



5-2018

# Somatic Mutation Detection in Leukemia-Derived Circulating DNA: Utility in Monitoring Clonal Dynamics and Disease Response in Pediatric Acute Lymphoblastic Leukemia

Sarah Hisham Abdelaziz Youssef  
*University of Tennessee Health Science Center*

Follow this and additional works at: <https://dc.uthsc.edu/dissertations>

 Part of the [Medical Genetics Commons](#), [Medical Molecular Biology Commons](#), and the [Pharmacy and Pharmaceutical Sciences Commons](#)

---

## Recommended Citation

Youssef, Sarah Hisham Abdelaziz (<http://orcid.org/https://orcid.org/0000-0001-8681-2680>), "Somatic Mutation Detection in Leukemia-Derived Circulating DNA: Utility in Monitoring Clonal Dynamics and Disease Response in Pediatric Acute Lymphoblastic Leukemia" (2018). *Theses and Dissertations (ETD)*. Paper 458. <http://dx.doi.org/10.21007/etd.cghs.2018.0453>.

This Thesis is brought to you for free and open access by the College of Graduate Health Sciences at UTHSC Digital Commons. It has been accepted for inclusion in Theses and Dissertations (ETD) by an authorized administrator of UTHSC Digital Commons. For more information, please contact [jwelch30@uthsc.edu](mailto:jwelch30@uthsc.edu).

---

# Somatic Mutation Detection in Leukemia-Derived Circulating DNA: Utility in Monitoring Clonal Dynamics and Disease Response in Pediatric Acute Lymphoblastic Leukemia

**Document Type**

Thesis

**Degree Name**

Master of Science (MS)

**Program**

Pharmaceutical Sciences

**Track**

Pharmacometrics

**Research Advisor**

Charles Gawad, MD, Ph.D.

**Committee**

Hassan Almoazen, Ph.D. Gerard P. Zambetti, Ph.D.

**ORCID**

<http://orcid.org/https://orcid.org/0000-0001-8681-2680>

**DOI**

10.21007/etd.cghs.2018.0453

**Somatic Mutation Detection in Leukemia-Derived Circulating DNA: Utility in  
Monitoring Clonal Dynamics and Disease Response in Pediatric Acute  
Lymphoblastic Leukemia**

A Dissertation  
Presented for  
The Graduate Studies Council  
The University of Tennessee  
Health Science Center

In Partial Fulfillment  
Of the Requirements for the Degree  
Master of Science  
From The University of Tennessee

By  
Sarah Hisham Abdelaziz Youssef  
May 2018

Copyright © 2018 by Sarah Youssef.  
All rights reserved.

## **DEDICATION**

I am grateful for every blessing I was ever granted, every challenge I faced and every hardship I overcame. I am grateful to those people who believed in me and motivated me to keep going when it seemed hopeless. I am grateful to my cancer patients, those lovely innocent kids, who taught me patience and resilience. I am more than grateful to my parents who brought me up to be who I am today. I dedicate every success I achieved or will ever achieve to my mother and grandmother without whom I would have never been writing this today.

## **ACKNOWLEDGEMENTS**

I would like to express my sincere gratitude to my advisor Prof. Charles Gawad for the continuous support, patience, and immense knowledge.

I would also like to thank members of my thesis committee: Prof. Hassan Almoazen, and Prof. Gerard Zambetti for their efforts and help.

I thank my colleagues at Charles Gawad Lab., St. Jude Children's Research Hospital. I especially thank Veronica Gonzalez, Yuki Inaba, and Ousman Mahmoud for their input and help to get this work done.

I am deeply grateful for the Binational Fulbright Commision in Egypt for giving me such a great opportunity.

Last but not the least, I would like to thank my family: my parents, my brother, and my friends for believing in me.

## ABSTRACT

Despite the improved outcome associated with current treatment strategies of pediatric acute lymphoblastic leukemia (ALL), relapse still represents a major challenge. Pediatric ALL demonstrates branched evolution in response to selective pressure exerted by therapy; relapse founder clones emerge from pre-leukemic clones or minor subclones present at diagnosis. It is hence crucial to develop biomarkers capable of tracking subclones throughout therapy. Current practices for monitoring disease response in leukemia rely on the analysis of BM biopsy sample at specific time points throughout therapy. Not only the invasiveness of the BM biopsy hinders the sequential sampling, but also, the currently implied techniques are associated with a lack of sensitivity to detect subclones other than the major diagnostic clone. Somatic mutation detection in circulating-tumor DNA (Ct-DNA) offers a new venue for non-invasive studying of genetic heterogeneity and tracking clonal dynamics throughout therapy. Here, we employ targeted Next-Generation Sequencing (NGS) using a specifically designed ALL custom gene panel for Ct-DNA analysis of sequential plasma samples of 14 pediatric ALL during remission induction therapy. Utilizing 1 ml of plasma, Ct-DNA successfully captured all the clinically relevant somatic single nucleotide variants (SNVs) detected by whole exome sequencing (WES) in bone marrow (BM) biopsy samples at diagnosis. Moreover, we were able to show the ability of Ct-DNA analysis to track the change in the mutant allele fraction (MAF) across multiple time points as well as, to detect mutations in Flowcytometry (FC) MRD-negative patients. Taken together, sequential analysis of Ct-DNA in plasma demonstrates a role, as a non-invasive technique, for detecting the clonal composition as well as, tracking clonal dynamics in pediatric ALL.

## TABLE OF CONTENTS

<b>CHAPTER 1. INTRODUCTION .....</b>	<b>1</b>
Overview .....	1
Aim of the Work .....	2
<b>CHAPTER 2. LITERATURE REVIEW .....</b>	<b>4</b>
Genetics and Pharmacogenomics in ALL .....	4
Introduction.....	4
Germline Variants Associated with Increased Risk of ALL Development.....	4
Recurrent Somatic Genetic Alterations of Precursor B-cell ALL (BCP-ALL) .....	6
Aneuploidy.....	6
Recurrent translocations of pediatric BCP-ALL.....	7
Recurrent genetic alterations of pediatric T-cell ALL.....	7
New ALL subtypes.....	7
Pathways Commonly Mutated in Pediatric ALL .....	8
Pathways commonly mutated in B-cell precursor ALL at diagnosis.....	8
Pathways commonly mutated in T-cell ALL at diagnosis.....	8
Somatic Mutations Associated with Relapse and Therapy Resistance .....	9
Tumor Heterogeneity, Clonal Evolution and Therapy Resistance.....	10
Clinical Implications of Monitoring Clonal Dynamics and Drug Resistance .....	11
Pharmacogenomics of ALL .....	12
Germline variants associated with poor outcome .....	12
Germline SNPs associated with increased toxicity to antileukemic therapy.....	13
Monitoring Disease Response in Pediatric Leukemia .....	14
Introduction.....	14
Detection of Minimal Residual Disease (MRD) in Pediatric ALL.....	15
Sampling.....	15
Analytical platforms for MRD measurement.....	15
Somatic Mutation Detection in Leukemia-derived Circulating DNA .....	16
Introduction.....	16
Characteristics of Cell-free and Circulating-tumor DNA.....	17
Pre-analytical Considerations Affecting Cf-DNA Analysis .....	18
Blood sampling.....	18
Anticoagulants.....	18
Handling of blood samples.....	18
Cf-DNA quantification.....	18
Technological Platforms Used in Circulating-tumor DNA (Ct-DNA) Analysis.....	19
Introduction.....	19
Targeted approaches.....	19
Massive Parallel Sequencing (MPS) approaches.....	21
Error Suppression for Deep Sequencing.....	25
Safe-Sequencing System (“Safe-SeqS”).....	25
Integrated Digital Error Suppression method (iDES).....	26
Ct-DNA in Hematological Malignancies .....	28



Ct-DNA in Pediatric Leukemia.....	28
Rationale of the Study .....	29
Study Objective.....	30
<b>CHAPTER 3. METHODOLOGY .....</b>	<b>32</b>
Patients and Samples.....	32
Separation of Plasma from Peripheral Blood .....	32
Cell-free DNA Extraction and Quantification.....	32
Construction of Adapter-ligated Libraries.....	32
Hybridization-based Capture and Sequencing.....	33
Design of the Mutation Capture Panel.....	33
Processing of Sequencing Data.....	34
<b>CHAPTER 4. RESULTS .....</b>	<b>35</b>
Study Design and Patient Enrollment .....	35
Somatic Mutation Detection in Leukemia-derived Cell-free DNA .....	35
Cf-DNA Isolation and Library Preparation .....	38
Samples for WES. ....	38
Samples for targeted capture.....	38
Gel electrophoresis for Cf-DNA libraries.....	38
Exome Capture.....	38
Sequencing Metrics .....	38
NextSeq. QC metrics for WES.....	38
Miseq. QC metrics for targeted capture. Miseq QC metrics.....	38
Non-synonymous Mutation Detection in Whole-exome Captured Libraries.....	46
Non-synonymous Mutation Detection in Targeted Captured Libraries .....	46
Correlation between WBCs and Cf-DNA yield.....	46
Concordance between PB plasma Ct-DNA and WES in BM biopsy at diagnosis. ....	46
Concordance between PB plasma Ct-DNA and BM plasma Ct-DNA at diagnosis. ....	50
Monitoring clonal dynamics and disease response throughout remission induction therapy.....	50
Comparison of Ct-DNA analysis and FC-MRD across multiple time points.....	50
<b>CHAPTER 5. DISCUSSION .....</b>	<b>69</b>
<b>LIST OF REFERENCES.....</b>	<b>73</b>
<b>APPENDIX A. LIST OF GENES INCLUDED IN ALL GENE PANEL .....</b>	<b>81</b>
<b>APPENDIX B. TOTAL XVII RISK STRATIFICATION .....</b>	<b>87</b>
Criteria for Low-Risk (LR) ALL .....	87
Criteria for Standard-Risk (SR) ALL.....	87
Criteria for High-Risk (HR) ALL .....	87

**APPENDIX C. TOTAL XVII REMISSION INDUCTION TREATMENT**

**SCHEDULE ..... 88**

    Precursor B-cell ALL Remission Induction Treatment Schedule ..... 88

    T-cell ALL Remission Induction Treatment Schedule ..... 88

**VITA..... 89**

## LIST OF TABLES

Table 4-1.	Patient Characteristics.....	36
Table 4-2.	Classification of Patients into Risk Groups by NCI Criteria and Total XVII Criteria.....	37
Table 4-3.	NextSeq. QC Metrics for WES-captured Libraries .....	40
Table 4-4.	Post-capture Miseq. QC Metrics for Targeted-captured Libraries .....	42
Table 4-5.	Median DNA Yield and TLC at Different Time Points During Remission Induction Therapy .....	48
Table 4-6.	Clinically Relevant Single Nucleotide Somatic Variants MAF in BM Biopsy WES and PBPL Ct-DNA at Diagnosis.....	48
Table 4-7.	MAF in Ct-DNA Across Multiple Induction Time Points.....	52
Table 4-8.	FC-MRD Detection at Multiple Induction Time Points .....	52

## LIST OF FIGURES

Figure 4-1. Gel Electrophoresis for Cf-DNA Adapter-ligated Libraries.....	39
Figure 4-2. Gel Electrophoresis for Exome-captured Pooled Libraries (Case#2).....	39
Figure 4-3. Non-synonymous Mutations (Filtered Variant Calls) for Case#2 Whole Exome-captured Libraries .....	47
Figure 4-4. Correlation Between TLC (cells*10e9/L) and Cf-DNA Yield (ng).....	47
Figure 4-5. MAF in Diagnostic BM Biopsy (WES) and PB Plasma Ct-DNA at Diagnosis .....	49
Figure 4-6. Correlation Between MAF in BM Biopsy (WES) and PB Plasma Ct-DNA at Diagnosis .....	49
Figure 4-7. MAF (%) in BM Plasma and PB Plasma in Ct-DNA at Diagnosis.....	51
Figure 4-8. Pearson's Correlation Between MAF Detected in Ct-DNA PB Plasma and BM Plasma .....	51
Figure 4-9. Case#1 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	53
Figure 4-10. Case#2 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	55
Figure 4-11. Case#3 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	56
Figure 4-12. Case#4 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	57
Figure 4-13. Case#7 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	58
Figure 4-14. Case#11 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	60
Figure 4-15. Case#13 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	61
Figure 4-16. Case#5 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	62
Figure 4-17. Case#6 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	63

Figure 4-18. Case#8 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	64
Figure 4-19. Case#9 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	65
Figure 4-20. Case#12 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	66
Figure 4-21. Case#14 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy.....	67

## LIST OF ABBREVIATIONS

AAP	Asparaginase Acute Pancreatitis
ABCB1	ATP Binding Cassette Subfamily B Member 1
ABCC2	ATP Binding Cassette Subfamily C Member 2
ABL1	Abelson Tyrosine-Protein Kinase 1
ABL2	Abelson Tyrosine-Protein Kinase 2
ACTG1	Actin Gamma 1
ADD3	Adducin 3
ADSL	Adenylosuccinate Lyase
AKT	v-akt Murine Thymoma Viral Oncogene Homolog
ALK	Anaplastic Lymphoma Receptor Tyrosine Kinase
ALL	Acute Lymphoblastic Leukemia
ANKS1B	Ankyrin Repeat and Sterile Alpha Motif Domain Containing 1B
ANTXR1	Anthrax Toxin Receptor 1
ARF1	ADP Ribosylation Factor 1
ARID4B	AT-Rich Interaction Domain 4B
ARID5B	AT-Rich Interaction Domain 5B
ASO	Allele Specific Oligo
ATIC	5-Aminoimidazole-4-Carboxamide Ribonucleotide
	Formyltransferase/IMP Cyclohydrolase
ATP8A2	ATPase Phospholipid Transporting 8A2
AUTS2	Autism Susceptibility Candidate 2
BCP	B-Cell Precursor
BCR	Breakpoint Cluster Region
BLNK	B Cell Linker
BM	Bone Marrow
BMPL	Bone Marrow Plasma
BRAF	v-raf Murine Sarcoma Viral Oncogene Homolog B
BRD1	Bromodomain Containing 1
BSAP	B-cell lineage-specific activator protein
BWA-aln	Burrows-Wheeler Aligner
CAPG	Capping Actin Protein, Gelsolin Like
CAPP-Seq.	Cancer Personalized Profiling by deep Sequencing
CBP	Csk-binding protein
CCND3	Cyclin D3
CCO	Cell-of-origin classifier
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
CEBPE	CCAAT/Enhancer Binding Protein Epsilon
CEP72	Centrosomal Protein 72
Cf-DNA	Cell Free DNA
CHFR	Checkpoint with Forkhead and Ring Finger Domains
CNS	Central Nervous System
CNV	Copy Number Variation
COSMIC	Catalogue of Somatic Mutations in Cancer

CRC	Colorectal Carcinoma
CREB	CAMP Responsive Element Binding Protein
CREBBP	CREB Binding Protein
CRLF2	Cytokine Receptor Like Factor 2
CSF1R	Colony Stimulating Factor 1 Receptor
CT	Computed Tomography
Ct-DNA	Circulating Tumor DNA
Ct-MRD	Circulating Tumor Minimal Residual Disease
CTNNB1	Catenin Beta 1
DACH1	Dachshund Family Transcription Factor 1
DCB	Durable Clinical Benefit
ddPCR	Digital Droplet PCR
DLBCL	Diffuse Large B-Cell Lymphoma
DNAH8	Dynein Axonemal Heavy Chain 8
ds DNA	Double stranded DNA
DUX4	Double Homeobox 4
EDTA	Ethylene Diamine Tetra Acetic Acid
EED	Embryonic Ectoderm Development
EFS	Event Free Survival
EGFR	Epidermal Growth Factor Receptor
ELN	Elastin
EOI	End of Induction
EP300	E1A Binding Protein P300
EPOR	Erythropoietin Receptor
ERBB2	Erb-B2 Receptor Tyrosine Kinase 2
ERG	ETS-related gene
ETP	Early T-precursor
ETV6	ETS Variant 6
EZH2	Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit
FA	Fractional Abundance
FC	Flowcytometry
FDR	False Discovery Rate
FLT3	Fms Related Tyrosine Kinase 3
FMN2	Formin 2
FOXL2	Forkhead Box L2
FOXP1	Forkhead Box P1
GART	Glycinamide Ribonucleotide Transformylase
GATA3	GATA Binding Protein 3
GATK	Genome Analysis Toolkit
GC	Glucocorticoids
GNAS	Guanine Nucleotide Binding Protein, alpha stimulating
GRIA1	Glutamate Ionotropic Receptor AMPA Type Subunit 1
GWAS	Genome Wide Association Studies
H3K36	Lys36 on histone H3
HIPK1	Homeodomain Interacting Protein Kinase 1
HLF	Hepatic Leukemia Factor

HOXA	Homeobox A
HS	High Sensitivity
ICI	Immune Checkpoint Inhibitors
iDES	Integrated Digital Error Suppression
IG	Immunoglobulin
Ig alpha	Immunoglobulin alpha
IKZF1	IKAROS Family Zinc Finger 1
IKZF2	IKAROS Family Zinc Finger 2
IKZF3	IKAROS Family Zinc Finger 3
IL7R	Interleukin 7 Receptor
INK4A	CDKN2
IPI	International Prognostic Index
IPT	Immunophenotyping
IRB	Institutional Review Board
ITSN1	Intersectin 1
JAK1	Janus Kinase 1
JAK2	Janus Kinase 2
JAK3	Janus Kinase 3
KDM6A	Lysine Demethylase 6A
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
LDH	Lactate Dehydrogenase
LMO1	LIM Domain Only 1
LMO2	LIM Domain Only 2
LRP1B	LDL Receptor Related Protein 1B
LYL1	Lymphoblastic Leukemia Associated Hematopoiesis Regulator 1
MAF	Mutant Allele Fraction
MAGI2	Membrane Associated Guanylate Kinase, WW And PDZ Domain Containing 2
MAPK1	Mitogen-Activated Protein Kinase 1
MDR1	ABCB1, ATP Binding Cassette Subfamily B Member 1
MDS	Myelodysplastic Syndrome
MEF2D	Myocyte Enhancer Factor 2D
MEK	MAP2K1 (Mitogen-activated Protein Kinase Kinase 1)
MLL2	Myeloid/Lymphoid Or Mixed-Lineage Leukemia 2
MM	Mutiple Myeloma
MPS	Massive Parallel Sequencing
MRD	Minimal Residual Disease
MSH6	MutS Homolog 6
mTOR	Mechanistic Target Of Rapamycin Kinase
MTR	5-Methyltetrahydrofolate-Homocysteine Methyltransferase
MTX	Methotrexate
MYRIP	Myosin VIIA And Rab Interacting Protein
ND	Not Detected
NF1	Neurofibromin 1
NFATC2	Nuclear Factor Of Activated T Cells 2
NGS	Next Generation Sequencing



NKX2-1	NK2 Homeobox 1
NOTCH1	Neurogenic Locus Notch Homolog Protein 1
NPAS3	Neuronal PAS Domain Protein 3
NPFFR1	Neuropeptide FF Receptor 1
NRAS	Neuroblastoma RAS Viral (v-ras) Oncogene Homolog
NRK	Nik Related Kinase
NSCLC	Non-Small Cell Lung Cancer
NT5C2	5'-Nucleotidase, Cytosolic II
NTRK3	Neurotrophic Receptor Tyrosine Kinase 3
NUDT15	Nudix Hydrolase 15
PAG1	Phosphoprotein Membrane Anchor with Glycosphingolipid Microdomains 1
PAX5	Paired Box 5
PB	Peripheral Blood
PBPL	Peripheral Blood Plasma
PBX1	Pre-B-Cell Leukemia Homeobox 1
PCR	Polymerase Chain Reaction
PDE4B	Phosphodiesterase 4B
PDGFRB	Platelet Derived Growth Factor Receptor Beta
PER1	Period Circadian Regulator 1
PET	Positron Emission Tomography
PFAS	Phosphoribosyl Formyl Glycinamide Synthase
PFS	Progression Free Survival
PI3K	Phosphatidylinositol 3-kinase
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PIP2	Phosphatidylinositol-5,4-bisphosphate
PIP3	Phosphatidylinositol (3,4,5)-trisphosphate
PIP4K2A	Phosphatidylinositol-5-Phosphate 4-Kinase Type 2 Alpha
PML	Promyelocytic Leukemia
POM121	Pore Membrane Protein Of 121 Kda
POM121C	POM121 transmembrane nucleoporin C
PRPS1	Phosphoribosyl Pyrophosphate Synthetase 1
PTK2B	Protein Tyrosine Kinase 2 Beta
PTPN11	Protein Tyrosine Phosphatase, Non-Receptor Type 11
PYGL	Glycogen Phosphorylase L
RAF	Rapidly Accelerated Fibrosarcoma
RAS	Rat Sarcoma
RB1	Retinoblastoma 1
RET	ret proto-oncogene
RGS6	Regulator of G Protein Signaling 6
RIT2	Ras Like without CAAX 2
ROS1	ROS Proto-Oncogene 1, Receptor Tyrosine Kinase
RQ PCR	Real-Time Quantitative Polymerase Chain Reaction
RTK	Receptor Tyrosine Kinase
RUNX1	Runt Related Transcription Factor 1

SETD2	SET Domain Containing 2
SH2B3	SH2B Adaptor Protein 3
SLCO1B1	Solute Carrier Organic Anion Transporter Family Member 1B1
SMARCA4	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin, Subfamily A, Member 4
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SOX2	SRY-Box 2
SOX5	SRY-Box 5
STAG1	Stromal Antigen 1
STAT	Signal Transducer And Activator Of Transcription
SUZ12	Suppressor of Zeste 12 Homolog
TAL1	TAL BHLH Transcription Factor 1
TAL2	TAL BHLH Transcription Factor 2
TAm-Seq	Tagged-amplicon deep sequencing
TBXAS1	Thromboxane A Synthase 1
TCF3	Transcription Factor 3
TCGA	The Cancer Genome Atlas
TCR	T-Cell Receptor
TET3	Tet Methylcytosine Dioxygenase 3
TKI	Tyrosine Kinase Inhibitor
TLC	Total Leucocytic Count
TLX1	T Cell Leukemia Homeobox 1
TLX3	T Cell Leukemia Homeobox 3
TP53	Tumor Protein P53
TPMT	Thiopurine S-Methyltransferase
TRRAP	Transformation/Transcription Domain Associated Protein
TSLP	Thymic Stromal Lymphopoietin
TYK2	Tyrosine Kinase 2
UID	Unique Molecular Index
ULK2	Unc-51 Like Autophagy Activating Kinase 2
USH2A	Usherin
USP7	Ubiquitin Specific Peptidase 7
VAF	Variant Allele Frequency
WBC	White Blood Count
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WHSC1	Wolf-Hirschhorn syndrome candidate 1
ZFP36L2	ZFP36 Ring Finger Protein Like 2
ZNF384	Zinc Finger Protein 384
ZNF521	Zinc Finger Protein 521
6-MP	6-Mercaptopurine

## CHAPTER 1. INTRODUCTION

### Overview

Acute Lymphoblastic Leukemia (ALL) shows the highest prevalence among children and adolescents in the United States, comprising one fifth of the cancers in persons aged below 20 years. [1] ALL is either B-cell or T-cell, both of which constitute different subtypes, harboring chromosomal alterations as well as, secondary sequence mutations. [2] ALL is polyclonal at diagnosis; most of B-cell of leukemias having multiple subclonal mutations at diagnosis.[3] Chromosomal alterations, which compromise an early initiating event, have been used to risk stratify children with ALL, together with treatment response (i.e. minimal residual disease (MRD) at specified time points)[4, 5] Still, within the favorable risk group, a substantial number of patients, relapse [5]. A number of studies have addressed the change in the clonal dynamics and mutational landscape at relapse using matched used matched diagnosis/remission/relapse samples [3, 6, 7] Mutant allele frequency/fraction (MAF) defined as "the number of mutant reads divided by the total number of reads – coverage – at a specific genomic position" [8] changes from diagnosis to relapse, where in most cases, one minor clone that survived therapy and acquired additional mutations becomes the relapse founder clone [3]. Mutations in different pathways have shown to be enriched at relapse, of which NT5C2 and CREBBP have been shown to confer resistance to 6-mercaptopurine (6-MP) & glucocorticoids, respectively.[6, 9].

It has been postulated that therapy contributes to clonal evolution through either differential sensitivity or through mass extinction [10]. Besides, an increase in subclonal MAF preceded clinical relapse In one study, ultra-deep sequencing of sequential samples (4 patients) during maintenance phase, showed an exponential increase in the percentage of PRSP1 mutation (conferring resistance to 6MP) preceded clinical relapse, hence, indicating that PRSP1 could drive relapse. [11] Additionally, spatial heterogeneity at relapse has been demonstrated; a recent study had shown that in a case with combined (BM and testicular relapse), the testicular relapse represented an independent subclone from the BM relapse. [12]

Despite the advancement in treatment protocols, the currently applied strategies for ALL diagnosis and follow up have two main limitations hindering the detection of clonal diversity as diagnosis and longitudinal follow up of clonal evolution. First, single site BM biopsy at diagnosis lack the ability to capture tumor heterogeneity associated with spatially separated subclones.[13, 14]. Second, flow cytometry (FC) as well as, Polymerase Chain Reaction (PCR)-based methods for disease monitoring fail to detect evolved leukemic clones that were not initially detected at diagnosis [15, 16]. Thus, prohibiting early detection of somatic mutations that confer therapy resistance and consequently, relapse.

In the light of the above findings, the incorporation of a sensitive biomarker capable of capturing temporal and spatial heterogeneity, would have a potential implication in real-time monitoring of the disease and guiding therapy accordingly.

Tumor derived Cell-free DNA or Circulating tumor DNA (Ct-DNA) has been isolated and characterized from the plasma of cancer patients, together with non-tumor derived Cell-free DNA (Cf-DNA). Ct-DNA is 167 bp double stranded DNA fragments released from apoptotic/necrotic tumor cells (12,13). Derived from different malignant subclones, Ct-DNA offers a representation of tumor heterogeneity, consequently, providing a better characterization of disease behavior than does currently used quantitative methods based on the detection of the most dominant diagnostic clone. Several proof-of-concept studies have demonstrated the utility of Ct-DNA in molecular profiling, monitoring of disease response, tracking therapy resistance, as well as, unraveling the heterogeneity of resistance mechanisms. [14, 17, 18]

Capturing heterogeneity at diagnosis, and more importantly, early detection of evolving resistant subclones have a number of therapeutic implications, i.e., the incorporation of newly discovered targeted therapies and small molecules to reverse a phenotype or to use different therapies sequentially to evade evolving clones. (3,10,11).

In this project, we evaluated the utility of detecting somatic mutations in plasma Ct-DNA as a non-invasive technique for molecular profiling at diagnosis, as well as monitoring tumor dynamics throughout therapy. We performed Next-generation sequencing (NGS) of Ct-DNA (whole exome sequencing (WES) or targeted sequencing) from sequential plasma samples withdrawn during remission induction therapy. We then compared the MAF across different time points as well as, between BM biopsy, and plasma samples at diagnosis.

Our preliminary data showed that Whole Exome Sequencing (WES) is capable of detecting the change in predominance of subclonal somatic mutations in Ct-DNA. We employed WES for the first recruited patients; however, data analysis was challenged by the large number of mutations detected, rendering it difficult to discriminate true mutations from false positive ones. Hence, for the next group of patients, we applied targeted sequencing using a specifically designed ALL customized gene panel. Our long-term goal is to incorporate Ct-DNA in the clinical protocols as a minimally invasive biomarker for comprehensive molecular profiling, and longitudinal monitoring of disease response in pediatric ALL.

### **Aim of the Work**

NGS was applied to study somatic mutations in Ct-DNA in peripheral blood (PB)/bone marrow (BM) plasma. Plasma samples were obtained at diagnosis, at specified time points during, and at the end of remission induction therapy. MAF was calculated for all the specified time points. Study specific aims are:

- 1) To test the utility of Ct-DNA in capturing ALL tumor heterogeneity at diagnosis. We compared MAF detected by Ct-DNA analysis of diagnostic PB plasma to MAF detected by clinical WES of diagnostic BM biopsy sample.
- 2) To test the utility of Ct-DNA in monitoring the change in clonal dynamics during remission induction therapy. We compared the results of somatic mutation detection in Ct-DNA to FC based clinical MRD, with a specific focus on patients with negative MRD.

## **CHAPTER 2. LITERATURE REVIEW**

### **Genetics and Pharmacogenomics in ALL**

#### **Introduction**

Acute Lymphoblastic Leukemia (ALL) shows the highest prevalence among children and adolescents in the United States, comprising one fifth of the cancers in persons aged below 20 years. [1] ALL comprise a malignant clonal proliferation of lymphoid progenitor cells of the B-cell or T-cell lineage; B-cell lineage (B-ALL) being the most common.[2, 19]. Recent advances in cytogenetic and molecular techniques have allowed for the unraveling of the genetic basis of leukemogenesis, clonal evolution and relapse. Moreover, it allowed for the study of the prognostic significance of those genetic alterations, thus improving risk stratification and treatment outcome. [20]. ALL, either B-cell or T-cell, constitute different subtypes, harboring chromosomal alterations as well as, secondary sequence mutations. [2]. It has been proposed that the genetic alterations associated with pathogenesis as well as, relapse occur sequentially. Common inherited genetic variants and less frequently highly penetrant germline mutations predispose to ALL [2, 19]. It follows that chromosomal alterations, in the form of aneuploidy, translocations or intrachromosomal re-arrangement occur as an initiating event. Finally, secondary genetic mutations or structural alterations in in genes involved in tumor suppression, cell cycle regulation, apoptosis, lymphoid development and chromatin modification lead to the clinical manifestation of leukemia. [2]

Genetic heterogeneity, plasticity and clonal evolution in response to therapy further complicates the genetic landscape of leukemia, where in most cases relapse emerges from a subclone present at diagnosis or moreover, a pre-leukemic clone. Additionally, pharmacogenomic determinants of toxicity or response affects the outcome. [3, 5, 21-23]

A number of advancements led to the improvement in the outcome of ALL; the use of multi-agent chemotherapy, the use of response-adapted therapy where biological features together with minimal residual disease evaluation are used for patient stratification, as well as the consideration of pharmacodynamics, pharmacokinetics and pharmacogenomic features. Despite the unprecedented improvement in the outcome, relapse still represents a challenge. [5, 24]

#### **Germline Variants Associated with Increased Risk of ALL Development**

Inherited susceptibility to ALL occurs through highly penetrant germline mutations, predisposing to high-risk childhood leukemia in rare familial cases or common susceptibility alleles, associated with lower risk for leukemia development. [23, 25]. During leukemogenesis, interaction occurs between inherited and somatic genetic

alterations, where specific inherited alleles predispose to specific chromosomal-defined subtypes, hence playing a role in disease pathogenesis and prognosis [26] Another synergistic interaction exists between inherited and acquired genetic alterations, where somatic mutations target ALL susceptibility loci [27]

Loss of function of mutations in TP53, associated with Li-Fraumeni syndrome, have been reported in more than 90% of children with low hypodiploid ALL. The mutations included missense, non-sense as well as, indels mostly in a homozygous state because of aneuploidy. The presence of TP53 in non-tumor cells in 43% of patients suggested that the mutation is either inherited or a denovo mutation acquired in germline or hematopoietic compartment. The inherited origin of TP53 was highly suggested due to the presence of family history in 2 patients and its confirmation in another patient through studying non-hematopoietic cells [28].

