% confidence interval was determined and used to define the Limit of Blank (LOB = 0.05%; blue hashed line). At a CPD = 0.2, statistically significant *BCR-ABL* frequencies can be obtained to an LOD of 0.25% based on the measured LOB and a paired Student's *t*-test.

The results presented here extend the range of potential clinical applications of ddPCR related to cancer diagnostics by describing a digital assay for initial diagnosis of CML that, when applied to *BCR-ABL* positive cell lines, provides a quantitative measure of *BCR-ABL* frequencies down to an LOD of 0.25% at a total CPD of 0.2. The ddPCR-based M-*BCR* status assay outperforms FISH in terms of detecting low amounts of the *BCR-ABL* gene fusion, but only if the raw data are appropriately analyzed. A robust data processing methodology was established that accounts for the impact of non-biologic (mechanical) fragmentation of the M-*BCR* on the number of positive data clusters observed and the distribution of read droplets among them. The method allows one to de-convolute the raw digital data to fully enumerate copies of M-*BCR* that have

been disrupted by non-biologic fragmentation or a biologically derived DSB, respectively. By also accounting for potential loss of that portion of *BCR* lying to the 3′ side of the M-*BCR*, the model allows quantification of *BCR-ABL* in cases where partial or complete deletion of derivative chromosome 9q has occurred. Studies have shown that such deletions occur at the onset of CML [69, 74-76, 71], and that the micro-deletions then observed in both 5′*ABL* and 3′*BCR* occur at the time of translocation [75]. The size of these deletions varies considerably between patients, but in many cases extends several mega-bases within 9q, 22q, or both [74, 78]. Patients carrying a classic PhC (t(9;22)(q34;q11)) have been found to have 9q deletions averaging *ca.* 12% [73-76, 72], while patients harboring a variant translocation (10% of the patient population) can exhibit more significant deletions – ~ 39% on average [73, 74, 77, 75, 76, 72]. As a result, derivative chromosome 9 deletions occur in *ca.* 15% – 16% of the total CML patient population [69-72]. Huntly et al., [78] analyzed patients lacking derivative chromosome 9 and found that all lacked the *ABL-BCR* transcript. Connections between the loss of 3′*BCR* and the loss of derivative chromosome 9 were also investigated by Fourier et al., [61], who reported that the losses occur simultaneously and correlate in *ca.* 90% of patients.

Without properly accounting for these potential variants, interpretation of raw data from the M-*BCR* status assay might prove challenging or erroneous, possibly triggering unnecessary or deleterious changes in therapy if applied clinically. Importantly, CML patients with deletions of or within derivative chromosome 9 have a poor prognosis if undergoing conventional treatment [74], possibly due to loss of one or more genes within the deleted region(s) of 9q. Losses in the number of the HEX^+ droplets may provide an indication of such deletions, highlighting the potentially rich information content of the digital assay.

When operated at a recommended CPD of 0.2, the M-*BCR* status assay described here offers improved specificity and may allow identification of variant patterns, including derivative chromosome 9 deletions, additional questions must be addressed prior to clinical use. For example, the observed linear correlation of assay results to *BCR-ABL* frequencies must be demonstrated in patient samples, and protocols must be established to address rare cases in which false positives are recorded, such as for a patient exhibiting an atypical chromosomal rearrangement in which either 5′*BCR* fuses to another partner, or *ABL-BCR* is present but not *BCR-ABL*. Though the latter case should be clearly identified in the assay through atypical loss

of both Alexa^{+} and FAM^{+} droplets, clinical results of the assay may nevertheless need to be interpreted together with standard cytogenetic data, and possibly other molecular genetic analyses. Nevertheless, the results reported here provide the first demonstration that ddPCR can be used to analyze genomic DNA for the presence of balanced translocations, as exemplified through quantitative enumeration of disruptions in the M-*BCR* leading to the *BCR-ABL* fusion oncogene and CML.

The results reported here were typically collected at a CPD between 0.2 and 0.3, in part to enable determination of the minimum number of copies of *BCR-ABL* required for a unequivocal positive call, which was found to be ~ 3 copies. The LOB and LOD values reported apply to those sampling conditions, and yield a coefficient of variation (CV) of ±1.9%. As the combined sub-sampling plus partitioning (Poisson) error is known to decrease with increasing CPD up to a CPD ~ 1.6, we determined what improvements might be realized by operating the assay at a higher CPD. Measured total errors, calculated from the standard deviation of 4 replicate ddPCR runs, and the underlying contribution of the combined sub-sampling and partitioning error to them where determined as a function of CPD (Figure 2-10). At low CPD, the total experimental error equals the sub-sampling + partition error, as expected (https://mcb.berkeley.edu/.../Statistics of ddPCR_v1DEC3.pdf) [212]. At CPDs above 1.5, the experimental error becomes larger due in part to increased tendency for overlap of clusters, which confounds cluster assignments. Both the total error, the subsampling + partitioning error, however, gradually decrease with CPD, such that a reduction in the CV to ±1.6% and an associated near 2-fold improvement in the LOD (LOD = 0.15%) can be realized by operating at a higher CPD (i.e., CPD ~ 1).

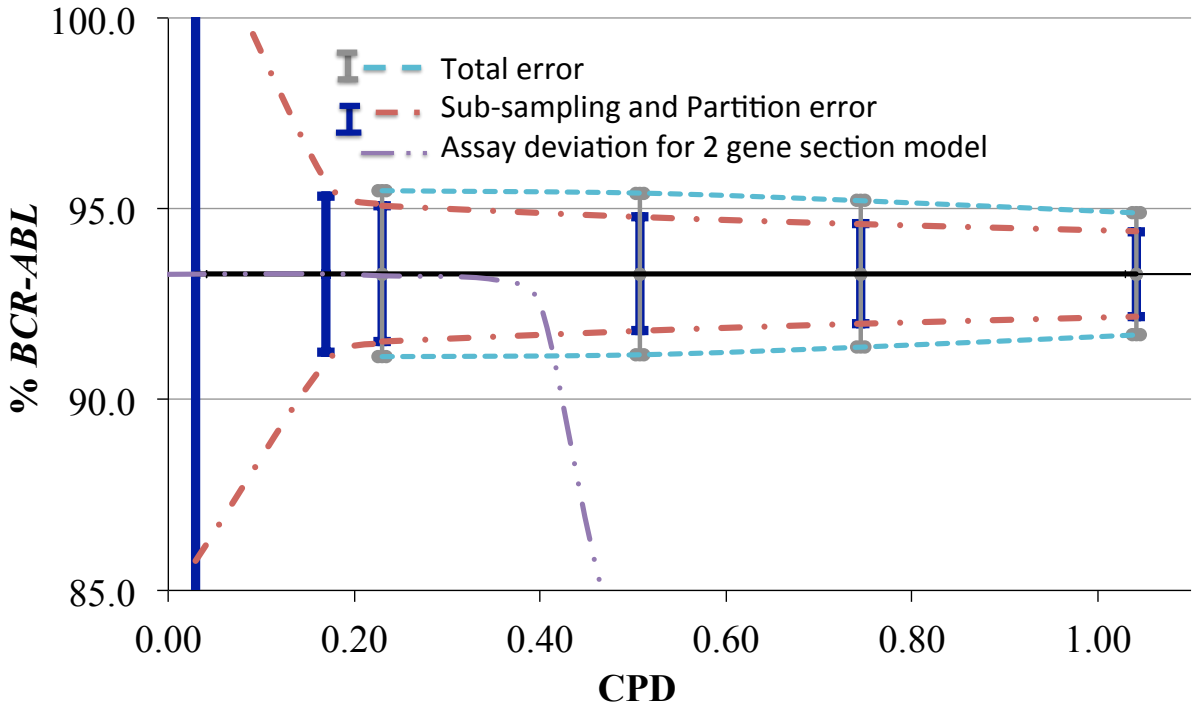


Figure 2-12. Sources of error in the M-*BCR* status assay and their dependence on CPD. Representative data set (KU812 gDNA; 93.3% biological DSB frequency of M-*BCR*). Blue error bars and red dashed dotted lines report the sub-sampling + partitioning (Poisson) error. Gray error bars and blue dotted lines report the total experimental error at each CPD. The combined sub-sampling + partitioning error (as expected) increases non-linearly with decreasing CPD, growing in a near exponential fashion at CPD < 0.2. The total experimental error was calculated from the standard deviation of 4 replicate ddPCR runs. At low CPD, the experimental error equals the sub-sampling + partition error [212, 273]. Above a CPD of ~ 1.5, the total error becomes larger than the subsampling + partition error due to increased overlap of clusters, which confounds cluster assignments. The purple dashed line shows as a function of CPD the *BCR-ABL* frequency predicted by the assay if the data-processing model is applied to the raw data set under the simplifying assumption that no more than two copies partition into a given droplet. The result shows that knowledge and use of the full Poisson distribution for each state of M-*BCR* is required to accurately analyze raw data at a CPD above ~ 0.35 due to the significant number of droplets containing three or more copies.

The %*BCR-ABL* frequency provided by the assay may be computed either using full Poisson statistic or by applying the data-processing model to the raw data set under the simplifying assumption that no more than two copies partition into a given droplet. The result (Figure 2-12) shows that knowledge and use of the full Poisson distribution for each state of M-

BCR is required to accurately analyze raw data at CPDs above ~ 0.35 due to the significant number of droplets containing three or more copies.

The results presented in this chapter demonstrate the validity of the proposed new digital method for detecting and quantifying rearrangements in (proto-)oncogenes associated with reciprocal translocation events. To achieve a successful workable assay, the method includes a new process to gently extract gDNA from cell lines was developed (See section 2.2.3). Double stranded gDNA is susceptible to mechanical fragmentation, particularly during extraction from cells [274]. DNA fragmentation and associated fragment sizes can significantly impact the ddPCR results [204, 218]. The new extraction method developed ensures an acceptably low amount of shearing of gDNA.

The method also requires careful design of primers and probes, as well as ddPCR operating conditions, which together serve to minimize the "rain effect" which produces less defined clusters in the raw ddPCR output. Artifacts that can occur in a ddPCR experiment include poorly formed droplets and droplet shearing (bubbles in sample mixture rise within the pipette tip and may burst and shear droplets) [191]. A probabilistic model that accounts for the effects of these potential sources of bias is included in the method to avoid false-negative and false-positive signals that might lead to an inappropriate clinical decision [209]. The model includes the assumption of independence between a mechanical DSBs (shear) and biological DSBs (translocation). Based on the concordance between FISH and ddPCR assay results, that assumptions appears valid, as should be expected since the translocation occurs within the tumor cell well before a sample is extracted for analysis, and there is no theory or evidence to suggest a translocated fusion gene to be more or less susceptible to mechanical shear.

Chapter 3: Initial Diagnosis of *ALK*-positive Non-small Cell Lung Cancer Based on *ALK* Gene Fragmentation Analysis Utilizing Droplet Digital PCR

The novel digital method proposed in this thesis for detecting and quantifying translocation events is applied to the detection of oncogene rearrangements associated with non-reciprocal (inversion) translocations. It is used here to create a new assay to detect rearrangements in *ALK* associated with *ALK*-positive non-small cell lung cancer (NSCLC). NSCLC patients may carry a non-reciprocal translocation on human chromosome 2 in which double stranded breaks (DSB) within the echinoderm microtubule-associated protein-like 4 (*EML4*) gene and the anaplastic lymphoma kinase (*ALK*) gene lead to an inversion of genetic material that forms the non-natural gene fusion *EML4-ALK*. The chimeric *EML4-ALK* gene encodes a highly active tyrosine kinase, EML4-ALK, which is associated with 3 to 7% of all NSCLCs – a subclass defined as *ALK*-positive NSCLC.

The clinical detection of *ALK* rearrangements in NSCLC is currently visualized via a FISH assay. The FISH assay utilizes two probes, both specific to the *ALK* gene, with a segregation (break apart) of fluorescence signal indicating an *ALK* rearrangement. That assay can detect *ALK* rearrangements to a limit of detection of 15%, while the ddPCR assay presented here provides a limit of detection of 0.25% at lower cost and faster turnaround times.

3.1 Introduction

Approximately 1.5 million mortalities worldwide are attributed each year to the two major forms of lung cancer [275] – non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC accounts for 85 to 90% of all lung cancers [125] and can be further classified into 3 major histological subtypes – adenocarcinoma, squamous-cell carcinoma and large cell lung cancer [126]. Many NSCLC adenocarcinomas harbor tumor-associated mutations in either the epidermal growth factor receptor (*EGFR*) or kirsten rat sarcoma viral oncogene homolog (*KRAS*) gene, or rearrangement of the anaplastic lymphoma kinase gene (*ALK*) on chromosome 2p23 encoding the tyrosine kinase receptor ALK [13, 79]. Constitutive activation of ALK has been shown to occur in a number of different ways in NSCLC, but most often arises from a paracentric translocation on chromosome 2 (inv(2)(p21;p23)) that inverts a 5′ fragment of

the echinoderm microtubule-associated protein-like 4 (*EML4*) gene and fuses it to a 3′ fragment of *ALK* to form the fusion oncogene *EML4-ALK* [13]. Other known *ALK* translocation partners include the TRK-fused gene (*TFG*) and the kinase family member 5B (*KIF5B*) gene [158, 136, 3]. The associated adenocarcinoma subtype is defined clinically as *ALK*-positive NSCLC [127], with the dominant *EML4-ALK* fusion observed in 2.6% to 6.7% of the total NSCLC population, equating to approximately 40,000 to 100,000 newly diagnosed patients worldwide each year [13, 14, 128-131]. *ALK*-positive NSCLC occurs in lung fibroblasts [126], and generally affects patients who have never smoked. The ALK fusion kinase constitutively activates MAP/ERK, PI3K/AKT and other pathways, leading to the malignant phenotype [128, 139]. The *EML4-ALK* gene fusion has also been identified in breast and colorectal cancer [137], but has been rarely observed in other types of lung cancer.

Clinical detection of alterations in *ALK* is typically achieved using an FDA-approved FISH assay [276]. That assay can detect (but not differentiate between) all relevant *ALK* rearrangements and is currently used in determining patient eligibility for treatment with crizotinib (Xalkori; Pfizer, La Jolla, Calif), a small-molecule ALK inhibitor [15, 277-279]. The FISH assay is based on hybridization of two long, uniquely labeled probes to specific sequences on the 5′ and the 3′ side, respectively, of intron 19/exon 20 of the *ALK* gene, where the double strand break (DSB) in *ALK* associated with an oncogenic paracentric translocation event is most commonly observed. Co-localization of the red reporter on the probe annealing to the 3′ side of the breakpoint region with the green reporter annealing the other, producing a yellow (“fused”) signal indicating a normal *ALK* gene state within an imaged cell nucleus. Separation (break-apart) of red and green signals by at least two signal diameters indicates an *ALK* rearrangement. In certain cases, loss of a green signal in nuclei displaying a red signal may be observed, being indicative of a rearranged *ALK* locus [280]. Regrettably, cell imaging and signal interpretation are quite difficult in the FISH assay [15, 183] due to the fact that the inv(2)(p21;p23) translocation results in a relatively small change in the separation of the 5′ and 3′ specific probe signals. The resulting scoring uncertainties and signal instabilities lead to significant inter-observer variability, which in turn makes the assay prone to both false positives and false negatives. A FISH against *ALK* is therefore typically considered at least equivocally positive when greater than 15% of the nuclei within a sample contain an *ALK* rearrangement. The break-

apart FISH assay is also time intensive, relatively expensive, and requires a highly skilled technician [281].

Alternative methods for detecting NSCLC-predictive rearrangements in *ALK* have been developed. They include real-time polymerase chain reaction (qPCR) analyses of that portion of *ALK* encoding the kinase domain [179], multiplexed reverse-transcription qPCR (RT-qPCR) based detection of the *EML4-ALK* fusion transcript [147, 180], and immunohistochemical (IHC) staining of the chimeric EML4-ALK kinase [130, 140, 181]. Neither the qPCR assays nor the RT-qPCR methods have proven sufficiently robust [147] or comprehensive [180] to establish their clinical use in *ALK*-positive NSCLC testing. For example, the RT-qPCR assay detects only certain *ALK* fusion-gene variants, and reproducible results have proven difficult to obtain in FFPE tissue sections [142], presumably due to the inherent instability of RNA. Poor sensitivity and reproducibility likewise limit IHC tests for EML4-ALK positive NSCLC, including the IHC assay developed by Roche [180], due in part to variability in the quality of anti-ALK antibodies used [130] and the fact that differential expression of the EML4-ALK protein occurs at a low level [282]. Results from IHC assays are of similar quality to those for the break-apart FISH assay [184], though significant scoring discrepancies between the two assays have been observed [140, 181].

A reliable, simple, inexpensive, and sensitive method for assaying NSCLC-predictive rearrangements in *ALK* remains a need that could provide significant clinical benefits. Ideally, that method would preserve the strengths of the break-apart FISH assay, most notably the ability to detect all clinically relevant *ALK* rearrangements. Toward that goal, we describe here a droplet digital PCR (ddPCR) assay that uses multiplexed droplet digital PCR (ddPCR) to analyze the break-point region within *ALK* for any DSB. The assay partitions individual copies of normal or rearranged *ALK* into isolated sub-nL droplets. Three short templates within each copy are then amplified in isolation, with the pattern of fluorescent signals generated during template amplification permitting the detection and quantification of droplets harboring either an intact *ALK* gene or a clinically relevant fragment of *ALK* derived from a DSB within the 2.4 kbp breakpoint region spanning intron 19 and exon 20. During the required gDNA isolation process, (chemico-)mechanical fragmentation of the *ALK* gene within the breakpoint region can occur due, for example, to shear. A new data analysis tool is presented to quantify copies of *ALK* lost

due to non-biologic fragmentation and thereby enable accurate quantification of the frequency of *ALK* rearrangements in a gDNA specimen, as demonstrated here for various *ALK* positive cell lines and FFPE reference standards. Comparison of ddPCR assay and FISH results for the same specimens shows that the measured frequency of *ALK* alterations matches quantitatively. Finally, we show that the digital assay, when operated at a copies per drop (CPD) of 0.2 – 0.35, can be used to accurately identify and quantify *ALK* rearrangements to a detection limit of 0.25%. Relative to FISH, this assay offers a significant improvement in sensitivity in a lower cost, more time-efficient format.

3.2 Materials and Methods

3.2.1 Oligonucleotides

Primers and dual-labeled hydrolysis probes were purchased from IDT, Inc. Probes were HPLC purified and the primers purified by desalting. Purified forward (FP) and reverse (RP) primers, along with probes, were re-suspended to 100 µM in IDTE (10 mM Tris, pH 8.0, 0.1 mM EDTA) buffer and stored at -20 °C prior to use.

3.2.2 Cell Lines and *EML4-ALK* Reference Samples

HL60 cell line (negative control) was kindly donated by the BC Cancer Agency, and H2228 cell line was purchased from ATCC (CRL-5935). The H2228 cell line is derived from a 1998 female patient suffering from a NSCLC adenocarcinoma, and has been shown to harbor the *EML4-ALK* fusion gene on chromosome 2 and to be a variant 3, indicating that *ALK* exons 20 to 29 join *EML4* exons 1-6 to form the *EML4-ALK* fusion gene. All cell lines were cultured in HyClone RPMI 1640 media (GE, Healthcare), with 10% fetal bovine serum, 1% glutamine and 1% penicillin/streptomycin (all from Invitrogen, Canada).

The following FFPE and gDNA reference samples were purchased from Horizon Diagnostics, Inc.: gDNA having 50% *EML4-ALK* allele frequency (HD664), FFPE slides with

25% *EML4-ALK* allele frequency (HD231), and FFPE slides (HDC170) with a 33% *EML4-ALK* allele frequency (3 core slide - a wild type core + cores B and C at 50% allele frequency). For each sample, the % reported reflects the frequency of rearranged *ALK*.

3.2.3 Genomic DNA Purification

The limit of detection for this assay is largely defined by the limit of blank, which was determined by applying the assay to gDNA specimens that do not harbor an *ALK*-mediated translocation. That limit of blank (LOB) will be shown to be largely defined by uncertainties in the amount of *ALK* in the specimen that has undergone non-biologic fragmentation within intron 19/exon 20 during sample processing, most notably during gDNA purification. As demonstrated in chapter 2, assay performance is improved when mechanical fragmentation (shear) is reduced, and various gDNA extraction protocols were therefore evaluated once again, including the Qiagen Gene-read DNA FFPE kit, which utilizes a column extraction method, and the Qiagen MagAttract HMW DNA kit that exploits the use of magnetic beads. None of these methods proved suitable, as on average greater than 60% of the *ALK* template was lost to mechanical fragmentation.

The new gDNA extraction method described in section 2.2.3 was applied. Briefly, it uses precipitation, minimal pipetting and no vortexing to minimize shear-induced fragmentation. In addition, all centrifugation steps during and following DNA extraction are performed at mild conditions (2000 g). The new protocol, which reduces total non-biologic fragmentation of *ALK* to *ca.* 10%, is described for cultured cells in section 2.2.3 and for FFPE samples below.

FFPE samples used in this study were provided as core section mounted on glass slides from Horizon Diagnostics (Cambridge, UK). The DNA-containing section was scraped from the slide using a clean razor blade, deposited into a 2 ml tube containing de-paraffinization solution (Qiagen) and incubated at 56 °C for 3 minutes. The mild gDNA purification procedure described in section 2.2.3 was then followed for the FFPE samples from the cell lysis step onwards with the addition of the removal of the blue de-paraffinization solution before the protein precipitation step. FFPE samples can undergo gDNA damage from the fixing process [283, 284]. For example, deamination can occur which may prevent binding of probes. Despite the milder process used to

extract gDNA, larger losses in amplifiable material are expected from FFPE samples. Damage to gDNA recovered from cellular tissue or FFPE samples (cross-linking caused by formalin fixation is known to degrade gDNA [285]) may lead to underestimation of the number of biological DSBs [252].

3.2.4 Primer and Probe Design

The *ALK*-status assay developed using the platform is a one-well multiplexed ddPCR comprised of three qPCR-type reactions, each targeting a specific region of *ALK* to produce an amplicon of *ca.* 100 bp and a unique fluorescent signal generated through hydrolysis of the associated dual-labeled Taqman-type probe. The sequences and chemistries of the primers and probe used for each reaction are reported in Table 3-1.

The sequence for human *ALK* was obtained from the NCBI database (sequence NG_009445.1; http://www.ncbi.nlm.nih.gov/). From that sequence, forward and reverse primers were designed using Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi), then analyzed by primer-BLAST to identify sequence homology within the human genome database (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). The human genome database browser (http://genome.ucsc.edu/cgi-bin/hgGateway) was then used to identify common single nucleotide polymorphisms (> 1% minor allele frequency) that could serve to diminish primer (or probe) performance. Finally, the self-complimentary of possible primer and probe pairs was scored using the Exiqon OligoAnalyzer tool (https://www.exiqon.com/oligo-tools). When annealed to their fully complementary template, primers were designed to melt at 60 – 63 °C under PCR conditions (50 mM K^+, 3 mM Mg^{2+} and a total strand concentration (C_T) of either 200 or 900 nM) (http://biophysics.idtdna.com/), while probes where designed to melt at 66 – 70 °C. Finally, the optimal annealing temperature (T_a) was identified (60 °C) using a T_a gradient study on the Bio-Rad QX-100 ddPCR instrument.

