Molecular profiling of cell free DNA in patients with paediatric solid tumours

Reda Stankunaite

Submitted to the University of London/The Institute of Cancer Research for the Degree of Doctor of Philosophy

November 2022

Author’s Declaration

I hereby declare that that the work presented in this thesis is my original work. Any contribution made by others with whom I have worked is explicitly acknowledged in the thesis.

Reda Stankunaite

November 2022

Abstract

Tumour tissue profiling studies show that a substantial proportion of paediatric cancer patients have potentially actionable alterations and minimally invasive molecular profiling tests using cell fee DNA (cfDNA) could provide a powerful platform to guide clinical decision-making and to deliver precision treatments. Clinical diagnostic sequencing of cfDNA is well advanced for adult patients, but application to paediatric cancer patients lags behind. To fill this gap, in this thesis I have described the development and validation of a clinically relevant pan-paediatric solid tumour NGS capture panel optimised for cfDNA analysis and an accompanying workflow with low coverage WGS (lcWGS) to molecularly profile paediatric patients with solid tumours. I applied this workflow to cfDNA samples from multiple clinical trials and showed that it is informative and yields comparable results to tissue biopsy molecular profiling in patients with extracranial tumours. Additionally, high number of cfDNA unique variants detected in patients at relapse showed the potential to complement tissue biopsy testing in many clinical diagnostics situations by allowing detection of tumour heterogeneity and identifying variants missed by tissue biopsy profiling, some of which are potentially targetable or aiding enrolment to clinical trials. In patients with CNS tumours plasma based cfDNA profiling was of limited success, however I showed potential to use cerebrospinal fluid cfDNA instead. Significant changes were observed between diagnostic and primary tissue biopsies with accumulation of SNVs and copy number changes at relapse and the ability to detect these changes in cfDNA was shown in patients with good purity ctDNA, highlighting the potential of cfDNA profiling to monitor tumour evolution in a minimally invasive way. Finally, several case studies showed the potential to track disease progress and identify relapse earlier than conventional methods using longitudinal cfDNA sampling. Overall, in this thesis I have shown that cfDNA is a good biomarker in paediatric cancer care and future molecularly enriched biomarker-driven interventional clinical trials utilising cfDNA are warranted.

Acknowledgements

I would like to express my gratitude to my supervisors for giving me the opportunity to pursue this PhD. Thank you for the support, guidance and for always encouraging me to become a better scientist. Many thanks to Prof. Michael Hubank for being the best supervisor any student could ask for, for always being available, for asking the tough questions and encouraging me to explore and always do a bit better. I am greatly thankful to my co-supervisors Prof. Andrea Sottoriva and Prof. Louis Chesler for showing me the different ways of thinking and approaching scientific problems. I have learned so much from all of you and it was a great privilege to be part of your labs.

I would like to thank all the members of Translational Research and Cancer Evolution and modelling labs, you have been an amazing family for me in London. Trans Res team, thank you for listening during the struggles and guiding me to find the solutions. Every single one of you are an absolute inspiration of how science can help the patients. Special thanks to Sally for always finding the time to read my drafts and make them infinitely better, Paula for all your NGS and life wisdom, Debbie and Claire for tackling the big projects together. Thanks Michaela, Sabri and Ridwan for all the coffee breaks and after work fun. Massive thank you to Inma, Erica and Rachel for all the walks and talks throughout these years, especially throughout the lockdowns. Thanks, George, for the pub trips and support and reassurance. And a massive thanks to my best dancing buddies – George, Kat, Erica, Rachel, Claire, Javi, Erika and Timon. I was so fortunate to be part of your teams.

I am eternally grateful for my best friends supporting me throughout this journey – Laurita, Vilija, Iveta, Nithya, Laura, Erica – thank you for always being there for me in your own and unique ways. Special thank you to Nithya, my best PhD friend, for all the tea breaks and lunches, for all the inspiration and the shared struggles. Your kindness and strength to look at life with such grace has been an inspiration.

I would like to thank NIRH for the studentship funding and Christophers smile, for trusting our work and inspiring me to look for better ways to help patients. Your determination and kindness inspired me throughout this project. This work would not be possible without all the patients and their families, involved in these studies. I would also like express my gratitude to all the clinicians I had the honour to work with and a special thank you to Sally and Paola who helped to put my findings into a wider clinical perspective.

Finally, I am most grateful to my family, my mom Ruta, my dad Rimantas and my sister Rimante, for their love and never-ending support, for inspiring me to love science and teaching the resilience needed to achieve this.

Table of Contents

Abstract **3**

Acknowledgements **4**

Table of Contents ***5***

List of Figures **8**

List of Tables **11**

List of abbreviations **12**

Authors contribution **13**

Chapter 1 Introduction ***14***

1.1 Paediatric cancers **14**
1.1.1 Epidemiology and characteristics of paediatric cancers 14
1.1.2 The evolution of paediatric solid tumours from diagnosis to relapse 23

1.2 Molecular profiling of cfDNA in paediatric cancer patients **24**
1.2.1 cfDNA biological features 25
1.2.2 ctDNA levels in patients with paediatric solid tumours 27
1.2.3 Potential applications of cfDNA analysis 28
1.2.4 Approaches to cfDNA analysis 35

1.3 Aims of the project **38**

Chapter 2 Materials and methods ***40***

2.1 Clinical samples and validation materials **40**
2.1.1 Validation materials 40
2.1.2 Clinical trials 40

2.2 Sample processing for cfDNA studies **43**
2.2.1 Blood sample processing 43
2.2.2 CSF sample processing 43

2.3 Extraction of DNA **44**
2.3.1 cfDNA extraction 44
2.3.2 DNA and RNA extraction form tissue and blood cell pellet 45

2.4 Targeted sequencing **45**
2.4.1 cfDNA targeted sequencing panel design 45
2.4.2 cfDNA library preparation and capture 47
2.4.3 Sequencing library preparation and capture for tissue and blood cell pellet DNA 49

2.5 lcWGS of cfDNA and tissue **50**

2.6 Sequencing **50**
2.6.1 cfDNA sequencing 50
2.6.2 Tissue DNA sequencing 51

2.7 Bioinformatics analysis **51**
2.7.1 cfDNA analysis pipeline using UMIs 51
2.7.2 DeepSNV error correction 54
2.7.3 Tissue analysis pipeline 54
2.7.4 Primary tissue analysis pipeline 54
2.7.5 lcWGS analysis 55

2.8 Variant curation **57**

2.9 ctPC validation statistical analysis **57**
2.9.1 Overall performance 57
2.9.2 Sensitivity, specificity and accuracy 58

2.9.3 Repeatability and reproducibility 58
2.9.4 Limit of detection 59

2.10 Other methods 59
2.10.1 Droplet digital PCR 59
2.10.2 RNA analysis 59
2.10.3 dNdS analysis 60

Chapter 3 Development and validation of the cfDNA analysis workflow 61

3.1 Introduction 61

3.2 Determining the optimal cfDNA extraction method 63
3.2.1 Further validation of cfDNA extraction methods 65

3.3 Development of the targeted cfDNA panel 66
3.3.1 Target selection 66
3.3.2 Target selection for ctPC_v2 71

3.4 Validation of the ctPC panel for SNV detection 72
3.4.1 Validation of the ctPC NGS capture panel 73
3.4.2 Performance of ctPC workflow in cfDNA 77

3.5 Verification of ctPC_v2 81
3.5.1 Verification of ctPC_v2 panel performance 81
3.5.2 Verification of sequencing performance and trueness 81

3.6 Variant interpretation in cfDNA: guidelines and limits of detection 83

3.7 Copy number changes detection by lcWGS analysis 86
3.7.1 Limit of detection for lcWGS 87

3.8 Optimisation of the workflow for challenging samples 88
3.8.1 Optimisation for low input samples 89
3.8.2 Optimisation for samples with high molecular weight DNA contamination 91

3.9 Proposed workflow of cfDNA sample processing 94

3.10 Detection of focal copy number changes by ctPC panel in cfDNA 95

3.11 Detection of fusions in cfDNA 95

3.12 Discussion 99

Chapter 4 Concordance of tissue and liquid biopsies in paediatric patients with solid tumours 102

4.1 Introduction 102

4.2 NGS tumour profiling study: cfDNA profiling in paediatric patients with solid tumours 104
4.2.1 ctDNA is detectable in most paediatric solid tumour patients with active disease 104
4.2.2 Variants detected in tissue sequencing are found in cfDNA in most patients with active extracranial disease 106

4.3 Stratified Medicine for Paediatrics study: cfDNA profiling at relapse 108
4.3.1 ctDNA levels in different cancer types 108
4.3.2 Single nucleotide variant and indel detection using panel sequencing in cfDNA and tissue 112
4.3.3 Genome wide copy number changes can be detected by lcWGS in cfDNA 129
4.3.4 Genomic changes acquired in paediatric solid tumour at relapse 133

4.4 Discussion 140

Chapter 5 Liquid biopsy for paediatric cancer patients with CNS tumours 146

5.1 Introduction 146

5.2 Blood based cfDNA profiling is of limited success in paediatric patients with CNS tumours 150

5.2.1 Early studies of blood based cfDNA profiling for patients with CNS tumours 150
5.2.2 Detection of ctDNA in plasma from patients with CNS tumours at relapse..... 153

5.3 CSF as an alternative source of cfDNA for patients with CNS tumours.....157
5.3.1 Preliminary results of CSF-cfDNA analysis..... 158
5.3.2 Practical and technical considerations for liquid biopsy implementation for children with CNS tumours..... 162

5.4 Discussion.....164

Chapter 6 Monitoring disease progress using cfDNA..... 170

6.1 Introduction.....170

6.2 Serial cfDNA profiling can be used to monitor the course of disease 172
6.2.1 cfDNA can monitor disease progress 172
6.2.2 Monitoring response to treatment in patients on targeted therapies using cfDNA. 177
6.2.3 cfDNA unique variants detected in patients at progression..... 182

6.3 Discussion.....185

Chapter 7 Discussion 190

7.1 Development and validation of the method 190

7.2 Clinically relevant and actionable genomic alterations can be detected in cfDNA ...192

7.3 Longitudinal cfDNA profiling can successfully monitor disease progress 196

7.4 Implementation into clinical practice and future directions 201

Appendix 203

Bibliography..... 217

List of Figures

Figure 1-1 The epidemiology of childhood cancer 16
Figure 1-2 Mutation burden of paediatric cancers compared to adult cancers 17
Figure 1-3 Schematic representation of the origins and the range of biological analytes for liquid biopsy approaches 26
Figure 1-4 T The potential applications of cfDNA analysis throughout disease course of a patient enabled by serial sample collection 29
Figure 1-5 Types of heterogeneity and the advantages of cfDNA profiling 31
Figure 1-6 Schematic of the of NGS with using UMIs approach for cfDNA molecular profiling. 38
Figure 2-1 Panel design workflow 47
Figure 2-2 Flow chart outlining the main steps of the Cell3™ Target workflow 48
Figure 2-3 ctDNA analysis pipeline 53
Figure 3-1 Comparison of different cfDNA extraction methods 64
Figure 3-2 Sequencing results in three paediatric plasma samples, extracted using either QIAamp or Qiasymphony method 65
Figure 3-3 Sequencing depth achieved in cfDNA samples depending on extraction method 66
Figure 3-4 The genes on the ctPC panel compared to tissue sequencing panel and paediatric cancer landscape studies 70
Figure 3-5 The genes on the ctPC panel compared to the recent high profile paediatric cancer landscape studies 71
Figure 3-6 The workflow for the implementation of a molecular genetic test for diagnostic use. 73
Figure 3-7 ctPC panel performance 74
Figure 3-8 Repeatability of ctPC panel sequencing 75
Figure 3-9 Reproducibility of ctPC panel sequencing 76
Figure 3-10 Correlation of observed and expected allelic frequency of SNVs in different types of samples. 77
Figure 3-11 Correlation of allelic frequency of SNVs and indels in Sera Care cfDNA control samples 79
Figure 3-12 Correlation of allelic frequency of SNVs in clinical cfDNA samples. 80
Figure 3-13 Allele Frequency (AF) comparison between SeraCare 2% control as determined by the two versions of ctPC panel for SNVs and indels. 81
Figure 3-14 Quality metrics of sequencing in cfDNA samples run on ctPC and ctPC_v2 panel 82
Figure 3-15 Trueness of VAF in clinical paediatric cfDNA samples 82
Figure 3-16 Variants called blindly in each VAF range 85
Figure 3-17 The two lcWGS analysis methods used throughout the studies produce similar purity estimates and CNV profiles 86
Figure 3-18 Limit of detection for lcWGS analysis. 87
Figure 3-19 An example of in silico size selection effect on lcWGS profile 88
Figure 3-20 Volume of plasma collected and cfDNA yield of paediatric plasma samples discussed in this thesis 89
Figure 3-21 Molecular profiling of low yield cfDNA samples 90

Figure 3-22 Size selection protocol to remove HMW contamination form cfDNA samples..... 92
Figure 3-23 Comparison of molecular profiling results for a clinical sample with high HMW contamination. 94
Figure 3-24 The proposed workflow for cfDNA sample NGS library preparation 95
Figure 3-25 Copy number detection using ctPC panel in cfDNA 95
Figure 3-26 The most common fusion break points in paediatric cancer patients in EWSR1 gene 97
Figure 3-27 Fusions can be detected in cfDNA 98
Figure 4-1 cfDNA yield from plasma from patients with various paediatric cancer types 105
Figure 4-2 The unique sequencing depth (UMIx) achieved depending on the input amount of cfDNA into the library preparation 106
Figure 4-3 Comparison of findings between paediatric cfDNA and tissue samples in extracranial tumours. 107
Figure 4-4 cfDNA samples included in the study 109
Figure 4-5 Characteristics of samples included in the study 110
Figure 4-6 ctDNA fraction in cfDNA in patients with various solid tumours 111
Figure 4-7 Median fragment size distributions of cfDNA from plasma 112
Figure 4-8 SNV and indel detection in cfDNA versus tissue. 113
Figure 4-9 Differences between cfDNA-unique and tissue and cfDNA overlapping variants 117
Figure 4-10 Characteristics of the patients with no SNVs, indels and focal CNVs detected 118
Figure 4-11 Comparison of SNV and indel detection between cfDNA and tissue biopsy in patients with neuroblastoma 119
Figure 4-12 Molecular profiling results of patient SMP0154 with neuroblastoma 121
Figure 4-13 Comparison of CNV detection using targeted panel sequencing in cfDNA and tissue biopsy in patients with neuroblastoma 122
Figure 4-14 Serial liquid biopsy sampling results compared to tissue biopsy 123
Figure 4-15 Comparison of SNV and indel detection between cfDNA and tissue biopsy in patients with sarcoma 125
Figure 4-16 Comparison of CNV detection using targeted panel sequencing in cfDNA and tissue biopsy in patients with sarcoma 126
Figure 4-17 Serial liquid biopsy sampling results compared to tissue biopsy 127
Figure 4-18 Comparison between SNV and CNV detection between tissue and cfDNA in less common cancer types 128
Figure 4-19 Genome wide copy number profiles of tissue and cfDNA in patients with high purity ctDNA and tissue samples (>10% purity) 130
Figure 4-20 The identification of likely driver somatic copy-number alterations 131
Figure 4-21 Comparison of CNV detection using lcWGS sequencing in cfDNA and tissue biopsy..132
Figure 4-22 Differences between genome wide copy number profiles in multiple cfDNA and tissue biopsy samples. 133
Figure 4-23 Characteristics of the tumours at diagnosis and relapse 134
Figure 4-24 Genome wide copy number changes observed in diagnostic and relapse paediatric solid tumours 135

Figure 4-25 Selective pressures in primary and relapsed paediatric tumours and detection of driver genes 138
Figure 4-26 Comparison of SNV and indel detection between primary and tissue biopsy in patients with neuroblastoma (A) sarcomas (B) and other extracranial solid tumours (C) 139
Figure 4-27 Summary flow chart of samples discussed in this chapter 140
Figure 5-1 Sample characteristics from the patients with CNS tumours in Stratified Medicine Paediatric Programme. 153
Figure 5-2 Comparison of panel sequencing results in tissue versus cfDNA from plasma in pateints with CNS tumours 154
Figure 5-3 Representative lcWGS profiles of tissue and cfDNA from patients with CNS tumours 157
Figure 5-4 NGS4294 study CSF samples 159
Figure 5-5 The three CSF-cfDNA samples analysed using ctPC panel and lcWGS 160
Figure 5-6 The proposed guidelines for pre-analytical CSF handling for cfDNA analysis. 162
Figure 5-7 Summary flow chart of blood derived cfDNA samples discussed in this chapter 164
Figure 5-8 Summary flow chart of CSF derived cfDNA samples discussed in this chapter. 166
Figure 6-1 Monitoring response to treatment using cfDNA in patients with neuroblastoma 173
Figure 6-2 Monitoring response to treatment using cfDNA in patients with neuroblastoma 176
Figure 6-3 Monitoring the response to treatment in a patient with neuroblastoma 179
Figure 6-4 Monitoring the response to treatment in a patient with neuroblastoma 181
Figure 6-5 Serial liquid biopsy sampling results compared to tissue biopsy at time points where tissue biopsy was available 184
Figure 7-1 Steps needed for integration of a liquid biopsy assay into the clinical practice. 199
Appendix Figure 1 The number of SNVs and indels reported by targeted panel sequencing in tissue (A) and cfDNA (B) in patients with different cancer types at relapse 216

List of Tables

Table 1-1 Features of selected PCR-based and next-generation sequencing (NGS) based approaches for ctDNA detection 36
Table 3-1 cfDNA extraction kits most often used in liquid biopsy studies 63
Table 3-2 Targets and their selection criteria for ctPC panel 67
Table 3-3 Differences between ctPC and ctPC_v2 panels. 72
Table 3-4 Variants called correctly in each SeraCare dilution sample 83
Table 3-5 cfDNA NGS panel analysis guidelines for tumour informed and uninformed analysis. .. 85
Table 3-6 cfDNA samples in Stratified Medicine Paediatric Programme where fusions covered by ctPC panel were identified in RNA 98
Table 4-1 cfDNA-unique variants present in the matched tissue below the limit of detection of tissue reporting..... 115
Table 5-1 cfDNA profiling of patients with CNS tumours on various trials 151
Table 5-2 SNVs detected in cfDNA in Stratified Medicine Paediatric Programme patients with CNS tumours..... 155
Appendix Table 1 Regions showing consistently poor capture in validation runs in HD and Promega control samples 203
Appendix Table 2 Comparison of single nucleotide variants expected from the tumour sequencing and detected in the plasma sequencing using ctPC panel..... 204
Appendix Table 3 Molecular profiling results for patients with ctDNA and/or tissue biopsies taken at more than one time-point. 208

List of abbreviations

AD	Allele depth
BBB	Blood brain barrier
cfDNA	Cell- free DNA
CHIP	Clonal haematopoiesis of indeterminate potential
CNV	Copy number variant
CSF	Cerebrospinal fluid
CSF-cfDNA	Cell- free DNA derived from cerebrospinal fluid
CTC	Circulating tumour cell
ctDNA	Circulating tumour DNA
ctPC	Circulating tumour DNA Paediatric Cancer panel
ddPCR	Digital droplet polymerase chain reaction
DP	Depth at the position
DIPG	Diffuse intrinsic pontine glioma
ETMR	Embryonal tumour with multi-layered rosettes
EWS	Ewing sarcoma
FF	Fresh Frozen
FFPE	Formalin-Fixed Paraffin-Embedded
HMW	High molecular weight
IGV	Integrative Genomic Viewer
Indel	Insertion and/or deletion
lcWGS	Low coverage whole genome sequencing
LoD	Limit of detection
LOH	Loss of heterozygosity
LP	Lumbar puncture
MRD	Minimal residual disease
MRI	Magnetic resonance imaging
NGS	Next generation sequencing
NHS	National Health Service
NSCLC	non-small cell lung cancer
SIOP	International Society of Paediatric Oncology
SNV	Single nucleotide variant
SV	Structural variant
TMB	Tumour mutation burden
UMI	Unique Molecular Identifier
VAF	variant allele frequency
WES	Whole exome sequencing
WGS	Whole genome sequencing

Authors contribution

This thesis describes the research I conducted during my PhD project at the Institute of Cancer Research, London, UK. The work presented in this thesis is my original research, with significant contributions from colleagues explicitly acknowledged in the thesis.

Due to the complex nature of the projects, they were conducted in collaborative manner. Tumour tissue analysis was performed either through clinical route via NHS or as part of the clinical trials conducted at the ICR, RMH, GOSH and other institutions. For the Stratified Medicine Paediatric Programme tissue analysis (targeted tissue sequencing and RNAseq) was performed by Molecular Diagnostics and Translational research teams and coordinated by Debbie Hughes. Some cfDNA sequencing runs were performed by Michaela Smalley, Paul Carter, Ama Brew and Paula Proszek. Some of the cfDNA extractions and library preparations were performed by Ama Brew. For BIOMEDE and SCOUT trials cfDNA extraction was performed and clinical information collected by Diana Martins Carvalho and Matt Clarke in Glioma team, ICR. Bioinformatics analysis was performed by Translational Research bioinformatics team - Ridwan Shaikh, Sabri Jamal, Lewis Gallagher, Xin Liu, Dolapo Ajayi, Lina Yuan - and Evolutionary Genomics and Modelling team - Claire Lynn, George Creswell, Chela James. Interpretation of the clinical findings for Stratified Medicine Paediatric Programme was performed by multidisciplinary Molecular Tumour Board, with results managed by Minou Oostveen. Clinical information about individual patients was provided by treating clinicians or study leads - Sally George, Paola Angelini, Louis Chesler, Fernando Carceller, John Anderson and Chris Jones.

Clinicians and researchers consulting on the panel design: Elisa Izquierdo Delgado, Sally George, Louis Chesler, Lynley Marshall, Janet Shipley, John Anderson, Chris Jones, Susanne Gatz, Paola Angelini, Fernando Carceller, Andrea Sottoriva, Michael Hubank.

Chapter 1 Introduction

1.1 Paediatric cancers

1.1.1 Epidemiology and characteristics of paediatric cancers

Cancer is a group of diseases caused by abnormally growing cells that divide uncontrollably and have the ability to invade the surrounding healthy tissue. Cancer is often described as an age-related disease as incidence of most cancers increases with age [1]. Paediatric cancers (arising in children and adolescents who are 0-19 years old) are relatively rare compared to adult cancers, with an estimated 400,000 cases each year worldwide [2,3]. In UK, there are roughly 1,800 new cases of paediatric cancer diagnosed each year [1] but the incidence of paediatric and young adult cancers has been increasing worldwide [3-5]. In high-income countries, the 5-year survival is ~80% for children with cancer, but in low- and middle-income counties, less than 30% of patients are cured [2]. Despite low incidence and high cure rates, paediatric cancers are the leading cause of death by disease for children in developed countries [6]. In addition, long-term side effects and reduced quality of life from highly toxic treatments are common in paediatric cancer survivors, consisting of but not limited to mental disabilities, organ toxicities and secondary cancers [7-10]. Therefore, better, and kinder treatments are needed in this patient population. Better understanding of the dependencies of these cancers and their genomic landscape could allow targeted treatments with less toxic side effects.

1.1.1.1 Epidemiology of paediatric cancers

The relatively low incidence of paediatric cancers coupled with the broad spectrum of tumour subtypes observed in this patient population makes them hard to study. The majority of paediatric cancers are of poorly understood aetiology - it is thought that paediatric cancers are caused by the mix of inherited or constitutional cancer predisposition factors, endogenous developmental mutational processes (such as impaired DNA damage repair pathways), disruption of cell differentiation and exposure to environmental mutagens or oncogenic pathogens [11,12]. Many paediatric cancers arise within developing tissues undergoing

substantial expansion during early organ formation, growth, and maturation therefore it is likely that that the spectrum of drivers of these cancers will be different from the drivers of adult cancers [13].

Indeed, several studies have highlighted multiple differences between paediatric and adult cancers. Firstly, the spectrum of cancer types occurring in paediatric population is different from the one seen in adults (Figure 1-1 A). Medulloblastoma, neuroblastoma, rhabdomyosarcoma, Ewing sarcoma, osteosarcoma and Wilms tumour which are common in children, are extremely rare in adults [13]. Furthermore, gastrointestinal, male genital, lung, and breast cancers, which are the most common types of adult cancers, are rarely diagnosed in paediatric cancer patients [13]. Interestingly, the types of tumours arising ranges markedly in different age groups of paediatric cancer patients (Figure 1-1 B). Overall, in children aged 0-14 years the most common cancers are leukaemia, followed by CNS tumours and lymphomas [3,5,13]. However, while leukaemia is the most common type of cancer in ages 0-14 years, lymphoma becomes the most common cancer type in older patients (aged 15-19) [3]. Epithelial tumours and melanoma are the second most common tumour group in the oldest group of paediatric cancer patients but represent just 1% of all cancer cases in children aged 0-4 years, with incidence increasing with age [3]. A similar trend is seen for lymphomas and bone tumours [3]. The opposite is true for leukaemia, sympathetic nervous system tumours, retinoblastoma, and renal tumours where incidence decreases in higher age groups [3]. In infants (age <1 year), the sympathetic nervous system tumours account for a third of all tumours (Figure 1-1B) with neuroblastoma being the most frequent in children aged 0-4 years (12.5% of all cancer cases in this age group) and very rare in those aged 15-19 years old (0.2% of all cancer cases in this age group) [3].

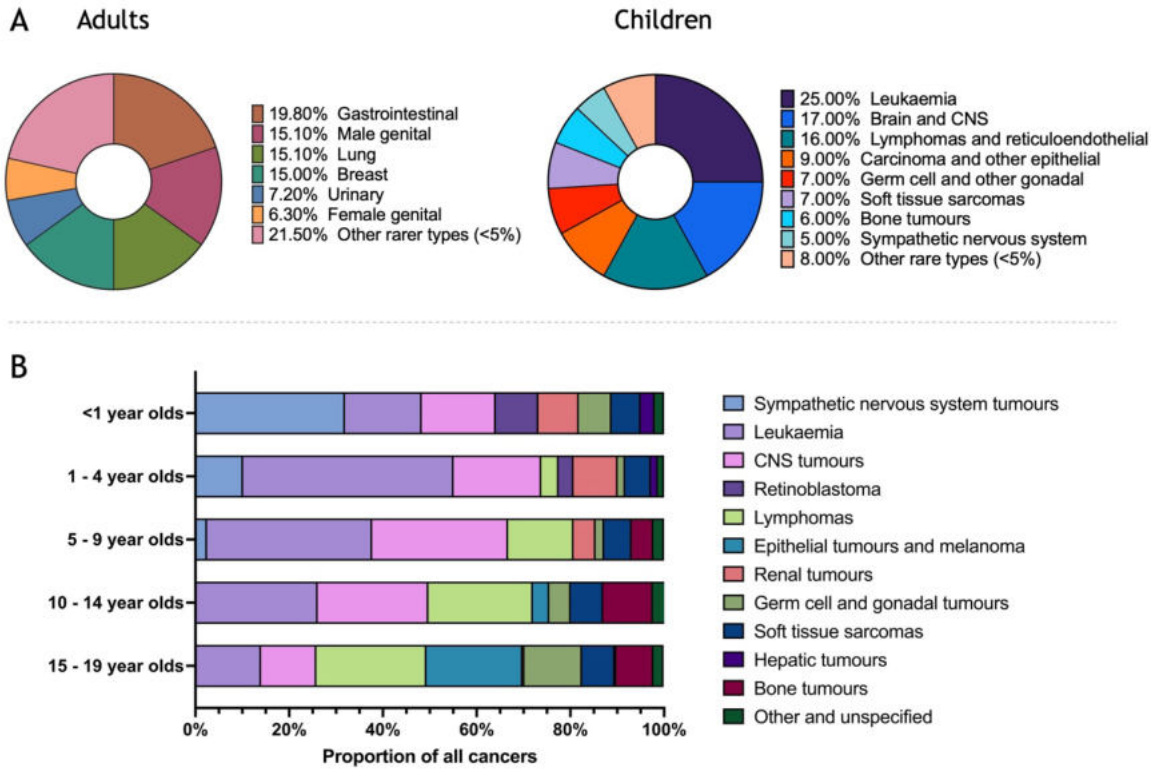


Figure 1-1 The epidemiology of childhood cancer. (A) Frequency of cancer diagnoses in children and adults based on 2000 and 2012 Surveillance, Epidemiology and End Results (SEER) data, Figure adapted from [13] (B) The distribution of cancer types by age for children 0-19 years-old: 0–14-year-olds data from 1998–2007 European cancer registry [5], 15-19-year-olds data combined from 1998–2007 European cancer registry and 2001-2010 worldwide cancer registry study [3]. ALL- Acute lymphoblastic leukaemia, AML - Acute myeloid leukaemia, CNS - central nervous system tumour. Figure adapted from [3,5].

1.1.1.2 The genomic landscape of paediatric solid cancers

Not only the cancer types observed but the genetic features of paediatric cancers differ from adult cancers. Over the last decade, many entity-specific and several pan-cancer sequencing studies have explored the landscape of somatic and germline single nucleotide variants (SNVs), small insertions and deletions (indels), copy number variants (CNVs) and structural variants (SVs) in paediatric cancer patients. It was shown that the mutation burden is significantly lower in paediatric than adult cancers (Figure 1-2) with an increasing number of mutations at relapse[14]. Depending on the study, the median mutation burden in paediatric cancers is around 0.1-1.7 coding mutations per million bases (Mb) [14-19], which is ~14 times lower than the average for adult cancers [14]. The somatic mutation burden increases with patient age except for hypermutated (more than 10 mutations per Mb) paediatric cancer cases [14]. The low mutation burden in paediatric cancers is likely explained by a limited exposure to environmental carcinogens (the highest mutation frequencies are observed in adult melanoma

and lung cancer with >100/Mb due to exposure to ultraviolet radiation and tobacco smoke respectively)[20]. Additionally, the embryonal origin and dysregulation of developmental pathways is thought to result in low mutation burden in these cancers [21]. However, low mutation burden does not imply simple genomes – many complex structural rearrangements (such as chromoplexy and chromothripsis) and copy number alterations are observed in paediatric cancer patients [14,16].

Figure 1-2 Mutation burden of paediatric cancers compared to adult cancers – number of coding mutations per Mb in 24 paediatric (n = 879 primary tumours) and 11 adult (n = 3,281) cancer types (TCGA). Hypermutated and highly mutated samples are separated by dashed grey lines and highlighted with black squares. Median mutation loads are shown as solid lines (black, cancer types; purple, all paediatric; green, all adult). Figure from [14].

The large-scale paediatric sequencing studies also revealed that the overall spectrum of mutations that occur in paediatric cancers are markedly different from adult tumours. Only 30% of significantly mutated genes identified by Gröbner et al. and only 45% of those reported by Ma and colleagues, overlap with adult pan-cancer analyses [14,16]. Even in tumours with similar histology, the mutations observed can be remarkably different. For example, a significant number of paediatric glioblastomas arise in the brainstem as diffuse intrinsic pontine gliomas (DIPGs) and carry K27M and G34R/V mutations in genes encoding histone H3 variants [22,23]. These alterations are extremely rare in adult glial brain tumours, which most often carry *TP53* and *PTEN* mutations [24].

Several pan-cancer paediatric molecular profiling studies were published before the start of this project. One of the biggest studies, analysing over 900 patients with 24 types of major childhood cancers described genetic alterations in 149 putative cancer driver genes [14]. *TP53* was the most common somatically mutated gene, followed by *KRAS, ATRX, H3F3A, NF1* and *RB1* [14]. In a number of smaller studies *TP53* also emerged as the most mutated gene in solid tumours [17,25]. Single nucleotide variants were rare in ependymomas, hepatoblastomas, Ewing sarcomas (driven by *EWSR1* fusions) and pilocytic astrocytomas [14]. Driver gene mutations were common in K27M driven high-grade gliomas, WNT medulloblastomas and Burkitt's lymphomas [14]. An accompanying pan-cancer paediatric study (looking at both leukaemias and solid tumours across 6 histotypes) reported 142 driver genes, with more than half of the driver genes specific to a single cancer type, such as *ALK* and *TERT* in neuroblastoma and *AMER1* in Wilms tumour [16]. *CDKN2A* and *NOTCH1* emerged as the most frequently altered genes in this study, but more than half of the patients in the study had leukaemia, thus explaining the difference in results when compared to other pan-cancer studies [16]. Additionally, whole genome sequencing (WGS) showed high levels of CNVs and structural variants (including chromothripsis) in a high number of paediatric solid tumours[16].

Although overall recurrent mutations are rare in paediatric solid tumours [14], some clinically impactful hot-spot mutations and recurrent genomic alterations exist. Notable examples would be *ALK* F1174 and R1275 mutations and *MYCN* amplification in neuroblastoma [26–29], histone *H3* K27M and G24R/V mutations in glioma[30] and *BRAF* V600E mutation in Langerhans cell histiocytosis [31]. Overall, genes linked to epigenetic modification, cell cycle and transcriptional regulation and MAP-kinase-associated pathways emerged as the most common pathways affected in paediatric solid tumours [14,16,17]. Bigger and more comprehensive studies have been published more recently, broadly supporting these results, and they will be discussed in detail in the discussion section.

Around 10% of paediatric cancer patients carry a pathogenic or likely pathogenic germline variant in cancer predisposition genes [14,32,33]. A high number of cancer predisposition syndromes are known, most common being retinoblastoma, characterised by mutations in RB1 gene, Fanconi anaemia caused by mutations in DNA repair pathways, Li-Fraumeni syndrome characterised by *TP53* mutations and the results of molecular profiling for each patient need to be interpreted in the

light of germline genetic predisposition [34]. Hereditary predisposition is most often reported in patients with adrenocortical carcinomas, hypodiploid B-ALL, K27wt high-grade glioma, osteosarcoma, ATRT, SHH medulloblastoma, and retinoblastoma [14,33]. The most commonly mutated genes in the affected patients are *TP53, APC, BRCA1/2, NF1, PMS2, RB1, LZTR1, TSC2*, and *CHEK2*, majority of which are related to either DNA mismatch repair or double-stranded break repair pathways [14,32,33]. Mismatch repair deficiency is of particular clinical interest, because it offers the potential of immune checkpoint inhibition-based therapy which is undergoing several clinical trials alone or in combination in paediatric cancer patients at the moment [35–38].

In agreement with low mutation burden, a high proportion of paediatric solid tumours had no known pathogenic (known to increase the risk of developing cancer) or potential driver (known to increase cancer clone fitness) mutations detected, even in broad profiling studies. For example, in a comprehensive analysis of 961 tumours of 24 different types of solid paediatric tumours, only 47% of the tumours harboured at least one potential driver mutation and 57% of those had only one driver identified [14]. In another study, only 26% of patients had more than one potentially driver/actionable alteration [17]. In contrast, 93% of adult cancers have at least one driver mutation and 76% have multiple [39]. Most of the significantly mutated genes were shown to be mutually exclusive between different cancer types, compared to more frequent co-mutation in adult cancers [14,16]. A high proportion of paediatric cancers was shown to be driven by structural or copy number changes, such as *MYC, MYCN* or *ACVR2B* amplifications or *CDKN2A/B, TP53* or *SMARCB1* deletions and gene fusions [14,16]. However, the lack of detectable drivers in ~10% of patients [14] raise the question if the current profiling approaches and our knowledge about paediatric cancer drivers is comprehensive enough.

Overall, paediatric cancers have lower mutation burden, lower number of driver mutations and different spectrum of mutated genes when compared to adult tumours. This highlights the need for paediatric specific genetic molecular profiling and tailored treatment strategies.

1.1.1.3 Landscape of actionable genomic alterations in paediatric solid tumours

The treatment of solid paediatric tumours has improved significantly, after moving from only surgical approaches to multi-modal treatments based on chemotherapy combined with radiotherapy [40]. Chemotherapy is the main element of therapy, but irradiation is still essential for many patients with solid tumours [40]. The survival of most paediatric cancer patients has increased gradually over the last decades, but with the intensification of treatment regimens the risk of treatment-related toxic late side effects and treatment-related death has increased as well [7,40]. Survivors of childhood cancer are reported to have a wide range of adverse physical and mental health outcomes, resulting from intensive treatments, such as increased rates of second malignant neoplasms [8-10], endocrinopathies [7], cardiovascular disease [7], and neurocognitive deficits [7]. One way to potentially reduce the exposure to harsh treatments and also provide opportunities for treatment for relapsed/refractory patients running out of standard of care options would be to identify genetic alterations that could be targeted with precision treatments.

Several national and international paediatric cancer molecular profiling platforms have been established to evaluate the clinical utility of identification of targetable alterations and the benefit of matching paediatric cancer patients with targeted therapies. At the time of this study conception, the results from these studies indicated that pathogenic variants can be detected in 40-51% of patients with solid tumours using single platform approach [25,32,41] with increase up to 43-87% when DNA sequencing is combined with other methods, such as RNAseq or methylation analysis [17,19,42,43].

Most studies evaluated the feasibility of a precision cancer medicine in paediatric patients with high risk, relapsed-refractory cancer. In iCat (NCT01853345) 43% of patients had a variant of potential clinical significance identified and 31% of patients received recommendation of therapy based on the results, however only 3 patients actually received the recommended therapy [41]. The most common actionable alterations were in genes of cancer-associated signalling pathways and cell cycle gene mutations in addition to copy number alterations in *MYC/MYCN* [41]. A study using targeted gene panel, specifically designed for paediatric cancer

patients, reported detection of actionable alterations in 51% of patients with significant differences between different cancer types [25]. Over 60% of patients with osteosarcoma, rhabdomyosarcoma, glioma, and other CNS tumours had actionable variants detected, compared to less than 10% of patients with medulloblastoma [25]. Tumour WES identified diagnostic or potentially actionable alterations in 40% of newly diagnosed paediatric cancer patients, with higher levels of detection of known diagnostically relevant alterations in CNS than other solid tumours [32]. In a different study combining WES with tumour RNA analysis to detect gene fusions potentially actionable alterations were identified in 43% of patients with solid tumours [43]. Importantly, in this study the detection of genetic changes resulted in change of treatment for 14 patients (14% of total participants of the study), 9 of which gained clinical benefit [43]. Similarly, in a different multi-platform (tumour tissue WES, lcWGS, methylation and expression microarrays and RNAseq) study, a potentially actionable alteration was detected in 50% of patients with 10 patients (18% of participants in the study) receiving targeted therapy based on the findings [42]. Another study of paediatric patients with recurrent or refractory solid tumours reported detection of actionable alterations in 61% of the patients based on WES and RNAseq analysis [17]. Out of these, 14 patients (19% of total participants of the study) changed therapy and 5 of these patients experienced objective tumour response [17]. A study focusing on hard to treat relapsed and refractory paediatric cancers (including ALL patients) utilised WES and RNAseq and identified actionable alterations in 87% of the study population [19]. A pan-cancer genomic profiling study also reported that nearly 50% of paediatric solid tumours harbour a potentially targetable alteration [14]. All these studies indicate that there is scope for molecular profiling of paediatric solid tumours and matching of patients with targeted therapies. However, only a small proportion of the patients received the therapy recommended based on the presence of specific genetic alterations, mainly due to lack of drug access or a choice of alternative therapy course [17,41].

The wide range of results is in part caused by the lack of standardisation of the definition of “an actionable variant”. The detection of a particular variant can change patient management by informing diagnosis, prognosis or identifying targets for treatment. Some studies include the evaluation of familial cancer risk or non-druggable alterations that can help with risk stratification. Even with a

standardised definition, real-life actionability and impact still depends on multiple factors such as drug and clinical trial availability, which differ in different counties and changes over time. Another reason for differences in the proportion of patients reported to have actionable alterations is variable cohort sizes and disease entities studied. For example, the recent report from GAIN/iCat2 study showed that 83% of the patients had one or more alteration with potential to impact care and 77% of these alterations were gene fusions, because two thirds of the cases in this study were patients with sarcomas [44]. In contrast, a study with a wider range of solid paediatric tumours reported that 47% of potentially actionable variants were SNVs, 46% were focal copy number alterations and only 5% were gene fusions [15]. As this project progressed, more and more evidence of the feasibility of these approaches accumulated. In more recent studies, combining extensive DNA and RNA profiling, up to 86% of patients were described to have actionable alterations [15,18,44,45]. Most of the studies aimed to evaluate the proportion of the patients that have actionable alterations and hence could have clinical benefit from the molecular profiling and results of these studies will be discussed in detail in the discussion section.

The biological pathways affected by alterations considered as actionable in paediatric genetic profiling studies also vary due to different definitions of actionability and the differences in cancer entities included in the studies. Despite that, recurrent mutations were most often identified in genes involved in the receptor tyrosine kinases signalling (especially MAPK, JAK-STAT and AKT/mTor pathways), cell cycle regulation and DNA damage repair pathways [14,17,19,42]. Notably, the signalling pathways affected were not specific to any one tumour type, with key pathways affected across different paediatric solid tumours [14,42].