Mutations of PAX5 show another example of inherited deleterious penetrant mutations in pediatric ALL [23]. PAX5 is a transcription factor encoding B-cell lineage-specific activator protein (BSAP). The transcriptional control of Pro-B cell development occurs through both activation of B-cell specific genes (BLNK, Ig $\alpha$ , CD19) as well as, repression of alternate pathway genes (MPO, Notch1, M-CSFR), thus playing a major role in B-cell commitment and fate. [29, 30]. A number of partner genes are involved; HIPK1, FOXP1, POM121, ELN, AUTS2, JAK2, ETV6, DACH1, PML, ZNF521, BRD1, CHFR, SOX5 and POM121C [27, 31, 32]. In 2013, inherited germline mutations of PAX5 were reported in two kindred of uncommon ancestry. The heterozygous germline variant affected PAX5 octapeptide domain and resulted in decreased transcriptional activity. PAX5 mutations were present together with deletion of 9p and loss of heterozygosity in all the affected children within both families. However, the mutation showed incomplete penetrance. [23, 29]

Other genes associated with increased susceptibility to ALL, among different groups, have been identified by GWAS association studies. The top list includes ARID5B, IKZF1, CEBPE, CDKN2A, PIP4K2A, GATA3 and TP63 [25]. ARID5B (AT rich interaction domain transcription factor), IKZF1 (Transcriptional regulator of the differentiation of lymphoid cells) and CEBPE showed significant association with the risk of ALL, and in addition retained significant association to B-cell ALL. ARID5B was shown to be associated with an increased risk of development of hyperdiploid ALL. [33-35]. CDKN2A encodes both p16 (INK4A) which negatively regulates cyclin-dependent kinases, as well as p14 (ARF1), an activator of p53. rs3731217 which localizes to intron 1 of CDKN2A showed significant association with both B-cell and T-cell ALL. [36]. However, ARID5B, IKZF1, CEBPE, CDKN2A altogether, underscore for 8% of B-cell precursor genetic variance. [37]. PIP4K2 belongs to a family of enzymes which phosphorylates phosphatidylinositol-5-phosphate to form phosphatidylinositol-5,4-bisphosphate (PIP2), a PIP3 precursor. The clustering of four SNPs within PIP4K2A gene intronic region was reported; with rs10828317 showing specific association with hyperdiploid ALL. [38, 39] TP63 (a member of TP53 family) and PTPRJ (Protein Tyrosine Phosphatase, Receptor Type J, involved in the regulation of several cellular processes including cell differentiation, cell growth, mitotic cell cycle as well as,



oncogenic transformation) have been identified as susceptibility genes to ETV6-RUNX1-rearranged pediatric ALL. [35].

Several studies have addressed the association between GATA3 (a transcription factor and putative tumor suppressor gene) and genetic susceptibility to B-cell leukemia. GATA3 rs3824662, which maps to a predicted enhancer, showed association with non-(Tel-AML) positive, non-hyperdiploid ALL. rs3824662 homozygous risk allele showed a significant correlation to older age as well as, worse event free survival [39]. In a discovery GWAS in adolescents and young adult patients, the frequency of GATA3 risk allele was higher in patients aged 16yrs or above, compared to younger patients irrespective of the subtype. [40] In a GWAS of Ph-like ALL, rs3824662 allele showed strong association with Ph-like associated genetic lesions (JAK mutation, CRLF2 rearrangement, and IKZF1 deletion). [26]. Additionally, rs3824662 A allele was associated with an increased risk of relapse, as well as higher level of MRD measured at end of induction. Compared to non-ALL controls, rs3824662 was also overrepresented in non-Ph-like-ALL implying a variant effect on ALL predisposition in general. [26].

### **Recurrent Somatic Genetic Alterations of Precursor B-cell ALL (BCP-ALL)**

Chromosomal alterations include aneuploidy, chromosomal rearrangements, and translocations [19]. Translocations either relocates oncogenes under the effect of regulatory regions of actively transcribed genes; hence leading to the dysregulated expression of the oncogenic protein or fuses two genes resulting in the expression of a chimeric functionally distinct protein. Several subtypes of ALL, lacking a single characterizing chromosomal alteration, have been characterized based on genomic (i.e. unique gene expression profile) or pathologic features. [2] Advancements in NGS and array-based cytogenetic techniques have allowed for the discovery of submicroscopic DNA alterations in genes involved in various cellular processes i.e. normal hematopoiesis, cell cycle regulation, tumor suppression and apoptosis. In addition, it allowed for a deeper insight into the known recurrent abnormalities, as well as the unraveling of gene targets involved in abnormal hematopoiesis and relapse. [20]

**Aneuploidy.** Aneuploidy is further classifier into the following:

**High hyperdiploid ALL.** defined as 51-65 chromosomes, confers good prognosis in 25-30% of BCP pediatric ALL. [20, 41] High hyperdiploid ALL genomic landscape showed involvement of histone modifiers and RTK-RAS pathway. Structural alterations included structural deletions, duplications, translocations, as well as complex rearrangements of a number of genes including ETV6, IKZF1, PAX5 and ADD3. Single nucleotide variations in the form of substitutions and indels were detected in nine genes including *FLT3*, *KRAS*, *NRAS*, *WHSC1*, *CREBBP*, and *PTPN11*, known to be associated with leukemia. [42]



**Hypodiploid ALL.** defined as less or equal 45 chromosomes, represents 6-7% of pediatric ALL. [43] Hypodiploid ALL is further classified into subgroups; namely high hypodiploid (40-43 chromosomes), low-hypodiploid (32-39 chromosomes) and near-haploid (24-31 chromosome) [44]. The first genomic analysis done to characterize hypodiploid leukemia has shown that near-haploid and low-hypodiploid show recognizably different sequence mutations and focal deletions, hence compromising two distinct subtypes. Near haploid ALL, harbored deletion in PAG1 (Csk-binding protein (CBP)), deletion of the lymphoid transcription factor gene *IKZF3*, deletions of a histone cluster at chromosome 6p22, activating genetic alteration (deletion, amplification and/or sequence mutation) affecting the RTK or RAS signaling pathway (of *NF1*, *NRAS*, *KRAS*, *MAPK1*, *FLT3* or *PTPN11* genes), *RBI* alterations. Low-hypodiploid cases, however, showed TP53 sequence mutations (nonsense, missense and indels) in 88.2, *IKZF2* deletion, *RBI* alteration. [28].

**Recurrent translocations of pediatric BCP-ALL.** Translocations arise from either the fusion of two genes or the relocation of an oncogene into a regulatory region of another active gene. This results in the production of a new chimeric protein with a recognizably different function from the joined genes or the dysregulation of expression of an oncogene, respectively. [2]

Common translocations of BCP ALL include ETV6-RUNX1 (TEL-AML1), TCF3-PBX1 (E2A-PBX1), BCR-ABL1 (Philadelphia chromosome), MLL gene rearrangements, IGH translocations and Intrachromosomal amplification of chromosome 21 (iAMP21). Various translocations exist at different frequencies among different age and ethnic groups. [23].

**Recurrent genetic alterations of pediatric T-cell ALL.** T-cell ALL can be subclassified into three groups: (1) Early T-cell precursor (ETP), (2) Near-ETP ALL and (3) Non-ETP ALL. *Detailed characteristics of ETP are discussed in the following section.* T-cell ALL is associated with dysregulation of the expression of transcription factor genes. Such dysregulation can be employed to classify T-cell ALL into 8 groups according to the deregulated gene (TAL1, TAL2, TLX1, TLX3, HOXA, LMO1/LMO2, LMO2/LYL1 and NKX2-1). TAL1 deregulation frequently occurs in both near- and non-ETP ALL; whereas TLX3 is more frequently detected in near-ETP cases, and deregulation of TLX1 deregulation showed to be the most common non-ETP ALL. [45]

**New ALL subtypes.** The newly discovered genetic alterations have a clinical implication on the diagnosis, risk assignment as well as, targeted therapy. [20] New ALL subtypes include Philadelphia chromosome-like ALL, Early T-cell precursor (ETP) ALL, DUX4-, and ERG-Deregulated ALL, and MEF2D, and ZNF384 gene fusions. [19]

**Philadelphia chromosome-like ALL (BCR-ABL1-like ALL.** BCR-ABL1-like ALL occurs at a higher frequency in young adults (27% vs 10% in children). This comprises a group of patients who possess leukemia cell gene expression profile similar to that of Ph-positive ALL, despite the lack of BCR-ABL1 fusion. However, the majority of patients in this group carries genetic alterations activating the kinase signaling

pathways. The most common genetic alterations detections are rearrangements in *ABL1*, *ABL2*, *CRLF2*, *PDGFRB*, *PTK2B*, *TSLP*, *TYK2*, *CSF1R*, *EPOR*, *NTRK3*, *JAK2*, and sequence mutations in *FLT3*, *IL7R*, or *SH2B3*. [23, 27, 46, 47]

**Early T-cell precursor (ETP) ALL.** ETP ALL Early T-cell precursor (ETP) represents 15% of T-cell ALL and is associated with poor outcome. ETP comprises a distinct entity of leukemia, with a characteristic pattern of gene expression different from human early T-cell precursor cells. ETP ALL expresses myeloid or stem cell markers. Additionally, the gene expression profile of ETP ALL showed enrichment for signatures associated with poor outcome ALL (leukemic stem-cell signature, IKZF1-mutated high-risk B-progenitor ALL signature). [19, 48, 49]. ETP ALL commonly demonstrates dysregulation in the expression of transcription factor genes LMO2/LYL1 [45]

Genes frequently mutated in ETP ALL comprise RAS signaling pathways, cytokine receptor regulation, epigenetic regulation and hematopoietic development. Activating mutations in genes involved in RAS signaling and cytokine receptor regulation (KRAS, NRAS, IL7R, FLT3, BRAF, JAK1, JAK3 and SH2B3) exist at the highest frequency. Genes harboring mutations that inactivate hematopoietic development include ETV6, RUNX1, IKZF, GATA3, and EP300 whereas histone modifying genes affected include SETD2, EZH2, EED, SUZ12, and EP300. [19, 48]

### **Pathways Commonly Mutated in Pediatric ALL**

**Pathways commonly mutated in B-cell precursor ALL at diagnosis.** In a study of clonal evolution of pediatric B-cell leukemia from diagnosis to relapse, high frequency of mutations was detected in six pathways at diagnosis and/or relapse. Different mutation frequency existed among different pathways; epigenetic modification (65%), Ras signaling (65%), JAK-STAT signaling, (25%), transcriptional regulation of lymphoid development (85%), nucleoside metabolism (45%), and cell cycle regulation (60%). In addition, NT5C2, CREBBP, WHSC1, TP53, USH2A, NRAS and IKZF1 mutations have shown to be enriched at relapse [3].

**Pathways commonly mutated in T-cell ALL at diagnosis.** In T-cell, ten pathways were recurrently mutated. As is the case with BCP, transcriptional regulation showed the highest mutation frequency (91%). Other pathways that showed recurrent mutations were cell cycle regulation and tumor suppression in 84% of the studied case, NOTCH1 signaling in 79%, epigenetic modification in 68% and to lesser extent JAK-STAT signaling (25%), PI3K-AKT-mTOR signaling (29%), ribosomal function (13%), RNA processing (9%) whereas RAS signaling was only mutated in 14% of the cases. [45]

## Somatic Mutations Associated with Relapse and Therapy Resistance

Mutation frequency has been shown to vary among different subtypes in relation to enrichment at diagnosis or relapse. Moreover, a number of studies have unraveled the association between specific mutations and therapy resistance. Importantly, the cytotoxic activity of novel small molecules in mutant cells has been shown, holding promise for future role in the clinic. Moreover, reversal of the resistant phenotype and regaining of sensitivity was demonstrated. [11, 50]

Several studies have addressed somatic mutations in epigenetic regulators in high risk groups of leukemia, however, the reported frequency varies in different subtypes of *denovo* ALL. 60% of hypodiploid cases carried somatic mutations in genes encoding epigenetic regulators. In contrast to other ALL subtypes where CREBPP is more enriched in relapse, mutation in CREBPP was the most common epigenetic regulator mutation reported in *denovo* hypodiploid ALL. *CREBBP* encodes a histone/non-histone acetyltransferases which plays a role in regulation of glucocorticoid gene expression, its mutation contributes to prednisolone/dexamethasone (glucocorticoid) resistance [9, 51]; [28] *CREBBP* mutations were also prevalent among hyperdiploid ALL relapsed cases. [52]. Missense mutations were also detected in *EZF2* gene; with 2 cases harboring *EZF2* mutations only at relapse [3, 48].

Matched diagnosis/remission/relapse samples of pediatric patients with B-cell ALL showed enrichment of mutations in genes coding epigenetic regulators at relapse, with 57% of the patients harboring epigenetic regulator mutations at relapse. Frequently mutated epigenetic regulators included *MLL2*, *KDM6A* and *MSH6*, as well as additional mutations in *SETD2* and *CREBBP*. The high frequency of mutations detected in relapsed cases, as well as *denovo* high risk ALL, plus the different mechanisms by which *SETD2* and *MSH6* can promote clonal survival, explains a possible role in chemotherapy resistance. The importance of such findings lies in the possible future role of epigenetic therapy aiming at inhibiting the enzymatic activity of the opposing regulators [53]. Another study also showed a relapse mutation rate in epigenetic regulator genes of 65% compared to 45% at diagnosis; with relapse-specific mutations found in *WHSC1* and *CREBBP*; with *CREBBP* showing multiclonal mutations in one case [3].

In a group of patients with ETP, 48% showed activating mutations in histone modifying genes, namely *EZH2*, *EED*, *SUZ12*, *SETD2* and *EP300* [28, 48]. Also, Mutations in *WHSC1*, *SUZ12*, *SMARCA4*, *ARID4B* and *USP7* were reported as relapse specific mutations in T-cell ALL pediatric patients. A finding that highlights the importance of somatic mutations in epigenetic regulators is the detection of E1099K activating mutation in the H3K36 histone methyl transferase (*WHSC1*) as the only mutational difference between primary leukemia and relapse in a patient who suffered early relapse [7].

Several studies have reported RAS mutations to be subclonal, with a heterogeneous pattern of clonal evolution; where associated RAS mutations can be gained, lost or even retained from diagnosis to relapse. [3, 54] In BCP ALL, RAS

mutations (KRAS, NRAS, FLT3, PTPN11 were highly prevalent (Incidence 37.9%) at relapse, and more importantly, demonstrated significant association with early relapse (KRAS/NRAS), as well as CNS involvement [50] The role of RAS-MAPK signaling in drug resistance has been studied both in human cell lines as well as, mouse primary isogenic leukemic lymphoblasts. Functional analysis of RAS mutations corroborates a role of this signaling pathway in methotrexate resistance; as well as increased sensitivity to vincristine. Thus, positive or negative selection of RAS mutant clones can occur at relapse. [54].

In vitro functional analysis as well as, genetic study of T-cell ALL reported an association between IL-7receptor signaling components (including KRAS and JAK-1) and poor prednisone response and clinical outcome. [55]. An important pharmacogenomic implication to that association, is the finding that combining ruxolitinib, a JAK1/2 inhibitor, and dexamethasone may enhance the efficacy of glucocorticoids (GC) in a subset of T-cell ALL, with IL-7 dependent GC resistance. [56]

Multiple relapse-specific activating mutations in NT5C2, conferring resistance to purine analogues (6-MP and TG), have been discovered in both BCP and T-cell. NT5C2 encodes a hydrolase which metabolizes and inactivates active metabolites of 6-MP, hence affecting 6-MP cytotoxicity. Several groups demonstrated NT5C2 to be exclusively present in relapse samples. However, the frequency of NT5C2 varied across different studies from 3%-45% in BCP [3, 57] . Ma et al. detected NT5C2 at 45% in relapse cases, with 3 cases showing multiclonal mutations whereas, in 4 out of 14 cases, mutations at relapse were subclonal from the relapse founder clone. [3]. NT5C2 demonstrated a significant association with early relapse (within 36 months of diagnosis). It was detected at a frequency of 16% in early relapsing patients and 100% of Individuals carrying NT5C2 mutations encountered early relapse [6].

In T-cell pediatric patients, NT5C2 was among the most common relapse mutations, detected at a frequency 19%-38.5% across different studies [7, 57]. As is the case in BCP, NT5C2 could be present as clonal or subclonal mutation at relapse, hence representing a late event in relapse. [7]

Mutually exclusive to NT5C2, multiple activating PRPS1 mutations have been identified as relapse specific in B-cell pediatric ALL. PRSP1 mutants modulate thiopurine activity indirectly through reduced feedback inhibition of de novo purine synthesis and hence, competitive inhibition of thiopurine activation. Mutant PRSP1 conferred early relapse in all patients who carried the mutation. Other relapse specific mutations in the denovo purine synthesis pathway have been detected; involving PRPS2, ATIC, ADSL, GART and PFAS. [11]

## **Tumor Heterogeneity, Clonal Evolution and Therapy Resistance**

The notion that tumors are formed of evolving clones predicts specific characteristics: the presence of clonal genotypes, the temporal change in clonal dynamics,

spatial variation across different sites, emergence of drug resistant clones, partial response to drug therapy, as well as relapse originating from minor subclones. [3, 58, 59]

Generally, ALL is polyclonal at diagnosis i.e. multiple genetically distinct clones, sharing the common clonal origin are present at diagnosis. It has been shown that 50% of the BCP ALL cases leukemias exhibit multiple subclonal mutations at diagnosis. Different subclones show various response to the selective pressure imposed by therapy [2, 3, 60]. It has been assumed that therapy exerts its effect on clonal evolution via one of three. The first model “Differential Sensitivity” model, in which active selection of therapy resistant clones takes place, resulting in convergent evolution, in which the drug resistant mutation become the predominant clone at relapse. Differential sensitivity model may show specific relevance to targeted therapies directed against genetic alterations not shared across various clones. [10]. The second “Mass Extinction and Competitive Release” model, assumes equal therapy sensitivity across various subclones, as evident by comparable proportion pre-and post cytoreduction. In this second model, the evolutionary landscape is continuously altered by the clearing niche, causing the best fitting clone to expand. [10, 61]. The third model is “Chemotherapy Induced Mutagenesis”, where novel mutations are induced by genotoxic chemotherapy [10]. The last model is less likely to explain clonal evolution in pediatric ALL as current treatment protocols limit the use of mutagenic agents and there is no shift in mutation burden nor an increase in the burden of sequence mutations. [3]

A recent study of the mutational profiling in two ALL kids with testicular relapse has shown that in one case (combined BM and testicular relapse), the relapse clone evolved from an ancestral clone. Importantly, the VAF analysis showed that although BM and testicular leukemia originate from the ancestral clone, yet the testicular relapse represents an independent subclone (that shares a fraction of mutations with the diagnostic clone which were absent in the marrow). [12]

In most cases, a minor subclone at diagnosis that either harbors or acquires additional mutations conferring therapy resistance or else from ancestral clones (pre-malignant clone) prior to the dominant diagnostic clone, survives therapy and expands to be the relapse founder clone. To a lesser extent, predominant relapse clones are identical to or genetically distinct from the diagnostic clones representing a secondary leukemia. [2, 3, 5, 60]. Hence, a better understanding of clonal evolution, drug resistance mechanisms is required, in order to tailor therapy in a way that counters clonal evolution, consequently improving the outcome. [58].

### **Clinical Implications of Monitoring Clonal Dynamics and Drug Resistance**

Backtracking of serial BM samples in relapsing ALL patients harboring PRSP1 mutation, conferring resistance to purine analogues, showed an exponential increase in the percentage of PRSP1 mutation, prior to clinical relapse. NT5C2 was also detected in a remission prior to clinical relapse. [11]



The detection of somatic mutations has a great role in advancing precision medicine. Ruxolitinib, a JAK1 inhibitor, has been incorporated in a recent clinical trial for patients harboring genetic alterations (mutations, deletions or re-arrangements) affecting the JAK/STAT signaling (e.g., *EPOR*, *IL7R*, *JAK1*, *JAK2*, *JAK3* or *SH2B3*). (unpublished data) Studies on small molecules also hold a great promise in combating resistance associated with specific pathways. Selumetinib, an allosteric selective 1/2 MEK-inhibitor, demonstrated significant in vitro cytotoxicity against RAS-mutant ALL cells harboring KRAS, G12D and G13D or NRAS, Q61R. Moreover, it showed significant decrease in the level of circulating blasts in mice xenografted with RAS-mutant primary ALL cells harboring KRAS, G12D or NRAS, Q61R; as well as significantly decreased the CNS disease burden. MEK inhibitors may carry a therapeutic role in the future in patients harboring RAS mutations. [50] Lometrexol, GART inhibitor, provides another example; it successfully inhibited denovo purine synthesis resulting in reversal of thiopurine resistance in PRPS1 mutant cells[11].

Not only small molecules show promising activity, conventional chemotherapy might still play a role in patients harboring mutations to as specific drug; i.e. in patients with NT5C2 mutations, the mutation didn't confer resistance to Nelarabine, used as second line therapy in T-cell ALL. (T-cell). [57]

## Pharmacogenomics of ALL

In addition to the aforementioned genetic alterations; several germline variants play a role as pharmacogenomic determinants of toxicity or drug response affecting the treatment outcome.

**Germline variants associated with poor outcome.** In a Genome-wide association (GWAS) studying the relationship between genotypes at single nucleotide polymorphisms and risk of relapse, 134 SNPs were found to be associated with relapse. 133 SNPs showed to be prognostic after adjusting for other risk factors. Top 25 SNPs associated with ALL relapse involved the following genes PYGL, MYRIP, LOC642340, PDE4B, TBXAS1, MAGI2, RIT2, NPFFR1, ATP8A2, PDE4B, ANTXR1, NPAS3, ATP8A2, STAG1, ANKS1B, ATP8A2 (validation frequency ranging between 37%-79%). [62]

One important finding of the study is the significant association between relapse-associated SNPs and pharmacokinetics and pharmacodynamics of 3 antileukemic drugs (L-asparaginase, dexamethasone and methotrexate) as follows: 2 relapse specific SNPs within FCHSD1 gene showed association with high L-asparaginase antibody levels and increased risk of relapse, 4 SNPs (2 in ABCB1 gene) showed an association with higher dexamethasone clearance, 5 SNPs (3 SNPs within the PDE4B gene) showed association with lower intracellular methotrexate polyglutamate accumulation, as well as 3 SNPs associated with higher MTX clearance. The role ABCB1 (MDR1) in the transport of other chemotherapeutic agents (glucocorticoids, anthracyclines and vincristine), raises the

possibility that the association between ABCB1 SNP and outcome is due to unfavorable host pharmacokinetic properties as well as, intrinsic leukemic blast resistance [62]

### **Germline SNPs associated with increased toxicity to antileukemic therapy**

***SNPs associated with Vincristine toxicity.*** Several studies evaluated the association of SNPs to Vincristine peripheral neuropathy. A CEP72 variant is associated with increased vincristine neuropathy during maintenance therapy in pediatric ALL patients treated with 1.5 mg/m<sup>2</sup> vincristine. The mutant protein (T allele at rs924607) increases sensitivity to vincristine through binding a transcription repressor, leading to lower level of CEP72 expression. Patients homozygous for the risk allele showed a significant increase in the cumulative incidence of Grade 3-4 neuropathy. [63]. CEP72 polymorphism was not related to the incidence of neuropathy during or by the end of remission induction in the same patient cohort. [64]. Polymorphisms in drug transporter ABCC2 may play a role in vincristine neuropathy during induction, where 3 SNPs were shown to be associated with neurotoxicity in early phases of treatment in the primary analysis yet didn't reach statistical significance after FDR correction. [65]. A higher risk of high grade neurotoxicity was associated with one SNP in the ACTG1 gene, which encodes a major cytoskeletal protein ACTG. 2 SNPs in CAPG and ABCB1 genes were shown to be protective against low grade neurotoxicity. [66]

***SNPs associated with 6-Mercaptopurine toxicity.*** The relationship between TPMT genetic variation and thiopurine toxicities has been well established and preemptive TPMT genotyping is used clinically to guide MP dosing during ALL therapy [67]. Despite the reduction in toxicity brought about by preemptive *TPMT*-guided dosing, clinically significant thiopurine-related myelosuppression is still observed in patients with so-called normal *TPMT* activity. [68]

Several recent studies addressed the association between genetic variants other than TPMT and 6MP toxicities. In the discovery GWAS of 657 children with ALL in the AALL03N1 cohort, a germline variant in NUDT15 (rs116855232) strongly associated with MP dose intensity in childhood ALL ( $P = 8.8 \times 10^{-9}$ ), where the number of copies of the T allele at the NUDT15 SNP rs116855232 showed a significant correlation with tolerated MP dose, indicating a gene dosage effect. Patients heterozygous for *NUDT15* variants (TC genotype) tolerate 63% of a standard 6-MP dose and *NUDT15*-deficient patients (TT genotype) tolerated only 8% of a standard dose. Yang et al. also showed that rs116855232 genotype distribution varied substantially by race/ethnicity, being more frequent in East Asians (9.8%). Consistent with these findings Tanaka et al. 2015 showed a statistically significant NUDT15 association with grade 3-4 leucopenia development within 60 days of the initiation of maintenance therapy (early leucopenia) in a cohort of 92 Japanese children with acute lymphoblastic leukemia. Although both TPMT and NUDT15 are associated with susceptibility to 6-MP toxicity, different mechanisms are hypothesized. [68] For TPMT, the risk variants affect metabolism leading to accumulation of TGNs, the active metabolite of 6-MP, in hematopoietic tissues while for NUDT15, susceptibility to toxicity involves DNA repair and apoptotic pathways.[69]

Mutation in ABCC4 (rs3765534), a drug efflux transporter, showed significant association with severe neutropenia in Japanese patients with childhood leukemia. The combined effect of ABCC4 and heterozygous NUDT15 variants was similar to that conferred by loss of NUDT15 activity. [70]

***SNPs associated with L-Asparaginase toxicity.*** A recent GWAS study detected an association between germline mutations in ULK2, a gene involved in autophagy as well as, in RGS6 gene (a negative regulator of G-protein signaling) and asparaginase associated pancreatitis (AAP). In patients <10 years, heterozygous for both risk alleles (ULK2 rs281366 and RGS6 rs17179470) showed AAP cumulative incidence of 85%, compared to 22.5% in patients harboring only one risk allele, whereas the incidence in patients without no risk allele was 4.1%. [71]

Few studies addressed associations between SNPs and L-asparaginase hypersensitivity. Asparaginase allergy was significantly associated with GRIA1 variants (rs10070447, rs6890057, rs4958676, and rs6889909) [72], HLA-DRB1\*07:01 allele (45% higher incidence of L-asparaginase hypersensitivity) [73] and intronic polymorphism in NFATC2 (rs6021191). It has also been shown that patients carrying both NFATC2 rs6021191 and HLA-DRB1rs17885382 variants are at higher risk of developing asparaginase hypersensitivity compared with patients with single or no risk alleles [74]

***SNPs associated with methotrexate toxicity.*** The most common toxicities following MTX administration are nephrotoxicity, hepatotoxicity and myelosuppression. Several genetic polymorphisms showed association with a specific MTX toxicity. Polymorphisms of the ARID5B gene (rs4948502, rs4948496, rs4948487) were associated with MTX and 7-OH-MTX serum levels, and toxicity. SLC19A1 rs7499 was significantly associated with hepatotoxicity; yet SLC9A1 rs1051266 didn't show correlation with toxicity. Granulocytopenia post 2gm/m<sup>2</sup> MTX dose was associated with MTR rs3768142 [75]. SLCO1B1 also shows association MTX pharmacokinetics. [76]

## **Monitoring Disease Response in Pediatric Leukemia**

### **Introduction**

In leukemia patients, minimal residual disease (MRD) represents leukemic cells undetectable by morphologic examination of bone marrow smears. Acute lymphoblastic leukemia (ALL) has been at the head of the development of MRD methods.[77] Testing of MRD has a great impact on the advancement of ALL treatment; it has shaped the definition of remission, proved to be the most important prognostic factor for treatment outcome and it measure during induction , as well as at specific points throughout therapy, has been used to risk-stratify the patients according to treatment response [4, 15]



## Detection of Minimal Residual Disease (MRD) in Pediatric ALL

**Sampling.** An important point to consider, when looking into the reliability of MRD measurement in representing the burden of the disease, is that studies are performed on a single small bone marrow aliquot, assuming heterogeneity throughout the marrow. In fact, it has been shown as early as 1966 that heterogeneity is associated with the distribution of leukemia, where 38% of patients who were morphologically in remission by peripheral blood and single bone marrow samples had persistent infiltration in different site. [78, 79]

**Analytical platforms for MRD measurement.** Currently applied methods for measuring MRD in lymphoblastic leukemia are based on specific molecular and cellular characteristics [80] Each of the mentioned techniques is associated with advantages and limitations.

**Flow Cytometry (FC).** Leukemic blasts differ immunophenotypically from normal cells, as they express abnormal cell markers, hence flow cytometry can be used to detect leukemic lymphoblasts (based on the cell markers) with a limit of detection of 0.01% (1leukemic cell in 10,000 normal cells). [79, 81] The advantages of Flow cytometry are the accuracy of detection, ability to detect the hematopoietic cell maturation, easy storage of data and less time and labor needed compared to PCR. [15, 82]

However, a number of limitations do exist. Flow cytometry is based on the phenotypic markers of the diagnostic dominant clone [16], the sensitivity is compromised by the difficulty inherited in the differentiation between ALL clones and regenerating blasts and importantly, it lacks the ability to detect emerging clones, that differ phenotypically from the diagnostic clone i.e. it doesn't take into consideration the phenotype alteration associated with therapy derived clonal evolution. [15, 16, 82]

**Polymerase Chain Reaction (PCR).** PCR is applied to detect either rearrangement of T-cell receptor (TCR) and immunoglobulin (IG), which can be measured at a sensitivity of 0.001% by PCR. [83] the presence of chromosomal alterations (PCR amplification of fusion transcripts) [84]

Detection of TCR and IG re-arrangement by PCR shows high sensitivity and applicability in 90-95% of the patients. Moreover, in one study, it was shown to better predict T-cell leukemia patients with high risk of relapse than does flow cytometry[85]. As is the case with flow cytometry, the inability to detect subclones (due to the use of specific primer sets for the dominant clone) represents a major drawback. [15, 16] Additionally, PCR is associated with more time and labor due to the design of patient specific primer sets.[79]

A main advantage of the fusion transcripts amplification is that gene fusions, being early initiating events, exist in stable association with the leukemic clone, regardless of clonal selection. On the other side, drawbacks are the inaccurate

quantification due to intra-patient variability in the number of fusion transcripts per leukemic cell as well as, interpatient variability throughout treatment duration, and existence in less than half of leukemia patients. [3]

***Next Generation Sequencing (NGS).*** Next generation sequencing is based on the use of consensus primers for parallel rapid sequencing. NGS overcomes a number of disadvantages associated with PCR. NGS is associated with higher sensitivity compared to PCR (up to  $10^6$ ), the use of consensus primer rather than patient specific primers decreased the time and labor associated with the test, the ability to quantitate clonal evolution, hence allowing a deeper insight into the biology of leukemia as well as, the kinetics of relapse. The challenges associated with NGS is the need for the presence of high frequency index clone in the initial diagnostic sample and the expertise needed to use complex bioinformatics approaches in the analysis. [15, 16, 86, 87]

## **Somatic Mutation Detection in Leukemia-derived Circulating DNA**

### **Introduction**

Tumor heterogeneity, spatial and temporal, has been documented in cancer patients including pediatric ALL [3, 7, 12]. Therapy exerts a selective pressure leading to clonal evolution and therapy resistance [6, 10, 12]. Hence, one of the greatest challenge, for precision medicine, is to overcome tumor heterogeneity [88], consequently improving the outcome.