3.2.5 Droplet Digital PCR (ddPCR) Assay Workflow

The droplet digital PCR procedure consists of three main steps: droplet generation, PCR amplification, and droplet reading. 20 µL samples for ddPCR analysis were prepared from 1X dUTP-free ddPCR™ Supermix for probes (Bio-Rad), 900 nM of each FP and RP, 250 nM of each FAM and HEX dual-labeled hydrolysis probe, 375 nM of the Alexa-488 dual labeled hydrolysis probe, and an aliquot purified gDNA (*ca*. 4 – 8 ng) containing 1500 – 3000 copies of *ALK*. Addition of gDNA was omitted for no template control (NTC) reactions.

For dynamic-range and limit of detection evaluation, serial dilutions of gDNA isolated from *EML4-ALK* positive H2228 cells into that from *ALK*-negative HL60 cells were prepared in IDTE buffer (IDT), with the total copies of *ALK* in each dilution quantified on a Bio-Rad QX100 ddPCR instrument [286].

A set of 20 µL samples prepared in this manner were then loaded into the sample well of the DG8 eight-channel disposable droplet generator cartridge (cat # 186-4008, Bio-Rad) using a 20 µL pipette tip (cat # 022491270, Eppendorf); 60 µL of droplet generation oil (cat # 186-3009, Bio-Rad) was loaded into the oil well for each channel. The cartridge was processed in a Bio-Rad QX-100 Droplet Generator to create an emulsion of each sample containing approximately 12,000 -15,000 readable droplets. The cartridge was removed from the droplet generator and 40 µL of each emulsified sample were transferred using a multichannel pipette to an Eppendorf Twin.tec semi-skirted 96 well PCR plate (cat # 951020362, Eppendorf). The plate was heat sealed with foil (cat #1814040, Bio-Rad) and the sample amplified in a CFX96™ thermocycler (95 ^{o}C for 10 min, 94 ^{o}C for 30 sec, 60 ^{o}C for 1 min) for 50 cycles, with the droplets then stabilized by a final incubation at 98 °C for 10 min. End-point fluorescence readings within each droplet were measured using the Bio-Rad QX100 droplet reader and presentation of the raw ddPCR data was performed using QuantaSoft analysis software (1.3.2.0).

For dynamic-range and limit of detection (LOD) evaluation, serial dilutions of gDNA isolated from *EML4-ALK* positive H2228 cells into that from *ALK*-negative HL60 cells were prepared in IDTE buffer (IDT), with the total copies of *ALK* in each dilution quantified on a Bio-Rad QX100 ddPCR instrument.

3.2.6 *ALK* Break-Apart FISH Assay

FISH was performed using the Vysis LSI *ALK* Break Apart Rearrangement Probes (Abbott Molecular). FISH consists of four main parts, fixation of cells to a surface, hybridization of DNA probes with DNA sample, washing and visualization. H2228 cells were cultured as described for 24 h, and then prepared for fixation by arresting during division using colcemid. Aliquots of ~ 3000 - 5000 arrested cells in 10ml RPMI 1640 culture medium were centrifuged (250 g) for 10 min and the supernatant removed. The cells were re-suspended in 10 mL of hypotonic solution (37 °C) and incubated for 25 min. Excess hypotonic solution was then removed by centrifugation (250 g for 10 min) and the cells fixed in up to 10 mL of a 3:1 methanol to acetic acid solution. That fixing process was repeated three times, with the resulting fixed cells then stored at 4 °C until use. When required for FISH, samples were centrifuged and supernatant removed to leave ca. 0.75 mL. The cells were re-suspended in fixative to a final volume of *ca.* 1 mL to achieve an opaque suspension. Aliquots of 8 µL were spotted onto glass microscope slides (Leica Surgipath Snowcoat Precleaned 1 x 3 x 1 mm). The FFPE reference slides (Horizon) were prepared along with the H2228 cells using the Vysis LSI *ALK* Break-Apart Rearrangement Probes, (Abbott Molecular) according to the manufacturer's instructions. Fifty interphase nuclei were analyzed using a Zeiss Axioimager Z2 epifluorescence microscope with a triple-band pass filter and DAPI counter stain added to aid visualization of nuclei.

3.2.7 Statistics

Basic metrics of assay performance (limit of detection (LOD), limit of blank (LOB), confidence interval (CI), dynamic range) were defined by computing means and standard deviations (SDs) from the mean from sets of *n* replicates, with the value of *n* specified. Statistical errors and the significance of recorded frequencies of biological disruption of *ALK* were determined using a paired Student's *t*-test.

3.3 Results

3.3.1 Assay Design

In this assay, which was again developed from the concepts described in section 2.3.1, individual copies of intact *ALK* or fragments of *ALK* are partitioned into isolated sub-nL droplets, typically at a CPD between 0.2 and 0.35. Into each droplet are also introduced the reagents needed to amplify three specific sequences within the *ALK* (proto-)oncogene (Figure 3-1) and then detect those amplicons with an associated uniquely labeled hydrolysis probe. The first reaction amplifies a short sequence lying 26 base pairs (bp) upstream of exon 19, with the amplification end-point detected by a 5′HEX labeled hydrolysis probe. The second amplifies a sequence 97 bp downstream of exon 20, with the end-point detected by a 5′(6-FAM) labeled probe. The sequences queried in these two reactions are *ca.* 2.4 kbp apart and flank the break-point region of *ALK*. Co-localized detection of both a FAM^+ and a HEX^+ signal in a droplet (denoted as a FAM^+HEX^+ droplet) is indicative of a copy of *ALK* having no disruption within the break-point region, while segregation of FAM^+ and HEX^+ signals represents either a biological or non-biologic (e.g., shear) disruption within that region. To estimate the frequency of non-biologic disruption of the *ALK* break-point region (which is expected to be less than that observed for *BCR* due to the shorter length of the *ALK* breakpoint region), the assay includes a reaction amplifying a third short sequence (detected by an Alexa-488 labeled probe) lying 2.4 kbp downstream of the second template queried. As the frequency of shear-induced fragmentation of gDNA is known to be length-dependent and stochastic [263, 264], the three reactions are spaced equidistantly to allow the separation of $Alexa^+$ and FAM^+ signals to serve as a surrogate measure of the frequency of non-biologic disruption of the *ALK* break-point region.

In this work, as noted, the assay is conducted at a CPD of between 0.2 and 0.35, and the distribution of copies into droplets is approximated by Poisson statistics, with most droplets harboring either 0 or 1 copy of intact *ALK* or a fragment of *ALK*. At a CPD of 0.2, for example, only 1.64% of all droplets contain 2 copies, 0.11% 3 copies, and 0.006% more than 3 copies.

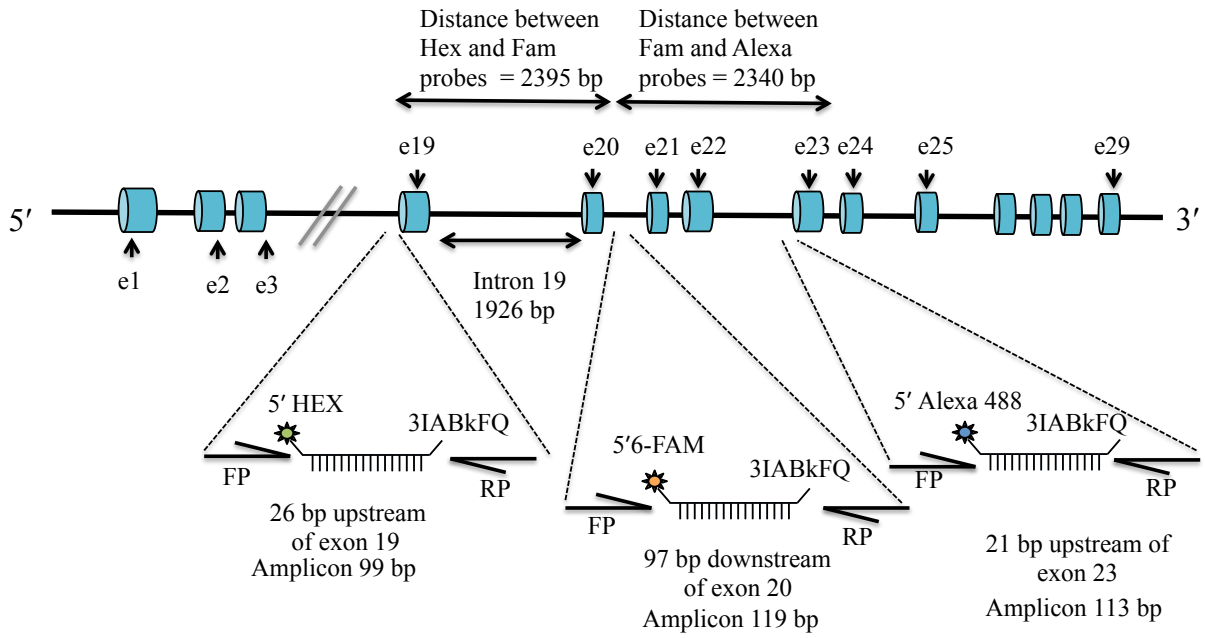


Figure 3-1. Digital PCR amplification targets within the *ALK* gene used to detect biologic and non-biologic cleavage within the breakpoint region.
The exon structure for the human anaplastic lymphoma receptor tyrosine kinase (*ALK*) gene on chromosome 2 (NCBI Reference Sequence: NG_009445.1). The assay amplifies short sequences 26 bp upstream of exon 19 (99 bp amplicon detected using a 5′HEX labeled hydrolysis probe) and 97 bp downstream of exon 20 (119 bp amplicon detected using a 5′6-FAM labeled probe). An additional control reaction monitored using an Alexa Fluor 488 labeled probe is used to quantify the quality (% shear) of the gDNA across the breakpoint region of *ALK*. The Alexa Fluor 488 labeled probe binds 2340 bp downstream of exon 20. The percentage of *ALK* copies that undergo shear between the Alexa Fluor 488 and 5′6-FAM probes during sample processing can thereby be quantified and used as a surrogate for shear frequency within the *ALK* breakpoint region of equal length.

Table 3-1. Forward primer (FP), reverse primer (RP) and probe sequences used in each amplification reaction comprising the ddPCR *ALK* status assay.

Amplicon Template	Reagent	Sequence
113 bp segment – 21 bp upstream of exon 23	FP	5′-gtatcctgttcctcccagtt-3′
	RP	5′-cccaatgcagcgaacaat-3′
	Probe[£]	5′-(Alex488N/acatccctctctgctctgcagca/3IABkFQ/-3′
119 bp segment – 97 bp downstream of exon 20	FP	5′-cagtgtaggggctgaatgt-3′
	RP	5′-cctgaatgtcaaggcttgtc-3′
	Probe[£]	5′-(6-FAM/agagccctc/ZEN/cctatgggcacc/3IABkFQ/-3′
99 bp segment – 26 bp upstream of exon 19	FP	5′-cgatgggaaggagcaagtag-3′
	RP	5′-cccactggggtattgacaac-3′
	Probe[£]	5′-HEX/tgggaccaa/ZEN/ctcaaaggagacc/3IABkFQ/-3′

[£] Alexa-488 (Alexa Fluor® 488), 6-FAM (6-carboxyfluorescein), HEX (hexachloro-fluorescein), and 3IABkFQ (3' Iowa Black® FQ) ZEN™ (IDT Internal Quencher) .

3.3.2 Quantification of *ALK* status

Figure 3-2 reports the raw data, presented as a 2D diagram, from application of our ddPCR *ALK*-status assay to a gDNA reference sample in which 50% of the total copies of *ALK* have translocated to form the *EML4-ALK* fusion gene. A cluster of empty droplets is observed in the lower left quadrant along with seven distinct non-overlapping positive clusters. The dense populations of both the HEX+ and FAM+Alexa+ droplet clusters provide clear visual evidence of ALK disruption. For a sample in which 50% of *ALK* has undergone a translocation event, one would expect to see equivalent populations in the $HEX^+FAM^+Alexa^+$ (intact *ALK*), HEX^+ and FAM^+Alexa^+ clusters in the absence of other disruption mechanisms. The somewhat larger populations of the HEX^+ and FAM^+Alexa^+ clusters relative to the $HEX^+FAM^+Alexa^+$ cluster, as well as the presence of a small $Alexa^+$ cluster, indicate non-biologic disruption of *ALK*, which must be taken into account to accurately analyze the frequency of biological disruption of *ALK* within its breakpoint region. Certain *ALK* translocations are known to result in partial or complete loss of the 5′ fragment of *ALK* (i.e., that portion that does not participate in the *EML4-ALK* fusion) [280, 287], which can eliminate the fragment of *ALK* detected by the HEX^+ labeled probe in the assay. Loss of signal can t occur, resulting in a concomitant change in the 2D diagram and the need to account for this effect in the data analysis.

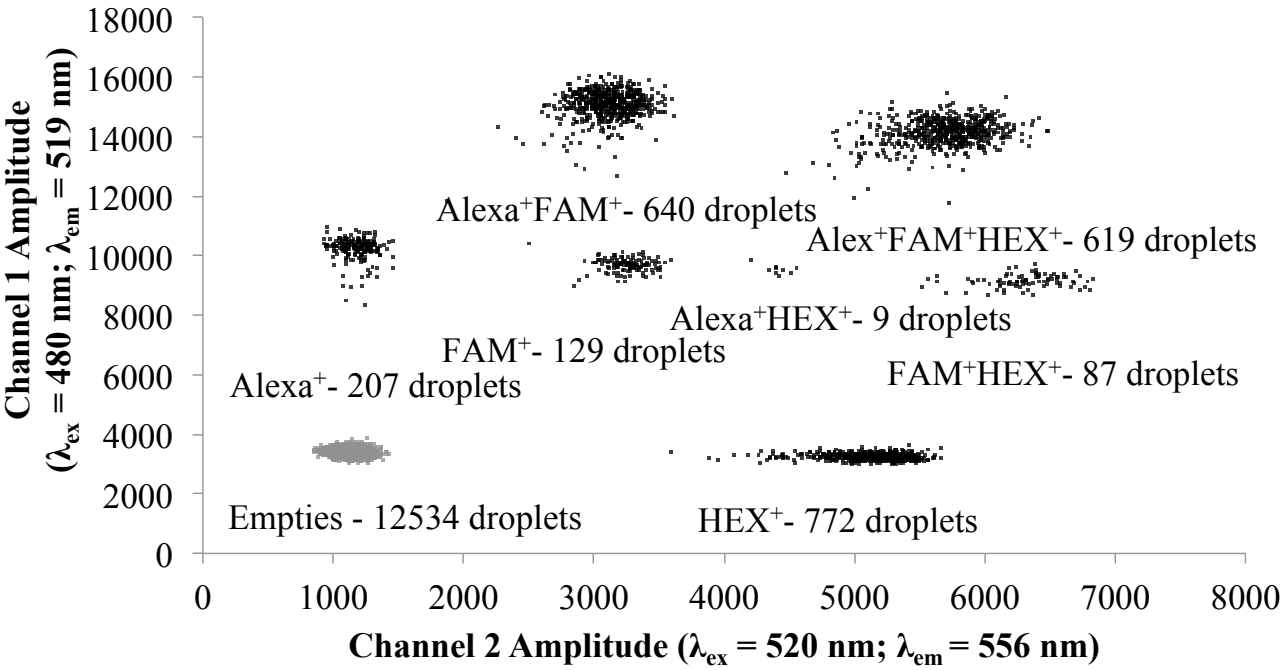


Figure 3-2. Droplet Digital PCR *ALK* status assay output for a gDNA reference sample in which 50% of *ALK* exhibits rearrangement.
Data recorded in the assay partition into 8 clearly defined non-overlapping clusters. The assay output for the 50% *ALK* positive (all in the form of *EML4-ALK*) gDNA reference sample: 2463 droplets contain at least one copy (CPD of 0.179); Analysis of the raw data using the model embodies in equations 3.1 – 3.3 yields values for *b* and *s* of 49.8% and 13.3%, respectively. The sample shows no loss of HEX^+ signal, and the % of rearranged *ALK* that was recorded in the ddPCR assay equals that specified by the supplier of the reference material.

The arguments made above indicate that alterations to a copy of *ALK* within the 4.8 kbp region queried (Figure 3-1) can arise through a biologic double stranded break within the breakpoint region or a non-biologic break (shear) within the queried region; those rearrangements in *ALK* may also result in loss of end-point HEX^+ signal. The possible states of each copy of *ALK* resulting from these three possible events are shown in Figure 3-3. It begins with segregation of *ALK* copies into those that have (*b*) or do not have ($(\bar{b})$ = (1-*b*)) a biological DSB. Each of those copies may (*s*) or may not ($\bar{s}$) be disrupted by a non-biologic mechanism, and/or may have undergone alterations resulting in loss (*l*) of HEX^+ signal. Figure 3-3 can be combined with the probabilistic modeling approach defined in chapter 2 to connect the unknown variables *b*, *s*, and *l* to the droplet populations for each cluster (e.g., Figure 3-2):

$$\frac{Alexa^{+}FAM^{+}HEX^{+}}{Total\ ALK} + \frac{FAM^{+}HEX^{+}}{Total\ ALK} = (1-b)(1-s) \quad (3.1)$$

$$\frac{HEX^{+}}{Total\ ALK} + \frac{Alexa^{+}FAM^{+}}{Total\ ALK} = b(2-l-2s^{2}) + s\,(2-4b-s) \quad (3.2)$$

$$\frac{FAM^{+}}{Total\ ALK} + \frac{Alexa^{+}}{Total\ ALK} = sb(3-2s-sb) + s(1+s) \quad (3.3)$$

The derivation of equation 3.1 – 3.3 again makes use of the fact that certain states, such as $\bar{s}s\bar{b}$ and $s\bar{s}\bar{b}$, cannot be distinguished. Those states are treated as one state in the model to avoid double counting. In addition to the droplet counts for each cluster, solution of equations 3.1 – 3.3 requires knowledge of the total copies of *ALK* in the sample, which is taken as the average of all $Alexa^{+}$ droplets and all FAM^{+} droplets. An iterative least squares fitting routine is used to produce optimal goodness of fit utilizing all six positive clusters to calculate the 3 variables with constraints on *s*, *b* and *l* of ≤ 1 and ≥ 0. Stable numerical solutions can be achieved using the "Solver" subroutine of Excel (Microsoft®) based on the generalized reduced gradient (GRG) method. Solution of the model for the data set shown in Figure 3-2 yields a mechanical disruption frequency of s = 13.1(±0.4)%, and a biological DSB frequency of b = 48.1(±1.1)% (SD for 6 replicates). The latter frequency (b) agrees quantitatively with that reported by Horizon Diagnostics Inc. for the reference material used. An example of method used to compute the right-hand sides of equations 3.1 – 3.3 is provided in Appendix B.

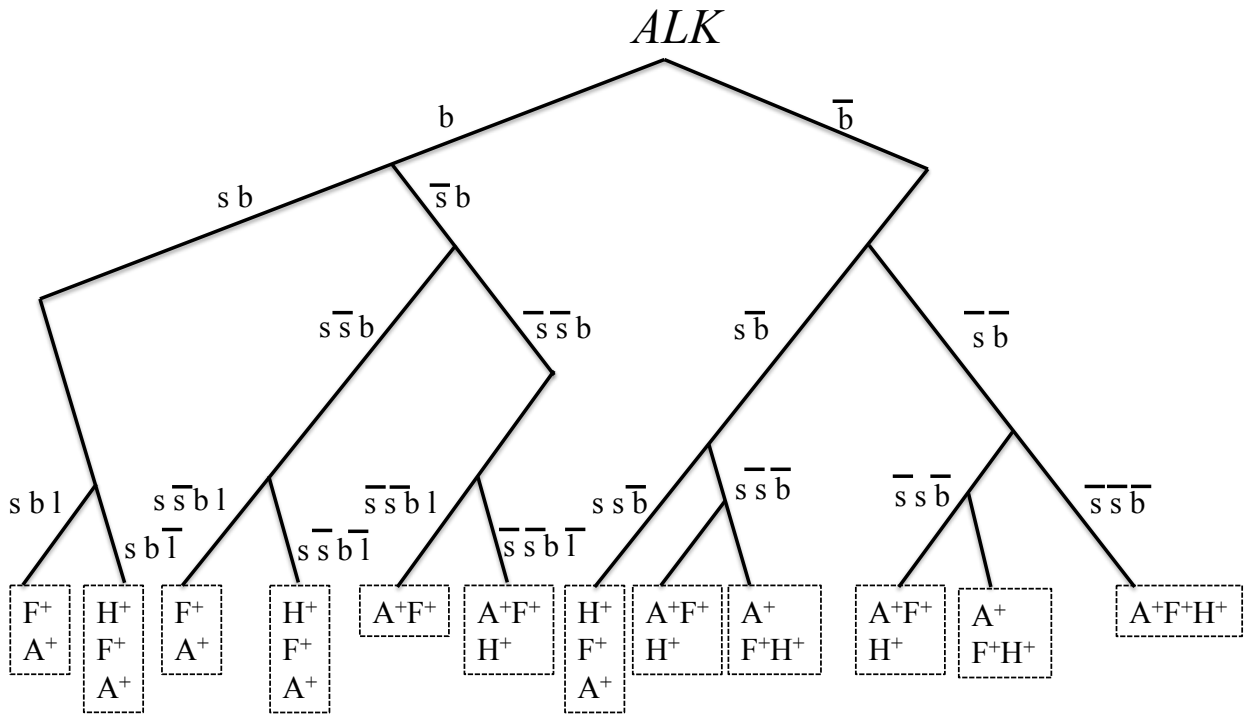


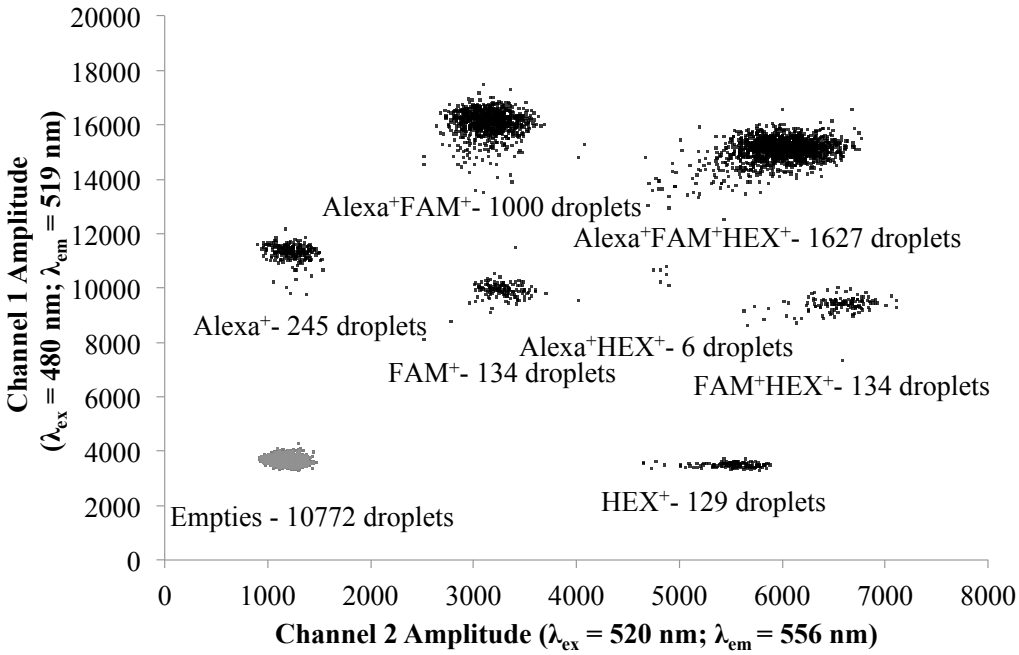
Figure 3-3. Map of all potential states of an *ALK* gene within the ddPCR based translocation assay.
Analysis of all possible states of a *ALK* gene copy within the assay resulting from all relevant combinations of translocation, shear and loss (disruption) of HEX signal, where b = biological DSB (translocation), s = shear (mechanical fragmentation or degradation) and l = disruption or loss of Hex signal. A bar above a letter indicates that the event has not occurred within the gene copy. The clusters within the ddPCR output data that are populated as a result of a particular set of events within the *ALK* gene copy are shown in the hashed boxes, where A^+ = Alexa$^+$, F^+ = FAM$^+$, A^+F^+= Alexa$^+$FAM$^+$, F^+H^+ = FAM$^+$HEX$^+$ and $A^+F^+H^+$ = Alexa$^+$FAM$^+$HEX$^+$

3.3.3 Analyzing *ALK* status in *EML4-ALK* Positive and Negative Samples

The H2228 cell line is heterozygous for *EML4-ALK* and is classified as having a variant 3 translocation (containing exons 1 to 6 of *EML4*) [280, 287] characterized by a loss of the 5′ fragment of *ALK*, which is observed in FISH [280]. A loss in HEX$^+$ signal in the ddPCR assay is expected. Figure 3-4 reports the output from the ddPCR *ALK* status assay applied to gDNA

purified from the H2228 cell line. Near complete loss (l = 99.7%) of the HEX^+ cluster is indeed observed, in accordance with FISH data [280].