Overall, these studies of genomic molecular profiling for children with solid tumours highlight that a high proportion of patients carry clinically relevant and/or actionable alterations affecting various biological pathways. This emphasises the potential of personalised medicine in this patient population. However, what proportion of these alterations could be detected using minimally invasive liquid biopsy based molecular profiling methods was unknown at the time of this study conception. In the projects described in this thesis I set out to evaluate the ability to detect genetic alterations using liquid biopsy approaches in patients with various paediatric cancer types at different stages of the disease.

1.1.2 The evolution of paediatric solid tumours from diagnosis to relapse

Despite intensive research, the outcomes for most paediatric cancer patients with high-risk disease at relapse are dismal, and better understanding of the processes driving recurrence are needed. The studies comparing genetic molecular profile of the same patient at diagnosis and relapse are limited, but it has been shown that certain paediatric cancers evolve under standard therapies prior to relapse. Accumulation of SNVs and/or structural rearrangements and copy-number changes at relapse was reported in patients with embryonal tumours with multi-layered rosettes (ETMRs) [46], Ewing sarcoma [47], medulloblastoma [48] and neuroblastoma [49-51]. In patients with ETMRs accumulation of SNVs was observed at relapse while structural variants were largely conserved throughout the disease course [46]. The enrichment of mutational signatures associated with cisplatin treatment at relapse indicate that accumulation of mutations might have been caused by intensive treatment [46]. Similarly, in paediatric Ewing sarcoma, which is driven by *EWS* gene fusion in most cases and has otherwise relatively simple genomic profile, increase in the number of mutations, structural rearrangements and copy-number changes has been reported in patients after treatment [47]. There are few recurrent mutations reported in patients with Ewing sarcoma [52], but the expansion of *STAG2* mutant clones at relapse is associated with highly aggressive disease [53,54]. In neuroblastoma patients, accumulation of SNVs, CNVs and SVs at relapse has been reported with a high fraction of relapse-unique mutations predicted to activate the RAS-MAPK pathway, especially in tumours after chemotherapy [50,51].

Minimal overlap between SNVs detected in primary tumours and subsequent relapses in some patients with EWS [47], osteosarcoma [55] and neuroblastoma [50] highlight the need for molecular profiling of the tumour at relapse to fully understand the disease at relapse. While medulloblastomas have been shown to retain their tumour subgroups (WNT, SHH, Group 3 and Group 4) at recurrence [56] and at metastasis [57], accumulation of new SNVs has been reported at relapse across all subgroups [48]. Most importantly, a switch of driver mutation post-therapy has been reported, where a clone driving the tumour at diagnosis has been eradicated by therapy, leading to recurrence at least in part through clonal selection of a minor clone that was present at the time of diagnosis [48]. In

neuroblastoma, increase in the number of mutations and enrichment of targetable mutations at relapse [50], especially emergence of *ALK* mutations [49–51], has been reported. Targeted analysis of the ALK gene in paired diagnostic and relapse samples showed that *ALK* mutations are enriched at relapse, but they can also be present at low levels at diagnosis in some patients with subsequent clonal expansion at relapse [49].

These studies, describing tumour evolution in paediatric cancer patients, provide strong rationale for molecular profiling of paediatric tumours at relapse with potential to identify targeted therapy options. It is further supported by the pan-cancer paediatric molecular profiling study where several relapse-specific significantly mutated genes, implicated in disease progression and chemotherapy resistance, have been described [14]. Despite this, obtaining a tissue biopsy at relapse is challenging and often not performed. Minimally invasive genomic profiling by liquid biopsy approaches could solve this problem, if proven to recapitulate the genetic features of the tumour at relapse.

1.2 Molecular profiling of cfDNA in paediatric cancer patients

Currently, precision medicine approaches rely on the biopsy of the tumour to identify actionable targets for patients with paediatric solid tumours. The benefits of extensive molecular profiling to detect targetable alterations are convincing, however the acquisition of adequate biopsy can be challenging in children with relapsed and refractory cancer [50,58]. Acquiring tumour biopsies in children exposes patients to additional anaesthesia, ionising radiation, pain and has a complication rate of 6% to 8% [59]. Therefore, utilising less invasive sampling approaches that could reduce the dependency on tissue biopsies and provide complementary information would be of great benefit to patients with paediatric cancer.

Liquid biopsies – the detection of biomarkers present in body fluids - is an attractive alternative approach to tissue molecular profiling. Liquid biopsy is less invasive than conventional tissue biopsy and longitudinal samples are more easily obtainable throughout the course of the disease which has the potential to allow monitoring of the dynamics of the disease. A range of analytes can be studied

using liquid biopsies - circulating tumour cells (CTCs), circulating nucleic acids (cell- free DNA (cfDNA) and the tumour-derived fraction of cfDNA - circulating tumour DNA (ctDNA)), cell-free RNAs (mRNAs, long non-coding RNAs and microRNAs), extracellular vesicles, tumour-educated platelets, proteins, and metabolites - that can be detected in blood, CSF, urine or other bodily fluids [60]. Cell- free DNA (cfDNA) is one of the easiest to collect and one of the best studied bioanalytes in the liquid biopsy field and it is the focus of this thesis.

1.2.1 cfDNA biological features

Circulating cell-free DNA (cfDNA) is the fragmented genomic DNA that is present in biological fluids. In patients with cancer, a subset of cfDNA is derived from tumour cells and hence is referred to as circulating or cell-free tumour DNA (both abbreviated ctDNA). In this thesis the term ctDNA will refer to the DNA present in biological fluids originating from the tumour. When discussing analysis methods, the term cfDNA is more appropriate, as we do not have the tools to separate the ctDNA from cfDNA in the laboratory, hence the analysis is performed on bulk cfDNA.

The detection of cfDNA in the blood was first reported in 1948 [61] and the clinical value of cfDNA by detection of foetal DNA in maternal plasma, allowing non-invasive prenatal testing was shown a few decades later [62]. Additionally, the presence of tumour-derived DNA in the blood of advanced cancer patients was being reported, paving the way for the idea of less invasive cancer diagnostics [63-66]. cfDNA is released by both healthy cells and cancer cells, with the major contribution from the blood cells under normal conditions [67-71]. It is thought to be released from cells undergoing apoptosis, necrosis, and possibly by active secretion (Figure 1-3). Increased cfDNA levels in the circulation are reported under various other physiological conditions, such as after intense exercise [68], acute trauma [72], infection [73], sepsis [71] and transplantation [71,73,74]. Depending on the release mechanism cfDNA can be found in circulation in the form of fragmented single or double-stranded DNA or in complexes with other molecules, such as nucleosomes or inside the extracellular vesicles, such as exosomes [69]. cfDNA is highly fragmented, with the main peak at 166bp which corresponds to the length of DNA wrapped around a nucleosome and a linker DNA associated with histone H1

[75,76]. The cfDNA profile shows periodicity of single and double nucleosome peaks with the predominant single nucleosome peak (Figure 1-3)[76]. cfDNA is subject to nuclease activity in the blood which results in fragment sizes showing a 10bp ladder pattern, consistent with nucleases cleaving the DNA periodically at exposed sites of each turn of the DNA double helix [76-78]. Several studies have explored the biological features of cfDNA, but the results have been contradictory. Most studies report ctDNA to have shorter fragments than cfDNA from non-cancerous cells [77,79] but some show enrichment of tumour derived DNA in higher molecular weight fragments [80]. Generally, it is agreed that ctDNA is more variable in length than cfDNA coming from non-cancerous cells [81], reflecting the changes in chromatin structure in cancer cells as well as high levels of necrosis.

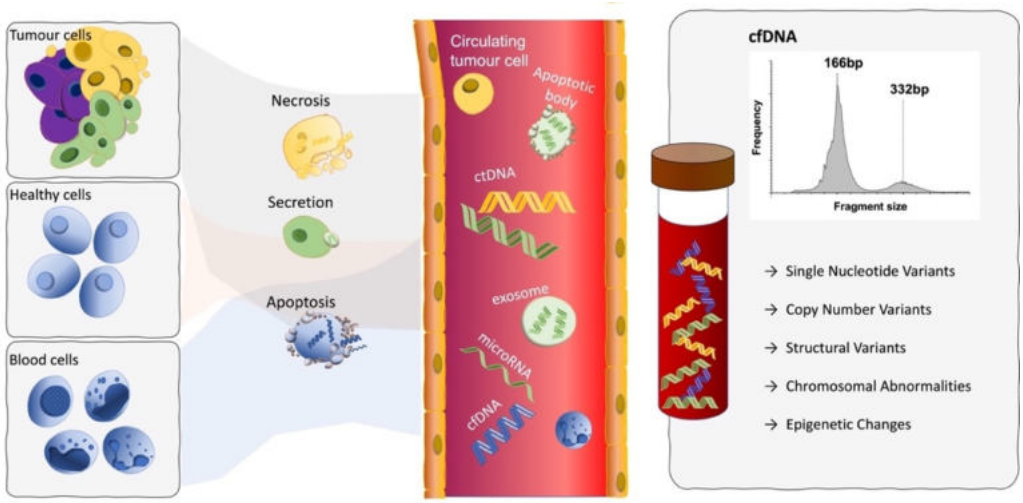


Figure 1-3 Schematic representation of the origins and the range of biological analytes for liquid biopsy approaches. Both healthy and cancer cells release cfDNA into the blood through apoptosis, necrosis, and active secretion. Additionally, other tumour components such as circulating tumour cells, apoptotic bodies, exosomes, and microRNA enter the bloodstream from the tumour. cfDNA exists as free DNA in the bloodstream, in extracellular vesicles or bound to nucleosomes which leads to characteristic fragment size pattern with main peak at 166base pairs (representing DNA bound to a nucleosome) and a smaller peak at 332base pairs (representing DNA from double nucleosome). Blood sample collection and cfDNA analysis allow detection of various genetic and epigenetic alterations that are present in the tumour. Figure adapted from [78,82].

The half-life of cfDNA in the circulation system is very short – estimated to be up to a few hours, therefore allowing close to real-time disease monitoring [78]. cfDNA is cleared from the circulation via nuclease action, renal excretion into the urine, uptake by liver and spleen followed by degradation by macrophages [78]. However, it is important to recognise that analogous to other biological variables, cfDNA shedding into the bloodstream and its metabolism might differ between children and adults [83]. For example, the levels of cfDNA are higher in older healthy

individuals [71], likely due to slower cfDNA clearance or underlying mild conditions. However, studies exploring cfDNA and ctDNA properties in paediatric population specifically are lacking. Therefore, the best practice for analysis and result interpretation is based on adult studies.

1.2.2 ctDNA levels in patients with paediatric solid tumours

In patients with cancer, high levels of cfDNA have been shown to be tumour derived, with increasing proportion of circulating tumour DNA (ctDNA) in patients with advanced and metastatic disease when compared to early-stage patients [69]. The fraction of ctDNA in the cfDNA varies highly between the patients and between different cancer types, ranging from non-detectable levels to almost 100% ctDNA [84–86]. The levels of ctDNA depend on disease stage, tumour load, tumour location and vasculature and it can significantly change during the course of the disease as reported in adult studies [82,84,87,88].

At the beginning of this project, the knowledge of cfDNA levels (and ctDNA fraction in cfDNA) in the blood of paediatric patients with solid tumours was limited. It was known that ctDNA can be detected in patients with solid tumours, such as Ewing sarcoma [89–92], lymphoma [93], osteosarcoma [89,90,94], Wilms tumour [90], renal tumours [95], rhabdomyosarcoma [90], diffuse intrinsic pontine glioma (DIPG) [96,97], with the highest levels reported in patients with neuroblastoma [90,98–103]. In agreement with adult studies, higher levels of ctDNA were observed in high stage, high risk paediatric cancer patients when compared to low stage patients [100,104] and in patients with metastatic versus localised disease [98]. Increased ctDNA levels were shown to be associated with poor survival in Ewing sarcoma and osteosarcoma [89] and with poor prognosis in Hodgkin's lymphoma [93]. Therefore, the ability to detect ctDNA in paediatric patients with solid tumours was established. However due to different methodologies used to quantify ctDNA levels and due to differing sample collection strategies direct comparison of ctDNA levels is complicated. To fill this gap in knowledge this project was designed to collect blood samples at relapse from a range of solid paediatric cancer patients, using standardised procedures that would allow direct comparison of ctDNA and cfDNA levels, including some patients with very rare cancer types.

Towards the end of this project, several other paediatric cfDNA molecular profiling studies were published. MAPPYACTS (NCT02613962) was the biggest study so far evaluating the possibility of cfDNA profiling in recurrent and refractory paediatric cancer patients with extra-cerebral tumours [15]. The feasibility of profiling paediatric cancer patients with CNS tumours using cfDNA from different bodily fluids - plasma, CSF and urine - was recently evaluated as well [105]. Additionally, the possibility of copy number aberration profiling in cfDNA from routinely collected blood samples from patients with various solid tumours was evaluated recently [106]. The results of these and other disease-type specific cfDNA profiling studies that were published recently will be discussed in more detail and in the context of the results of this project in the discussion section.

1.2.3 Potential applications of cfDNA analysis

Cell-free DNA analysis can be utilised in several clinical scenarios, starting from early cancer detection, all the way to molecular profiling and prognostication, monitoring the response to treatment, detecting minimal residual disease, monitoring tumour evolution and emergence of resistance to treatment (Figure 1-4). Liquid biopsy techniques offer the possibility to overcome sampling limitations inherent to tissue biopsy and the minimally invasive nature of sample collection allows serial sample collection and therefore close monitoring of the disease course. Both quantitative and qualitative analysis is possible from a single sample, and it can provide complementary information. This project was focused on molecular profiling and monitoring clonal evolution using liquid biopsy techniques therefore the following sections will focus on these aspects.

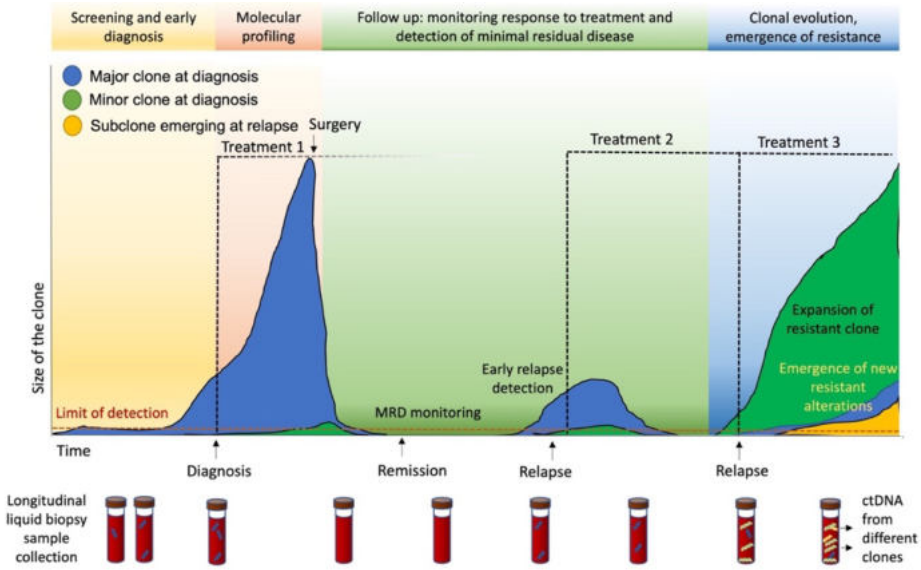


Figure 1-4 The potential applications of cfDNA analysis throughout disease course of a patient enabled by serial sample collection. During the different stages of the disease, different information can be acquired – from screening and early diagnosis to molecular profiling to identify relevant alterations at diagnosis and relapse, to monitoring the response to treatment and early relapse detection. At relapse, molecular profiling of cfDNA can identify emergence of resistant clones and potentially inform treatment options. Additionally, tracking the size of different (sub)clones in the tumour can allow studying clonal evolution. The different cancer clones and their changing proportions in the tumour are indicated by different colours in this schematic. Figure adapted from [78].

1.2.3.1 Molecular profiling of genomic alterations in cfDNA for childhood cancer patients

The isolation and molecular profiling of cfDNA has shown great potential for identification of actionable biomarkers in various adult cancers where mutation detection in cfDNA when compared to time matched tissue samples has been estimated to be highly sensitive (65%-98% detection sensitivity) [78,86,107]. International studies have shown that large-scale standardised cfDNA testing is possible and allows stratifying patients to matched targeted therapies for adult cancer patients [78]. At the time of this project conception, no studies of similar scale were present for paediatric cancer patients, but the evidence for the clinical utility of cfDNA analysis in patients with different types of paediatric tumours was accumulating [83,108,109].

In early studies, the copy number profiling of cfDNA in most common paediatric solid malignancies showed high concordance with tumour tissue profiling [90,98,100]. In patients with neuroblastoma early cfDNA studies showed the possibility to detect MYCN amplification (one of the key poor outcome markers) in ctDNA from

the serum [103]. In fusion positive cancers, patient specific PCR-based assays (informed on the identification of patient-specific breakpoints in the tissue) were able to detect ctDNA even at very low levels (VAF=0.1-0.01%) [90–92,110]. A study comparing ctDNA detection using CNV profile from lcWGS or fusion detection by hybrid capture, highlighted that detection of patient-specific fusion was a more sensitive approach [90]. However, this approach relies on tissue profiling to identify the patient-specific breakpoint.

Targeted molecular profiling also allowed for detection of genomic alterations expected from tissue sequencing in cfDNA in patients with osteosarcoma [94], and Ewing sarcoma [91]. In patients with malignant kidney tumours WES of cfDNA was able to detect tumour tissue specific alterations — CNAs, SNVs or both — in 10 (55.6%), 17 (94.4%) and 17 (94.4%) of the 18 patients in the cohort, respectively [95]. In patients with neuroblastoma WES of cfDNA and tissue biopsy at diagnosis showed overlap of 41% for SNVs and 93% overlap for CNVs [99]. A small study (n=12) where pan-cancer adult specific cfDNA panel was used showed the ability to detect variants in cfDNA expected from tissue profiling for patients with various cancer types in time-matched samples [25]. In another study, BCOR internal tandem duplications were detected in cfDNA, allowing minimally invasive diagnosis of clear cell sarcoma of the kidney (CCSK)[111]. However, in most studies some discrepancies between tissue and cfDNA profiling have been observed – either tissue specific and/or cfDNA unique alterations have been reported.

1.2.3.2 Detection of tumour heterogeneity using cfDNA

Given that solid tumours are often spatially heterogeneous, needle-biopsy sampling as a primary means of diagnosis is problematic as it often provides only a limited snapshot of the tumour's genetic composition [112–114]. Therefore, analysis of cfDNA, as a potential filtrate of a wider sample of the tumour mass could represent a better surrogate indicator of heterogeneity (Figure 1-5), but the utility of this remains to be fully investigated and validated.

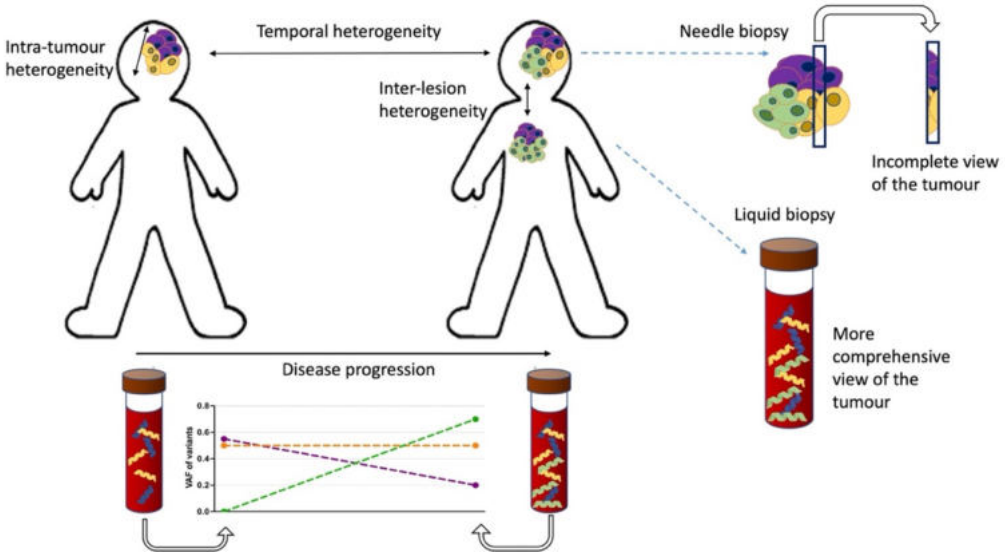


Figure 1-5 Types of heterogeneity and the advantages of cfDNA profiling. There are two types of tumour heterogeneity: spatial and temporal. Spatial heterogeneity is characterised by the presence of distinct variants in different parts of the same lesion (intra-tumour variation) or between different lesions and metastatic sites in the patient (inter-lesion heterogeneity). Temporal heterogeneity is described as the variation in genetic profile over the time course of the disease. Both types of heterogeneity could potentially be evaluated and monitored using liquid biopsy tools. The temporal heterogeneity could be tracked using serial sampling. The problem of subsampling by tissue biopsies in heterogenous cancers could also be alleviated by liquid biopsy methods that represent the genetic diversity of the tumour better. In this schematic the molecularly distinct (sub)clones of the tumour are indicated in different colours.

Firstly, the extent to which ctDNA represents all subclones of a tumour is unknown. Several possibilities can be envisioned, where the more aggressive, highly proliferating clones with high propensity to metastasise are overrepresented in ctDNA. Recent studies in adult malignancies, comparing multi-region and/or multi-lesion tissue sequencing with time matched cfDNA indicate that cfDNA is representative of the diversity of intra-tumoural and intra-lesional heterogeneity [115-117]. However, highly sensitive methods are needed to detect low level subclonal variants in cfDNA. Increased shedding of ctDNA from the more aggressive, resistant clones into the blood has been reported in adult patients with breast, gastro-intestinal, and lung cancer [117-119]. Multi-regional tissue sequencing compared to cfDNA in adult non-small cell lung cancer (NSCLC)[118] and hepatocellular carcinoma [120-122] showed that cfDNA reflects the truncal-branching hierarchy determined by tissue sequencing but it does so incompletely. Additionally, the location of metastatic sites plays a role in the levels of ctDNA and the ability to detect subclonal variants [123,124].

In the paediatric cancer setting, several studies reported the presence of cfDNA unique alterations, not present in the tissue biopsy from the same patient [25,98,99].

In many cases, the tissue and cfDNA samples have been collected at different timepoints, making the interpretation difficult, as the differences between the genomic profiles of tissue and cfDNA can be caused by technical issues (including low sample purity) or could arise due to spatial or temporal tumour heterogeneity. In a pan-cancer paediatric molecular profiling study, better agreement between tissue and cfDNA was observed in time matched samples, with emergence of cfDNA unique variants at relapse [25]. Importantly, in two patients, relapse specific cfDNA unique mutations were retrospectively detected at very low levels in the diagnostic tissue samples [25].

In neuroblastoma, where intra-tumoural heterogeneity and ctDNA levels are high, cfDNA profiling studies are pointing towards the ability of ctDNA to represent spatial heterogeneity but with a potential overrepresentation of the aggressive clones [15,98,99,106]. For example, in a study of patients with neuroblastoma, copy number profiles in 19 cases out of 70 had disagreements between tissue and cfDNA profiles due to technical issues (mainly low tumour purity in either tissue or cfDNA), but several cfDNA and tissue specific CNV breakpoints were also described [98]. In one patient, the cfDNA profile was distinct from the matching diagnostic tissue biopsy, but closely resembled an invaded bone marrow metastasis, highlighting the possibility of cfDNA to represent more aggressive metastatic cell profiles [98]. Another study comparing WES of primary tumours and cfDNA in patients with neuroblastoma reported 93% overlap of CNVs, but only 41% overlap of SNVs between tissue and cfDNA, highlighting higher heterogeneity of SNVs in these tumours [99]. Multi region or multi-site tissue samples were not available in this study to confirm the origin of cfDNA unique variants. However, high levels of overlap between cfDNA unique SNVs at diagnosis and relapse suggest that tumour driving alterations were missed in tissue profiling in these cases. Interestingly, when deep targeted sequencing was performed on the diagnostic tissue, relapse cfDNA unique SNVs were detected in minor subclones in the tissue (MAF mean of 0.84%), giving insight into the evolution of these tumours [99]. The only study so far looking at multi-region sequencing of high-risk neuroblastoma and comparing copy number alterations in tissue with cfDNA from the blood came out recently and revealed high intra-tumoural heterogeneity and multiple differences between aberrations detected in the tissue and cfDNA [125]. Interestingly, even with multi-region biopsy (though mainly only two different sites per tumour) some cfDNA

unique alterations were detected, arguing for added value of combined tissue and cfDNA profiling [125]. High agreement between CNV profiles and more discrepancies in SNV profiles in tissue and cfDNA suggest that CNVs may occur earlier in neuroblastoma progression and therefore contribute less to intra-tumoural heterogeneity. Alternatively, it can be influenced by higher sensitivity of the methods for SNV detection in these studies.

The evidence to date points to the ability of cfDNA to reflect tumour heterogeneity in adult cancer patients, with increased ctDNA shedding from more aggressive, resistant clones, but more studies are needed to fully understand the origin of cfDNA unique variants. The key hypothesis is that cfDNA unique mutations can represent tumour heterogeneity missed by single tissue biopsy, but there is very little information to date about the potential of cfDNA to assess the spatial heterogeneity of paediatric tumours. Larger scale studies, comparing multi region and/or multi-lesion tissue biopsies with time matched cfDNA will be able to answer the question better and inform on the potential actionability of cfDNA unique variants.

1.2.3.3 Disease monitoring in paediatric cancer patients using cfDNA

Another benefit of liquid biopsies is the potential for use as a minimally invasive means of assessing treatment response, surveillance of relapse and identification of resistance mechanisms. This is highly relevant for children where tumours are likely to acquire changes between initial diagnosis and relapse [48–50,125–127]. Less invasive molecular profiling methods could be of high importance for patients with brain tumours, where repeated tissue biopsies are particularly challenging. In the setting of acquired resistance, serial profiling of cfDNA could identify the emergence of clinically relevant resistance alterations, as reported in adult cancers [128,129] and expand our understanding of how the tumours evolve under treatment.

At the time of this project conception, the evidence for the ability to use cfDNA to monitor disease progress in paediatric cancer patients was beginning to accumulate. Changes in ctDNA levels have been shown to correlate with clinical disease status in neuroblastoma [99,101], Ewing sarcoma [90,92,130], osteosarcoma [90,94] and rhabdomyosarcoma [90,110]. Early studies suggested the use of total cfDNA levels

as a surrogate biomarker to evaluate tumour burden and minimal residual disease in patients with neuroblastoma [102]. The levels of cfDNA were shown to correlate with tumour burden during early therapy periods but fluctuations in cfDNA levels caused by other clinical factors such as inflammation and transfusion, hinder the reliability of this approach [102]. Therefore, most later studies rely on more specific methods, allowing quantification of ctDNA levels either by SNV, CNV, SV levels or other cancer specific markers for disease monitoring.

Minimal residual disease monitoring using patient-specific genetic molecular markers is a fundamental tool for assessment of treatment response and risk stratification in acute lymphoblastic leukaemia patients [131]. Similarly, molecular profiling of tumour tissue of patients with solid tumours could provide targets for highly sensitive monitoring by ddPCR of ctDNA in the blood. Indeed, ctDNA levels (as measured by lcWGS copy number profiles) correspond to treatment response, rapidly declining after initiation of chemotherapy but increasing at later points in patients with a clinical relapse in various cancer types [90]. In patients with Ewing sarcoma and rhabdomyosarcoma harbouring gene fusions the changes in ctDNA levels could be tracked sensitively using patient specific fusion detection by ddPCR [90]. Another study utilising patient specific ddPCR assays demonstrated that ctDNA can be detected in the majority of patients with Ewing sarcoma at diagnosis and can be a suitable biomarker for monitoring tumour burden, response to therapy and identifying disease relapse [92]. Several case studies showed the potential of dynamic ctDNA profiling (by assessing the copy number of patient-specific fusion) where ctDNA levels tracked with tumour burden and spiked at the development of new metastases [92]. In a case study of a patient with *PAX3-FOXO1* fusion positive rhabdomyosarcoma the fusion was tracked in cfDNA and allowed early detection of tumour recurrence [110]. Similarly, as in patients with Ewing sarcoma, the fusion gene dosage estimated by quantitative PCR reflected the tumour volume during the disease course [110].

The potential to study tumour evolution using cfDNA was shown in patients with neuroblastoma, where cancer specific alterations decreased in cfDNA in patients responding to treatment and remained undetectable throughout remission [90,99]. In patients experiencing relapse the ctDNA levels (as measured by VAF of tumour-specific alterations) increased during the treatment which corresponded to

disease progression [99]. Importantly, in this study the identification and quantification of different subclones in cfDNA by WES allowed modelling tumour evolution [99]. In one patient the expansion of treatment resistant subclone over the major clone present at diagnosis was shown. In another patient, expansion of multiple subclones together with the major clone at relapse was observed, indicating a different evolutionary pattern. This study highlighted the possibility of cfDNA profiling to shed more light on how paediatric cancers evolve and further studies are needed to determine what factors might affect different evolutionary trajectories in patients with the same disease.

Together these studies show the potential of serial cfDNA sampling to monitor disease course and gain knowledge about tumour evolution. However, limited numbers of patients in each study raise the question of how broadly applicable these methods could be, and more studies are needed to understand what proportion of the patients could benefit from this minimally invasive disease monitoring strategy.

1.2.4 Approaches to cfDNA analysis

Recent technological advances have significantly increased the success of ctDNA detection and new methods, and approaches are constantly developed. The chosen method of cfDNA analysis will depend on the desired application, as much more sensitive methods are needed for the detection of minimal residual disease (MRD) than for molecular profiling of aggressive tumours with likely high ctDNA levels. The major types of assays used for the detection of genomic alterations in cfDNA are NGS and PCR based, each with its own strengths and limitations (Table 1-1). Copy number profiling using single nucleotide polymorphism (SNP) array approach, which is routine in tissue analysis, has also been explored in cfDNA [132], but to a limited extent due to lower limit of detection compared with the aforementioned technologies (Table 1-1). Generally, PCR-based assays are low cost, highly sensitive and perfect for monitoring known variants, but they only allow simultaneous evaluation of limited number of alterations. Untargeted NGS methods (WGS and WES) allow detection of a much broader range of targets without prior knowledge, require more complex analysis and have lower sensitivity, especially in low ctDNA purity samples and they become highly

expensive when higher sensitivity is needed. Targeted NGS methods fall in between the two, with some approaches matching the sensitivity of PCR-based approaches and also allowing detection of unknown genetic variants for cancer detection and monitoring. Additionally, while PCR-based methods can be highly sensitive, they use up all the material, while NGS sequencing libraries generated for targeted profiling can be used for repeated/alternative profiling methods. All of these methods can be used for cancer detection and monitoring with NGS approaches also allowing detection of novel resistance alterations [133]. On the other hand, untargeted NGS approaches allow discovery of unknown genomic alterations, including genome wide copy number alterations and more complex structural variants [133].

Table 1-1 Features of selected PCR-based and next-generation sequencing (NGS) based approaches for ctDNA detection. Abbreviations: PCR—polymerase chain reaction; ARMS—amplification refractory mutation system; qPCR—quantitative real-timePCR; ddPCR—droplet digital PCR; BEAMing—beads, emulsion, amplification, magnetics; SERS—surface-enhanced Raman spetroscopy; NGS—next-generation sequencing; Tam-Seq—Tagged-amplicon deep sequencing; CAPP-Seq—Cancer Personalized Profiling by Deep Sequencing; iDES—Integrated Digital Error Suppression; WES—whole exome sequencing; WGS—whole genome sequencing; LOH – loss of heterozigosity. Table adapted from [132-134].

Analysis method		Technique	Limit of detection	Targets	Advantages	Limitations
PCR based	qPCR	ARMS-PCR, Clamp PCR, COLD PCR	0.01-0.1%	Hotspot mutations	High specificity and sensitivity, low cost, rapid	No multiplexing, limited to detection of known mutations, use up the material
	Digital PCR	ddPCR, BEAMing	0.001-0.1%	Hotspot mutations, gene fusions, CNVs	High specificity and sensitivity, low cost, rapid, absolute quantification	Limited multiplexing capability, limited to detection of known mutations, use up the material
	PCR coupled to spectrometry	SERS	0.1-1%	Hotspot mutations	High multiplexing capacity	Limited to detection of known mutations, lowest sensitivity of PCR-based methods, use up the material
		UltraSEEK				
NGS based	Targeted NGS	eTam-Seq	0.02%	Predefined mutations, indels, CNVs, structural variants	High number of alterations profiled simultaneously, include error correction, allowing high sensitivity, allow detection of novel variants, NGS library can be used for repeated/expanded tests	Amplification based
		Safe-SeqS	0.01-0.05%			
		SPE-duplex UMI	0.1-0.2%			
		CAPP-Seq	0.02%			High input amount needed, capture based
		iDES-eCAPPSeq	0.00025-0.004%			High number of variants (>200) need to be present in the sample

	Untargeted NGS	WES	5%	All coding regions, intron-exon junctions, promoters	Comprehensive profiling, allow detection of novel variants, mutation signature and tumour mutational burden prediction, NGS library can be used for repeated/expanded tests	Low sensitivity (high purity samples needed), bioinformatics expertise needed for variant curation, high amount of data generated
		WGS	5-10%	All coding and non-coding regions, structural genome-wide variants		
Array based	SNP array	SNP Chromosomal Microarray	10-20%	CNVs, structural variants	High resolution for LOH and uniparental disomy	Less comprehensive, low resolution especially for focal changes

The levels of ctDNA can be very low, especially in patients with low tumour burden, therefore highly sensitive methods are required. The key to improving the limit of detection for targeted NGS assays is the ability to discriminate between true and false variants. Even with improved NGS library preparation techniques using high fidelity polymerases with low PCR error rates (~1 error in 10^7 nucleotides) and NGS platforms reporting random error rate up to 1.5% per base call, it adds up to high risk of false positive variant detection, especially in deep-sequencing approaches [133]. To tackle this problem, many computational error-correction methods have been developed to increase the accuracy of NGS [135,136]. Various methods to deal with adapter contamination, PCR biases during library preparation, local sequence context, capture biases and sequencing biases and errors are continuously developed [133,136,137]. The simplest error correction models use uniform error probability models for respective sequencing platform, while more advanced ones employ empirically determined base confusion matrix for each locus [137]. One of the key techniques allowing highly sensitive targeted NGS approaches is the use of unique molecular identifiers (UMIs) - short sequences that incorporate a unique barcode onto each DNA molecule within a given sample library (Figure 1-6). UMIs allow reliable error correction by removal of PCR and sequencing errors and hence reduce the rate of false-positive variants. This is achieved by grouping sequencing reads containing the same UMI and same start-stop position into a single family and creating a consensus read. Variant calling is then performed on the consensus UMI reads, highly reducing the number of false positive calls. Combination of molecular barcoding with UMIs with advanced computational algorithms allowing removal of stereotypical errors associated with

deep targeted sequencing, such as Integrated Digital Error Suppression (iDES) improve the sensitivity of variant detection by ~15fold [138].

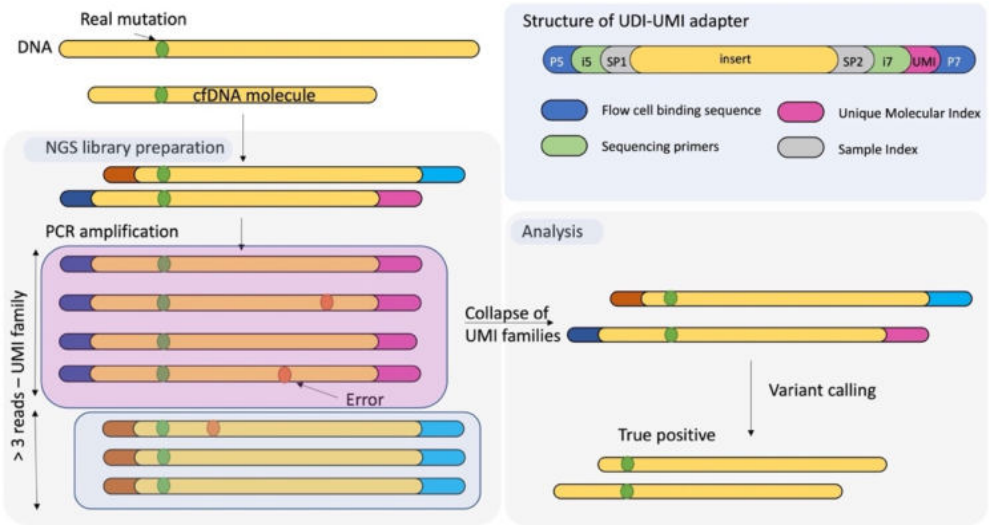


Figure 1-6 Schematic of the of NGS with using UMIs approach for cfDNA molecular profiling. The structure of sequencing adapter with UMIs shown on the right. The cfDNA molecule with a real mutation entering library preparation is molecularly barcoded with a unique molecular identifier (UMI). After PCR amplification, all molecules coming from the original cfDNA molecule will have the same UMI barcode. After sequencing, all molecules mapping to the same genomic location carrying the same UMI are collapsed into UMI families (at least 3 reads needed to create a family). Only variants present in most reads of the same UMI family are called by the bioinformatics pipeline, which allows to remove sequencing and PCR errors. Ultimately true positive variants can be called with higher confidence and errors are removed. Figure adapted from IDT website [139].

1.3 Aims of the project

The knowledge supporting the feasibility of cfDNA monitoring in children with solid tumours has been accumulating and the main aim of this project was to develop a cfDNA based assay that could be used to detect pathogenic mutations and copy number changes relevant to a range of paediatric solid tumours, monitor patients throughout the disease progress and study tumour evolution in a wide range of paediatric solid tumours.

Aim 1 Develop and validate pan-cancer paediatric cancer cfDNA profiling methodology

1.1 Develop a clinically relevant, easily implementable, cost-effective, pan-cancer profiling platform for paediatric solid tumour profiling using cfDNA extracted from plasma and CSF

1.2 Validate the method using control materials with known variants and previously profiled clinical cfDNA samples

1.3 Perform on-going evaluation using previously uncharacterised paediatric cfDNA samples

Aim 2 Define the landscape of genomic alterations detectable by liquid biopsy in a representative sample of paediatric cancer patients diagnosed in UK

2.1 Define the levels of cfDNA and ctDNA and frequency of actionable genomic alterations (SNVs, indels, copy number changes) detected in paediatric patients with different solid tumours at diagnosis and relapse

2.2 Assess the concordance between tumour and cfDNA molecular profiling

2.3 Evaluate the feasibility of cfDNA sequencing to facilitate the identification of targetable changes that can direct the selection of treatment or enrolment in clinical trials

2.4 Evaluate blood based and CSF based cfDNA profiling for patients with CNS tumours

Aim 3 Assess the feasibility to monitor the disease progress and study the evolution of paediatric tumours using serial cfDNA samples

3.1 Assess concordance of genomic alterations between tumour and serial cfDNA samples. Define methods for tracking specific genomic alterations in cfDNA throughout treatment and after relapse in order to monitor treatment response in longitudinal studies

3.2 Determine if liquid biopsy approach allows detection of emerging resistance mutations in patients undergoing targeted treatment

3.3 Determine if liquid biopsy approach allows studying clonal evolution during treatment

3.4 Establish specific evolutionary trajectories of individual patients over their disease course using serial liquid biopsy samples.

Chapter 2 Materials and methods

2.1 Clinical samples and validation materials

2.1.1 Validation materials

Different sample types with variants (SNVs and indels) known from previous experiments or provided by manufacturer were used for panel validation:

1. DNA from four cell blends containing 46 cancer-specific SNVs at known variant allelic frequencies (VAF) of 4-30%, 528 background SNVs and 108 background indels (Tru-Q1-4 HorizonDiscovery, Cambridge, UK).

2. Three clinical paediatric formalin fixed paraffin embedded (FFPE) DNA samples from tumour tissue containing 46 known SNVs of 2-87% VAF.

3. Two artificial cfDNA control blends containing 25 SNVs, 15 indels and 3 focal amplifications in a dilution series resulting in 0.125%-5% VAF (Seraseq Complete and Mutation mix v2, SeraCare, USA).

4. Sixteen clinical cfDNA samples containing 11 SNVs detected by Avenio ctDNA Expanded panel sequencing (Roche, USA) and 10 SNVs confirmed by ddPCR (custom design; Bio-Rad, Applied Biosystems, Thermo Scientific and IDT).