Comprehensive characterization of the disease has been limited by sampling and technical issues incorporated in disease follow-up. Limitations associated with sampling include: single site biopsies lack the ability to capture heterogeneity among spatially separated malignant clones [14], the invasiveness of the procedures used limits the ability to obtain serial samples hence, the detection of the change mutational landscape throughout treatment [18], the sufficiency of the sample to perform multiple assays [88] and quality of the specimens, i.e. BM hypoplasia or fibrosis can lead to low quality samples in MDS patients [17]. The criteria used to evaluate the disease, lack the ability to detect the change in the dynamics of the evolving malignant clones [17].

Ct-DNA analysis serves a minimally invasive technique, which can be repeated at multiple time points, comprehensively characterize contributions from heterogeneous clones [88]. and capture the differential response of different disease subclones throughout therapy and predict disease progression [17]. Moreover, the discovery of actionable mutations in Ct-DNA has potential important implications in the advancement of precision medicine. [14]

## Characteristics of Cell-free and Circulating-tumor DNA

As early as 1948, Mendel and Mètais were the first to describe cell-free DNA (Cf-DNA); fragments of DNA circulating freely in plasma. [89-91]. The detection of circulating DNA in the plasma of cancer patients occurred later, in 1987, when M. Stroun et al. isolated an extractable quantity of DNA from 27% of studied cancer patients as compared to none in normal controls, as well as, confirmed the human origin of the isolated DNA.[92]

Using murine models xenografted with human derived tumors, human derived circulating DNA was detected in the plasma of mice, along with an increased level of mouse derived DNA, demonstrating that CF-DNA is derived from both normal and tumor cells. [93]. The exact mechanism of release of circulating DNA into the plasma of cancer patients requires further elucidation. [94] Several mechanisms have been demonstrated mainly; apoptosis, necrosis, in addition to active secretion. [95] Apoptosis leads to the generation of fragments with size 180 bp or multiples, whereas necrosis produces much larger fragments. It has, hence, been suggested that different sizes of DNA fragments can be applied to distinguish between apoptosis and necrosis as sources of circulating DNA [96] However, studies demonstrated that 90% out of the total circulating DNA exist in a size range of 150-180bp [89] with a significant peak at 166 bp, representing a mono-nucleosome DNA length (147 bp) plus the length of the Histone H1 linker DNA (20bp) [97]

The kinetic properties of Cf-DNA clearance have been derived from the studying of fetal Cf-DNA clearance from maternal plasma. The clearance of Cf-DNA occurs in two phases [98, 99], an initial rapid phase (mean half-life 1 hr.) corresponding to distribution/elimination followed by a second slower phase (mean half-life 13 hr.) corresponding to total elimination. [98, 100] It was also shown that several mechanisms contribute to clearance; with DNase and trans-renal eliminations playing a minor role. [97, 98, 101]

It has been demonstrated that the change in circulating DNA level is correlated to treatment response after radiation therapy i.e. persistence of circulating DNA is associated with a poor treatment response whereas a decrease in the level of circulating DNA is associated with radiological signs of remission in different solid tumors. [102] Years later, it was shown, through the advancement of molecular techniques, that circulating nucleic acid displays a wide range of tumor-associated genetic alterations i.e. chromosomal alterations (Inversions & deletions), microsatellite alterations, point mutations, microsatellite alterations as well as, promotor hypermethylation. [10, 103] These finding, opened the door for studying the utility of Ct-DNA analysis in a wide range of applications in various tumor types.

## Pre-analytical Considerations Affecting Cf-DNA Analysis

The variability in the levels of Cf-DNA detected across different studies can be attributed not only to the biological variability between patients, but also to the origins of contaminating DNA as well as, the sensitivity of the analysis methods. Hence, it is crucial to consider the pre-analytical factors that affect the stability and yield of Cf-DNA as well as, those that minimize contamination from other sources of DNA (i.e. genomic DNA). [89]

**Blood sampling.** Plasma was shown to be a better matrix compared to serum for studying Cf-DNA. It has been shown that the increased concentration of Cf-DNA in serum occurs due to clotting of WBCs and release of genomic DNA which contributes to the higher concentration of Cf-DNA in serum. [104] It has also been demonstrated that plasma is a better source for study tumor derived Cf-DNA. In mice xenografted with CRC cells, Thierry et al. were able to show that a higher concentration of Cf-DNA was detected in serum, as compared to plasma, when using murine specific KRAS wild-type primers. On the contrary, using human wild-type KRAS primers (targeting tumor derived Cf-DNA), the concentration measured was higher in plasma; indicating that plasma serves as a more reliable source for studying tumor derived Cf-DNA. [100, 105]

**Anticoagulants.** Lam et al. studied the effect of different anticoagulants on the concentration of Cf-DNA using quantitative real-time PCR. They showed that no significant difference in the concentration of Cf-DNA detected using heparin, citrate and EDTA as an anticoagulant as long as when plasma separation is done within 6 hr. from blood collection ( $P > 0.05$ , Friedman test). However, after 24 hr., EDTA produced higher Cf-DNA concentration, compared to citrate or heparin. Hence, the authors recommend EDTA as the anticoagulant of choice if samples are to be stored for 24 hr. before analysis and considers all the three anticoagulants as acceptable options. [100, 106]

On the other hand, others recommend against the use of heparin as it might inhibit the polymerase chain reaction of downstream applications. [107] This later result was confirmed by Ginkel et al., reporting that the concentration of Cf-DNA concentrations can remain stable in EDTA plasma over 24 hr. until fractionation. [108] Another study was able to detect a significantly higher concentration in EDTA compared to heparin at 0 and 3 hr. ( $P < 0.05$ ). [109]

**Handling of blood samples.** The length of time and temperature of storage of plasma has been studied. Lam et al. showed that the total concentration of Cf-DNA increases after 24 hr. compared to the concentration at 0hr and 6 hr, irrespective of the anticoagulant used, likely due to apoptosis or necrosis of leucocytes. [106]

**Cf-DNA quantification.** Fluorospectroscopy and fluormetry quantification methods are used to measure isolated Cf-DNA. Mauger et al. studied the correlation between the Cf-DNA quantification carried by Nanodrop, Qubit to ddPCR measurements. They were able to demonstrate a significant strong correlation between the measurement obtained from Qubit show a significant strong correlation to the

measurements obtained by ddPCR. On the contrary, Nanodrop measurements lacked correlation with ddPCR measurements. Hence, the authors recommend the use of low-cost Qubit quantification, to confirm that sufficient amount of DNA is isolated, prior to proceeding to PCR analysis. [109]

## **Technological Platforms Used in Circulating-tumor DNA (Ct-DNA) Analysis**

**Introduction.** Advancements in the molecular techniques and sequencing methods have allowed for an upsurge of studies of a wide range of applications of tumor derived Cf-DNA in various adult and pediatric cancers. [14, 17, 18]. Analysis of tumor derived Cf-DNA allows for the detection of the entire spectrum of tumor genetic alterations. The analysis of tumor derived Cf-DNA, however, is complicated by the low amount present in plasma (especially with low disease burden), the presence of non-tumor derived CF-DNA in the plasma, the high degradation and additionally the tumor heterogeneity. Hence, sensitive detection techniques and analysis pipelines are needed. [89]

The detection of genetic alterations in Ct-DNA implies two main approaches; methods with a capability to detect all possible alteration and those targeting specific changes. [89]

**Targeted approaches.** The selection of targeted approaches for tumor derived CF-DNA analysis requires previous knowledge of hotspot mutations (specific positions at which mutations concentrate) and genetic alterations (indels and mutations). [89]

**Digital Droplet PCR.** Digital PCR, first described by Vogelstein et al., is based on mutation analysis by fluorescent probes. [110] The technique involves pairing digital PCR to Nanoliter-sized droplet technology where the PCR reaction is partitioned into 10,000 to 20,000 independent polymerase reactions per tube, eliminating the error associated with reverse transcription and normalization. [111] ddPCR is associated with higher precision and less variability when compared to real-time PCR. [112] The technique is associated with increased sensitivity, limit of detection as low as 1 in 10,000 copies (0.01%). [111], ddPCR can be applied in SNV, CNV analysis, rare variants detections and quantification of transcripts. [113] It has been implied in the analysis of Ct-DNA in a wide range of cancers as well as, for the validation of other methods used for Ct-DNA analysis [14, 18]

**BEAMing.** The workflow involves multiple steps: The workflow involves multiple steps:

- 1) Coupling biotinylated oligonucleotides to streptavidin beads.
- 2) Formation of microemulsions (an aqueous mix of PCR reagents, primer-bound
- 3) beads. and DNA template mixed together with an oil/detergent mix).
- 4) Conventional PCR.
- 5) Breaking the emulsion, purification of the beads using a magnet.

- 6) Incubation of the beads with different oligonucleotides.
- 7) FC to differentiate between different fluorescent beads.[114]

The technique has not been widely applied due to high cost, laborious workflow and low throughput. [113]

***Clinical utility of BEAMING and ddPCR.*** In Colorectal Cancer patients, CF Tumor DNA has been used to identify actionable drug targets, track tumor dynamics throughout therapy as well as, monitor resistance to EGFR-blocking agents. Oncogenic mutations in RAS, NRAS and BRAF detect the lack of response to EGFR-blockade. By studying the mutations in the RAS pathway in matched tissue and blood samples, Ct-DNA showed 97% concordant results. Additionally, it was able to capture mutations not detected in tissue samples in 8% of the patients, confirming its postulated capability to capture tumor heterogeneity. In a patient cohort refractory to EGFR, NGS has also allowed for the identification of somatic mutations in p53 and APC (known to be commonly mutated in CRC), somatic mutations in MEK gene, in addition to successfully detecting amplifications in ERBB2 and FLT3.

Ct-DNA successfully detected mutations associated with acquired (secondary) resistance to anti-EGFR antibodies (panitumumab or cetuximab) in colorectal cancer patients who progressed on treatment, Ct-DNA analysis was able to detect the presence of KRAS mutations in 73% of the patients, EGFR ectodomain mutations as well as, KRAS and MEK amplifications.

Longitudinal analysis of Ct-DNA in 5 the patients with KRAS mutations showed that the fractional abundance of KRAS mutations emerging at time of relapse, decreases upon suspension of the first line drug and remain below the level of detection with subsequent treatment.

Moreover, the study addressed the change in clonal evolution in a limited number of patients (3 patients) who received multiple rounds of anti-EGFR antibodies. The change in percentage of KRAS mutated alleles was carried in tissue or plasma samples. The KRAS clones decline upon suspension of EGFR blockage and increase after re-challenge with anti-EGFR, despite partial response to the second line EGFR, showed the plasticity of CRC treated with EGFR blockade. This plasticity clearly shows the implication longitudinal real-time monitoring would have on treatment decisions which would otherwise depend an outdated molecular profile. It also supports the approach of adaptive therapy. The authors recommend that each new round of therapy should be selected based on a recent comprehensive molecular profile of the disease. Hence, the plasma-based detection of Cf Tumor DNA may provide a more comprehensive interrogation of the mutational status of the RAS pathway than does the tissue-based molecular studies. [18]

Hence, Ct-DNA has a potential role in monitoring both intrinsic resistance due to primary tumor heterogeneity, as well as to dynamically tracking the acquisition of



secondary resistance to therapy through the selection of additional molecular alterations. [18]

**Massive Parallel Sequencing (MPS) approaches.** The study of tumor heterogeneity and evolution has been limited by the number of genetic lesions studied and/or the sensitivity of the techniques applied. [10] The advent of MPS, and associated improvement in sequence quality, read length and throughput have allowed it to provide a reliable method for Ct-DNA quantification. Although some limitations do exist; namely the efficiency of capturing the regions of interest and the high error rate of sequencing, different approaches to overcome associated limitations and decrease the error have been successfully applied. [89]

Techniques for quantifying Ct-DNA by MPS are classified into two main approaches, PCR-based on the amplification of the regions of interest using highly multiplexed specific primers and hybridization capture approaches in which to oligonucleotide probes complementary to the region of interest are used for capture followed by library amplification. Finally, deep sequencing is carried out to allow accurate quantitation of the relative amount of both wild type and mutant DNA at every specific locus. [89] The detection limit of Ct-DNA by ultrasensitive profiling is challenged by Cf-DNA recovery and errors introduced during library preparation and sequencing. [115]

***Hybridization-capture based techniques.*** This approach involves barcoded library construction, followed by hybridization of the library(s) to oligonucleotide probes complementary to the region of interest. This is followed by isolation of the captured molecules (i.e. using streptavidin bead) then amplification using primer sets complementary to the library barcodes. [89]

Hybrid capture techniques comprise either on-array capture or in-solution capture, both being based on the use of oligonucleotide probes complementary to the regions of interest of the barcoded library. The main difference, however, is the amount of DNA library and oligonucleotide probes. On-array capture use excess amount of library to oligonucleotide probes, hence requiring a library amount of 10–15 $\mu$ g as an input. [116] On the contrary, in solution capture utilizes excess probes over the DNA library, allowing capture from low input DNA library, [117] hence making it capable for use in Ct-DNA analysis. [89]

The advantage is this approach is the decreased contamination from non-apoptotic sources, due to lack of shearing step, which increases the Ct-DNA content of the library leading to an increased percentage of mutant Ct-DNA entering the analysis pipeline as compared to PCR-based methods. However, the technique is challenged by the low DNA quantity and quality. [89]

A remarkably successful application of the use of hybridization capture enrichment strategy for NGS quantification of Ct-DNA is “cancer personalized profiling by deep sequencing (CAPP-Seq.)”, developed by Newman and colleagues. [115]

However, it requires a priori knowledge of recurrently mutated genes in the cancer of interest. [89]

***Whole Exome Sequencing (WES).*** Whole exome sequencing can be utilized when a priori knowledge about mutations are lacking. [89] In a study of 6 adult patients with different solid tumors (Breast, Ovarian and Lung), Murtaza et al. 2013 used exome sequencing of cell free tumor DNA to follow the genomic evolution of metastatic cancer. They were able to demonstrate that Ct-DNA represents the tumor genome. Additionally, they successfully tracked the change in MAF of the mutations detected in plasma after different courses of therapy; providing evidence for the utility of WES of Ct-DNA in monitoring clonal evolution and emergence of therapy resistance. The small number of patients recruited represents as limitation. Additionally, the selection of samples with high allele frequency (as detected by ddPCR and tagged-amplicon deep sequencing (TAm-Seq.) limits the generalization of such findings to diseases with low Ct-DNA as well as, cases with low MAF.

***Cancer Personalized Profiling by Deep Sequencing (CAPP-Seq).*** The main aim for developing CAPP-Seq. was the design of a method capable of detecting somatic alterations in the majority of patients without the need of patient specific optimization at a reasonable cost. Newman and colleagues designed a selector for NSCLC, optimized the method and tested its clinical applicability in different stages of NSCLC.

***CAPP-Seq. selector design.*** The selector is made up of biotinylated DNA oligonucleotides complementary to recurrently mutated regions in a specific cancer. Designing the selector comprises the combination of library preparation techniques from low input DNA to a 6-phase bioinformatics approach. The initial phase of the selector design involved the selection of genomic regions carrying known driver mutations in NSCLC from Catalogue of Somatic Mutations in Cancer (COSMIC). This was followed by the addition of exons that contain recurrent SNPs detected in 407 NSCLC by WES from The Cancer Genome Atlas (TCGA). The priority for exons to be included in the CAPP-Seq. NSCLC selector was done based on 2 metrics; Recurrence index (defined as “the number of unique patients (i.e., tumors) with somatic mutations per kilobase of a given genomic unit (i.e. exon) as well as, the number of unique patients (i.e. tumors) with mutations in a given genomic unit. This allowed for keeping the selector size at minimum, while maximizing the number of missense mutations per patient. The authors then added exons of potential driver mutations. Finally, the authors added exons and introns with breakpoints for ROS1, ALK and RET rearrangements. [115]

The total selector size is 125 kb targeting mutations in 139 genes (521 exon and 13 intron), offering the ability to detect 4 SNPs and to cover 96% of adenocarcinoma and squamous cell carcinoma patients. [115] The performance of CAPP-Seq. selector for NSCLC was carried by analyzing samples from cell lines, matched circulating DNA/leucocytes from NSCLC patients, Cf-DNA from healthy adults as well as, NSLC patients. The selector capture demonstrated uniformity and efficiency as shown by the size of the sequenced fragments. [115]



The technique was optimized through studying different factors that would affect the limit of detection as well as accuracy of the approach have been studied. The recovery rate was in accordance with that of PCR, the sample cross contamination was 0.06% in multiplexed circulating DNA, the allelic bias towards reference alleles was minimum (as studied by analysis of heterozygous germline SNVs in peripheral blood) whereas the PCR and sequencing errors as measured by non-reference alleles distribution across the selector, showed to be reduced as compared to other techniques (mean background rate 0.006% and median background rate 0.0003%). Further steps showed that the biological background doesn't highly affect the quantitation of Ct-DNA at limits of detection > 0.01%. Importantly, the detection of fractional abundances at a range of > 0.02% (0.025% and 10%) were accurate and highly linear. [115]

***Clinical utility of CAPP-Seq.*** CAPP-Seq. demonstrated utility in different aspects i.e. MRD monitoring, molecular profiling, assessment of response to therapy as well as, prediction of relapse free survival.

CAPP-Seq. Ct-DNA quantitation pre- and post- treatment accurately detected MRD in primary Stage I-III lung cancer patients. In the group of patients with Ct-DNA detected pre-treatment, MRD status (+ve/-ve) post treatment showed significant correlation with 3-yr progression-free as well as, event-free survival. The Ct-DNA MRD could actually predict disease progression (positive predictive value 100% and a negative predictive value 93%).[118]

In DLBCL, CAPP-Seq. showed utility in subtyping and risk stratification. First, somatic variants in genomic DNA were used to design Cell-of-origin classifier (CCO). CAPP-Seq. was then used to assess the performance of the classifier in tumor and plasma samples. CAPP-Seq. results for tumor and plasma were highly concordant (88%-93%). Initial Ct-DNA levels prior to treatment showed significant correlation with disease biomarker (LDH) and imaging (PET/CT metabolic tumor volume), International Prognostic Index (IPI) and, disease stage, hence, allowing for risk stratification of this group of patients. It was also shown that CCO by CAPP-Seq., in either plasma or tumor, can significantly predict PFS in this group of patients. [119]

Studies also showed that Ct-DNA analysis by CAPP-Seq. can carry a role in selecting patients most likely to benefit from therapy, hence guide the clinical decision, consequently leading to a better outcome. In metastatic NSCLC, Ct-DNA quantitation by CAPP-Seq. was compared to CT scans (standard of care) for post immune checkpoint inhibitors (ICI) treatment response. Both Ct-DNA analysis by CAPP-Seq. and CT scans were done pre-treatment, and repeated post treatment every 6 weeks for CAPP-Seq. and every 2 months for CT scans. The first response assessment by CAPP-Seq. was done after a mean of 41 days from commencement of treatment compared to a mean of 60 days for CT imaging. Patients with partial response or stable disease (durable clinical benefit (DCB)), showed significant reduction in Ct-DNA as compared to patients with no durable benefit (NDB). Moreover, 100% of the patients who had undetectable Ct-DNA were in the DCB group. Hence, Ct-DNA would allow for earlier assessment of response to therapy and consequently patients who would benefit from ICI. [120]

***CAPP-Seq. utility in defining drug resistance mechanisms.*** Analysis of Ct-DNA by CAPP-Seq. proved beneficial in unveiling inter- and intra-patient heterogeneity of drug resistance mechanisms. Chabon. et al. applied a 302 kb CAPP-Seq. selector targeting 771 regions in 252 genes, followed by Illumina Hi-seq. 2500 High Output sequencing (median depth app. 5,500 x) for the analysis of sequential samples from T790M-positive NSCLC patients. They discovered inter- and intra-patient differences in the mechanism of resistance to different generations of tyrosine kinase inhibitors (TKIs) utilized in first as well as, second line treatment. [88]

In agreement with other studies, analysis of Ct-DNA by CAPP-Seq. was successfully applied for the molecular profiling. Pre-treatment (Pre-Rociletinib treatment, i.e. at progression on a first or second generation TKI) tumor and plasma samples showed high level of concordance for EGFR activating mutations and T790M (conferring sensitivity to Rociletinib). [88]

The authors showed that pre-treatment samples (resistant to first or second generation TKIs) harbor T790M in addition to other alterations (SNVs and CNV including MET), at a high frequency. The effect of such alterations (other than T790M) on response to Rociletinib was further investigated.

The change in clonal dynamics during Rociletinib therapy was also shown. Analysis of Ct-DNA captured the change in the relative ratio of T790M subclone to the EGFR activating mutation after the initiation of Rociletinib therapy. The relative ratio of T790M subclone to the activating mutation, in concordance with the results obtained from tissue biopsy, significantly declined at progression on Rociletinib therapy, indicating that Rociletinib is preferentially active against this subclone. Additionally, the ratio of T790M to the activating mutation before Rociletinib initiation was significantly correlated with the treatment response; where patients with ratio  $\leq 0.5$  showing poor response compared to other patients, thus carrying a potential prognostic value. [88]

The group utilized the detected change in clonal dynamics, as measured by the change in the fractional abundance of putative resistance mutations, to unravel new mechanisms of resistance. As was the case with first line TKIs, Ct-DNA analysis revealed the heterogeneity of resistance mechanisms including MET copy number gain, concurrent SCNAs and SNVs in 65% of the patients developing resistance to Rociletinib. Based on the hypothesis that mutations driving resistance would be positively selected with therapy; an increase in the fractional abundance of the subclone harboring MET-copy-number-gain led to further investigation of patients harboring the alteration. The authors were successfully able to demonstrate that MET-copy-number-gain confers resistance to Rociletinib, and that patients harboring MET-copy-number-gain experience a worse outcome. Interestingly, they were able to show that combination of Rociletinib and Crizotinib (MET inhibitor) leads to restoration of Rociletinib sensitivity in RR cells and patient-derived NSCLC xenograft model. [88]

More importantly, the study demonstrated a significant correlation between different mechanisms of resistance and PFS. Antecedent copy number gain in MET,

EGFR and ERBB2 was associated with PFS < 3 months (described by the authors as innate resistance), whereas a (PFS > 3 month) was associated with SNVs (described as acquired resistance). Another seminal finding was the discovery of novel EGFR (L798I) tertiary mutation a KRAS activating mutation after TKI treatment in EGFR mutated NSCLC. [88]

Taken together, the aforementioned findings clearly demonstrate the therapeutic implication of Ct-DNA analysis in guiding therapy. In cases where multiple resistance mutations/mechanisms co-exist, single-pathway inhibition might be insufficient to elicit the desired response.[88] Ct-DNA holds a great promise in identifying patients who are more likely to benefit from single drug versus those who require a combination regimen. It follows that, incorporation of Ct-DNA analysis at diagnosis and throughout therapy, would not only impact the outcome; but also, the cost associated with administration of non-effective therapy. [88]

### **Error Suppression for Deep Sequencing**

For all applied sequencing approaches, a PCR step is included for Cf-DNA libraries enrichment. Despite using high-fidelity polymerases, which combine proofreading activity (3'→5' exonuclease activity) with low error rates and keeping PCR cycles at minimal; the errors generated from PCR and sequencing (i.e. errors existing with bridge amplification at illumina platform) [121] remain a challenge for accurate detection of low frequency alleles. [89, 115] Accordingly, the goal of error suppression methods is to be able to discriminate between reads originating from an original duplex and those originating from other fragments covering the same locus (i.e. representing different alleles or cells).[89]

Molecular indices comprise either random nucleotides, partially degenerate or defined nucleotides (in case of low template input) have been successfully employed to tag. [122]; single DNA strands or double-stranded DNA molecules. Single-strand molecular barcoding is associated with higher efficiency; whereas duplex molecular barcoding better suppresses errors.[123] Two error suppression methods are introduced here, Safe-SeqS and iDES.

**Safe-Sequencing System (“Safe-SeqS”).** This approach mainly utilizes tagging of individual templates using endogenous or exogenous unique identifier (UID) and nested PCR amplification to amplify each uniquely tagged DNA molecules; for each UID tagged molecule, amplification results in UID family; each strand of the original duplex, produces a uniquely identified fragment. Accordingly, a true mutation that exist in the original template should be present every daughter molecule i.e. shared across the UID family. This is followed by redundant sequencing; UID reads are used to organize the reads into a “UID family”. UID is designed as “supermutant” if the mutation is shared in 95% of the family members. [121]

The introduction of an endogenous UID is employed through introducing an exogenous sequence by PCR, ligation to the end of the DNA template; or shearing (in case of genomic DNA) and using original template ends as tags. The detailed method introduced by Kindle and colleagues utilizes shearing, which can't be employed in the case of Ct-DNA as the fragment end distribution is not random. Briefly, shearing is followed library preparation, solid-surface capture, amplification and sequencing. Upon determination of the error rate with and without Safe-SeqS, the authors concluded that Safe-SeqS was associated with a 9-fold decrease in the error rate. [121]

Exogenous UID strategy for Safe-SeqS involves amplification with a set of primers targeting the gene of interest. The UID (random sequence 12-14 N) is attached to one of the primers at its 5' end; both primers have universal sequence for later PCR amplification. Two PCR cycles are carried out to assign UID; as in the case of endogenous UID, each duplex produces two uniquely identified fragments. This is followed by PCR, using universal primers; followed by sequencing where the generated reads are used for UID family assignment and supermutants detection. [121]

Advantages of Safe-SeqS exogenous strategy is that it allows not only for the detection of the fraction of templates harboring the variants; but also, the quantitation of the number of templates analyzed. The technique can analyze low input samples with minimal template loss. Comparing Safe-SeqS to conventional analysis of illumina sequencing data showed an error reduction of 24- and 15-fold respectively when analyzing the prevalence of rare mutations in a single gene from human cells or short mitochondrial gene segment respectively. [121]

The applicability of the technique in analyzing multiple amplicons from a low-input sample has not been studied; multiplexing is proposed to combat such a drawback. Additionally, Safe-SeqS success is dependent on the efficiency of early amplification cycles for UID assignment, which can be affected by the presence of inhibitors in the clinical sample. Also, during library enrichment, polymerase error can generate a new UID family; consequently, creating an artificial supermutant. Additional criteria for the analysis would help mitigate the effect of such factor. [121]

**Integrated Digital Error Suppression method (iDES).** Low concentration of Ct-DNA, especially post-treatment, represents a challenge for ultrasensitive detection techniques. CAPP-Seq. had > 50% sequencing artifacts below an allele frequency of 0.02%. Newman et al. 2016 developed iDES error suppression method to overcome the sequencing errors challenging CAPP-Seq. at detection limit below 0.02%. [123]

Newman and colleagues designed “Index adapters” and “Tandem adapters”; detailed as follows.: (1) “Index Adapters” comprise a re-design of the original 8-base multiplexing barcode; an 4-mer “index barcode” is incorporated into the single stranded part of the adapter, adjacent to the now 4-base multiplexing barcode. During sequencing, Index barcodes are read with index reads. (2) “Tandem Adapters” include two exogenous index reads (to uniquely barcode each single strand of the original duplex) and two 2-base-double stranded-UID “insert barcodes” integrated into the ds part, adjacent to the

ligating side, of the adapter. Insert barcodes are sequenced as part of the main read (at the beginning and the end of the read). Hence, for paired-end sequencing, 4-base UID is assigned to every single DNA strand. The authors note that the position of insert and/or index barcode, as well as the length can be altered in different adapter designs to achieve the desired level of molecular diversity. [123]

The effect of “Index” and “Insert” barcodes on error suppression was evaluated using the NSCLC CAPP-Seq. selector previously described (under CAPP-Seq. Selector Design). Analysis done incorporating “index barcode” only resulted in 2.5-fold improvement in the mean error rate detected in the CAPP-Seq. approach; whereas the incorporation of “insert” barcode only exceeded “index barcode” by 1.25 -fold. Employing both barcodes, the overall error rate was  $9 \times 10^{-5}$  error/base. [123]

Cell-free sequencing data showed that G>T transversions represent the majority of background errors. In their analysis for possible causes; Newman and colleagues were able to exclude in vivo oxidative damage and strand bias in sequencing as causes for such increase in G>T transversions. Evaluating different times associated with hybrid capture, they demonstrated that most of the errors is generated post ligation and during the enrichment step of strand-specific target sequence. To mitigate background error, the authors designed a computational method “Background Polishing” through employing position-specific error modeling in a control cohort to allow error suppression in independent samples. [123]

Combining barcoding with background polishing “iDES” resulted in a 15-fold reduction of overall error rate ( $1.5 \times 10^{-5}$  errors per base) achieving 98% error-free positions (as compared to 90%). [123]

The effect of iDES on the recovery rate (a known important factor affecting the detection limit) was evaluated using 32 ng of Cf-DNA. Post-capture recovery rates of 60% were achieved for ~ 10,000 input hGE, hence reducing errors without negatively affecting the efficiency of recovery. [123]

Furthermore, the effect of iDES on CAPP-Seq. detection limit was validated. A Glioblastoma multiforme personalized CAPP-Seq. selector was designed to cover 1502 non-synonyms SNV; 32 ng of healthy Cf-DNA were spiked various proportions of GBM DNA molecules. The iDES-enhanced-CAPP-Seq. limit of detection was 10-fold below that reported for CAPP-Seq. i.e. 0.0025% (using 30 mutations and ~ 3000 hGE). [123]

In samples obtained from NSCLC patients prior to clinical progression, a detection limit of 0.004% was achieved. iDES-enhanced-CAPP-Seq. successfully detected Ct-DNA in 93% of patient prior to treatment; and significantly detected in 73% of pre-and post-treatment samples (Specificity 100%). [123]



## **Ct-DNA in Hematological Malignancies**

Employing advanced molecular approaches, proof-of-concept studies have demonstrated the utility of Ct-DNA analysis for mutational profiling and longitudinal monitoring in different hematological malignancies. [14, 17]

In patients with Myelodysplastic Syndrome (MDS), targeted deep sequencing of 55 genes (known to be mutated in MDS) detected the driver mutations with 100% concordance between matched BM and plasma samples. Moreover, the MAF detected in Ct-DNA showed an excellent correlation ( $r^2 = 0.84$ ) with that measured in BM across most of the matched points, providing evidence for Ct-DNA as a non-invasive biomarker for disease monitoring. Importantly, monitoring of somatic mutations and karyotype abnormalities throughout therapy; successfully predicted treatment failure. [17]

The utility of somatic mutation detection in Ct-DNA in monitoring disease progression was also demonstrated in Multiple Myeloma (MM), where the change in the fractional abundance of a clone specific mutation showed concordance with the clinical disease status i.e. increase in FA coincided with disease relapse or moreover, occurred prior to serological markers detection. [14]

Furthermore, Ct-DNA analysis showed applicability in the comprehensive characterization of MM mutational landscape, as an adjunct to BM. Mithraprabhu et al. utilized OnTarget mutation detection platform which includes 96 mutations in the KRAS, NRAS, CTNNB1, EGFR, PIK3CA, TP53, FOXL2, GNAS and BRAF genes, to study BM and plasma of newly diagnosed and relapsed/refractory MM patients. Out of total 128 mutations detected, 46% showed concordance between BM and plasma, 30% were detected in BM only, whereas 24% were detected only in plasma; hence giving a supporting evidence for the presence of spatial heterogeneity in MM. [14]

## **Ct-DNA in Pediatric Leukemia**

Studies of Cf-DNA in pediatric leukemia focused on the quantification of leukemia MRD, done by RQ PCR using allele specific oligonucleotide (ASO) primers for Ig/TCR targeting the junctional region, during early phase of treatment.[124, 125] In accordance with previous studies showing a higher Cf-DNA concentration in cancer patients, higher total plasma DNA was detected at diagnosis in pediatric ALL, as compared to healthy children, which decreased back to normal levels by D4. Sequential quantification of total plasma DNA, and leukemic MRD in pediatric ALL patients at specific time points during induction phase, demonstrated significantly higher plasma total DNA level initially and at D3 versus later time points, which correlated to the leucocyte count. Slight elevation in total DNA concentration was detected on D33 and week 12, alongside with marrow regeneration. Leukemia MRD detected qualitatively showed high concordance (86%) with MRD detection in leukocytes. [124] Leukemia MRD was below the limit of detection (cutoff of  $10^{-5}$ ) at D33 and week 12. Despite the lack of correlation between FC-MRD and plasma leukemic MRD, plasma leukemic MRD

(D15, cutoff of  $10^{-4}$ ) showed an independent correlation to the risk of relapse in both univariate and multivariate analysis in Chinese children. [125]

Limited by the sensitivity of the techniques used [126], the studies couldn't solidify a role for Ct MRD quantification of Ct-DNA, yet, they clearly demonstrated the feasibility of analyzing Ct-DNA in this group of patients, showing that the analysis of Ct-DNA in pediatric leukemia carries great promise if more advanced sensitive techniques are incorporated in the analysis.