The frequency of biological disruption within the breakpoint region of *ALK* for the H2228 cell line was also recorded using the ddPCR assay and compared to that provided by the break-apart FISH assay, typical results from which are shown in Figure 3-4. Data analysis using the probabilistic model embodied in equations 3.1 – 3.3 yields a non-biologic disruption frequency of 10.3% (significantly less than that observed for *BCR*, as expected), and a biologically disruption frequency of 35.7%, which again matches the value recorded by FISH (35%).



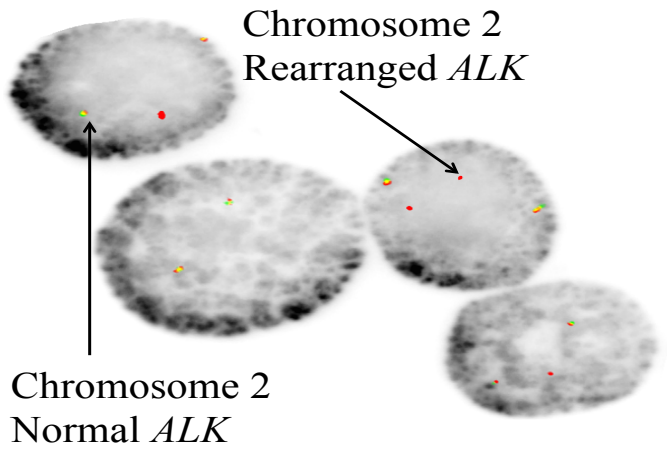


Figure 3-4. Droplet digital PCR and FISH assay output for the *EML4-ALK*$^+$ H2228 cell line. DdPCR assay operated at a CPD of 0.266 ± 0.006, (SD based on 6 replicates). The high density of the Alexa$^+$FAM$^+$ cluster (1000 droplets) provides direct visual evidence of *ALK* rearrangement(s), with the associated loss of signal (droplets) within the HEX$^+$ cluster consistent with the fact that the translocation process in this cell line eliminates a fragment of *ALK* lying downstream of exon 19 that includes the sequence detected in the assay by the HEX-labeled probe. The ddPCR assay for the H2228 sample reports 35.7% of *ALK* has undergone rearrangement, and 10.3% of *ALK* in the sample has been non-biologically fragmented. FISH assay results for 50 probed and imaged nuclei of H2228. Most cells contained a single break-apart signal (isolated red), as well as loss of green signal indicating genetic disruption downstream of exon 19; two fused signals (yellow) corresponding to an intact *ALK* gene are generally also observed.

As a negative control, the ddPCR assay was applied to gDNA from the HL60 cell line (Figure 3-5). A densely populated $Alexa^{+}FAM^{+}HEX^{+}$ cluster (2162 droplets) and sparsely populated HEX^{+} and $FAM^{+}Alexa^{+}$ clusters are observed, with comparison of, for example, Figures 3.2 and 3.5 showing that simple visual inspection of cluster populations provides clear evidence of biological disruption within the *ALK* breakpoint region. Note that sparse population of the $Alexa^{+}FAM^{+}$ and $FAM^{+}HEX^{+}$ clusters, as well as within the secondary clusters (e.g., $Alexa^{+}HEX^{+}$), is observed in the negative control due to a mechanical disruption frequency of 7.8% for this sample. The model records a biologic disruption frequency of 0.01(±0.02)% (mean ± SD calculated from $n = 24$ replicates) for HL60 gDNA.

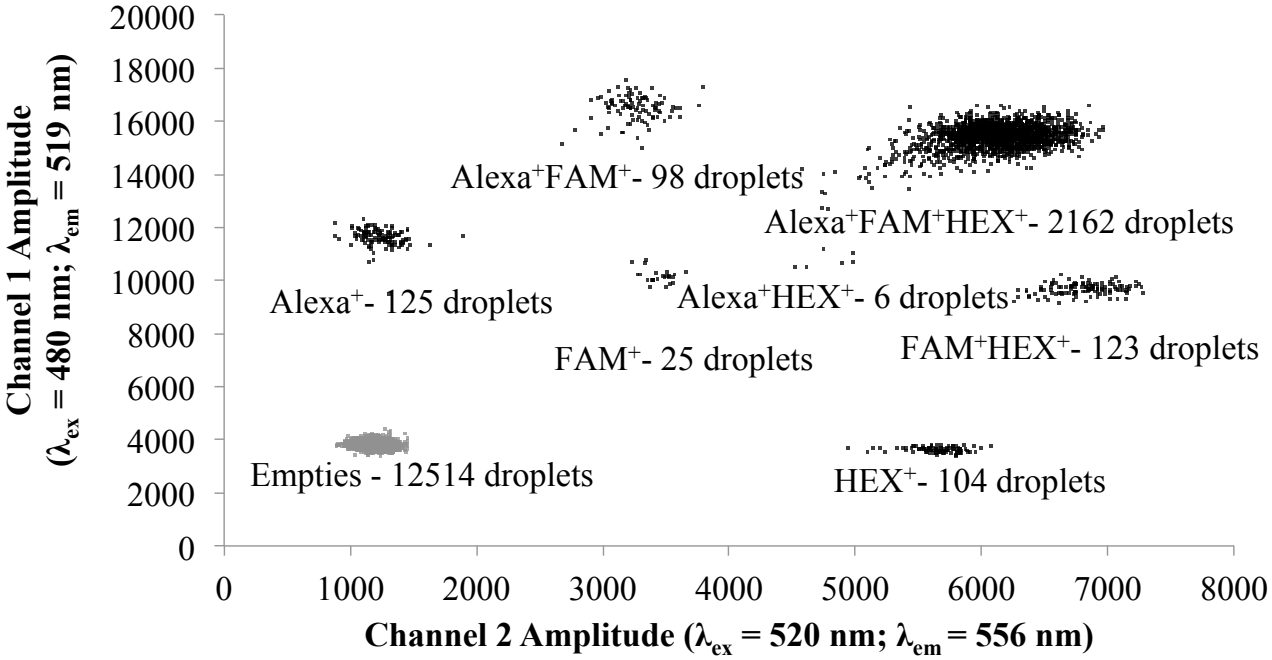


Figure 3-5. Droplet digital PCR assay output for the (*EML4-ALK* negative) HL60 cell line. DdPCR assay operated at a CPD of 0.192, the high density of the $Alexa^{+}FAM^{+}HEX^{+}$ cluster (2162 droplets) indicates an absence of *ALK* rearrangement(s). The ddPCR assay for the HL60 sample reports 0.01(± 0.02)% of *ALK* has undergone rearrangement, with the low level population of the remaining clusters due to mechanical fragmentation or degradation of *ALK* during sample processing.

3.3.4 Assay Limit of Detection when applied to gDNA from Cell Lines

Serial dilutions (n = 3 for each dilution) of H2228 gDNA in HL60 gDNA down to an *ALK* frequency of 0.25% were used to define the limit of detection of the ddPCR assay (Figure 3-6). The assay was first applied to gDNA (n = 24) from HL60 alone (negative control), from which the limit of blank (= 0.05%) was determined as the mean + 95% CI of the set of b values recorded for the negative-control replicates. Measured *ALK* disruption frequencies correlate linearly with expected disruption frequencies down to the measured limit of detection of ~ 0.25% determined based on the limit of blank and a paired Student's t-test. In these experiments, the CPD was set between 0.2 and 0.3; the reported LOB and LOD apply to those conditions. The results show that the ddPCR assay can reliably identify and quantify an *ALK* positive sample through the detection of as few as *ca.* 3 copies of rearranged *ALK* in a sample containing at least 1500 total copies of *ALK*.

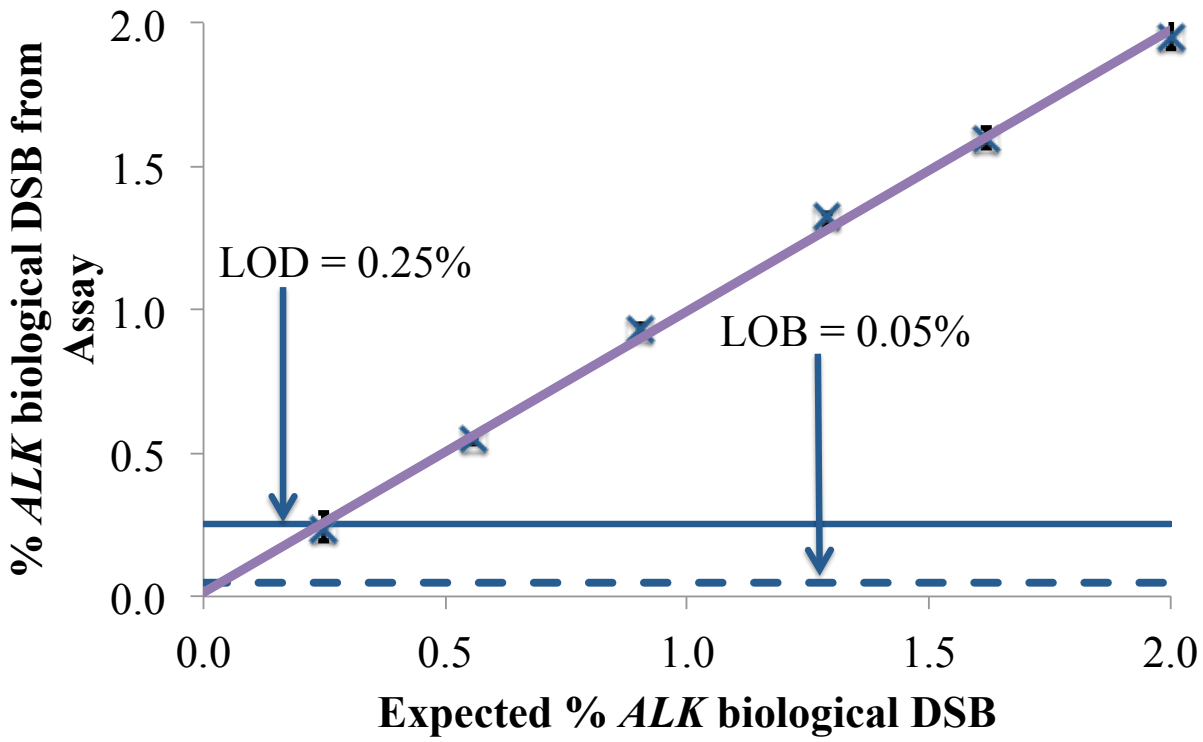


Figure 3-6. Accuracy, precision and detection limit of the ddPCR-based *ALK* status assay. Measured biologic disruption frequencies and standard deviations (n = 3) are plotted versus expected *ALK* rearrangement frequencies for serial dilutions of H2228 in HL60 gDNA. Significant linear correlation ($R^2 \geq 0.99775$; $P < 0.0001$) between the measured and expected *ALK* rearrangement frequencies is observed down to 0.25%(limit of detection (LOD). Replicates (n = 24) of gDNA from the *ALK* negative HL60 cell line were used to define the mean and standard deviation of false positives, from which the 95% confidence interval was determined and used to define the limit of blank (LOB = 0.05%; blue hashed line). At a CPD = 0.2, statistically significant *ALK* rearrangement frequencies can be obtained to an LOD of 0.25% based on the measured LOB and a paired Student's *t*-test.

3.3.5 *ALK* Status Assay on FFPE Reference Samples

Fine-needle and other biopsies of lung tumors are typically prepared and stored as FFPE specimens, and the associated fixation and embedding chemistries are known to degrade gDNA quality [288, 289]. Two FFPE reference samples were analyzed using both the ddPCR assay and FISH. For each sample, the frequency of biological disruption recorded by the ddPCR assay matches that provided by the reference material vendor (Horizon Diagnostics) and by FISH (Table 3-2). For example, for the 25% biological disruption FFPE reference, the ddPCR *ALK*

status assay records a *b* of 25.4% (raw data in Figure 3-7), while FISH (representative images in Figure 3-7) yields a value of 22% (Table 3-2). The non-biologic disruption of *ALK* in this sample was 28.8% as recorded by the ddPCR assay. For reasons explained in chapter 2, a higher frequency of non-biologic disruption in expected in samples.

Extraction of gDNA from the second FFPE reference (33% biologic disruption frequency) was less efficient due to the manner in which the core sections were presented on the slide (i.e., as 3 independent sections, each having to be scraped from the slide separately). An unusually high mechanical disruption frequency of 43.3% was consequently recorded from the ddPCR assay data (Figure 3-8). Nevertheless, the ddPCR assay reports a *b* of 33.1%, in quantitative agreement with the reference specifications; a value of 27% was recorded in the corresponding FISH assay (Figure 3-8). These results suggest that the ddPCR assay can tolerate relatively poor quality gDNA without significant loss in accuracy. Example of raw data and probabilistic analysis is provided in Appendix B.

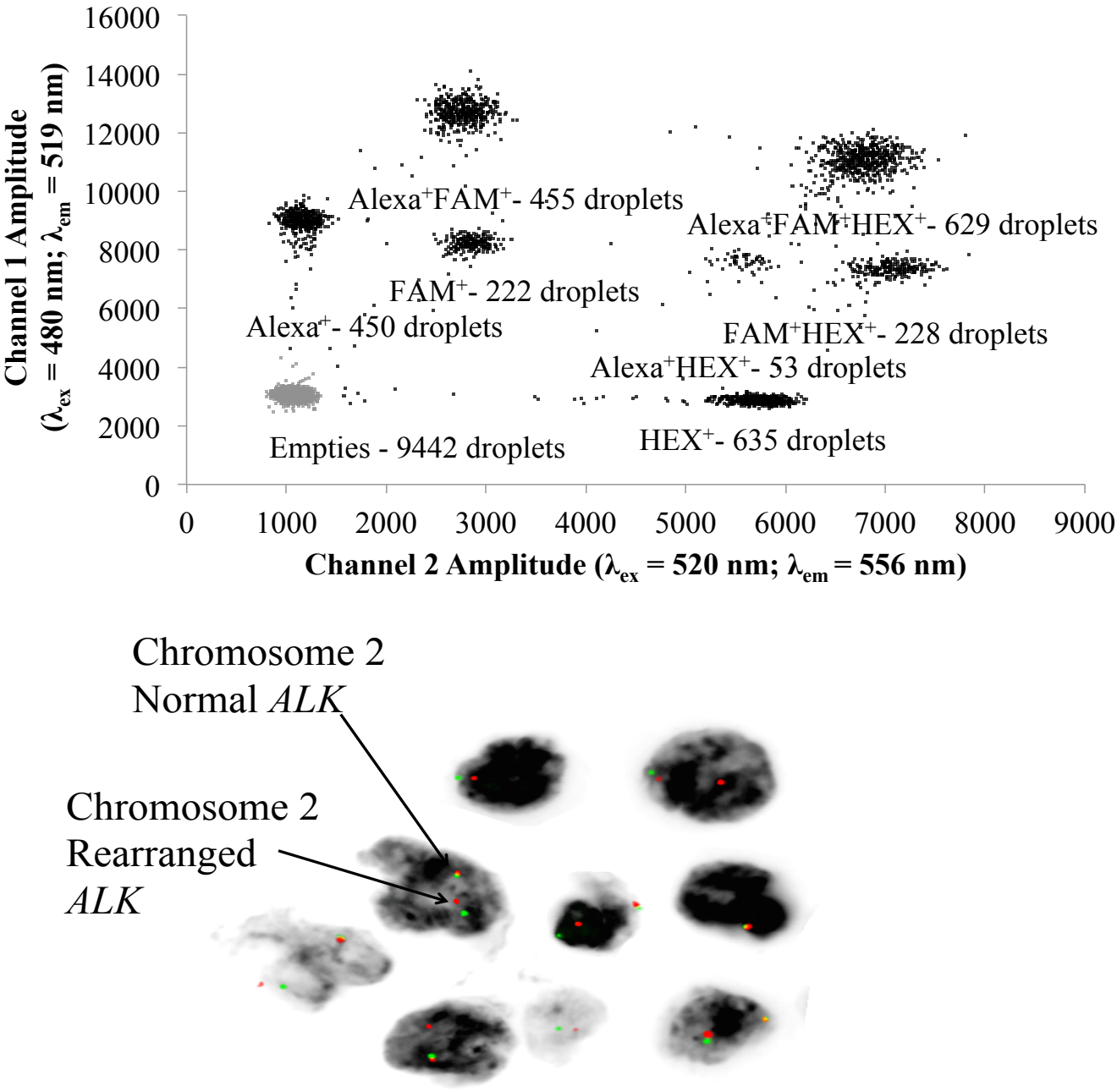


Figure 3-7. Droplet digital PCR and FISH assay output for FFPE sample having an *ALK* rearrangement frequency of 25%.
The raw ddPCR data (CPD of 0.183) and associated data analysis record that 25.4% of *ALK* is rearranged, in accordance with the reference value, and that 28.8% of the *ALK* has undergone non-biologic fragmentation within the probed region. FISH assay results for 50 probed and imaged nuclei. 50% of the cells have 2 normal fused signals, while the remaining 50% have one normal chromosome 2 fused signal and one translocated (separate red and green) signal. The FISH assay records a rearranged *ALK* frequency of 22% (Table 3-2).

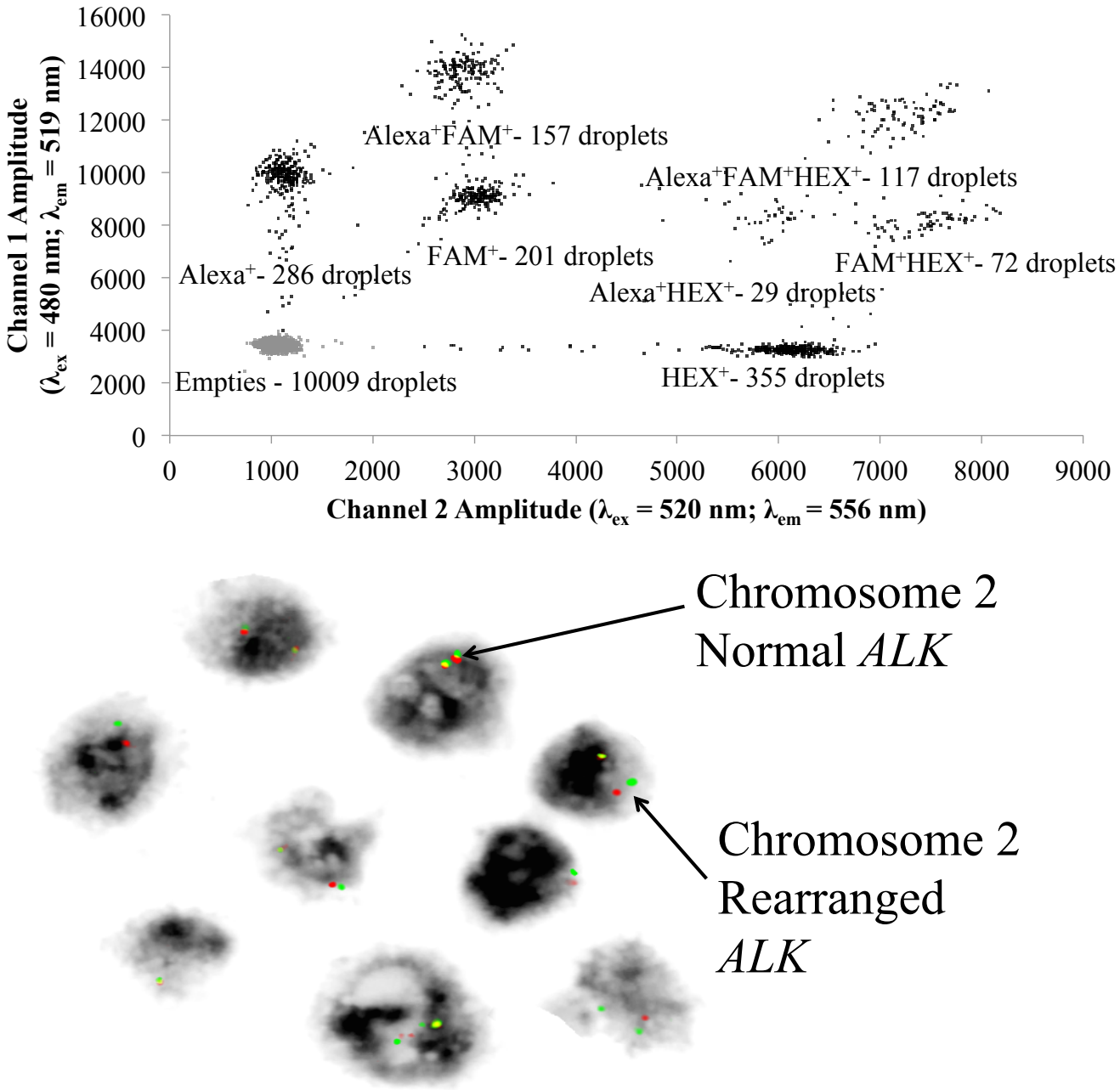


Figure 3-8. Droplet digital PCR and FISH assay output for FFPE sample having an *ALK* rearrangement frequency of 33%.

The raw ddPCR data (CPD of 0.115) and associated data analysis record that 33.1% of *ALK* is rearranged, in accordance with the reference value, and that 43.3% of the *ALK* has undergone non-biologic fragmentation within the probed region. FISH assay results for 50 probed and imaged nuclei. The cores where imaged, with the first core negative for *ALK* rearrangements, while in the remaining two cores 50% had one normal chromosome 2 fused signal and one translocated (separate red and green) signal.