5. For lcWGS limit of detection analysis, DNA from T24 (urinary bladder derived cell line, hypotetraploid to hypertetraploid with multiple copy number abnormalities) was diluted into wild type MRC-5 (diploid human fibroblasts) cell line DNA.

2.1.2 Clinical trials

Further validation was performed on paediatric cfDNA samples with matched tissue sequencing (where available). National Research Ethical approval was obtained for all studies and participants and/or guardians gave informed consent and age-appropriate assent for patients outside of clinical trials as well. Recommended whole blood volume was up to 20ml with smaller volumes collected

according to EMA guidelines for children weighing less than 10kg or as per clinical team discretion.

Plasma samples for cfDNA analysis were obtained from patients enrolled on:

1. The Royal Marsden Hospital (RMH) clinical sequencing pilot study CCR-4294 (REC reference 15/LO/07). Patients were eligible to enrol at any time, including diagnosis, relapse/progression, and remission. Archival tissue was retrieved from the most recent surgery, or a repeat biopsy could be requested at the treating clinician's discretion. The patients had diagnostic and/or relapse tissue (FFPE or FF) sequenced on paediatric tissue capture panel [25,140].

2. Stratified Medicine Paediatric Programme (reference: 246557/264925) - plasma and time-matched tissue samples at relapse were obtained from hospitals across the UK. More than 400 tissue (FFPE or FF, depending on availability) and plasma samples were collected with an aim to genomically characterise relapsed paediatric cancers for improved diagnosis and stratification to therapy. Where possible, archival tissue was retrieved from diagnostic biopsies for these patients. The trial is ongoing at the time of submission of this thesis.

3. The Liquid biopsies for biomarker development, molecular tumour profiling and disease monitoring in paediatric solid tumour patients: a feasibility study (Liquid Biopsies in solid tumours) (CCR-5395). Blood, urine, saliva, CSF, bone marrow and pleural or ascitic fluid was collected for patients at diagnosis and multiple timepoints during treatment and surveillance to evaluate the feasibility of disease monitoring using various bodily fluids. Bone marrow aspirates, cerebrospinal fluid, pleural and/or ascitic fluid were collected opportunistically at the time of planned procedures as part of routine of care. Tissue samples (fresh tumour tissue, formalin fixed paraffin embedded (FFPE) tumour tissue or slides or snap frozen tumour tissue) were collected at diagnosis and surgery, for the patients who underwent surgery.

4. IMI2 (CCR-4892) - a comprehensive paediatric preclinical package of care platform to enable clinical molecule development for children and young adults with cancer (part of an ITCC EU Framework Programme). Blood, bone marrow and tissue samples were collected for patients at diagnosis and multiple timepoints during treatment and surveillance. Tumour samples were collected preferably at diagnosis, surgery or at relapse. Bone marrow samples were collected any time a bone marrow examination was undertaken for clinical reasons. Serial blood samples for cfDNA analysis were collected at clinically meaningful timepoints, depending on cancer type (for example, for neuroblastoma patients - at diagnosis, end of induction, pre-chemotherapy pre-radiotherapy, pre-immunotherapy and at the end of treatment).

5. A Study to Test the Safety and Efficacy of Larotrectinib for the Treatment of Tumours With NTRK-fusion in Children (SCOUT) (NCT02637687). As a secondary objective, blood samples at diagnosis, before surgery and throughout the treatment were collected from some patients for cfDNA analysis and disease monitoring.

6. The Biological Medicine for DIPG Eradication trial (BIOMEDE) (NCT02233049) – a multicentre, randomised phase II study with molecular stratification of DIPG after upfront biopsy. Within the UK, snap-frozen tissue, whole-blood, live tissue, and optionally plasma, taken at diagnosis were collected. As a secondary objective, cfDNA molecular profiling was performed for a subset of patients.

Participants and/or guardians gave informed consent and age-appropriate assent for cerebrospinal fluid (CSF) sample collection for cfDNA analysis during standard of care treatment or as part of clinical trials:

1. The Royal Marsden Hospital (RMH) clinical sequencing pilot study CCR-4294 (REC reference 15/LO/07). Patients were eligible to enrol at any time, including diagnosis, relapse/progression, and remission. The surplus CSF samples from the routine diagnostic by microbiology laboratory were used, which led to suboptimal sample processing (samples kept at room temperature for extended periods of time) and small sample volumes.

2. The Liquid biopsies for biomarker development, molecular tumour profiling and disease monitoring in paediatric solid tumour patients: a feasibility study (Liquid Biopsies in solid tumours) (CCR-5395). Cerebrospinal fluid samples were collected specifically for cfDNA analysis. Sample collection was allowed at any point when lumbar puncture (LP) was performed for diagnosis, assessment or follow up.

3. A sample from cyst fluid was collected for research use form a patient with DIPG undergoing standard of care treatment.

A subset of patients was enrolled on multiple clinical trials throughout their disease course which allowed plasma and/or tissue biopsy collection for research purposes. Where possible, the results were combined to allow monitoring of the disease course by cfDNA analysis.

2.2 Sample processing for cfDNA studies

2.2.1 Blood sample processing

Blood collected into cell-free DNA blood collection tubes (Streck) was separated into plasma no later than 7 days after collection, as advised by the manufacturer and recent studies [141,142]. Blood collected into EDTA blood collection tubes (Camlab) was separated for plasma no later than 4 hours after collection. The separation was performed by double centrifugation - centrifuging the tube at 1600g for 10 min, collecting the plasma fraction and centrifuging again at 1600g for 10 min, followed by plasma storage at -80°C. A 2ml aliquot of blood cell pellet was taken for germline DNA analysis and stored in the fridge for next day extraction or stored at -80°C for later extraction. For blood samples separated in local centres, the plasma samples were transported on dry ice and any additional freeze-thaw cycles were noted for each sample.

2.2.2 CSF sample processing

For the patients on the RMH clinical sequencing pilot study CCR-4294 surplus CSF samples from the routine microbiology diagnostic laboratory were used. CSF was collected into plain universal container tubes, stored at room temperature or 4°C while routine microbiology tests were performed, and surplus CSF was then

collected for cfDNA extraction. When the extraction was not possible on the day of sample receipt, sample was stored at -80°C for later extraction.

For the patients on Liquid Biopsies in solid tumours study, a separate aliquot of CSF was collected during lumbar puncture, spun down at 1600g for 10 min and both supernatant and the pellet were stored at -80°C for later extraction.

2.3 Extraction of DNA

2.3.1 cfDNA extraction

cfDNA was extracted from the plasma using QIAamp Circulating Nucleic Acid Kit (Qiagen) and QIAvac 24 Plus Vacuum Manifold system according to manufacturer's instructions. In cases where the kit was not available, QIAsymphony Circulating DNA Kit on the QIAsymphony automated extraction system (Qiagen) was used instead. After the extraction, cfDNA concentration was measured by Qubit fluorometric assay (Invitrogen) with dsDNA HS (High Sensitivity) or BR (Broad range) Assay Kit, depending on the need. The total cfDNA concentration per ml of plasma was calculated and indicated in ng/ml of plasma. Samples with high ng/ml cfDNA yield were analysed using Agilent TapeStation with the genomic DNA or cfDNA ScreenTape assay to evaluate the level of contamination by high molecular weight DNA.

2.3.1.1 Testing different cfDNA extraction kits

To determine the optimal protocol for cfDNA extraction from plasma, five commercial kits were compared (QIAamp Circulating Nucleic Acid Kit (Qiagen), QIAsymphony Circulating DNA Kit (Qiagen), Quick-cfDNA Serum & Plasma Kit (Zymo), Plasma/serum cell-free circulating DNA Purification midi kit (Norgen Biotek), QIAamp MinElute ccfDNA Mini Kit (Qiagen)). Surplus anonymised plasma from prior adult liquid biopsy experiments was thawed, mixed, and pooled together to generate three samples of 5ml. Each sample was then split into five aliquots of 1ml and frozen at -80°C in cryovials. On the day of extraction, one aliquot of each sample was thawed and extracted according to manufacturer protocol. The cfDNA quantity and quality was assessed using Qubit quantification and by Genomic DNA ScreenTape on Agilent TapeStation (to detect possible high

molecular weight contamination) and High Sensitivity D1000 ScreenTape (to determine the fragment profile).

2.3.1.2 Evaluating the reproducibility of the chosen extraction methods

Surplus plasma from archival liquid biopsy experiments was thawed, mixed, and pooled together to generate three samples as described above. The three samples (1ml each) were extracted in two technical replicates for each technique: 1ml protocol performed as per manufacturer's instructions. The cfDNA quantity and quality was assessed using Qubit (Invitrogen) quantification and by Genomic DNA ScreenTape on Agilent TapeStation (to detect possible high molecular weight contamination) and High Sensitivity D1000 ScreenTape (to determine the fragment profile).

2.3.2 DNA and RNA extraction form tissue and blood cell pellet

DNA from blood cell pellet, cell lines, FF and FFPE samples was extracted using the QIAamp DNA blood mini kit, the QIAamp DNA mini kit and the QIAamp DNA FFPE tissue kit (Qiagen), respectively. For samples from Stratified Medicine Paediatric Programme, the extractions were performed at Specialist Integrated Haematology and Malignancy Diagnostic Service- Acquired Genomics (SIHMDS-AG) laboratory, using the following kits on Promega Maxwell machines (RSC and LEV): Maxwell RSC Blood DNA kit for blood and frozen tissue, Maxwell RSC FFPE Plus DNA extraction kit for FFPE tissue, Maxwell® 16 LEV RNA FFPE Kit for FFPE tissue and Maxwell® 16 LEV SimplyRNA Cells Kit for blood and frozen tissue.

The DNA quantity and quality was assessed using Qubit (Invitrogen) quantification and by Genomic DNA ScreenTape assay on Agilent TapeStation to evaluate the degree of fragmentation of the DNA prior to library preparation.

2.4 Targeted sequencing

2.4.1 cfDNA targeted sequencing panel design

The ctPC panel was designed to enable sensitive and reliable detection of low-frequency variants in cfDNA in the most clinically relevant genes in common paediatric solid cancer types. The design was based on a clinically accredited 91-

gene Paediatric Solid Tumour Panel [25,140], but the size of the region covered had to be reduced to allow the deep sequencing to be economically feasible (Figure 2-1 A). To do that, I focused on the most commonly mutated genes, but also took into account the potential actionability of variants identified. To do that I have:

1. Removed genes with low clinical actionability - all Level 6 (research interest only) and Level 5 (high risk germline SNP) genes (as defined by [140])

2. Removed genes with moderate clinical actionability and low incidence - Level 3 genes (Potentially targetable) with low incidence (<4% in solid/brain cancers reported in PeCan and PedcBioPortal databases and reported in <3 cases in pan-paediatric studies [14,16])

3. Remove big genes (>500kb) with limited clinical actionability (assessed on individual basis)

4. Reconsider if hot-spot regions can be added for any of the removed genes (based on mutations reported in PeCan and PedcBioPortal databases)

After final gene selection, a bed file containing exact chromosome location of each region of interest (ROI) was created. For genes with full coverage, all coding exons with 5bp padding at the end of the exons was included, for genes with hot spots, only the hot spot were included. For two genes (ALK and EWSR1) relative hotspot region of introns where fusions are most often reported were included. To detect Copy Number Variants (CNVs), common heterogenous SNPS in the population (based on gnomAD database) in the gene were included (4-10 SNPs per gene). The bed file was then shared with research scientists at Nonacus, who created an in-silico design of all the probes for ROIs. The in-silico design was again checked to make sure no important regions were missed and regions requiring boosting for even coverage were identified. Improved final version of the panel was then manufactured and tested in validation runs (Figure 2-1 B).

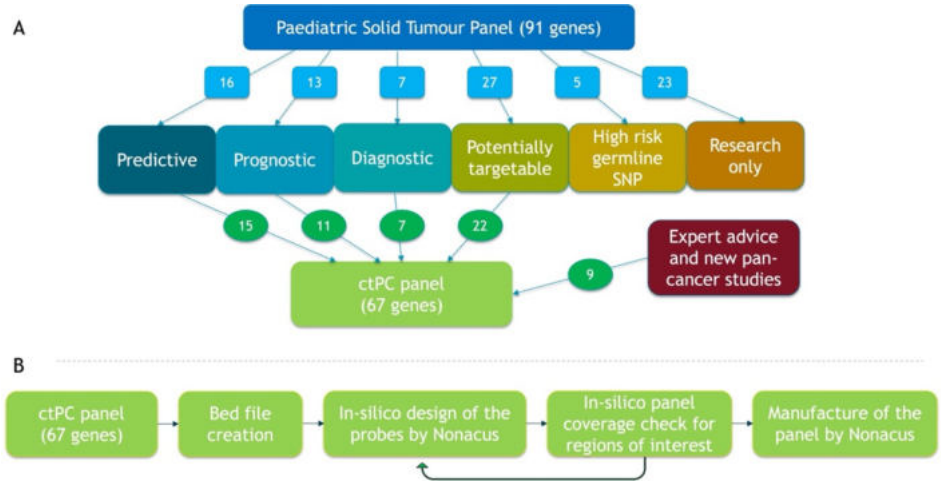


Figure 2-1 Panel design workflow. (A) Gene selection for the panel was based on Paediatric Solid Tumour Panel (91 genes in total), the number of genes in each clinical relevance group indicated. (B) The workflow from gene selection to final manufacture of the panel by Nonacus.

2.4.2 cfDNA library preparation and capture

Sequencing libraries were generated using Cell3™Target (Nonacus) library preparation kit with Dual Index Unique Molecular Identifier (UMI) adapters (IDT) as per manufacturer protocol (FFPE input of 50-200ng, cfDNA 5-50ng) (Figure 2-2). cfDNA and artificial cfDNA samples were prepared according to cell free DNA protocol (with no fragmentation) and genomic DNA (unfragmented cell line DNA, FFPE samples and blood cell pellet samples) was prepared using genomic DNA protocol as per manufacturer instructions. Samples containing < 20ng of cfDNA were amplified for 8 cycles at the pre-capture PCR step, cfDNA samples with 20-50ng – 6 cycles. Sequencing libraries were then pooled (up to 16 libraries per pool) and captured using ctPC (**c**irculating **t**umour DNA in **P**aediatric **C**ancers) custom design panel (Cell3™Target, Nonacus). Where possible, cfDNA samples were pooled separately from blood cell pellet DNA samples to avoid preferential capture of higher input genomic DNA. At least one pool in each sequencing run included one of artificial cfDNA controls 0.125%-5% VAF (Seraseq Complete or Mutation mix v2, SeraCare, USA) for intra-run control and blank control consisting of water. After the hybridisation, non-targeted regions were washed away. Remaining pool was amplified and quality of the pool was assessed using Qubit (in duplicate) and High Sensitivity ScreenTape on Agilent TapeStation.

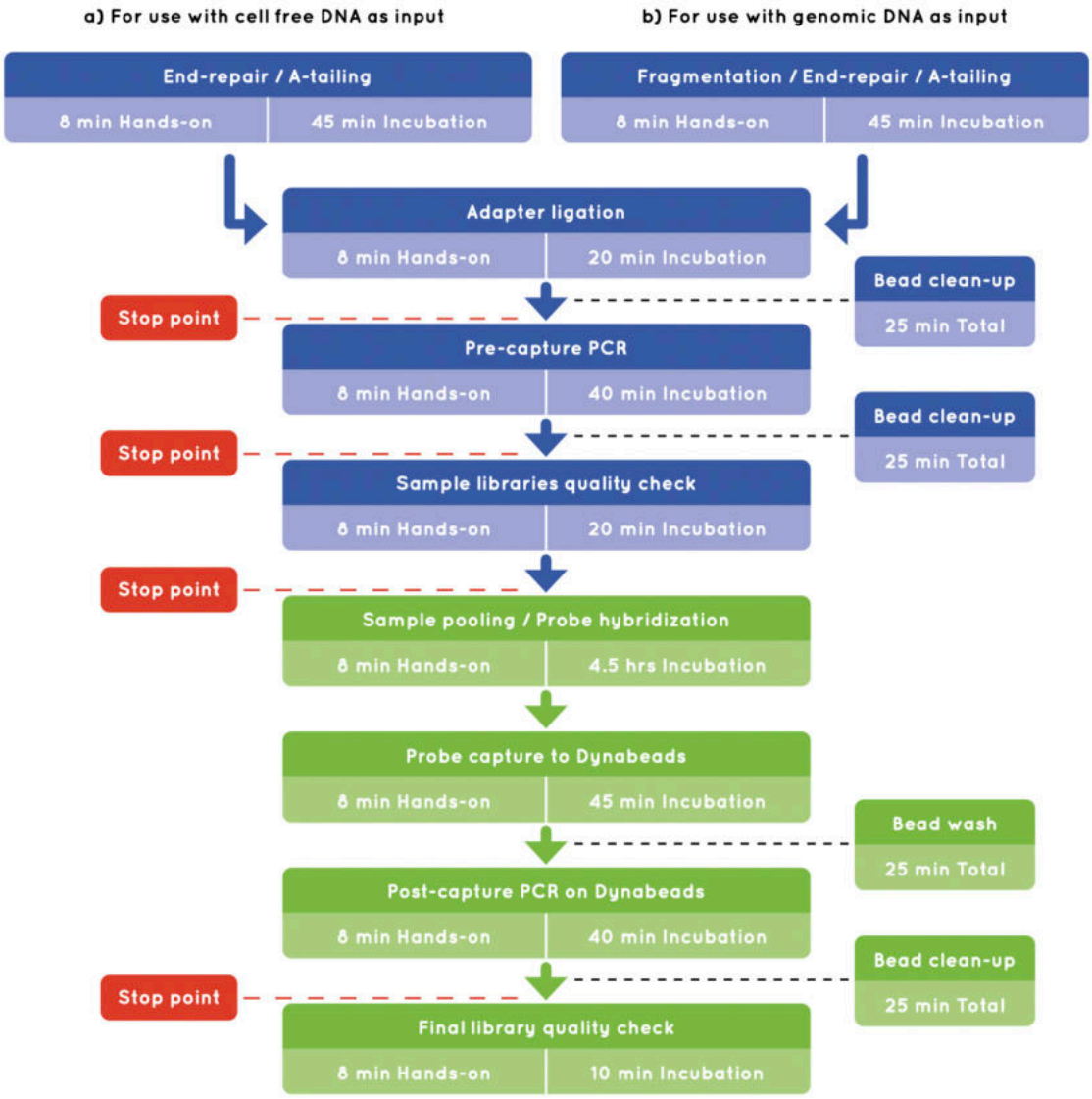


Figure 2-2 Flow chart outlining the main steps of the Cell3™ Target workflow (Nonacus, UK). Blue boxes refer to library preparation steps (~3h); while green boxes refer to probe hybridization / capture and target enrichment steps (~8h). Figure from Nonacus Protocol Guide v1.2.4

2.4.2.1 Modifications to library preparation for suboptimal quality samples

In our studies I have encountered a number of cfDNA samples with extremely high DNA yield due to high molecular weight (HMW) DNA contamination. Preparing sequencing libraries from samples with HMW contamination would lead to low library yields (as the PCR steps in the library preparation are optimised for short length fragments) and highly reduced sensitivity, therefore bead-based size selection was used for samples with >500ng total DNA and <50% of DNA in cfDNA range as determined by cfDNA ScreenTape on Agilent TapeStation.

The bead-based size selection was performed by a modified double size selection protocol using Target Pure™ NGS clean-up beads. Briefly:

1. The 0.6x ratio of beads to sample volume (for example 30μl of beads to 50μl of sample) was added to cfDNA sample eluted in water, mixed well by pipetting, incubated on the magnetic stand for 5 minutes to pellet the beads and the supernatant was recovered to a fresh tube. The tube containing the beads bound to large size DNA fragments (>700bp) was discarded

2. Final 1.8x beads ratio (for example 60μl of beads for the starting 50μl of sample) was added to the supernatant, mixed well by pipetting, incubated on the magnetic stand for 5 minutes to pellet the beads and the supernatant was discarded. Two ethanol washes of the beads were performed, followed by air drying the beads and recovering the fragments by elution in water.

In cases with lower total DNA amount, the loss of DNA during the size selection would lead to lower input of cfDNA into the prep, therefore up to 250ng was used as input for these samples, which would equate to adding ~60ng of cfDNA (given average purity of 25%).

2.4.3 Sequencing library preparation and capture for tissue and blood cell pellet DNA

DNA from the tissue and blood cell pellet for germline analysis were processed by previously established methods [25,140]. Briefly, library preparation was performed using the KAPA HyperPlus Kit and SeqCap EZ adapters (Roche, NimbleGen, Madison WI, USA), following the manufacturer's protocol, including dual-SPRI size selection of the libraries (250-450 bp). Based on the level of DNA fragmentation, 200ng of DNA (samples yielding DNA with median fragment length > 1000 bp) or 400ng of DNA if there was sufficient DNA (for samples with DNA < 1000 bp) was used as an input into the library preparation. 1 μg of the pooled library DNA was hybridised to a custom panel of 92 genes (473kb) (NimbleGen SeqCap EZ library, Roche, Madison, WI, USA).

Sequencing libraries for a subset of blood cell pellet samples were prepared using Cell3™Target (Nonacus) library preparation kit with a standard fragmentation protocol as described above.

2.5 lcWGS of cfDNA and tissue

To perform lcWGS, the amplified libraries from the library preparation used for the panel capture were used. Equal amounts (in DNA mass) of each library were used to create a sequencing pool (no limit on number of samples pooled, as long as adapter clashes were avoided). Libraries of similar quality (based on qubit value) were pooled together where possible to allow the lowest amount of DNA to be used for each pool (the total DNA amount based on the number of samples, to allow accurate pipetting, minimum 1μl).

2.6 Sequencing

2.6.1 cfDNA sequencing

All pools passing quality control metrics (>1ng/μl concentration in 30μl total volume and absence of significant lower or higher molecular weight peaks on the fragment size analyser) were sequenced using Illumina technology. The first validation run of 4 samples was run on Illumina MiSeq (v3 150cycle kit, 150bp paired end reads), the next validation run was run on NextSeq (High output 300 cycle kit, 150bp paired end reads) and all the following runs were performed on NovaSeq (300 or 200 cycle kits, 150bp and 100bp paired end reads). The sequencing runs were filled with samples from other projects, ensuring optimal flow cell loading. For ctPC panel, 8Gb were allocated per sample (which ensured ~1500xUMI coverage), for lcWGS – 4Gb per sample (which ensured ~1x coverage). For lcWGS, the pre-capture amplified libraries from the same prep used for the panel capture, were independently pooled and sequenced separately on a different flow cell using NovaSeq6000 (200 cycle kit).

We experimented with running the ctPC enriched and lcWGS sample simultaneously on the same flow cell. It was possible, but the bioinformatic analysis proved complicated due to the shared indices. While removing the captured regions allows chromosome level CNV to be determined, we lost resolution for individual amplified genes and the profiles were noisy. It proved

more practical to sequence separately on a parallel flow cell. This can be performed alongside other samples captured on the panel, so long as the indices do not clash.

2.6.2 Tissue DNA sequencing

DNA from the tissue and blood cell pellet for germline analysis was sequenced as previously described [25,140]. Briefly, sequencing libraries were pooled with up to 24 samples per pool and sequenced on the NextSeq (2x75bp, mid-output kit). Capture panel pools and lcWGS pools were run on separate flow cells.

2.7 Bioinformatics analysis

2.7.1 cfDNA analysis pipeline using UMIs

A bespoke pipeline for ctDNA analysis has been developed in-house to analyse ctDNA data. The ctDNA pipeline (Figure 2-3) differs from the tissue pipeline [140] in the following two steps:

1. Alignment - where Unique Molecular Identifiers (UMIs) are used as an error correction tool to increase sensitivity and reduce PCR bias.

2. Variant calling - a different variant caller, VarDict was used due to the possibility of calling low allelic fractions, an essential feature when performing variant calling in ctDNA.

The bioinformatic analysis can be divided into four steps: demultiplexing of raw sequencing data, alignment, quality control and variant calling. Demultiplexing is performed using Bcl2fastq2version 2.20.0 (https://emea.support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html) with 0 mismatch in index reads setting. This step separates raw sequencing data into 3 reads and converts them to FASTQ format for downstream analysis. R1 containing the first fragment of DNA, R2 containing the 9bp unique molecular index (UMI) sequence, and R3 which contains the second fragment of sequenced DNA. The three FASTQ files for each sample are then passed into the alignment step of the pipeline. The first step is to generate an unmapped .bam file with each read corresponding UMI sequence annotated in the RX tag. This is

done through the NGS suite of tools Fgbio (https://fulcrumgenomics.github.io/fgbio/) version 0.6.1 with the FastqTo.bam module. The second step aligns reads to the reference genome (hg19) using Burrows-Wheeler Aligner - MEM (BWA-MEM) version 0.7.15 (http://bio-bwa.sourceforge.net/). Next, reads which appear to have come from the same source molecule are grouped together, determined by template and UMI. This is performed with the fgbioGroupReadsByUmi module. Consensus reads are then generated using the fgbioCallMolecularConsensusReads module which examines base-by-base the likelihood of each base at that position within reads with the same unique molecular tag. This outputs a consensus read in a new unaligned .bam file. It is at this stage where the –min-reads option can be set to determine the minimum number of reads to produce a consensus base. This pipeline sets the –min-reads parameter to 3. The following stage filters this consensus .bam with fgbio FilterConsensusReads, filters used are –min-reads 3 –max-read-error-rate 0.05 –max-base-error-rate 0.1 –min-base-quality 30 –max-no-call-fraction 0.1. Filtered consensus reads are once again aligned to the reference genome with BWA-MEM. Finally, before variant calling can be performed, the aligned-consensus .bam file is clipped using fgbio ClipBam. This clips overlapping reads from the same template, so that during variant calling, evidence for a variant is not counted twice when the variant appears on each strand of a single read. Instead, this counts as only a single piece of evidence. Quality metrics are recorded using Picard Tool version 2.8.1 (https://broadinstitute.github.io/picard/) and that native QC files generated by fgbio GroupReadsByUmi. Variant calling is performed with VarDictJava version 1.5.8 (https://github.com/AstraZeneca-NGS/VarDictJava). VarDictJava is supplied with an input .bam file and the BED file of the target panel. The -V/freq argument is set to 0.0001 (0.001%). The pipeline is executed using the Molecular Diagnostics Information Management System (MDx), a web tool that allows users to submit analysis as well as view and report results from next-generation sequencing data.

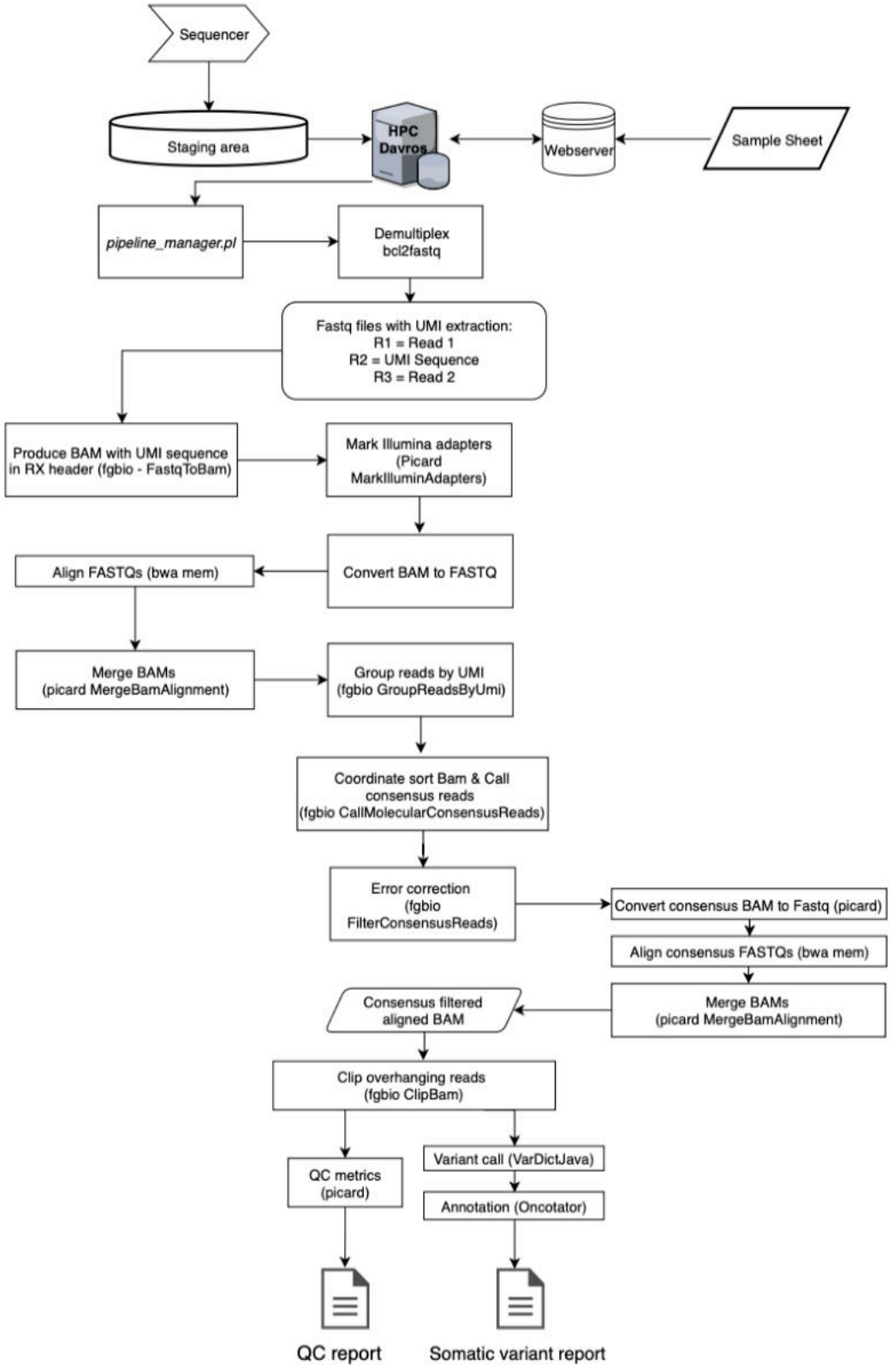


Figure 2-3 ctDNA analysis pipeline. Pipeline development and figure by L. Gallagher

2.7.2 DeepSNV error correction

The background error correction model was created using DeepSNV version 1.34.1 [143], using 13 blood cell pellet samples from paediatric cancer patients with varying diagnoses. The blood cell pellet DNA was extracted using QIAamp DNA blood mini kit and libraries were prepared using the Nonacus fragmentation protocol as per manufacturer instructions. The pool was sequenced to the same depth as cfDNA samples to create the model of noise in each genomic location. The error correction model was then run on a number of cfDNA samples, evaluating the confidence of each variant called by the pipeline.

2.7.3 Tissue analysis pipeline

The analysis of tissue and blood cell pellet DNA for CCR – 4924 and Stratified Medicine Paediatric clinical trials has been performed using Molecular Diagnostics Information Management System v4.0, based on genome build hg19. In summary raw data is converted to FASTQ files using the bcl2fastq2 software, FASTQs are then aligned to the hg19 reference genome using BWA-MEM and quality metrics are generated using Picard Tools. Single nucleotide variants are calling using Mutect2 from the GATK4 suite of tools and are annotated using PCGR which contains a number of variant databases including but not limited to VEP, TCGA, gnomAD and dbSNP. For tumour samples, structural variants are calling using the Manta software and are annotated with AnnotSV and internal tandem duplications are called with Pindel. CNVs were called using a proprietary tool developed by the Clinical Genomics team at the Royal Marsden. The results were interpreted and reported by the Clinical Genomics team at the Royal Marsden.

For patients on other trials, where tissue analysis was performed at other centres or as part of standard of care, the results of clinical reports were used for analysis.

2.7.4 Primary tissue analysis pipeline

The analysis of primary tissue panel and lcWGS sequencing for Stratified Medicine Paediatric clinical trial was performed by Dr. Claire Lynn. Somatic variants occurring at the time of diagnosis were called jointly with relapse tissue and cfDNA using using GATK Mutect2 (v4.2.3.0) in multi-sample mode, ensuring greatest sensitivity. Following calling, variants were annotated using vep101.0 and GATK

FilterMutectCalls (v4.2.3.0) was used to mark variants for downstream filtering. Prior to filtering, 13/597 reported variants were not called in the joint analysis. Variants were imported into R with VcfR 1.13.0 and filtered with R scripts as described. Blacklisted, synonymous and non-coding variants were removed, with the exception of upstream TERT mutations and whitelisted variants. Variants required "MODERATE" or "HIGH" severity, VAF >=5%, AD >=5, DP >=10 and frequency <0.0002 in gnomad. Ultimately variants that had not previously been reported in relapse were manually inspected, and likely artefacts were removed. To reduce manual inspection, whilst conserving all reported variants, FilterMutectCalls flags were first used as a guide as follows: "normal_artifact" removed if germline allele fraction was greater than 15% of the allele fraction in the tumour; "slippage" removed if AD < 20 and VAF < 0.3%; "strand_bias" up to AD=10, variants with at least 1 read on the opposing strand were kept, for AD > 10 variants were removed if the count difference between strands exceeded 150%. Variants are displayed using modifications of the oncoprint function from ComplexHeatmap v2.6.2 [144].

For comparison between relapse tissue and cfDNA, VAFs for reported variants were extracted from previously processed vcf files. Mutect2 and Vardict were used to call variants in tissue and cfDNA respectively in a proprietary pipeline from the Clinical Genomics team at the Royal Marsden.

2.7.5 lcWGS analysis

Two lcWGS analysis pipelines have been used throughput the project - ichorCNA [145] and in-house ASCFLP - both showing similar results.

2.7.5.1 IchorCNA

The ichorCNA analysis was performed as per authors guidelines - the base call (bcl) files were demultiplexed using bcl2fastq v2.17.14. Reads were aligned to the human reference genome build GRch37 (Hg19) using Burrows Wheeler Aligner (BWA) v0.7.12. The aligned reads were segmented into 500kb windows and counted using the readCounter function as part of HMMcopy v0.99.0 and the resulting .wig file was used as input for ichorCNA v0.2.0 [145]. IchorCNA utilises a

hidden Markov model (HMM), to predict large-scale copy number alterations, and estimate the tumour fraction.

2.7.5.2 ASCFLP

ASCFLP(https://github.com/georgecresswell/CNA_stability/blob/master/Figure3/1_CNA_calling/00_runASCFLP.R) analysis pipeline developed by Dr. G. Cresswell uses the absolute copy number status calculated by utilising the approach of ASCAT [146]. The first analysis steps are the same as for ichorCNA, except allowing for allow mappability to be =<0.85 to allow calling MYCN (which has mappability of 0.89) and several other relevant bins, followed by the ASCAT equation to describe LogR ratios, a range of purities from 0.1 to 1 is searched for a custom range of ploidies:

$$AveragePloidy = (2 \times NormalFraction) + (TumourPloidy \times TumourFraction)$$
$$CopyNumber = ((AveragePloidy \times (2^{LogR})) - (2 \times NormalFraction) / TumourFraction$$

For each parameter combination the continuous copy number status was calculated of each bin and calculated the sum of squared differences of these values to the nearest positive integer of the modulus. Local minima were then identified in a 3x3 grid search across the parameter space (a minimum is defined by the centre of the grid having the lowest distance) and the local minima with the smallest sum of squared differences is taken as the solution. The absolute copy number state is taken as the integer that is closest to the value calculated for each bin, using the purity and ploidy parameters. If no local minimum is found the purity is assumed to be 0. If the solution produces negative copy number states, these are capped at zero to avoid impossible copy number states. Finally, the purity and ploidy value with the best fit is chosen and checked manually to evaluate if the fit is reasonable (to avoid over or under fitting).

2.7.5.3 In silico size selection on lcWGS

To test if in-silico size selection improves the resolution of CNVs on lcWGS data, we removed the reads in the .bam file coming from the fragments shorter than 150bp and 100bp (two different sets of analysis) and re-run the ASCFLP algorithm.

2.8 Variant curation

After going through the bioinformatics pipeline, the variants called were manually inspected by two independent scientists, who then issued a final report with variants passed. The .vcf files were inspected by the 1st checker and variants were manually curated using IGV_2.11. software [147] to exclude false positives. Read uniqueness was based on i) minimum of 3 UMI reads for tissue informed and a minimum of 5 UMI reads for novel SNV calling, ii) clearly unique UMIs and unique genomic alignment positions for read pairs, iii) absence of clear strand bias iv) absence of mapping to a different genomic location, v) absence of variant in matched blood cell pellet vi) absence of the variant in unmatched deeply sequenced set of blood cell pellet DNA (DeepSNV version 1.34.1 analysis [143]) and wild type regions of positive cfDNA control (SeraCare artificial cfDNA). The results were interpreted and reported by the Clinical Genomics team at the Royal Marsden.

2.9 ctPC validation statistical analysis

2.9.1 Overall performance

The four HD cell line blends were processed and sequenced as described above. Reads were aligned and the depth and coverage at each region of interest (ROI) was determined. The log mean depth across the panel was compared to the log depth of each ROI and for each gene. To call the region robustly covered, the depth needed to be no lower than 2 SD of the mean. The calculation was based on log2 normalized depth to remove effect due to amplification and deletion.

$$Log2(ROI) > mean(log2(ROI)) - 2 x sd(log2(ROI))$$

Quality and coverage metrics were calculated for all the samples analysed - total reads, percentage of reads mapped to the expected sequence, percentage of duplicates, mean depth (UMIx) across all regions and for each gene on the panel. GC content of the genes covered by the panel and the number of probes per gene was compared against the mean coverage of the gene to evaluate if these parameters influence the achieved depth.

2.9.2 Sensitivity, specificity and accuracy

Variants (SNVs and indels) were called and the presence compared to the intersection of the HD cell line and artificial cfDNA blends list of variants with the regions of interest on the panel. This allowed for the sensitivity and specificity of the panel to be determined within a 95% CI. True positives (TP) are known variants called correctly in a cell blend. False negatives (FN) are known variants that were not called in a cell blend. A false positive (FP) is a variant called in the cell blend, where the variant is not known to exist. Finally, a true negative (TN) is a known variant that is present in one horizon blend and is (designed to be) absent in another. Therefore, we can confirm both its presence and its absence in analysed samples.

Sensitivity, specificity, accuracy, positive predictive value and negative predictive value calculations follow the formulas below.

$$Sensitivity = \frac{TP}{TP+FN} \quad Specificty = \frac{TN}{TN+FP} \; Accuracy = \frac{TP+TN}{(TP+FP+FN+TN)}$$

$$Rule\ of\ 3 = \frac{3}{N}, \qquad N = Number\ of\ replicates$$

$$Positive\ predictive\ value = \frac{TP}{TP+FP} \quad Negative\ predictive\ value = \frac{TN}{FN+TN}$$

However, when designed experiments, the lack of FN or FP leads to a sensitivity and specificity of 100%. The rule of 3 was used to compensate for the lack of variability in the dataset. For example, for a sensitivity of 100%, the compensated sensitivity will be equal to 100 – (3/N)*100 with 95% CI (N = Number of replicates).

2.9.3 Repeatability and reproducibility

The repeatability (within-run precision) was determined by comparing the background variant data (for variant detection and VAF) in the same run between different samples. Two sets of repeatability data were generated by analysing two independent runs prepared by different users.

Reproducibility (between-run precision) was determined by comparing the HD background variant data (for variant detection and VAF) in four samples prepared by different users and sequenced on different sequencing runs.

2.9.4 Limit of detection

To assess the limit of detection and determine a reliable cut off for the reporting (in terms of VAF and number of reads in the .bam file), SeraCare artificial cfDNA with known cancer-specific variants at defined VAFs (range of 5%-0.125% VAF) was used to evaluate the cfDNA analysis workflow. Wild type artificial cfDNA blend was used as a negative control to determine the background noise level.

2.10 Other methods

2.10.1 Droplet digital PCR

The Bio-Rad QX200 Droplet Digital PCR (ddPCR™) system (QX200 Automated Droplet Generator (AutoGD), C1000 Touch Thermal cycler and QX200 Droplet reader) was used to perform ddPCR. The procedure set up followed manufacturer's instructions. Briefly, AutoGD partitions DNA into ~20,000 nanoliter-sized droplets containing ddPCR supermix, a single nucleotide genotyping assay (a pair of primers and probes), water and target DNA. The PCR reaction is then performed in individual droplets and the plate is read by QX200 Droplet reader. In the reader, the droplets are streamed individually through a detector and signals form mutant (positive) and wild type (negative) droplets are counted to provide absolute quantification of target DNA. Custom design TaqMan-based quantitative PCR genotyping assays (Bio-Rad; Applied Biosystems, Thermo Scientific and IDT, Integrated DNA Technologies) were used to detect mutations.