### **Rationale of the Study**

Multiple genetic factors play a role in the pathogenesis and prognosis of pediatric leukemia. Interaction occurs between inherited and somatic genetic alterations, where specific inherited alleles predispose to specific chromosomal-defined subtypes. [1] The genetic landscape of leukemia is further complicated by the genetic heterogeneity; the disease reported to be polyclonal in 50% of the patients at diagnosis. [2] Additionally, pharmacogenomic determinants of drug response and toxicity affects the outcome [3].

MRD-adapted therapy, in which risk stratification is based on biological features and response to therapy has been associated with an improvement in the treatment outcome. [4] Nevertheless, within the favorable risk group, a substantial number of patients, relapse. [5]

Advancement in the molecular techniques have provided a deeper insight into the biology of relapse. In most cases, relapse emerged from a minor clone at diagnosis or rather a pre-ancestral clone which survived therapy, acquired additional mutations to become the relapse founder clone. Moreover, mutations enriched at relapse were unraveled as well as, in many cases their association with drug resistance. [9, 11, 57]

Although several studies have addressed the role of clonal evolution in relapse [3, 7], the change in clonal dynamics has been only studied at diagnosis, remission and relapse. Hence, the knowledge of real-time change in the clonal dynamics, in response to therapy, remains obscure.

Also, despite the demonstrated impact of heterogeneity on clonal evolution and relapse, comprehensive characterization of the disease at diagnosis or throughout therapy is still limited. This is attributed to the inherent limitations in the currently applied strategies for ALL diagnosis and follow up. Two main limitations hinder the detection of clonal diversity at diagnosis and longitudinal follow up of clonal evolution. Firstly, samples for response assessment are obtained through single-site BM biopsy. The invasiveness of the technique hinders its use for sequential analysis. [6,7] Additionally, a single BM aliquot fails to capture tumor heterogeneity associated with spatially separated subclones. Secondly, MRD assessment is done either by Flowcytometry or PCR, both of which lack the ability to detect subclones that differ from the major diagnostic clone. [8,9]

The finding that an increase in the percentage of mutations conferring resistance precedes clinical relapse, together with the successful reversal of resistance using small molecules or targeted therapies, implies the importance of incorporating techniques capable of tracking clonal dynamics into clinical management protocols. [10]

Hence, the development of biomarker that is capable of real-time follow up of clonal dynamics and therapy resistance carries great promise in leukemia monitoring, consequently guiding therapy.

Ct-DNA was shown to be derived from different malignant subclones and hence offers a representation of tumor heterogeneity and would thus provide a better characterization of disease behavior than does currently used quantitative methods based on the detection of the most dominant diagnostic clone. Importantly, it would serve as a minimally invasive technique, which can be repeated at multiple time points. [11]

Several proof-of-concept studies have demonstrated the utility of Ct-DNA in different application in various hematological and solid cancers. [11,12,13,14]. The analysis of Ct-DNA showed high degree of concordance to BM or tissue biopsy; demonstrating its utility in molecular profiling. [11,12]. Additionally, measurement of MAF in Ct-DNA showed utility in longitudinal monitoring of the disease [14]. Moreover, analysis of Ct-DNA allowed for tracking drug resistance and discovering drug resistance mechanisms. [11,13]

However, studies of Ct-DNA in pediatric leukemia were limited by the sensitivity of the techniques used yet, they were still able to demonstrate its feasibility in sequential analysis during early treatment phase of treatment. [15,16]

### **Study Objective**

In this project, we study the utility of somatic mutation detection in plasma Ct-DNA as a method for molecular profiling at diagnosis, as well as monitoring tumor dynamics throughout therapy. We employed WES for the first recruited patients; however, data analysis was challenged by the large number of mutations detected and the difficulty to discriminate true mutations from false positives. For the next group of patients, we applied targeted sequencing using ALL customized gene panel. The MAF is compared across different time points as well as, between BM and Plasma samples at matched time points.

Our long-term goal is to incorporate Ct-DNA in the clinical protocols as a minimally invasive biomarker for comprehensive molecular profiling and longitudinal monitoring of the disease in pediatric ALL

NGS was applied to study somatic mutations in peripheral blood (PB)/bone marrow (BM) plasma Ct-DNA at diagnosis, at specified time points during, and at the



end of remission induction therapy. MAF was calculated for all the specified time points. The specific aims of the current work are:

- 1) Testing the utility of Ct-DNA in capturing ALL tumor heterogeneity at diagnosis. MAF detected by Ct-DNA analysis of diagnostic PB plasma is compared to MAF detected by clinical WES of diagnostic BM biopsy sample.
- 2) Testing the utility of Ct-DNA in monitoring the change in clonal dynamics during remission induction therapy. Somatic mutation detection in Ct-DNA was compared to FC based clinical MRD, with a specific focus on patients with negative MRD.

## **CHAPTER 3. METHODOLOGY**

### **Patients and Samples**

All patients were enrolled in Total Therapy Study XVII (TOTXVII), St. Jude Children's Research Hospital. Samples were collected with informed consent for research use and approval of IRB. BM samples were collected at time of diagnosis, D15 and at the end of induction. Peripheral blood samples were collected at diagnosis, Induction therapy D3, 8, 15, 22 and at the end of induction.

### **Separation of Plasma from Peripheral Blood**

Blood samples were collected, at the specified time points during induction therapy, in K2EDTA tubes and plasma separated as follows: samples were diluted with PBS in a ratio 1:1, followed by centrifugation at 1600 g (10 min), plasma is then separated and re-centrifuged at 14000 rpm (10 min.) to remove any residual cells, then stored at -80°C as 1 ml aliquots until Cf-DNA extraction. BM and blood samples were fractionated by the biorepository staff.

### **Cell-free DNA Extraction and Quantification**

Cf-DNA was extracted from 650 µl – 1000 µl (average 950 µl) of plasma or non-cellular BM (BM plasma) using Maxwell RSC ccfDNA Plasma Kit, according to the manufacturer's protocol using 40 µl Qiagen elution buffer. Following extraction, Cf-DNA was quantified using Qubit dsDNA High Sensitivity Kit.

Technical details for using Maxwell RSC ccfDNA Plasma Kit can be accessed at <https://www.promega.com/resources/protocols/technical-manuals/101/maxwell-rsc-ccfdna-plasma-kit-protocol/>. [127]

### **Construction of Adapter-ligated Libraries**

For each sample, the whole amount of Cf-DNA isolated was input into library preparation. Library preparation was carried out by NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (NEB #E7645) according to manufacturer's protocol with few modifications: (1) no size selection was done irrespective of the input (2) Kappa 10x primer mix was used for PCR enrichment of the adapter-ligated libraries (3) For WES, dual indexed adapters were used whereas for targeted capture, index barcoded adapters were employed (4) Qiagen EB buffer was used for elution.

Since no pooling guidelines were available for Nextseq, we selected adapter combinations based on TruSeq DNA Sample Preparation Guide (Pooling Guidelines) for

Hiseq./Miseq. ([https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry\\_documentation/samplepreps\\_truseq/truseqdna/TruSeq\\_DNA\\_SamplePrep\\_Guide\\_15026486\\_C.pdf](https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqdna/TruSeq_DNA_SamplePrep_Guide_15026486_C.pdf)) [128] as well as, general pooling guidelines (<https://support.illumina.com/bulletins/2016/10/library-pooling-guidelines-for-the-nextseq-and-miniseq-systems.html>). [129]

Post PCR clean-up was carried as single or double-sided cleanup. Libraries were quantified using Qubit ds-DNA High Sensitivity Kit. Following library preparation and quantification, QC Miseq run was performed for assessment of the quality of the sequencing libraries prior to hybridization capture.

### Hybridization-based Capture and Sequencing

For each case, libraries (equivalent to pre-defined induction time points) are pooled (multiplexed) together (in a ratio 1:1) into one capture.

For WES, hybridization-based exome capture was done according to IDT protocol *Hybridization capture of DNA libraries using xGen® Lockdown® Probes for Illumina HT- Adaptor Ligated Libraries* according to manufacturer's protocol, using *xGen® Exome Research Panel* which consists of 429,826 probes, spans a 39 Mb target region (19,396 genes) of the human genome and covers 51 Mb of end-to-end tiled probe space. Enriched libraries were quantified using Qubit™ dsDNA BR Assay Kit. QC Miseq run was performed for assessment of the quality of the pooled libraries prior to full Nextseq run employing Illumina Nextseq High Output lane.

For targeted capture, the same protocol was applied using the custom gene panel which covering  $\cong$  1668 exons and 319.4 kb.

### Design of the Mutation Capture Panel

A specific ALL custom gene panel was designed by our lab. computational biologist. Initially, the list combined a number of genes known to harbor somatic mutations in ALL derived from different pathways/sources (1) Oncogenic pathway, (2) Tumor suppressor genes, (3) Relapse-specific genes, (4) Drug metabolism, (5) CRISPR/Cas9 screening (unpublished data), (6) Genes associated with glucocorticoid resistance, (7) xGen® Pan-Cancer Panel (which includes list of genes derived from TCGA; known to be mutated across multiple cancer types. Finally, the genes which harbored  $\geq 3$  non-synonyms SNVs in B-cell or T-cell (either at diagnosis or relapse) in the SiTH database (derived from  $\cong$  500 patients) were selected for the final panel. Probes were designed to the genomic regions surrounding the mutations (Personal communication with Ousman Mahmoud on 17<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital) (**Appendix A. List of Genes of Custom ALL Panel**).

## Processing of Sequencing Data

For WES, Trimmomatic was employed for trimming adapters and illumina sequences. Sequence reads were aligned to the reference genome using Burrows-Wheeler Aligner (BWA-aln/sampe). Alignments were sorted by co-ordinate using Picard.SortSam. GATK was employed for marking duplicates, local realignment and base-recalibration. Variant calling and filtering was performed using Mutec2. Annotation was carried out using Annovar.

For targeted sequencing using the ALL customized gene panel, sequence reads were de-multiplexed using Picard, FgBio, Samtools. Sequence reads were aligned to the reference genome using Burrows-Wheeler Aligner (BWA-aln/sampe). Fgbio was employed for UMI consensus calling. Reads were grouped by UMIs (Fgbio – GroupReadsByUmi) Consensus reads generated (Fgbio –CallMolecularConsensusReads) and filtered (Fgbio– FilterConsensusReads). Sequence reads were aligned to the reference genome using Burrows-Wheeler Aligner (BWA-aln/sampe). GATK was employed for marking duplicates, local realignment and base-recalibration. Varscan was applied for mutation detection. Variant filtering was done according to specific criteria including: the presence in  $\geq$  two time-points, location within the target region, Non-synonyms mutation and absence from germline. Additionally, mutations detected in sequence repeats are removed as well as, those present across multiple samples.

## CHAPTER 4. RESULTS

### Study Design and Patient Enrollment

14 patients recruited in St. Jude Total Therapy Study XVII were included in the study. B-cell precursor comprised 11/14 (78.6%) whereas T-cell represented 3/14 (21.4%). The patients' age ranged from 2.6 to 17.3 years (**Table 4-1**). Using NCI risk criteria for classification of precursor B-cell cases, 4 patients were classified as High risk (WBCs  $\geq$  50,000 cells and/or Age  $\geq$  10 years) and 7 patients were Standard risk (WBCs  $<$  50,000 cells and/or Age  $<$  10 years). Provisional risk stratification as well as, final risk stratification according to Total XVII study is summarized in (**Table 4-2**). Total XVII provisionally classifies patients into three risk groups (low-, standard-, or high risk) according to several presenting features (age, TLC, CNS status, molecular, cytogenetic diagnosis and immunophenotyping (IPT). Early response to therapy, as measured by MRD during and at the end-of-induction therapy dictates the final risk stratification. (**Appendix B. Total XVII Risk Stratification**).

In Total study XVII, NGS is implemented as a diagnostic clinical standard-of-care for all patients enrolled in the protocol. WGS, WES and transcriptome sequencing are carried by the clinical pathology lab.; genetic alterations with relevance to diagnosis or management are reported in patients' clinical records. Total XVII therapy implies Flowcytometry based MRD for response assessment during induction therapy and throughout treatment. During induction, MRD is evaluated in BM on D15, D22 (in patients with D15 MRD  $\geq$  1%) and at end of induction (D42); whereas MRD evaluation in PB is carried out on D8.

We performed Ct-DNA analysis in sequential plasma samples withdrawn at equivalent time points. BM-plasma was evaluated on D1, D15 and at end of induction; whereas PB-plasma was evaluated on D1, D3, D15, D22 and at end of induction.

### Somatic Mutation Detection in Leukemia-derived Cell-free DNA

Cf-DNA was isolated from PB plasma and BM plasma using Maxwell® RSC ccfDNA Plasma Kit. Extracted DNA was quantified by Qubit™ dsDNA HS Assay Kit. whole amount of Cf-DNA isolated/sample was used as input into library construction. Molecular barcoding was incorporated into library preparation for targeted sequencing. Libraries were quantified using Qubit™ dsDNA HS Assay Kit, analyzed on agarose gel to confirm the size of the fragments. This was followed by QC Miseq. run for libraries. Following library construction, samples were either exome captured using IDT Exome Panel or targeted capture. Post exome or target enrichment, libraries were sequenced on Illumina NextSeq. 550-platform. Analysis of sequencing data was done by a bioinformatics pipeline for the detection of somatic SNV.

**Table 4-1. Patient Characteristics**

Case#	IPT	CRLF-2 Expression	Age (yrs)	Initial WBCs (x10e9/L)	CNS Status	Stemline DNA Index	t(9;22) CML	t(9;22) ALL	t(1;19)	t(4;11)	t(12;21)
Case 1	B-cell	Negative	2.9	4.4	Negative	1.25	Negative	Negative	Negative	Negative	Negative
Case 2	B-cell	Positive	3.1	50.6	Negative	1.05	Negative	Negative	Negative	Negative	Negative
Case 3	B-cell	Negative	11.1	1	Negative	1.20	Negative	Negative	Negative	Negative	Negative
Case 4	T-cell	-----	12.8	189.5	Positive	1	-----	-----	-----	-----	-----
Case 5	B-cell	Negative	12.3	6.4	Negative	1.18	Negative	Negative	Negative	Negative	Negative
Case 6	B-cell	Negative	2.6	25.7	Negative	1	Negative	Negative	Negative	Negative	Positive
Case 7	B-cell	Positive	17.3	49.7	Positive	1	Negative	Negative	Negative	Negative	Negative
Case 8	B-cell	Negative	3.6	2.5	Negative	1	Negative	Negative	Negative	Negative	Positive
Case 9	B-cell	Negative	7.6	1	Positive	1.24	Negative	Negative	Negative	Negative	Negative
Case 10	ETP	-----	4.8	0.2	Negative	1	-----	-----	-----	-----	-----
Case 11	B-cell	Positive	3.3	5.3	positive	1	ND	ND	Negative	ND	Negative
Case 12	B-cell	Negative	9.3	25.7	Negative	1	ND	ND	Negative	ND	Negative
Case 13	B-cell	Negative	7.7	3.5	positive	1.16	ND	ND	Negative	ND	Negative
Case 14	T-cell	-----	11.3	32.5	Negative	----	-----	-----	-----	-----	-----

**Table 4-2. Classification of Patients into Risk Groups by NCI Criteria and Total XVII Criteria**

<b>Case#</b>	<b>NCI Risk</b>	<b>Total XVII Provisional Risk</b>	<b>D8 PB MRD (%)</b>	<b>D15 BM MRD (%)</b>	<b>D15 PB MRD (%)</b>	<b>D22 BM MRD (%)</b>	<b>D22 PB MRD (%)</b>	<b>EOI BM MRD (%)</b>	<b>EOI PB MRD (%)</b>	<b>Total XVII Final Risk</b>
Case 1	Standard	Low	0.408	6.820	----	2.12	----	<0.01	----	Standard
Case 2	High	Standard	0.071	0.01	----	-----	----	<0.01	----	Standard
Case 3	High	Low	0.013	2.150	----	<0.01	----	<0.01	----	Standard
Case 4	----	Standard	<0.01	<0.01	----	----	----	<0.01	----	Standard
Case 5	High	Low	7.930	0.983	----	----	----	<0.01	----	Low
Case 6	Standard	Low	0.003	0.016	----	----	----	<0.01	----	Low
Case 7	High	Standard	0.738	1.502	----	0.421	----	<0.01	----	Standard
Case 8	Standard	Low	0.030	0.022	----	----	----	<0.01	----	Low
Case 9	Standard	Standard	0.045	0.191	----	----	----	<0.01	----	Standard
Case 10	----	Standard	4.370	----	<0.01	0.958	1.609	2.981	4.73	High
Case 11	Standard	Standard	0.018	0.014	----	----	----	<0.01	----	Standard
Case 12	Standard	Low	16.030	9.936	----	2.059	----	<0.01	----	Standard
Case 13	Standard	Standard	<0.01	<0.01	----	----	----	<0.01	----	Standard
Case 14	----	Standard	38.65	55.203	47.04	31.665	30.246	0.139	0.043	High

## Cf-DNA Isolation and Library Preparation

**Samples for WES.** An average of 952  $\mu$ l (800 $\mu$ l-1100 $\mu$ l) plasma/sample was used as an input to Maxwell® RSC ccfDNA Plasma Kit followed by quantification by Qubit™ dsDNA HS Assay Kit. The mean concentration of Cf-DNA isolated/sample was 1.35 ng/ $\mu$ l; median= 0.27 ng/ $\mu$ l; range 0.05-34.2 ng/ $\mu$ l); For diagnostic samples, mean= 1.5 ng/ $\mu$ l; median 0.332 ng/ $\mu$ l). The total yield of Cf-DNA isolated/sample was used as input into library preparation (mean Cf-DNA input to library preparation = 41.5 ng; median=9.6 ng; range=1.6-950 ng).

**Samples for targeted capture.** For peripheral blood plasma and BM plasma samples analyzed, the mean yield of Cf-DNA extracted was 56.6 ng, median=9 ng. The yield of extraction for peripheral blood plasma samples was 54.5 ng, median = 8.93ng whereas, for BM plasma samples; the mean yield of Cf-DNA extracted was 58.3 ng and the median 9.46 ng.

**Gel electrophoresis for Cf-DNA libraries.** The fragment size was ~ 300 bp as expected for Cf-DNA. Of notice that in some cases bi-nucleosomal inserts yield libraries of size ~ 450 bp; ~334 bp of the bi-nucleosomal Cf-DNA insert + 120 bp (2 \*60-bp Adapter) (**Figure 4-1**).

## Exome Capture

Gel electrophoresis repeated after exome capture confirmed the 300 bp size was confirmed for the captured fragments (**Figure 4-2**).

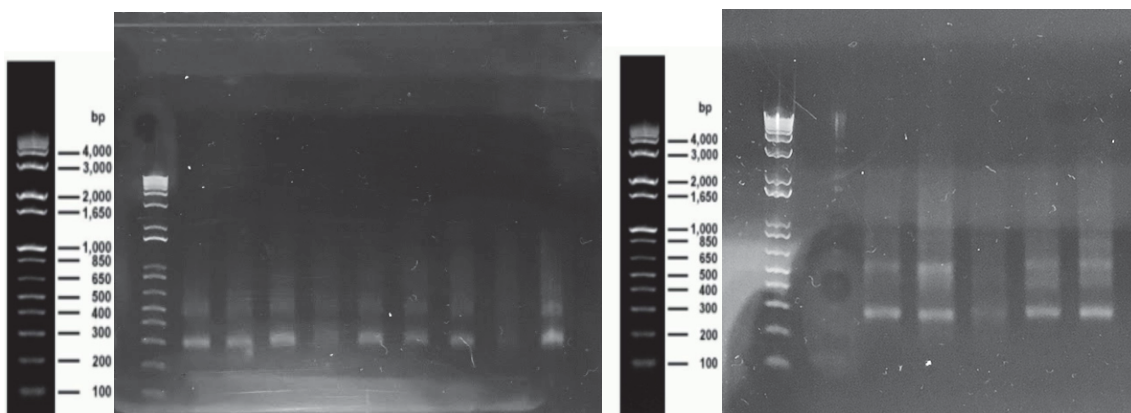
## Sequencing Metrics

Post exome capture, Miseq QC run was carried followed by NextSeq. complete run. QC metrics are described below:

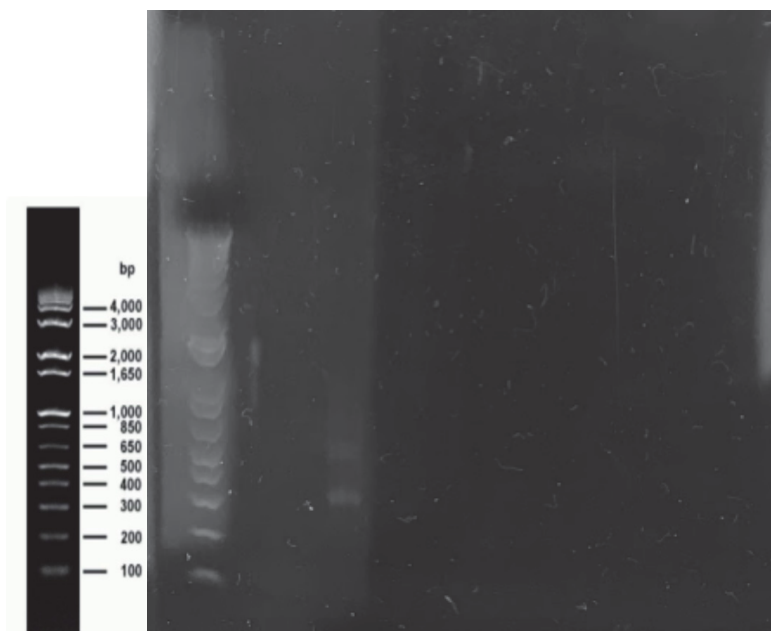
**NextSeq. QC metrics for WES.** QC metrics for the NextSeq. runs (**Table 4-3**) are summarized as follows: average number of reads =110717316.7, average % reads lligned = 99.39454545, average % read duplicates= 33.39136364 and average % selected bases (xgen Exome) = 87.79.

**Miseq. QC metrics for targeted capture.** Miseq QC metrics for the targeted capture (**Table 4-4**) are summarized as follows: average % of aligned reads=88.7% whereas the average % of selected bases=93.94%.





**Figure 4-1. Gel Electrophoresis for Cf-DNA Adapter-ligated Libraries**  
Case#8 (left) and Case#10 (right)



**Figure 4-2. Gel Electrophoresis for Exome-captured Pooled Libraries (Case#2)**

**Table 4-3. NextSeq. QC Metrics for WES-captured Libraries**

<b>Case #</b>	<b>Sample Name</b>	<b>Number of Reads</b>	<b>% Reads Aligned</b>	<b>% Read Duplicates</b>	<b>%Selected Bases (xgen Exome)</b>
<b>Case 01</b>	Case01_D01_BMPL	155653732	99.56	73.42	88.97
	Case01_D01_PBPL	84983694	97.53	14.98	92.24
	Case01_D03_PBPL	117168598	99.45	64.00	89.62
	Case01_D08_PBPL	8008298	99.29	3.53	92.04
	Case01_D15_PBPL	107283528	99.44	53.65	90.42
	Case01_D15_BMPL	153753698	99.53	48.28	91.13
	Case01_D22_PBPL	106452792	99.43	13.91	91.96
	Case01_D42_PBPL	111422134	99.50	32.16	91.68
	Case01_D42_BMPL	95648770	99.42	44.80	90.94
<b>Case 02</b>	Case02_D01_BMPL	166646680	99.71	14.00	92.63
	Case02_D01_PBPL	143217210	99.71	12.83	92.55
	Case02_D03_PBPL	131912546	99.66	14.60	91.89
	Case02_D08_PBPL	144250060	99.21	20.82	82.97
	Case02_D14_PBPL	90042306	99.75	40.96	91.16
	Case02_D15_BMPL	79915688	99.40	30.39	90.87
	Case02_D22_PBPL	88835376	99.74	83.13	84.53
	Case02_END_PBPL	112411712	99.73	67.40	90.90
	Case02_END_BMPL	65827244	99.22	32.92	88.91
<b>Case 03</b>	Case03_D01_BMPL	122430274	99.22	19.81	82.94
	Case03_D01_PBPL	110666618	99.33	21.45	84.53
	Case03_D03_PBPL	93000222	99.22	22.08	83.61
	Case03_D08_PBPL	101185822	99.31	35.93	82.44
	Case03_D15_BMPL	143407628	99.35	78.32	65.89
	Case03_D15_PBPL	81486274	99.31	22.66	82.97
	Case03_D22_PBPL	76349870	99.33	22.60	83.27
	Case03_END_BMPL	116908314	99.35	39.51	81.72
	Case03_END_PBPL	124942240	98.51	19.64	84.54
<b>Case 05</b>	Case05_D01_PBPL	164377728	99.64	14.76	88.98
	Case05_D03_PBPL	135878598	99.56	13.39	88.24
	Case05_D08_PBPL	151060668	99.57	14.01	89.02
	Case05_D15_BMPL	124653950	99.32	12.06	88.25
	Case05_D15_PBPL	119711104	99.56	13.42	88.16
	Case05_D22_PBPL	180899414	99.63	70.73	79.90
	Case05_END_BMPL	100656310	99.18	11.58	88.70
	Case05_END_PBPL	136924480	99.58	49.21	84.22

**Table 4-3. (Continued)**

<b>Case #</b>	<b>Sample Name</b>	<b>Number of Reads</b>	<b>% Reads Aligned</b>	<b>% Read Duplicates</b>	<b>%Selected Bases (xgen Exome)</b>
<b>Case 06</b>	Case06_D01_BMPL	131321820	99.62	31.37	91.28
	Case06_D01_PBPL	137796606	99.62	12.82	91.90
	Case06_D03_PBPL	49788616	99.60	8.61	92.09
	Case06_D08_PBPL	1761754	99.53	17.28	92.21
	Case06_D15_BMPL	119909784	99.25	54.59	89.78
	Case06_D15_PBPL	16765330	99.38	82.45	85.21
	Case06_D22_PBPL	153520806	99.63	84.49	84.51
	Case06_END_BMPL	92246978	99.81	11.73	91.29
	Case06_END_PBPL	120476660	99.67	18.94	91.70

**Table 4-4. Post-capture Miseq. QC Metrics for Targeted-captured Libraries**

<b>Case #</b>	<b>Sample</b>	<b>% Reads Aligned</b>	<b>%Selected Bases</b>
<b>Case 01</b>	D1_PB	77.20	93.70
	D15_PB	78.58	93.90
	D22_PB	77.80	93.15
	D42_PB	79.19	93.79
<b>Case 02</b>	D1_BM	77.82	93.49
	D1_PB	76.95	93.99
	D3_PB	77.05	93.68
	D8_PB	76.73	93.51
	D14_PB	76.10	94.01
	D15_BM	70.77	93.55
	D22_PB	76.44	92.12
	D42_BM	76.19	93.91
	D42_PB	76.07	93.78
<b>Case 03</b>	D1_BM	77.13	93.91
	D1_PB	77.45	94.09
	D3_PB	76.89	94.55
	D8_PB	73.50	94.16
	D15_BM	75.61	93.44
	D15_PB	76.88	93.39
	D22_PB	76.60	93.86
	D42_PB	82.18	94.25
	D42_BM	68.10	94.03
<b>Case 04</b>	D1_PB	73.73	94.01
	D3_PB	74.60	94.43
	D8_PB	84.27	90.74
	D15_BM	83.81	91.05
	D15_PB	84.97	90.80
	D22_PB	85.94	91.05
	D42_BM	83.20	91.40
	D42_PB	85.03	90.97

**Table 4-4. (Continued)**

<b>Case #</b>	<b>Sample</b>	<b>% Reads Aligned</b>	<b>%Selected Bases</b>
<b>Case 05</b>	D1_PB	85.89	91.56
	D3_PB	85.26	91.50
	D8_PB	85.00	91.56
	D15_BM	85.56	91.65
	D15_PB	86.07	90.90
	D22_PB	85.50	91.46
	D42_BM	80.09	90.15
	D42_PB	83.92	89.33
<b>Case 06</b>	D1_BM	84.38	90.07
	D1_PB	84.63	89.80
	D3_PB	84.30	89.1
	D8_PB	84.09	90.38
	D15_PB	85.67	89.85
	D15_BM	84.19	90.22
	D22_PB	85.09	90.27
	D42_BM	81.84	89.93
	D42_PB	84.81	90.56
<b>Case 07</b>	D1_PB	83.81	88.12
	D3_PB	94.29	93.90
	D8_PB	93.70	94.88
	D15_BM	94.27	93.64
	D15_PB	94.16	94.75
	D22_PB	94.19	95.94
	D42_BM	94.28	95.03
	D42_PB	94.18	94.49
	D73_PB	92.42	94.94
<b>Case 08</b>	D1_PB	94.46	96.35
	D3_PB	94.21	97.43
	D8_PB	93.25	95.48
	D15_BM	94.55	96.06
	D15_PB	94.29	96.65
	D22_PB	94.84	96.02
	D49_PB	93.99	96.10
	D52_BM	95.65	93.72
	D52_PB	94.49	96.12

**Table 4-4. (Continued)**

<b>Case #</b>	<b>Sample</b>	<b>% Reads Aligned</b>	<b>%Selected Bases</b>
<b>Case 09</b>	D1_BM	94.55	96.65
	D8_PB	94.90	94.88
	D22_PB	95.16	95.92
	D50_BM	93.80	94.92
	D50_PB	94.11	95.46
<b>Case 10</b>	D15_PB	94.92	95.49
	D22_PB	94.40	90.84
	D22_BM	97.27	94.33
	D43_PB	97.32	95.13
	D59_PB	96.98	94.42
<b>Case 11</b>	D1_PB	97.44	91.25
	D3_PB	97.48	93.41
	D8_PB	97.32	95.13
	D14_PB	96.80	95.86
	D15_BM	96.67	94.93
	D22_PB	96.74	93.40
	D42_BM	96.99	95.14
	D42_PB	97.29	95.66
	D98_PB	97.02	94.36
<b>Case 12</b>	D1_BM	97.44	91.02
	D1_PB	97.62	94.00
	D3_PB	97.52	91.33
	D8_PB	97.37	94.75
	D15_BM	97.11	94.80
	D15_PB	97.12	95.11
	D22_PB	97.51	97.11
	D42_BM	97.71	93.75
	D42_PB	95.86	94.11
	D71_PB	96.89	93.99

**Table 4-4. (Continued)**

<b>Case #</b>	<b>Sample</b>	<b>% Reads Aligned</b>	<b>%Selected Bases</b>
<b>Case 13</b>	D1_PB	97.71	97.96
	D3_PB	97.67	91.92
	D8_PB	95.47	97.81
	D15_BM	96.40	96.95
	D15_PB	95.36	96.54
	D22_PB	96.18	98.42
	D42_PB	95.71	97.77
<b>Case 14</b>	D1_BM	96.10	96.80
	D1_PB	95.94	96.01
	D3_PB	95.83	98.00
	D8_PB	95.62	96.96
	D15_BM	95.95	96.15
	D15_PB	96.33	97.85
	D22_PB	93.85	97.71
	D42_PB	64.76	97.62
	D64_PB	95.92	97.49

## Non-synonymous Mutation Detection in Whole-exome Captured Libraries

Mutation detection in whole exome captured samples was challenged by the extremely high number of mutations detected per sample, the low MAF detected, the low depth of coverage, and high error rate; hence, rendering the discrimination of false positive from true positive calls difficult (**Figure 4-3**).