Table 3-2. Comparison of ddPCR-based *ALK* status assay results to benchmark FISH data for H2228 and HL60 cell line and FFPE reference samples.

Cell line	FISH[£] % translocation (*EML4-ALK*)/ ((*EML4-ALK*)+*ALK*)	Assay % translocation
H2228	35	35.1 ± 0.6
25% FFPE	22	26.1 ± 1.3
33% FFPE	27	33.7 ± 0.9
50% gDNA		49.2 ± 1.1
HL60	(Negative sample) 15	0.01 ± 0.02

[£] FISH data (50 nuclei) were analyzed by calculating translocation as the number of *EML4-ALK* fusions divided by the sum of the normal *ALK* transcripts and *EML4-ALK* fusions. The results show a close comparison between the FISH and the *ALK* ddPCR assay. The gDNA was only run in the *ALK* status ddPCR assay.

3.4 Discussion

Obtaining biopsy specimens of potentially tumorous lung tissue is difficult and painful for patients. Those biopsies tend to be small, making identification of a translocation difficult. To lower costs and ease patient discomfort [147], surgical resection of the tumor is generally avoided, particularly at advanced stages of NSCLC, in favor of a fine needle biopsy, a trans-bronchial biopsy, bronchial washing or pleural fluid collection. Much of the FFPE specimen prepared from such samples is typically devoted to pathological testing, so only minimal tumor tissue may be available for molecular testing of *ALK* [142]. To satisfy Medicare insurance requirements in the US, testing for oncogenic *EGFR* mutations is performed before *EML4-ALK* testing [182], further reducing the available sample.

A measured *ALK* frequency of ≤ 15% is taken as *ALK*-negative for the FISH assay, mostly due to the inability to access sufficient intact cells from a FFPE tissue sample to enable more sensitive detection - 50 cells are generally counted due to the limited access to biopsy

material. For a non-reciprocal translocation such as that involving disruption of *ALK*, application of the testing platform described in this thesis could provide a clinically applicable solution to this challenge. Here, a ddPCR assay based on that platform has been created for initial diagnosis of *ALK*-positive NSCLC. When applied to *EML4-ALK* positive cell lines and *ALK*-positive translocated reference samples, it provides a quantitative measure of biological disruption of *ALK* to frequencies down to a LOD of 0.25%. The ddPCR *ALK*-status assay is nearly two orders of magnitude more sensitive than the break-apart FISH assay or current IHC tests [290], providing a significantly improved detection limit with a much smaller degree of uncertainty, due in large part to the fact that sufficient gDNA can generally be extracted from available biopsy material to operate the ddPCR assay at a CPD between 0.2 and 0.35. Moreover, the assay is able to quantify *ALK* positive samples in cases where the translocation event results in loss of some or all of the *ALK* sequence lying to the 5′ side of the breakpoint, as demonstrated on the H2228 cell line. Such losses are known to occur in certain *ALK* positive NSCLC cases, indicating a rearranged *ALK* locus [280].

The robust data processing methodology essential to the platform has been shown to properly account for the impact of non-biologic fragmentation of *ALK* on the number of positive data clusters observed and the distribution of droplets among them. The method allows the de-convolution of the raw digital data to fully evaluate and discriminate between the number of copies of *ALK* that have been disrupted either biologically or non-biologically within the breakpoint region. Although the data processing method accounts for non-biologic fragmentation of *ALK*, the LOB (and thus the LOD) of the ddPCR assay is largely defined by the uncertainty in the % non-biologic fragmentation recorded.

Improved detection of *ALK* modifications leading to an aberrant ALK protein could directly benefit NSCLC patients whose tumors are positive for the *EML4-ALK* fusion gene and qualify for treatment with crizotinib [277-279]. A recent study by Johung et al., [291] found that patients with *ALK*-positive NSCLC and a brain metastasis have prolonged survival when compared to NSCLC patients with no *ALK* rearrangement and a brain metastasis. This further highlights the importance of identifying the *ALK* rearrangement in patients to ensure that they receive the most appropriate and effective treatment.

Finally, Zheng et al., [292] analyzed 319 FFPE specimens from a cohort of NSCLC adenocarcinoma patients using in-depth anchored multiplex PCR coupled with next generation sequencing. All *ALK* rearrangements identified resulted in fusion to *EML4*. However, subsequent studies have found rearrangements that create other *ALK* gene fusions, as well as fusions involving different tyrosine kinase genes, such as *ROS1*, in tumors of patients with NSCLC [293]. Similarly, Rikova et al., [158] have reported on a *TFG-ALK* fusion, while Takeuchi et al., [136] discovered one *KIF5B-ALK* fusion among FFPE samples from 96 NSCLC patients. Thus, although *ALK* translocations in NSCLC patients generally result in fusion to *EML4*, other rare *ALK* fusion partners may occur within this disease. Like the break-apart FISH assay, the ddPCR assay reported here analyses *ALK* status and is not affected by the fusion partner. The assay could therefore be applied to other *ALK* fusion gene partners where there is a clinical benefit in determining the translocation frequency.

To conclude, the ddPCR *ALK* status assay offers vastly improved specificity relative to the FISH and IHC assays currently used in clinics. As it is significantly cheaper (~ $6 per sample) and less labor intensive (~ 8 hour turnaround time) than FISH, the assay should be of value to clinics and hospitals for initial diagnoses of *ALK*-positive NSCLC. Planned clinical validation studies will be required as part of that technology translation. More fundamentally, the results reported here and in chapter 2 provide the first demonstration that ddPCR assay constructed using the platform concepts central to this thesis can be used to analyze gDNA for the presence of either reciprocal or non-reciprocal (inversion) translocations.

Chapter 4: Conclusions and Future Work

4.1 Conclusions

Digital forms of the polymerase chain reaction are being exploited in many areas of cancer analysis due to their high level of sensitivity and ability to precisely quantify genomic DNA (gDNA). The commercialization of dPCR machines has enabled analysis of various cancer biomarkers, including somatic mutations [193, 214, 215], oncogenes [294], copy number variations (CNVs) [195, 216, 217, 295], and loss of heterozygosity (LOH) [210]. For instance, breast cancer metastasis can be identified by quantifying the abundance of somatic mutations in circulating tumor DNA (ctDNA) using chip-based digital PCR (cdPCR) coupled with next generation sequencing (NGS) [219]. This and other examples exploit the ability of dPCR [204, 295] to quantify low concentrations of allelic biomarkers with increased precision relative to conventional qPCR [16, 204, 296]. However, analyses of translocations within gDNA using dPCR have been minimal, with the only study reported to date by Shuga et al., [223], who successfully used a form of dPCR in combination with NGS to detect the t(14;18) reciprocal translocation associated with follicular lymphoma. In that work, dPCR was used only for biomarker amplification, not the detection and quantification of the translocation, which was achieved by the coupled NGS instrument.

This thesis describes the development of a general ddPCR based platform for detecting oncogene rearrangements and associated translocation events at the gDNA level. We show that the method can be used to detect rearrangements associated with either reciprocal or paracentric (inversion) translocations, which represent the two most common translocation classes associated with cancer and cancer progression [1, 2]. The utility of this platform has been demonstrated through the development of two new assays: the first providing highly sensitive detection of the *BCR-ABL* fusion gene and the associated reciprocal t(9:22)(q34;q11) translocation that are the hallmark of chronic myelogenous leukemia, and the second providing detection of rearrangements in *ALK* associated with a paracentric translocation, most notably the (inv(2)(p21;p23)) translocation and coupled formation of the *EML4-ALK* fusion gene, associated with *ALK*-positive non-small-cell lung cancer (NSCLC). In both of these cases, the platform yielded an assay capable of identifying and quantifying the relevant rearrangement down to a

detection limit of 0.25%, making each assay far more sensitive than the corresponding FISH or IHC methods currently used clinically. More importantly, the platform reported here represents the first demonstration that a commercial ddPCR instrument can be used on its own to analyze gDNA for the presence of either reciprocal or non-reciprocal chromosomal translocations associated with cancer.

These assays exploit the single molecule counting capabilities of ddPCR to detect minute amounts of genetic material at a performance that surpasses other, often considerably more expensive, quantitation methods. The basic concept of absolute quantification of oncogenes by single molecule counting was first forwarded by Sykes et al., [297], who utilized Poisson statistical analyses in a limiting dilution assessment. They postulated that by diluting and distributing DNA copies into a very large number of individual PCR reactions, quantification of the number of copies in a specimen can be achieved through analysis of the number of partitions displaying positive amplification using an appropriate amplification reporter. This partitioning occurs randomly and independently [194], allowing for the dilution of a sample such that each partition initially contains no more than one target molecule to be analyzed. The distribution of copies among partitions can then be reasonably well estimated by Poisson statistics. As has been demonstrated in this thesis work for translocation events, a key advantage of ddPCR is that the raw data can often be interpreted directly to gain useful and reliable clinical information, as has been shown in other dPCR applications [298].

Current commercial dPCR instruments create several thousand to several million partitions – the droplet digital PCR instrument used in this work creates approximately 15,000 readable droplets per well [213]. This exceptional multiplexed reaction density has been shown to deliver greatly improved sensitivity, precision and reproducibility relative to qPCR [201, 204]. This is due in part to the fact that the signal-to-noise ratio tends to be much higher because of the end-point detection method used and the fact that the limiting dilution and partitioning process generally reduces the concentrations of contaminant species, including non-target background DNA, that might inhibit template amplification.

The platform developed and presented in this work exploits all of these advantages, and adds to them the concept of simultaneously interrogating multiple segments of a target (onco)gene or region within a chromosome by analyzing the resulting distribution of end-point signals from amplicons among partitions.

The ddPCR *BCR* status assay (Chapter 2) shares strengths with the current *BCR-ABL* FISH assay. Similar to FISH, for example, the ddPCR assay allows for direct visual identification of the *BCR* rearrangement in the raw data. By coupling that data with a general and powerful probabilistic model that connects the raw data to the true distribution of states of *BCR* copies, a greatly improved detection limit and level of precision is realized. Moreover, the assay is highly reproducible and can be conducted at lower cost (factor of *ca.* 10) and with faster turn-around time (assay takes *ca.* 6 to 8 hours to complete, as opposed to 2 days for FISH).

Those same strengths are retained in the ddPCR *ALK* status assay (described in Chapter 3) that arguably addresses a more pressing clinical need. The current *ALK* break-apart FISH assay achieves a relatively poor limit of detection (15%) at a high degree of uncertainty, while also being both time-consuming and requiring considerable technical expertise. Alternative assays such as immunohistochemical (IHC) methods fair no better. The ddPCR assay developed in this work therefore has the potential to greatly improve detection of *ALK*-positive NSCLC and enable prognosis earlier in the disease progression.

There are recognized clinical benefits of robust, highly sensitive and highly specific diagnostic assays for detecting mutations, in part because they can allow for timely intervention that improves health outcomes for cancer patients [299] while minimizing the stress of uncertainty and waiting on the patient [300]. By providing those clinical benefits in a lower cost and faster turn-around format, the ddPCR based platform described here may also provide improved health economics [301].

The primary competing technology, NGS, may prove clinically applicable to translocation monitoring as well, but currently does not offer comparable detection limits [302], even though a larger amount of sample is required compared to ddPCR [214, 303, 304]. Finally, ddPCR technology is advancing in much the same ways as NGS, with hardware, chemistry and throughput improving at rates approaching Moore's Law. Thus, these two technologies soon may both be considered essential tools in monitoring and diagnosing diseases, and in the future will likely find both complementary and partnered applications [204, 205, 207].

4.2 Future Work

The *BCR* status assay developed here and described in chapter 2 is specific to the M-*BCR* region of the *BCR* gene. While the vast majority of CML patients will have a breakpoint within this region, approximately 2% are known to instead have a breakpoint within the m-*BCR* (minor-*BCR*) region (see Figure 2-3). Extension of the ddPCR *BCR* status assay to include the m-*BCR* region is in theory possible through a multiplexed approach that analyses both the M-*BCR* and the m-*BCR* (54.4 kbp) regions using an appropriate set of equidistantly spaced templates, which in the case of the m-*BCR* might require up to 6 nested reactions. The issue of loss of copies due to shear (non-biologic) fragmentation would obviously become more significant in this case, so further research and development would be required to ensure the amplicon sequences chosen and the reporter designs preserve good sensitivity and reliability. But if realized, the extended assay has the potential to be of significant benefit to patients with either CML or AML who carry a Philadelphia chromosome formed via a double-stranded break in the M-*BCR* or m-*BCR* region.

A potentially more impactful advance, at least from the perspective of clinical need, would be the development of a ddPCR assay offering highly sensitive detection in gDNA of the *BCR-ABL* fusion gene; that assay would be suitable for use in minimal residual disease (MRD) monitoring. Regularly scheduled MRD monitoring of CML in patients undergoing or following treatment is currently conducted either by FISH, which offers poor sensitivity, or RT-qPCR, which provides an adequate LOD (frequency ~ 10^{-4}) [207, 305], but suffers from intrinsic difficulties in data standardization and interpretation [306]. In particular, absolute quantification of *BCR-ABL* abundance is not possible, and clinical evaluation is made using an internationally adopted relative (calibration) scale. Interpretation of the RT-qPCR results close to the LOD is quite difficult due to the high coefficient of variation (CV) of the assay; a 1 log increase in *BCR-ABL* abundance is required before a relapse is defined [199, 202]. These problems could be avoided by establishing a means of directly detecting the fusion, and thus leukemic load, within chromosomal DNA, as opposed to less stable mRNA. This could in theory be achieved by ddPCR and its associated use of limited dilution and partitioning [199]. Any improvements realized in quantification sensitivity and CV could serve to avoid unnecessary testing and to improve patient therapy and overall patient management [247]

Achieving a LOD in a ddPCR assay that is comparable with the current RT-qPCR assays used for CML monitoring would almost certainly require a larger amount of input gDNA (approximately 5 μg) than was used in the two assays described in this thesis. With this larger amount of DNA, a LOD of 10^{-5} is possible if *ca.* 5 – 10 positive droplets can be reliably identified in a gDNA sample partitioned among ~ 1 million droplets. While no attempt to conduct MRD monitoring of CML by the ddPCR concept outlined above has been attempted, some insights into the scale of the problem are provided by Bartley et al., [175], who reported on a RT-qPCR assay against gDNA of CML patients that uses over 600 primer sets to locate the fusion junction, which is generally unknown in these patients. Though informative, that method is clearly not suitable for clinical use due to the complexity of the assay, the very high loads of gDNA required and the high rate of false positives recorded. These limitations can, at least in theory, be overcome using ddPCR. For example, the Bio-Rad QX-200 ddPCR instrument generates *ca*. 15,000 readable droplets per well, so that a filled 96 well plate running a 5 μg gDNA sample at the CPD of 1 could be used to screen for a small population of *BCR-ABL* (amplicon) positive droplets among the *ca.* 1.4 million total read droplets. The challenge would then be to identify ways to properly nest the set of several hundred primer pairs needed to ensure amplification across the fusion and subsequent detection of the fusion-gene amplicons.

It is possible that a larger level of partitioning would be required as a result of this primer multiplexing issue, making the current Bio-Rad instrument unsuitable. However, a ddPCR-based MRD monitoring assay might still be achieved using an alternative digital PCR platform, such as the RainDrop instrument (RainDance Technologies, Inc.), which segregates a gDNA sample into 10 million 5 pL-volume droplets [213].

Finally, the most logical and valuable extension of this thesis work is to validate the platform and the two assays developed from it on clinical samples through both retrospective and prospective double-blind studies. Though they are beyond the fundamental goals of this thesis, those studies should serve to validate the probabilistic data analysis method at the core of the platform, as well as set limits of acceptability on shear, rain, and CPD during a given run. In addition, the platform could be and should be applied to other reciprocal and inversion translocations that have been shown to be prognostic or theranostic of cancer, including specific genotypes of lung cancer, prostate cancer, breast cancer, ALCL, promyelocytic leukemia and thyroid cancer [3, 307]. Efforts could also be made to extend the platform to the analysis of

ctDNA, as this could potentially reduce patient suffering associated with tumor biopsy [308-310] and enable increased testing frequency. Sundaresan et al., [311] have recently shown that NGS can be used to analyze ctDNA for oncogenic somatic mutations, demonstrating the potential for less invasive routes to cancer monitoring.

Finally, *ALK* is translocated with other fusion partners in cancers other than NSCLC. Fourteen different fusion partners have been identified [3] to date and ongoing research may identify more, making *ALK* a potentially important target in many cancers and cancer treatment decisions. Validating the ddPCR *ALK* status assay on those fusions could potentially be a future goal.

References

1. Roukos V, Misteli T. The biogenesis of chromosome translocations. Nature Cell Biology. 2014;16(4):293-300.
2. Nambiar M, Kari V, Raghavan SC. Chromosomal translocations in cancer. Biochimica et Biophysica Acta. 2008;1786(2):139-52.
3. Mitelman F, Johansson B, Mertens F. The Impact of Translocations and Gene Fusions on Cancer Causation. Nature Reviews Cancer. 2007;7(4):233-45.
4. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature. 1973;243(5405):290-3.
5. Puil L, Liu J, Gish G, Mbamalu G, Bowtell D, Pelicci PG, Arlinghaus R, Pawson T. Bcr-Abl oncoproteins bind directly to activators of the Ras signalling pathway. EMBO journal. 1994;13(4):764-73.
6. Nowell PC, Hungerford D. A minute chromosome in human chronic granulocytic leukemia. Science. 1960;142(1497).
7. Horita M, Andreu EJ, Benito A, Arbona C, Sanz C, Benet I, Prosper F, Fernandez-Luna JL. Blockade of the Bcr-Abl kinase activity induces apoptosis of chronic myelogenous leukemia cells by suppressing signal transducer and activator of transcription 5-dependent expression of Bcl-xL. Journal of Experimental Medicine. 2000;191(6):977-84.
8. Lobo I. Chromosome abnormalities and cancer cytogenetics. Nature Education. 2008;1(1):68.
9. Zhang Y, Rowley JD. Chronic myeloid leukemia: current perspectives. Clinics in Laboratory Medicine. 2011;31(4):687-98, .
10. Clarkson B, Strife A, Wisniewski D, Lambek CL, Liu C. Chronic myelogenous leukemia as a paradigm of early cancer and possible curative strategies. Leukemia. 2003;17(7):1211-62.
11. Melo JV, Barnes DJ. Chronic myeloid leukaemia as a model of disease evolution in human cancer. Nature Reviews Cancer. 2007;7(6):441-53.
12. Melo JV, Ross DM. Minimal residual disease and discontinuation of therapy in chronic myeloid leukemia: can we aim at a cure? Hematology American Society of Hematology Education Program. 2011;2011:136-42.
13. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H. Identification of the Transforming EML4-ALK Fusion Gene in Non-Small-Cell Lung Cancer. Nature. 2007;448(7153):561-6.
14. Takeuchi K, Choi YL, Soda M, Inamura K, Togashi Y, Hatano S, Enomoto M, Takada S, Yamashita Y, Satoh Y, Okumura S, Nakagawa K, Ishikawa Y, Mano H. Multiplex Reverse Transcription-PCR Screening for EML4-ALK Fusion Transcripts. Clinical Cancer Research. 2008;14(20):6618-24.
15. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, Ou SH, Dezube BJ, Janne PA, Costa DB, Varella-Garcia M, Kim WH, Lynch TJ, Fidias P, Stubbs H, Engelman JA, Sequist LV, Tan W, Gandhi L, Mino-Kenudson M, Wei GC, Shreeve SM, Ratain MJ, Settleman J, Christensen JG, Haber DA, Wilner K, Salgia R, Shapiro GI, Clark JW, Iafrate AJ. Anaplastic Lymphoma Kinase Inhibition in Non-Small-Cell Lung Cancer. New England Journal of Medicine. 2010;363(18):1693-703.
16. 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. Analytical Chemistry. 2011;83(22):8604-10.
17. Mitelman F, Johansson B, Mertens F. Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. 2015. http://cgap.nci.nih.gov/Chromosomes/Mitelman. Accessed March 2015.
18. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science. 2015;347(6217):78-81.
19. Lieber MR, Ma Y, Pannicke U, Schwarz K. Mechanism and regulation of human non-homologous DNA end-joining. Nature Reviews Molecular Cell Biology. 2003;4(9):712-20.
20. Dudley DD, Chaudhuri J, Bassing CH, Alt FW. Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. Advances in Immunology. 2005;86:43-112.
21. Ensminger M, Iloff L, Ebel C, Nikolova T, Kaina B, Lbrich M. DNA breaks and chromosomal aberrations arise when replication meets base excision repair. The Journal of Cell Biology. 2014;206(1):29-43.
22. Budman J, Chu G. Processing of DNA for nonhomologous end-joining by cell-free extract. EMBO journal. 2005;24(4):849-60.
23. Ross DM, O'Hely M, Bartley PA, Dang P, Score J, Goyne JM, Sobrinho-Simoes M, Cross NC, Melo JV, Speed TP, Hughes TP, Morley AA. Distribution of genomic breakpoints in chronic myeloid leukemia: analysis of 308 patients. Leukemia. 2013;27(10):2105-7.
24. Richardson C, Jasin M. Coupled homologous and nonhomologous repair of a double-strand break preserves genomic integrity in mammalian cells. Molecular and Cellular Biology. 2000;20(23):9068-75.
25. Orthwein A, Fradet-Turcotte A, Noordermeer SM, Canny MD, Brun CM, Strecker J, Escribano-Diaz C, Durocher D. Mitosis inhibits DNA double-strand break repair to guard against telomere fusions. Science. 2014;344(6180):189-93.
26. Downs JA, Nussenzweig MC, Nussenzweig A. Chromatin dynamics and the preservation of genetic information. Nature. 2007;447(7147):951-8.
27. Kruhlak MJ, Celeste A, Dellaire G, Fernandez-Capetillo O, Muller WG, McNally JG, Bazett-Jones DP, Nussenzweig A. Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. The Journal of Cell Biology. 2006;172(6):823-34.
28. Roukos V, Voss TC, Schmidt CK, Lee S, Wangsa D, Misteli T. Spatial dynamics of chromosome translocations in living cells. Science. 2013;341(6146):660-4.
29. Misteli T. Beyond the sequence: cellular organization of genome function. Cell. 2007;128(4):787-800.
30. Roukos V, Burman B, Misteli T. The cellular etiology of chromosome translocations. Current Opinion in Cell Biology. 2013;25(3):357-64.
31. Nikiforova MN, Stringer JR, Blough R, Medvedovic M, Fagin JA, Nikiforov YE. Proximity of chromosomal loci that participate in radiation-induced rearrangements in human cells. Science. 2000;290(5489):138-41.
32. Neves H, Ramos C, da Silva MG, Parreira A, Parreira L. The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. Blood. 1999;93(4):1197-207.