2.10.2 RNA analysis

RNA was analysed using the TruSight RNA Pan-Cancer Panel (Illumina, San Diego, CA, USA) targeting 1385 genes according to the manufacturer's protocol. Bioinformatic analysis was performed using the RNA-Seq Alignment App v2.0.1 (BaseSpace Sequencing Hub) using STAR aligner (to RefSeq Homo Sapiens/ hg19 genome) and Manta for gene fusion calling with default parameters. Paired end sequencing (2x75bp) generated a minimum of 3 million unique aligned reads per sample, in line with supplier's recommendations. The results were interpreted and reported by the Clinical Genomics team at the Royal Marsden.

2.10.3 dNdS analysis

Primary and relapse variants were called jointly as described above. Briefly, variants were filtered for VAF >= 0.05, AD >= 3, DP >=10 and normal DP >=10. To reduce artefacts without manual inspection, variants were discarded with any FilterMutectCalls flag other than “PASS” as all mutations including synonymous or low impact are required. SMP0342 was removed from the analysis as it was identified as a hypermutator, leaving 251 primary/relapse pairs. dNdS analysis was performed on a consensus panel of 82 genes using dndscv v0.0.1.0 [148]. Three groups of variants were analysed separately: variants in primary samples, variants in relapse samples and variants exclusively occurring in relapse samples.

Chapter 3 Development and validation of the cfDNA analysis workflow

Parts of this chapter were published as part of [149]:

3.1 Introduction

The implementation of liquid biopsies into clinical practice is well underway for adult patients, but its application to paediatric cancer patients lags behind. There are multiple commercial assays for liquid biopsy molecular profiling, encompassing the most common alterations in a range of adult cancers (such as FDA approved Guardant360CDx and FoundationOne Liquid CDx), however, as highlighted in the introduction, paediatric cancers differ from adult cancers both in their complexity and the genes that are most often altered [14,16,18,21,44]. Only 30% of significantly mutated genes identified by Gröbner et al. and only 45% of those reported by Ma and colleagues, overlap with adult pan-cancer analyses [14,16]. This limits the benefits of the application of these assays into the paediatric oncology space. Therefore, the goal of this project was to develop a liquid biopsy focused method that would be:

1) Applicable to a wide range of paediatric solid tumours
2) Sensitive and specific enough so the assay could be used as a diagnostic tool
3) Economically viable so it could be proposed as an assay for the National Health Service (NHS), if shown to be of benefit

In the initial stages, various methods to achieve these objectives were considered. A number of studies in liquid biopsy field focus on a specific cancer type, which allows to narrow down the target gene list and use highly sensitive methods such as ddPCR to detect a limited number of relevant variants [150–153]. Combinations of different ddPCR assays, specific either for the patient or disease type were considered, due to extremely high sensitivity of ddPCR methods and relative ease and speed of the assay. However, not all paediatric solid tumours have recurrent targets and in the case of tumour suppressor genes, the tumour driving mutation can be situated anywhere in the gene. We were also interested in studying paediatric cancer evolution and evaluating if we can detect the emergence of resistance mutations in patients on treatment, therefore a broader profiling method was needed. WGS or WES which allows for extremely wide scope of

profiling are common in liquid biopsy research [15,95,99]. However, WES or WGS on cfDNA requires either high sequencing depth (to account for low ctDNA purity in most samples) which comes with high costs, or it risks missing variants in low ctDNA purity samples. At the time of project conception, there were no commercial or research NGS assays that would be specific for cfDNA analysis for a range of solid paediatric tumours. Having considered this, I decided to focus the project on the development of a paediatric specific NGS capture panel, optimised for cfDNA. This approach would allow high sensitivity and specificity profiling (using UMIs to further increase the sensitivity) of multiple targets using a single assay. Additionally, when thinking about implementation into clinical lab workflows, practical pre-requisites, such as rapid turnaround times, manageable analysis times for clinical scientist, financial viability, and diagnostic yield need to be met. Pan-cancer approach (one method that would be applied to all paediatric cfDNA samples) would make workflows in the laboratory more manageable by saving hands-on time for laboratory scientists, improving turnaround times, and reducing costs when compared to personalised assays.

Somatic copy number variants (CNVs) - gains and losses of genomic regions - including individual genes, chromosome arms or whole chromosomes are common in paediatric cancers [14,16]. Therefore, we included copy number probes to detect focal CNVs in clinically relevant genes into the panel design. To get information on the chromosome level copy number changes, we incorporated low coverage WGS (lcWGS) to the cfDNA analysis workflow as well. lcWGS analysis offers a cost-effective method to detect CNVs and also provides an estimate of ctDNA purity (the fraction of ctDNA in the total cfDNA) which aids the interpretation of panel results.

The development and validation of the panel, optimisation for difficult sample types and full workflow, all the way from cfDNA extraction to result analysis and clinical reporting is discussed in detail in this chapter with the focus on SNV and indel detection.

3.2 Determining the optimal cfDNA extraction method

For paediatric cancer patients the amount of blood that can be safely taken can be a limiting factor sometimes [99]. Therefore, the most efficient cfDNA extraction method should be used. Given a wide selection of cfDNA extraction kits on the market, after a literature review, I selected five kits (Table 3-1) that were compatible with the equipment available and sample types expected in the project for testing, to confirm if the results of the comparative studies could be reproduced in our lab as well. Using surplus adult clinical plasma samples available in the laboratory, I extracted cfDNA using each kit and compared the yield in terms of DNA quantity and quality (by fragment size profile).

Table 3-1 cfDNA extraction kits most often used in liquid biopsy studies, based on literature review in 2019 and previous experience in the laboratory.

cfDNA extraction kit (supplier)	Extraction chemistry	Rationale for selection	Abbreviation
QIAamp Circulating Nucleic Acid Kit (QIAGEN)	Column based	[99,154-156]	QIAamp
Quick-cfDNA Serum & Plasma Kit (Zymo research)	Column based	Experience in the laboratory	Zymo
QIAamp MinElute ccfDNA mini kit (QIAGEN)	Column and bead based	[99]	ccfDNA MinElute
Plasma/serum cell-free circulating DNA Purification midi kit (Norgen Biotek)	Column based	[99]	Norgen
Qiasymphony DSP Circulating DNA Kit (QIAGEN)	Magnetic beads based (automated)	Routinely used in the diagnostic laboratory, automated extraction, minimal hands-on time	Qiasymphony

The Qiagen QIAamp Circulating Nucleic Acid Kit (further referred to as QIAamp) produced the highest yields of cfDNA (Figure 3-1 A), closely followed by Qiasymphony. The reproducibility of the two best performing methods (Qiasymphony and QIAamp) was assessed and both methods showed consistent and reproducible results with QIAamp reliably producing higher cfDNA yields (Figure 3-1 B). In one sample, the extraction using QIAamp resulted in 12-fold higher DNA yield (19ng vs 236ng) when compared to Qiasymphony and high molecular weight

DNA trace was seen in sample extracted using QIAamp method (Figure 3-1 C and D). This is not unexpected, as Qiasymphony method includes size selection for DNA fragments in cfDNA range during the last stages of extraction and QIAamp does not. Higher probability of co-extracting genomic DNA together with cfDNA due to lack of size selection during the extraction process was deemed to not be a significant problem as all plasma samples in the trials will be processed according to the standardised protocol with two centrifugation steps which should result in minimal or no contamination from gDNA. The lack of available plasma samples limited the scope of testing, but the results obtained were consistent with literature review and at this point, the QIAamp kit was selected as the optimal extraction method. Qiasymphony automated method was shown to be the second-best option, especially considering the reduced hands-on time for extraction. It was chosen as a contingency method and could be considered as a favoured method if high numbers of samples with relatively high plasma volumes need to be extracted in future clinical trials.

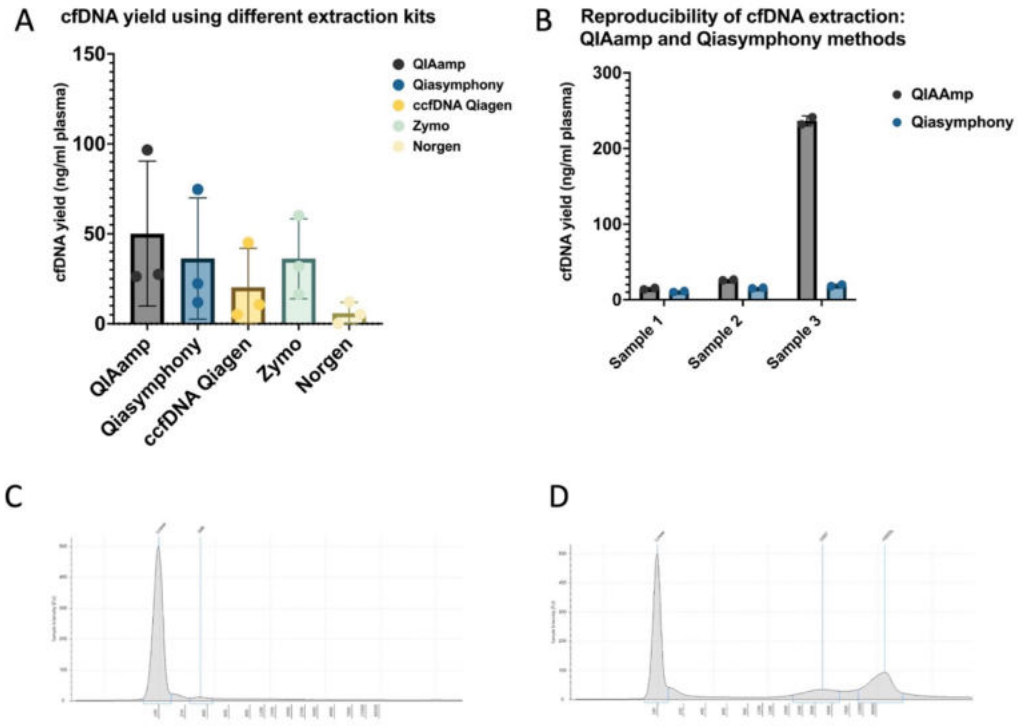


Figure 3-1 Comparison of different cfDNA extraction methods. (A) cfDNA yield from 3 plasma samples extracted using 5 different cfDNA extraction methods, yield of DNA in ng/ml of plasma. Each dot representing one sample B. Reproducibility of cfDNA extraction – cfDNA yield from three samples extracted in duplicate using QIAamp and Qiasymphony methods. C Fragment size analysis of sample 3 extracted using Qiasymphony – no significant high molecular weight DNA contamination visible. D Fragment size analysis of sample 3 extracted using QIAamp – significant high molecular weight DNA contamination.

3.2.1 Further validation of cfDNA extraction methods

At a later stage in the project, following optimisation cfDNA analysis method and the custom design panel, I also tested if the extraction method has any influence on the downstream results of panel sequencing. Our hypothesis was that manual extraction using QIAamp kit results in cleaner cfDNA eluate and higher depth of sequencing with less background noise in sequencing data. To assess this, three paediatric plasma samples that had >8ml of plasma were split in half and extracted using manual extraction with QIAamp and automated extraction using Qiasymphony. The NGS sequencing libraries were prepared simultaneously and sequenced in the same sequencing run to evaluate the effect of cfDNA extraction method. The sequencing depth achieved (both total number of reads, number of unique reads and UMIx depth after UMI analysis) was higher when manual extraction was performed for all three samples (Figure 3-2 A). The number of variant calls to manually inspect (indicating higher background noise) varied – two samples had slightly less variants to check in the manually extracted sample (Figure 3-2 B), indicating that hands-on time for analysis would be lower for samples extracted using QIAamp method. Additionally, the overall results of variants reported were consistent between the two extraction methods (a variant expected from tissue analysis reported in two samples in both extraction methods, and one sample where mutation was not detected after either extraction method).

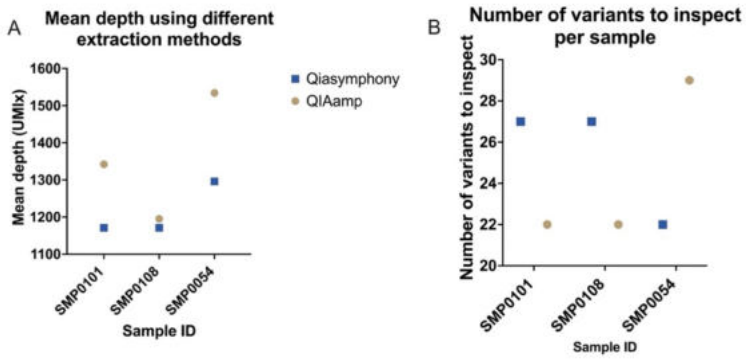


Figure 3-2 Sequencing results in three paediatric plasma samples, extracted using either QIAamp or Qiasymphony method. (A) the mean depth of sequencing achieved using cfDNA sample extracted by either QIAamp or Qiasymphony method (the same DNA amount input into the library preparation) B. Number of variants so manually inspect after bioinformatics pipeline, comparing the cfDNA samples extracted by either QIAamp or Qiasymphony method (the same DNA amount input into the library preparation). The higher the number, the higher the background noise in the sample.

During the time period when QIAamp kit was not available for purchase due to Covid-19 pandemic, Qiasymphony extraction kit was used instead. Afterwards I investigated the quality metrics and confirmed that QIAamp kit does not only produce higher cfDNA yields, but also results in better quality of cfDNA that leads to significantly higher depth of coverage in panel sequencing (Figure 3-3 A). In this case, only samples with 50ng cfDNA input into the library preparation and no significant high molecular weight contamination were compared. Even when removing an outlier (one sample with extremely high cfDNA levels but uncharacteristic fragment size peak which achieved >8000x coverage), and looking at similar number of samples (collected for the same clinical trials), the significantly higher depth of sequencing achieved by QIAamp extracted samples remains (Figure 3-3 B).

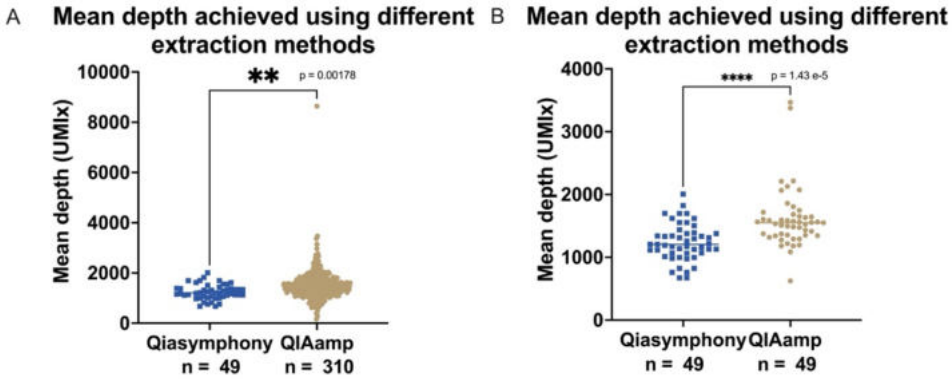


Figure 3-3 Sequencing depth achieved (UMIx) in cfDNA samples depending on extraction method. Only libraries prepared from 50ng cfDNA are shown for comparison purposes. (A) All samples B. Samples from the same trial, profiled at similar time, numbers and time matched. Unpaired t-test, ** p =0.00178, **** p=0.000014.

3.3 Development of the targeted cfDNA panel

3.3.1 Target selection

The NGS capture panel for analysis of cfDNA from paediatric patients with solid tumours (further referred to as ctPC panel) was designed to enable sensitive and reliable detection of low-frequency variants in cfDNA in the most clinically relevant genes in common paediatric solid cancer types. The design is based on a clinically accredited 91-gene Paediatric Solid Tumour Panel [25,140], which is currently in routine diagnostic use in the UK (Figure 3-4 A). To maximise the diagnostic benefit, while maintaining a practical cost for clinical application,

known hot-spot regions of oncogenes and full coverage for important tumour suppressors was considered. I included genes with predictive, prognostic and diagnostic SNVs and CNVs in the most commonly reported solid paediatric tumours [14,16], incorporated the data from publicly available large-scale childhood cancer genomics data sets (pecan.stjude.cloud/, pedcbioportal.kidsfirstdrc.org/ and cancer.sanger.ac.uk/cosmic), the National genomic test directory for cancer in England and consensus opinion from paediatric oncologists. The final panel design covers 67 genes: 33 genes with full coverage of all exons, 25 genes with partial coverage and 9 genes which are only assessed for copy number. Twenty-two genes with coverage of at least one coding exon had supplementary probes to enable CNV detection (Table 3-2).

Table 3-2 Targets and their selection criteria for ctPC panel. Region covered column indicates the extent of gene covered by the capture probes: all coding exons are covered unless otherwise indicated by the exons listed. CNV indicates genes with probes for copy number variant calling and Tx indicates genes with selected translocation regions covered (with the list of introns covered). Basis and Clinical Actionability criteria indicate the clinical relevance of genes included and are based on previous work on Paediatric Solid tumour sequencing [8, 19].

Basis criteria: Level 1- Predictive Biomarker, Level 2 - Prognostic Biomarker, Level 3 - Diagnostic Biomarker, Level 4 - Potentially targetable, inhibitors available or in development, Level 5 - Research only.

Clinical Actionability criteria: Tier 1 - Recognised (FDA/EMA approved) predictive biomarker for response to drug in that indication (OncoKB Level 1), Tier 2A - Recognised standard of care predictive biomarker for response to FDA/EMA approved drug in this indication (OncoKB Level 2A), Tier 2B - Recognised standard of care predictive biomarker for response to FDA/EMA approved drug in another indication (OncoKB Level 2B), Tier 3 - Open clinical trial for predictive biomarker for paediatric solid tumours, Tier 4 -Compelling biological evidence supports biomarker as being predictive of response to drug (OncoKB Level 4).

Trials for Paediatric Solid tumours list currently active clinical trials recruiting paediatric patients where the variant in gene is either used as biomarker, eligibility criteria or is being evaluated as biomarker. For genes with >5 active clinical trials (**) see clinicaltrials.gov for full list.

This resulted in a panel covering 212kb of genomic sequence, including 724 exons, 4 introns and containing 323 copy number probes.

Gene	Region covered exons/CNV/Tx	Basis criteria	Clinical Actionability criteria	Trials for Paediatric Solid Tumours
ACVR1	Exons 6-9	Level 2,3,4	4	
AKT1	Exon 3, CNV	Level 4	3	NCT02813135, NCT04589845, NCT02693535
ALK	Exons 19-29, CNV, Tx (inrt9)	Level 1	3 (mutation), 2A (translocation)	NCT03155620, NCT03107988, NCT03194893, NCT03107988, NCT04094610 **
AMER1	All exons	Level 3		
ARID1A	All exons	Level 5	4	NCT03718091
ATM	All exons, CNV	Level 4	3	NCT02813135, NCT02693535, NCT03155620, NCT03718091, NCT03233204
ATRX	All exons, CNV	Level 3,4	3	NCT03718091

BCOR	All exons	Level 3, 5		
BRAF	Exon 15	Level 1,3	2B (V600E, V600K)	NCT03155620, NCT01677741, NCT02684058, NCT01748149, NCT03155620 **
CCND1	All exons, CNV	Level 4	4	NCT04169074
CCND2	Exon 5, CNV	Level 4	4	
CCNE1	Exon 7, CNV	Level 4	4	NCT03718091
CDK4	CNV	Level 1	3 (amplification)	NCT02813135, NCT02693535, NCT02644460, NCT04238819, NCT03434262
CDK6	CNV	Level 1	3	NCT02813135, NCT02693535, NCT02644460, NCT04238819, NCT03434262
CDKN2A	All exons, CNV	Level 4	3	NCT02813135, NCT02693535
CDKN2B	All exons, CNV	Level 4	3	NCT02813135,
CREBBP	All exons	Level 2		
CTNNB1	Exon 3	Level 2,4	4	NCT01265030, NCT04195399
EGFR	All exons, CNV	Level 1	2B (mutation), 4 (amplification)	NCT02813135, NCT01582191
ERBB2	All exons, CNV	Level 1	2B (amplification)	NCT02693535, NCT04589845
EWSR1	Tx (intr8-intr10, 12)	Level 3	3	NCT03132155 (fusion)
FBXW7	Exons 8, 9, 11	Level 4	4	NCT03718091
FGFR1	All exons, CNV	Level 1	3	NCT02813135, NCT02693535, NCT03155620, NCT03210714
FGFR2	All exons, CNV	Level 1	3	NCT02813135, NCT02693535, NCT03155620,NCT03210714
FGFR3	All exons, CNV	Level 1	3	NCT02813135, NCT02693535, NCT03155620, NCT03210714
FGFR4	Exons 11-18, CNV	Level 4	3	NCT02813135, NCT03155620, NCT03210714
GLI2	CNV	Level 3		
H3F3A	Exons 2, 3	Level 2,3,4	3	NCT02717455
HIST1H3A	All exons	Level 2,3,4	3	NCT02717455
HIST1H3B	All exons	Level 2,3,4	3	NCT02717455
HIST1H3C	All exons	Level 2,3,4	3	NCT02717455
HIST2H3C	All exons	Level 2,3,4	3	
HRAS	All exons	Level 4	3	NCT02813135, NCT03155620, NCT04284774, NCT02285439
IDH1	Exons 4, 8	Level 2,3,4	2B (oncogenic)	NCT03434262, NCT04195555, NCT03749187, NCT04164901
IDH2	Exon 4	Level 4	2B (oncogenic)	NCT04164901, NCT03749187
IGF1R	CNV	Level 4	4	NCT03041701, NCT02306161
KIT	Exons 9-14, CNV	Level 1	2B	NCT02693535, NCT00942877
KRAS	Exons 1-4	Level 4	3	NCT02813135, NCT02285439, NCT03155620, NCT02639546

LIN28B	CNV	Level 2		
MAP2K1	Exons 2,3	Level 4	4	NCT03326388, NCT03434262, NCT03433183, NCT02639546
MDM2	CNV	Level 4	4	NCT04589845, NCT03654716
MDM4	CNV	Level 2		
MET	CNV	Level 1	2B	NCT02693535, NCT02034981, NCT02867592, NCT03598244 **
MYC	CNV	Level 2,4	4	NCT03718091, NCT03838042, NCT04718675, NCT03936465, NCT03434262
MYCN	Exon 2, CNV	Level 2,4	3	NCT02813135, NCT03838042, NCT03936465, NCT01704716, NCT02095132
MYOD1	All exons	Level 2		NCT03462888
NF1	All exons, CNV	Level 4,5	3	NCT03155620, NCT02840409, NCT03326388, NCT02390752, NCT04201457 **
NRAS	Exons 2-4	Level 4	3	NCT02813135, NCT03155620, NCT04417621
PDGFRA	Exons 9-15, CNV	Level 1,2	2B	NCT02813135, NCT03155620, NCT02389309
PIK3CA	All exons	Level 1	3	NCT02813135, NCT03155620, NCT04589845, NCT03213678
PIK3R1	Exons 2,9-11,13	Level 4	3	NCT02813135, NCT03213678
PPM1D	All exons	Level 4		NCT03654716
PTCH1	All exons	Level 3	4	NCT03434262, NCT01878617
PTEN	All exons, CNV	Level 4	3	NCT02813135, NCT03155620, NCT04690725, NCT03213678
PTPN11	Exons 3,8,13	Level 2, 4		
RB1	All exons, CNV	Level 3,4	3	NCT02813135
SMARCA4	All exons	Level 4,5	3	NCT03155620, NCT03654716
SMARCB1	Exons 1-9, CNV	Level 4,5	3	NCT03155620, NCT03654716
SMO	Exons 6, 9	Level 1	4	NCT03434262, NCT04402073
STAG2	All exons	Level 2		
STAG3	All exons	Level 3		
SUFU	All exons	Level 3		
TERT	Upstream promoter	Level 2		
TP53	All exons	Level 2,5	3	NCT04589845, NCT03434262, NCT04585750
TSC1	All exons	Level 1	2B	NCT02813135, NCT03155620, NCT03213678, NCT02693535
TSC2	All exons	Level 1	2B	NCT02813135, NCT03155620, NCT03213678, NCT02693535
WT1	Exons 7,9	Level 3		NCT03652545, NCT02789228

The ctPC panel covers 41% and 52% of genes reported in more than three patients by Gröbner et al. and by Ma et al., respectively (Figure 3 4 B). The purpose of the panel would be to identify targetable alterations and as shown in Figure 3 4 C, the panel covers most of the genetic alterations described as actionable in paediatric malignancies[14]. The genes where translocations are described as actionable were excluded, because detection of high number of translocations in cfDNA is out of

scope for current panel. However, as a proof of concept, selected introns of two genes that have well defined recurrent fusion break points (ALK and EWSR1)[140] were included into this version of the panel. The genes covered are involved in potentially targetable receptor tyrosine kinases signalling pathways, epigenetic and cell cycle regulation and DNA damage repair pathways. To reduce the size of the panel and make it economically viable, genes where mutations are most often of germline origin (such as CHECK2, MLH1, MSH2/6 etc.) were excluded from ctPC panel design, as they would already have been detected by standard of care testing in these patients.

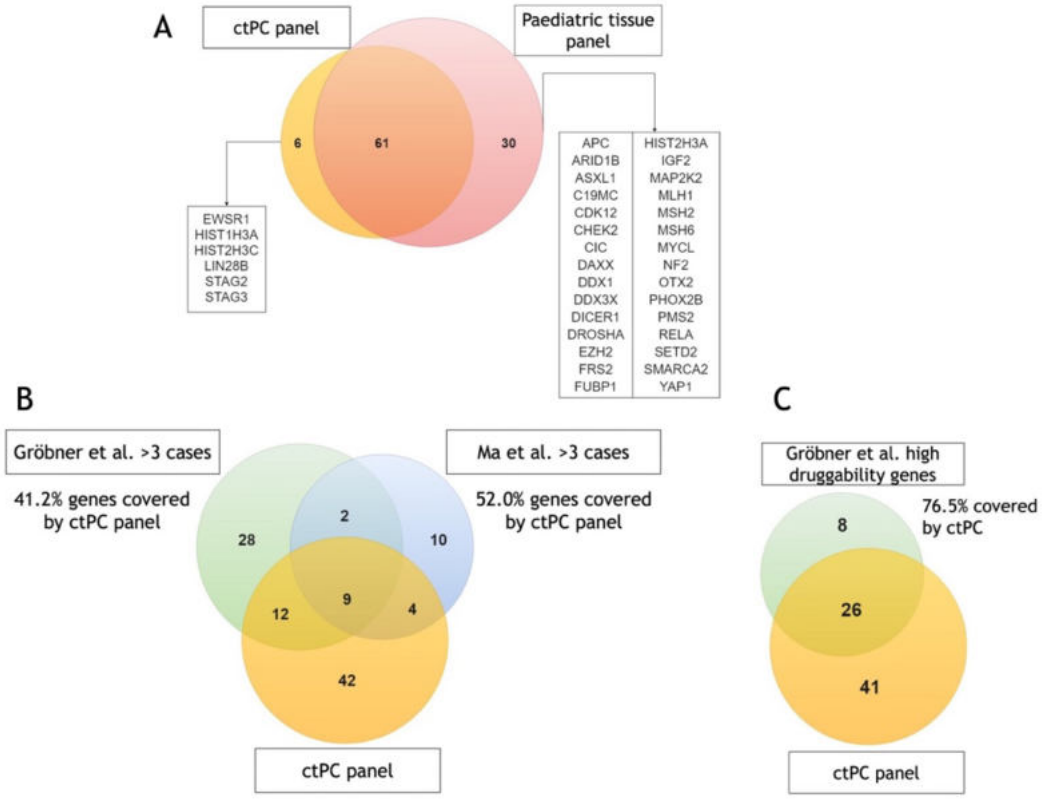


Figure 3-4 The genes on the ctPC panel compared to tissue sequencing panel and paediatric cancer landscape studies. (A) Venn diagram of the overlap of genes covered by ctPC panel and paediatric tissue panel[140] (B) Venn diagram of ctPC panel and two large-scale paediatric landscape studies [14,16]. The numbers indicate the number of genes overlapping between the two studies and with the genes covered by ctPC panel. The genes with variants reported in more than 3 separate patients were intersected with the panel. (C) The overlap between genes classified as highly druggable in paediatric cancers[14] and the ctPC panel.

Toward the very end of this project, several other substantial molecular profiling studies in paediatric cancers have been published [15,44,45]. ctPC panel covers 50%-92% of genes where SNVs and/or CNVs were described as actionable in these studies (Figure 3-5 A-C), highlighting the clinical value of the current panel. Interestingly, the genes not covered by the ctPC panel, did not overlap between

the different studies, possibly because of differences in the cohorts (GAIN/iCat2 consisted mostly of patients with various sarcomas, whereas Zero Childhood Cancer Program focused more on poor-outcome, high risk cancers) or the differences in the definitions of pathogenicity/actionability in different programs. Importantly, 81% (13/16) SNVs and/or CNVs that allowed to match patients to the treatment in GAIN/iCat2 study are covered by ctPC panel (Figure 3-5 D). In the light of this, some of the genes involved in DNA mismatch repair (such as MSH2/6, MLH1) could be considered for the updated design of the ctPC panel.

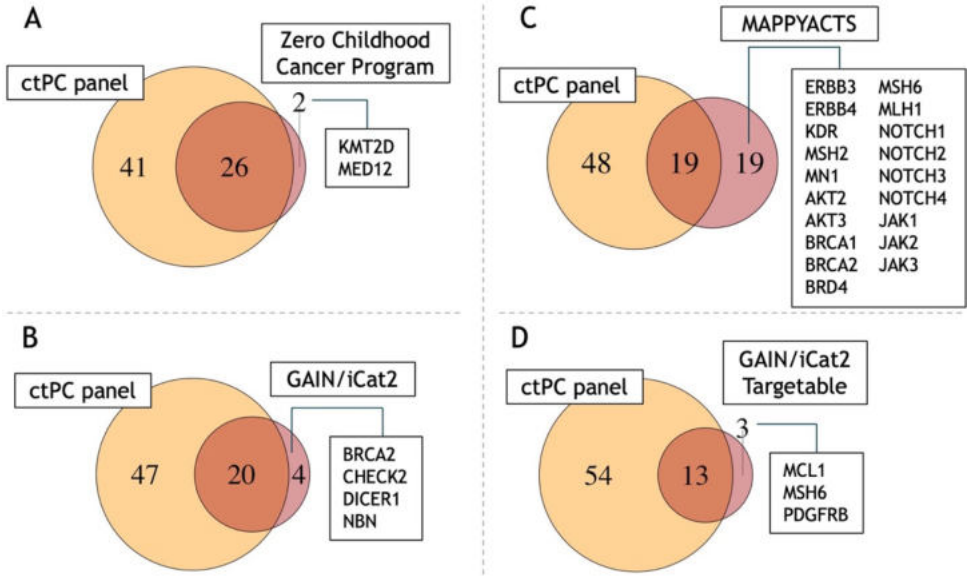


Figure 3-5 The genes on the ctPC panel compared to the recent high profile paediatric cancer landscape studies. The numbers indicate the number of genes overlapping between the ctPC panel and a named study with genes not covered by ctPC panel listed. (A) Venn diagram of the overlap of genes covered by ctPC panel and Zero Childhood Cancer Program [45] (Pathogenic and likely pathogenic SNVs and CNVs detected in >=3 patients in the study) (B) Venn diagram of the overlap of genes covered by ctPC panel and GAIN/iCat2 study [44] (clinically relevant SNVs and CNVs detected in >=3 patients in the study) (C) Venn diagram of the overlap of genes covered by ctPC panel and MAPPYACTS study [15](Potentially actionable SNVs and CNVs detected in >=5 patients in the study) (D) (B) Venn diagram of the overlap of genes covered by ctPC panel and genes where targeted therapy was offered on SNV or CNV detected in this gene in GAIN/iCat2 study [44].

3.3.2 Target selection for ctPC_v2

A few minor changes have been introduced into the second iteration of the capture panel (Table 3-3) to remove a gene that was not required (*STAG3*) and add some additional regions. Full list of regions covered and supporting evidence is provided in Table **3-2**.

Table 3-3 Differences between ctPC and ctPC_v2 panels.

	ctPC	ctPC_v2
Number of genes covered	67	67
Number of genes covered for CNV	30	30
Selected introns covered for translocations	*ALK* and *EWSR1*	*ALK* and *EWSR1*
Additional exons covered		*FBXW7* exons 10 and 12 *KIT* exons 17 and 18
Additional genes covered		*EZH2* (all exons)
Genes removed		*STAG3* (all exons)

3.4 Validation of the ctPC panel for SNV detection

Implementing a new molecular genetic test into diagnostic use requires many levels of assessment – analytical and clinical validation, clinical utility, and social and ethical implications [157]. The ctPC panel was validated according to the standardised clinical assay validation framework (Figure 3-6) [157]. This was the first test of this type developed in our laboratory therefore a full validation was performed - the sensitivity (the proportion of positive results correctly identified), specificity (the proportion of negative results correctly identified), positive predictive value (the proportion of detected variants that are true positives), negative predictive value (the proportion of negative results that are true negatives), analytical accuracy (the proportion of correctly called variants among all variants), the limit of detection (the lowest level of analyte that can reliably be detected by the test) and limitations of critical parameters (DNA input amount, quality) were defined.

The ideal test for cfDNA based molecular profiling would have high accuracy, sensitivity, specificity, limit of detection in the range of ctDNA (<0.1% VAF) and would perform reliably with minute DNA amounts. However, there is an inverse relationship between sensitivity and specificity – as more stringent cut-offs are used to reduce the number of false positives (increase specificity), the likelihood of false negatives increases (reduced sensitivity). Therefore, a balance between the two measurements needs to be achieved to ensure the highest analytical accuracy (the measure of how close the measurement is to the true value). To assess the analytical validity of ctPC method we focused on the validation and verification of laboratory processes and the ability of the test to deliver reliable

and consistent results for SNV and indel detection. The assessment of clinical utility requires high numbers of real-world patient samples and preliminary data on it will be discussed in later chapters.

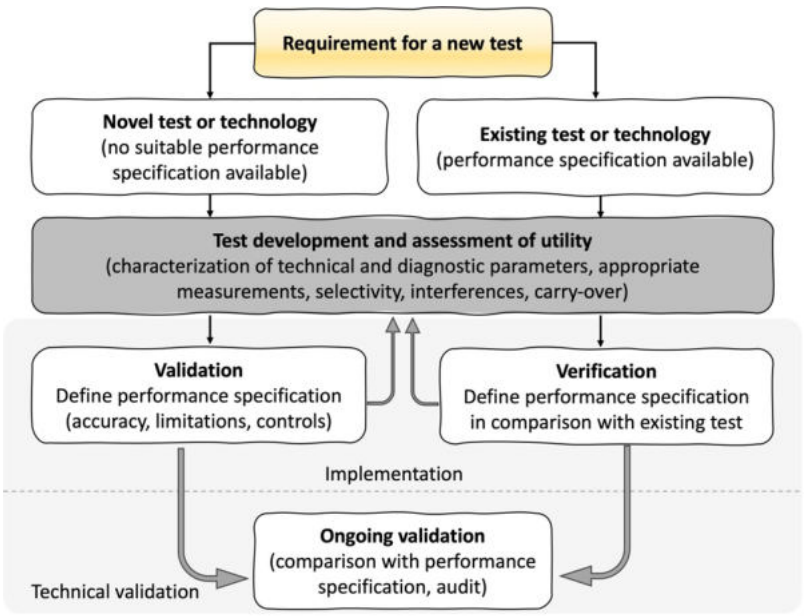


Figure 3-6 The workflow for the implementation of a molecular genetic test for diagnostic use. The shaded arrows represent the two general routes to implementation: validation and verification. Figure adapted from [157].

3.4.1 Validation of the ctPC NGS capture panel

As described in detail in the methods, commercially available reference material was used for the validation: Horizon Discovery® TruQ1-4 blends of cell line DNA with cancer specific variants at known VAF and known background variants (further referred to as HD controls TruQ1-4), Seraseq® ctDNA Reference material with known cancer specific variants at known VAF (further referred to as SeraCare). Additionally, DNA from FFPE tumour samples and previously profiled clinical cfDNA samples were used for the validation.

3.4.1.1 Overall performance

Firstly, to establish if the test is fit for the intended purpose, I assessed if the regions which the panel was designed against are captured sufficiently to detect variants. The capture probe performance showed normal coverage distribution for all relevant sample types (Figure 3-7 A-D) with coverage at each region of interest consistent between the samples and adequate for the purposes of the test (96.6%

of the probes consistently having mean normalised read depth higher than 2 SD of the mean) (Figure 3-7 E). We have identified 34 mutation calling coding probes that showed consistently poor capture (a list of probes in Appendix Table 1), however in most cases the coverage was still sufficient to confidently call variants. The mean depth achieved did not correlate with the GC content of the regions covered by the probes or the number of probes per gene (Figure 3-7 F, G). Only one gene, multi-copy *HIST2H3C*, was consistently underperforming, but given homogenous/repetitive nature of histone genes it was expected, and these probes were kept in the panel for detection of mutations with a caveat that they are unlikely to pass sequencing metrics in most samples.

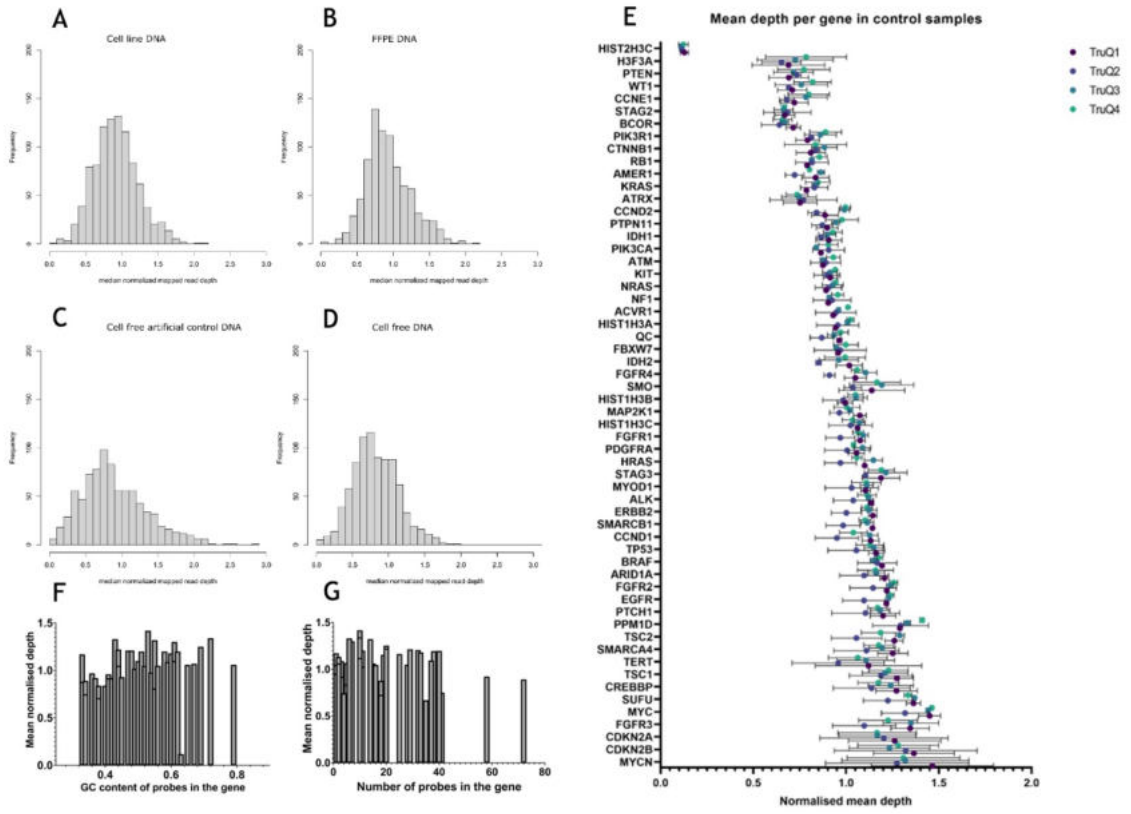


Figure 3-7 ctPC panel performance. (A-D) A representative capture probe coverage distribution for all sample types validated, showing normal distribution. (E) Mean depth for each gene in the panel normalised to mean depth of the all probes in the sample. (F) Mean depth achieved in regions of differing GC content. (G) Mean depth achieved based on number of probes per gene. Data based on cell line control samples TruQ1-4, sequenced in duplicate.