## Non-synonymous Mutation Detection in Targeted Captured Libraries

**Correlation between WBCs and Cf-DNA yield.** A moderate significant correlation existed between total leukocytic count (TLC) and Cf-DNA yield (spearman's coefficient:  $r=0.33$ ;  $p\text{-value}=0.0029$ ) (**Figure 4-4**).

However, when samples were grouped according to WBCs into low ( $TLC < 1$ ) and high ( $TLC \geq 1$ ), no significant difference existed between the two groups. The mean Cf-DNA yield for the high group was 43.53 ng ( $Sd=99.48$  ng); whereas for the low group the mean Cf-DNA yield was 12.8 ng ( $Sd=13.69$  ng) ( $p\text{-value} = 0.3352$ ). The median Cf-DNA yield for the High group was 8.545 ng ( $IQR=17.79$  ng); whereas for the low group, the median Cf-DNA yield was 6 ng ( $IQR=6.9975$ ng) ( $p\text{-value} = 0.21$ ).

Using multivariate analysis by Induction day; D22 yield was significantly different from each of D1 ( $p=0.021$ ), D3 ( $p=0.0091$ ) and D8 ( $p=0.018$ ) (**Table 4-5**).

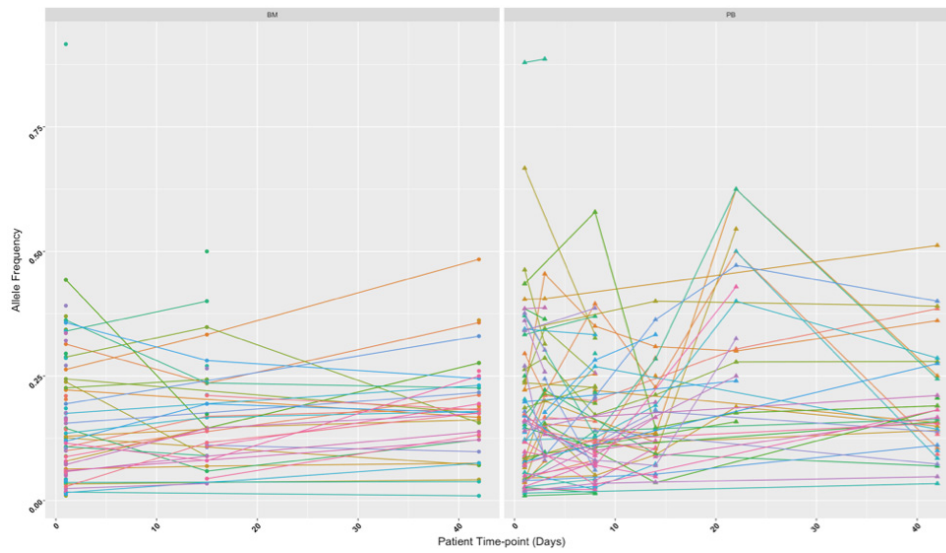
**Concordance between PB plasma Ct-DNA and WES in BM biopsy at diagnosis.** WES of bone marrow biopsy obtained at diagnosis reported 13 clinically relevant somatic SNV in 8 patients (**Table 4-6**). The MAF of SNVs detected in WES BM biopsy ranged from 6-74% (mean 37.2%; SD 17.13% and median of 30%; IQR 15%).

Ct-DNA analysis of diagnostic PB plasma detected 9 out the 13 mutations. The MAF of SNVs ranged from 2.22-41.27% (mean 24.9%; SD 19.6% and median of 27.6%; IQR 29.3%). PIK3CD was not included in our ALL custom gene panel nor were the specific exons of CCND3 or TRRAP in which the mutations were detected. For Case10, diagnostic PB plasma was not banked. Interestingly, analysis of Case10\_D15 plasma sample detected the mutation at an AF of 0.79%. Hence Ct-DNA analysis in PB plasma showed 100% concordance with BM biopsy.

In Ct-DNA; somatic SNVs in the RAS/MAPK pathway represented ~ 44% (3 NRAS and 1 KRAS mutation) of the detected mutations followed by JAK2 representing ~ 33% of the detected mutations.

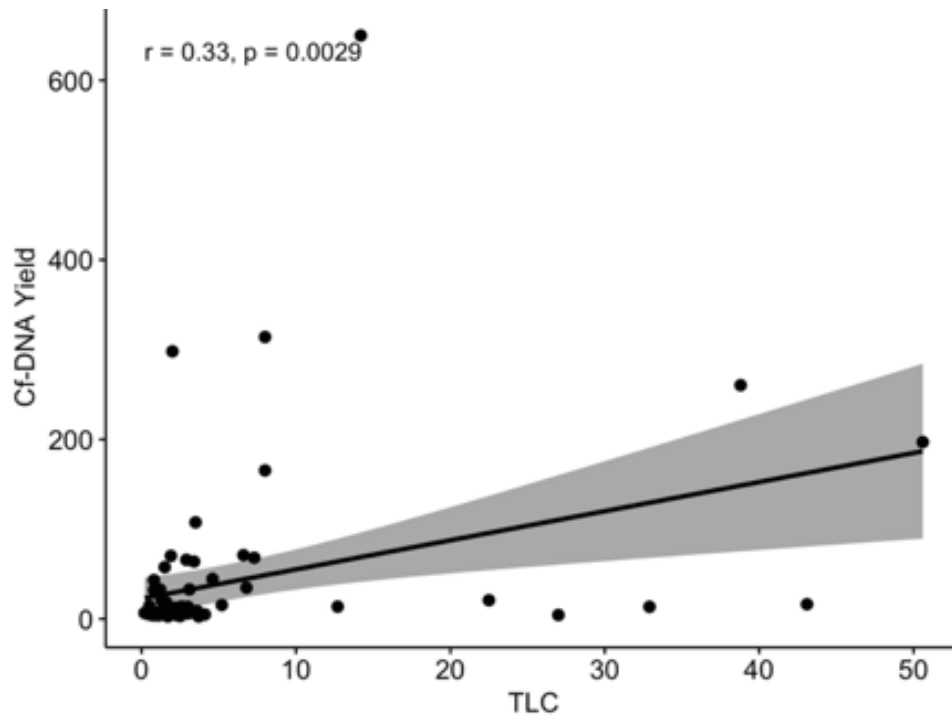
Regarding the correlation between MAF detected in PB plasma Ct-DNA and that detected in BM biopsy samples (whole exome clinical sequencing) at diagnosis, a moderate correlation was detected (Pearson coefficient  $r=0.53$ ). (**Figure 4-5** and **Figure 4-6**).





**Figure 4-3. Non-synonymous Mutations (Filtered Variant Calls) for Case#2 Whole Exome-captured Libraries**

(Personal communication from Ousman Mahmoud Charles Gawad Lab. St. Jude Children's Research Hospital on March 6th 2018).



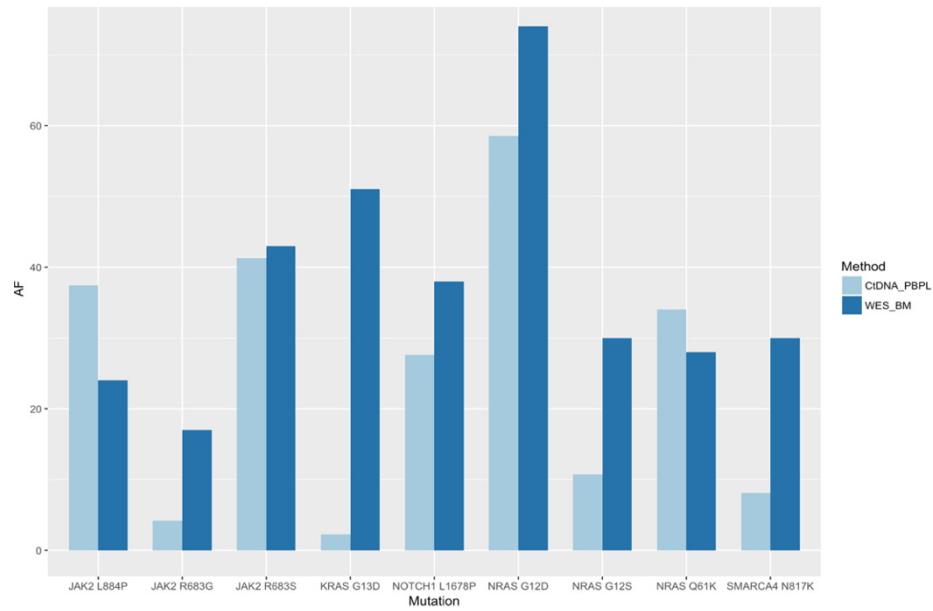
**Figure 4-4. Correlation Between TLC (cells\*10e9/L) and Cf-DNA Yield (ng)**

**Table 4-5. Median DNA Yield and TLC at Different Time Points During Remission Induction Therapy**

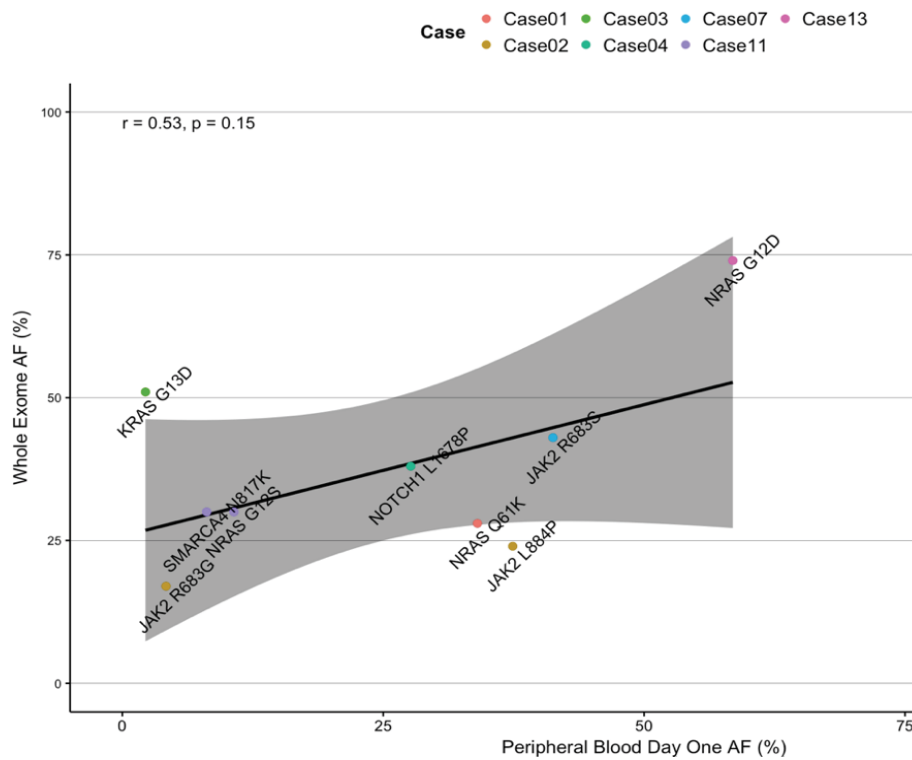
Induction Day	Median TLC (cells*10e <sup>9</sup> /L)	Median Cf-DNA Yield (ng)
D1	4.90	17.62
D3	2.90	32.94
D8	1.40	12.78
D15	1.30	7.02
D22	2.05	5.99
EOI	2.20	10.71

**Table 4-6. Clinically Relevant Single Nucleotide Somatic Variants MAF in BM Biopsy WES and PBPL Ct-DNA at Diagnosis**

Case#	Alteration	HGVS Nomenclature	WES BM Biopsy AF (%)	Ct-DNA PBPL AF (%)
Case 1	NRAS Q61K	NM_002524:c.181C>A; p.Gln61Lys	28	34.03
Case 2	JAK2 R683G	NM_004972:c.2047A>G; p.Arg683Gly	17	4.19
	JAK2 L884P	NM_004972:c.2651T>C; p.Leu884Pro	24	37.41
Case 3	KRAS G13D	NM_033360:c.38G>A; p.Gly13Asp	51	2.22
Case 4	CCND3 E236K	NM_001760:c.706G>A;p.Glu236Lys	33	-----
	NOTCH1 L1678P	NM_017617:c.5033T>C;p.Leu1678Pro	38	27.64
	PIK3CD E1021K	NM_005026:c.3061G>A;p.Glu1021Lys	45	-----
Case 7	JAK2 R683S	NM_004972:c.2049A>T; p.Arg683Ser	43	41.27
Case 10	KRAS A146V	NM_033360:c.437C>T; p.Ala146Val	6	-----
Case 11	NRAS G12S	NM_002524:c.34G>A; p.Gly12Ser	30	10.7
	SMARCA4 N817K	NM_001128849:c.2451C>A; p.Asn817Lys	30	8.09
	TRRAP E3107K	NM_003496:c.9319G>A; p.Glu3107Lys	27	-----
Case 13	NRAS G12D	NM_002524:c.35G>A; p.Gly12Asp	74	58.5



**Figure 4-5. MAF in Diagnostic BM Biopsy (WES) and PB Plasma Ct-DNA at Diagnosis**



**Figure 4-6. Correlation Between MAF in BM Biopsy (WES) and PB Plasma Ct-DNA at Diagnosis**

**Concordance between PB plasma Ct-DNA and BM plasma Ct-DNA at diagnosis.** For the 9 cases in which clinically relevant mutations were detected, only 2 (Case#2 and #3) had BM plasma analyzed at diagnosis. The MAF of the 3 mutations detected in Ct-DNA of PB plasma and BM plasma showed high concordance (Pearson's coefficient  $r=1$ ;  $p\text{-value}=0.0099$ ) (**Figure 4-7** and **Figure 4-8**).

**Monitoring clonal dynamics and disease response throughout remission induction therapy.** For those SNVs shared between diagnostic BM biopsy samples (WES) and PB plasma, Ct-DNA analysis in sequential plasma samples successfully captured the change in MAF across multiple induction time points. For all cases, the mutations were still detected in D3 plasma sample at a comparable or higher level for most cases. For all 7 cases, none of the SNVs re-emerged at a later time point after being undetected (**Table 4-7**).

**Comparison of Ct-DNA analysis and FC-MRD across multiple time points.** For cases in which mutations detected in clinical WES of BM biopsy, comparing MRD detection as a categorical variable (i.e. FC-MRD positivity/negativity and detection/absence of the initially specified mutation in Ct-DNA) at different induction time points (**Table 4-8**) demonstrated the following:

**D8 Induction.** FC-MRD in PB demonstrated 50% concordance with PB plasma Ct-DNA.

**D15 Induction.** FC-MRD in BM demonstrated 50% concordance with PB plasma Ct-DNA.

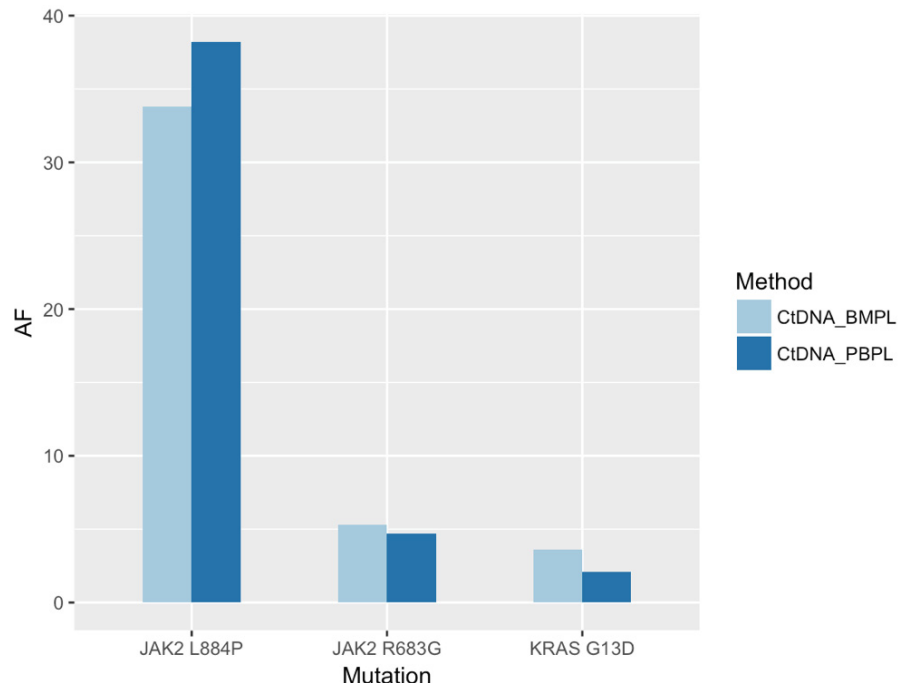
**D22 Induction.** FC-MRD in BM showed 100% concordance PB Ct-DNA; Ct-DNA analysis successfully captured mutations in the two FC-MRD positive cases.

**EOI.** All samples were negative by both techniques.

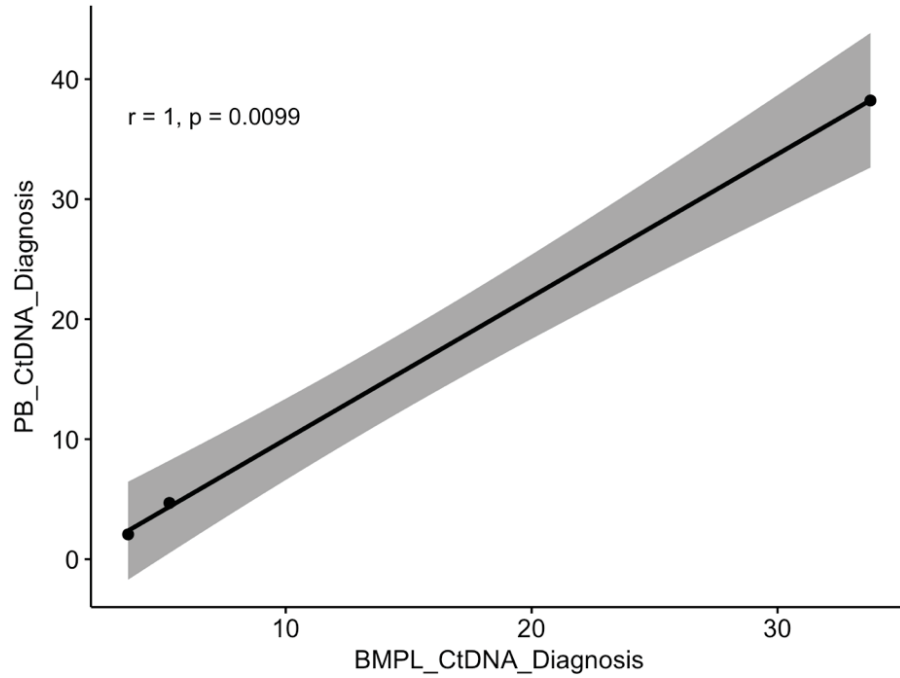
Detailed description by case is provided below:

**Case 1** Ct-DNA analysis of case 1 (B-cell hyperdiploid) demonstrated the presence of NRAS Q61K mutation at diagnosis at an allele frequency of 32.8% in PB plasma (BM plasma samples was not available for this case) in concordance with the BM-NGS clinical report. Reports previously detected the existence of RAS mutation at its highest prevalence in hyperdiploid ALL and were previously shown to confer resistance to methotrexate, prednisone; and sensitivity to VCR.

The change in FC-MRD and MAF in Ct-DNA followed the same pattern. In concordance with FC-MRD positivity on D15 and D22, NRASQ61K was detected (at a decreased allele frequency). At EOI, the patient achieved a negative FC-MRD and the NRAS Q61K was undetected at the assay threshold (**Figure 4-9A** and **B**).



**Figure 4-7. MAF (%) in BM Plasma and PB Plasma in Ct-DNA at Diagnosis**



**Figure 4-8. Pearson's Correlation Between MAF Detected in Ct-DNA PB Plasma and BM Plasma**

**Table 4-7. MAF in Ct-DNA Across Multiple Induction Time Points**

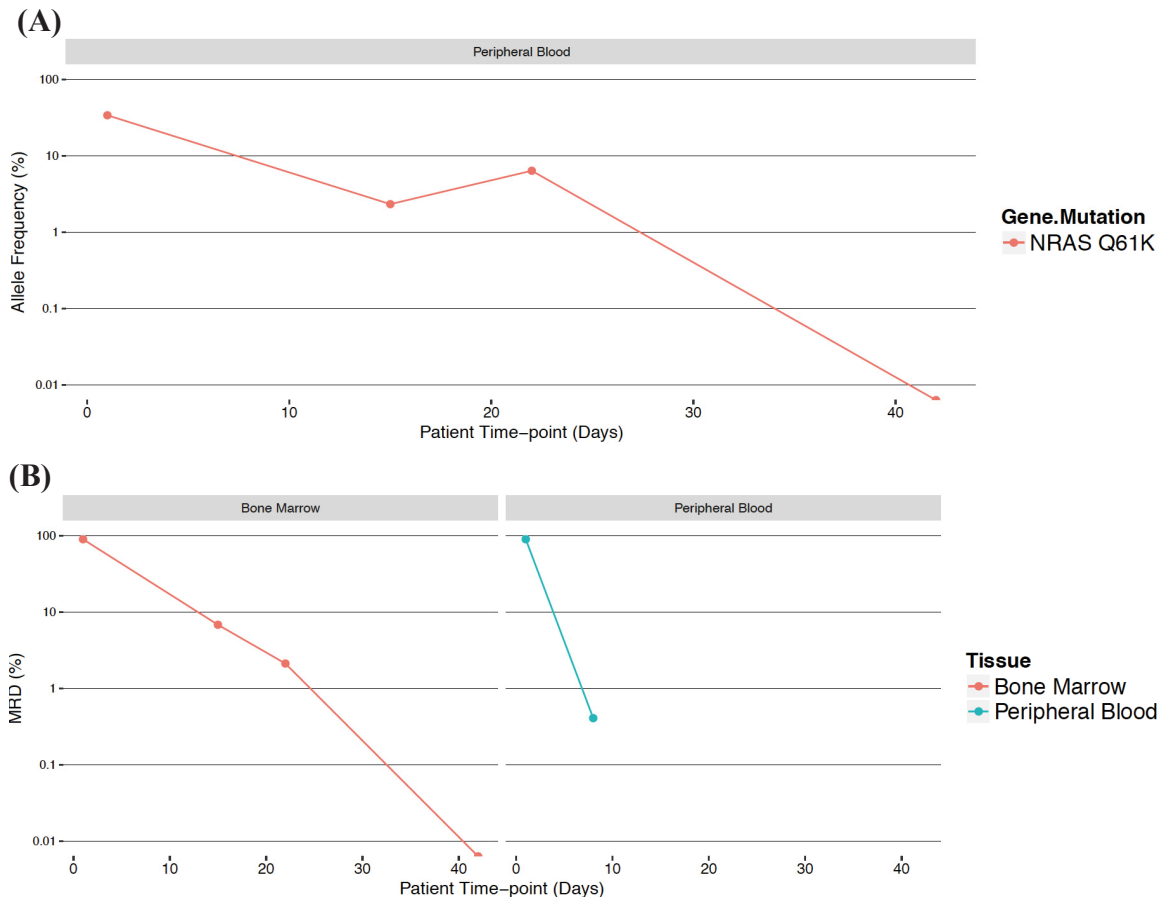
Case#	Mutation	MAF (%)					EOI
		D1 Ind.	D3 Ind.	D8 Ind.	D15 Ind.	D22 Ind.	
Case1	NRAS Q61K	34.03	-----	-----	2.33	6.38	0
Case2	JAK2 R683G	4.19	2.48	2.67	0	0	0
	JAK2 L884P	37.41	37.80	0	0	-----	0
Case 3	KRAS G13D	2.22	3.16	43.38	0	0	0
Case4	NOTCH1 L1678P	27.64	11.41	3.33	4.76	0	0
Case 7	JAK2 R683S	41.27	38.01	15.29	0.98	1.13	0
Case 11	NRAS G12S	10.70	44.71	0	0	0	0
	SMARCA4 N817K	8.09	43.69	0	0	0	0
Case 13	NRAS G12D	58.50	48.74	24.16	0	0	0

Only cases with mutations detected in clinical WES of BM biopsy are included.

**Table 4-8. FC-MRD Detection at Multiple Induction Time Points**

Case#	MRD (%)			
	D8 PB	D15 BM	D22 BM	EOI BM
Case 1	0.408	6.82	2.12	<0.01
Case 2	0.071	0.01	NA	<0.01
Case 3	0.013	2.15	<0.01	<0.01
Case 4	<0.01	<0.01	NA	<0.01
Case 7	0.738	1.502	0.421	<0.01
Case 11	0.018	0.014	NA	<0.01
Case 13	<0.01	<0.01	NA	<0.01

Only cases with mutations detected in clinical WES of BM biopsy are included.



**Figure 4-9. Case#1 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in NRAS Q61K MAF in PB plasma Ct-DNA throughout induction.

(B) FC-MRD in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 28<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).

**Case 2.** Samples for all the study specified time points were analyzed. At diagnosis, the two JAK2 mutations (JAK2 L884P and JAK2 R683G) were detected in both BM plasma and PB blood plasma samples, at comparable allele frequency. On D8 Induction, JAK2 R683G was still detected, coinciding with a positive PB FC-MRD. However, D15 Ct-DNA analysis revealed no mutations in either BM plasma or PB plasma, whereas FC-MRD was still detected at a borderline value of 0.01%. At end of induction, both assays were concordant showing a negative result (**Figure 4-10A and B**).

**Case 3.** At diagnosis, Ct-DNA analysis of this case (B-cell hyperdiploid) detected KRASG13D in BM plasma and PB plasma at diagnosis. The MAF in BM biopsy sample was 51% whereas Ct-DNA captured the same mutation at an allele frequency of 2.2%. KRASG13D was not detected in D15 samples (BM plasma and PB plasma) whereas FC-MRD was positive (2.15%).

Other mutations detected in the plasma samples at different time points require further computational analysis work to confirm them as true variants. Taking into consideration those mutations, Ct-DNA analysis showed concordance with FC-MRD, being detectable on both D8 and D15.

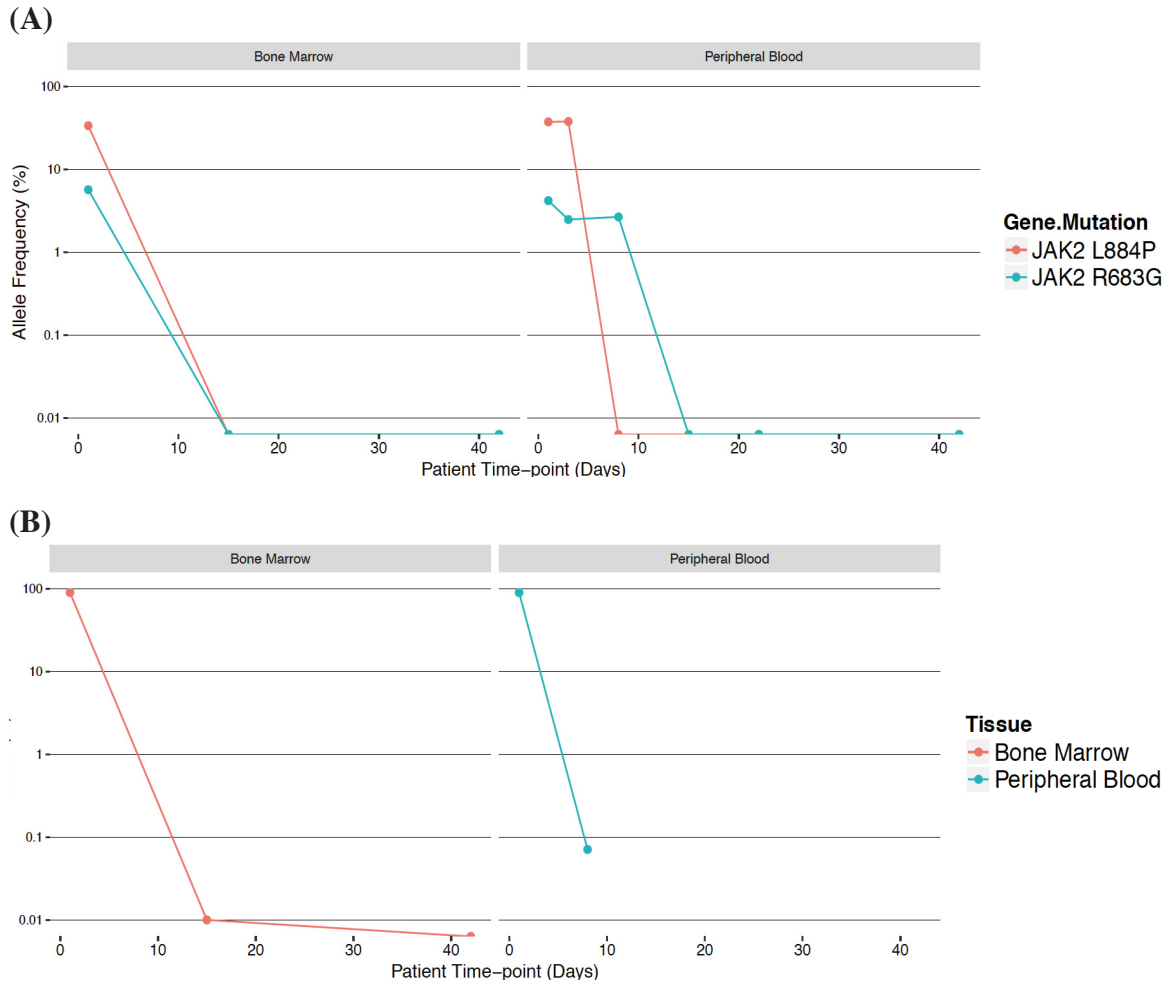
Additionally, two of the captured mutations remained positive at D22 and EOI in discordance with FC-MRD (**Figure 4-11A and B**).