33. Lukasova E, Kozubek S, Kozubek M, Kjeronska J, Ryznar L, Horakova J, Krahulcova E, Horneck G. Localisation and distance between ABL and BCR genes in interphase nuclei of bone marrow cells of control donors and patients with chronic myeloid leukaemia. Human Genetics. 1997;100(5-6):525-35.
34. Rhind N, Gilbert DM. DNA replication timing. Cold Spring Harbor Perspectives in Biology. 2013;5(8):a010132.
35. Ryba T, Battaglia D, Chang BH, Shirley JW, Buckley Q, Pope BD, Devidas M, Druker BJ, Gilbert DM. Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. Genome Research. 2012;22(10):1833-44.
36. De S, Michor F. DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. Nature Biotechnology. 2011;29(12):1103-8.
37. Kawasaki ES, Clark SS, Coyne MY, Smith SD, Champlin R, Witte ON, McCormick FP. Diagnosis of chronic myeloid and acute lymphocytic leukemias by detection of leukemia-specific mRNA sequences amplified in vitro. Proceedings of the National Academy of Sciences. 1988;85(15):5698-702.
38. Martiat P, Ifrah N, Rassool F, Morgan G, Giles F, Gow J, Goldman JM. Molecular analysis of Philadelphia positive essential thrombocythemia. Leukemia. 1989;3(8):563-5.
39. Martiat P, Mecucci C, Nizet Y, Stul M, Philippe M, Cassiman JJ, Michaux JL, Van den Berghe H, Sokal G. P190 BCR/ABL transcript in a case of Philadelphia-positive multiple myeloma. Leukemia. 1990;4(11):751-4.
40. Mitani K, Sato Y, Tojo A, Ishikawa F, Kobayashi Y, Miura Y, Miyazono K, Urabe A, Takaku F. Philadelphia chromosome positive B-cell type malignant lymphoma expressing an aberrant 190 kDa bcr-abl protein. British Journal of Haematology. 1990;76(2):221-5.
41. Moorman AV, Harrison CJ, Buck GA, Richards SM, Secker-Walker LM, Martineau M, Vance GH, Cherry AM, Higgins RR, Fielding AK, Foroni L, Paietta E, Tallman MS, Litzow MR, Wiernik PH, Rowe JM, Goldstone AH, Dewald GW. Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial. Blood. 2007;109(8):3189-97.
42. Jones LK, Saha V. Philadelphia positive acute lymphoblastic leukaemia of childhood. British Journal of Haematology. 2005;130(4):489-500.
43. Melo JV. The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. Blood. 1996;88(7):2375-84.
44. Jamieson CH, Ailles LE, Dylla SJ, Muijtjens M, Jones C, Zehnder JL, Gotlib J, Li K, Manz MG, Keating A, Sawyers CL, Weissman IL. Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. New England Journal of Medicine. 2004;351(7):657-67.
45. Shepherd P, Suffolk R, Halsey J, Allan N. Analysis of molecular breakpoint and m-RNA transcripts in a prospective randomized trial of interferon in chronic myeloid leukaemia: no correlation with clinical features, cytogenetic response, duration of chronic phase, or survival. British Journal of Haematology. 1995;89(3):546-54.
46. Chissoe SL, Bodenteich A, Wang YF, Wang YP, Burian D, Clifton SW, Crabtree J, Freeman A, Iyer K, Jian L, Ma Y, McLaury H-J, Pan H-Q, Sarhan OH, Toth S, Wang Z, Heisterkamp N, Groffen J, Roe BA. Sequence and analysis of the human ABL gene, the BCR gene, and regions involved in the Philadelphia chromosomal translocation. Genomics. 1995;27(1):67-82.

47. Ben-Neriah Y, Daley GQ, Mes-Masson AM, Witte ON, Baltimore D. The chronic myelogenous leukemia-specific P210 protein is the product of the bcr/abl hybrid gene. Science. 1986;233(4760):212-4.
48. Melo JV. BCR-ABL gene variants. Baillieres Clinical Haematology. 1997;10(2):203-22.
49. Saglio G, Guerrasio A, Rosso C, Zaccaria A, Tassinari A, Serra A, Rege-Cambrin G, Mazza U, Gavosto F. New type of Bcr/Abl junction in Philadelphia chromosome-positive chronic myelogenous leukemia. Blood. 1990;76(9):1819-24.
50. Score J, Calasanz MJ, Ottman O, Pane F, Yeh RF, Sobrinho-Simoes MA, Kreil S, Ward D, Hidalgo-Curtis C, Melo JV, Wiemels J, Nadel B, Cross NC, Grand FH. Analysis of genomic breakpoints in p190 and p210 BCR-ABL indicate distinct mechanisms of formation. Leukemia. 2010;24(10):1742-50.
51. Li S, Ilaria RL, Jr., Million RP, Daley GQ, Van Etten RA. The P190, P210, and P230 forms of the BCR/ABL oncogene induce a similar chronic myeloid leukemia-like syndrome in mice but have different lymphoid leukemogenic activity. Journal of Experimental Medicine. 1999;189(9):1399-412.
52. Verstovsek S, Lin H, Kantarjian H, Saglio G, De Micheli D, Pane F, Garcia-Manero G, Intrieri M, Rotoli B, Salvatore F, Guo JQ, Talpaz M, Specchia G, Pizzolo G, Liberati AM, Cortes J, Quackenbush RC, Arlinghaus RB. Neutrophilic-chronic myeloid leukemia: low levels of p230 BCR/ABL mRNA and undetectable BCR/ABL protein may predict an indolent course. Cancer. 2002;94(9):2416-25.
53. Melo JV. The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. Blood. 1996;88:2375-84.
54. Jackson RC, Radivoyevitch T. A pharmacodynamic model of Bcr-Abl signalling in chronic myeloid leukaemia. Cancer Chemotherapy and Pharmacology. 2014;74(4):765-76.
55. Zhao X, Ghaffari S, Lodish H, Malashkevich VN, Kim PS. Structure of the Bcr-Abl oncoprotein oligomerization domain. Nature Structural Biology. 2002;9(2):117-20.
56. Maru Y, Witte ON. The BCR gene encodes a novel serine/threonine kinase activity within a single exon. Cell. 1991;67(3):459-68.
57. Melo JV, Deininger MW. Biology of chronic myelogenous leukemia--signaling pathways of initiation and transformation. Hematoogy/Oncology Clinics of North America. 2004;18(3):545-68, vii-viii.
58. Pendergast AM, Muller AJ, Havlik MH, Maru Y, Witte ON. BCR sequences essential for transformation by the BCR-ABL oncogene bind to the ABL SH2 regulatory domain in a non-phosphotyrosine-dependent manner. Cell. 1991;66(1):161-71.
59. He Y, Wertheim JA, Xu L, Miller JP, Karnell FG, Choi JK, Ren R, Pear WS. The coiled-coil domain and Tyr177 of bcr are required to induce a murine chronic myelogenous leukemia-like disease by bcr/abl. Blood. 2002;99(8):2957-68.
60. Zhang X, Subrahmanyam R, Wong R, Gross AW, Ren R. The NH(2)-terminal coiled-coil domain and tyrosine 177 play important roles in induction of a myeloproliferative disease in mice by Bcr-Abl. Molecular and Cellular Biology. 2001;21(3):840-53.
61. Fournier M, Lacrosse S, Jamar M, Bours V, Herens C. Deletions of the 3'BCR and 5'ABL regions in patients with Philadelphia-positive chronic myeloid leukemia: a one-step process occurring in about 10% of the cases without any evidence of genetic instability in the target cells. Cancer Genetics and Cytogenetics. 2005;160(2):184-7.

62. Gonzalez FA, Anguita E, Mora A, Asenjo S, Lopez I, Polo M, Villegas A. Deletion of BCR region 3' in chronic myelogenous leukemia. Cancer Genetics and Cytogenetics. 2001;130(1):68-74.
63. Melo JV, Gordon D, Cross N, Goldman J. The ABL-BCR fusion gene is expressed in chronic myeloid leukemia. Blood. 1993;81:158-.
64. American Society of Clinical Oncology. http://www.cancer.net. Accessed March 2014.
65. Canadian Cancer Society. http://www.cancer.ca. Accessed June 2014.
66. Chen Y, Li S. Molecular signatures of chronic myeloid leukemia stem cells. Biomarker Research. 2013;1(1):21.
67. Gribble SM, Reid AG, Roberts I, Grace C, Green AR, Nacheva EP. Genomic imbalances in CML blast crisis: 8q24.12-q24.13 segment identified as a common region of over-representation. Genes Chromosomes Cancer. 2003;37(4):346-58.
68. Brazma D, Grace C, Howard J, Melo JV, Holyoke T, Apperley JF, Nacheva EP. Genomic profile of chronic myelogenous leukemia: Imbalances associated with disease progression. Genes Chromosomes Cancer. 2007;46(11):1039-50.
69. Herens C, Tassin F, Lemaire V, Beguin Y, Collard E, Lampertz S, Croisiau C, Lecomte M, De Prijk B, Longree L, Koulischer L. Deletion of the 5'-ABL region: a recurrent anomaly detected by fluorescence in situ hybridization in about 10% of Philadelphia-positive chronic myeloid leukaemia patients. British Journal of Haematology. 2000;110(1):214-6.
70. Mohr B, Bornhauser M, Platzbecker U, Freiberg-Richter J, Naumann R, Prange-Krex G, Mohm J, Kroschinsky F, Ehninger G, Thiede C. Problems with interphase fluorescence in situ hybridization in detecting BCR/ABL-positive cells in some patients using a novel technique with extra signals. Cancer Genetics and Cytogenetics. 2001;127(2):111-7.
71. Cohen N, Rozenfeld-Granot G, Hardan I, Brok-Simoni F, Amariglio N, Rechavi G, Trakhtenbrot L. Subgroup of patients with Philadelphia-positive chronic myelogenous leukemia characterized by a deletion of 9q proximal to ABL gene: expression profiling, resistance to interferon therapy, and poor prognosis. Cancer Genetics and Cytogenetics. 2001;128(2):114-9.
72. Morel F, Ka C, Le Bris MJ, Herry A, Morice P, Bourquard P, Abgrall JF, Berthou C, De Braekeleer M. Deletion of the 5' abl region in Philadelphia chromosome-positive chronic myeloid leukemia. Leukemia. 2003;17(2):473-4.
73. Dewald GW, Wyatt WA, Silver RT. Atypical BCR and ABL D-FISH patterns in chronic myeloid leukemia and their possible role in therapy. Leukemia Lymphoma. 1999;34(5-6):481-91.
74. Sinclair PB, Nacheva EP, Leversha M, Telford N, Chang J, Reid A, Bench A, Champion K, Huntly B, Green AR. Large deletions at the t(9;22) breakpoint are common and may identify a poor-prognosis subgroup of patients with chronic myeloid leukemia. Blood. 2000;95(3):738-43.
75. Huntly BJP, Reid AG, Bench AJ, Campbell LJ, Telford N, Shepherd P, Szer J, Prince HM, Turner P, Grace CN, E.P., Green AR. Deletions of the derivative chromosome 9 occur at the time of the Philadelphia translocation and provide a powerful and independent prognostic indicator in chronic myeloid leukemia. Blood. 2001;98(6):1732-8.
76. Kolomietz E, Al-Maghrabi J, Brennan S, Karaskova J, Minkin S, Lipton J, Squire JA. Primary chromosomal rearrangements of leukemia are frequently accompanied by extensive submicroscopic deletions and may lead to altered prognosis. Blood. 2001;97(11):3581-8.
77. Markovic VD, Bouman D, Bayani J, Al-Maghrabi J, Kamel-Reid S, Squire JA. Lack of BCR/ABL reciprocal fusion in variant Philadelphia chromosome translocations: a use of double fusion signal FISH and spectral karyotyping. Leukemia. 2000;14(6):1157-60.

78. Huntly BJP, Bench AJ, Delabesse E, Reid AG, Li J, Scott MA, Campbell L, Byrne JP, E. Brizard, A. Niedermeiser, D. Nacheva, E. P. Guilhot, F., Deininger M, Green AR. Derivative chromosome 9 deletions in chronic myeloid leukemia: poor prognosis is not associated with loss of ABL-BCR expression, elevated BCR-ABL levels, or karyotypic instability. Blood. 2002;99(12):4547-53.
79. Albano F, Anelli L, Zagaria A, Coccaro N, D'Addabbo P, Liso V, Rocchi M, Specchia G. Genomic Segmental Duplications on the basis of the t(9;22) Rearrangement in Chronic Myeloid Leukemia. Oncogene. 2010;29(17):2509-16.
80. Zagaria A, Anelli L, Albano F, Storlazzi CT, Liso A, Roberti MG, Buquicchio C, Liso V, Rocchi M, Specchia G. A fluorescence in situ hybridization study of complex t(9;22) in two chronic myelocytic leukemia cases with a masked Philadelphia chromosome. Cancer Genetics and Cytogenetics. 2004;150(1):81-5.
81. Albano F, Specchia G, Anelli L, Zagaria A, Storlazzi CT, Buquicchio C, Roberti MG, Liso V, Rocchi M. Genomic deletions on other chromosomes involved in variant t(9;22) chronic myeloid leukemia cases. Genes Chromosomes Cancer. 2003;36(4):353-60.
82. Albano F, Anelli L, Zagaria A, Coccaro N, Casieri P, Rossi AR, Vicari L, Liso V, Rocchi M, Specchia G. Non random distribution of genomic features in breakpoint regions involved in chronic myeloid leukemia cases with variant t(9;22) or additional chromosomal rearrangements. Molecular Cancer. 2010;9:120.
83. Malvestiti F, Agrati C, Chinetti S, Di Meco A, Cirrincione S, Oggionni M, Grimi B, Maggi F, Simoni G, Grati FR. Complex variant of Philadelphia translocation involving chromosomes 9, 12, and 22 in a case with chronic myeloid leukaemia. Case Reports in Genetics. 2014;2014:691630.
84. Marzocchi G, Castagnetti F, Luatti S, Baldazzi C, Stacchini M, Gugliotta G, Amabile M, Specchia G, Sessarego M, Giussani U, Valori L, Discepoli G, Montaldi A, Santoro A, Bonaldi L, Giudici G, Cianciulli AM, Giacobbi F, Palandri F, Pane F, Saglio G, Martinelli G, Baccarani M, Rosti G, Testoni N. Variant Philadelphia translocations: molecular-cytogenetic characterization and prognostic influence on frontline imatinib therapy, a GIMEMA Working Party on CML analysis. Blood. 2011;117(25):6793-800.
85. Richebourg S, Eclache V, Perot C, Portnoi M-F, Van den Akker J, Terré C, Maareck O, Soenen V, Viguié F, Laï J-L. Mechanisms of genesis of variant translocation in chronic myeloid leukemia are not correlated with ABL1 or BCR deletion status or response to imatinib therapy. Cancer Genetics and Cytogenetics. 2008;182(2):95-102.
86. Deininger M, Buchdunger E, Druker BJ. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood. 2005;105(7):2640-53.
87. Sacha T. Imatinib in Chronic Myeloid Leukemia: an Overview. Mediterranean Journal of Hematology and Infectious Diseases. 2014;6(1):e2014007.
88. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science. 2000;289(5486):1938-42.
89. Deininger MW. Milestones and monitoring in patients with CML treated with imatinib. Hematology American Society og Hematol ogy Education Program. 2008:419-26.
90. Hughes T, Deininger M, Hochhaus A, Branford S, Radich J, Kaeda J, Baccarani M, Cortes J, Cross NC, Druker BJ, Gabert J, Grimwade D, Hehlmann R, Kamel-Reid S, Lipton JH, Longtine J, Martinelli G, Saglio G, Soverini S, Stock W, Goldman JM. Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: review and recommendations for

harmonizing current methodology for detecting BCR-ABL transcripts and kinase domain mutations and for expressing results. Blood. 2006;108(1):28-37.
91. Hughes TP, Hochhaus A, Branford S, Muller MC, Kaeda JS, Foroni L, Druker BJ, Guilhot F, Larson RA, O'Brien SG, Rudoltz MS, Mone M, Wehrle E, Modur V, Goldman JM, Radich JP. Long-term prognostic significance of early molecular response to imatinib in newly diagnosed chronic myeloid leukemia: an analysis from the International Randomized Study of Interferon and STI571 (IRIS). Blood. 2010;116(19):3758-65.
92. Druker BJ, Sawyers CL, Kantarjian H, Resta DJ, Reese SF, Ford JM, Capdeville R, Talpaz M. Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. New England Journal of Medicine. 2001;344(14):1038-42.
93. Deininger M, O'Brien SG, Guilhot F, Goldman JM, Hochhaus A, Hughes TP, Radich JP, Hatfield AK, Mone M, Filian J, editors. International randomized study of interferon Vs STI571 (IRIS) 8-year follow up: Sustained survival and low risk for progression or events in patients with newly diagnosed chronic myeloid leukemia in chronic phase (CML-CP) treated with Imatinib. ASH Annual Meeting Abstracts; 2009.
94. Jain P, Kantarjian H, Cortes J. Chronic myeloid leukemia: overview of new agents and comparative analysis. Current Treat ment Options in Oncology. 2013;14(2):127-43.
95. Huntly BJP, Guilhot F, Reid AG, Vassiliou G, Hennig E, Franke C, Byrne J, Brizard A, Niederwieser D, Freeman-Edward J, Cuthbert G, Bown N, Clark RE, Nacheva EP, Green AR, Deininger MW. Imatinib improves but may not fully reverse the poor prognosis of patients with CML with derivative chromosome 9 deletions. Blood. 2003;102(6):2205-12.
96. Baccarani M, Cortes J, Pane F, Niederwieser D, Saglio G, Apperley J, Cervantes F, Deininger M, Gratwohl A, Guilhot F, Hochhaus A, Horowitz M, Hughes T, Kantarjian H, Larson R, Radich J, Simonsson B, Silver RT, Goldman J, Hehlmann R. Chronic myeloid leukemia: an update of concepts and management recommendations of European LeukemiaNet. Journal of Clinical Oncology. 2009;27(35):6041-51.
97. Hochhaus A, Baccarani M, Deininger M, Apperley JF, Lipton JH, Goldberg SL, Corm S, Shah NP, Cervantes F, Silver RT, Niederwieser D, Stone RM, Dombret H, Larson RA, Roy L, Hughes T, Muller MC, Ezzeddine R, Countouriotis AM, Kantarjian HM. Dasatinib induces durable cytogenetic responses in patients with chronic myelogenous leukemia in chronic phase with resistance or intolerance to imatinib. Leukemia. 2008;22(6):1200-6.
98. Cilloni D, Saglio G. Molecular pathways: BCR-ABL. Clinical Cancer Research. 2012;18(4):930-7.
99. Saglio G, Kim DW, Issaragrisil S, le Coutre P, Etienne G, Lobo C, Pasquini R, Clark RE, Hochhaus A, Hughes TP, Gallagher N, Hoenekopp A, Dong M, Haque A, Larson RA, Kantarjian HM. Nilotinib versus imatinib for newly diagnosed chronic myeloid leukemia. New England Journal of Medicine. 2010;362(24):2251-9.
100. Castagnetti F, Palandri F, Amabile M, Testoni N, Luatti S, Soverini S, Iacobucci I, Breccia M, Rege Cambrin G, Stagno F, Specchia G, Galieni P, Iuliano F, Pane F, Saglio G, Alimena G, Martinelli G, Baccarani M, Rosti G. Results of high-dose imatinib mesylate in intermediate Sokal risk chronic myeloid leukemia patients in early chronic phase: a phase 2 trial of the GIMEMA CML Working Party. Blood. 2009;113(15):3428-34.
101. Maru Y. Molecular biology of chronic myeloid leukemia. Cancer Science. 2012;103(9):1601-10.

102. O'Hare T, Shakespeare WC, Zhu X, Eide CA, Rivera VM, Wang F, Adrian LT, Zhou T, Huang WS, Xu Q, Metcalf CA, 3rd, Tyner JW, Loriaux MM, Corbin AS, Wardwell S, Ning Y, Keats JA, Wang Y, Sundaramoorthi R, Thomas M, Zhou D, Snodgrass J, Commodore L, Sawyer TK, Dalgarno DC, Deininger MW, Druker BJ, Clackson T. AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. Cancer Cell. 2009;16(5):401-12.
103. Cortes J, Lipton JH, Rea D, Digumarti R, Chuah C, Nanda N, Benichou AC, Craig AR, Michallet M, Nicolini FE, Kantarjian H. Phase 2 study of subcutaneous omacetaxine mepesuccinate after TKI failure in patients with chronic-phase CML with T315I mutation. Blood. 2012;120(13):2573-80.
104. Chen Y, Hu Y, Michaels S, Segal D, Brown D, Li S. Inhibitory effects of omacetaxine on leukemic stem cells and BCR-ABL-induced chronic myeloid leukemia and acute lymphoblastic leukemia in mice. Leukemia. 2009;23(8):1446-54.
105. Bueno-da-Silva A, Brumatti G, Russo F, Green D, Amarante-Mendes G. Bcr-Abl-mediated resistance to apoptosis is independent of constant tyrosine-kinase activity. Cell Death and Differentiation. 2003;10(5):592-8.
106. Holyoake T, Jiang X, Eaves C, Eaves A. Isolation of a highly quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia. Blood. 1999;94(6):2056-64.
107. Dierks C, Beigi R, Guo GR, Zirlik K, Stegert MR, Manley P, Trussell C, Schmitt-Graeff A, Landwerlin K, Veelken H, Warmuth M. Expansion of Bcr-Abl-positive leukemic stem cells is dependent on Hedgehog pathway activation. Cancer Cell. 2008;14(3):238-49.
108. Yang L, Xie G, Fan Q, Xie J. Activation of the hedgehog-signaling pathway in human cancer and the clinical implications. Oncogene. 2010;29(4):469-81.
109. Chen JK, Taipale J, Cooper MK, Beachy PA. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. Genes and Development. 2002;16(21):2743-8.
110. Zhao C, Chen A, Jamieson CH, Fereshteh M, Abrahamsson A, Blum J, Kwon HY, Kim J, Chute JP, Rizzieri D, Munchhof M, VanArsdale T, Beachy PA, Reya T. Hedgehog signalling is essential for maintenance of cancer stem cells in myeloid leukaemia. Nature. 2009;458(7239):776-9.
111. Jiang BH, Liu LZ. PI3K/PTEN signaling in angiogenesis and tumorigenesis. Advances in Cancer Research. 2009;102:19-65.
112. Geoerger B, Kerr K, Tang CB, Fung KM, Powell B, Sutton LN, Phillips PC, Janss AJ. Antitumor activity of the rapamycin analog CCI-779 in human primitive neuroectodermal tumor/medulloblastoma models as single agent and in combination chemotherapy. Cancer Research. 2001;61(4):1527-32.
113. Carvajal RD, Tse A, Schwartz GK. Aurora kinases: new targets for cancer therapy. Clinical Cancer Research. 2006;12(23):6869-75.
114. Harrington EA, Bebbington D, Moore J, Rasmussen RK, Ajose-Adeogun AO, Nakayama T, Graham JA, Demur C, Hercend T, Diu-Hercend A, Su M, Golec JM, Miller KM. VX-680, a potent and selective small-molecule inhibitor of the Aurora kinases, suppresses tumor growth in vivo. Nature Medicine. 2004;10(3):262-7.
115. Hayette S, Chabane K, Michallet M, Michallat E, Cony-Makhoul P, Salesse S, Maguer-Satta V, Magaud JP, Nicolini FE. Longitudinal studies of SRC family kinases in imatinib- and dasatinib-resistant chronic myelogenous leukemia patients. Leukemia Research. 2011;35(1):38-43.