3.4.1.2 Repeatability and reproducibility

To measure precision of the assay, we assessed a set of known variants supplied by manufacturer in the 4 HD cell line DNA control materials (analysis based on 390

SNPs and 96 indels per sample). The repeatability (within-run precision) was determined by comparing the background variant data in the same run between different samples. Two sets of repeatability data were generated by analysing two independent runs prepared by different users. The pairwise correlation of allele frequencies of the variants was consistent among four cell lines within the same run, with 0.986 [95%CI: 98.3% – 99.0%] correlation of SNPs and 0.927 [95%CI: 89.5% – 95.9%] correlation of indels, indicating that the workflow can consistently detect the variants (Figure 3-8).

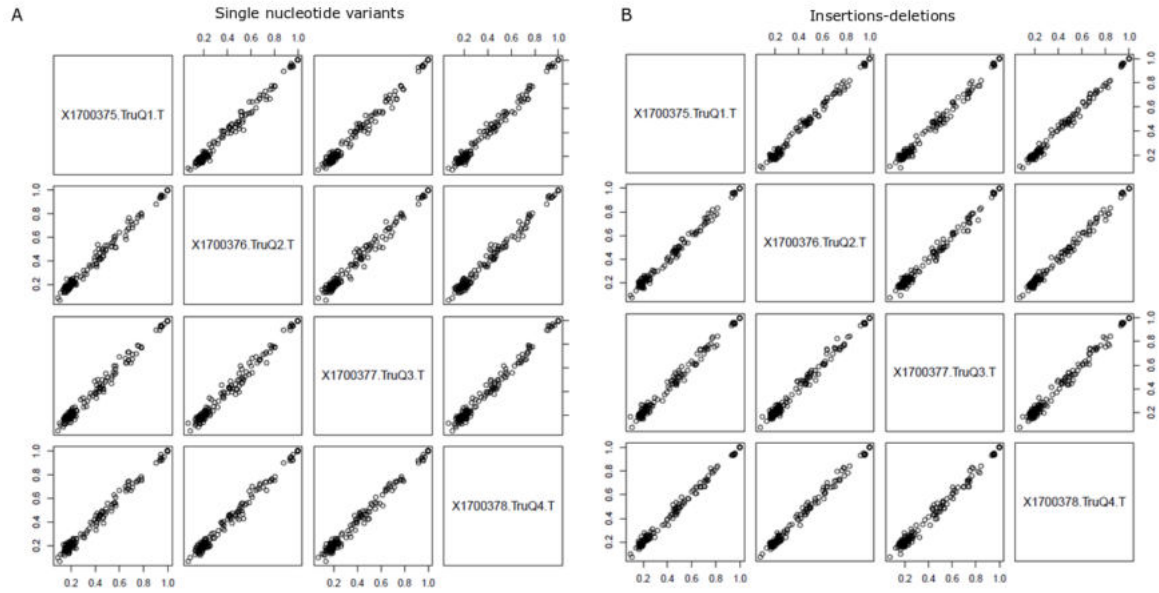


Figure 3-8 Repeatability of ctPC panel sequencing was determined by comparing the Horizon blend background variant data in the same run. All The variants were analysed for detection and variant allele frequency. This indicates if the panel can repeat data from the same samples on the same run. This was done for two independent sequencing runs, generated by different users generating two sets of repeatability data. Consistency of single nucleotide variant (A) and insertions-deletions (B) allele frequency among four Horizon cell lines with identical background variants in the same sequencing run is shown.

The reproducibility (between-run precision) was determined by comparing the HD background variant data in four samples prepared by different users and sequenced on different sequencing runs. The pair-wise correlation of > 0.984 [95%CI: 98.0% – 98.9%] for SNPs and correlation of > 0.896 [95%CI: 86.7% – 94.8%] for indels (Figure 3-9) was reported indicating high reproducibility.

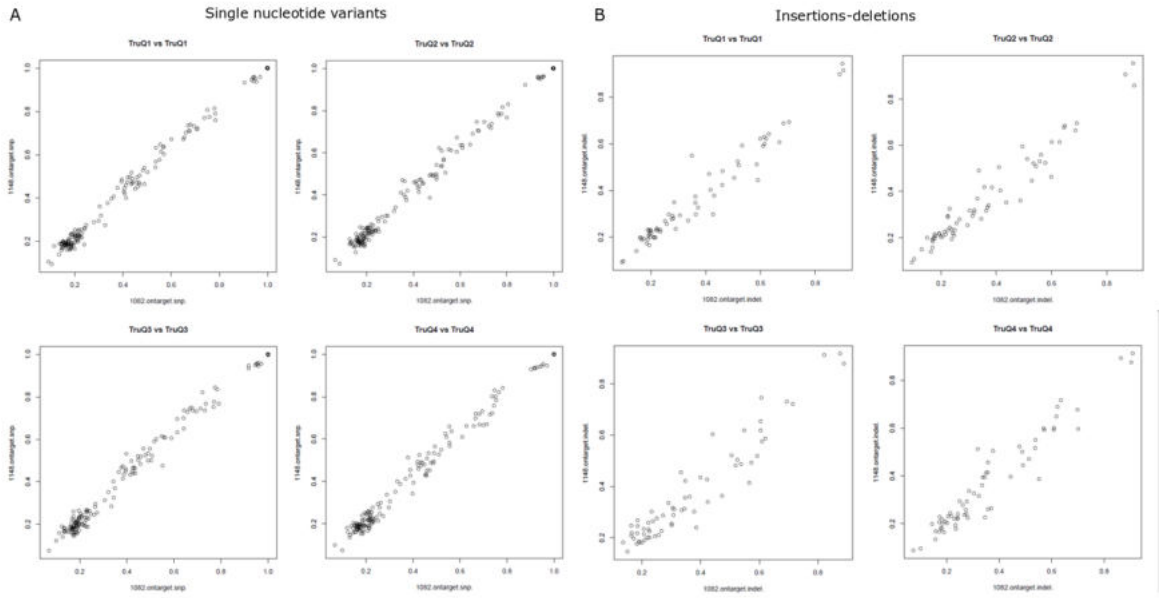
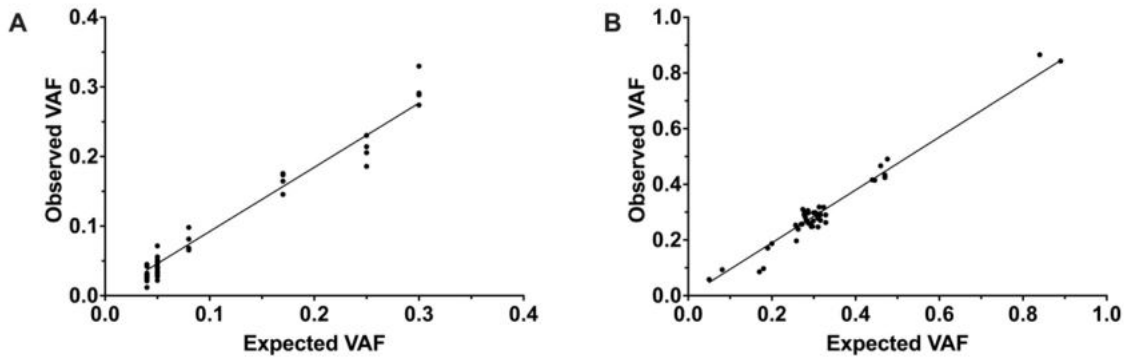


Figure 3-9 Reproducibility of ctPC panel sequencing. The reproducibility was determined by comparing the Horizon blend (TruQ1- TruQ4) variant data in two different runs prepared by different users and sequenced on different sequencing machines. Pairwise correlation of single nucleotide variants (A) and insertions – deletions (B) of each of four samples between two independent runs is shown.

3.4.1.3 Sensitivity, specificity, PPV, NPV and accuracy

Four HD cells blends with known SNVs at known variant allele frequencies and 3 FFPE samples from paediatric cancer patients previously profiled using paediatric tissue panel [140] were used for this part of the validation. Good quality libraries were produced from all the samples, resulting in successful sequencing where all expected SNVs and indels were detected by ctPC analysis (Figure 3 10) with high correlation of VAF values ($R^2 = 0.96$ in FFPE samples and $R^2 = 0.97$ in HD cell line blends). High sensitivity and specificity was achieved for SNV detection (>0.96) with Positive Predictive Value (PPV) of 0.98 (Figure 3 10 C), resulting in high analytical accuracy of 0.99 for SNV detection.



C	SNVs designed to be present	SNVs designed to be absent	
SNVs detected by ctPC panel	61 (TP)	1 (FP)	
SNVs not detected by ctPC panel	0 (FN)	90 (TN)	
	Values	95% CI Lower	95% CI Upper
Sensitivity (TP/(TP+FN))	0.999	0.941	1.00
Specificity (TN/(FP+TN))	0.969	0.940	0.999
Positive Predictive Value (TP/(TP+FP))	0.984	0.914	0.999
Negative Predictive Value (TN/(FN+TN))	1.00	0.959	1.00
Number of replicates	152	-	-

Figure 3-10 Correlation of observed and expected allelic frequency of SNVs in different types of samples. (A) Correlation of allelic frequency of SNVs in HD control samples. VAF observed using ctPC panel vs expected as reported by the supplier (R^2 = 0.97). n=4, SNVs= 46, mean depth = 756x (B) Correlation of allelic frequency of SNVs in paediatric patient FFPE samples. VAF observed using ctPC panel vs expected from Paediatric Solid Tumour Panel (R^2 = 0.96). n=3, SNVs= 47, mean depth = 1080x (C) Contingency table for SNV detection in the HD control and FFPE samples. TP – true positive, FN – false negative, FP – false positive, TN – true negative.

3.4.2 Performance of ctPC workflow in cfDNA

Since the main application of the panel is to detect tumour derived pathogenic mutations in cfDNA, which potentially could be present at very low allelic frequency, we enhanced the limit of detection by use of molecular barcoding with UMIs and a customised bioinformatics pipeline based on the CAPseq method [158]. The specificity, sensitivity and limit of detection were evaluated in the cfDNA context using commercially available cfDNA material (SeraCare) and our in-house bioinformatics analysis pipeline. Additionally, clinical cfDNA samples with known mutations were analysed on this workflow to calculate accuracy and trueness.

3.4.2.1 Limit of detection

To determine the lowest quantity of the analyte that could be reliably detected above the background noise level (in terms of VAF and number of reads in the .bam file) the limit of detection analysis was performed. SeraCare artificial cfDNA with known cancer-specific variants at known VAFs (range of 5%-0.125% VAF) was used to evaluate the cfDNA analysis workflow.

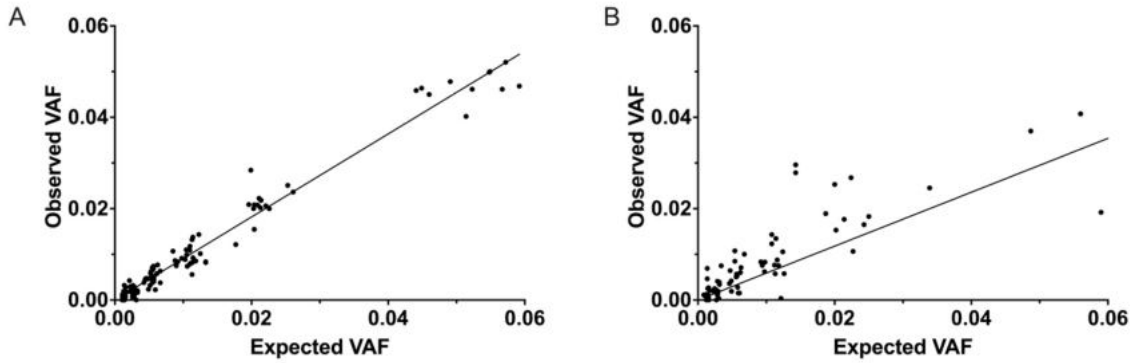
All variants were detected for the 5%, 2% 1% and 0.5% VAF samples, but some of the variants in 1% and lower VAF dilution produced lower VAF than reported by the manufacturer. The biggest deviation in 1% dilution was *EGFR* T790M that was observed at 0.559%. The biggest deviation in 0.5% dilution was *KRAS* G12D that was observed at 0.227%. However, this may be production batch number specific and slight variation between controls from different batches is expected. For the 0.25% sample, 24/25 mutations were detected. *NRAS* Q61R was not detected, but the variant was present in .bam file with 2/3719 reads on Integrative genomics viewer (IGV). For the 0.125% sample, 21/24 mutations were detected. *KRAS* p.G12F was not detected but was present in .bam file in 2/4840 reads when visualised on IGV. *PDGFRA* D842V and *PIK3C*A E454K were not detected and were absent in .bam files. With the filter of at least 3 reads no variants were detected in WT sample.

For indels all variants were detected for the 5%, 1% and 0.5% samples. For the 0.125% and 0.25% samples, 14/15 indels were detected when a filter of 3 consensus reads was applied, with the same relatively big indel (*EGFR* S752_I759delISPKANKEI) not detectable in both samples. The same variant was detected at much lower frequency in higher AF samples (detected at 1.92% in 5% dilution, 0.24% in 1% dilution and 0.15% in 0.5% dilution) so it is expected to be lost in lower dilutions. Again, with the filter of at least 3 reads no variants were detected in the WT sample. Therefore, we established the limit of detection for known variants to be 0.125% with ≥ 3 mutant reads.

3.4.2.2 Sensitivity, specificity, PPV, NPV and accuracy

Based on these results (Figure 3-11), we show that a high proportion of detected SNVs are true positives (Positive Predictive Value of 0.98) and the test has high analytical accuracy of 0.90 for SNV detection. SNVs can be correctly identified

with high technical sensitivity of 0.965 and acceptable specificity 0.815 at a VAF range of 100%-0.125% (Figure 3-11 C). Equally good accuracy (0.98) was demonstrated for indel detection (Figure 3-11 D) with high technical sensitivity of and specificity (both >0.97) at a VAF range of 100%-0.125%. Negative Predictive Value (NPV) was >0.9 for both SNVs and indels.



C	SNVs designed to be present	SNVs designed to be absent	
SNVs detected by ctPC panel	83 (TP)	10 (FP)	
SNVs not detected by ctPC panel	4 (FN)	44 (TN)	
	Values	95% CI Lower	95% CI Upper
Sensitivity	0.965	0.901	0.993
Specificity	0.820	0.686	0.908
Positive Predictive Value (TP/(TP+FP))	0.983	0.813	0.940
Negative Predictive Value (TN/(FN+TN))	0.917	0.805	0.967
Number of replicates	140	-	-

D	indels designed to be present	indels designed to be absent	
indels detected by ctPC panel	73 (TP)	0 (FP)	
indels not detected by ctPC panel	2 (FN)	57 (TN)	
	Values	95% CI Lower	95% CI Upper
Sensitivity	0.973	0.907	0.997
Specificity	0.978*	0.937	1.00
Positive Predictive Value (TP/(TP+FP))	1.00	0.950	1.00
Negative Predictive Value (TN/(FN+TN))	0.966	0.885	0.994
Number of replicates	132	-	-

Figure 3-11 (A) Correlation of allelic frequency of SNVs in Sera Care cfDNA control samples. VAF was observed using the ctPC panel versus expected as reported by the supplier (R 2 = 0.97) n = 10, SNVs = 125, mean depth = 2316 ×UMI. (B) Correlation of allelic frequency of indels in Sera Care cfDNA control samples. VAF was observed using the ctPC panel versus expected as reported by the supplier (R 2 = 0.49). n = 10, SNVs = 75. Mean depth = 2316 ×UMI. (C) Contingency table for SNV detection in the SeraCare control samples. (D) Contingency table for indel detection in the SeraCare control samples TP – true positive, FN – false negative, FP – false positive, TN – true negative. *specificity for indels is calculated using the rule of 3, as there were no false positives detected in any of the samples.

3.4.2.3 ctPC compared to other molecular profiling methods

Next, the performance of deep sequencing with UMIs was further validated using 16 clinical cfDNA samples previously characterised using alternative methods [159]. Six cfDNA samples from adult cancer patients with a total of 11 SNVs covered by our panel that previously were characterised by Avenio ctDNA targeted panel sequencing (Roche) were interrogated by ctPC workflow (Figure 3-12 A). All mutations detected by Avenio were detected using ctPC method (VAF range of 0.30% - 47.1%). Two mutations below the limit of detection of our assay (*PTCH1* c.3191C>T p.Thr1064Met at 0.11% VAF and *ALK* c.3522C>A p.Phe1174Leu at 0.02% VAF) were reported by Avenio but were not detected by ctPC sequencing. Interestingly, two *TP53* mutations were detected by ctPC panel that were not reported by Avenio. The *TP53* (c.855delG p.Glu287fs at 59.9% AF) highlights the capability of our method to detect frameshift mutations. The *TP53* (c.818G>A p.Arg273His at 48.97% AF) was found by the Avenio assay and software, but did not pass the default filters as it was not a part of the build-in databases. However, the review of recent literature classed the variant as pathogenic, again highlighting the flexibility of our method and ability to confidently call variants.

Additionally, we showed good concordance between ctPC panel and custom designed and validated ddPCR assays (Figure 3-12 B). Eight variants were detected by both methods, however two variants in KRAS were not detected by the ctPC panel (VAF of 0.18% and 0.3% in ddPCR). Additionally, two *TP53* mutations below the current limit of detection of ctPC assay were not detected as well (VAF of 0.06% and 0.07% by ddPCR).

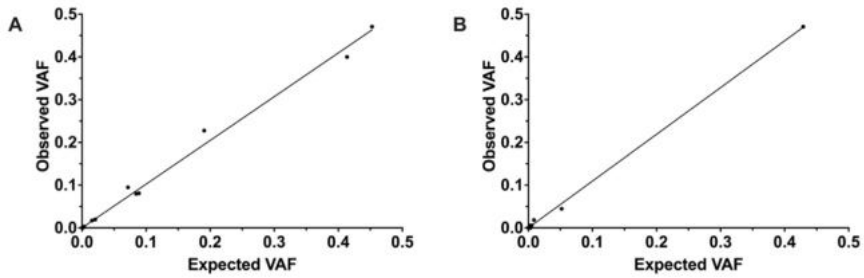


Figure 3-12 Correlation of allelic frequency of SNVs in clinical cfDNA samples. A VAF observed using the ctPC panel versus expected as reported by clinically validated Avenio sequencing panel (R 2 = 0.99) n = 7, SNVs = 11 Mean depth = 1284 ×UMI. (B) VAF observed using the ctPC panel versus expected as determined by individually designed ddPCR assays for each variant (R 2 = 0.99) n = 6, SNVs = 11 Mean depth = 1314 ×UMI.

3.5 Verification of ctPC_v2

The performance of the updated ctPC panel, ctPC_v2 was verified as per verification guidelines for clinical molecular genetic tests [157]. The performance metrics (on target sequencing and depth achieved) were expected to remain the same, as no new GC boosting strategies have been implemented, however the probe design has changed at the manufacturer's side – probes that have been empirically validated in their exome design were used to fill gaps in coverage.

3.5.1 Verification of ctPC_v2 panel performance

SeraCare commercial cfDNA samples with known VAFs of cancer mutations were used to verify accuracy in the same way as in original validation. The panel showed high accuracy with all SNVs and indels present in SeraCare 2% control sample detected using ctPC_v2 capture panel with comparable VAFs (Figure 3-13). Higher variability for indels is expected in capture panels and also is illustrated by AF spread reported by manufacturer of the control material.

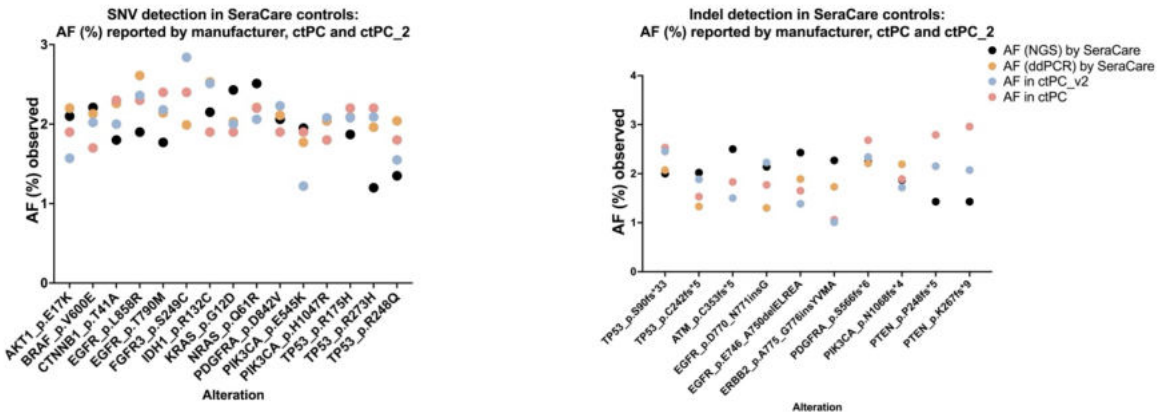


Figure 3-13 Allele Frequency (AF) comparison between SeraCare 2% control as determined by the two versions of ctPC panel for SNVs and indels. The AF supplied by manufacturer (by NGS and ddPCR) plotted next to AF determined using ctPC and ctPC_v2 panels.

3.5.2 Verification of sequencing performance and trueness

Five previously profiled cfDNA samples from paediatric patients (containing total of 11 SNVs) were used to measure trueness of the updated version of the panel. The NGS libraries previously run on ctPC panel have been re-pooled and captured using ctPC_v2 panel. The overall sequencing quality metrics were comparable

between the different panels: the percentage on targets reads was >80% for most samples, the percentage of PCR duplicates was within the recommended range (70-95%) for all samples and the mean depth was comparable between the two panels (Figure 3-14). All variants were detected using the new version of the capture panel with comparable VAF (R^2=0.992) in the VAF range from 0.001 to 0.73 (Figure 3-15). Therefore, we concluded that ctPC_v2 panel can accurately and reliably detect variants in ctDNA in plasma samples with comparable quality metrics as ctPC and it was used for most of the samples analysis described in this thesis.

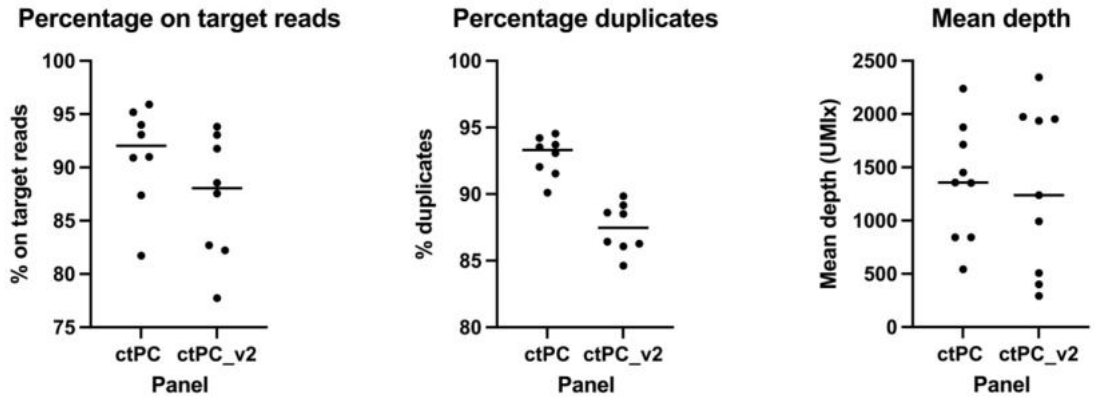


Figure 3-14 Quality metrics of sequencing in 8 cfDNA samples run on both ctPC and ctPC_v2 panel. (A) The percentage on target reads in each sample, run on different capture panels. (B) The percentage duplicates in each sample, run on different capture panels. (C) The mean depth achieved for each sample, run on different capture panels.

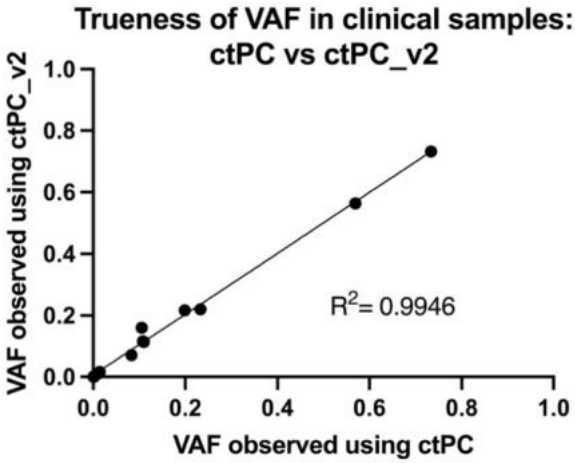


Figure 3-15 Trueness of VAF in clinical paediatric cfDNA samples – VAF detected by ctPC versus ctPC_v2 panel.

As a positive control we include one SeraCare control sample in each sequencing run. This allowed retrospective evaluation of the performance of the panel. Twelve sequencing runs had been audited and 100% sensitivity for variants with VAF above 0.5% was confirmed (Table 3-4). The variants most affected were indels, which also had lower than expected VAFs in higher VAF dilutions as discussed above in the validation. Out of the 10 missed variants, only one had 0 reads in the .bam file when inspected on IGV, the other variants were present with 1-2 reads.

Table 3-4 Variants called correctly in each SeraCare dilution sample. VAF expected, number of variants correctly called in this VAF range out of total number of variants expected, percentage of variants correctly called in that VAF range and cumulative percentage across VAF ranges (5%-0.5% and <0.5%).

VAF	Number of correct calls	Percentage	Combined percentage
5%	48/48	100%	100%
2%	23/23	100%	
1%	62/62	100%	
0.50%	38/38	100%	
0.25%	24/25	96%	89.36%
0.125%	60/69	87%	

3.6 Variant interpretation in cfDNA: guidelines and limits of detection

As described previously, the limit of detection for the assay was found to be VAF=0.125% with at least 3 unique reads in the .bam file inspected on IGV. That assured high sensitivity and specificity (>95%) of the method. The analysis was based on tissue-informed methodology, i. e. looking for the variants we know are expected to be in the sample, either based on control material design or in cases of clinical cfDNA samples, based on presence of the variant in the matching tissue sample. However, in some clinical situations tissue material is likely to be unavailable, therefore we need guidelines for tissue uninformed analysis as well. Even with use of UMIs and custom design bioinformatics pipeline, there remains some noise in the data, and it quickly became clear that for tissue uninformed analysis, the guidelines for variant reporting needed to be more stringent.

Extensive analysis of this problem had been performed by P. Carter, M. Smalley and A. Feber in the lab, looking at cfDNA samples from patients with gastrointestinal cancer. A different NGS capture panel (but based on the same methodology, produced by the same company, and using the same bioinformatics analysis pipeline, therefore generalisable to ctPC panel as well) had been used to profile 208 plasma samples in a tumour agnostic way and compared to Signatera™ Testing for Circulating Tumor DNA (ctDNA) Detection (Natera) and later unblinded to the time-matched tissue. Then, assuming that real variants are the variants present both in tissue and cfDNA in the patient, the features of correctly called real variants have been explored. The correctly called variants had quite broad distribution of VAFs - range 0.11 – 20.6% (median = 0.77%) with 64.2% of real variants having VAF ≥0.5%, confirming high sensitivity at this range of VAF. Importantly 157 variants called blind in cfDNA were not present in the tissue (VAF range 0.1 – 0.7%; median = 0.17%,) and were considered false positives in this analysis. The majority of the false positives were of VAF <0.3% and only 3 variants were of VAF ≥0.5%, highlighting reduced specificity at lower VAF ranges. It is clearly illustrated by looking at the percentage of variants correctly called in each VAF range (Figure 3-16). Based on that, the following variant calling guidelines have been defined (Table 3-5) and followed for the analysis of cfDNA samples described in this thesis. The only disadvantage of this analysis was the assumption that variant is not real in cfDNA if it was not present in tissue, which fails to account for tumour heterogeneity and possible tissue subsampling. However, since there is no gold standard method for cfDNA analysis at the moment, this is the best approximation we could have, and it should be explored further in the future. Well defined pathogenic variants and tumour-type specific variants were prioritised in comparison to variants of unknown significance in the final variant reporting.

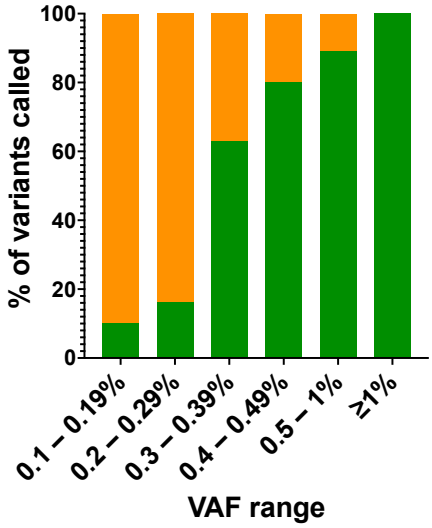


Figure 3-16 Variants called blindly in each VAF range. Green – variants correctly called blind, orange – variants incorrectly called in blind analysis (false positives) (Data and analysis by P. Carter).

Table 3-5 cfDNA NGS panel analysis guidelines for tumour informed and uninformed analysis.

cfDNA NGS panel analysis guidelines
Tumour informed
Limit of detection = 0.1%, 3 reads, for research purposes 0.05%, 2 reads Call any variant ≥0.05% called by the pipeline if the variant was also detected in tissue If variant is not called by the pipeline, check .bam in IGV, and call any variant which was also detected in tissue if ≥3 reads, ≥0.05%, if all variant reads : • Have a MAPQ of 60 • Mate is mapped • Variant base has QV of 45
Tumour uninformed (agnostic)
Limit of detection = 0.1%, 3 reads Depending on the VAF of the variant: **VAF = 0.1 – 0.3%:** ≥5 unique reads of variant strongly linked with cancer type, or ≥3 reads of very common hotspot needed (e.g. KRAS in CRC) **VAF = 0.3 – 0.49%:** ≥5 reads, preferably >9 reads; more confidence ≥0.4%; only call variants in genes strongly linked to cancer type, or pathogenic mutations **VAF ≥0.5%:** ≥6 reads, call all variants passing QC filters Overall confidence at different VAF ranges: 0.1 – 0.29% = low confidence 0.3 – 0.49% = medium confidence ≥5% = high confidence Special cases: **Pseudogenes:** >7 reads, >1% (guidance only) **Homopolymers >4 bases in length:** >10 reads, >1% (guidance only) **Strand bias:** ≥7 reads, ≥0.3%; don't automatically fail variants with some degree of strand bias

I have also explored the possibility to reduce the background of errors in ultra-deep sequencing by creating a background noise model, based on deepSNV algorithm [143]. This algorithm is based on comparison of a set of controls and the interrogated sample at each genomic locus. A beta-binomial model and a likelihood ratio test then is used to discriminate sequencing errors from low level real SNVs. This method was designed for detection of subclonal SNVs, but given low VAFs in cfDNA samples, it could be utilised in this setting as well. We created the background model using 13 blood cell pellet DNA samples form the paediatric cancer patients and used it as additional guidance for low level variant interpretation.

3.7 Copy number changes detection by lcWGS analysis

To allow comprehensive molecular profiling ctPC panel sequencing was combined with low coverage whole genome sequencing (lcWGS) from the same NGS library (therefore without the need for additional DNA). lcWGS provides two sets of information: genome wide copy number profile for each sample and estimation of ctDNA purity (the fraction of cfDNA that is coming from the tumour, i.e., ctDNA). The methods for lcWGS analysis are quite well established and ichorCNA package has been used for early analyses. As the projects progressed, we switched to ASCFLP pipeline, which allowed more flexibility and easier adjustment of analysis parameters. Both pipelines performed reasonably well, giving similar profiles and estimated tumour fraction in most samples (Figure 3-17), with the highest differences in low purity, low confidence samples.

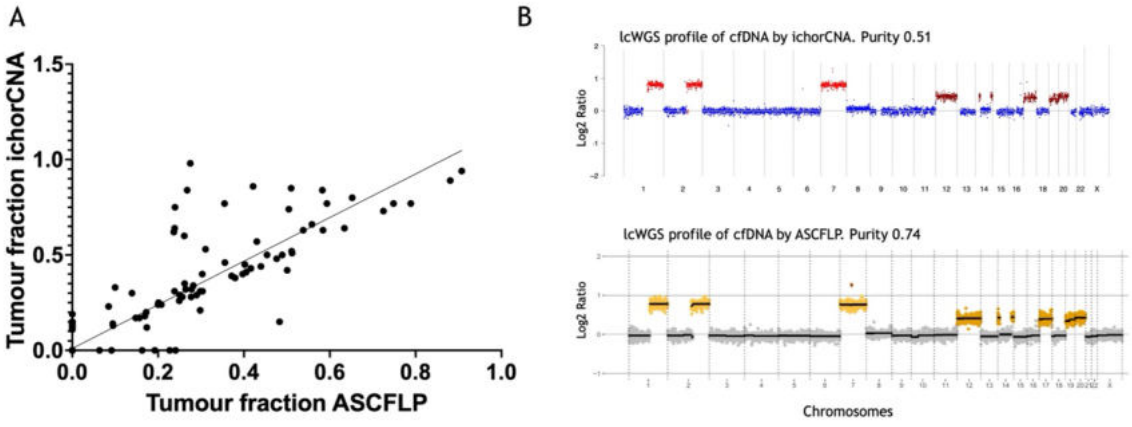


Figure 3-17 The two lcWGS analysis methods used throughout the studies produce similar purity estimates and CNV profiles. (A) Comparison of tumour fraction estimate by ichorCNA versus ASCFLP (B) An example of CNV profile of the same cfDNA sample analysed by ichorCNA and ASCFLP.

3.7.1 Limit of detection for lcWGS

To determine the limits of detection for lcWGS analysis in low input samples, I have performed a dilution series experiment. I created 100%, 30%, 10% and 5% purity samples, by mixing cancer cell line DNA (with known CNVs) with normal fibroblasts that have no CNVs and can act as wild type control in various fractions. Sequencing libraries were prepared for each dilution using low input protocol (from 5ng of starting DNA) and profiled using lcWGS analysis pipelines to compare cancer cell fraction (purity) estimate. The CNV profile was easily resolved at 30% purity, with lower confidence at 10% purity and not-detectable reliably at 5% purity, by both lcWGS analysis pipelines (Figure 3-18). The limit of detection of 10% is consistent with simulations performed in the lab, where CNV detection in lower purity samples is often less reliable.

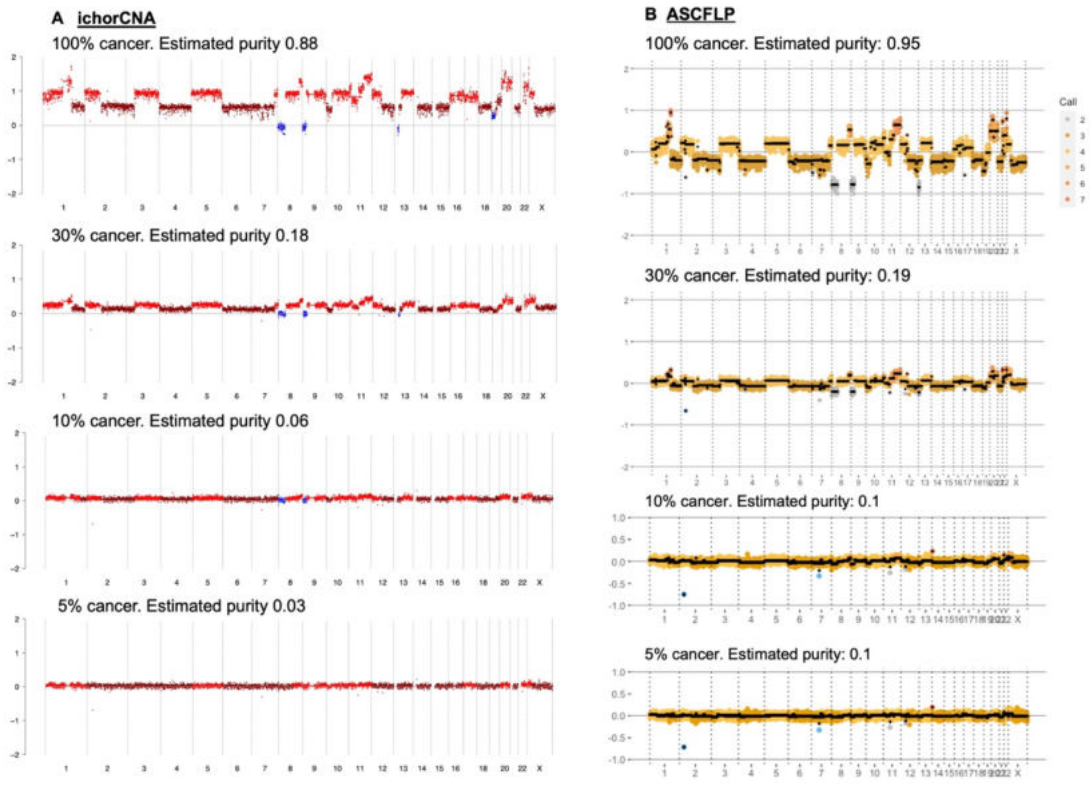


Figure 3-18 Limit of detection for lcWGS analysis. (A) ichorCNA analysis pipeline (B) ASCFLP analysis pipeline. A 100% cancer cell line DNA and it's serial dilutions in WT cell line DNA down to 5% cancer purity.

Several studies have shown that ctDNA is enriched in shorter fragments of cfDNA [77,79,160]. Therefore, we performed in silico size selection on alignment files of lcWGS to explore if this would lead to increase in the limit of detection. In our hands the removal of fragments above 150bp or 100bp did not improve the CNV resolution. In contrast, the loss of coverage due to removal of larger size fragments led to noisier profiles instead (Figure 3-19).

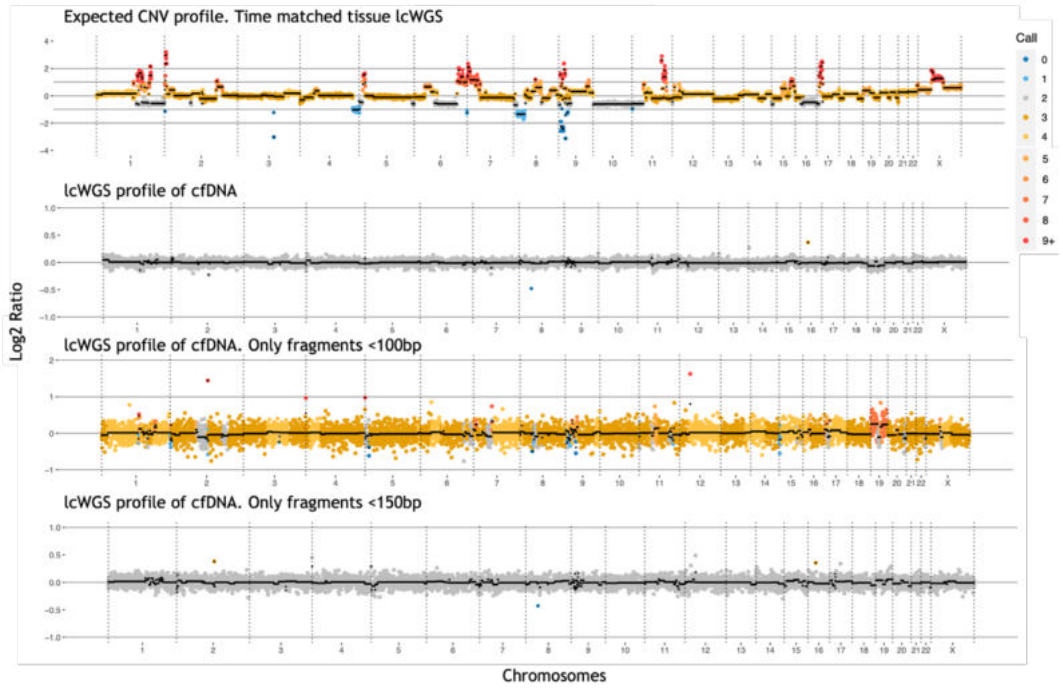


Figure 3-19 An example of in silico size selection effect on lcWGS profile. CNV plot (ASCFLP) of a cfDNA sample (SMP0042) before (2nd plot) and after in silico size selection to the range <100 bp (3rd plot) and <150bp (4th plot) as compared to the expected profile form matched tissue lcWGS profile (top plot).

3.8 Optimisation of the workflow for challenging samples

Even in clinical trials designed specifically for cfDNA collection, some of the samples that have reached us have been of suboptimal quality. During the duration of this project, cfDNA from 850 paediatric plasma samples have been extracted from patients on different clinical trials, with mean plasma volume of 4.79ml (min 0.1ml max 16ml), mean cfDNA yield of 102ng/ml of plasma (min 0.79ng/ml, max 7213ng/ml) resulting in mean total cfDNA yield of 417ng (min 2.4ng max 22880ng) (Figure **3-20**). These samples were collected from numerous clinical trials, from patients with different disease types, disease stages and timepoints (including some surveillance samples from patients on treatment or in remission, which were

expected to have low cfDNA yields). Some (but not all) extreme cfDNA yield values at the high end were due to high molecular weight contamination, which is discussed in more detail in the next section.