**Case 4.** This patient comprises a T-cell ALL CNS-positive patient. At diagnosis, PB Ct-DNA analysis revealed KRAS G12D and NOTCH1 L1678P mutations (diagnostic BM plasma sample was not banked). Interestingly; Negative FC-MRD (<0.01%) was detected in D8 PB, D15 BM and at end of induction; whereas, Ct-DNA analysis detected NOTCH1 L1678P in D8 and D15 PB plasma and KRAS G12D in D15 BM plasma and PB plasma. Later during the course therapy, this patient experienced CNS relapse and an increase in BM FC-MRD (0.003%) (**Figure 4-12A and B**).

**Case 7.** Ct-DNA analysis detected a single JAK2 mutation (JAK2 R683S) in PB plasma at diagnosis (diagnostic BM plasma sample was not banked). The MAF of the aforementioned mutation was comparable to that detected in diagnostic BM biopsy by clinical WES. The patient achieved a negative FC-MRD (<0.01%) as well as, an undetectable JAK2R683S only by the end of induction hence, all other measured points were considered positive by both techniques. However, FC-MRD decreased between D15 and D22 whereas the MAF in Ct-DNA showed a slight increase between the same time points.

Ct-DNA detected KRAS G12D with changing allele frequency across different time points for which further computational analysis is needed, to determine whether KRAS G12D is a false positive call or not. KRAS G12D was still detected at EOI (**Figure 4-13A and B**).



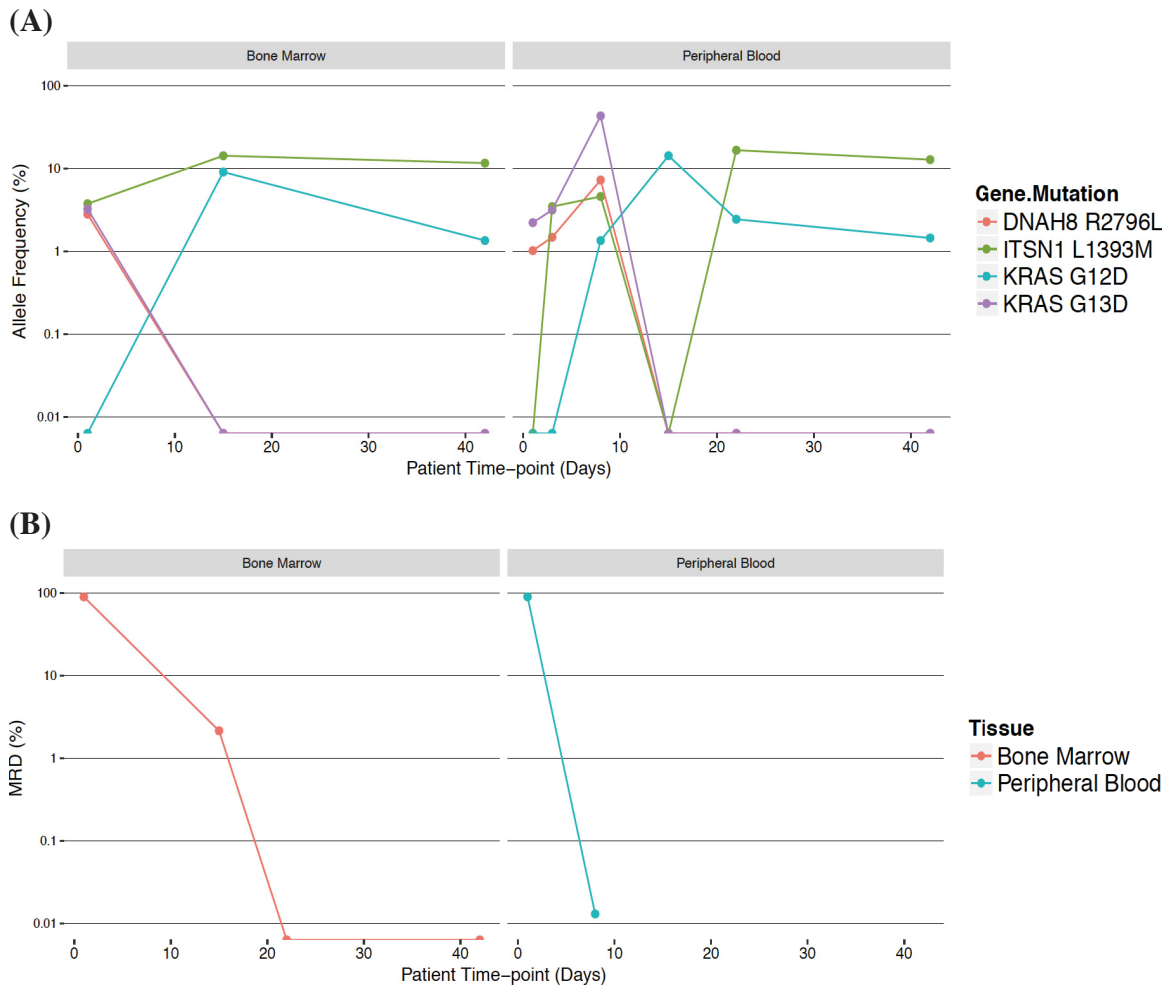


**Figure 4-10. Case#2 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in JAK2 MAF in Ct-DNA throughout induction in BM samples (left) and PB samples (right).

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 28<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).

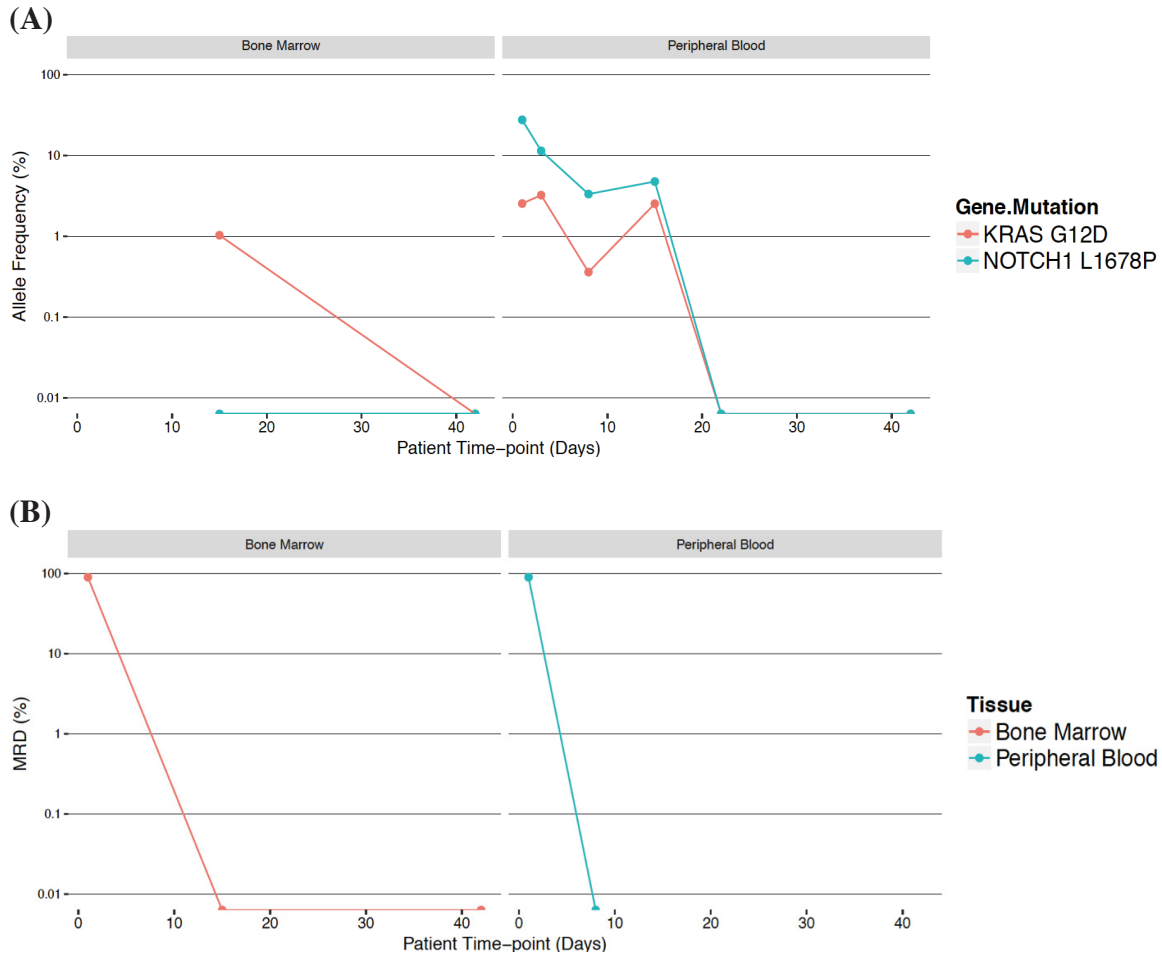


**Figure 4-11. Case#3 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in MAF in Ct-DNA throughout induction therapy in BM samples (left) and PB samples (right).

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 28<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).

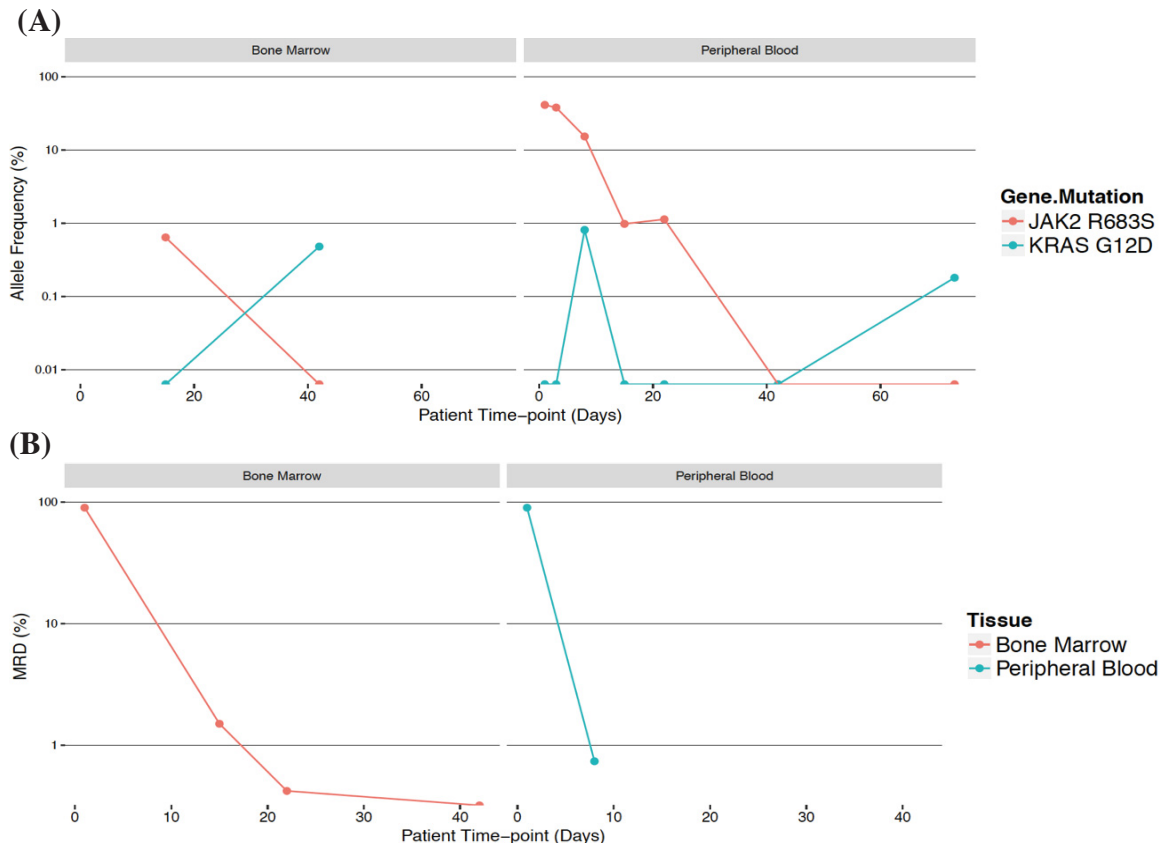


**Figure 4-12. Case#4 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in MAF in Ct-DNA throughout induction therapy in BM samples (left) and PB samples (right).

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 28<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).



**Figure 4-13. Case#7 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in MAF in Ct-DNA throughout induction therapy in BM samples (left) and PB samples (right).

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 28<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).

**Case 11.** The clinical WES detected 3 mutations in the diagnostic BM sample (NRAS G12S, SMARCA4 N817K and TRRAP E3107K). Ct-DNA analysis in PB plasma detected NRAS G12S and SMARCA4 N817K (the exon in which the TRRAP mutation was detected was not included in our ALL customized gene panel). In discordance with FC MRD, both mutations were undetected on D8 and D15 induction by Ct-DNA analysis (**Figure 4-14A and B**).

**Case 13.** NRASG12D was detected in the PB plasma at diagnosis. Despite that peripheral blood FC-MRD on D8 was negative, NRASG12D mutation was detected in Ct-DNA at a relatively high allele frequency of 24%. On D15, both techniques showed a negative result (**Figure 4-15A and B**).

Single nucleotide variants captured in cases#5, 6, 8, 9, 12 and 14 require further analysis to discriminate true variants from false positives. (**Figure 4-16A and B through 4-21A and B**).

The following analysis considers the whole number of mutations detected in the studied in 13 patients (**Table 4-2 and Figure 4-9A and B through Figure 4-21A and B**). This data is liable to change upon confirming that the detected variants as true calls.

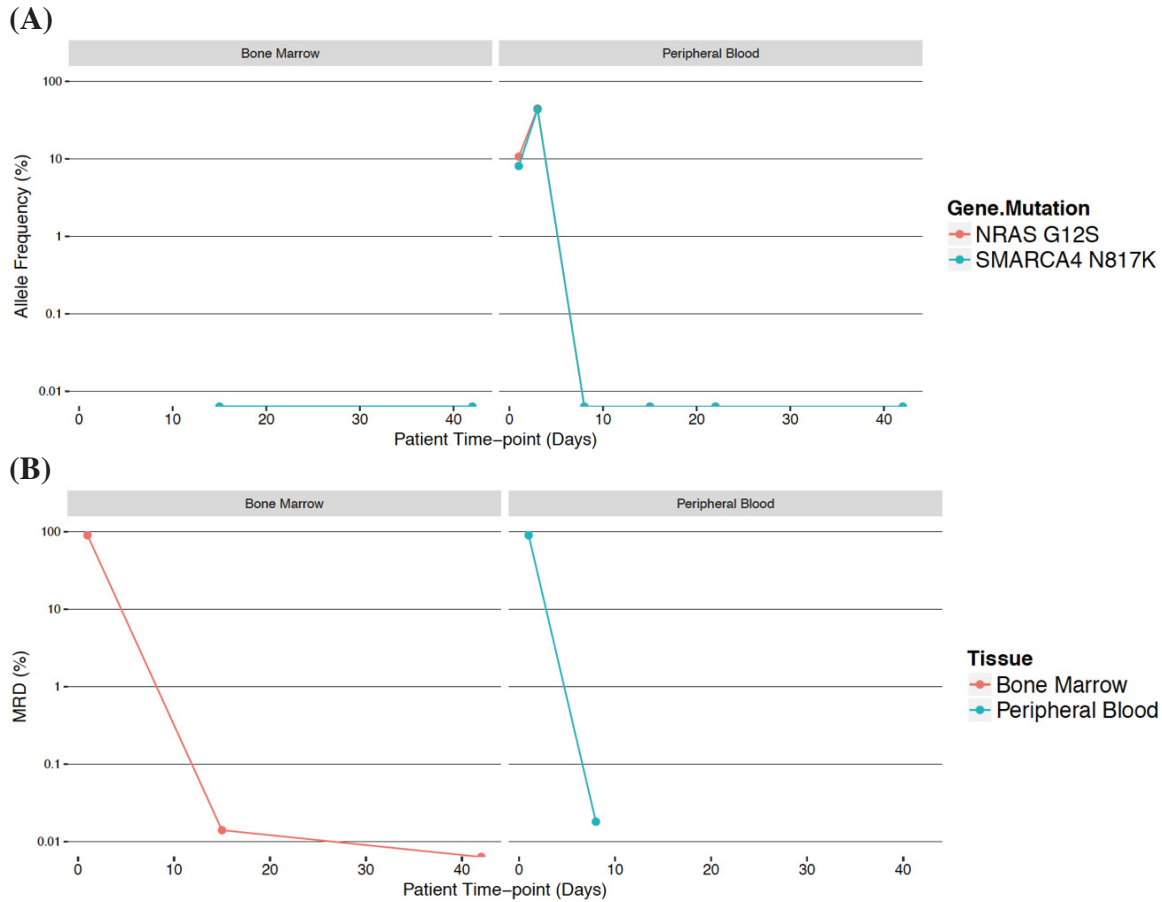
**D8 Induction.** Ten patients who had detectable FC-MRD on D8. Ct-DNA analysis was not done for one case at this time point, whereas Ct-DNA analysis in PB plasma showed a positive result in 8 out of 9 patients (~88.8%). Only 3 patients had MRD <0.01% by FC; in one of the two patients FC-MRD was still detected at a level of 0.003%. Ct-DNA analysis showed to be positive in the 3 patients detecting a range of 1-3 mutations.

**D15 Induction.** FC-MRD was evaluated in BM for the 13 cases. FC-MRD level was undetected (<0.01%) in 2 cases,  $\geq 0.01\%$ - < 1% in 6 cases and  $\geq 1\%$  in 5 cases.

In the first group of patients where FC-MRD <0.01%, Ct-DNA analysis and FC-MRD results were discordant. Ct-DNA analysis of BM plasma as well as, PB plasma showed a positive result for both cases.

In the group of patients whose FC-MRD was  $\geq 0.01\%$ - < 1%; 5 out of 6 patients had samples available for Ct-DNA analysis. For the 5 patients; Ct-DNA analysis was carried out in both BM plasma and PB plasma (10 samples). Ct-DNA showed a positive result only in 2 cases (40%). Ct-DNA analysis results were concordant between PB plasma and BM plasma in all 5 cases.

In patients with FC-MRD  $\geq 1$ , Ct-DNA in PB plasma analysis was positive in all of the 4/5 cases (80%) whereas Ct-DNA in BM plasma analysis was positive in the 5/5 cases (100%).

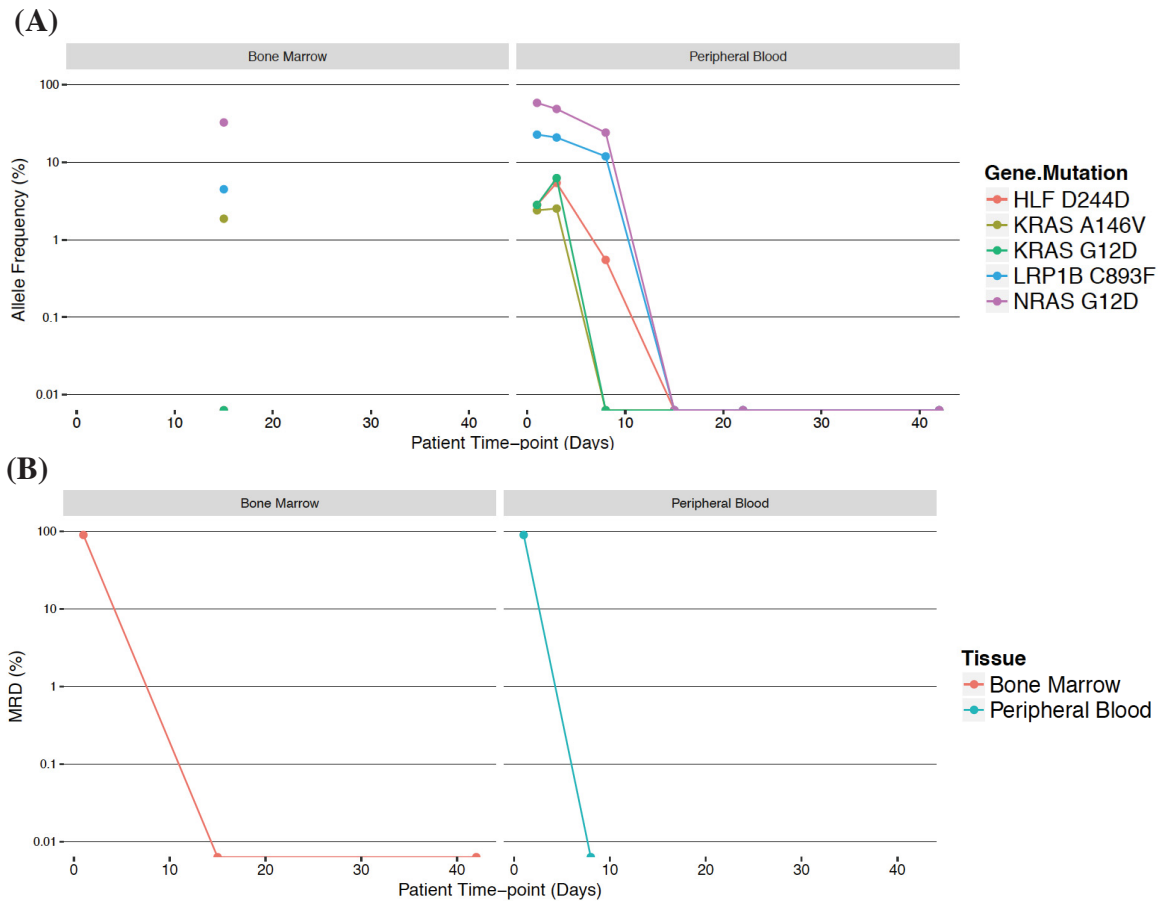


**Figure 4-14. Case#11 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in MAF in Ct-DNA throughout induction therapy in PB samples.

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 28<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).



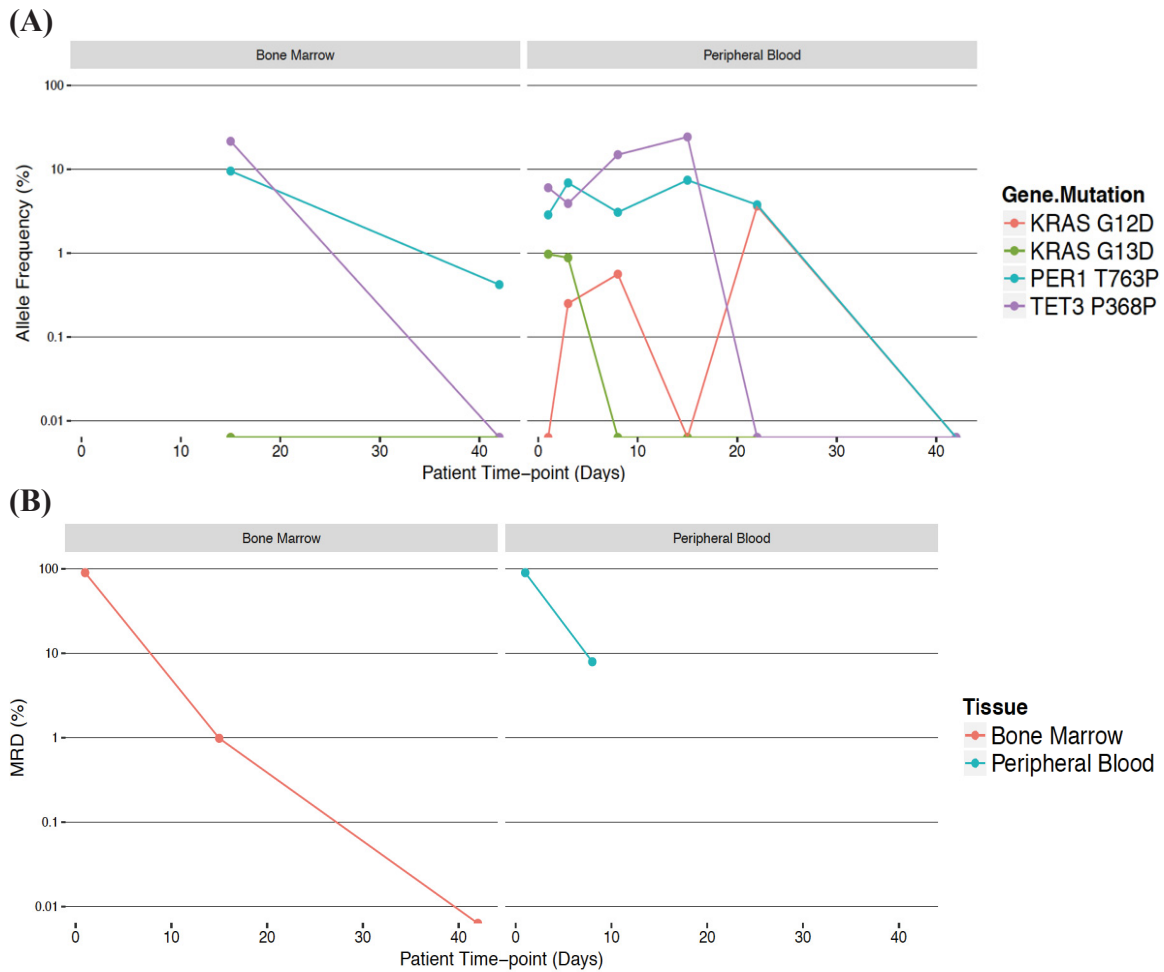
**Figure 4-15. Case#13 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) MAF in Ct-DNA in D15 BM sample (left), Change in MAF in Ct-DNA throughout induction therapy in PB samples (right).

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 28<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).



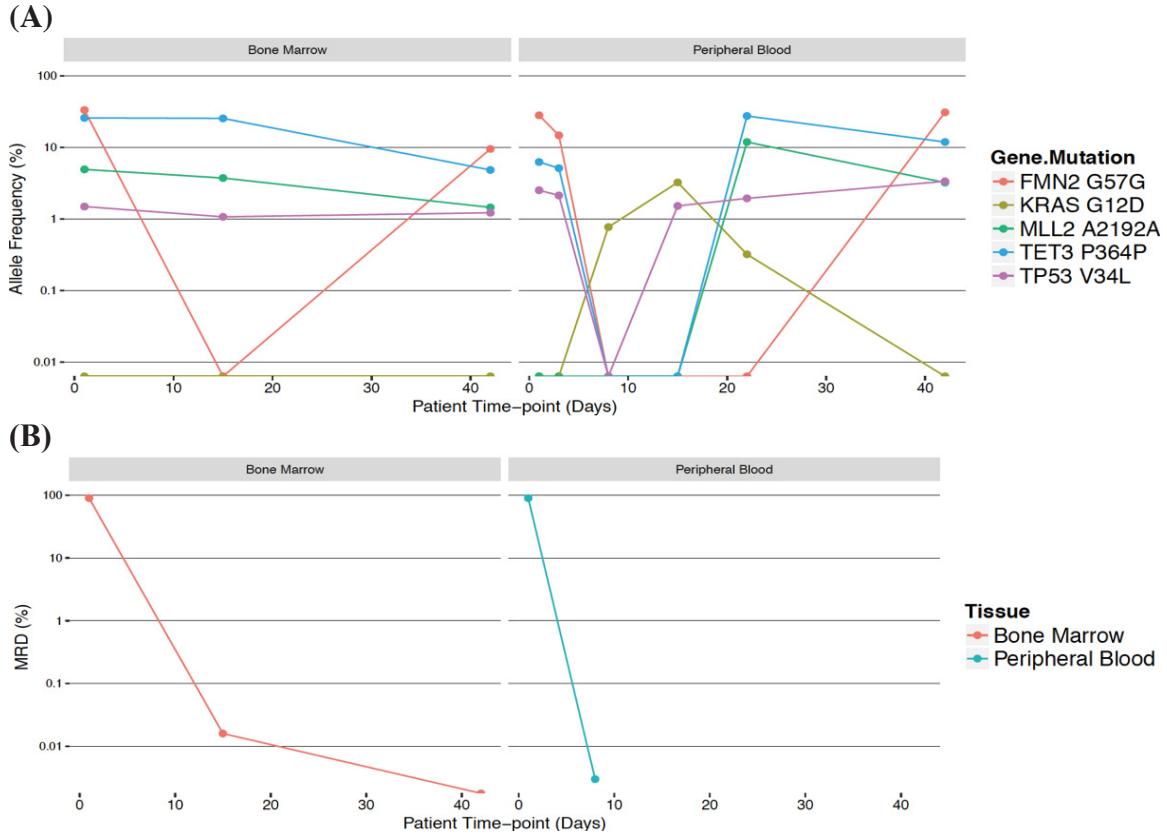


**Figure 4-16. Case#5 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in MAF in Ct-DNA throughout induction therapy in BM samples (left) and PB samples (right).

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 28<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).

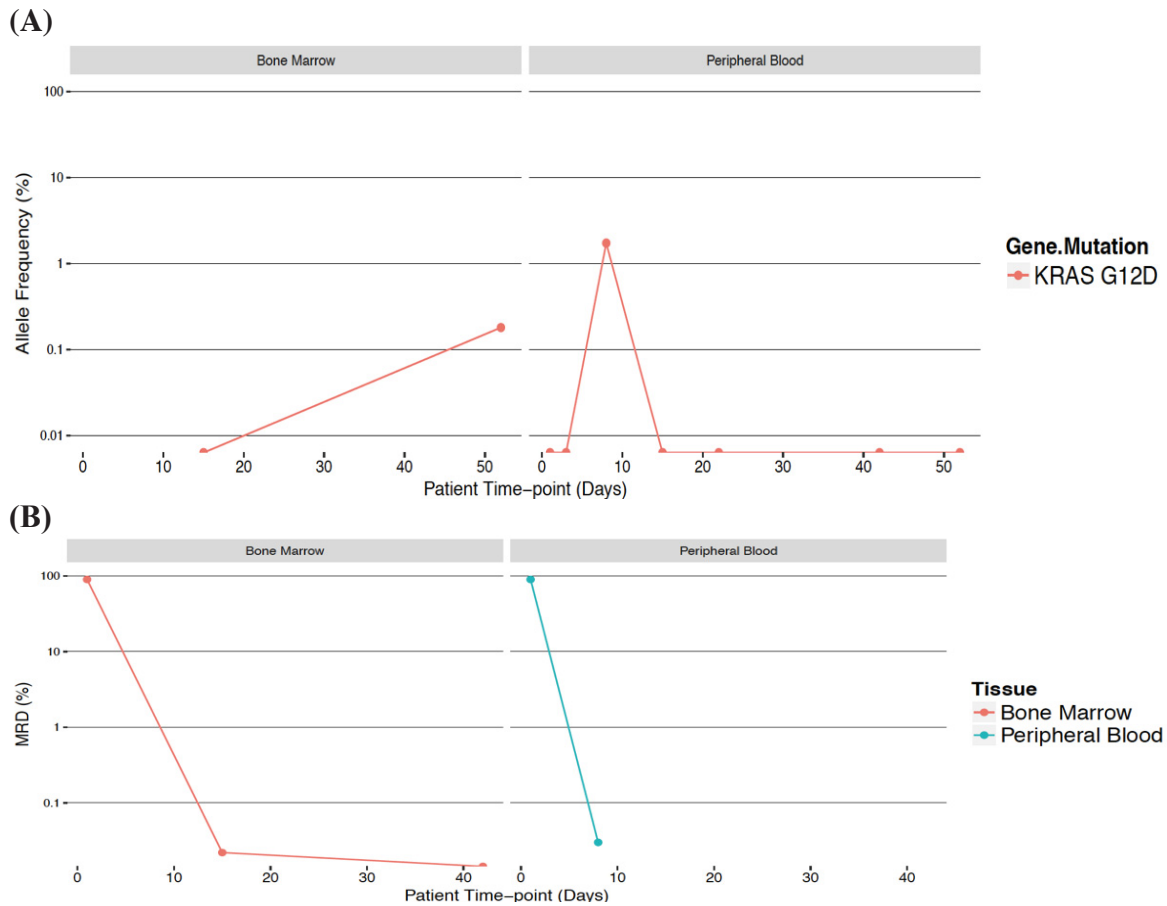


**Figure 4-17. Case#6 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in MAF in Ct-DNA throughout induction therapy in BM samples (left) and PB samples (right).