116. Neviani P, Santhanam R, Trotta R, Notari M, Blaser BW, Liu S, Mao H, Chang JS, Galietta A, Uttam A, Roy DC, Valtieri M, Bruner-Klisovic R, Caligiuri MA, Bloomfield CD, Marcucci G, Perrotti D. The tumor suppressor PP2A is functionally inactivated in blast crisis CML through the inhibitory activity of the BCR/ABL-regulated SET protein. Cancer Cell. 2005;8(5):355-68.
117. Duy C, Hurtz C, Shojaee S, Cerchietti L, Geng H, Swaminathan S, Klemm L, Kweon SM, Nahar R, Braig M, Park E, Kim YM, Hofmann WK, Herzog S, Jumaa H, Koeffler HP, Yu JJ, Heisterkamp N, Graeber TG, Wu H, Ye BH, Melnick A, Muschen M. BCL6 enables Ph+ acute lymphoblastic leukaemia cells to survive BCR-ABL1 kinase inhibition. Nature. 2011;473(7347):384-8.
118. Ito K, Bernardi R, Morotti A, Matsuoka S, Saglio G, Ikeda Y, Rosenblatt J, Avigan DE, Teruya-Feldstein J, Pandolfi PP. PML targeting eradicates quiescent leukaemia-initiating cells. Nature. 2008;453(7198):1072-8.
119. Peng C, Chen Y, Yang Z, Zhang H, Osterby L, Rosmarin AG, Li S. PTEN is a tumor suppressor in CML stem cells and BCR-ABL-induced leukemias in mice. Blood. 2010;115(3):626-35.
120. Reya T, Duncan AW, Ailles L, Domen J, Scherer DC, Willert K, Hintz L, Nusse R, Weissman IL. A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature. 2003;423(6938):409-14.
121. Mahon FX, Rea D, Guilhot J, Guilhot F, Huguet F, Nicolini F, Legros L, Charbonnier A, Guerci A, Varet B, Etienne G, Reiffers J, Rousselot P. Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre Stop Imatinib (STIM) trial. Lancet Oncology. 2010;11(11):1029-35.
122. Graham SM, Jorgensen HG, Allan E, Pearson C, Alcorn MJ, Richmond L, Holyoake TL. Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to STI571 in vitro. Blood. 2002;99(1):319-25.
123. Rousselot P, Huguet F, Rea D, Legros L, Cayuela JM, Maarek O, Blanchet O, Marit G, Gluckman E, Reiffers J, Gardembas M, Mahon FX. Imatinib mesylate discontinuation in patients with chronic myelogenous leukemia in complete molecular remission for more than 2 years. Blood. 2007;109(1):58-60.
124. World Health Organization: Estimated Cancer Incidence, Mortality and Prevalence in Worldwide in 2012. In: International Agency for Research on Cancer. 2012. http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx?cancer=lung. Accessed January 2015.
125. Ettinger DS, Bepler G, Bueno R, Chang A, Chang JY, Chirieac LR, D'Amico TA, Demmy TL, Feigenberg SJ, Grannis FW, Jr., Jahan T, Jahanzeb M, Kessinger A, Komaki R, Kris MG, Langer CJ, Le QT, Martins R, Otterson GA, Robert F, Sugarbaker DJ, Wood DE. Non-Small Cell Lung Cancer Clinical Practice Guidelines in Oncology. Journal of National Comprehensive Cancer Network. 2006;4(6):548-82.
126. Herbst RS, Heymach JV, Lippman SM. Lung cancer. New England Journal of Medicine. 2008;359(13):1367-80.
127. Sun S, Schiller JH, Gazdar AF. Lung Cancer in Never Smokers--A Different Disease. Nature Reviews Cancer. 2007;7(10):778-90.
128. Koivunen JP, Mermel C, Zejnullahu K, Murphy C, Lifshits E, Holmes AJ, Choi HG, Kim J, Chiang D, Thomas R, Lee J, Richards WG, Sugarbaker DJ, Ducko C, Lindeman N, Marcoux JP, Engelman JA, Gray NS, Lee C, Meyerson M, Janne PA. EML4-ALK Fusion Gene and Efficacy of an ALK Kinase Inhibitor in Lung Cancer. Clinical Cancer Research. 2008;14(13):4275-83.

129. Perner S, Wagner PL, Demichelis F, Mehra R, Lafargue CJ, Moss BJ, Arbogast S, Soltermann A, Weder W, Giordano TJ, Beer DG, Rickman DS, Chinnaiyan AM, Moch H, Rubin MA. EML4-ALK Fusion Lung Cancer: A rare acquired event. Neoplasia. 2008;10(3):298-302.
130. Inamura K, Takeuchi K, Togashi Y, Hatano S, Ninomiya H, Motoi N, Mun MY, Sakao Y, Okumura S, Nakagawa K, Soda M, Choi YL, Mano H, Ishikawa Y. EML4-ALK Lung Cancers are Characterized by Rare other Mutations, a TTF-1 Cell Lineage, an Acinar Histology, and Young Onset. Modern Pathology. 2009;22(4):508-15.
131. Shinmura K, Kageyama S, Tao H, Bunai T, Suzuki M, Kamo T, Takamochi K, Suzuki K, Tanahashi M, Niwa H, Ogawa H, Sugimura H. EML4-ALK Fusion Transcripts, but no NPM-, TPM3-, CLTC-, ATIC-, or TFG-ALK Fusion Transcripts, in Non-Small Cell Lung Carcinomas. Lung Cancer. 2008;61(2):163-9.
132. Shaw AT, Yeap BY, Mino-Kenudson M, Digumarthy SR, Costa DB, Heist RS, Solomon B, Stubbs H, Admane S, McDermott U, Settleman J, Kobayashi S, Mark EJ, Rodig SJ, Chirieac LR, Kwak EL, Lynch TJ, Iafrate AJ. Clinical features and outcome of patients with non-small-cell lung cancer who harbor EML4-ALK. Journal of Clinical Oncology. 2009;27(26):4247-53.
133. Rodig SJ, Mino-Kenudson M, Dacic S, Yeap BY, Shaw A, Barletta JA, Stubbs H, Law K, Lindeman N, Mark E, Janne PA, Lynch T, Johnson BE, Iafrate AJ, Chirieac LR. Unique clinicopathologic features characterize ALK-rearranged lung adenocarcinoma in the western population. Clinical Cancer Research. 2009;15(16):5216-23.
134. Husain H, Rudin CM. ALK-targeted therapy for lung cancer: ready for prime time. Oncology 2011;25(7):597-601.
135. Iwahara T, Fujimoto J, Wen D, Cupples R, Bucay N, Arakawa T, Mori S, Ratzkin B, Yamamoto T. Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. Oncogene. 1997;14(4):439-49.
136. Takeuchi K, Choi YL, Togashi Y, Soda M, Hatano S, Inamura K, Takada S, Ueno T, Yamashita Y, Satoh Y, Okumura S, Nakagawa K, Ishikawa Y, Mano H. KIF5B-ALK, a Novel Fusion Oncokinase Identified by an Immunohistochemistry-Based Diagnostic System for ALK-positive Lung Cancer. Clinical Cancer Research. 2009;15(9):3143-9.
137. Lin E, Li L, Guan Y, Soriano R, Rivers CS, Mohan S, Pandita A, Tang J, Modrusan Z. Exon Array Profiling Detects EML4-ALK Fusion in Breast, Colorectal, and Non-Small Cell Lung Cancers. Molecular Cancer Research. 2009;7(9):1466-76.
138. Soda M, Takada S, Takeuchi K, Choi YL, Enomoto M, Ueno T, Haruta H, Hamada T, Yamashita Y, Ishikawa Y, Sugiyama Y, Mano H. A mouse model for EML4-ALK-positive lung cancer. Proceedings of the National Academy of Sciences. 2008;105(50):19893-7.
139. Aggarwal C. Targeted Therapy for Lung Cancer: Present and Future. Annals of Palliative Medicine. 2014;3(3):229-35.
140. McDermott U, Iafrate AJ, Gray NS, Shioda T, Classon M, Maheswaran S, Zhou W, Choi HG, Smith SL, Dowell L, Ulkus LE, Kuhlmann G, Greninger P, Christensen JG, Haber DA, Settleman J. Genomic Alterations of Anaplastic Lymphoma Kinase may Sensitize Tumors to Anaplastic Lymphoma Kinase Inhibitors. Cancer Research. 2008;68(9):3389-95.
141. Shaw AT, Yeap BY, Solomon BJ, Riely GJ, Gainor J, Engelman JA, Shapiro GI, Costa DB, Ou SH, Butaney M, Salgia R, Maki RG, Varella-Garcia M, Doebele RC, Bang YJ, Kulig K, Selaru P, Tang Y, Wilner KD, Kwak EL, Clark JW, Iafrate AJ, Camidge DR. Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: a retrospective analysis. Lancet Oncology. 2011;12(11):1004-12.

142. Saltman D. Regulatory Approval and Funding of Precision Medicine in Non-Small Cell Lung Cancer. Precision Medicine. 2013:16.
143. Schiller JH, Harrington D, Belani CP, Langer C, Sandler A, Krook J, Zhu J, Johnson DH. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. New England Journal of Medicine. 2002;346(2):92-8.
144. Choi YL, Soda M, Yamashita Y, Ueno T, Takashima J, Nakajima T, Yatabe Y, Takeuchi K, Hamada T, Haruta H, Ishikawa Y, Kimura H, Mitsudomi T, Tanio Y, Mano H. EML4-ALK mutations in lung cancer that confer resistance to ALK inhibitors. New England Journal of Medicine. 2010;363(18):1734-9.
145. Katayama R, Shaw AT, Khan TM, Mino-Kenudson M, Solomon BJ, Halmos B, Jessop NA, Wain JC, Yeo AT, Benes C, Drew L, Saeh JC, Crosby K, Sequist LV, Iafrate AJ, Engelman JA. Mechanisms of acquired crizotinib resistance in ALK-rearranged lung Cancers. Science Translational Medicine. 2012;4(120):120ra17.
146. Doebele RC, Pilling AB, Aisner DL, Kutateladze TG, Le AT, Weickhardt AJ, Kondo KL, Linderman DJ, Heasley LE, Franklin WA, Varella-Garcia M, Camidge DR. Mechanisms of resistance to crizotinib in patients with ALK gene rearranged non-small cell lung cancer. Clinical Cancer Research. 2012;18(5):1472-82.
147. Gerber DE, Minna JD. ALK Inhibition for Non-Small Cell Lung Cancer: from Discovery to Therapy in record time. Cancer Cell. 2010;18(6):548-51.
148. Iacono D, Chiari R, Metro G, Bennati C, Bellezza G, Cenci M, Ricciuti B, Sidoni A, Baglivo S, Minotti V, Crino L. Future options for ALK-positive non-small cell lung cancer. Lung Cancer. 2015.
149. Gainor JF, Tan DS, De Pas T, Solomon B, Ahmad A, Lazzari C, De Marinis F, Spitaleri G, Schultz K, Friboulet L, Yeap BY, Engelman JA, Shaw AT. Progression-Free and Overall Survival in ALK-Positive NSCLC Patients Treated with Sequential Crizotinib and Ceritinib. Clinical Cancer Research. 2015.
150. Seto T, Kiura K, Nishio M, Nakagawa K, Maemondo M, Inoue A, Hida T, Yamamoto N, Yoshioka H, Harada M, Ohe Y, Nogami N, Takeuchi K, Shimada T, Tanaka T, Tamura T. CH5424802 (RO5424802) for patients with ALK-rearranged advanced non-small-cell lung cancer (AF-001JP study): a single-arm, open-label, phase 1-2 study. Lancet Oncology. 2013;14(7):590-8.
151. Khozin S, Blumenthal GM, Zhang L, Tang S, Brower M, Fox E, Helms W, Leong R, Song P, Pan Y, Liu Q, Zhao P, Zhao H, Lu D, Tang Z, Al Hakim A, Boyd K, Keegan P, Justice R, Pazdur R. FDA Approval: Ceritinib for the Treatment of Metastatic Anaplastic Lymphoma Kinase-Positive Non-Small Cell Lung Cancer. Clinical Cancer Research. 2015.
152. R. M, Camidge DR, Sharma S, Felip E, Tan DS, Vansteenkiste JF, De Pas TM, Kim D-W, Santoro A, Liu G, Goldwasser M, Dai D, Radona M, Boral AL, Shaw AT. First-in-human phase I study of the ALK inhibitor LDK378 in advanced solid tumors. Journal of Clinical Oncology. 2012;30 Supplimentary Abstract 3007.
153. Friboulet L, Li N, Katayama R, Lee CC, Gainor JF, Crystal AS, Michellys PY, Awad MM, Yanagitani N, Kim S, Pferdekamper AC, Li J, Kasibhatla S, Sun F, Sun X, Hua S, McNamara P, Mahmood S, Lockerman EL, Fujita N, Nishio M, Harris JL, Shaw AT, Engelman JA. The ALK inhibitor ceritinib overcomes crizotinib resistance in non-small cell lung cancer. Cancer Discovery. 2014;4(6):662-73.
154. Ou SH, Greenbowe J, Khan ZU, Azada MC, Ross JS, Stevens PJ, Ali SM, Miller VA, Gitlitz B. I1171 missense mutation (particularly I1171N) is a common resistance mutation in

ALK-positive NSCLC patients who have progressive disease while on alectinib and is sensitive to ceritinib. Lung Cancer. 2015;88(2):231-4.
155. Normant E, Paez G, West KA, Lim AR, Slocum KL, Tunkey C, McDougall J, Wylie AA, Robison K, Caliri K, Palombella VJ, Fritz CC. The Hsp90 inhibitor IPI-504 rapidly lowers EML4-ALK levels and induces tumor regression in ALK-driven NSCLC models. Oncogene. 2011;30(22):2581-6.
156. Lovly CM, Shaw AT. Molecular pathways: resistance to kinase inhibitors and implications for therapeutic strategies. Clinical Cancer Research. 2014;20(9):2249-56.
157. Sang J, Acquaviva J, Friedland JC, Smith DL, Sequeira M, Zhang C, Jiang Q, Xue L, Lovly CM, Jimenez JP, Shaw AT, Doebele RC, He S, Bates RC, Camidge DR, Morris SW, El-Hariry I, Proia DA. Targeted inhibition of the molecular chaperone Hsp90 overcomes ALK inhibitor resistance in non-small cell lung cancer. Cancer Discovery. 2013;3(4):430-43.
158. Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, Nardone J, Lee K, Reeves C, Li Y, Hu Y, Tan Z, Stokes M, Sullivan L, Mitchell J, Wetzel R, Macneill J, Ren JM, Yuan J, Bakalarski CE, Villen J, Kornhauser JM, Smith B, Li D, Zhou X, Gygi SP, Gu TL, Polakiewicz RD, Rush J, Comb MJ. Global Survey of Phosphotyrosine Signaling Identifies Oncogenic Kinases in Lung Cancer. Cell. 2007;131(6):1190-203.
159. Wang Y, Wang S, Xu S, Qu J, Liu B. Clinicopathologic features of patients with non-small cell lung cancer harboring the EML4-ALK fusion gene: a meta-analysis. PLoS One. 2014;9(10):e110617.
160. Zhang X, Zhang S, Yang X, Yang J, Zhou Q, Yin L, An S, Lin J, Chen S, Xie Z, Zhu M, Zhang X, Wu YL. Fusion of EML4 and ALK is associated with development of lung adenocarcinomas lacking EGFR and KRAS mutations and is correlated with ALK expression. Molecular Cancer. 2010;9:188.
161. Morse HG, Humbert JR, Hutter JJ, Robinson A. Karyotyping of bone-marrow cells in hematologic diseases. Human Genetics. 1977;37(1):33-9.
162. National Comprehensive Cancer Network - Chronic Myelogenous Leukemia. 2014. (http://www.nccn.org/professionals/physician_gls/pdf/cml.pdf) Accessed 2015 February.
163. Tkachuk DC, Westbrook CA, Andreeff M, Donlon TA, Cleary ML, Suryanarayan K, Homge M, Redner A, Gray J, Pinkel D. Detection of bcr-abl fusion in chronic myelogeneous leukemia by in situ hybridization. Science. 1990;250(4980):559-62.
164. Chase A, Grand F, Zhang JG, Blackett N, Goldman J, Gordon M. Factors influencing the false positive and negative rates of BCR-ABL fluorescence in situ hybridization. Genes Chromosomes Cancer. 1997;18(4):246-53.
165. Sinclair PB, Green AR, Grace C, Nacheva EP. Improved sensitivity of BCR-ABL detection: a triple-probe three-color fluorescence in situ hybridization system. Blood. 1997;90(4):1395-402.
166. Siu LL, Ma ES, Wong WS, Chan MH, Wong KF. Application of tri-colour, dual fusion fluorescence in situ hybridization (FISH) system for the characterization of BCR-ABL1 fusion in chronic myelogenous leukaemia (CML) and residual disease monitoring. BMC Blood Disorders. 2009;9:4.
167. Das K, Tan P. Molecular cytogenetics: recent developments and applications in cancer. Clinical Genetics. 2013;84(4):315-25.
168. Wan TS. Cancer cytogenetics: methodology revisited. Annals of Laboratory Medicine. 2014;34(6):413-25.
169. Bishop R. Applications of fluorescence in situ hybridization (FISH) in detecting genetic aberrations of medical significance. Bioscience Horizons. 2010;3(1):85-95.

170. Savola S, Nardi F, Scotlandi K, Picci P, Knuutila S. Microdeletions in 9p21.3 induce false negative results in CDKN2A FISH analysis of Ewing sarcoma. Cytogenet Genome Res. 2007;119(1-2):21-6.
171. Baccarani M, Saglio G, Goldman J, Hochhaus A, Simonsson B, Appelbaum F, Apperley J, Cervantes F, Cortes J, Deininger M, Gratwohl A, Guilhot F, Horowitz M, Hughes T, Kantarjian H, Larson R, Niederwieser D, Silver R, Hehlmann R. Evolving concepts in the management of chronic myeloid leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. Blood. 2006;108(6):1809-20.
172. Hochhaus A, Lin F, Reiter A, Skladny H, Mason PJ, van Rhee F, Shepherd PC, Allan NC, Hehlmann R, Goldman JM, Cross NC. Quantification of residual disease in chronic myelogenous leukemia patients on interferon-alpha therapy by competitive polymerase chain reaction. Blood. 1996;87(4):1549-55.
173. Minimal Residual Disease (MRD-A) Survey. Northfield, IL: College of American Pathologists (CAP). 2012.
174. Nutz S, Doll K, Karlovsky P. Determination of the LOQ in real-time PCR by receiver operating characteristic curve analysis: application to qPCR assays for Fusarium verticillioides and F. proliferatum. Analytical and Bioanalytical Chemistry
. 2011;401(2):717-26.
175. Bartley PA, Latham S, Budgen B, Ross DM, Hughes E, Branford S, White D, Hughes TP, Morley AA. A DNA Real-time Quantitative PCR Method Suitable for Routine Monitoring of Low Levels of Minimal Residual Disease in Chronic Myeloid Leukemia. The Journal of Molecular Diagnostics. 2014.
176. Bartley PA, Ross DM, Latham S, Martin-Harris MH, Budgen B, Wilczek V, Branford S, Hughes TP, Morley AA. Sensitive detection and quantification of minimal residual disease in chronic myeloid leukaemia using nested quantitative PCR for BCR-ABL DNA. International Journal of Laboratory Hematology. 2010;32(6 Pt 1):e222-8.
177. Shibata Y, Malhotra A, Dutta A. Detection of DNA fusion junctions for BCR-ABL translocations by Anchored ChromPET. Genome Medicine. 2010;2(9):70.
178. Branford S, Cross NC, Hochhaus A, Radich J, Saglio G, Kaeda J, Goldman J, Hughes T. Rationale for the recommendations for harmonizing current methodology for detecting BCR-ABL transcripts in patients with chronic myeloid leukaemia. Leukemia. 2006;20(11):1925-30.
179. Robesova B, Bajerova M, Liskova K, Skrickova J, Tomiskova M, Pospisilova S, Mayer J, Dvorakova D. TaqMan based Real Time PCR Assay Targeting EML4-ALK Fusion Transcripts in NSCLC. Lung Cancer. 2014;85(1):25-30.
180. Sanders HR, Li HR, Bruey JM, Scheerle JA, Meloni-Ehrig AM, Kelly JC, Novick C, Albitar M. Exon Scanning by Reverse Transcriptase-Polymerase Chain Reaction for Detection of Known and Novel EML4-ALK Fusion Variants in Non-Small Cell Lung Cancer. Cancer Genetics. 2011;204(1):45-52.
181. Blackhall FH, Peters S, Bubendorf L, Dafni U, Kerr KM, Hager H, Soltermann A, O'Byrne KJ, Dooms C, Sejda A, Hernandez-Losa J, Marchetti A, Savic S, Tan Q, Thunnissen E, Speel EJ, Cheney R, Nonaka D, de Jong J, Martorell M, Letovanec I, Rosell R, Stahel RA. Prevalence and Clinical Outcomes for Patients with ALK-positive resected stage I to III Adenocarcinoma: results from the European Thoracic Oncology Platform Lungscape Project. Journal of Clinical Oncology. 2014;32(25):2780-7.
182. Centre for Medicare services. Comprehensive Genomic Profiling for Non-Small Cell Lung Cancer 2014. http://www.cms.gov/medicare-coverage-database/details/lcd-

details.aspx?LCDId=35895&ContrId=229&ver=7&ContrVer=1&name=229*1&bc=AQAAAgAAAAAAAA%3d%3d&. Accessed July 2014.
183. Camidge DR, Kono SA, Flacco A, Tan AC, Doebele RC, Zhou Q, Crino L, Franklin WA, Varella-Garcia M. Optimizing the Detection of Lung Cancer Patients Harboring Anaplastic Lymphoma Kinase (ALK) Gene Rearrangements Potentially Suitable for ALK Inhibitor Treatment. Clinical Cancer Research. 2010;16(22):5581-90.
184. Paik JH, Choe G, Kim H, Choe J, Lee H, Lee C, Lee J, Jheon S, Chung J. Screening of Anaplastic Lymphoma Kinase Rearrangement by Immunohistochemistry in Non-Small Cell Lung Cancer: Correlation with Fluorescence in situ Hybridization. Journal of Thoracic Oncology. 2011;6(3):466-72.
185. Mano H. Non-solid oncogenes in solid tumors: EML4-ALK fusion genes in lung cancer. Cancer Science. 2008;99(12):2349-55.
186. Gonzalez-Angulo AM, Hennessy BT, Mills GB. Future of personalized medicine in oncology: a systems biology approach. Journal of Clinical Oncology. 2010;28(16):2777-83.
187. Brisco MJ, Latham S, Bartley PA, Morley AA. Incorporation of measurement of DNA integrity into qPCR assays. Biotechniques. 2010;49(6):893-7.
188. Morisset D, Štebih D, Milavec M, Gruden K, Žel J. Quantitative analysis of food and feed samples with droplet digital PCR. 2013.
189. Rački N, Morisset D, Gutierrez-Aguirre I, Ravnikar M. One-step RT-droplet digital PCR: a breakthrough in the quantification of waterborne RNA viruses. Analytical and Bioanalytical Chemistry. 2014;406(3):661-7.
190. Dube S, Qin J, Ramakrishnan R. Mathematical Analysis of Copy Number Variation in a DNA Sample using Digital PCR on a Nanofluidic Device. PLoS One. 2008;3(8):e2876.
191. Whale AS, Cowen S, Foy CA, Huggett JF. Methods for applying accurate digital PCR analysis on low copy DNA samples. PLoS One. 2013;8(3):e58177.
192. 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. Diagnostic Molecular Pathology. 2012;21(1):1-8.
193. Vogelstein B, Kinzler KW. Digital PCR. Proceedings of the National Academy of Sciences. 1999;96(16):9236-41.
194. Morley AA. Digital PCR: a brief history. Biomolecular Detection and Quantification. 2014;1(1):1-2.
195. Whale AS, Huggett JF, Cowen S, Speirs V, Shaw J, Ellison S, Foy CA, Scott DJ. Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. Nucleic Acids Research. 2012;40(11):e82.
196. Hoshino T, Inagaki F. Molecular quantification of environmental DNA using microfluidics and digital PCR. Systematic and Applied Microbiology. 2012;35(6):390-5.
197. 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. Nature Methods. 2013;10(10):1003-5.
198. Heyries KA, Tropini C, VanInsberghe M, Doolin C, Petriv I, Singhal A, Leung K, Hughesman CB, Hansen CL. Megapixel digital PCR. Nature methods. 2011;8(8):649-51.
199. Ross DM, Branford S. Minimal residual disease: the advantages of digital over analog polymerase chain reaction. Leukemia Lymphoma. 2011;52(7):1161-3.