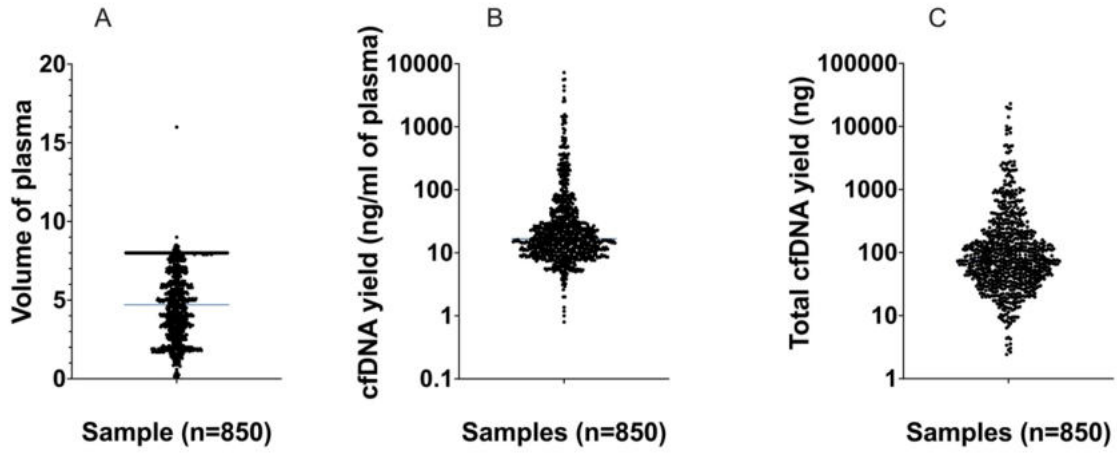


Figure 3-20 Volume of plasma collected and cfDNA yield of paediatric plasma samples discussed in this thesis. (A) Volume of plasma extracted per patient, mean indicated as blue line (mean 4.79, SD 2.27). The high number of samples at 8ml mark is due to the maximum input of the extraction protocol being 8ml (B) cfDNA yield per sample (ng/ml of plasma), median indicated as blue line (mean 102 ng/ml, SD 470, median 16.7 ng/ml) (C) Total cfDNA yield per sample (mean 417ng, SD 1721, median 76.6ng).

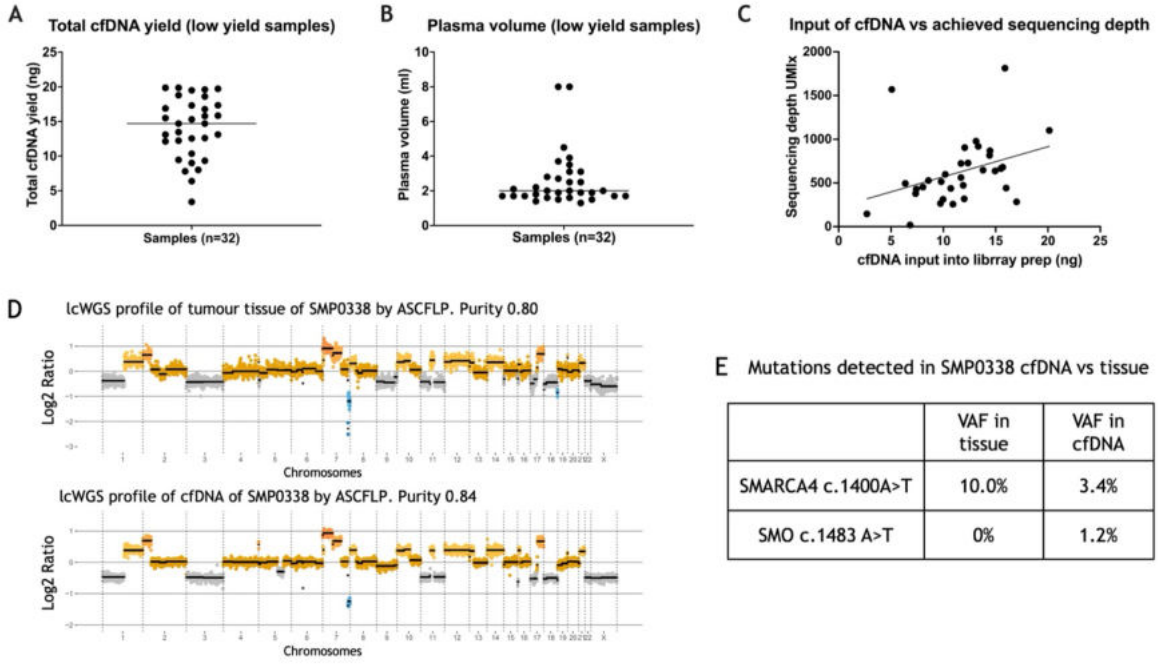
A number of sub-optimal samples are expected in real-life diagnostic labs and to implement this testing into the clinic we need to have ways to deal with poor quality samples and have pre-defined quality control metrics for excluding samples from the analysis. The strategies to deal with low cfDNA yield and HMW contaminated samples are explored in this section.

3.8.1 Optimisation for low input samples

One of the challenges for implementing liquid biopsy approaches in paediatric cancers is the low volume of blood collected from patients and/or low cfDNA yield in some cases. Over the span of this project, 32 low DNA yield (<20ng total DNA) samples have been molecularly profiled on ctPC panel. Low sample volumes (mean 2.6ml of plasma) have been the main cause of low yields (Figure 3-21 A,B). However, a few of the low yielding samples had been extracted from 8ml of plasma and these were patients at relapse (one patient with neuroblastoma and one with rhabdomyosarcoma), highlighting high variability of cfDNA levels between patients (which will be explored in depth in Chapter 4). To improve the method for low input samples, I have investigated the addition of carrier DNA into

the library preparation to increase the amount of DNA available for the capture, but that did not lead to any improvements. Therefore, the only modification for library preparation for low input (<20ng of DNA) samples was increased number of PCR cycles (8 PCR cycles instead of 6) at the first amplification step.

Despite the low DNA input into the sequencing library preparation, an average sequencing depth of 623 xUMI was achieved, with only two samples failing to have adequate sequencing depth (Figure 3-21 C). The depth of sequencing achieved increased with increasing DNA input amount to a certain degree, but most importantly, a few of these samples had SNVs detected, highlighting the possibility to perform analysis even with low input DNA. In some cases with high ctDNA purity, lcWGS profile could be successfully determined as well. For example, in SMP0338 – a cfDNA sample from a patient with neuroblastoma at relapse - where sequencing library was prepared form 6ng of cfDNA, the variant expected from tissue sequencing was detected and a distinctive lcWGS profile, matching tissue profile was obtained (Figure 3-21 D, E).



E Mutations detected in SMP0338 cfDNA vs tissue

	VAF in tissue	VAF in cfDNA
SMARCA4 c.1400A>T	10.0%	3.4%
SMO c.1483 A>T	0%	1.2%

Figure 3-21 Molecular profiling of low yield cfDNA samples. (A) The total yield of low yield (defined as <20ng total cfDNA) samples (B) Plasma volume that has been collected for these samples (C) The correlation of cfDNA input amount into the library preparation with the sequencing depth achieved by ctPC. (D) An example of lcWGS profile achieved from low input cfDNA sample (6ng into library prep). E. Summary of SNVs detected in the same sample, comparing VAF in tissue and cfDNA.

3.8.2 Optimisation for samples with high molecular weight DNA contamination

In our studies I have encountered a number of cfDNA samples with extremely high DNA yield. Upon inspection on fragment analyser, it was shown that a fraction of samples contained high molecular weight (HMW) DNA contamination. It is thought that the main reason for HMW DNA contamination is inappropriate sample processing – if blood collected into the EDTA tube is not separated into plasma by the double spin protocol in 4 hours (or in 7 days for blood in STRECK tubes), the chances of leucocyte lysis and subsequent release of HMW DNA into the sample increases. Preparing sequencing libraries from samples with HMW contamination would lead to low library yields (as the PCR steps in the library preparation are optimised for short length fragments) and highly reduced sensitivity, therefore I investigated ways to improve workflow and analysis of these samples with the goal of removing as much of HMW DNA as possible without losing much cfDNA. Two strategies were compared – adding higher amount of DNA into the library preparation versus performing size selection on the sample before the library preparation. Each have their drawbacks – a substantial amount of cfDNA is lost during size selection (and it is more time consuming), while DNA input >100ng into the library preparation risks exhausting the UMI adapters, which would lead to not enough UMI families and lower final sequencing depth.

To compare the two approaches, I tested a simple bead based size selection protocol to remove HMW DNA. An example is shown in Figure 3-22 A and B, where the fraction of cfDNA was increased from 23% to 63% after the size selection. I have tested several dilution protocols and compared Nonacus Target Pure™ NGS Clean-up Beads with AMPure XP beads, showing that dilution of the sample to 50μl volume and using Nonacus beads leads to acceptable level of cfDNA clean up. To establish this protocol, I have compared size selection on neat sample (normally 15-20μl) versus diluted sample (hypothesising that higher volumes would lead to better pipetting accuracy). Additionally, I compared two types of beads routinely used for DNA clean up and size selection. The dilution of the sample increased the recovery of cfDNA (with no significant difference between 50μl and 100μl dilution) and there was no significant difference between the different types of beads (Figure 3-22 C), therefore the most economical option was chosen.

In the testing stage samples with varying levels of HMW contamination were tested (10-23% cfDNA in the original sample), leading to acceptable levels of cfDNA for library preparation in all cases (55-87% cfDNA after the size selection) with the most extreme example – 16 % cfDNA to 80% cfDNA after clean-up (increase of 64%). However, the size selection protocol involves several capture steps and the loss of DNA is a significant concern. The amount of DNA lost depends on the level of contamination, so in these samples ~90% of DNA amount was lost. However, some of the DNA in the cfDNA size range will be lost as well. In these samples, 53% of cfDNA was lost as estimated from the fragment proportion in the cfDNA range and DNA concentration.

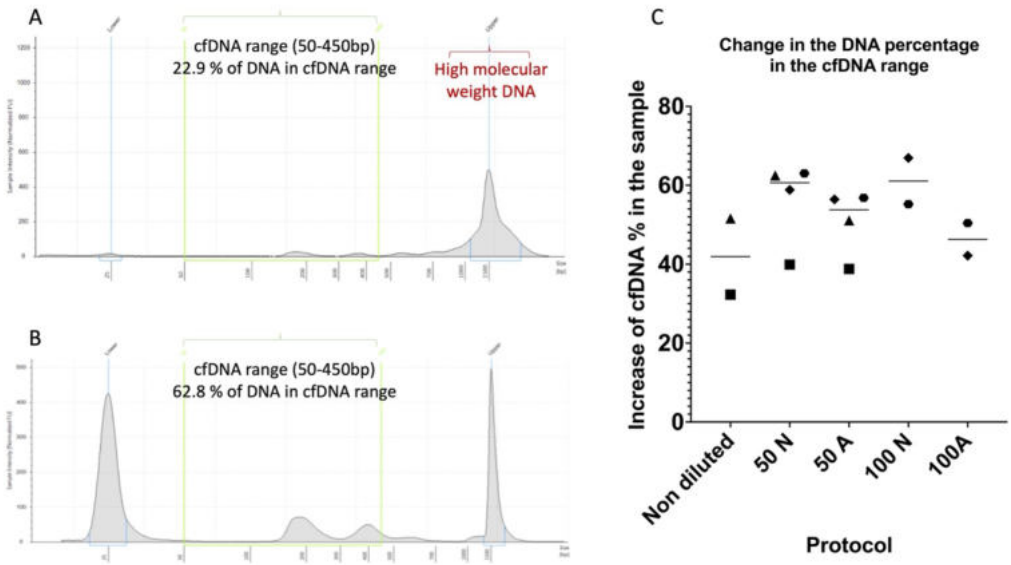


Figure 3-22 Size selection protocol to remove HMW contamination form cfDNA samples. (A) An example of electropherogram of size distribution of cfDNA sample with high HMW contamination (only 23% of the DNA in cfDNA range) by Tapestation with the High Sensitivity DNA ScreenTape assay. The y-axis showing the signal intensity (FU) and the x-axis showing the DNA fragment size (base pairs). cfDNA range shown in green, assay markers at (50bp and 1500bp marked in blue). (B) The same cfDNA sample after size selection using beads (resulting in 63% DNA in the cfDNA range). (C) The change of the DNA percentage in cfDNA range after various size selection protocols. Non diluted sample – 15μl of sample used for size seletion, 50 N – the 15μl of sample diluted to 50μl and then size selected using Nonacus beads, 100 N – the 15μl of sample diluted to 100μl and then size selected using Nonacus beads, 50 A – the 15μl of sample diluted to 50μl and then size selected using Ampure beads, 100 N – the 15μl of sample diluted to 100μl and then size selected using Nonacus beads. Each unique sample indicated by unique symbol on the plot.

A good opportunity to compare the size selection approach versus increased starting DNA input amount into the library preparation came around when a sample with very high HMW contamination has been received from one of the patients I have been monitoring though serial blood sampling. LB001 – a patient with neuroblastoma with a known pathogenic *ALK* variant – has been monitored

for several months while on treatment using ctPC panel (discussed in detail in Chapter 6). One of the serial samples had a DNA yield of 1000ng and high HMW contamination (15.3% of DNA in cfDNA range as determined by cfDNA Tape on Agilent Tapestation) (Figure 3-22 A). Given such high yield and prior knowledge of the variant tracked, I have prepared the sequencing library from the same sample in 3 different ways to compare the effect it would have on molecular profiling results:

1. Standard protocol: 50ng input of heavily contaminated sample

2. 330ng input (to adjust for the level of high molecular weight contamination)

3. Bead based size selection prior to library preparation followed by 50ng input of size selected sample into the library preparation

All approaches led to successful sequencing library preparation, with 50ng input having the weakest library (Figure 3- 23 B), likely due to the size selection inherent to the library preparation protocol which removes the HMW contamination and leads to reduced representation of cfDNA in the total DNA input. Therefore, the adjusted input protocol was tested alongside it. Most importantly, the VAF of the pathogenic ALK variant was the same in all three samples (Figure 3- 23 E), despite differences in overall sequencing depth and depth at the target location and therefore the variant would have been reported in all 3 instances (Figure 3- 23 C,D). Both increased input to the library preparation and size selection produced similar results, however, in the size selected sample more low level variants have been detected by the bioinformatics pipeline, most of them noise. That is most likely due to higher overall sequencing depth. However, this patient had several low VAF variants detected in previous cfDNA samples that had been stable over time and likely coming from leucocytes and not the tumour, and these variants were only detectable in the size selected sample and not the other two, indicating higher sensitivity. A few more cfDNA samples with high HMW contamination have been prepared both with high input and size selection protocol and the results for both have been very similar. Therefore, it seems that both increased input amount and size selection are acceptable options, with size selection likely giving better sensitivity.

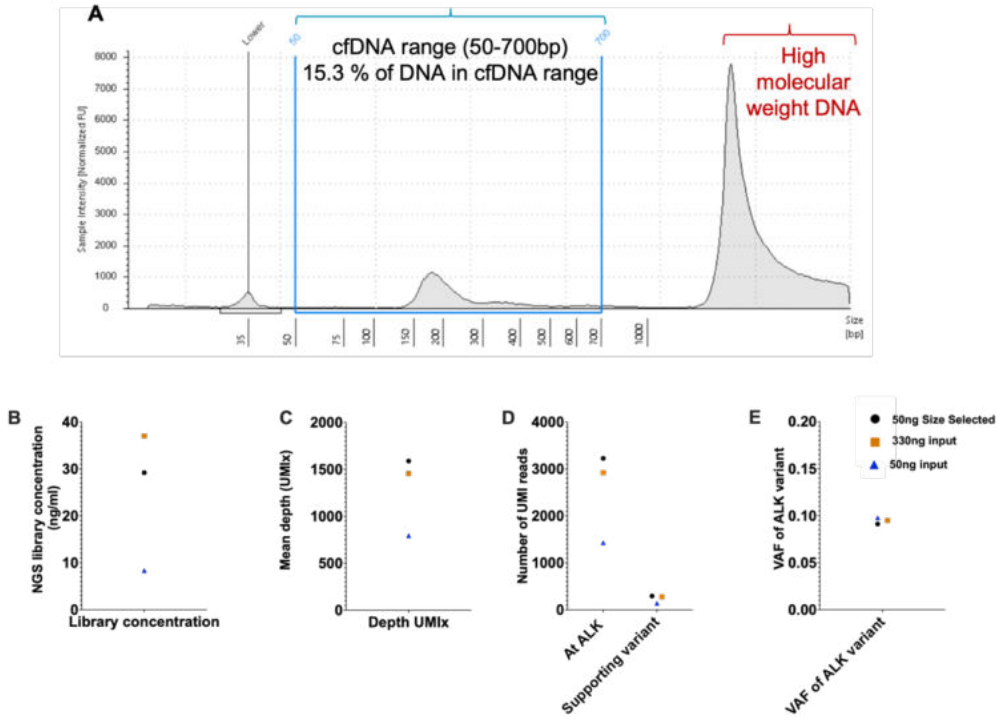


Figure 3-23 Comparison of molecular profiling results for a clinical sample with high HMW contamination. (A) An electropherogram of size distribution of cfDNA sample by Tapestation with the cfDNA ScreenTape assay. The y-axis showing the signal intensity (FU) and the x-axis showing the DNA fragment size (base pairs). B-E Varying molecular profiling metrics achieved in sample prepared in 3 different ways, each method colour and symbol coded. (B) NGS library preparation achieved (C) Mean depth (UMIx) achived in ctPC panel sequencing (D) The number of unique reads at the ALK location of the varinat and the number of reads supporting the pathogenic variant (D) VAF of the pathogenic ALK variant.

3.9 Proposed workflow of cfDNA sample processing

To implement this methodology into the diagnostic laboratories, clear guidelines for triaging and streamlining the more difficult samples are necessary. The proposed workflow (Figure 3-24) addresses both the cfDNA quantity and quality and is generalised enough to allow high throughput sample processing. For low input samples, as low as 5ng can be successfully used on this method, with the caveat of lower sensitivity and higher false negative rates. The cfDNA yield depends on multiple clinical factors, but as a rule of thumb, after extraction, all the samples with cfDNA yield higher than 30ng/ml of plasma should be run on fragment size analyser to assess the level of HMW DNA contamination. If the sample is estimated to have >50% of DNA in the cfDNA range, 50ng can be used for the library preparation using standard protocol. Samples with higher levels of HMW contamination can be either size selected to enrich cfDNA fraction or increased DNA input can be used for the library preparation. From the provisional experiments performed and theoretical calculations based on them the following

workflow is proposed - given the average purity of 25% cfDNA in samples with HMW contamination and the loss of 50% of cfDNA during the size selection, the size selection protocol should be performed on DNA samples with >500ng total DNA. In cases with lower total DNA amount, the loss of DNA during the size selection would lead to lower input of cfDNA into the prep, therefore 250ng should be used as input for these samples, which would equate to adding ~60ng of cfDNA (given average purity of 25%). In extreme cases, if time allows further adjustments could be made. For example, if total DNA <10ng - increase the number of PCR cycles to 9 or 10, or in extremely high HMW contamination (such as cfDNA fraction <10%) perform size selection in samples that have >250ng total DNA.

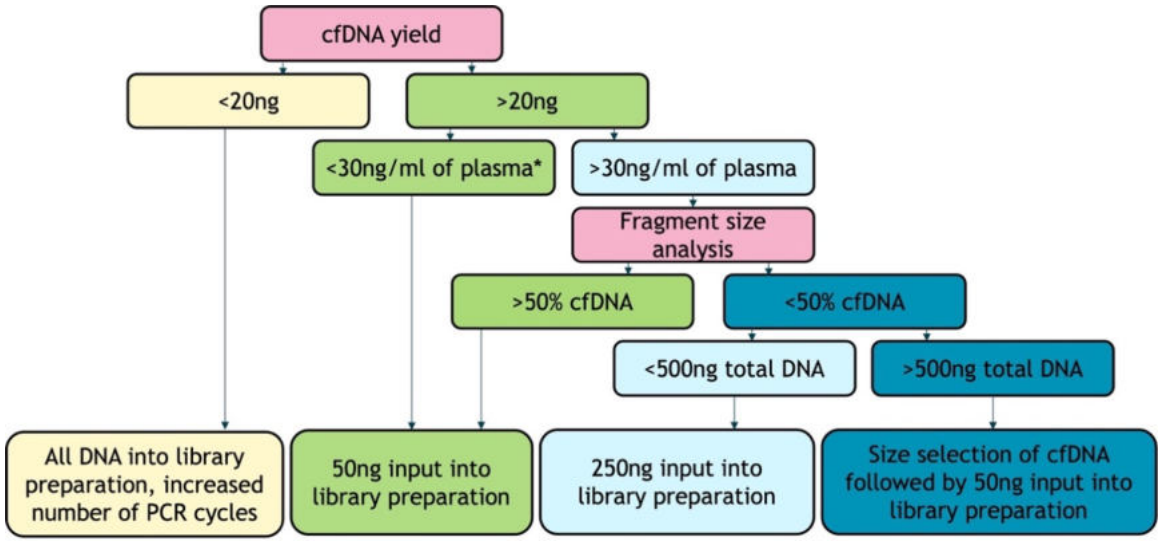


Figure 3-24 The proposed workflow for cfDNA sample NGS library preparation. The QC steps of cfDNA quantification (by Qubit) and fragment size analysis (by Agilent Tapestation) highlighted in pink. In samples with >20ng of cfDNA, fragment size analysis should be performed in samples where high molecular weight DNA contamination is suspected. The 30ng/ml of plasma proposed as a threshold, but it is dependent on clinical characteristics of the sample (such as disease type and clinical timepoint) and is due to scientists discretion.

3.10 Detection of focal copy number changes by ctPC panel in cfDNA

Focal copy number changes are important in a number of paediatric cancers, therefore probes specifically designed to detected focal amplifications and deletions in selected genes have been included into the ctPC panel design. The initial validation of focal copy number variants (CNVs) was performed using artificial cfDNA controls, as described above for SNV and indel detection. The validation was of limited scope for CNV detection due to the characteristics of commercial controls available at the time – only focal amplifications of three

genes were present in these samples with no focal gene deletions available. The ability to detect *ERBB2, MET* and *MYC* amplification in the same control sample, sequenced in five independent sequencing runs was evaluated. The amplifications in all three genes were consistently detected in all samples tested (amplification defined as log2 ratio >1 for >80% of the probes of the gene) showing good capability for CNV detection (Figure 3-25). The results were highly reproducible, with in standard deviation of 0.07-0.29 between repeats. Further validation was performed on clinical samples, where information about CNVs detected in tissue was available and therefore could be used as the ground truth, and it is outlined in detail in further results chapters.

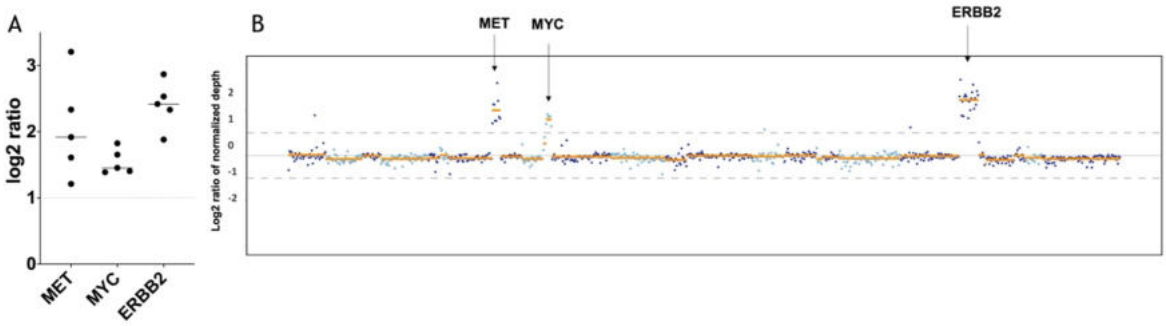


Figure 3-25 Copy number detection using ctPC panel in cfDNA. A. The log2 ratio of three genes known to be amplified in the control sample. Log2>1 (indicated by grey dashed line) indicates successfully detected amplification; each dot represents independently sequenced sample. B. Representative copy number profile produced by ctPC analysis pipeline, highlighting *MET, MYC* and *ERBB2* amplifications.

3.11 Detection of fusions in cfDNA

Some paediatrics cancers are driven by fusions, but detection of those is more challenging in cfDNA by NGS capture panels, that are not individual for each patient. Patients with Ewing sarcoma tend to have recurrent break points, in *EWRS1* in introns 8-12 (Figure 3-26) therefore these regions were covered by ctPC panel to evaluate the possibility of fusion detection in cfDNA. Additionally, *ALK* intron 19 was included to aid evaluation in patients that were anticipated in future studies and where *ALK* rearrangements would be diagnostically relevant.

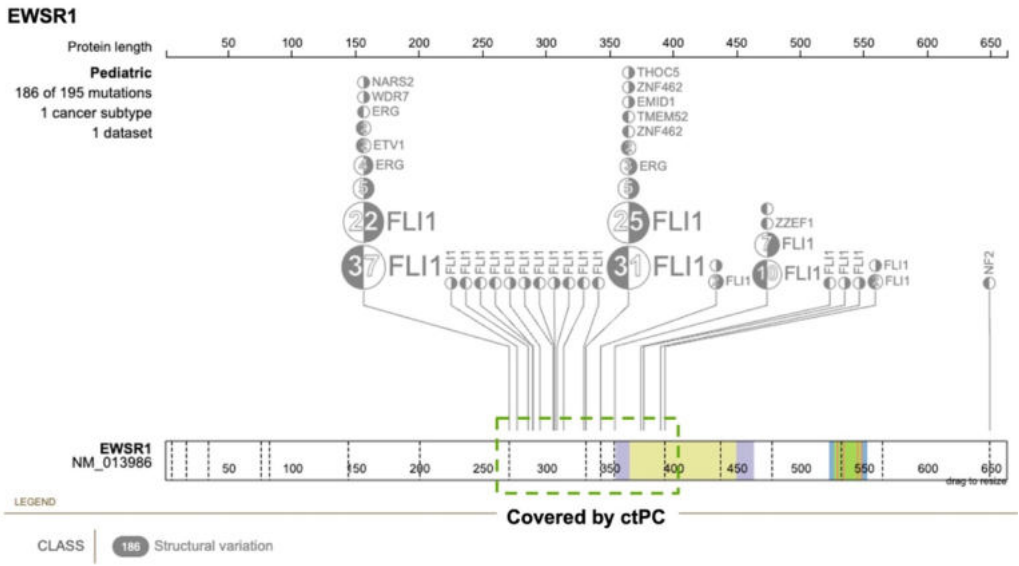


Figure 3-26 The most common fusion break points in paediatric cancer patients in EWSR1 gene. The schematic of EWSR1 gene and fusion partners indicated introns noted in dashed lines. Numbers in each bubble indicate the number of reported cases for the fusion in that genomic location with that particular fusion partner. The region covered by ctPC panel indicated in green. Figure from St. Jude Children's Research Hospital — PeCan Data Portal, accessed Sept. 2022 [161].

Fusion detection was beyond the scope of the validation of ctPC panel, but I investigated the possibility of detecting fusions in cfDNA in the Stratified Medicine Paediatric Programme where RNA from relapse tissue was available for a subset of patients and fusions were profiled using TruSight RNA Pan-Cancer Panel (Illumina) which identifies an exact break point for each fusion gene. Using that information, I have investigated the cfDNA sequencing results for patients with fusion break points covered by ctPC panel by inspecting the alignment files manually at the known breakpoint location using IGV software. Twenty nine patients in Stratified Medicine Paediatric Programme had fusions detected in RNA involving *ALK* and *EWRS1*, which are covered by ctPC panel. cfDNA samples of ten of these patients had ctDNA purity >10% and the initial proof-of-principle analysis was focused on these samples (Table 3-6). Fusion was detectable in four out of ten cfDNA samples. Our bioinformatics pipeline was unable to detect these structural variants, but manual inspection showed the presence of reads, with an example shown in Figure 3-27. Therefore, detection of fusions in cfDNA is possible in some high purity samples, however improvement of the bioinformatic pipeline followed by full validation and evaluation of sensitivity, specificity and limit of detection is needed in the future.

Table 3-6 cfDNA samples in Stratified Medicine Paediatric Programme where fusions covered by ctPC panel were identified in RNA. Only samples with detectable ctDNA by lcWGS shown. The disease for each patient, fusion and fusion breakpoint, ctDNA purity from lcWGS and the ability to detect the fusion in cfDNA shown.

Trial ID	Disease type	Fusion detected in RNA	Fusion breakpoint at EWSR1 or ALK	ctDNA purity by lcWGS	Fusion detected in cfDNA
SMP0197	Ewing Sarcoma	*FLI1:EWSR1*	chr22:29,688,156	0.880	yes
SMP0090	Ewing Sarcoma	*FLI1:EWSR1*	chr22:29,684,593	0.558	yes
SMP0237	Ewing Sarcoma	*FLI1:EWSR1*	chr22:29,683,123	0.421	no
SMP0128	PNET	*FLI1:EWSR1*	chr22:29,683,123	0.261	no
SMP0117	Ewing Sarcoma	*FLI1:EWSR1*	chr22:29,683,123	0.249	yes
SMP0232	Ewing Sarcoma	*FLI1:EWSR1*	chr22:29,683,123	0.238	no
SMP0048	Ewing Sarcoma	*FLI1:EWSR1*	chr22:29,688,155	0.236	yes
SMP0163	Ewing Sarcoma	*EWSR1:FLI1*	chr22:29,683,123	0.225	no
SMP0094	Ewing Sarcoma	*FLI1:EWSR1*	chr22:29,683,123	0.084	no

Figure 3-27 Fusions can be detected in cfDNA. An example of fusion detection in patient SMP0048. An IGV view at the expected EWSR1 break point. Mismatched bases shown in colour, bases matching reference shown in gey. BLAT of mismatched reads and their end showing EWSR1:FLI1 reads displayed below the alignments. BLAT feature marks directionality relative to the original search sequence as displayed by IGV.

3.12 Discussion

The implementation of liquid biopsies into clinical management of patients with paediatric cancer requires custom approaches, due to different genomic landscapes between adult and paediatric cancers. At research setting, several cancer type specific cfDNA molecular profiling approaches have been developed for neuroblastoma [132,150,151,162], glioma [96,152,153], paediatric sarcomas [89,90,94,160]. However, due to relatively low patient numbers (~1838 new children's cancer cases per year in UK, according to Cancer Research UK[1]), a pan-cancer approach would help to bring the liquid biopsy based molecular profiling closer to clinical application. The first step towards the implementation of NGS assays into the clinical practice requires thorough validation in clinically accredited laboratories, showing that the test is fit for the intended purpose and technically robust.

In this chapter I outlined the development and validation of pan-cancer clinically relevant NGS based molecular profiling approach for cfDNA extracted from blood from patients with paediatric cancers. The ctPC panel, covering most clinically relevant genes in most common paediatric solid tumours, has been validated to be repeatable and reproducible with high sensitivity and specificity for SNV and indel detection at low variant allele fractions that are expected in cfDNA analysis. The validation is ongoing for the CNV and SV detection, but the preliminary results are promising. Importantly, the targets on the panel are easily amendable with minimal verification required after the changes, making this a flexible and sustainable approach. The current panel is designed to be used as tool to identify therapeutic targets, help diagnosis and prognostication, and identify resistance mechanisms to treatment. Therefore, only genes implicated in oncobiology of the most common solid paediatric tumours are covered. The resulting relatively small size of the panel allows cost-effective molecular profiling and manageable analysis time. I have outlined the guidelines for SNV calling based on the validation and other experiments performed in the lab, but quite high background noise in the cfDNA analyses is still a problem, which leads to the need of extensive manual variant curation. To implement this test to the clinic, improved background suppression methods would be needed. Additionally, the current limit of detection of 0.125% VAF could be improved if better error correction models were implemented. We have explored the application of DeepSNV [143], but the results of it could be further improved with more appropriate control materials to construct

the background error model. Collection of the blood of healthy age-matched controls for creation of the background model would be ideal. Another approach would be to use machine learning approaches to better understand the nature of artefacts in UMI sequencing data and the work is continuing in the lab to develop the models, based on high numbers of samples analysed in all the different paediatric trials.

The CNV profiling using lcWGS is more universal molecular profiling method and it is not novel in paediatric liquid biopsy field. The lcWGS analysis has been shown to have good concordance between tissue and cfDNA in various paediatric cancer types (including osteosarcoma, Ewing sarcoma, rhabdomyosarcoma, Wilms tumour, brain tumours and neuroblastoma) [90,100,106]. Here I have shown that lcWGS profiling can be performed together with high depth panel sequencing, without the requirement of additional DNA. The preliminary results show that detection of CNVs is possible, both from ctPC panel sequencing and lcWGS, especially in high ctDNA purity samples. However, for clinical implementation, a more extensive validation is needed, looking at reproducibility, repeatability, sensitivity, and specificity as well as limits of detection for different sample types [157]. The full validation of CNV calling was out of scope for this thesis.

Gene fusions are also a common driver in paediatric cancers[14,163–165]. Due to the fragmented nature of cfDNA, fusions are more difficult to detect in this bioanalyte. However, the ctPC panel was designed to cover introns of *ALK* and *EWSR1* that have break point hot spots, as a proof of concept for fusion detection in cfDNA by NGS panel sequencing. I have shown preliminary data for detection of fusions in cfDNA, but more extensive work, following the clinical implementation guidelines, would be needed to evaluate the possibility of implementation of this part of the assay into the clinic.

Throughout the studies described in later chapters, I have encountered a number of low quality cfDNA samples, mainly due to low total cfDNA yields or significant HMW contamination from normal cells. To apply our workflow to samples with HMW contamination, bead-based size selection protocol has been tested and shown good results. The method was not extensively validated and different approaches have been tested on clinical cfDNA samples, because at the time it was thought that HMW contamination is a rare occurrence caused by inadequate

sample processing. We do not have the data to confirm this (as data on general health of the patient at the time of blood collection was not collected as part of any of the trials), but from unofficial correspondence we know that patients who have infections at the time of blood collection can show substantial HMW DNA contamination, presumably coming from the rapidly dying blood cells. Given this knowledge, a more controlled experiment could be designed to assess the efficiency of the different approaches of processing samples with different levels of HMW contamination. Artificial cfDNA control spiked into large fragment size wild type cell line DNA at varying fractions could be used to assess the loss of DNA at different size fractions and the effect it has on the specific variant detection. Despite this limitation, I have described a method to process these samples and proposed guidelines to streamline the decisions in the lab based on the quality of cfDNA sample to achieve optimal results (Figure 3-24).

To sum up, I have designed a pan-cancer paediatric specific NGS capture panel-ctPC – optimised for cfDNA analysis and an accompanying workflow with lcWGS to molecularly profile paediatric patients with solid tumours. The panel is validated to clinical standard for SNV detection and shows promise for CNV and specific fusion detection. The following chapters will describe the application of this methodology in several paediatric clinical trials, looking at concordance of molecular profiling in cfDNA versus tissue and the utility of monitoring the patients while on treatment to assess response and track tumour evolution.

Chapter 4 Concordance of tissue and liquid biopsies in paediatric patients with solid tumours

Parts of this chapter have been published as part of [149]

Parts of this chapter are from a manuscript in progress

4.1 Introduction

The isolation and molecular profiling of cfDNA has shown great potential for identification of actionable biomarkers in various cancers. Given the inherent difficulties of acquiring tissue biopsies in children with cancer, this less invasive method offers not only the possibility of providing the molecular profile of the tumour when tissue biopsy is not possible, but also the potential to assess tumour heterogeneity that might be missed by single tissue biopsy [115,128]. ctDNA is shed by tumour cells into the blood stream and its potential to be used as a molecular profiling tool, due to its concordance with tissue profiling is well described in adult cancer [78,86,107]. However, the shedding of ctDNA into the blood depends on multiple factors, such as tumour stage [166–168], tumour volume [101,129,168,169] and metastatic status [98,170,171]. These factors vary between individual patients [85]. Therefore, the clinical utility of cfDNA molecular profiling needs to be explored for each cancer type before recommendations for clinical implementation can be issued.

The evidence for the clinical value of cfDNA analysis in patients with different types of paediatric tumours has been recently reviewed [83,108,109], highlighting the highest potential in neuroblastoma, osteosarcoma, rhabdomyosarcoma, malignant renal tumours and Ewing sarcoma. However so far, studies exploring the use of liquid biopsies in paediatric cancers are restricted to small patient cohorts and varied methodologies (from custom designed droplet digital PCR (ddPCR) assays [91,92,96], custom next generation sequencing (NGS) panels [90,162], whole exome sequencing (WES) [99], low coverage whole genome sequencing (WGS)[90,100] to methylation assays [172,173]). The use of liquid biopsies in paediatric oncology is emerging, however to our knowledge there are still no clinically implemented assays.

At the time of the conception of this study, not much information was known about the utility of cfDNA profiling for paediatric patients with solid tumours. Most research on liquid biopsy applications in solid paediatric tumours was focused on neuroblastoma due to the relatively high yields of ctDNA in this paediatric cancer type [90]. Early proof-of-concept study had shown the detection of *MYCN* amplification in cfDNA extracted from plasma in neuroblastoma patients at diagnosis [103]. Further studies had shown increased levels of ctDNA in neuroblastoma patients with active disease [101,102]. Shallow WGS of cfDNA had shown good concordance with tissue profiling in neuroblastoma patients [98,100]. WES of the primary tumour and cfDNA at diagnosis showed good agreement for SNV and CNV detection (with 41% and 93% of all detected alterations common to the primary neuroblastoma and cfDNA, respectively) and the potential to detect relapse specific alterations and hence study tumour evolution using cfDNA profiling [99]. More recently, cfDNA optimised neuroblastoma specific NGS capture panel has been created and showed good concordance of tissue and cfDNA profiling for stage 4 neuroblastoma patients at diagnosis [162].

For other paediatric cancer types, relatively small and cancer type specific studies were showing promise. The association of increased ctDNA levels and poor prognosis/survival has been shown in Ewing sarcoma and osteosarcoma [89] and Hodgkin's lymphoma [93]. In patients with paediatric diffuse glioma detection of major driver mutations using ddPCR in cfDNA from CSF and plasma correlated with tumour volume measurements by MRI [96]. Targeted profiling showed good concordance between tissue and cfDNA profiles in osteosarcoma [94], malignant kidney tumours [95], neuroblastoma [99] and Ewing sarcoma [91]. Copy number profiling of various solid paediatric malignancies was also showing high concordance with tissue profiling [90,98-100]. In our lab we have also shown a high detection rate of somatic variants in paediatric plasma samples, using a generic adult cancer cfDNA panel [25]. In this project we wanted to test a pan-cancer molecular profiling approach in a much broader context, aiming to collect adequate data allowing clinical implementation.

In this chapter I will describe our work comparing NGS panel sequencing results of cfDNA compared with tissue in a wide range of solid paediatric tumours, using a pan-cancer NGS capture panel. We hypothesise that molecular profiling of cfDNA

has the potential to provide the same information as tissue based profiling, and hence replace or supplement tissue biopsy profiling. To that end the next chapter will describe the concordance between tissue and cfDNA profiling and explore if it is universal for all paediatric solid tumours studied. Throughout the span of these projects, various studies have come out supporting cfDNA profiling for paediatric cancer patients and the results of these in relation to this project will be described in detail in the discussion.

4.2 NGS tumour profiling study: cfDNA profiling in paediatric patients with solid tumours

To test the performance of our proposed cfDNA profiling workflow (ctPC capture panel combined with lcWGS) and its agreement with tissue profiling, I have analysed a set of cfDNA and tissue samples from retrospective molecular profiling study of paediatric patients with solid tumours (CCR-4294). Plasma samples have been collected for patients with neuroblastoma, various sarcomas, Wilms tumour, CNS and other tumours at various disease stages. The tissue profiling for these patients has been previously performed using our paediatric specific tissue NGS capture panel [25,140] and based on these results I have selected patients with known SNVs and CNVs to address the question if we can detect the same aberrations in cfDNA.

4.2.1 ctDNA is detectable in most paediatric solid tumour patients with active disease

The first cohort from the NGS Tumour profiling study (CCR-4294) consisted of 39 patients (67 cfDNA samples) representing a range of paediatric solid tumours. The cfDNA yield was highly variable between the patients (5.7-1452ng/ml of plasma) (Figure 4-1 A). The possibility of HMW DNA contamination was excluded by fragment size analysis and samples with high HMW fraction (>50% of DNA outside of cfDNA range) were excluded from these comparison plots. Blood was collected at various time points throughout the patient's cancer treatment and a higher ctDNA fraction in the total cfDNA (as calculated from lcWGS) was observed in patients with active or refractory disease (Figure 4-1 B).