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 28<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).

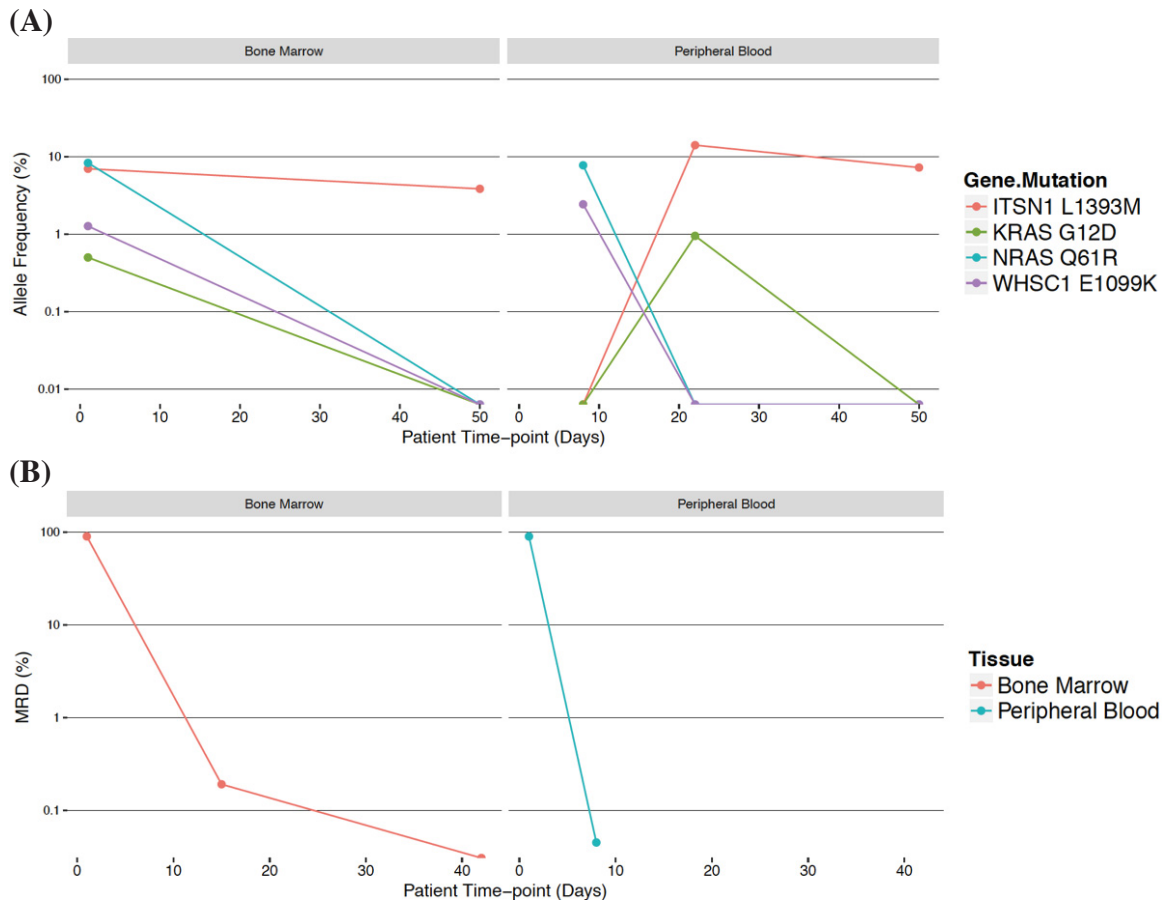


**Figure 4-18. Case#8 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in MAF in Ct-DNA throughout induction therapy in BM samples (left) and PB samples (right).

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 28<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).

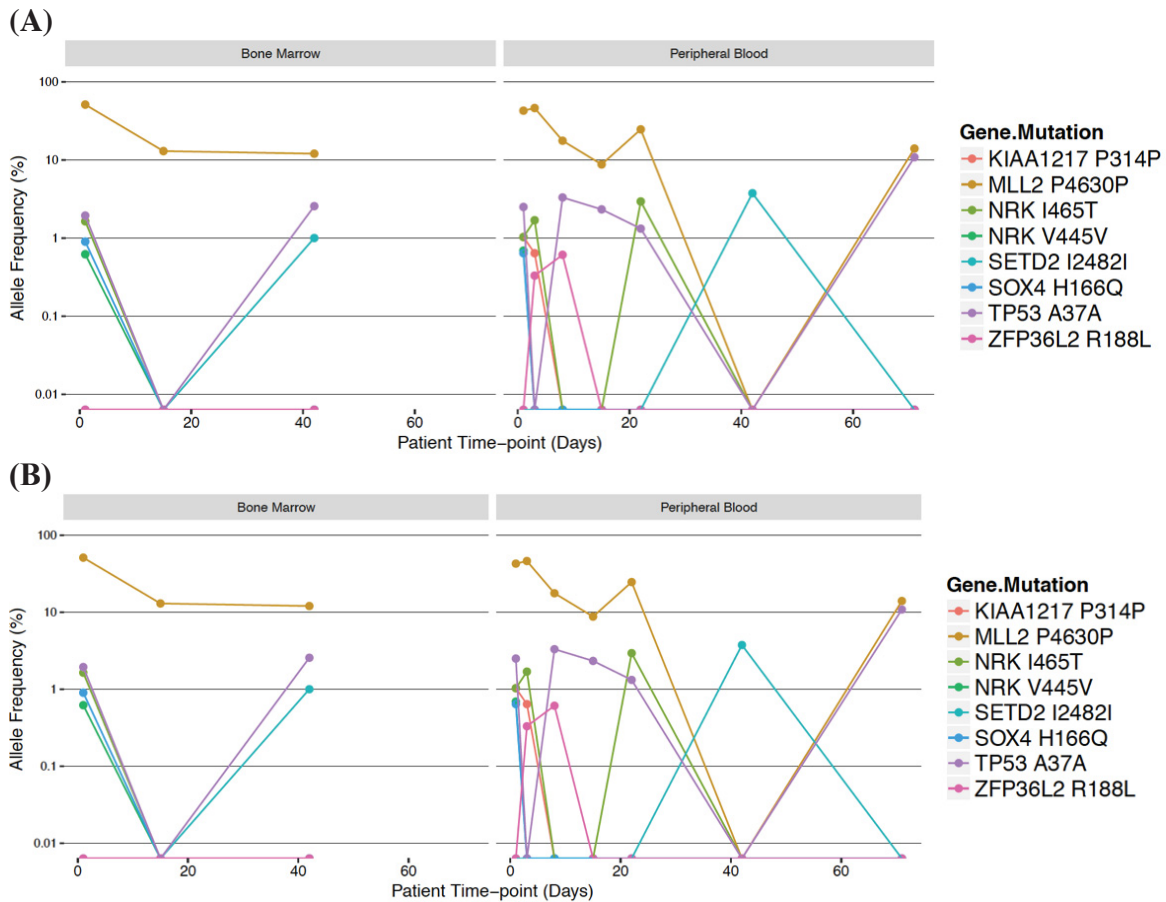


**Figure 4-19. Case#9 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in MAF in Ct-DNA throughout induction therapy in BM samples (left) and PB samples (right).

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 28<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).

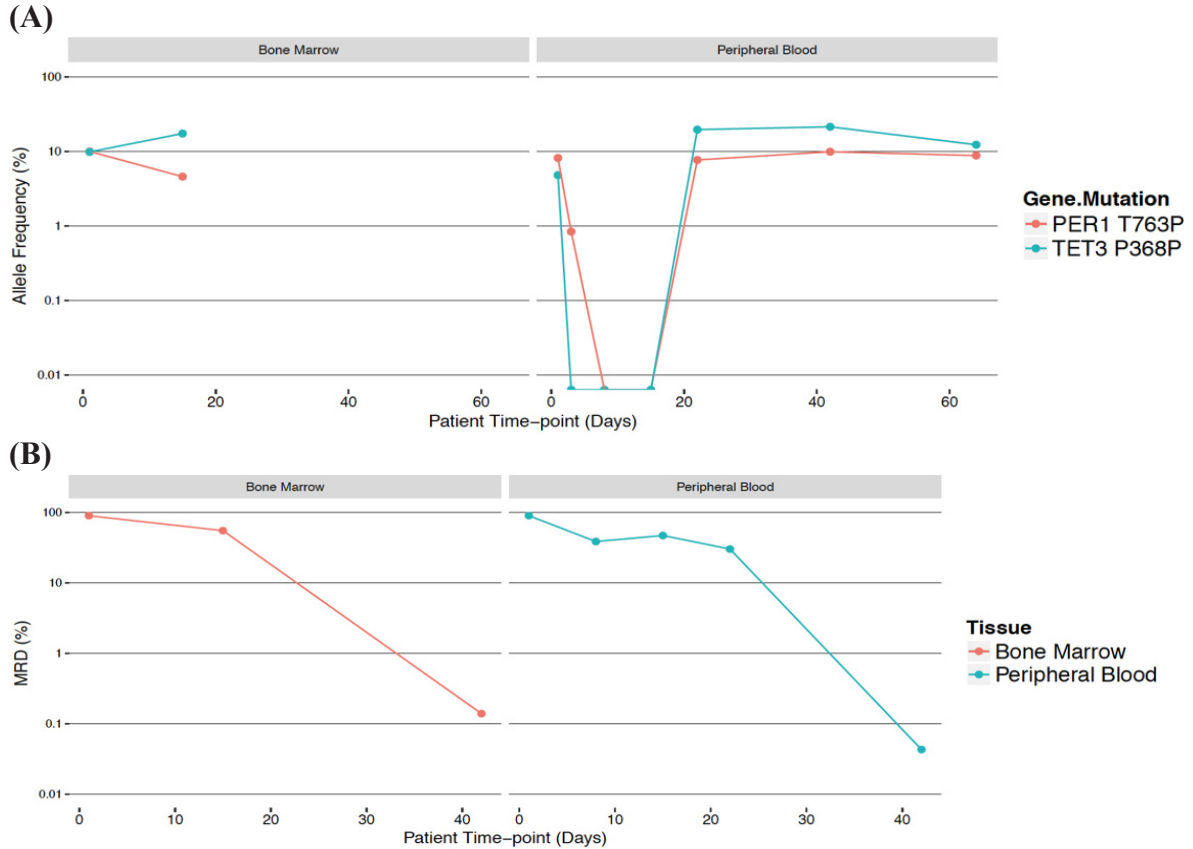


**Figure 4-20. Case#12 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in MAF in Ct-DNA throughout induction therapy in BM samples (left) and PB samples (right).

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 5<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).



**Figure 4-21. Case#14 Change in Ct-DNA MAF and FC-MRD Throughout Induction Therapy**

(A) Change in MAF in Ct-DNA throughout induction therapy in BM samples (left) and PB samples (right).

(B) Change in FC-MRD throughout induction therapy in BM samples (left) and PB samples (right).

(Personal communication with Ousman Mahmoud on 12<sup>th</sup> March 2018, Charles Gawad Lab. St. Jude Children's Research Hospital).

***End of induction (EOI).*** At end of induction, 12/13 patients achieved FC-MRD negativity (<0.01%). In 6/12 cases (50%). Ct-DNA analysis showed a positive result in either BM plasma (2 cases) or in both BM plasma and PB plasma (4 cases). Ct-DNA analysis of PB plasma for case 14 showed a positive result in concordance with FC-MRD positivity (BM plasma was not analyzed).

## CHAPTER 5. DISCUSSION

Clonal evolution represents a dynamic process; in which the genetic landscape continuously changes in response to changes in the evolutionary pressure (i.e. therapy selective pressure). [22] Discovering the extent of intratumor heterogeneity has far been limited by the focus on large chromosomal aberrations to decipher the tumor clonal structure as well as, the reduced sensitivity of the experimental techniques which lacked the ability to detect smaller subclones. The introduction of second generation sequencing has allowed for the deciphering of subclonal somatic mutations leading to a number of remarkable studies that demonstrated the ALL relapse follows a pattern of branched clonal evolution; linear evolution is rarely involved. [54] However, in most cases, studies addressing the clonal dynamics in ALL relapse compared matched diagnosis/relapse/ $\pm$  remission BM samples. In the exceptional cases, where relapse mutations (in few patients) were backtracked to earlier remission time points [11], it was demonstrated that an increase in MAF of relapse-mutations preceded clinical relapse and that subclonal mutations persisting post induction therapy comprise clonal mutations at relapse. [7, 11]

Despite the seminal findings of these studies; techniques currently implemented for the measurement of minimal residual disease lack the ability to capture the genetic heterogeneity; providing a quantitative rather than a qualitative description of the change in MRD. The ability to unveil the mutational profile and clonal dynamics and composition in serial MRD samples would complement the analysis; hence improving its predictive value. Moreover, the detection of drug-resistance variants in MRD samples, will have a direct clinical implication in tailoring therapy.[22] Ct-DNA analysis in sequential plasma samples have proved utility in mutational profiling and monitoring change in clonal dynamics in various solid and liquid malignancies.

This study represents the first proof-of-concept study to assess the utility of Ct-DNA in tracking tumor dynamics in pediatric ALL. Initially, WES sequencing was employed using a hybrid capture technique utilizing a whole exome gene panel targeting 19,396 genes and covering 51Mb of end-to-end tile probe space. Several studies have shown the utility of WES in tracking mutational change in Ct-DNA in plasma of patients with different types of cancer. However, those studies evaluated patients with advanced solid tumors [130]; in a seminal study using WES, Ct-DNA was quantified using ddPCR and Tam-Seq. and only patients with high allele frequency were selected for WES. In our study however, the mean total yield of Cf-DNA from the samples ranged from 6ng to 32 ng on D22 and D3 respectively. WES detected a large number of SNVs at a low MAF, was associated with a high false positive error rate, rendering true SNVs calling problematic.

Next, our group developed an ALL custom gene panel to analyze Ct-DNA in sequential peripheral blood and matched BM plasma samples, obtained at specified time points during induction therapy. The study evaluated 14 pediatric ALL patients of different B-cell and T-cell subtypes. The ALL customized gene panel was designed by the lab. computational biologist to target 806 genes covering  $\cong$  1668 exons and 319.4 kb.



The specific ALL custom gene panel designed includes genes with at least 3 mutations in SiTH database (derived from app. 500 patients). Mutations selected are somatic non-synonymous SNVs found in T-ALL and B-ALL diagnostic and relapse samples.

Genes included showed to overlap with at least one of the following categories (1) Oncogenic pathway (2) Tumor suppressor genes (3) Relapse-specific genes (4) Drug metabolism (5) CRISPR/Cas9 screening (unpublished data) (6) Genes associated with glucocorticoid resistance (7) xGen® Pan-Cancer Panel (which includes list of genes derived from TCGA; known to be mutated across multiple cancer types).

Implementation of CAPP-Seq. in NSCLC was associated with artifacts in 50% of the positions sequenced, below a detection limit of 0.02%. In their efforts to suppress the error associated with CAPP-Seq., molecular barcoding was employed and successfully reduced the error.[123] Hence, we additionally used index barcoding for all target captured cases.

The cases enrolled in this pilot study comprise different ALL subtypes; 11 cases are BCP and 3 cases are T-cell; one of which with ETP. According to NCI risk criteria, 36.4% (4 cases) of the BCP were high-risk whereas 63.6% (7 cases) were standard risk. According to Total XVII risk groups, 3 patients were stratified as low-risk, 2 patients as high-risk whereas 9 patients comprised the standard- (intermediate) risk group. In Total study XVII, NGS is implemented as a diagnostic clinical standard-of-care for all patients enrolled in the protocol. WGS, WES and transcriptome sequencing are carried by the clinical pathology lab.; genetic alterations with relevance to diagnosis or management are reported in the patient clinical record. Total XVII therapy implies Flowcytometry based MRD for response assessment during induction therapy and throughout treatment. During induction, MRD is evaluated in BM on D15, D22 (in patients with D15 MRD  $\geq 1\%$ ) and at end of induction (D42); whereas MRD evaluation in PB is carried out on D8. Sequential analysis of Ct-DNA in plasma was employed at equivalent time points; BM-plasma was evaluated on D1, D15 and at end of induction; whereas PB-plasma was evaluated on D1, D3, D15, D22 and at end of induction.

Total XVII induction therapy utilizes a 4-drug regimen during early induction (prednisone, daunorubicin, vincristine and peg-asparaginase) followed by cytarabine, cyclophosphamide and 6-mercaptopurine. Rituximab, bortezomib and dasatinib are administered to select patients (**Appendix C. Total XVII Remission Induction Treatment Schedule**). Hence, the study of somatic mutations in Ct-DNA offers a venue for monitoring the change in clonal composition and dynamics in response to the multi-agent regimen applied.

The first question was whether Ct-DNA analysis would capture those mutations detected in BM biopsy samples or not. NGS (WES and WGS) reported 13 clinically relevant mutations in the diagnostic BM samples of 8 out of 14 patients. Somatic mutation detection in peripheral plasma Ct-DNA showed 100% concordance with WES of BM biopsy at diagnosis. The ability of Ct-DNA analysis to capture all the detected

clinically relevant mutations utilizing  $\cong 1$  ml of plasma demonstrates the utility of Ct-DNA application in ALL.

Next, we evaluated the correlation between the MAF detected by both techniques. Studying the allele frequency of the 9 mutations detected in both BM biopsy and plasma samples yielded a moderate correlation (Pearson's coefficient = 0.53). The discrepancy can be attributed to a number of general factors namely, the low amount of Cf-DNA extracted (Median  $\sim 17$  ng) together with the fact that Cf-DNA is derived from various sources. Additional factors related to our study cohort that might have led to the modest correlation between MAF detected in BM WES and PB plasma Ct-DNA, are the small number of patients with detectable mutations and the presence of more than one mutation per sample (i.e. case 11 WES shows 3 mutations, each at an AF of  $\sim 30\%$  whereas plasma sample showed only 2 out of the 3 mutations at a much lower AF hence, one sample contributed to 22% of the dataset). Importantly, this discrepancy might in fact reflect a real difference in the MAF between different body compartments hence, reflecting tumor heterogeneity.

Additionally, Ct-DNA MAF was evaluated throughout the aforementioned time points and change in MAF was studied in relation to FC-MRD at matched time points. Although the study has not revealed the expected correlation between MAF and the quantitative level of FC MRD; Ct-DNA and FC-MRD were qualitatively concordant for 50% of samples on D8 and D15 and 100% concordant at D22 and EOI.

Regarding the 7 cases in which mutations were shared between BM biopsy and PB Ct-DNA; 2 cases (Case 1 and Case 7) showed concordance between FC-MRD and Ct-DNA across all time points.

Two cases (Case 2 and Case 11) had no mutations detected in Ct-DNA at FC-MRD positive time points. Case 11 harbored NRAS G12S, SMARCA4 N817K and TRRAP E3107K as detected by clinical WES of the diagnostic BM biopsy. TRAPP exon in which the mutation was detected was not included in our ALL custom gene panel. Hence, the inability of Ct-DNA to track TRAPP mutation detected in the BM biopsy sample might offer an explanation i.e. the positive FC-MRD might be attributed to the clone harboring the TRAPP mutation. Further analysis for the samples including that TRAPP exon might pose an answer. Also, the improvement of the analysis pipeline for capturing true positive calls might reveal other mutations that were present in Case 2 yet filtered out with the currently applied analysis pipeline.

For case 3, although the KRAS G13D was undetected on D15 in discordance with a positive MRD (2.15%), KRAS G12D mutation was detected at an allele frequency of 15.9%. Importantly, KRAS G12D was not detected in the PB plasma nor the BM plasma diagnostic samples; yet it was initially detected in D8 PB plasma. KRAS G12D detected in PB plasma as well as BM plasma on D15 and persisted till end of induction (at a lower allele frequency). ITSN1 L1393M was also detected at different allele frequencies across multiple time points and persisted till end of induction. Although further confirmation of these mutations as true positive is required, they might provide an

explanation for D15 discordance between FC-MRD and Ct-DNA on D15. Additionally, the fact that such mutations were still detected at EOI whereas the FC-MRD was negative. The aforementioned findings show the ability of the technique to unveil the clonal composition in FC-MRD positive patients as well as, its sensitivity to capture mutations at lower MAF in FC-MRD negative patients.

Case 4 showed a negative FC-MRD on D15 however NOTCH1 L1678P (detected at diagnosis) was still detected by Ct-DNA analysis in both PB plasma and BM plasma. Hence, it provides a confirmation for the sensitivity of the technique to unravel the clonal composition and detect mutations at low mutation allele frequency in FC-MRD negative patients. The same is true for Case 13 where NRASG12D mutation was detected in Ct-DNA at a relatively high allele frequency of 24% on D8 whereas the FC-MRD was negative.

The different scenarios mentioned above demonstrate a potential role for sequential analysis Ct-DNA as a complementary technique to FC MRD; providing an insight into the clonal composition of the MRD and consequently guiding treatment decisions. The non-invasiveness of the technique as well as the relative low cost further supports the utility of incorporation in the clinical management.

Ct-DNA has a potential therapeutic implication in guiding therapy. Despite the small number of patients studied, Ct-DNA successfully captured clinically relevant mutations at diagnosis as well as throughout induction therapy. As such; several future therapeutic options might be incorporated into clinical practice. Case 2 provides a good example; as noted in the clinical sequencing report, the assay (WES) lacks capability to detect whether the 2 JAK mutations exist in cis or trans yet; it is important to consider that the presence of L884P mutation confers resistance to type 2 JAK2 inhibitors in CRLF2/JAK2 R683G cells (case\_2 is CRLF2 positive). In such case, combination of type I/II JAK2 inhibitors might offer a treatment option to overcome resistance.[131] Monitoring the relevant change in MAF of both mutations throughout therapy might provide a deeper insight into the possible therapeutic options.

Further work would be needed to study a larger number of patients and advance the technique and the analysis pipelines in order to correctly detect real mutations and decrease false positive calls.

## LIST OF REFERENCES

1. Ward, E., et al., *Childhood and adolescent cancer statistics, 2014*. CA Cancer J Clin, 2014. **64**(2): p. 83-103.
2. Hunger, S.P. and C.G. Mullighan, *Acute Lymphoblastic Leukemia in Children*. N Engl J Med, 2015. **373**(16): p. 1541-52.
3. Ma, X., et al., *Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia*. Nat Commun, 2015. **6**: p. 6604.
4. Campana, D., *Minimal residual disease monitoring in childhood acute lymphoblastic leukemia*. Curr Opin Hematol, 2012. **19**(4): p. 313-8.
5. Ding, L.W., et al., *Mutational Landscape of Pediatric Acute Lymphoblastic Leukemia*. Cancer Res, 2017. **77**(2): p. 390-400.
6. Meyer, J.A., et al., *Relapse-specific mutations in NT5C2 in childhood acute lymphoblastic leukemia*. Nat Genet, 2013. **45**(3): p. 290-4.
7. Kunz, J.B., et al., *Pediatric T-cell lymphoblastic leukemia evolves into relapse by clonal selection, acquisition of mutations and promoter hypomethylation*. Haematologica, 2015. **100**(11): p. 1442-50.
8. Dienstmann, R., et al., *Analysis of mutant allele fractions in driver genes in colorectal cancer - biological and clinical insights*. Mol Oncol, 2017. **11**(9): p. 1263-1272.
9. Mullighan, C.G., et al., *CREBBP mutations in relapsed acute lymphoblastic leukaemia*. Nature, 2011. **471**(7337): p. 235-9.
10. Landau, D.A., et al., *Clonal evolution in hematological malignancies and therapeutic implications*. Leukemia, 2014. **28**(1): p. 34-43.
11. Li, B., et al., *Negative feedback-defective PRPS1 mutants drive thiopurine resistance in relapsed childhood ALL*. Nat Med, 2015. **21**(6): p. 563-71.
12. Ding, L.W., et al., *Mutational profiling of acute lymphoblastic leukemia with testicular relapse*. J Hematol Oncol, 2017. **10**(1): p. 65.
13. Kubackova, V., et al., *Cell-free DNA - Minimally invasive marker of hematological malignancies*. Eur J Haematol, 2017. **99**(4): p. 291-299.
14. Mithraprabhu, S., et al., *Circulating tumour DNA analysis demonstrates spatial mutational heterogeneity that coincides with disease relapse in myeloma*. Leukemia, 2017. **31**(8): p. 1695-1705.
15. Short, N.J. and E. Jabbour, *Minimal Residual Disease in Acute Lymphoblastic Leukemia: How to Recognize and Treat It*. Curr Oncol Rep, 2017. **19**(1): p. 6.
16. Gawad, C., et al., *Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia*. Blood, 2012. **120**(22): p. 4407-17.
17. Yeh, P., et al., *Molecular disease monitoring using circulating tumor DNA in myelodysplastic syndromes*. Blood, 2017. **129**((12)): p. 1685-1690.
18. Siravegna, G., et al., *Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients*. Nat Med, 2015. **21**(7): p. 795-801.
19. Iacobucci, I. and C.G. Mullighan, *Genetic Basis of Acute Lymphoblastic Leukemia*. J Clin Oncol, 2017. **35**(9): p. 975-983.

20. Woo, J.S., M.O. Alberti, and C.A. Tirado, *Childhood B-acute lymphoblastic leukemia: a genetic update*. Exp Hematol Oncol, 2014. **3**: p. 16.
21. Paulsson, K., *Genomic heterogeneity in acute leukemia*. Cytogenet Genome Res, 2013. **139**(3): p. 174-80.
22. Ferrando, A.A. and C. Lopez-Otin, *Clonal evolution in leukemia*. Nat Med, 2017. **23**(10): p. 1135-1145.
23. Pui, C.H., *Genomic and pharmacogenetic studies of childhood acute lymphoblastic leukemia*. Front Med, 2015. **9**(1): p. 1-9.
24. Inaba, H., M. Greaves, and C.G. Mullighan, *Acute lymphoblastic leukaemia*. The Lancet, 2013. **381**(9881): p. 1943-1955.
25. Moriyama, T., M.V. Relling, and J.J. & Yang, *Inherited genetic variation in childhood acute lymphoblastic leukemia*. Blood, 2015. **125**(26): p. 3988-3995.
26. Perez-Andreu, V., et al., *Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse*. Nat Genet, 2013. **45**(12): p. 1494-8.
27. Mullighan, C.G., et al., *Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia*. Nature, 2007. **446**(7137): p. 758-64.
28. Holmfeldt, L., et al., *The genomic landscape of hypodiploid acute lymphoblastic leukemia*. Nat Genet, 2013. **45**(3): p. 242-52.
29. Shah, S., et al., *A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia*. Nat Genet, 2013. **45**(10): p. 1226-1231.
30. Busslinger, M., *Transcriptional control of early B cell development*. Annu Rev Immunol, 2004. **22**: p. 55-79.
31. Nebral, K., et al., *Incidence and diversity of PAX5 fusion genes in childhood acute lymphoblastic leukemia*. Leukemia, 2009. **23**(1): p. 134-43.
32. Strehl, S., et al., *PAX5/ETV6 fusion defines cytogenetic entity dic(9;12)(p13;p13)*. Leukemia, 2003. **17**(6): p. 1121-3.
33. Papaemmanuil, E., et al., *Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia*. Nat Genet, 2009. **41**(9): p. 1006-10.
34. Trevino, L.R., et al., *Germline genomic variants associated with childhood acute lymphoblastic leukemia*. Nat Genet, 2009. **41**(9): p. 1001-5.
35. Ellinghaus, E., et al., *Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia*. Leukemia, 2012. **26**(5): p. 902-9.
36. Sherborne, A.L., et al., *Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk*. Nat Genet, 2010. **42**(6): p. 492-4.
37. Enciso-Mora, V., et al., *Common genetic variation contributes significantly to the risk of childhood B-cell precursor acute lymphoblastic leukemia*. Leukemia, 2012. **26**(10): p. 2212-5.
38. Xu, H., et al., *Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations*. J Natl Cancer Inst, 2013. **105**(10): p. 733-42.
39. Migliorini, G., *Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype*. . Blood, 2013: p. 3298-3307.



40. Virginia Perez-Andreu, K.G.R., Heng Xu, Colton Smith, Hui Zhang, Wenjian Yang, Richard C. Harvey, Debbie Payne-Turner, Meenakshi Devidas, I-Ming Cheng, William L. Carroll, Nyla A. Heerema, Andrew J. Carroll, Elizabeth A. Raetz, Julie M. Gastier-Foster, Guido Marcucci, Clara D. Bloomfield, Krzysztof Mr'ozek, Jessica Kohlschmidt, Wendy Stock, Steven M. Kornblau, Marina Konopleva, Elisabeth Paietta, Jacob M. Rowe, Selina M. Luger, Martin S. Tallman, Michael Dean, Esteban G. Burchard, Dara G. Torgerson, Feng Yue, Yanli Wang, Ching-Hon Pui, Sima Jeha, Mary V. Relling, William E. Evans, Daniela S. Gerhard, Mignon L. Loh, Cheryl L. Willman, Stephen P. Hunger, Charles G. Mullighan, and Jun J. Yang, *A genome-wide association study of susceptibility to acute lymphoblastic leukemia in adolescents and young adults*. Blood, 2015. **125**(4): p. 680-686.
41. Ito, C., et al., *Hyperdiploid Acute Lymphoblastic Leukemia With 51 to 65 Chromosomes: A Distinct Biological Entity With a Marked Propensity to Undergo Apoptosis*. Blood, 1999. **93**: p. 315-320.
42. Paulsson, K., et al., *The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia*. Nat Genet, 2015. **47**(6): p. 672-6.
43. Raimondi, S.C., et al., *Reassessment of the prognostic significance of hypodiploidy in pediatric patients with acute lymphoblastic leukemia*. Cancer, 2003. **98**(12): p. 2715-22.
44. Nachman, J.B., et al., *Outcome of treatment in children with hypodiploid acute lymphoblastic leukemia*. Blood, 2007. **110**(4): p. 1112-5.
45. Liu, Y., et al., *The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia*. Nat Genet, 2017. **49**(8): p. 1211-1218.
46. Roberts, K.G., et al., *Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia*. N Engl J Med, 2014. **371**(11): p. 1005-15.
47. Monique L Den Boer, M.v.S., Renée X De Menezes, Meyling H Cheok, Jessica G C A M Buijs-Gladdines, Susan T C J M Peters, Laura J C M Van Zutven, H Berna Beverloo, Peter J Van der Spek, Gaby Escherich†, Martin A Horstmann†, Gritta E Janka-Schaub, and W.E.E. Willem A Kamps‡, Rob Pieters, *A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study*. The Lancet Oncology, 2009. **10**(2): p. 125 - 134.
48. Zhang, J., et al., *The genetic basis of early T-cell precursor acute lymphoblastic leukaemia*. Nature, 2012. **481**(7380): p. 157-63.
49. Coustan-Smith, E., et al., *Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia*. The Lancet Oncology, 2009. **10**(2): p. 147-156.
50. Julie Irving, et al., *Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition*. Blood, 2014. **124**(23): p. 3420-3430.
51. Pasqualucci, L., et al., *Inactivating mutations of acetyltransferase genes in B-cell lymphoma*. Nature, 2011. **471**(7337): p. 189-95.
52. Inthal, A., et al., *CREBBP HAT domain mutations prevail in relapse cases of high hyperdiploid childhood acute lymphoblastic leukemia*. Leukemia, 2012. **26**(8): p. 1797-803.