200. Last A, Molina-Gonzalez S, Cassama E, Butcher R, Nabicassa M, McCarthy E, Burr SE, Mabey DC, Bailey RL, Holland MJ. Development and evaluation of a next-generation digital PCR diagnostic assay for ocular Chlamydia trachomatis infections. Journal of Clinical Microbiology. 2013;51(7):2195-203.
201. Dingle TC, Sedlak RH, Cook L, Jerome KR. Tolerance of droplet-digital PCR vs real-time quantitative PCR to inhibitory substances. Clinical Chemistry. 2013;59(11):1670-2.
202. Huggett JF, Foy CA, Benes V, Emslie K, Garson JA, Haynes R, Hellemans J, Kubista M, Mueller RD, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT, Bustin SA. The digital MIQE guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments. Clinical Chemistry. 2013;59(6):892-902.
203. Pohl G, Shih Ie M. Principle and applications of digital PCR. Expert Review of Molecular Diagnostics. 2004;4(1):41-7.
204. Sanders R, Huggett JF, Bushell CA, Cowen S, Scott DJ, Foy CA. Evaluation of digital PCR for absolute DNA quantification. Analytical Chemistry. 2011;83(17):6474-84.
205. Day E, Dear PH, McCaughan F. Digital PCR strategies in the development and analysis of molecular biomarkers for personalized medicine. Methods. 2013;59(1):101-7.
206. Sutton K, Witham J, Stead L, Taylor G, Quirke P. Next Generation Sequencing: A highly Sensitive Test for Mutation Detection. In: Leeds Institute of Molecular Medicine UK, editor. Horizon Diagnostics. 2013.
207. Huggett JF, Cowen S, Foy CA. Considerations for Digital PCR as an Accurate Molecular Diagnostic Tool. Clinical Chemistry. 2015;61(1):79-88.
208. Ståhlberg A, Håkansson J, Xian X, Semb H, Kubista M. Properties of the reverse transcription reaction in mRNA quantification. Clinical Chemistry. 2004;50(3):509-15.
209. Strain MC, Lada SM, Luong T, Rought SE, Gianella S, Terry VH, Spina CA, Woelk CH, Richman DD. Highly precise measurement of HIV DNA by droplet digital PCR. PLoS One. 2013;8(4):e55943.
210. Cochran RL, Cravero K, Chu D, Erlanger B, Toro PV, Beaver JA, Zabransky DJ, Wong HY, Cidado J, Croessmann S. Analysis of BRCA2 loss of heterozygosity in tumor tissue using droplet digital polymerase chain reaction. Human Pathology. 2014;45(7):1546-50.
211. Heredia NJ, Belgrader P, Wang S, Koehler R, Regan J, Cosman AM, Saxonov S, Hindson B, Tanner SC, Brown AS. Droplet Digital™ PCR quantitation of HER2 expression in FFPE breast cancer samples. Methods. 2013;59(1):S20-S3.
212. Milbury CA, Zhong Q, Lin J, Williams M, Olson J, Link DR, Hutchison B. Determining lower limits of detection of digital PCR assays for cancer-related gene mutations. Biomolecular Detection and Quantification. 2014;1(1):8-22.
213. Baker M. Digital PCR hits its stride. Nature Methods. 2012;9(6):541-4.
214. Pekin D, Skhiri Y, Baret J-C, Le Corre D, Mazutis L, Salem CB, Millot F, El Harrak A, Hutchison JB, Larson JW. Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. Lab on a Chip. 2011;11(13):2156-66.
215. Wang J, Ramakrishnan R, Tang Z, Fan W, Kluge A, Dowlati A, Jones RC, Ma PC. Quantifying EGFR alterations in the lung cancer genome with nanofluidic digital PCR arrays. Clinical Chemistry. 2010;56(4):623-32.
216. Roberts CH, Jiang W, Jayaraman J, Trowsdale J, Holland MJ, Traherne JA. Killer-cell Immunoglobulin-like Receptor gene linkage and copy number variation analysis by droplet digital PCR. Genome Medicine. 2014;6(3):20.

217. Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S, Emslie KR. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. Analytical Chemistry. 2012;84(2):1003-11.
218. Bhat S, Herrmann J, Armishaw P, Corbisier P, Emslie KR. Single molecule detection in nanofluidic digital array enables accurate measurement of DNA copy number. Analytical and Bioanalytical Chemistry. 2009;394(2):457-67.
219. Dawson S-J, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin S-F, Dunning MJ, Gale D, Forshew T, Mahler-Araujo B. Analysis of circulating tumor DNA to monitor metastatic breast cancer. New England Journal of Medicine. 2013;368(13):1199-209.
220. Roberts CH, Last A, Molina-Gonzalez S, Cassama E, Butcher R, Nabicassa M, McCarthy E, Burr SE, Mabey DC, Bailey RL, Holland MJ. Development and evaluation of a next-generation digital PCR diagnostic assay for ocular Chlamydia trachomatis infections. Journal of Clinical Microbiology. 2013;51(7):2195-203.
221. Lun FM, Chiu RW, Chan KC, Leung TY, Lau TK, Lo YM. Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. Clinical Chemistry. 2008;54(10):1664-72.
222. Oehler VG, Qin J, Ramakrishnan R, Facer G, Ananthnarayan S, Cummings C, Deininger M, Shah N, McCormick F, Willis S, Daridon A, Unger M, Radich JP. Absolute quantitative detection of ABL tyrosine kinase domain point mutations in chronic myeloid leukemia using a novel nanofluidic platform and mutation-specific PCR. Leukemia. 2009;23(2):396-9.
223. Shuga J, Zeng Y, Novak R, Lan Q, Tang X, Rothman N, Vermeulen R, Li L, Hubbard A, Zhang L, Mathies RA, Smith MT. Single molecule quantitation and sequencing of rare translocations using microfluidic nested digital PCR. Nucleic Acids Research. 2013;41(16):e159.
224. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA: A Cancer Journal for Clinicians. 2011;61(2):69-90.
225. Claire OC. Fluorescence in situ Hybridization (FISH), Cytogenetics can now go "FISH-ing" for chromosomal abnormalities, which are deletions an duplications that can cause disease. How exactly does FISH work? Nature Education, Chromosomes and Cytogenetics. 2008;1(1):1-8.
226. Trask BJ. Human cytogenetics: 46 chromosomes, 46 years and counting. Nat Rev Genet. 2002;3(10):769-78.
227. Nashed AL, Rao KW, Gulley ML. Clinical applications of BCR-ABL molecular testing in acute leukemia. J Mol Diagn. 2003;5(2):63-72.
228. Wolff DJ, Bagg A, Cooley LD, Dewald GW, Hirsch BA, Jacky PB, Rao KW, Rao PN. Guidance for fluorescence in situ hybridization testing in hematologic disorders. J Mol Diagn. 2007;9(2):134-43.
229. Ross JS, Cronin M. Whole cancer genome sequencing by next-generation methods. American Journal of Clinical Pathology. 2011;136(4):527-39.
230. Su Z, Ning B, Fang H, Hong H, Perkins R, Tong W, Shi L. Next-generation sequencing and its applications in molecular diagnostics. Expert Rev Mol Diagn. 2011;11(3):333-43.
231. Schaefer A, Jung M, Mollenkopf HJ, Wagner I, Stephan C, Jentzmik F, Miller K, Lein M, Kristiansen G, Jung K. Diagnostic and prognostic implications of microRNA profiling in prostate carcinoma. Int J Cancer. 2010;126(5):1166-76.
232. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T, Calin GA, Liu CG, Croce CM, Harris CC. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006;9(3):189-98.

233. Capper D, Preusser M, Habel A, Sahm F, Ackermann U, Schindler G, Pusch S, Mechtersheimer G, Zentgraf H, von Deimling A. Assessment of BRAF V600E mutation status by immunohistochemistry with a mutation-specific monoclonal antibody. Acta Neuropathol. 2011;122(1):11-9.
234. Janku F, Tsimberidou AM, Garrido-Laguna I, Wang X, Luthra R, Hong DS, Naing A, Falchook GS, Moroney JW, Piha-Paul SA, Wheler JJ, Moulder SL, Fu S, Kurzrock R. PIK3CA mutations in patients with advanced cancers treated with PI3K/AKT/mTOR axis inhibitors. Mol Cancer Ther. 2011;10(3):558-65.
235. Pantel K, Brakenhoff RH, Brandt B. Detection, clinical relevance and specific biological properties of disseminating tumour cells. Nat Rev Cancer. 2008;8(5):329-40.
236. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, Fulton LA, Locke DP, Magrini VJ, Abbott RM, Vickery TL, Reed JS, Robinson JS, Wylie T, Smith SM, Carmichael L, Eldred JM, Harris CC, Walker J, Peck JB, Du F, Dukes AF, Sanderson GE, Brummett AM, Clark E, McMichael JF, Meyer RJ, Schindler JK, Pohl CS, Wallis JW, Shi X, Lin L, Schmidt H, Tang Y, Haipek C, Wiechert ME, Ivy JV, Kalicki J, Elliott G, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson MA, Baty J, Heath S, Shannon WD, Nagarajan R, Link DC, Walter MJ, Graubert TA, DiPersio JF, Wilson RK, Ley TJ. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med. 2009;361(11):1058-66.
237. Weir B, Zhao X, Meyerson M. Somatic alterations in the human cancer genome. Cancer Cell. 2004;6(5):433-8.
238. van Eijk R, Licht J, Schrumpf M, Talebian Yazdi M, Ruano D, Forte GI, Nederlof PM, Veselic M, Rabe KF, Annema JT, Smit V, Morreau H, van Wezel T. Rapid KRAS, EGFR, BRAF and PIK3CA mutation analysis of fine needle aspirates from non-small-cell lung cancer using allele-specific qPCR. PLoS One. 2011;6(3):e17791.
239. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, Maher E, Kaye FJ, Sasaki H, Tepper JE, Fletcher JA, Tabernero J, Baselga J, Tsao MS, Demichelis F, Rubin MA, Janne PA, Daly MJ, Nucera C, Levine RL, Ebert BL, Gabriel S, Rustgi AK, Antonescu CR, Ladanyi M, Letai A, Garraway LA, Loda M, Beer DG, True LD, Okamoto A, Pomeroy SL, Singer S, Golub TR, Lander ES, Getz G, Sellers WR, Meyerson M. The landscape of somatic copy-number alteration across human cancers. Nature. 2010;463(7283):899-905.
240. Chiang DY, Getz G, Jaffe DB, O'Kelly MJ, Zhao X, Carter SL, Russ C, Nusbaum C, Meyerson M, Lander ES. High-resolution mapping of copy-number alterations with massively parallel sequencing. Nat Methods. 2009;6(1):99-103.
241. Lee J, Seo JW, Jun HJ, Ki CS, Park SH, Park YS, Lim HY, Choi MG, Bae JM, Sohn TS, Noh JH, Kim S, Jang HL, Kim JY, Kim KM, Kang WK, Park JO. Impact of MET amplification on gastric cancer: possible roles as a novel prognostic marker and a potential therapeutic target. Oncol Rep. 2011;25(6):1517-24.
242. Berggren P, Kumar R, Sakano S, Hemminki L, Wada T, Steineck G, Adolfsson J, Larsson P, Norming U, Wijkstrom H, Hemminki K. Detecting homozygous deletions in the CDKN2A(p16(INK4a))/ARF(p14(ARF)) gene in urinary bladder cancer using real-time quantitative PCR. Clinical Cancer Research. 2003;9(1):235-42.

243. Lundberg P, Karow A, Nienhold R, Looser R, Hao-Shen H, Nissen I, Girsberger S, Lehmann T, Passweg J, Stern M, Beisel C, Kralovics R, Skoda RC. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. Blood. 2014;123(14):2220-8.
244. Mathews J, Duncavage EJ, Pfeifer JD. Characterization of translocations in mesenchymal hamartoma and undifferentiated embryonal sarcoma of the liver. Exp Mol Pathol. 2013;95(3):319-24.
245. Grossmann V, Kohlmann A, Klein HU, Schindela S, Schnittger S, Dicker F, Dugas M, Kern W, Haferlach T, Haferlach C. Targeted next-generation sequencing detects point mutations, insertions, deletions and balanced chromosomal rearrangements as well as identifies novel leukemia-specific fusion genes in a single procedure. Leukemia. 2011;25(4):671-80.
246. Branford S, Hughes T. Diagnosis and monitoring of chronic myeloid leukemia by qualitative and quantitative RT-PCR. Methods Mol Med. 2006;125:69-92.
247. Jennings LJ, George D, Czech J, Yu M, Joseph L. Detection and Quantification of BCR-ABL1 Fusion Transcripts by Droplet Digital PCR. The Journal of Molecular Diagnostics. 2014;16(2):174-9.
248. Ren R. Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. Nat Rev Cancer. 2005;5(3):172-83.
249. Baxter EJ, Hochhaus A, Bolufer P, Reiter A, Fernandez JM, Senent L, Cervera J, Moscardo F, Sanz MA, Cross NC. The t(4;22)(q12;q11) in atypical chronic myeloid leukaemia fuses BCR to PDGFRA. Human Molecular Genetics. 2002;11(12):1391-7.
250. Roumiantsev S, Krause DS, Neumann CA, Dimitri CA, Asiedu F, Cross NC, Van Etten RA. Distinct stem cell myeloproliferative/T lymphoma syndromes induced by ZNF198-FGFR1 and BCR-FGFR1 fusion genes from 8p11 translocations. Cancer Cell. 2004;5(3):287-98.
251. Demiroglu A, Steer EJ, Heath C, Taylor K, Bentley M, Allen SL, Koduru P, Brody JP, Hawson G, Rodwell R, Doody ML, Carnicero F, Reiter A, Goldman JM, Melo JV, Cross NC. The t(8;22) in chronic myeloid leukemia fuses BCR to FGFR1: transforming activity and specific inhibition of FGFR1 fusion proteins. Blood. 2001;98(13):3778-83.
252. Sanders R, Mason DJ, Foy CA, Huggett JF. Evaluation of digital PCR for absolute RNA quantification. PLoS One. 2013;8(9):e75296.
253. Kishi K. A new leukemia cell line with Philadelphia chromosome characterized as basophil precursors. Leukemia Research. 1985;9(3):381-90.
254. Ross DM, Schafranek L, Hughes TP, Nicola M, Branford S, Score J. Genomic translocation breakpoint sequences are conserved in BCR-ABL1 cell lines despite the presence of amplification. Cancer Genetics and Cytogenetics. 2009;189(2):138-9.
255. Virgili A, Nacheva EP. Genomic amplification of BCR/ABL1 and a region downstream of ABL1 in chronic myeloid leukaemia: a FISH mapping study of CML patients and cell lines. Molecular Cytogenetics. 2010;3:15.
256. Gahmberg CG, Jokinen M, Andersson LC. Expression of the major red cell sialoglycoprotein, glycophorin A, in the human leukemic cell line K562. The Journal of Biological Chemistry. 1979;254(15):7442-8.
257. Lozzio CB, Lozzio BB. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. Blood. 1975;45(3):321-34.
258. Naumann S, Reutzel D, Speicher M, Decker HJ. Complete karyotype characterization of the K562 cell line by combined application of G-banding, multiplex-fluorescence in situ hybridization, fluorescence in situ hybridization, and comparative genomic hybridization. Leukemia Research. 2001;25(4):313-22.

259. Gribble SM, Roberts I, Grace C, Andrews KM, Green AR, Nacheva EP. Cytogenetics of the chronic myeloid leukemia-derived cell line K562: karyotype clarification by multicolor fluorescence in situ hybridization, comparative genomic hybridization, and locus-specific fluorescence in situ hybridization. Cancer Genetics and Cytogenetics. 2000;118(1):1-8.
260. Brazma DV, A. Grace, C. Howard, J. Chanalaris, A. Melo, J.V. Apperley, J.F. Nacheva, E.P. When is a Double Philadelphia not a Double Philadelphia. Poster, Leukemia Research. 2010.
261. Burke BA, Carroll M. BCR-ABL: a multi-faceted promoter of DNA mutation in chronic myelogeneous leukemia. Leukemia. 2010;24(6):1105-12.
262. Ogura M, Morishima Y, Ohno R, Kato Y, Hirabayashi N, Nagura H, Saito H. Establishment of a novel human megakaryoblastic leukemia cell line, MEG-01, with positive Philadelphia chromosome. Blood. 1985;66(6):1384-92.
263. Dancis BM. Shear Breakage of DNA. Biophysical Journal. 1978;24(2):489-503.
264. Bowman RD, Davidson N. Hydrodynamic Shear Breakage of DNA. Biopolymers. 1972;11(12):2601-24.
265. Blin N, Stafford DW. A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Research. 1976;3(9):2303-8.
266. Brown AM. A step-by-step guide to Non-Linear Regression Analysis of Experimental Data using a Microsoft Excel Spreadsheet. Computer Methods and Programs in Biomedicine. 2001;65(3):191-200.
267. Bowen WP, Jerman JC. Nonlinear Regression using Spreadsheets. Trends Pharmacological Sciences. 1995;16(12):413-7.
268. Wim T, Matthijs V, De Neve Jan BP, Ward S. ddpcRquant: threshold determination for single channel droplet digital PCR experiments 2.
269. Taly V, Pekin D, El Abed A, Laurent-Puig P. Detecting biomarkers with microdroplet technology. Trends in Molecular Medicine. 2012;18(7):405-16.
270. McCaughan F, Dear PH. Single-molecule genomics. The Journal of Pathology. 2010;220(2):297-306.
271. Yung TK, Chan KA, Mok TS, Tong J, To K-F, Lo YD. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non–small cell lung cancer patients. Clinical Cancer Research. 2009;15(6):2076-84.
272. Hayden R, Gu Z, Ingersoll J, Abdul-Ali D, Shi L, Pounds S, Caliendo A. Comparison of droplet digital PCR to real-time PCR for quantitative detection of cytomegalovirus. Journal of Clinical Microbiology. 2013;51(2):540-6.
273. Bio-Rad. Measurement and statistics of ddPCR Experiments: Bio-Rad2012.
274. Mullis KB. The unusual origin of the polymerase chain reaction. Scientific American. 1990;262(4):56-61.
275. Torre L, Bray A, Siegel F, Ferlay R, Lortet-Tieulent J, Ahmedin J. Global Cancer Statistics, 2012. CA: A Cancer Journal for Clinicians. 2015;65(2):87-108.
276. Martinez P, Hernandez-Losa J, Montero MA, Cedres S, Castellvi J, Martinez-Marti A, Tallada N, Murtra-Garrell N, Navarro-Mendivill A, Rodriguez-Freixinos V, Canela M, Ramon y Cajal S, Felip E. Fluorescence in situ Hybridization and Immunohistochemistry as Diagnostic Methods for ALK positive Non-Small Cell Lung Cancer Patients. PLoS One. 2013;8(1):e52261.
277. Antoniu SA. Crizotinib for EML4-ALK Positive Lung Adenocarcinoma: A hope for the advanced disease? Evaluation of Kwak EL, Bang YJ, Camidge DR, et al. Anaplastic lymphoma

kinase inhibition in non-small-cell lung cancer. New England Journal of Medicine 2010;363(18):1693-703. Expert Opinion on Therapeutic Targets. 2011;15(3):351-3.
278. Camidge DR, Bang YJ, Kwak EL, Iafrate AJ, Varella-Garcia M, Fox SB, Riely GJ, Solomon B, Ou SH, Kim DW, Salgia R, Fidias P, Engelman JA, Gandhi L, Janne PA, Costa DB, Shapiro GI, Lorusso P, Ruffner K, Stephenson P, Tang Y, Wilner K, Clark JW, Shaw AT. Activity and Safety of Crizotinib in Patients with ALK-positive Non-Small-Cell Lung Cancer: Updated Results from a Phase 1 Study. Lancet Oncology. 2012;13(10):1011-9.
279. Shaw AT, Kim DW, Nakagawa K, Seto T, Crino L, Ahn MJ, De Pas T, Besse B, Solomon BJ, Blackhall F, Wu YL, Thomas M, O'Byrne KJ, Moro-Sibilot D, Camidge DR, Mok T, Hirsh V, Riely GJ, Iyer S, Tassell V, Polli A, Wilner KD, Janne PA. Crizotinib versus Chemotherapy in Advanced ALK-Positive Lung Cancer. New England Journal of Medicine. 2013;368(25):2385-94.
280. Martelli MP, Sozzi G, Hernandez L, Pettirossi V, Navarro A, Conte D, Gasparini P, Perrone F, Modena P, Pastorino U, Carbone A, Fabbri A, Sidoni A, Nakamura S, Gambacorta M, Fernandez PL, Ramirez J, Chan JK, Grigioni WF, Campo E, Pileri SA, Falini B. EML4-ALK Rearrangement in Non-Small Cell Lung Cancer and Non-Tumor Lung Tssues. American Journal of Pathology. 2009;174(2):661-70.
281. Wolff DJ, Bagg A, Cooley LD, Dewald GW, Hirsch BA, Jacky PB, Rao KW, Rao PN. Guidance for Fluorescence in situ Hybridization Testing in Hematologic Disorders. Journal of Molecular Diagnostics. 2007;9(2):134-43.
282. Kim H, Yoo S-B, Choe J-Y, Paik JH, Xu X, Nitta H, Zhang W, Grogan TM, Lee C-T, Jheon S. Detection of ALK Gene Rearrangement in Non-Small Cell Lung Cancer: A Comparison of Fluorescence in situ Hybridization and Chromogenic in situ Hybridization with correlation of ALK Protein Expression. Journal of Thoracic Oncology. 2011;6(8):1359-66.
283. 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(5):546-58.
284. Nambiar M, Choudhary B, Rao CR, Raghavan SC. Amplification of Chromosomal Translocation Junctions from Paraffin-Embedded Yissues of Follicular Lymphoma Patients. Biomedical Materials. 2008;3(3):034103.
285. McCaughan F, Darai-Ramqvist E, Bankier AT, Konfortov BA, Foster N, George PJ, Rabbitts TH, Kost-Alimova M, Rabbitts PH, Dear PH. Microdissection molecular copy-number counting (microMCC)--unlocking cancer archives with digital PCR. Journal of Pathology. 2008;216(3):307-16.
286. Lund L, Hughesman C, McNeil K, Clemens S, Hocken K, Pettersson R, Karsan A, Foster L, Haynes C. Initial Diagnosis of Chronic Myelogenous Leukemia Based on Quantification of M-BCR Status Using Droplet Digital PCR. Analytical and Bioanalytical Chemistry. 2015:DOI: 10.1007/s00216-015-9204-2.
287. Li Y, Ye X, Liu J, Zha J, Pei L. Evaluation of EML4-ALK Fusion Proteins in Non-Small Cell Lung Cancer using Small Molecule Inhibitors. Neoplasia. 2011;13(1):1-11.
288. Xie R, Chung J-Y, Ylaya K, Williams RL, Guerrero N, Nakatsuka N, Badie C, Hewitt SM. Factors Influencing the Degradation of Archival Formalin-Fixed Paraffin-Embedded Tissue Sections. Journal of Histochemistry & Cytochemistry. 2011;59(4):356-65.
289. Wang G, Maher E, Brennan C, Chin L, Leo C, Kaur M, Zhu P, Rook M, Wolfe JL, Makrigiorgos GM. DNA Amplification Method Tolerant to Sample Degradation. Genome Research. 2004;14(11):2357-66.