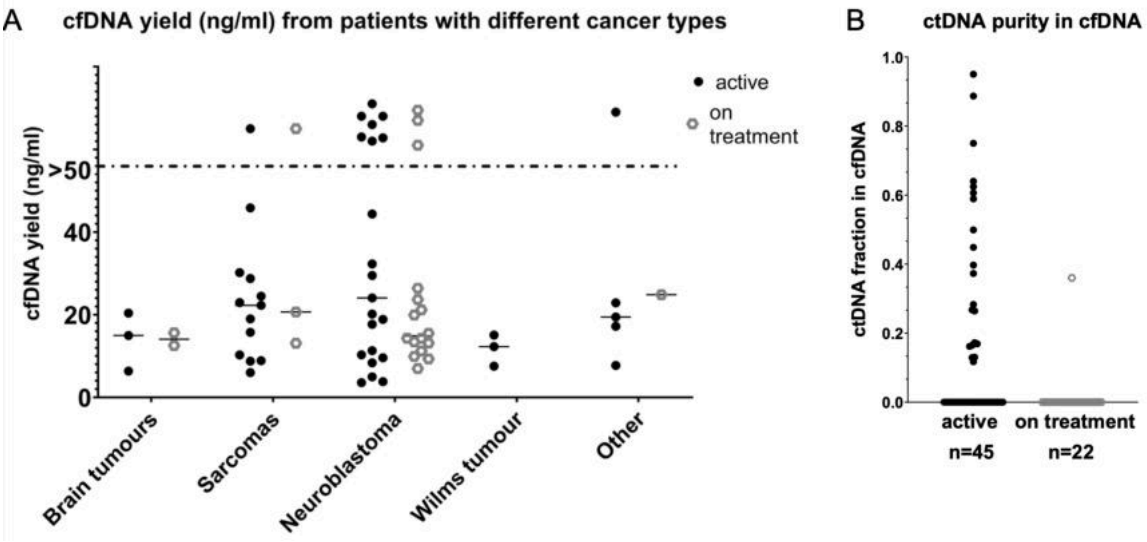


Figure 4-1 cfDNA yield from plasma from patients with various paediatric cancer types. A. cfDNA yield (ng/ml) in different cancer types. Brain tumours include germinoma, ependymoma and glioma. Sarcomas include Ewing sarcoma and rhabdomyosarcoma. All disease types with less than three samples available are grouped into "other" (adrenocortical carcinoma, metastatic carcinoma, malignant neoplasm of kidney, nasopharyngeal sarcoma). For each group blood samples that were taken in patients with active disease are shown separately from samples taken while the patient was on treatment. Samples above the dotted line containing >50ng of cfDNA/ml of plasma. B ctDNA purity estimate in the total cfDNA in all of the samples in this cohort. Active was classed as bloods taken at diagnosis, relapse or refractory to treatment, on treatment – blood sample taken when patient is responding to the treatment or in remission. ctDNA purity in cfDNA estimated from lcWGS.

Sequencing libraries were generated using 5-50 ng of cfDNA and sequenced to a mean depth of 1069xUMI (unique read families consolidated based on UMI per target). Interestingly, unique coverage correlated moderately with cfDNA input (R^2 = 0.49 for negative samples and R^2 = 0.33 for positive samples), but the wide distribution of coverage at optimal DNA input (50 ng) suggests that DNA quality has an impact on these results as well (Figure 4-2). Importantly, cfDNA levels do not seem to be dependent on disease status and even in some very low yield samples (as low as 5-10ng total cfDNA) positive molecular profiling results were achieved, as discussed in detail in the next section.

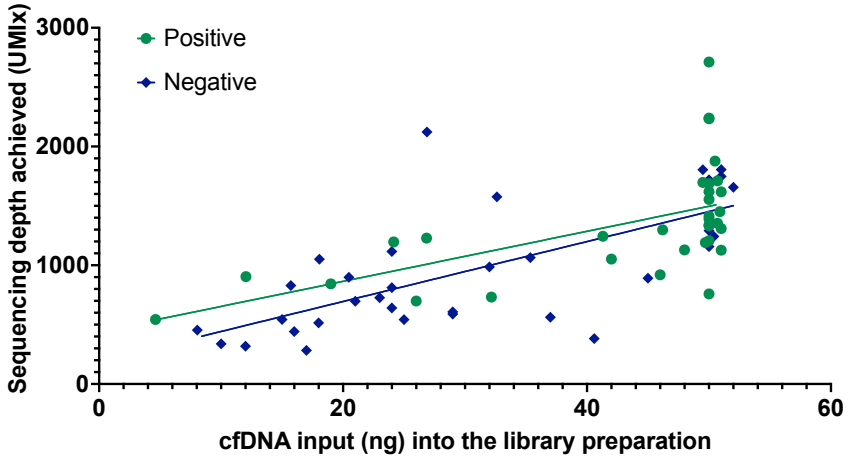


Figure 4-2 The unique sequencing depth (UMIx) achieved depending on the input amount of cfDNA into the library preparation. Coloured by alteration detection -green dots where alterations have been detected in ctDNA by sequencing, blue rhombus – no alterations were detected in ctDNA. Linear regression for both positive and negative samples shown in lines of corresponding colour.

4.2.2 Variants detected in tissue sequencing are found in cfDNA in most patients with active extracranial disease

The cohort of the NGS Tumour profiling study (CCR-4294) included 37 patients with both tissue and cfDNA samples. Good concordance of SNV and focal CNV detection between tumour and cfDNA in patients with active extracranial disease (concordant in 25/29 patients) was shown. Of those, 18 patients that had SNVs detected in tissue, 94.4% (17/18), had one or more tissue-confirmed SNV detectable in the matched cfDNA sample (Figure 4-3 A) with VAF=0.1% - 57% (Appendix Table 2). In one patient with Wilms tumour, cfDNA had low ctDNA fraction (<10% ctDNA) and the two SNVs expected from tissue sequencing were not detected by cfDNA profiling. In two other patients (patients 13 and 29), the SNV with higher VAF in tissue was detected in cfDNA, but the lower one was not (Appendix Table 2). For patient 29 it is most likely due to low ctDNA purity. However, patient 13 had very high ctDNA levels and the absence of the subclonal variant in ctDNA might be explained by tumour evolution, as the time between tissue biopsy collection and the blood sample for liquid biopsy was more than 5 years.

Several cfDNA-unique SNVs were detected, most of them in patients with neuroblastoma. These differences between tissue and cfDNA results were most likely related to the large gap in time between tissue and blood collection, that

might have led to tumour evolution and/or due to the heterogeneity of the tumours that might not be fully represented in single tissue biopsy. The tissue samples matching the time of blood collection were not available in these cases to confirm any of these hypotheses. In addition, no oncogenic genomic alterations were detectable in the plasma of patients with brain tumours and patients with minimal or no tumour burden at the time of blood collection (Appendix Table 2). The limitations of blood based cfDNA analysis for patients with brain tumours will be discussed in detail in the next chapter.

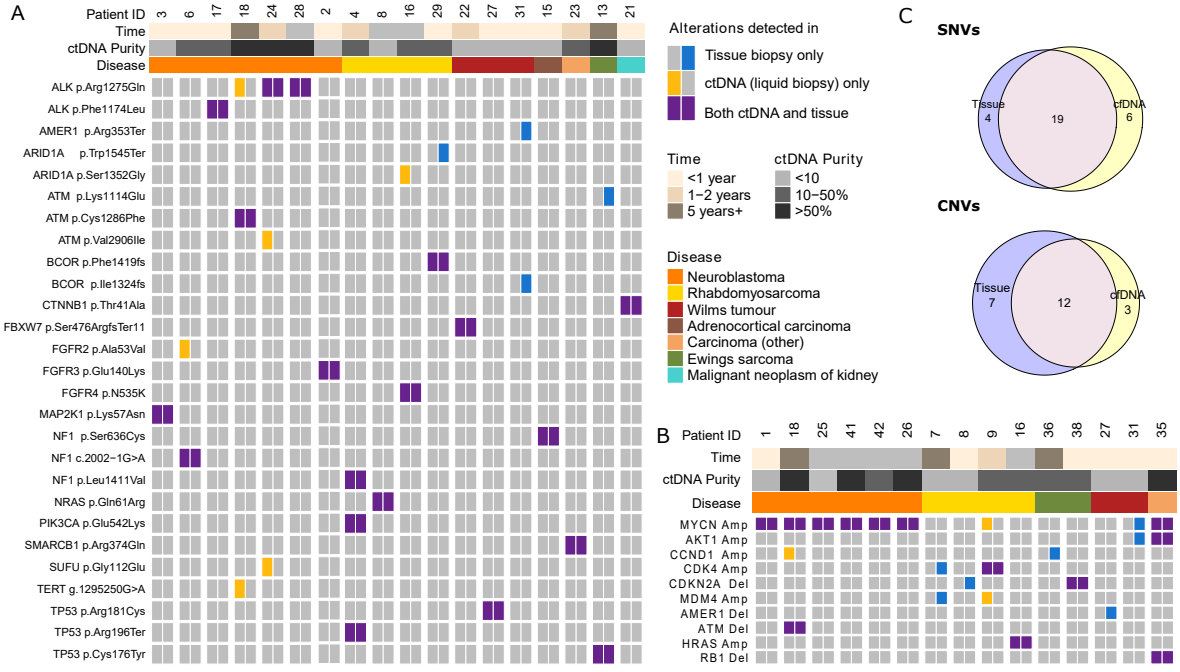


Figure 4-3 Comparison of findings between paediatric cfDNA and tissue samples in extracranial tumours. A. Comparison of SNVs detected in plasma vs tissue sequencing B. Comparison of CNVs detected in plasma vs tissue sequencing. Each column represents single patient, colour coded for disease type at the top and ctDNA purity and time between tissue and liquid biopsy indicated at the scale on top for each patient. Alterations in purple have been detected in both tissue and cfDNA, blue detected in tissue but not cfDNA and orange detected in cfDNA but not tissue. C Venn diagram comparing numbers of alterations detected in in tissue versus cfDNA.

Focal copy number variant detection by ctPC panel sequencing was combined with lcWGS from the same library preparation to allow comprehensive molecular profiling of CNVs. Focal CNVs expected from tissue sequencing were detected in 66.7% (10/15) of patients with active extracranial disease (Figure 4-3 B). All negative cases had very low ctDNA purity (<10% ctDNA) except for one (patient 36 with CCND1 amplification, Figure 4-3 B). In this case there was ~3 years between tumour and plasma sequencing, suggesting that the differences might be due to tumour evolution.

lcWGS allows estimation of the ctDNA fraction in cfDNA, which can help with clinical interpretation of ctPC results. In this cohort, we detected variants expected from tissue sequencing in all samples with ctDNA purity above 10%, which is consistent with studies in adult cancers where the reported cut-off of ctDNA purity is 5% ctDNA [145,174,175]. It is important to highlight however, that samples of low purity should not be excluded from panel analysis, as evident by detection of *MYCN* amplifications in neuroblastoma even in samples with <10% ctDNA purity. We conclude that false negatives are more likely in low purity samples, but detection of variants is still possible.

4.3 Stratified Medicine for Paediatrics study: cfDNA profiling at relapse

The initial cfDNA NGS profiling study (CCR-4294) showed that molecular profiling of cfDNA is possible for paediatric patients with solid tumours. However, the tissue and blood samples were not always collected at the same timepoint, making it difficult to interpret the cases with discordant results. The differences could represent tumour evolution or heterogeneity, but without time-matched samples we could not prove either hypothesis. Therefore, to better understand the dynamics of ctDNA in the blood of paediatric cancer patients with solid tumours I have analysed a cohort of time-matched cfDNA and tissue samples from patients at relapse in the Stratified Medicine for Paediatrics study, where plasma samples were obtained at the time of tissue collection, enabling direct comparison of the results. Additionally, tissue biopsy taken at diagnosis was available for a subset of patients allowing to evaluate tumour evolution and assess if cfDNA profiling at relapse can reliably detect these changes.

4.3.1 ctDNA levels in different cancer types

The first 370 patients enrolled to the Stratified Medicine for Paediatrics study were analysed to evaluate the levels of ctDNA and the ability to detect clinically relevant genetic alterations in cfDNA in patients with various solid tumours at relapse. 370 plasma samples (303 with a successfully sequenced matched tissue biopsy) at the time of relapse were analysed using targeted panel sequencing (with paired germline sample to exclude germline variants) and 358 samples were

analysed using lcWGS (297 with a matched tissue biopsy) (Figure 4-4). Additionally, 23 plasma samples were collected from the patients who experienced second relapse while on trial and one patient with a third relapse while on the trial. Five patients with only cfDNA sample and no tissue profiling results were also included in this analysis as well as 16 on-trial relapse samples of cfDNA without the tissue sample at the 2nd relapse.

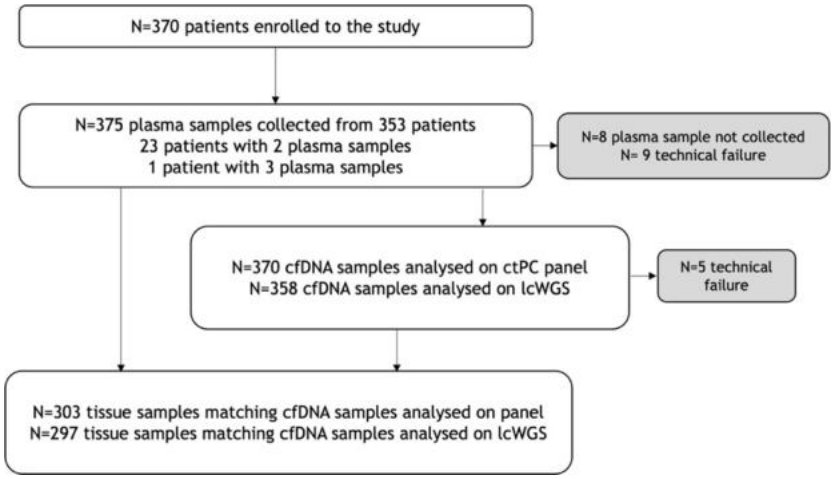


Figure 4-4 cfDNA samples included in the study. Technical failure includes failure in cfDNA extraction or NGS library preparation.

In total, 375 plasma samples were collected, with the mean volume of plasma of 6.2 ml (range of 0.6-9ml), with no significant differences in the volume of plasma collected in different age groups (Figure 4-5 A, B; Wilcoxon rank sum test, p=0.014). The mean total cfDNA yield of 365 ng (range of 6.3ng-19μg) was observed with no significant differences between different cancer types (Figure 4-5 C, D; Wilcoxon rank sum test, p=0.019). High fraction of HMW contamination from genomic DNA (<50% of DNA in cfDNA range) was observed in 42 cfDNA samples. However, it is important to note that fragment size analysis was performed only on samples with high yield (higher than 30ng of cfDNA/ml of plasma) and the actual levels of HMW contamination may be higher. The samples with high HMW fraction were processed by either performing bead-based size selection (21 samples) as described in the validation chapter or with increased input into the library preparation (21 samples). These samples were excluded from cfDNA yield analysis comparison in different cancer types. The total cfDNA yield varied from 5ng to >10μg with 85% of patients having at least 50ng of cfDNA available for analysis (50ng is the optimal input into the ctPC library preparation),

indicating that additional experiments on these samples would be possible in the future (Figure 4-5 E).

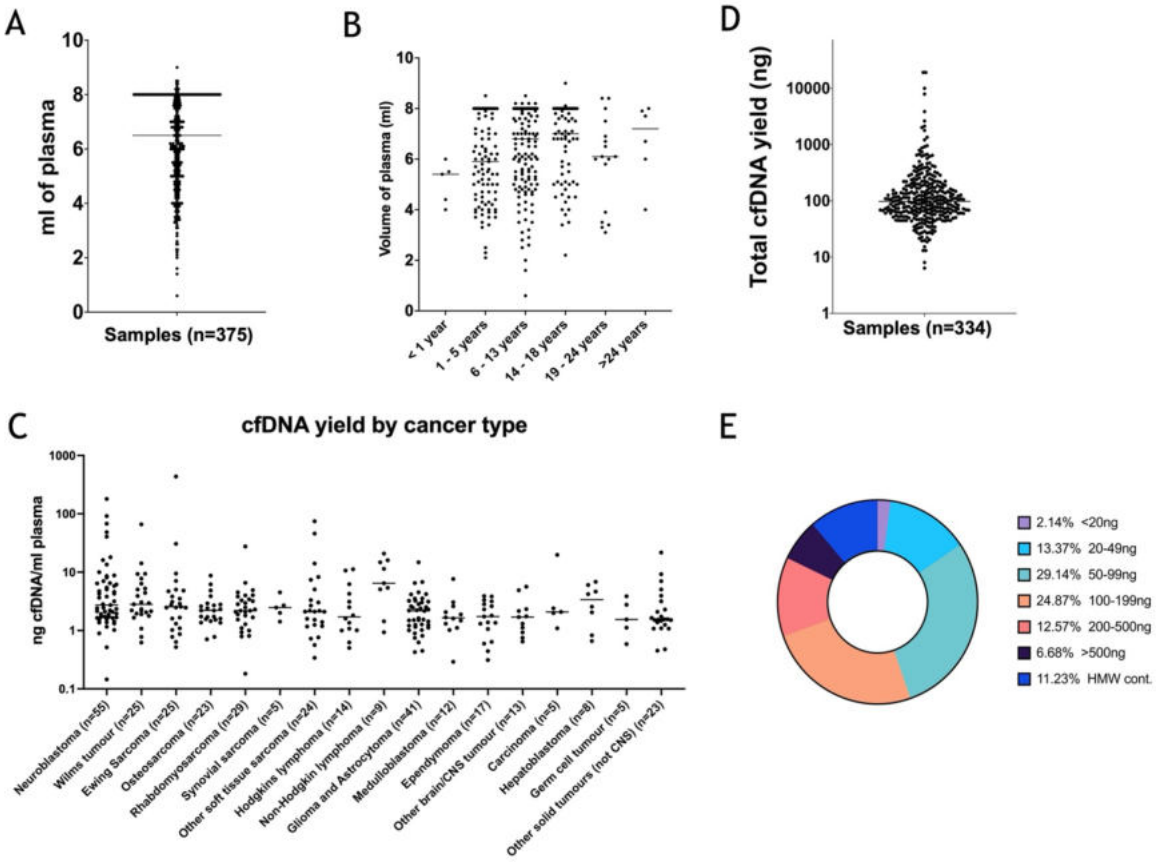


Figure 4-5 Characteristics of samples included in the study. (A) Volume of plasma extracted per patient, mean indicated as black line (mean 6.2ml). The high number of samples at 8ml mark, creating the second black line is due to the maximum input of the extraction protocol being 8ml. (B) Volume of plasma collected per patient in different age groups in this cohort. (C) cfDNA yield per ml of plasma in different cancer types. Samples with high molecular weight DNA contamination (as determined by fragment size analysis) excluded from the plot. (D) Total cfDNA yield per sample, median indicated as black line (median 97.7ng, mean 365ng); samples with high molecular weight DNA contamination (as determined by fragment size analysis) excluded from the plot. (E) Pie chat of total cfDNA yield (total ng), HMW cont. – high molecular weight contamination.

Even though cfDNA levels were similar in all cancer types included in this study, the ctDNA fraction was shown to be highly variable between different cancer types. Using lcWGS to estimate the purity of ctDNA (the fraction of ctDNA in cfDNA) it was shown that patients with neuroblastoma have significantly higher ctDNA levels when compared to all other cancer types (Wilcoxon test, $p<0.0001$) and patients with CNS tumours have significantly lower ctDNA levels (Wilcoxon test, $p<0.002$) (Figure 4-6 A). High ctDNA levels have been observed in several patients with other cancer types, such as various sarcomas, Wilms tumour, lymphoma and hepatoblastoma, but the differences in these cancer types were

not statistically significant. The main limitation of using lcWGS to estimate ctDNA purity is low resolution, as the limit of detection is ~10% and it depends on the presence of genome-wide copy number alterations. Another way to evaluate the levels of ctDNA could be by calculating cancer cell fraction from the VAF of clonal variants known from tissue detected by ctPC panel sequencing, which has the limit of detection of 0.125%, giving much higher resolution. However, this method is only applicable in patients whose tumours harbour SNVs or indels and without the knowledge of copy number status of the genomic locus of the variant or purity of the sample, cancer cell fraction is difficult to estimate. However, simply reporting the highest VAF of variant reported in each sample highlights the ability to detect low level ctDNA in various cancer types (Figure 4-6 B).

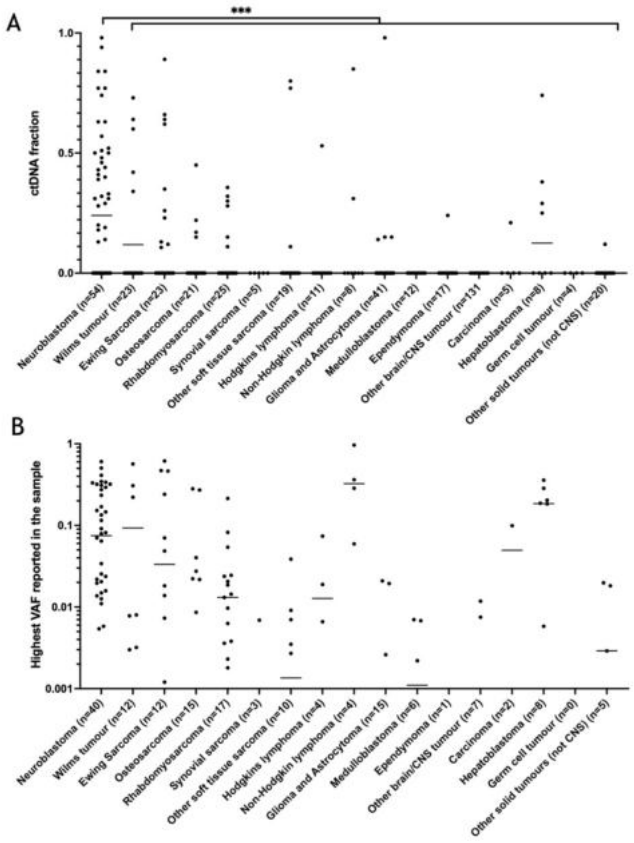


Figure 4-6 ctDNA fraction in cfDNA in patients with various solid tumours. (A) ctDNA fraction in cfDNA calculated using ASCFLP equations using lcWGS data. *** indicates statistically significant difference (Wilcoxon test, $p<0.0001$) (B) ctDNA fraction in cfDNA in patients with various solid tumours estimated as the VAF of the highest reported somatic variant in the sample in the targeted panel sequencing (germline variants excluded by analysis of matching blood cell pellet sample). VAF shown on log scale for better separation of low-level variants.

Neither method for ctDNA fraction estimation is perfect - they rely on genome wide CNV or SNV detection and low purity samples (<10% ctDNA) are underestimated using lcWGS. Another way to detect ctDNA could be by fragment

size analysis. It has been reported that ctDNA has shorter fragment size than cfDNA from healthy tissues. I have explored the fragmentation profiles of the samples in this study and observed a small shift in the peak of median density of the fragment sizes in high ctDNA purity samples when compared to low purity samples (Figure 4-7). On its own fragmentation analysis is of limited clinical use, but it may be helpful in interpreting samples with simple genomic profiles.

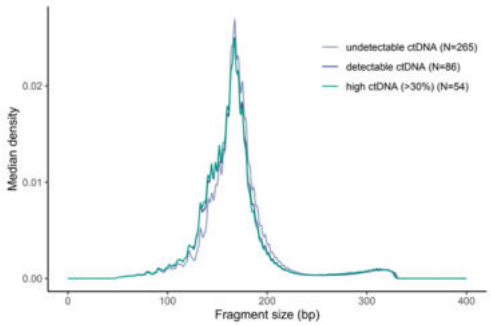


Figure 4-7 Median fragment size distributions of cfDNA from plasma, divided into three groups based on ctDNA fraction estimation by lcWGS: undetectable ctDNA, detectable ctDNA ctDNA (>10% ctDNA fraction) and high ctDNA fraction (>30% ctDNA fraction).

4.3.2 Single nucleotide variant and indel detection using panel sequencing in cfDNA and tissue

To systematically evaluate the possibility of detecting tumour-specific molecular alterations in cfDNA, panel sequencing and lcWGS was performed on time matched tissue and cfDNA samples from paediatric cancer patients experiencing disease relapse. In this cohort, 163 patients had SNVs or indels detected in tissue or cfDNA with a total of 313 alterations detected by panel sequencing: 31% of alterations were detected in both tissue and cfDNA, 41% of alterations were detected only in tissue and 26% of alterations were detected only in cfDNA (Figure 4-8 A). Overall, 42% of alterations expected from tissue sequencing were detected in cfDNA, with very good concordance between tissue and cfDNA in high purity (>10% ctDNA) samples, where 91% of alterations detected in tissue were also detected in cfDNA. Even in low ctDNA purity samples, variants were detected in cfDNA using panel sequencing. In fact, 40% of overlapping variants were detected in low purity ctDNA samples.

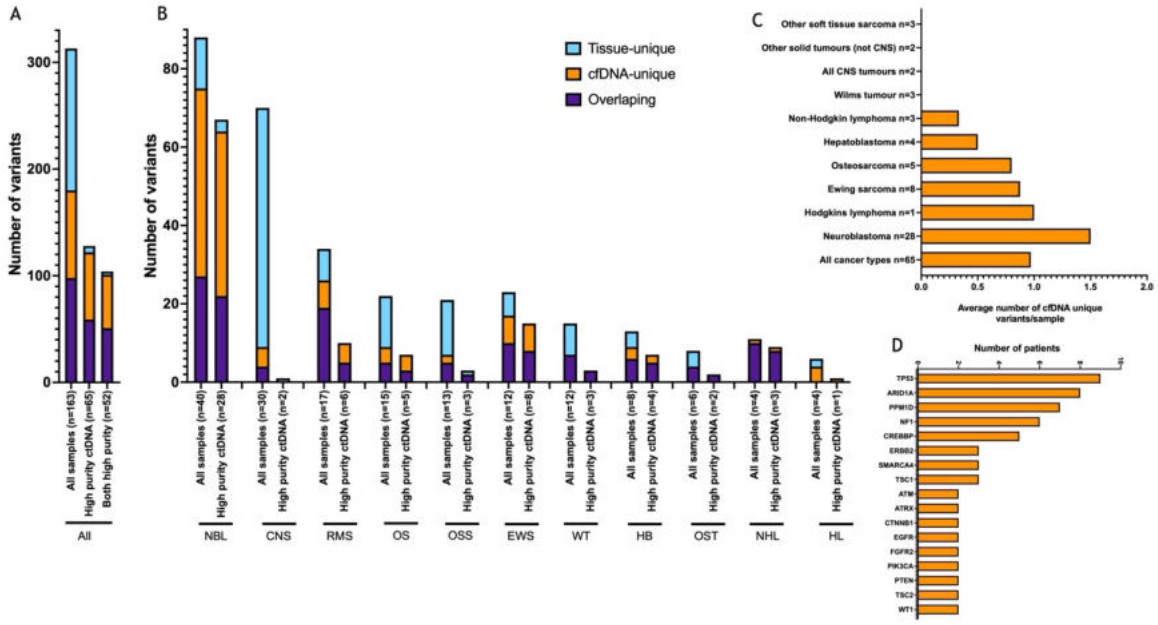


Figure 4-8 SNV and indel detection in cfDNA versus tissue. (A) The number of variants detected in cfDNA, tissue and both in all samples included in the study, high purity ctDNA samples (>10% ctDNA by lcWGS) and high purity ctDNA and tissue samples (>10% ctDNA by lcWGS). (B) The number of variants detected in cfDNA, tissue and both in different cancer types. All samples and only high ctDNA purity samples. (C) The average number of cfDNA unique variants per sample in different cancer types, based only on high ctDNA purity samples. (D) The number of cfDNA unique variants reported in different genes, only genes with more than one cfDNA-unique variant shown, all cancer types combined. NBL – neuroblastoma, CNS – CNS tumours, RMS – rhabdomyosarcoma, OS – osteosarcoma, OSS – other soft tissue sarcoma, EWS – Ewing sarcoma, WT – Wilms tumour, HB – hepatoblastoma, OST – other solid tumours, NHL – non-Hodgkin lymphoma, HL - Hodgkin lymphoma.

The best agreement between tissue and cfDNA profiling (regardless of ctDNA purity) was observed in patients with non-Hodgkin lymphoma (n=4), were all variants expected in the tissue were detected in cfDNA (Figure 4-8 B). However, low number of patients raise the need for further validation. Good detection of variants known form tissue profiling was also shown in patients with rhabdomyosarcoma (n=17), neuroblastoma (n=40), Ewing sarcoma (n=12) and hepatoblastoma (n=12) where 70%, 68%, 63% and 60% of tissue variants were detected in cfDNA, respectively (Figure 4-8 B). Very poor detection of tissue variants was observed in patients with Hodgkin lymphoma (n=4) where no tissue-specific variants were detected in cfDNA and CNS tumours (n=30) where only 6% of tissue-specific alterations were detected in cfDNA (Figure 4-8 B). The highest number of variants detected only in tissue was observed in patients with CNS tumours, which can be explained by low ctDNA levels in the plasma of these patients. Quite high fraction of tissue-unique variants was also observed in

patients with osteosarcoma and other soft tissue sarcomas, all explained by low ctDNA levels in these samples (Figure 4-8 B).

Importantly, a high number of cfDNA unique variants was present in all analyses, even if low purity tissue samples were excluded (Figure 4-8 A). The highest number of cfDNA-unique variants were reported in patients with neuroblastoma, which was also the most common cancer type in this cohort (Figure 4-8 B). The average number of cfDNA unique variants per sample was highest in patients with neuroblastoma, followed by patients with Hodgkin lymphoma, Ewing sarcoma and osteosarcoma, albeit much lower numbers of samples were included in the analysis for these cancer types and further validation in bigger cohorts might be useful (Figure 4-8 C). Interestingly, none or very few cfDNA-unique variants were observed in patients with soft tissue sarcomas, Wilms tumour, osteosarcoma and other solid tumours (which include immature blastic tumour, Langerhans cell histiocytosis, giant cell tumour of bone, Desmoid-type fibromatosis) (Figure 4-8 B, C), which was not influenced by the differences in mutation burden across different cancer types (Appendix Figure 1).

The presence of cfDNA unique variants potentially indicates the ability of cfDNA profiling to better represent tumour heterogeneity by detecting different (sub)clones present in the tumour. Out of 81 cfDNA-unique variants, the most commonly altered genes were *TP53, ARID1A, PPM1D* and *CREBBP* (Figure 4-8 D). Even though *TP53* and *PPM1D* are commonly reported in clonal haematopoiesis, the haematopoietic origin of these variants was excluded by inspecting matching blood cell pellet sample for each patient. Importantly, in 6 patients, cfDNA-unique variants were present in the time-matched tissue, but below the limit of detection for tissue panel sequencing (Table 4-1), showing added value of cfDNA profiling.

Table 4-1 cfDNA-unique variants present in the matched tissue below the limit of detection of tissue reporting.

Trial ID	Disease	Gene	Change	Change protein	VAF in cfDNA	ctDNA purity	VAF in tissue	Tissue purity
SMP0349	Hodgkin lymphoma	*CREBBP*	c.4348T>A	p.Tyr1450Asn	0.0739	0.53	0.027	0.53
SMP0122	Hepatoblastoma	*CTNNB1*	c.95A>G	p.Asp32Gly	0.3584	0.74	0.028	0
SMP0204	Synovial sarcoma	*PIK3CA*	c.371C>T	p.Pro124Leu	0.0069	0	0.023	0.66
SMP0026	Rhabdomyosarcoma	*PTCH1*	c.3095T>C	p.Ile1032Thr	0.0032	0.15	0.048	NA
SMP0332	Neuroblastoma	*SMARCA4*	c.2808C>A	p.Cys936Ter	0.0158	0.39	0.028	0.37
SMP0154	Neuroblastoma	*TSC1*	c.2368T>C	p.Tyr790His	0.1519	0.48	0.042	0.37

To further investigate the characteristics of cfDNA unique variants, I have compared the VAF of cfDNA-unique variants versus the variants co-detected in tissue and cfDNA. The average VAF of cfDNA-unique variants was significantly lower than VAF of variants overlapping between tissue and cfDNA (Figure 4-9 A), even though higher threshold for reporting cfDNA-unique variants was applied, supporting the hypothesis of high presence of subclonal variants in cfDNA. Additionally, the VAF of cfDNA-unique variants was not correlated with the ctDNA purity of the sample, with high numbers of very low VAF variants detected even in high ctDNA purity samples (Figure 4-9 B). In contrast, much better correlation of tissue and cfDNA overlapping variants with the sample purity was observed (Figure 4-9 B), further supporting this hypothesis. The stringent variant filtering, use of UMIs in the sequencing library preparation and further manual curation to only include cancer related, likely pathogenic low frequency variants minimise the risk that these cfDNA unique variants could be technical artefacts.

Cancer-type specific differences were observed, with the highest number of cfDNA unique variants and the most significant difference between VAF of cfDNA unique and overlapping variants seen in patients with neuroblastoma (Figure 4-9 C), which are highly heterogenous tumours [176,177]. Clear differences between cfDNA unique and overlapping VAFs were also seen in patients with Ewing sarcoma and non-Hodgkin lymphoma, where overlapping variants are most likely clonal, and the

cfDNA-unique ones are likely subclonal and therefore not detected by single tissue biopsy. Interestingly, in patients with various sarcomas and hepatoblastoma the VAF of cfDNA unique and overlapping variants was similar (Figure 4-9 C). This could indicate that single tissue biopsy is under representative even of clonal variants in these cancer types or metastatic sites and cfDNA findings might be more all-encompassing. However, the information on the sites of tumours and the metastatic status was not available in this cohort to test this hypothesis. It is possible that in heterogenous tumours the more aggressive subclones are shedding more ctDNA into the blood while the heterogeneity is not captured by the tissue biopsy. Also, it is important to note that in most cases only one variant per patient was reported and the differences might be caused by varying purities of ctDNA in different patients. In most patients with several alterations detected the variants overlapping between tissue and cfDNA profiling were of highest VAF and cfDNA unique variants were detected at lower levels. However, there were several exceptions where cfDNA-unique variants were of higher VAF than the overlapping ones and these cases will be explored in more detail in following sections. Overall, this data strongly supports the added value of cfDNA profiling in addition to tissue profiling, with good correlation between variant detection in high purity ctDNA samples and tissue and high numbers of cfDNA-unique variants.

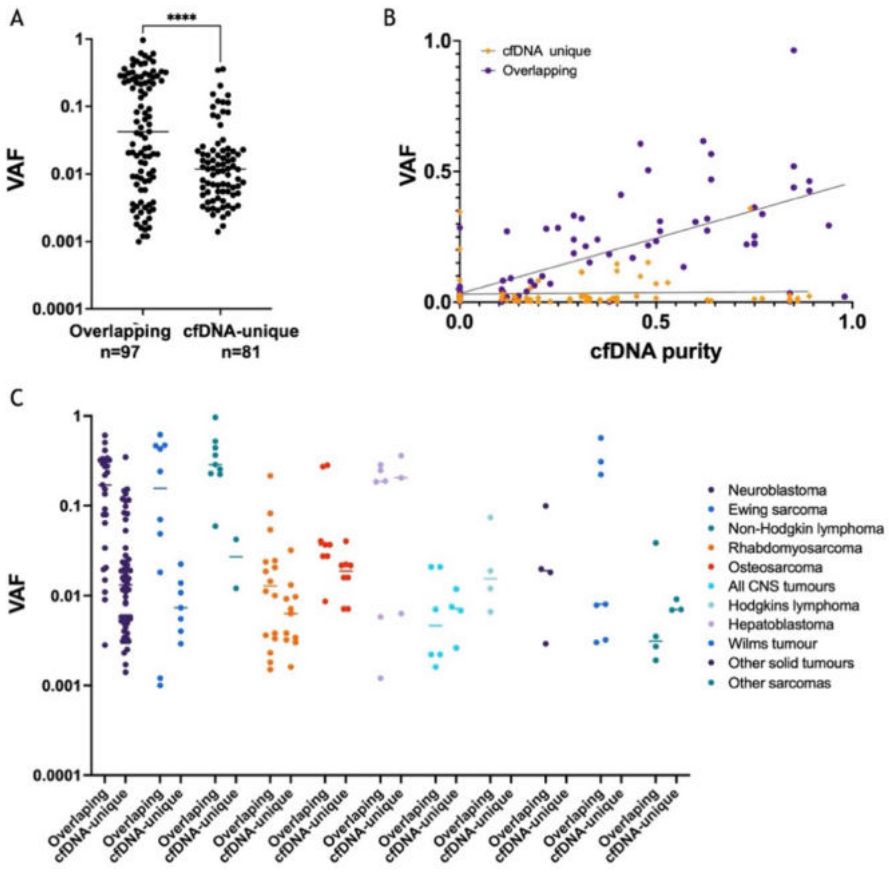
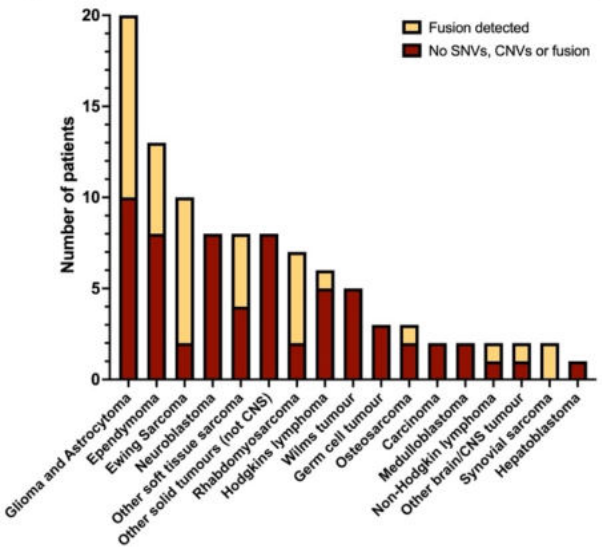


Figure 4-9 Differences between cfDNA-unique and tissue and cfDNA overlapping variants. (A) The VAF of overlapping and cfDNA-unique variants, **** p<0.0001 (t test). (B) The correlation of VAF and ctDNA purity of the given sample. cfDNA unique variants in orange rhombuses, variants overlapping between tissue and cfDNA in purple dots. (C) The differences between VAF of cfDNA-unique and overlapping variants in different cancer types.

Ultimately, panel based NGS profiling relies on the presence of SNVs and indels in the tumour. Paediatric cancers are known to have relatively simple genomic profiles with low numbers of mutations and in this cohort 102 patients did not have any SNVs, focal CNVs or indels reported using panel sequencing in tissue or cfDNA. This included patients with a range of tumours, with majority being CNS tumours and sarcomas (Figure 4-10 A). Not surprisingly, a high number of sarcomas were driven by fusion genes (as detected by RNA targeted profiling) as well as half of gliomas and astrocytomas (Figure 4-10 B). However, in this cohort 64 patients had no clear driver identified by either targeted DNA panel or RNA panel testing of tissue biopsy, highlighting that more comprehensive profiling methods might be needed for some patients. lcWGS profiling was successfully performed for 52 out of these 64 patients and 33 had genome wide copy number changes detected. This included alterations with prognostic value, such as 1q gain detected both in tissue and cfDNA in patient SMP0306 with Wilms tumour and 11q loss in four patients with neuroblastoma detected in tissue, but not cfDNA (<10% ctDNA by lcWGS in

these samples). The full spectrum of alterations detected by lcWGS in different cancer types will be discussed in detail in the next section.

A



B

Disease	Fusion	Number of patients
Glioma and Astrocytoma	BRAF:KIAA1549	7
	ALK:CCDC88A	1
	BRAF:RNF130	1
	PLAG1:EWSR1	1
Ependymoma	MAML2:C11orf95	1
	FRMD8:GRHPR	1
	ATP11A:NPM1	1
	BCOR:ELF4	1
	MAMLD1:YAP1	1
Ewing Sarcoma	FLI1:EWSR1	7
	NFATC2:EWSR1	1
Other soft tissue sarcoma	SSX1:SS18	1
	WT1:EWSR1	1
	CAMTA1:WWTR1	1
Rhabdomyosarcoma	FOXO1:PAX3	3
	FOXO1:PAX7	1
	FUS:TFCP2	1
Hodgkin lymphoma	ROS	1
Osteosarcoma	ZMIZ1:KDM5A	1
Non-Hodgkin lymphoma	NPM1:ALK	1
Other brain/CNS tumour	BEND2:EWSR1	1
Synovial sarcoma	SS18:SSX1	2

Figure 4-10 Characteristics of the patients with no SNVs, indels and focal CNVs detected. (A) The number of patients with no alterations detected in panel sequencing in different cancer types. The number of patients with gene fusions detected by RNA profiling shown in yellow. (B) Detailed list of fusions detected in different cancer types in this cohort.