53. Mar, B.G., et al., *Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia*. Nat Commun, 2014. **5**: p. 3469.
54. Oshima, K., et al., *Mutational landscape, clonal evolution patterns, and role of RAS mutations in relapsed acute lymphoblastic leukemia*. Proc Natl Acad Sci U S A, 2016. **113**(40): p. 11306-11311.
55. Li, Y., et al., *IL-7 Receptor Mutations and Steroid Resistance in Pediatric T cell Acute Lymphoblastic Leukemia: A Genome Sequencing Study*. PLoS Med, 2016. **13**(12): p. e1002200.
56. Delgado-Martin, C., et al., *JAK/STAT pathway inhibition overcomes IL7-induced glucocorticoid resistance in a subset of human T-cell acute lymphoblastic leukemias*. Leukemia, 2017.
57. Tzoneva, G., et al., *Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL*. Nat Med, 2013. **19**(3): p. 368-71.
58. Aparicio, S. and C. Caldas, *The implications of clonal genome evolution for cancer medicine*. N Engl J Med, 2013. **368**(9): p. 842-51.
59. Mengelbier, L.H., et al., *Intratumoral genome diversity parallels progression and predicts outcome in pediatric cancer*. Nat Commun, 2015. **6**: p. 6125.
60. Mullighan, C.G., *The molecular genetic makeup of acute lymphoblastic leukemia*. Am Soc Hematol Educ Program, 2012. **2012**: p. 389-396.
61. Nowak MA, S.K., *Evolutionary dynamics of biological games*. . Science 2004. **2004**(303): p. 793–799.
62. Yang, J.J., et al., *Genome-Wide Association Study Identifies Germline Polymorphisms Associated with Relapse of Childhood Acute Lymphoblastic Leukemia*. Blood, 2012. **120**(21): p. 4197-4204.
63. Diouf, B., et al., *Association of an inherited genetic variant with vincristine-related peripheral neuropathy in children with acute lymphoblastic leukemia*. JAMA, 2015. **313**(8): p. 815-23.
64. Diouf, B., K.R. Crews, and W.E. Evans, *Vincristine pharmacogenomics: 'winner's curse' or a different phenotype?* Pharmacogenet Genomics, 2016. **26**(2): p. 51-2.
65. Lopez-Lopez, E., et al., *Vincristine pharmacokinetics pathway and neurotoxicity during early phases of treatment in pediatric acute lymphoblastic leukemia*. Pharmacogenomics, 2016. **17**(7): p. 731-41.
66. Ceppi, F., et al., *Polymorphisms of the vincristine pathway and response to treatment in children with childhood acute lymphoblastic leukemia*. Pharmacogenomics, 2014. **15**(8): p. 1105-16.
67. Ramsey, M.V.R.a.L.B., *Pharmacogenomics of acute lymphoid leukemia: new insights into treatment toxicity and efficacy*. Hematology 2013. **2013**: p. 126-130.
68. Yang, J.J., et al., *Inherited NUDT15 variant is a genetic determinant of mercaptopurine intolerance in children with acute lymphoblastic leukemia*. J Clin Oncol, 2015. **33**(11): p. 1235-42.
69. Tanaka, Y., et al., *Susceptibility to 6-MP toxicity conferred by a NUDT15 variant in Japanese children with acute lymphoblastic leukaemia*. Br J Haematol, 2015. **171**(1): p. 109-15.

70. Tanaka, Y., et al., *Interaction between NUDT15 and ABCC4 variants enhances intolerance of 6-mercaptopurine in Japanese patients with childhood acute lymphoblastic leukemia*. Pharmacogenomics J, 2017.
71. Wolthers, B.O., et al., *Asparaginase-associated pancreatitis: a study on phenotype and genotype in the NOPHO ALL2008 protocol*. Leukemia, 2017. **31**(2): p. 325-332.
72. Chen, S.H., et al., *Genetic variations in GRIA1 on chromosome 5q33 related to asparaginase hypersensitivity*. Clin Pharmacol Ther, 2010. **88**(2): p. 191-6.
73. Fernandez, C.A., Smith, C., Yang, W., Daté, M., Bashford, D., Larsen, E., Bowman, W. P., Liu, C., Ramsey, L. B., Chang, T., Turner, V., Loh, M. L., Raetz, E. A., Winick, N. J., Hunger, S. P., Carroll, W. L., Onengut-Gumuscu, S., Chen, W., Concannon, P., Rich, S. S., Scheet, P., Jeha, S., Pui, C., Evans, W. E., Devidas, M., & Relling, M. V. , *HLA-DRB1\*07:01 is associated with a higher risk of asparaginase allergies*. . Blood, 2014. **124**(8): p. 1266-1276. .
74. Fernandez, C.A., et al., *Genome-wide analysis links NFATC2 with asparaginase hypersensitivity*. Blood, 2015. **126**(1): p. 69-75.
75. Csordas, K., et al., *Associations of novel genetic variations in the folate-related and ARID5B genes with the pharmacokinetics and toxicity of high-dose methotrexate in paediatric acute lymphoblastic leukaemia*. Br J Haematol, 2014. **166**(3): p. 410-20.
76. Trevino, L.R., et al., *Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects*. J Clin Oncol, 2009. **27**(35): p. 5972-8.
77. Campana, D., *Minimal Residual Disease in Acute Lymphoblastic Leukemia*. Hematology, 2010. **2010**(1): p. 7-12.
78. G. MATHt, et al., *Extensive Histological and Cytological Survey of Patients with Acute Leukaemia in "Complete Remission"*. British Medical Journal, 1966. **1966**(1): p. 640-642.
79. Campana, D., *Determination of Minimal Residual Disease in Leukemia Patients* British Journal of Haematology, 2003. **2003**(121): p. 823–838.
80. Campana, D., *Status of minimal residual disease testing in childhood haematological malignancies*. Br J Haematol, 2008. **143**(4): p. 481-9.
81. Coustan-Smith, E., et al., *Prognostic importance of measuring early clearance of leukemic cells by flow cytometry in childhood acute lymphoblastic leukemia*. Blood, 2002. **100**(1): p. 52-8.
82. van Dongen, J.J., et al., *Minimal residual disease diagnostics in acute lymphoblastic leukemia: need for sensitive, fast, and standardized technologies*. Blood, 2015. **125**(26): p. 3996-4009.
83. Velden, V.v.d., et al., *Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects*. Leukemia 2003. **2003** (17): p. 1013–1034.
84. Dongen, J.v., et al., *Standardized RT-PCR analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of minimal residual disease*. Leukemia 1999. **1999**(13 ): p. 1901–1928.
85. Thorn, I., et al., *Minimal residual disease assessment in childhood acute lymphoblastic leukaemia: a Swedish multi-centre study comparing real-time*



- polymerase chain reaction and multicolour flow cytometry*. Br J Haematol, 2011. **152**(6): p. 743-53.
86. Malek Faham, et al., *Deep-sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia*. Blood, 2012. **120**(26): p. 5173-5180.
  87. Ladetto, M., et al., *Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders*. Leukemia, 2014. **28**(6): p. 1299-307.
  88. Chabon, J.J., et al., *Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients*. Nat Commun, 2016. **7**: p. 11815.
  89. Volik, S., et al., *Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped By Emerging Technologies*. Mol Cancer Res, 2016. **14**(10): p. 898-908.
  90. Krishnamurthy, N., et al., *Liquid Biopsies for Cancer: Coming to a Patient near You*. J Clin Med, 2017. **6**(1).
  91. Mandel P, M.P., *Les acides nucléiques du plasma sanguin chez l'homme*. C R Seances Soc Biol Fil. , 1948 **1948**(142): p. 241-243.
  92. Stroun M1, A.P., Lyautey J, Lederrey C, Maurice PA., *Isolation and characterization of DNA from the plasma of cancer patients*. Eur J Cancer Clin Oncol. , 1987. **23**(6): p. 707-12.
  93. Cheng, C., et al., *Quantification of circulating cell-free DNA in the plasma of cancer patients during radiation therapy*. Cancer Sci, 2009. **100**(2): p. 303-9.
  94. Siravegna, G., et al., *Integrating liquid biopsies into the management of cancer*. Nat Rev Clin Oncol, 2017. **14**(9): p. 531-548.
  95. M. Stroun a, et al., *About the possible origin and mechanism of circulating DNA Apoptosis and active DNA release*. Clinica Chimica Acta, 2001. **313**: p. 139–142.
  96. Jahr, S., et al., *DNA Fragments in the Blood Plasma of Cancer Patients: Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells1*. Cancer Research, 2001. **61**: p. 1659–1665.
  97. Wan, J.C.M., et al., *Liquid biopsies come of age: towards implementation of circulating tumour DNA*. Nat Rev Cancer, 2017. **17**(4): p. 223-238.
  98. Yu, S.C., et al., *High-resolution profiling of fetal DNA clearance from maternal plasma by massively parallel sequencing*. Clin Chem, 2013. **59**(8): p. 1228-37.
  99. Lo, Y.M.D., et al., *Rapid Clearance of Fetal DNA from Maternal Plasma*. Am. J. Hum. Genet., 1999. **64**: p. 218–224.
  100. Thierry, A.R., et al., *Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts*. Nucleic Acids Res, 2010. **38**(18): p. 6159-75.
  101. T. M. Chused, A.D.S.a.N.T., *The Clearance and Localization of Nucleic Acids by New Zealand and Normal Mice* Clin. exp. Immunol., 1972. **1972** (12): p. 465-476.
  102. Leon, S.A., et al., *Free DNA in the Serum of Cancer Patients and the Effect of Therapy*. Cancer Research, 1977. **37**: p. 646-650.
  103. Fleischhacker, M. and B. Schmidt, *Circulating nucleic acids (CNAs) and cancer--a survey*. Biochim Biophys Acta, 2007. **1775**(1): p. 181-232.

104. Tzong-Hae Lee, et al., *Quantitation of genomic DNA in plasma and serum samples: higher concentrations of genomic DNA found in serum than in plasma*. Transfusion 2001. **41**: p. 276-282.
105. Bronkhorst, A.J., J. Aucamp, and P.J. Pretorius, *Cell-free DNA: Preanalytical variables*. Clin Chim Acta, 2015. **450**: p. 243-53.
106. Lam, N.Y.L., et al., *EDTA Is a Better Anticoagulant than Heparin or Citrate for Delayed Blood Processing for Plasma DNA Analysis*. Clin Chem, 2004. **50**(1): p. 255-6.
107. Beutler E, G.T., Kuhl W., *Interference of heparin with the polymerase chain reaction*. Biotechniques, 1990. **1990** (9): p. 166.
108. van Ginkel, J.H., et al., *Preanalytical blood sample workup for cell-free DNA analysis using Droplet Digital PCR for future molecular cancer diagnostics*. Cancer Med, 2017. **6**(10): p. 2297-2307.
109. Mauger, F., et al., *Comprehensive evaluation of methods to isolate, quantify, and characterize circulating cell-free DNA from small volumes of plasma*. Anal Bioanal Chem, 2015. **407**(22): p. 6873-8.
110. Han, X., J. Wang, and Y. Sun, *Circulating Tumor DNA as Biomarkers for Cancer Detection*. Genomics Proteomics Bioinformatics, 2017. **15**(2): p. 59-72.
111. Rapisuwon, S., E.E. Vietsch, and A. Wellstein, *Circulating biomarkers to monitor cancer progression and treatment*. Comput Struct Biotechnol J, 2016. **14**: p. 211-22.
112. Hindson, C.M., et al., *Absolute quantification by droplet digital PCR versus analog real-time PCR*. Nat Methods, 2013. **10**(10): p. 1003-5.
113. Hindson, B.J., et al., *High-throughput droplet digital PCR system for absolute quantitation of DNA copy number*. Anal Chem, 2011. **83**(22): p. 8604-10.
114. Dressman, D., et al., *Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations*. PNAS. **100**(15): p. 8817-8822.
115. Newman, A.M., et al., *An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage*. Nat Med, 2014. **20**(5): p. 548-54.
116. Mamanova, L., et al., *Target-enrichment strategies for next-generation sequencing*. Nat Methods, 2010. **7**(2): p. 111-8.
117. Gnirke, A., et al., *Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing*. Nat Biotechnol, 2009. **27**(2): p. 182-9.
118. Chaudhuri, A., et al., *Analysis of circulating tumor DNA in localized lung cancer for detection of molecular residual disease and personalization of adjuvant strategies*. Journal of Clinical Oncology 2017 **35**:**15\_suppl** p. 8519.
119. Scherer, F., et al., *Noninvasive molecular subtyping and risk stratification of DLBCL*. Journal of Clinical Oncology 2016. **34**:**15\_suppl**, **7554-7554**: p. 7554-7554
120. Merriott, D.J., et al., *Circulating Tumor DNA Quantitation for Early Response Assessment of Immune Checkpoint Inhibitors for Lung Cancer*. International Journal of Radiation Oncology\*Biophysics, 2017. **99**(2): p. S20-S21.
121. Kinde, I., et al., *Detection and quantification of rare mutations with massively parallel sequencing*. Proc Natl Acad Sci U S A, 2011. **108**(23): p. 9530-5.

122. Kou, R., et al., *Benefits and Challenges with Applying Unique Molecular Identifiers in Next Generation Sequencing to Detect Low Frequency Mutations*. PLoS One, 2016. **11**(1): p. e0146638.
123. Newman, A.M., et al., *Integrated digital error suppression for improved detection of circulating tumor DNA*. Nat Biotechnol, 2016. **34**(5): p. 547-555.
124. Schwarz, A.K., et al., *Quantification of free total plasma DNA and minimal residual disease detection in the plasma of children with acute lymphoblastic leukemia*. Ann Hematol, 2009. **88**(9): p. 897-905.
125. Cheng, S.H., et al., *Minimal residual disease-based risk stratification in Chinese childhood acute lymphoblastic leukemia by flow cytometry and plasma DNA quantitative polymerase chain reaction*. PLoS One, 2013. **8**(7): p. e69467.
126. Hocking, J., et al., *Liquid biopsies for liquid tumors: emerging potential of circulating free nucleic acid evaluation for the management of hematologic malignancies*. Cancer Biol Med, 2016. **13**(2): p. 215-25.
127. Maxwell® RSC ccfDNA Plasma Kit Technical Manual. Available from: <https://www.promega.com/resources/protocols/technical-manuals/101/maxwell-rsc-ccfdna-plasma-kit-protocol/>.
128. TruSeq® DNA Sample Preparation Guide. Available from: [https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry\\_documentation/samplepreps\\_truseq/truseqdna/TruSeq\\_DNA\\_SamplePrep\\_Guide\\_15026486\\_C.pdf](https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqdna/TruSeq_DNA_SamplePrep_Guide_15026486_C.pdf).
129. Library pooling guidelines for the NextSeq and MiniSeq systems. Available from: <https://support.illumina.com/bulletins/2016/10/library-pooling-guidelines-for-the-nextseq-and-miniseq-systems.html>.
130. Murtaza, M., et al., *Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA*. Nature, 2013. **497**(7447): p. 108-12.
131. Nadja Kopp, J.C.V.d.Z., Jacob Layer, Oliver Weigert, Eric Vangrevelinghe, Akinori Yoda, Thomas Radimerski and David Weinstock, *AK2 L884P Mutation Confers Resistance To The Type II JAK2 Inhibitor NVP-BBT594 When Co-Occurring With JAK2 R683G But Not JAK2 V617F*. Blood, 2013. **122**(21): p. 1429.

# APPENDIX A. LIST OF GENES INCLUDED IN ALL GENE PANEL

Gene Name					
ABP1	ALPK2	ATP13A4	C11orf41	CD6	CHRM2
ACE2	ANAPC1	ATP5A1	C11orf82	CDC20B	CHRNA4
ACOX1	ANK3	ATP8A1	C12orf51	CDC6	CLCA4
ACP5	ANKRD12	ATP8A2	C12orf64	CDH18	CLCN3
ACSM2B	ANKRD30B	ATRNL1	C12orf66	CDH23	CLEC4M
ACSM4	ANKRD43	ATRX	C13orf40	CDH8	CLIC3
ACTL6B	AOX1	AXIN2	C14orf4	CDH9	CMYA5
ACTN2	APC	AXL	C16orf52	CDK16	CNOT1
ADD1	APOB48R	AZI1	C1QTNF3	CDKN2A	CNTN1
ADH6	AR	BAI2	C6orf97	CDKN2AIP	CNTN3
ADNP	ARAP1	BAI3	CABLES1	CDR2L	CNTN5
ADRBK1	ARHGAP20	BAZ2A	CAD	CDYL2	CNTN6
AFAP1	ARID1A	BBS9	CAMSAP1L1	CELSR2	CNTNAP2
AFF2	ARID1B	BCOR	CAND1	CENPC1	CNTNAP5
AFF3	ARID2	BCORL1	CASP4	CENPE	COPS2
AGXT2L1	ARID5B	BCR	CBS	CEP290	CPB1
AHNAK	ASH1L	BIRC6	CBX4	CES1	CREBBP
AHNAK2	ASXL1	BMP1	CCBE1	CFTR	CREBBP,CREBBP
AHRR	ASXL2	BRAF	CCDC114	CHD1	CRLF2
AKR1B1	ASXL3	BRCA1	CCDC147	CHD2	CSMD1
AKR1C2	ATF7IP	BRPF1	CCND3	CHD4	CSMD2
AKT1	ATG4A	BRSK1	CD101	CHD7	CSMD3
ALG13	ATM	BVES	CD36	CHD9	CSPG4

Gene Name					
CTCF	DLC1	DST	ETV6	FGFR2	GALNTL4
CTNND1	DMBT1	DTX3L	ETV6,ETV6,ETV6	FGFR3	GAS2L3
CTNND2	DNAH2	DYNC2H1	EVC2	FHOD1	GATA2
CTR9	DNAH5	DYSF	EZH2	FLNC	GATA3
CUBN	DNAH8	ECT2L	F13B	FLRT2	GBP7
CXorf22	DNAH9	EED	FAAH	FLT3	GCKR
CXorf23	DNM2	EEF1A2	FAT1	FLT3,FLT3	GGT1
CXorf57	DNMT3A	EGFLAM	FAT4	FMN2	GIGYF2
CXorf59	DNMT3B	EGFR	FBN1	FMN2,FMN2	GIGYF2,GIGYF2
CYP2C8	DOCK11	EGR3	FBN2	FMOD	GIPR
CYP4F22	DOCK5	EHMT2	FBN3	FMR1	GLIS2
DAAM2	DOCK6	EIF4A1	FBXL19	FN1	GLTSCR1
DACH1	DPH1	ELAVL2	FBXO18	FNDC1	GNAQ
DACT2	DPP3	ELF1	FBXW7	FOXA1	GNB1
DCAF10	DPT	ENPP5	FBXW7,FBXW7	FOXO4	GPC6
DCHS2	DPYS	ENPP6	FCGBP	FOXP2	GPR116
DCST2	DRD1	EP300	FCRL2	FRAS1	GPR142
DDX3X	DRP2	EPPK1	FCRLA	FREM2	GPR98
DEPDC5	DSC1	ERC2	FER1L6	FRMD4A	GREB1
DHX33	DSCAM	ERCC6	FEZF1	FRMPD2	GRIA1
DIAPH2	DSG2	ERG	FGF12	FSIP2	GRIA2
DIAPH3	DSG4	ESX1	FGF13	FURIN	GRID1
DISC1	DSP	ETV1	FGF9	GAD1	GRID2

Gene Name					
GRIN2A	HRNR	JAG1	KIAA0922	LGALS12,LGALS12	MAMDC2
GRM5	HUWE1	JAK1	KIAA1024	LGALS4	MAP1B
GRPR	HYDIN	JAK2	KIAA1109	LIFR	MAP2
GSTA4	IDH1	JAK3	KIAA1217	LIMD1	MAP2K4
GYS1	IFIH1	JAKMIP1	KIAA1244	LOXL3	MAP3K4
H3F3A	IGSF9B	JARID2	KIAA1409	LPHN1	MAP4K5
HADHA	IKZF1	JMY	KIAA1715	LPHN2	MAPK9,MAPK9
HAT1	IKZF3	KAL1	KIF15	LPO	MARCKS
HCN4	IL15RA	KCNA4	KIF21A	LRFN2	MATN4
HCRT1	IL1RAPL1	KCNG2	KIF4A	LRP1B	MBD1
HDAC2	IL7R	KCNH1	KLF4	LRP2	MCF2
HDAC5	IL7R,IL7R	KCNH2	KLK9	LRRK2	MCM2
HDAC7	IL7R,IL7R,IL7R	KDM1A	KNDC1	LSG1	MCTP1
HDAC9	INPP4B	KDM2B	KPNB1	LTBP1	MDGA2
HGF	IQCB1	KDM5B	KRAS	LTBP2	MECP2
HIST1H1B	ITGA6	KDM5C	KRAS,KRAS	LUC7L2	MED12
HIST1H1C	ITGAD	KDM6A	KRT33A	LYPD2	MED13
HIVEP3	ITGB4	KDM6A,KDM6A	L3MBTL1	LZTR1	MGA
HK2	ITGB8	KDM6A,KDM6A,KDM6A	L3MBTL4	MADD	MGAT1
HLF	ITIH2	KIAA0240	LAMA2	MAGEB2	MGAT4C
HMCN1	ITPKC	KIAA0368	LAMB1	MAGED2	MGAT5B
HNF4G	ITPR1	KIAA0430	LAMC1	MAGI3	MID1
HNRNPA1	ITSN1	KIAA0802	LEPREL2	MAGI1	MLL,MLL

Gene Name					
MLL2	NAV3	NNMT	NUDT10	PCNX	PKHD1L1
MLL2,MLL2	NBEA	NOD1	NXF1	PCSK2	PKN3
MLL3	NCCRP1	NOS1	NYNRIN	PDCD2L	PKP2
MLL4	NCKAP1	NOS1AP	OGT	PDE11A	PLCB2
MME	NCKAP5	NOS2	OGT,OGT	PDE3A	PLCB3
MMS19	NCOR1	NOS3	OLFM3	PDE7B	PLCE1
MPO	NCOR2	NOTCH1	OPHN1	PDS5B	PLS3
MRC2	NDUFS6	NOTCH2	OPRM1	PDZD2	PLXNA2
MSH6	NEDD4L	NOTCH3	OR5I1	PDZRN4	PNMA2
MSN	NEGR1	NOTUM	OTUD7A	PEG3	POLB
MTA1	NEIL3	NOVA1	PAK2	PER1	POLD1
MTHFD1	NEK10	NOX1	PAMR1	PHACTR2	POLQ
MTNR1A	NELL1	NR3C1	PAPPA2	PHF12	POM121L12
MTOR	NEXN	NRAP	PARK2	PHF6	PON3
MTUS2	NF1	NRAS	PAX5	PHF6,PHF6	POT1
MVP	NF1,NF1	NRAS,NRAS	PAX9	PHF8	PPFIA2
MXRA5	NHSL1	NRAS,NRAS,NRAS	PBRM1	PHLDB1	PPFIBP1
MYB	NIPBL	NRK	PCDH1	PIK3CA	PPFIBP2
MYC	NKX2-8	NRXN1	PCDH15	PIK3R1	PPP1R10
MYCBP2	NLGN4X	NSD1	PCDHA3	PIK3R3,PIK3R3	PPP2R1A
MYST3,MYST3	NLRC4	NT5C1B	PCDHAC1	PIWIL4	PPP2R2C
NALCN	NLRP3	NT5C2	PCDHGA5	PKDREJ	PPYR1
NAT1	NLRP8	NT5C2,NT5C2	PCLO	PKHD1	PRAM1

Gene Name					
PRDM1	PTPN11	RUNX1	SMARCC1	AMOTL2	FAM96B
PRDM15	PTPN13	RXFP2	SMC1A	ARNT	FARSB
PRDM2	PTPN14	RYR2	SMC3	ATG7	FOLR1
PRDM9	PTPN3	SAE1	SNRNP200	BAK1	FTH1
PRKCD	PTPRF	SALL3	SOX4	BRD4	G6PC2
PRKCE	PTPRG	SCN5A	SOX6	C12orf10	GHRL
PRKCQ	PTPRN	SERPINA1	SOX9	C14orf178	GLIS1
PRKCZ	PTPRT	SERPINA4	SPAST	C2orf82	GPR149
PRKD1	PTPRZ1	SERPINB10	SPEN	CAPN3	GPR6
PRKD2	PXDNL	SETD2	SPOP	CASP5	HEY2
PRKDC	RAB32	SETDB1	SPTA1	CBX8	HIST1H3A
PRMT3	RAD21	SH2B3	SPTBN1	CCNDBP1	HIST1H3C
PRPS1	RAI1	SH3KBP1	SPTBN5	CD5L	HSD3B1
PRRX1	RB1	SHISA2	SPZ1	CD79A	IRGM
PRSS22	RBMX	SHROOM2	SRRM2	CPNE1	JPH3
PRSS3	RELN	SI	SSPO	CPSF6	KLF12
PRUNE2	RIMBP2	SIN3A	ST14	DDT	KLHDC4
PRX	RIPK4	SLFN5	STAG2	DNHD1	LIMS2
PSD4	ROBO1	SLIT3	STARD9	ECE2	LPA
PSMA7	ROBO2	SLITRK2	STK11	EIF4EBP3	LRIG2
PTCH1	ROR1	SLITRK4	ACTC1	ENPEP	LRRC34
PTGIS	RPL22	SMARCA4	ADAMTS3	ERBB2	LRRC58
PTPN11	RPL5	SMARCB1	ADAP2	EXTL3	LYPD4



Gene Name					
MAP10	TACC2	TMEM205	TUBGCP6	WDR17	ZSCAN20
MYCN	TACC3	TMTC1	TXNRD1	WHSC1	
NEUROG3	TAF1	TMTC2	U2AF1	WNK3	
NRIP1	TAF15	TNC	UBA2	WRN	
P2RY12	TAF3	TNFRSF13B	UBR2	WSCD2	
PCDHB16	TATDN1,TATDN1	TNIK	UBR4	WT1	
PDCD6	TBC1D9	TNN	UGT2A3	WWC1	
PLCH2	TBL1XR1	TP53	UHRF1	XBP1	
PMAIP1	TBX15	TP53,TP53	UHRF1BP1	XIRP2	
RPL22L1	TBX18	TP53,TP53,TP53	UNC80	XRRA1	
RRP9	TBX3	TPCN1	USH2A	ZAN	
SERPINB7	TCF3	TRIM39	USP5	ZBTB3	
SHC2	TECTA	TRMT1L	USP7	ZEB2	
SHF	TET2	TRPM1	USP9X	ZFHX4	
SPATA31E1	TET3	TRPM3	UTY	ZFP36L2	
SUN2	TGM4	TRPV2	VANGL1	ZKSCAN4	
SUPT6H	THBS2	TRRAP	VEZF1	ZMYM2	
SUZ12	THSD7A	TSC1	VPS13A	ZNF292	
SVIL	TIGD7	TSHZ2	VSIG2	ZNF445	
SYNE1	TLN2	TSKS	VWA3A	ZNF483	
SYNE2	TMC6	TTC17	VWA5A	ZNF521	
SYT12	TMEM127	TTC28	WBSCR17	ZNF687	
TAB2	TMEM184A	TTC9	WDR16	ZNRF2	

(Personal communication with Ousman Mahmoud on 13<sup>th</sup> February 2018, Charles Gawad Lab. St. Jude Children's Research Hospital)

## APPENDIX B. TOTAL XVII RISK STRATIFICATION

### Criteria for Low-Risk (LR) ALL

- B-ALL with DNA index  $\geq 1.16$ , t(12;21) **Or** age 1 to 9.9 years and presenting WBC count  $< 50 \times 10^9/L$  *plus*
- Absence CNS 3 status, testicular leukemia, adverse genetic characteristics (t(9;22); t(1;19); MLL gene re-arrangement, hypodiploidy and poor early response defined as MRD  $\geq 1\%$  on D15 or  $\geq 0.01\%$  at end of remission induction.

### Criteria for Standard-Risk (SR) ALL

- Patients with T- or B-ALL who do not meet the criteria for low-risk or high-risk ALL.

### Criteria for High-Risk (HR) ALL

- **End of early intensification:** MRD  $\geq 0.1\%$  + inadequate decrease in MRD levels after 1-2 courses of Consolidation treatment
- **End of remission induction:** MRD  $\geq 0.01\%$  in hypodiploid patients.
- **After re-induction II (Continuation W17):** Persistent  $\geq 0.01\%$  MRD
- Increasing in MRD at  $\geq 0.01\%$  after remission induction
- Previously MRD-negative patients: Re-emergence of MRD at  $\geq 0.01\%$

## APPENDIX C. TOTAL XVII REMISSION INDUCTION TREATMENT SCHEDULE

### Precursor B-cell ALL Remission Induction Treatment Schedule

Agent	Schedule
Prednisone or Prednisolone	Days 1-28
Vincristine	Days 1#, 8, 15, 22
DAUNOrubicin	Days 2# +/- 8
Pegaspargase	Days 4 and 23
Cyclophosphamide	Day 22
Cytarabine	Days 22-25 & 29-32
Mercaptopurine	Days 22-35

Bortezomib (*No targetable lesions and Day 15 MRD  $\geq 5\%$* ) on D29 and D32  
 Dasatinib (*Ph+, Ph-like ABL1-class fusion only*) starting on D15 except for select cases  
 Rituximab is given on D3 (un-blinded randomization)

### T-cell ALL Remission Induction Treatment Schedule

Agent	Schedule
Prednisone or Prednisolone	Days 1-28
VinCRISTine	Days 1#, 8, 15, 22
DAUNOrubicin	Days 2#, 8, 15
Pegaspargase	Days 4 and 23†
Cyclophosphamide†	Day 22
Cytarabine†	Days 22-25 and 29-32
Mercaptopurine†	Days 22-35

Bortezomib (*No targetable lesions and Day 15 MRD  $\geq 5\%$* ) on D29 and D32  
 Dasatinib (*ABL1-class fusion only*) starting on D15 except for select cases

## VITA

Sarah was born in Egypt on 1981. She earned her bachelor's degree in pharmaceutical sciences with a very good honors grade in 2003. She achieved her board of pharmacy specialty in Oncology pharmacy on 2010 and passed her re-certification exam in 2017. In 2008, She had the opportunity to join the largest pediatric cancer hospital in the Middle East (Children's Cancer Hospital Egypt, CCHE 57357) where she worked as a clinical pharmacist and got promoted to the position of Clinical Pharmacist Supervisor for the inpatient hematology oncology wards. In 2016, she was granted a Fulbright scholarship to study in the US for a MS. degree in Pharmaceutical Sciences which would be awarded to her from University of Tennessee Health Science Center May 2018.