290. Ilie MI, Bence C, Hofman V, Long-Mira E, Butori C, Bouhlel L, Lalvee S, Mouroux J, Poudenx M, Otto J, Marquette CH, Hofman P. Discrepancies between FISH and Immunohistochemistry for Assessment of the ALK status are associated with ALK 'borderline'-positive rearrangements or a high copy number: a potential major issue for anti-ALK therapeutic strategies. Annals of Oncology. 2015;26(1):238-44.
291. Johung KL, Yeh N, Desai NB, Williams TM, Lautenschlaeger T, Arvold ND, Ning MS, Attia A, Lovly CM, Goldberg S, Beal K, Yu JB, Kavanagh BD, Chiang VL, Camidge DR, Contessa JN. Extended Survival and Prognostic Factors for Patients With ALK-Rearranged Non-Small-Cell Lung Cancer and Brain Metastasis. Journal of Clinical Oncology. 2015.
292. Zheng Z, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, Chen J, Robinson HE, Shim HS, Chmielecki J, Pao W, Engelman JA, Iafrate AJ, Le LP. Anchored Multiplex PCR for Targeted Next-Generation Sequencing. Nature Medicine. 2014;20(12):1479-84.
293. Awad MM, Katayama R, McTigue M, Liu W, Deng Y-L, Brooun A, Friboulet L, Huang D, Falk MD, Timofeevski S. Acquired Resistance to Crizotinib from a Mutation in CD74–ROS1. New England Journal of Medicine. 2013;368(25):2395-401.
294. Chen R, Mias George I, Li-Pook-Than J, Jiang L, Lam Hugo YK, Chen R, Miriami E, Karczewski Konrad J, Hariharan M, Dewey Frederick E, Cheng Y, Clark Michael J, Im H, Habegger L, Balasubramanian S, O'Huallachain M, Dudley Joel T, Hillenmeyer S, Haraksingh R, Sharon D, Euskirchen G, Lacroute P, Bettinger K, Boyle Alan P, Kasowski M, Grubert F, Seki S, Garcia M, Whirl-Carrillo M, Gallardo M, Blasco Maria A, Greenberg Peter L, Snyder P, Klein Teri E, Altman Russ B, Butte AJ, Ashley Euan A, Gerstein M, Nadeau Kari C, Tang H, Snyder M. Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes. Cell.148(6):1293-307.
295. Bhat S, Curach N, Mostyn T, Bains GS, Griffiths KR, Emslie KR. Comparison of methods for accurate quantification of DNA mass concentration with traceability to the international system of units. Analytical Chemistry. 2010;82(17):7185-92.
296. 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.
297. Sykes PJ, Neoh SH, Brisco MJ, Hughes E, Condon J, Morley AA. Quantitation of targets for PCR by use of limiting dilution. Biotechniques. 1992;13(3):444-9.
298. Zhou W, Galizia G, Lieto E, Goodman SN, Romans KE, Kinzler KW, Vogelstein B, Choti MA, Montgomery EA. Counting alleles reveals a connection between chromosome 18q loss and vascular invasion. Nature Biotechnology. 2001;19(1):78-81.
299. Smith RA, Cokkinides V, Eyre HJ. American Cancer Society guidelines for the early detection of cancer, 2006. CA: A Cancer Journal for Clinicians. 2006;56(1):11-25.
300. Slade C, Talbot R. Sustainability of cancer waiting times: the need to focus on pathways relevant to the cancer type. Journal of the Royal Society of Medicine. 2007;100(7):309-13.
301. Schnipper LE, Davidson NE, Wollins DS, Tyne C, Blayney DW, Blum D, Dicker AP, Ganz PA, Hoverman JR, Langdon R, Lyman GH, Meropol NJ, Mulvey T, Newcomer L, Peppercorn J, Polite B, Raghavan D, Rossi G, Saltz L, Schrag D, Smith TJ, Yu PP, Hudis CA, Schilsky RL. American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options. Journal of Clinical Oncology. 2015;33(23):2563-77.
302. Totoki Y, Tatsuno K, Yamamoto S, Arai Y, Hosoda F, Ishikawa S, Tsutsumi S, Sonoda K, Totsuka H, Shirakihara T, Sakamoto H, Wang L, Ojima H, Shimada K, Kosuge T, Okusaka T, Kato K, Kusuda J, Yoshida T, Aburatani H, Shibata T. High-resolution characterization of a hepatocellular carcinoma genome. Nature Genetics. 2011;43(5):464-9.

303. Wilson KD, Schrijver I. Transitioning Diagnostic Molecular Pathology to the Genomic Era: Cancer Somatic Mutation Panel Testing. Molecular Testing in Cancer. Springer; 2014. p. 3-13.
304. Kiss MM, Ortoleva-Donnelly L, Beer NR, Warner J, Bailey CG, Colston BW, Rothberg JM, Link DR, Leamon JH. High-throughput quantitative polymerase chain reaction in picoliter droplets. Analytical Chemistry. 2008;80(23):8975-81.
305. Rawer D, Borkhardt A, Wilda M, Kropf S, Kreuder J. Influence of stochastics on quantitative PCR in the detection of minimal residual disease. Leukemia. 0000;17(12):2527-8.
306. Sanders R, Mason DJ, Foy CA, Huggett JF. Considerations for accurate gene expression measurement by reverse transcription quantitative PCR when analysing clinical samples. Analytical and Bioanalytical Chemistry. 2014;406(26):6471-83.
307. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458(7239):719-24.
308. 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. Clinical Chemistry. 2013;59(12):1722-31.
309. Beck J, Bierau S, Balzer S, Andag R, Kanzow P, Schmitz J, Gaedcke J, Moerer O, Slotta JE, Walson P, Kollmar O, Oellerich M, Schutz E. Digital droplet PCR for rapid quantification of donor DNA in the circulation of transplant recipients as a potential universal biomarker of graft injury. Clinical Chemistry. 2013;59(12):1732-41.
310. Devonshire AS, Whale AS, Gutteridge A, Jones G, Cowen S, Foy CA, Huggett JF. Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. Analytical and Bioanalytical Chemistry. 2014;406(26):6499-512.
311. Sundaresan TK, Sequist LV, Heymach JV, Riely GJ, Janne PA, Koch WH, Sullivan JP, Fox DB, Maher R, Muzikansky A, Webb A, Tran HT, Giri U, Fleisher M, Yu H, Wei W, Johnson BE, Barber TA, Walsh JR, Engelman JA, Stott SL, Kapur R, Maheswaran S, Toner M, Haber DA. Detection of T790M, the acquired resistance EGFR mutation, by tumor biopsy versus noninvasive blood-based analyses. Clinical Cancer Research. 2015.

Appendices

Appendix A Examples of Raw Data and Data Analysis for the ddPCR *BCR* Status Assay

Table A-1. Raw data from the ddPCR *BCR* status assay applied to independent replicates of KU812 cell line gDNA, including the sample shown in Figure 2-4A (sample 8).

cpd	no copies	probability all droplets
0.256	0	0.7740
0.256	1	0.1983
0.256	2	0.0254
0.256	3	0.0022
		0.9999

f2
0.1135684

mean CPD	Alexa+	FAM+	HEX+	FAM+HEX+	HEX+ Alexa+	Fam+ Alexa+	FAM+ HEX+ Alexa+	empty droplets	total droplets	doubles of A+F+H+	A+	F+	H+	A+F+	F+H+	A+ F+ H+	total doubles of A+F+	A+	F+	A+F+	total doubles of F+H+	F+	H+	F+H+	total doubles of A+H+	A+	H+	total doubles of A+	A+	total doubles of F+	F+	total doubles of H+	H+	Total A+	Total F+	Total H+	Total A+F+	Total F+H+	Total A+F+ H+
0.268	408	399	1322	57	46	894	112	10520	13758	13	3	1	3	4	4	9	102	58	58	87	6	4	4	6	46	46	46	46	45	45	150	150	566	512	1531	884	61	108
0.265	431	372	1407	55	47	982	133	11291	14718	15	4	2	4	5	5	11	112	64	64	96	6	4	4	5	47	47	47	49	49	42	42	160	160	598	487	1626	971	60	128
0.231	321	323	1255	41	40	902	103	11500	14485	12	3	1	3	4	4	8	102	59	59	88	5	3	3	4	40	40	40	36	36	37	37	143	143	450	413	1430	892	44	100
0.223	308	308	1282	31	30	900	123	11953	14935	14	3	2	3	5	5	10	102	58	58	88	4	2	2	3	30	30	30	35	35	35	35	146	146	423	394	1445	891	35	119
0.222	257	246	1189	27	20	819	101	10721	13380	11	3	1	3	4	4	8	93	53	53	80	3	2	2	3	20	20	20	29	29	28	28	135	135	352	320	1331	811	30	98
0.222	281	295	1303	37	31	987	115	12273	15322	13	3	2	3	5	5	9	112	64	64	96	4	2	2	4	31	31	31	32	32	34	34	148	148	399	384	1469	977	41	112
0.219	327	342	1292	39	29	984	101	12719	15833	11	3	1	3	4	4	8	112	64	64	96	4	3	3	4	29	29	29	37	37	39	39	147	147	446	435	1453	974	42	98
0.256	568	562	1272	63	73	715	94	11460	14807	11	3	1	3	4	4	8	81	46	46	70	7	4	4	6	73	73	73	65	65	64	64	144	144	754	678	1496	707	66	91

Table A-2. Representative model analysis results for *BCR* status assay applied to replicate samples of KU812 gDNA.

KU812										
CPD	START	A+	F+	H+	A+F+	F+H+	A+F+H+	Total Alexa+	Total FAM+	Total *ALK*
0.268	1561	566	512	1531	884	61	108	1557	1564	1561
0.265	1672	598	487	1626	971	60	128	1697	1646	1672
0.231	1446	450	413	1430	892	44	100	1442	1450	1446
0.223	1437	423	394	1445	891	35	119	1434	1439	1437
0.222	1260	352	320	1331	811	30	98	1261	1259	1260
0.222	1501	399	384	1469	977	41	112	1488	1513	1501

trans **b**	Shear **s**	model 2 **l**
0.9212	0.2100	0.0000

	x+(1-x)	SUM	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt
	models		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
A	0.363	0.062	0.362	0.358	0.311	0.295	0.279	0.266	0.064	0.112	0.100	0.107	0.102	0.101	0.000	0.000	0.000	0.000	0.000	0.000
F	0.350	1.526	0.328	0.291	0.286	0.274	0.254	0.256	1.547	1.554	1.606	1.626	1.701	1.630	0.000	0.000	0.000	0.000	0.000	0.000
H	0.938	0.713	0.981	0.973	0.989	1.006	1.057	0.979	0.714	0.649	0.597	0.569	0.533	0.522	0.000	0.000	0.000	0.000	0.000	0.000
AF	0.588		0.566	0.581	0.617	0.620	0.644	0.651												
FH	0.013		0.039	0.036	0.031	0.024	0.024	0.027												
AFH	0.049		0.069	0.077	0.069	0.083	0.078	0.074												

sum	0.0001	0.0002	0.0002	0.0001	0.0002	0.0002
SQRT	0.00830	0.01424	0.01256	0.00902	0.01371	0.01275

expt	trans **b**	Shear **s**	loss **l**
1	0.921	0.210	0.000
2	0.932	0.209	0.000
3	0.932	0.178	0.000
4	0.933	0.171	0.000
5	0.964	0.160	0.000
6	0.927	0.155	0.000

Appendix B Examples of Raw Data and Data Analysis for the ddPCR *ALK* Status Assay

Below is an example computation of values on the right-hand side of Equations 3.1 – 3.3. We apply our data analysis method at the $n = 2$ level (i.e., no more than 2 copies per droplet considered) to the 2D output data shown in Figure 3.4A for gDNA purified from H2228 cells. $Alexa^+$ cluster = 245 droplets, FAM^+ cluster = 134 droplets, HEX^+ cluster = 129 droplets, $Alexa^+FAM^+$ cluster = 1000 droplets, FAM^+HEX^+ cluster = 134 droplets, $Alexa^+HEX^+$ cluster = 6 droplets, and $Alexa^+FAM^+HEX^+$ cluster = 1627 droplets. The total number of empty droplets = 10772 droplets, and the total number of read droplets = 14058.

From this, the CPD ($= -ln(empty\ droplets/total\ droplets)$ = 0.266) is first computed and then combined with Poisson statistics to estimate the fraction of total read droplets containing n = 0, 1, 2, 3, … copies as $\frac{CPD^n exp^{-CPD}}{n!}$. For this sample, the fraction of empty droplets is 0.76625, n = 1 copy droplets is 0.20401, and n = 2 droplets is 0.02716.

Using these values, one then calculates the fraction $f2$ of all filled droplets that contain 2 copies (0.02716/(0.20401+0.02716) = 0.1175) to enable determination of the number of droplets within each cluster containing 2 copies. Beginning with the $Alexa^+FAM^+HEX^+$ cluster:

$Alexa^+FAM^+HEX^+$ droplets containing 2 copies = cluster count x $f2$ = 1627 x 0.1175 = 191

Then all pairs of copies that would generate a combined $Alexa^+FAM^+HEX^+$ signal within a droplet can be defined. Each copy may either be intact *ALK* (which on its own generates an $Alexa^+FAM^+HEX^+$ signal within a droplet) or a fragment of *ALK*. The 5 possible fragments of interest generate an $Alexa^+$ signal, a FAM^+ signal, a HEX^+ signal, an $Alexa^+FAM^+$ signal, or a FAM^+HEX^+ signal, respectively. There are 17 combinations that generate an $Alexa^+FAM^+HEX^+$ signal within a droplet (e.g., $Alexa^+$/ $Alexa^+FAM^+HEX^+$ and $Alexa^+FAM^+HEX^+$/$Alexa^+$ are two possible combinations, while the $Alexa^+FAM^+HEX^+$/$Alexa^+FAM^+HEX^+$ combination can only be formed in one way due to the two contributing signals being indistinguishable). From this, the

total abundance of each unique signal within the $Alexa^+FAM^+HEX^+$ droplets containing two copies is computed as:

$Alexa^+$ = 191 x 4/17 = 45 FAM^+ = 191 x 2/17 = 22

HEX^+ = 191 x 4/17 = 45 $Alexa^+FAM^+$ = 191 x 6/17 = 67

FAM^+HEX^+ = 191 x 6/17 = 67 $Alexa^+FAM^+HEX^+$ = 191 x (10/17 + 2x1/17) = 135

This analysis is repeated for the set of 5 positive clusters displaying a signal for one of the 5 possible *ALK* fragments generated by a biologic double stranded break (DSB) and/or a shear event:

$Alexa^+$ cluster droplets containing 2 copies = 245 x 0.1175 = 29

$Alexa^+$ = 29 x (2x1/1) = 58

FAM^+ cluster droplets containing 2 copies = 134 x 0.1175 = 16

FAM^+ = 16 x (2x1/1) = 32

HEX^+ cluster droplets containing 2 copies = 129 x 0.1175 = 15

HEX^+ = 15 x (2x1/1) = 30

$Alexa^+FAM^+$ cluster droplets containing 2 copies = 1000 x 0.1175 = 117

$Alexa^+$ = 117 x 4/7 = 67 FAM^+ = 117 x 4/7 = 67

$Alexa^+FAM^+$ = 117 x (4/7 + 2x1/7) = 101

FAM^+HEX^+ cluster droplets containing 2 copies = 134 x 0.1175 = 16

FAM^+ = 16 x 4/7 = 9 HEX^+ = 16 x 4/7 = 9

FAM^+HEX^+ = 16 x (4/7 + 2x1/7) = 13

From this information and the fact that the $Alexa^+HEX^+$ cluster (6 droplets) must be comprised of 6 pairs of $Alexa^+$ + HEX^+, the total abundance of each unique signal is then computed.

Total $Alexa^+FAM^+HEX^+$ = 1627 – 191 + 135 = 1571

Total $Alexa^+$ = 245 – 29 + 45 + 6 + 67 + 58 = 392

Total FAM^+ = 134 – 16 + 22 + 9 + 67 + 32 = 248

Total HEX^+ = 129 – 15 + 45 + 9 + 6 + 30 = 204

Total $Alexa^+FAM^+$ = 1000 – 117 + 67 + 101 = 1051

Total FAM^+HEX^+ = 134 – 16 + 13 + 67 = 198

The total abundance of all forms of the *ALK* gene in the sample is estimated as

Total of all $Alexa^+$-containing signals = 392 + 1051 + 1571 = 3014

Total of all FAM^+-containing signals = 248 + 1057 + 198 + 1571 = 3074

Total *ALK* gene = (3014 + 3074)/2 = 3044

These results are used to compute the required value on the left-hand side of equations 3.1 – 3.3. For equation 3.1, for example:

$$\frac{Alexa^+FAM^+HEX^+}{Total\ ALK} + \frac{FAM^+HEX^+}{Total\ ALK} = \frac{1571}{3044} + \frac{198}{3044} = (1-b)(1-s)$$

Equations 3.1 to 3.3 may then be solved as described in the main document to determine values for *b* (fraction of total *ALK* copies that have undergone a biological DSB), *s* (fraction of total copies that have undergone a shear event), and *l* (fraction of total copies that exhibit a loss of HEX signal).

Table B-1. Raw data from the ddPCR *ALK* status assay applied to independent replicates of H2228 cell line gDNA.

cpd	no copies	probability all droplets
0.276	0	0.7591
0.276	1	0.2092
0.276	2	0.0288
0.276	3	0.0027
		0.9998

f2
0.1211351

mean CPD	Alexa+	FAM+	HEX+	FAM+HEX+	HEX+ Alexa+	Fam+Alexa+	FAM+HEX+ Alexa+	empty droplets	total droplets	doubles of A+F+H+	A+	F+	H+	A+F+	F+H+	A+F+H+	total doubles of A+F+	A+	F+	A+F+	total doubles of F+H+	F+	H+	F+ H+	total doubles of A+H+	A+	H+	total doubles of A+	A+	total doubles of F+	F+	total doubles of H+	H+	Total A+	Total F+	Total H+	Total A+F+	Total F+H+	Total A+F+ H+
0.289	299	119	167	156	11	904	1411	9139	12206	171	40	20	40	60	60	121	110	63	63	94	19	11	11	16	11	11	11	36	36	14	14	20	20	455	232	252	951	216	1359
0.294	268	140	157	179	14	930	1474	9234	12396	179	42	21	42	63	63	126	113	64	64	97	22	12	12	19	14	14	14	32	32	17	17	19	19	429	262	249	980	242	1418
0.281	266	117	174	165	15	973	1547	10053	13310	187	44	22	44	66	66	132	118	67	67	101	20	11	11	17	15	15	15	32	32	14	14	21	21	427	234	267	1023	229	1491
0.278	249	113	144	137	16	870	1377	9064	11970	167	39	20	39	59	59	118	105	60	60	90	17	9	9	14	16	16	16	30	30	14	14	17	17	396	217	227	914	194	1328
0.285	218	134	109	141	10	846	1373	8601	11432	166	39	20	39	59	59	117	102	59	59	88	17	10	10	15	10	10	10	26	26	16	16	13	13	356	241	183	891	199	1323
0.283	260	128	149	132	14	920	1485	9445	12533	180	42	21	42	63	63	127	111	64	64	96	16	9	9	14	14	14	14	31	31	16	16	18	18	415	240	234	969	195	1431
0.283	224	113	141	153	8	853	1362	8742	11596	165	39	19	39	58	58	116	103	59	59	89	19	11	11	16	8	8	8	27	27	14	14	17	17	360	218	217	897	210	1312
0.276	297	120	173	144	8	943	1503	10044	13232	182.07	43	21	43	64.3	64	129	114	65	65	98	17	10	10	15	8	8	8	36	36	15	15	21	21	449	231	255	991	206	1449

Table B-2. Representative model analysis results for *ALK* status assay applied to replicate samples of gDNA from the H2228 cell line

H2228										
CPD	START	A+	F+	H+	A+F+	F+H+	A+F+H+	Total Alexa+	Total FAM+	Total *BCR*
0.289	2761	455	232	252	951	216	1359	2764	2757	2761
0.294	2865	429	262	249	980	242	1418	2827	2902	2865
0.281	2959	427	234	267	1023	229	1491	2941	2977	2959
0.278	2645	396	217	227	914	194	1328	2637	2652	2645
0.285	2612	356	241	183	891	199	1323	2570	2654	2612
0.283	2824	415	240	234	969	195	1431	2814	2834	2824

trans **b**	Shear **s**	Loss **l**
0.3550	0.1248	0.9693

	x+(1-x)	SUM	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt	expt
	models		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	
a	0.164	0.564	0.165	0.150	0.144	0.150	0.136	0.147	0.570	0.580	0.581	0.575	0.583	0.576	0.000	0.000	0.000	0.000	0.000	0.000	
f	0.093	0.434	0.084	0.091	0.079	0.082	0.092	0.085	0.436	0.429	0.436	0.431	0.411	0.426	0.000	0.000	0.000	0.000	0.001	0.000	
h	0.091	0.257	0.091	0.087	0.090	0.086	0.070	0.083	0.249	0.241	0.223	0.232	0.228	0.232	0.000	0.000	0.001	0.000	0.000	0.000	
AF	0.342		0.344	0.342	0.346	0.346	0.341	0.343													
FH	0.070		0.078	0.085	0.077	0.073	0.076	0.069													
AFH	0.494		0.492	0.495	0.504	0.502	0.506	0.507													

sum	0.0000	0.0002	0.0008	0.0003	0.0012	0.0004
SQRT	0.00000	0.01391	0.02769	0.01838	0.03405	0.02027

expt	trans **b**	Shear **s**	Loss **l**
1	0.355	0.125	0.969
2	0.354	0.122	0.977
3	0.356	0.113	0.951
4	0.357	0.115	0.967
5	0.354	0.112	1.000
6	0.354	0.114	0.973