Overall, 17 of the 64 patients with no driver identified by either DNA or RNA profiling had low sample purity as estimated by lcWGS both in tissue and matched cfDNA. In these cases, it is hard to know if both samples were of poor quality or indeed no large scale CNVs were present in the tumours. A more comprehensive profiling, such as WGS or methylation profiling of the tissue samples (which are expected to have higher purity than ctDNA normally) would be able to answer the question and potentially identify cancer related alterations. Interestingly, no enrichment of a particular cancer type was observed in this subset of patients.

4.3.2.1 Neuroblastoma: cfDNA is representative of tumour alterations and high numbers of cfDNA-unique variants are reported

Neuroblastoma was the cancer type with the highest ctDNA levels in this cohort (average of 0.32 ctDNA purity, range 0-0.94) and 60 patients with neuroblastoma at relapse were analysed using ctPC panel and lcWGS. Out of these, 40 patients had SNVs detected either in tissue or cfDNA, with at least one SNV expected from tissue profiling detected in cfDNA in 83.9% (26/31) of patients (Figure 4-11). Only in five patients (all with low ctDNA purity), SNVs expected from tissue profiling

were not detected in cfDNA. As a part of Stratified Medicine Paediatric Programme, molecular tumour board discussed the tissue profiling results with an aim to match patients with available treatments and clinical trials. Only findings that would make a patient eligible for an open (or to be expected to be open in following few weeks/months) clinical trial or would make them eligible for compassionate access to a targeted agent were considered actionable. In this cohort, molecular tumour board determined 15 SNVs to be potentially actionable (mutations in *ALK, ATRX, BRAF, FGFR1, MAP2K1, NF1, PIK3CA* and *PTPN11*) and 88% of these variants were detected in cfDNA as well. Importantly, in 10 patients no SNVs were detected in tissue, but cfDNA unique variants were detected, including 9 potentially targetable variants, such as pathogenic *ALK* and *NF1* mutations. Low purity in the tissue samples was observed in most of these patients (mean purity 0.29, two samples with 0% purity), highlighting the possibility to combine cfDNA and tissue results and provide actionable targets when tissue-based profiling is unsuccessful (Figure 4-11).

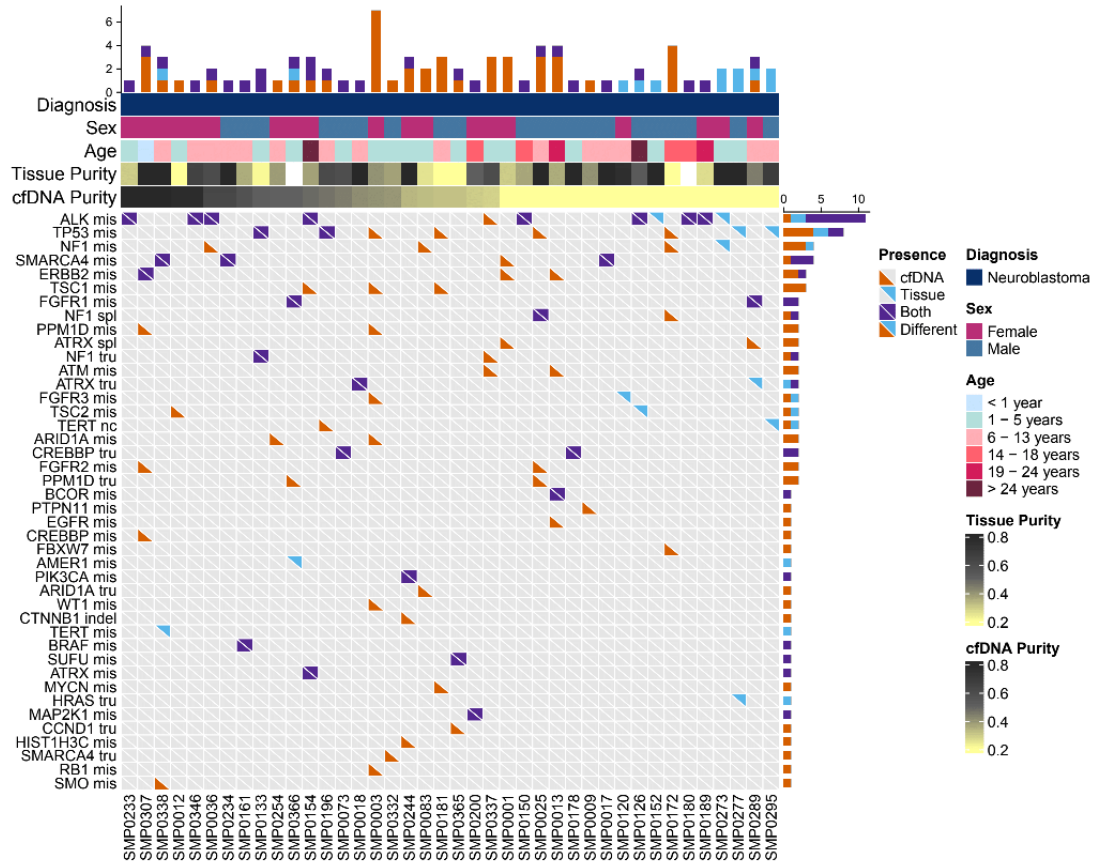


Figure 4-11 Comparison of SNV and indel detection between cfDNA and tissue biopsy in patients with neuroblastoma (only cases with successful tissue and cfDNA analysis with at least one SNV or indel are shown). Each column represents single patient, ordered by ctDNA purity (by lcWGS). The histograms on the top indicate the total number of SNVs and indels in each patient (blue – detected only in tissue, orange – detected only in cfDNA, purple - detected in both); histograms on the right indicate the number of patients with alterations in the indicated gene.

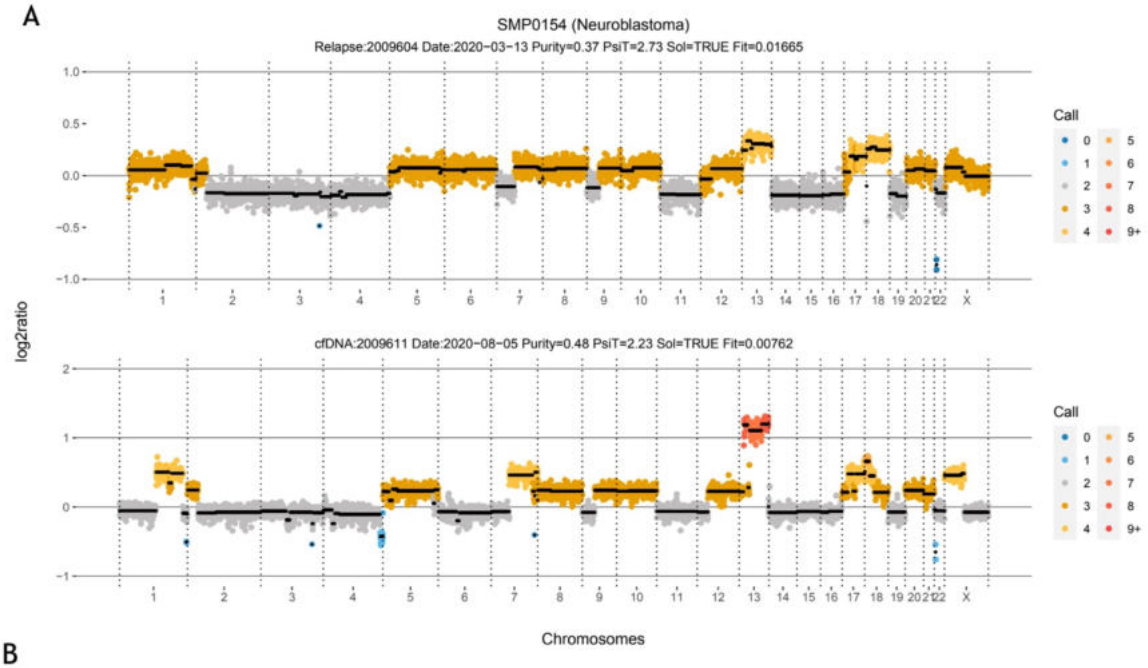
Pathogenic missense variants in *ALK* were the most common alteration, detected in both tissue and cfDNA, except in two patients where they were not detected in cfDNA and one patient where it was not detected in the tissue (Figure 4-11). The second most common alteration was *TP53* missense mutations followed by *NF1* and *SMARCA4* missense mutations. Interestingly, *TP53* variants were more frequently cfDNA unique, with 2 patients carrying 2 different *TP53* cfDNA-unique variants. A wide range of variants was reported in his cohort, with very few other recurrent alterations between the patients.

High risk neuroblastoma is often heterogenous, presenting with metastatic disease, therefore the presence of cfDNA unique variants likely indicates the potential of cfDNA to identify subclones present in the tumour that are not captured by tissue profiling. This is supported by two patients where after inspecting tumour sequencing reads, variants considered to be cfDNA unique (*SMARCA4* C936Ter in SMP0332 and *TSC1* Y790H in SMP0154) were detected in the time-matched tissue below the limit of detection of the tissue sequencing assay (Table 4-1). Additionally, both variants were also present in the diagnostic tissue biopsy in these patients. Furthermore, in patient SMP0172 four cfDNA unique variants were detected. However, there was no evidence of cfDNA unique variants in time matched tissue, but the *TP53* p.R273P variant was present in the diagnostic tissue biopsy from this patient, further highlighting the added value of cfDNA profiling and the variability of tissue biopsies.

In this cohort, 8 patients with neuroblastoma had the variant expected from tissue sequencing detected in cfDNA and additional cfDNA unique variants. In most cases the variant overlapping between tissue and cfDNA profiling had the highest VAF, while the cfDNA unique variants were of lower VAF. This provides further support to the hypothesis that cfDNA can better represent tumour heterogeneity and allow detection of subclones. However, in three patients a more complex picture emerged. In patients SMP0013 and SMP0307 the cfDNA unique variants had higher VAF that the variant detected both in tissue and cfDNA. Another illustrative example of cfDNA/tissue discordant results was SMP0154 (Figure 4-12) where clonal pathogenic *ALK* variant was detected in both tissue and cfDNA profiling. However, only one of the two subclonal variants reported in the tissue was detected in cfDNA. Additionally, a cfDNA unique variant in *TSC1* was reported,

which was present in tissue, but below the reporting limit (present at VAF=0.04). Interestingly, all variants detected in cfDNA were also present in the primary diagnostic tissue of this patient. lcWGS was available for cfDNA and the tissue at relapse and while the profiles were similar, several differences between the two samples were observed (Figure 4-12 A). The absence of some amplifications in cfDNA sample (such as chr6 and parts of chr1) strongly suggest that while the changes look clonal within the tissue biopsy, they are not truly clonal in the whole tumour. The cfDNA profile is more informative, allowing to characterise CNV changes and speculate of branched evolutionary pattern in this patient.

A



B

PATIENT	Gene	Change	Change protein	Reported in cfDNA	VAF in cfDNA	Reported in tissue at relapse	VAF in tissue at relapse	VAF in primary tissue
SMP0154	ALK	c.3824G>A	p.Arg1275Gln	YES	0.505	YES	0.476	0.57
SMP0154	ATRX	c.6809A>G	p.Glu2270Gly	YES	0.218	YES	0.124	0.162
SMP0154	ATRX	c.5408G>A	p.Arg1803His	NO	0.000	YES	0.075	0
SMP0154	TSC1	c.2368T>C	p.Tyr790His	YES	0.152	NO	0.042*	0.229

Figure 4-12 Molecular profiling results of patient SMP0154 with neuroblastoma. (A) lcWGS profiles of tissue at relapse (top plot) and cfDNA (bottom plot). (B) The summary of variants reported by panel sequencing in cfDNA, primary and relapse tissues. *Variant present in the sequencing reads, but below the limit of reporting for tissue panel sequencing (LoD=0.05).

Detection of focal copy number changes for selected genes was also evaluated using panel sequencing. Focal CNV detection is more challenging in cfDNA, especially in low purity samples, with only 40.0% (10/25) of patients having at least one focal CNV expected from the tissue detected in cfDNA (Figure 4-13). *MYCN* amplification/gain was the most commonly detected alteration in patients

with neuroblastoma in this cohort and it was detected in cfDNA in 70.6% (12/17) of the patients. Similar to SNVs, the variant was mostly missed in samples with low ctDNA purity, except in two cases.

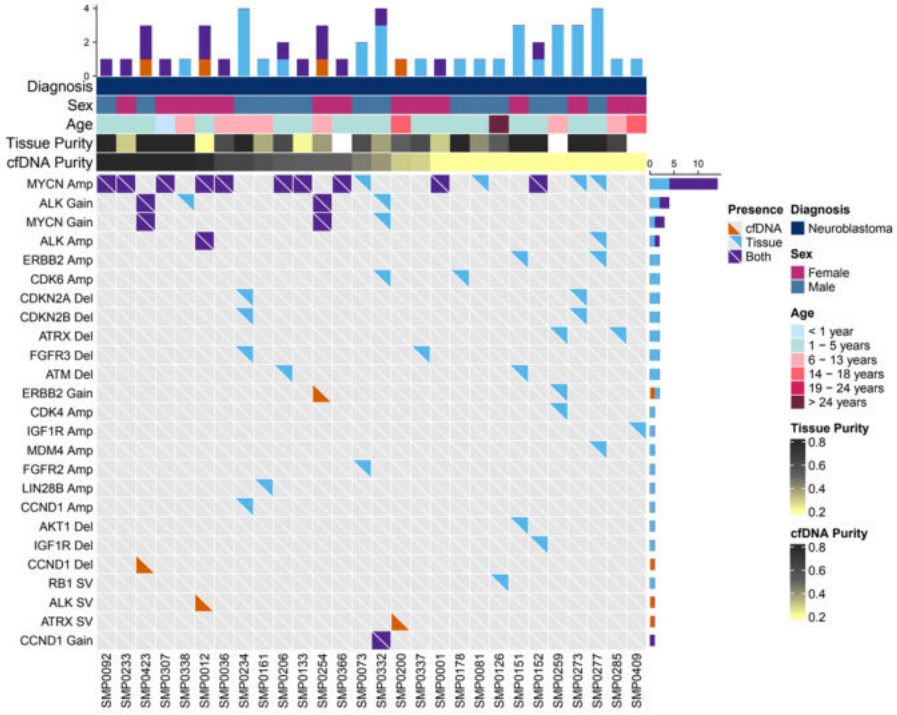


Figure 4-13 Comparison of CNV detection using targeted panel sequencing in cfDNA and tissue biopsy in patients with neuroblastoma (only cases with successful tissue and cfDNA analysis with at least one CNV in an actionable gene are shown). Each column represents single patient, ordered by ctDNA purity (by lcWGS, white colour indicates lcWGS was not successful and no purity estimate is available). Gain defined as sample ploidy +1, amplification as sample ploidy +2. The histograms on the top indicate the total number of CNVs in each patient (blue – detected only in tissue, orange – detected only in cfDNA, purple - detected in both); histograms on the right indicate the number of patients with alterations in the indicated gene.

For two patients who experienced a second relapse while on the trial, additional cfDNA samples were collected at the time of second relapse (Figure 4-14). In patient SMP0120 the *FGFR3* variant was detected in the tissue at diagnosis and at relapse, however it was not detected in a time matched cfDNA sample. However, the same variant was detected in cfDNA sample collected when patient experienced an on-trial relapse, highlighting the benefit of repeated sample collection for previously negative cfDNA samples. For one patient (SMP0289) two bone marrow aspirate samples were collected at relapse and plasma cfDNA unique variants were detected in cfDNA from both bone marrow aspirate cfDNA samples as well, adding confidence in their clinical significance and potential of bone marrow supernatant was cfDNA source.

An extraordinary case showing the accumulation of multiple new variants at relapse detected in cfDNA was observed in patient SMP0018. A clonal *ATRX* mutation was detected in tissue at relapse and subsequent on trial relapse and in two cfDNA samples collected at both relapses. At second relapse the patient acquired a somatic *MLH1* mutation, which was detected in the tissue (*MLH1* gene is not covered by ctDNA panel), and six additional lower VAF alterations, none of which were detected in cfDNA. However, at second relapse 32 cfDNA unique variants were detected. Multiple variants in the same gene have been observed with for example 4 variants in *NF1* and *ERBB2* and 3 variants in *CREBBP, EGFR, SMARCA4* and *TP53*, suggesting convergence towards acquiring mutations in key oncogenes. MLH1 plays a key role in mismatch repair, therefore the acquisition of somatic mutation at the time of second relapse potentially explains the accumulation of numerous alterations. ATRX mutations are also associated with chemoresistance and genome instability, due to DNA damage repair deficiencies associated with ATRX loss [178]. Additionally, this patient received multiple lines of therapy, including Temozolomide which has been reported to induce hypermutation in gliomas [179,180].

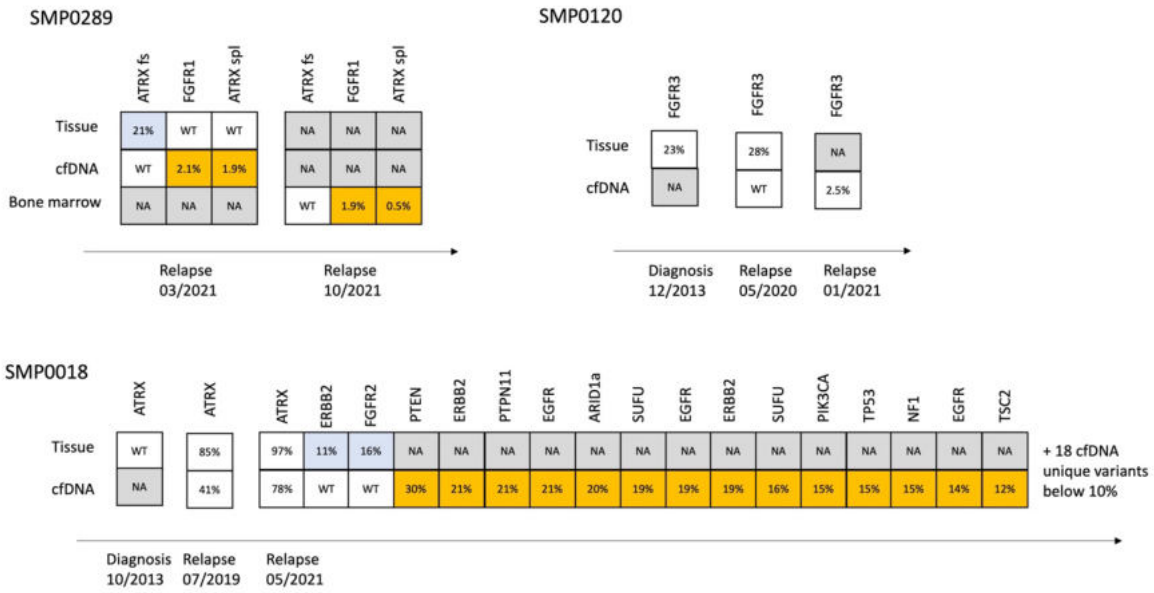


Figure 4-14 Serial liquid biopsy sampling results compared to tissue biopsy. The top boxes represent tissue sequencing results at a given time point, the bottom ones – cfDNA sequencing. NA – sample was not available. The cfDNA unique variants highlighted in orange, variants detected in tissue, but not cfDNA in blue.

4.3.2.2 Sarcoma: cfDNA is representative of tumour alterations in rhabdomyosarcoma and Ewing sarcoma

More variation in detecting variants found in the tissue sequencing was observed in patients with various sarcomas. The average ctDNA levels were much lower when compared to patients with neuroblastoma, with mean ctDNA fraction of 0.17 in patients with Ewing sarcoma, and an average below 0.1 in all other types of sarcoma (Figure 4-6). In this cohort, 57 patients out of 106 had had SNVs or indels detected in tissue or cfDNA and 13 had only focal CNVs detected by panel sequencing. Out of these, 58.9% (30/51) of patients had at least one SNV expected from tissue profiling detected in cfDNA (Figure 4-15). SNVs expected from the tissue were not detected in cfDNA in 21 patients (all with low ctDNA purity). Detection of focal copy number changes using panel sequencing was very difficult in cfDNA, with CNVs expected from tissue sequencing detected in cfDNA only in 10.8% (4/37) of patients (Figure 4-16).

Striking differences were observed between different types of sarcoma: 86.7% and 80% of patients with rhabdomyosarcoma and Ewing sarcoma respectively had at least one variant expected from tissue detected in cfDNA (Figure 4-15). In contrast, only 38.5% of patients with osteosarcoma had at least one variant expected from tissue detected in cfDNA. Importantly, in 4 patients (two rhabdomyosarcoma patients with *FOXO1:PAX3* fusion, one Ewing sarcoma patient with *EWSR1:FLI1* fusion and one patient with osteosarcoma) no SNVs were detected in tissue, but cfDNA unique variants were detected. Potentially clinically relevant alterations, such as *TP53* mutation in patient SMP0341 with rhabdomyosarcoma and were detected only in cfDNA.

Missense pathogenic variants in *TP53* were the most common alteration and they were detected in both tissue and cfDNA, except in three patients where variants were not detected in cfDNA and one patient where it was not detected in the tissue (Figure 4-15). This is particularly encouraging given that *TP53* mutations could be potentially actionable or act as prognostic and predictive biomarkers [181]. However, it is important to acknowledge that 54% of patients with sarcomas in this cohort (100% of patients with Ewing Sarcoma, 53% of patients with rhabdomyosarcoma) had fusion driver detected by RNA profiling and the value of SNVs detected needs to be interpreted in this context. Overall, in this cohort

molecular tumour board (MTB) evaluating tissue biopsy results determined 22 patients to carry actionable SNVs (such as *TP53*, *RB1*, *NF1* variants). The cfDNA profiling was able to detect at least one of these alterations in 63.6% (14/22) of these patients, with a total of 57.1% (16/28) of the alterations detected. In two patients (one with Ewing sarcoma and one with rhabdomyosarcoma) the potentially actionable *TP53* mutations expected from the tissue sequencing were detected in cfDNA and additional cfDNA unique *TP53* mutations were also detected (at the time of MTB discussion there was an arm of ESMART clinical trial (NCT02813135) including patients with TP53 variants).

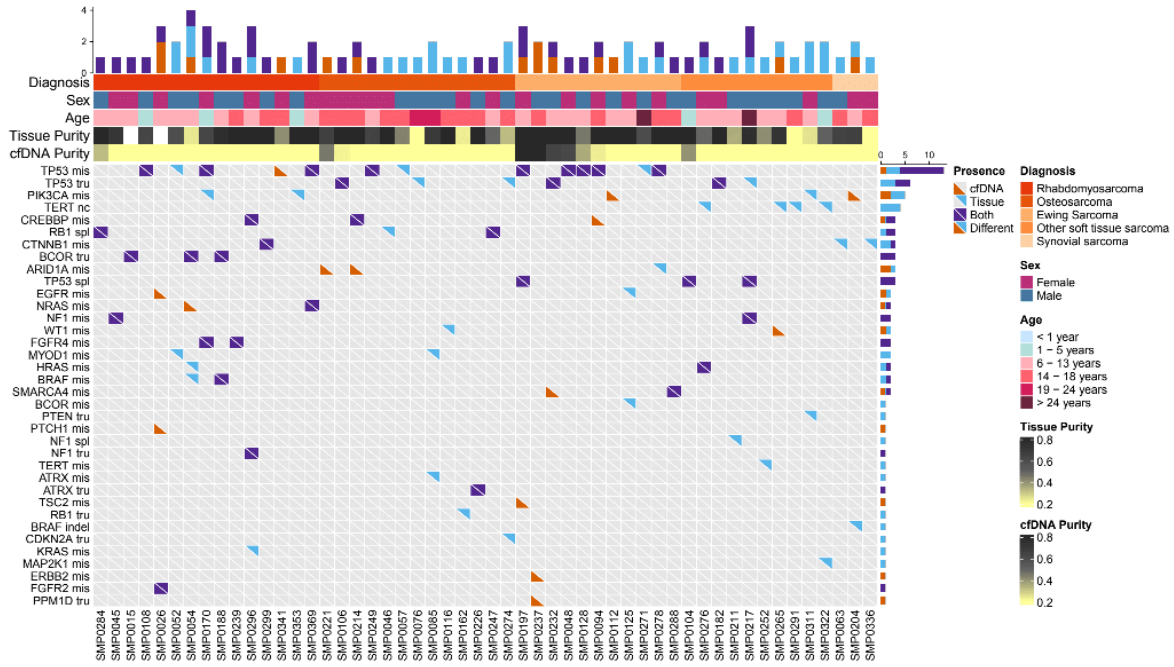


Figure 4-15 Comparison of SNV and indel detection between cfDNA and tissue biopsy in patients with sarcoma (only cases with successful tissue and cfDNA analysis with at least one SNV or indel are shown). Each column represents single patient, ordered by ctDNA purity (by lcWGS). The histograms on the top indicate the total number of SNVs and indels in each patient (blue – detected only in tissue, orange – detected only in cfDNA, purple - detected in both); histograms on the right indicate the number of patients with alterations in the indicated gene.

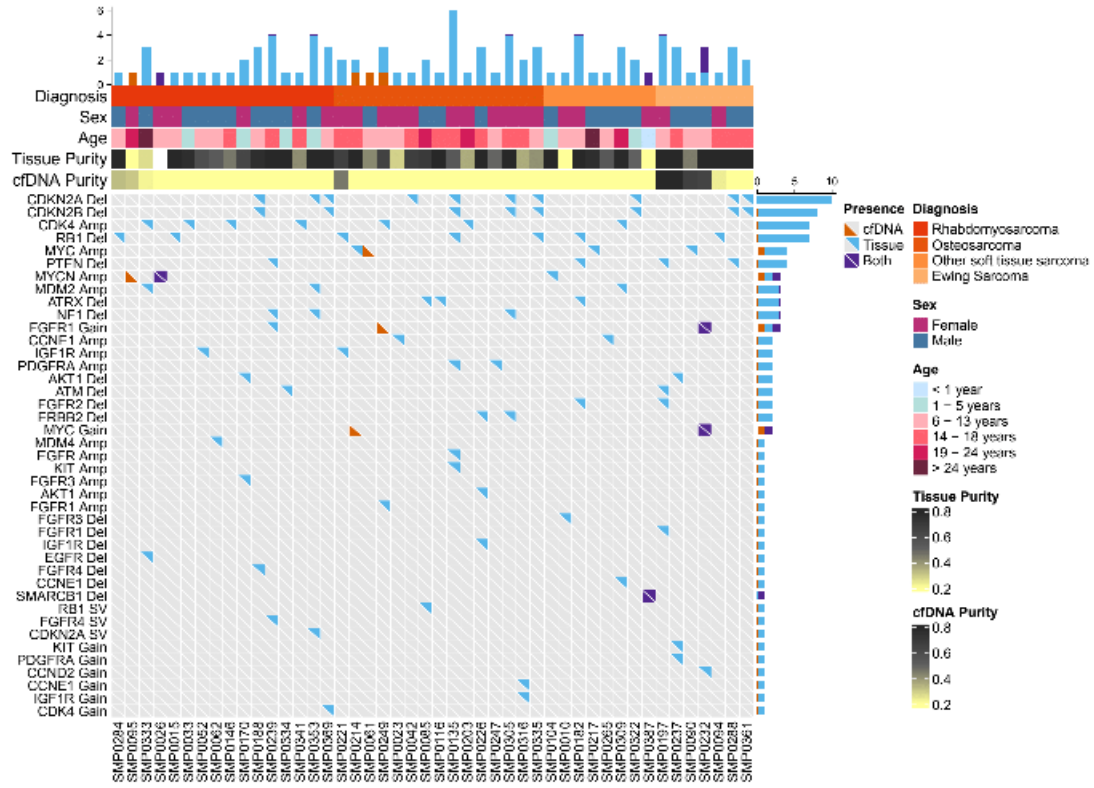


Figure 4-16 Comparison of CNV detection using targeted panel sequencing in cfDNA and tissue biopsy in patients with sarcoma (only cases with successful tissue and cfDNA analysis with at least one CNV in an actionable gene are shown). Each column represents single patient, ordered by ctDNA purity (by lcWGS, white colour indicates lcWGS was not successful and no purity estimate is available). The histograms on the top indicate the total number of CNVs in each patient (blue – detected only in tissue, orange – detected only in cfDNA, purple - detected in both); histograms on the right indicate the number of patients with alterations in the indicated gene.

For four patients who experienced a second relapse while on the trial, additional cfDNA samples were collected at the time of second relapse (Figure 4-17). For two patients (SMP0361 and SMP0090) cfDNA unique variants were detected only at second relapse and for one (SMP0057) a variant expected from the tissue was detected only in the on-trial relapse sample. For SMP0361 however a *TP53* variant was detected in tissue and cfDNA at the first on-trial relapse and two lower AF cfDNA unique variants in the same gene were also detected. A cfDNA sample collected a month later showed the same three *TP53* variants, supporting the robustness and added value of cfDNA profiling.

Figure 4-17 Serial liquid biopsy sampling results compared to tissue biopsy. The top boxes represent tissue sequencing results at a given time point, the bottom ones – cfDNA sequencing. NA – sample was not available. The cfDNA unique variants highlighted in orange.

4.3.2.3 Other extracranial paediatric solid tumours: cfDNA profiling is informative

One of the potential benefits of pan-cancer panel was the ability to detect clinically relevant and actionable variants in a wide range of paediatric solid tumours, including less common cancer types. Indeed, quite good concordance of SNV detection between tissue and cfDNA was observed in this cohort (Figure 4-18 A). All variants detected in tissue were also detected in cfDNA in 100% (4/4), 66.7% (4/6) and 58.3% (7/12) of patients with non-Hodgkin lymphoma, hepatoblastoma and Wilms tumours respectively. Interestingly, no cfDNA unique variants were detected in patients with Wilms tumours, even the high purity ones. Additionally, for eight patients a second cfDNA sample was collected during on-trial relapse and for six patients the same variants were detected in both cfDNA samples. For two patients variants expected from tissue were detected only in one of cfDNA samples.

In this subset of patients, 64.3% (9/14) of alterations described as actionable by molecular tumour board were detected in cfDNA as well. For example, an

actionable *BRAF* V600E mutation was detected in both tissue and cfDNA in a patient with Langerhans cell histiocytosis (SMP0286), and pancreaticoblastoma/acinar cell carcinoma (SMP0340). In patients with non-Hodgkin lymphoma all SNVs and indels expected from tissue sequencing were detected in cfDNA, including 4 actionable variants in two patients. Additionally, for three patients tissue profiling did not reveal any alterations, however cfDNA profiling detected *CTNNB1, ARID1A, MYC* and *PPM1D* variants. As in other cancer types discussed previously, the detection of CNVs was much more challenging with *CDKN2A/B* deletion detected in both tissue and cfDNA in one patient, but no other CNVs expected from the tissue confirmed in cfDNA (Figure 4-18 B).

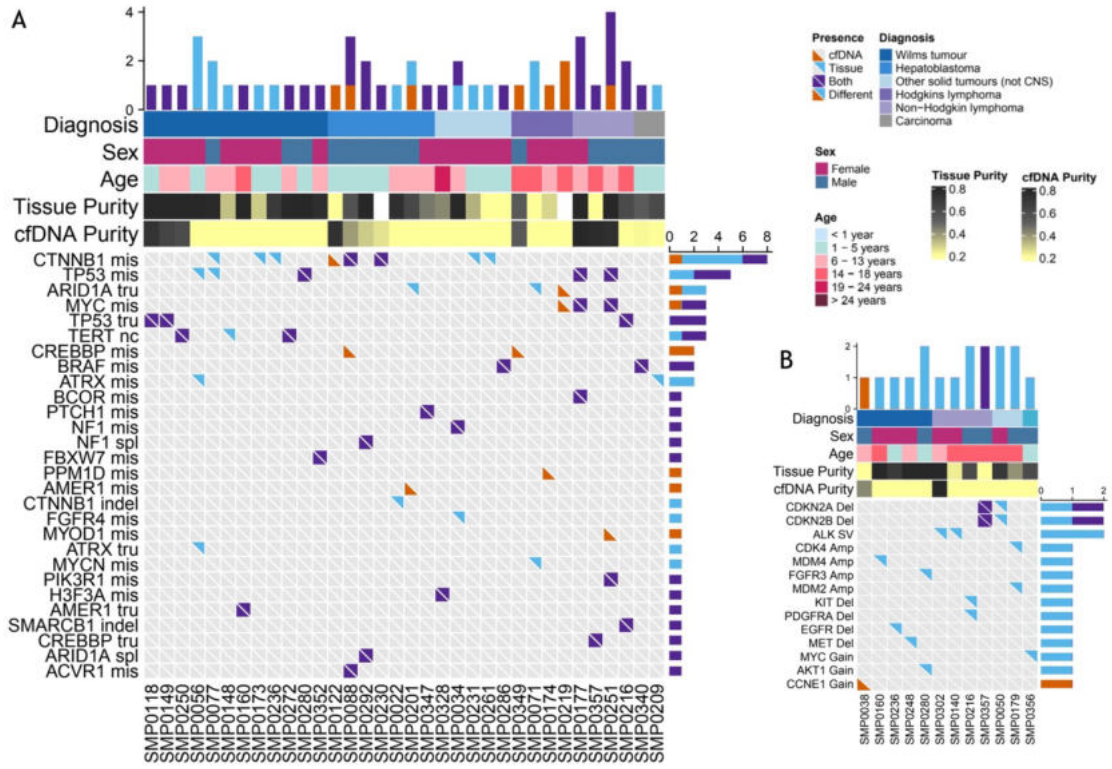


Figure 4-18 Comparison between SNV and CNV detection between tissue and cfDNA in less common cancer types. (A) Comparison of SNV and indel detection between cfDNA and tissue biopsy (only cases with successful tissue and cfDNA analysis with at least one SNV or indel are shown). Each column represents single patient, ordered by ctDNA purity (by lcWGS). The histograms on the top indicate the total number of SNVs and indels in each patient (blue – detected only in tissue, orange – detected only in cfDNA, purple - detected in both); histograms on the right indicate the number of patients with alterations in the indicated gene. (B) Comparison of CNV detection using targeted panel sequencing in cfDNA and tissue biopsy.

4.3.3 Genome wide copy number changes can be detected by lcWGS in cfDNA

Genome wide copy number profiles have been analysed for tissue and cfDNA pairs from 297 patients in the Stratified Medicine Paediatrics study. Good concordance between the two analytes has been observed (Figure 4-19), with an average of 87.1% of bins matching in samples with high purity (n= 63, >10% purity by ASACTlp). While low ctDNA purity was an issue in a high proportion of the patients (for 228 cfDNA samples ctDNA purity was estimated to be <10%), 49 tissue samples were also of low purity. Importantly, in 6 of the patients with no CNVs detected in the tissue sample, cfDNA profiling was successful. For example, in patient with neuroblastoma (SMP0181) poor prognosis markers - 11q deletion and 17q gain - were detected in cfDNA (Figure 4-21). Additionally, lcWGS allows detection of segmented chromosomal changes both in tissue and cfDNA samples in this cohort. Interestingly, highly fragmented copy number profiles were observed in a high fraction of patients with osteosarcoma with the CNV patterns very similar between tissue and cfDNA (Figure 4-19).

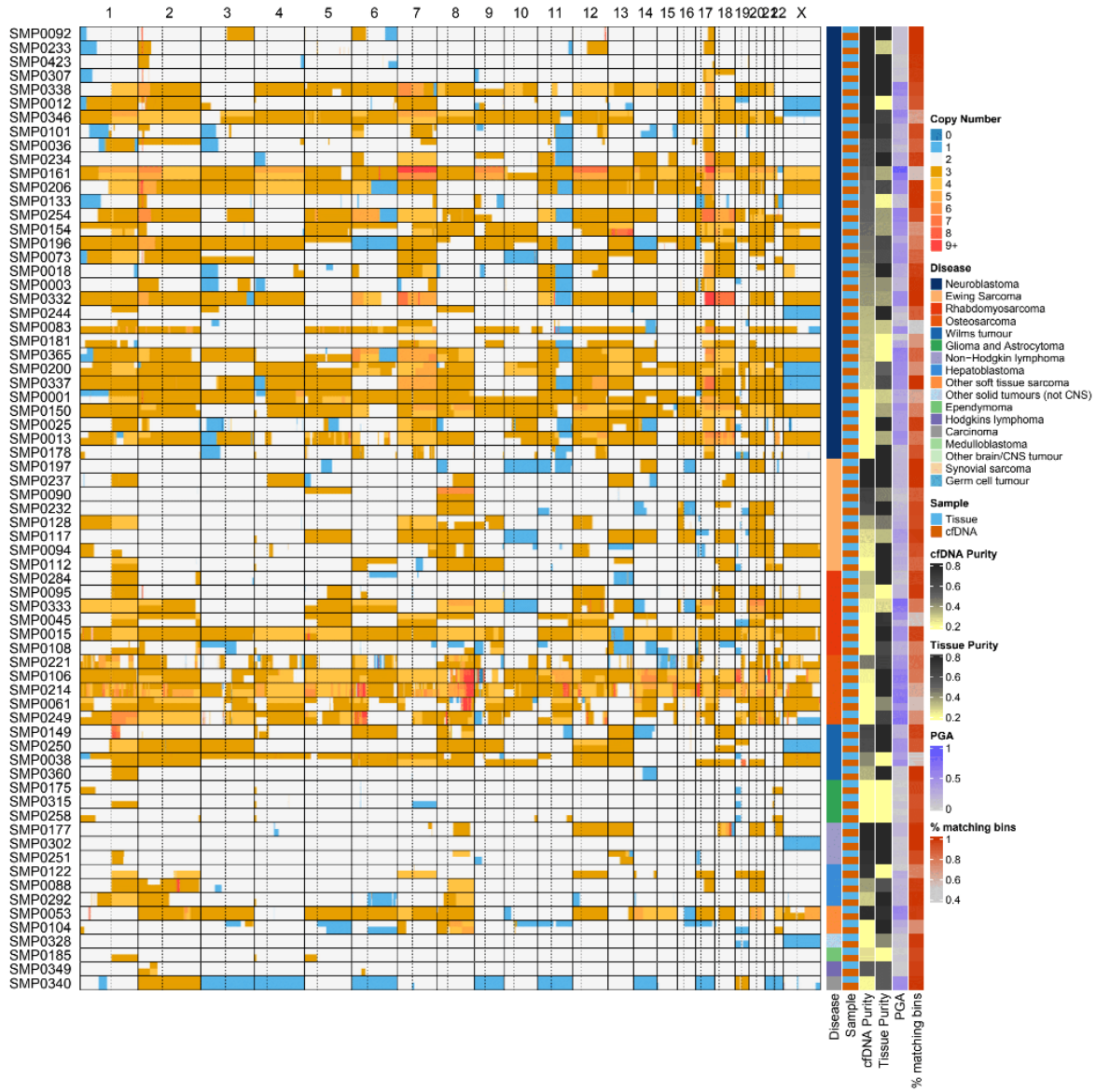


Figure 4-19 Genome wide copy number profiles of tissue and cfDNA in patients with high purity ctDNA and tissue samples (>10% purity). Each patient is represented by two rows – tissue (top) and cfDNA (bottom). Percentage genome altered (PGA) for each sample and Percentage of bins matching between tissue and cfDNA shown in the bar on the right. Bioinformatics analysis and figure by Dr. C. Lynn.

Segmented copy numbers for all samples with detectable tumour DNA (>10% purity) were analysed using GISTIC2.0 [148] and 48 regions were found to have recurrent copy number alterations: 17 amplified and 31 deleted (Figure 4-20). The regions included well known prognostic biomarkers, such as amplification of chr2p24.3, which contains *MYCN* and *ALK*. The detection of these recurrent segments in addition to clinically relevant genes covered by RMH200 gene panel (but not covered for copy number alterations by panel sequencing) between tissue and cfDNA was evaluated (Figure 4-21). When looking at all these alterations, 52% of alterations were detected both in tissue and cfDNA, 26% were detected in tissue

only and 21% was detected in cfDNA only in high purity samples (>10% ctDNA and tissue purity) with no significant difference between the detection of different types of alterations. Clinically relevant copy number changes, such as chromosome 1p, 3p, 4p, 6q and 11q loss and 1q, 2p and 17q gains in patients with neuroblastoma (Figure 4-20 A), 1q gain with 16q loss in patients with Ewing sarcoma (Figure 4-20 B) and 1q gain in patients with Wilms tumour (Figure 4-20 C) have been detected both in tissue and cfDNA.

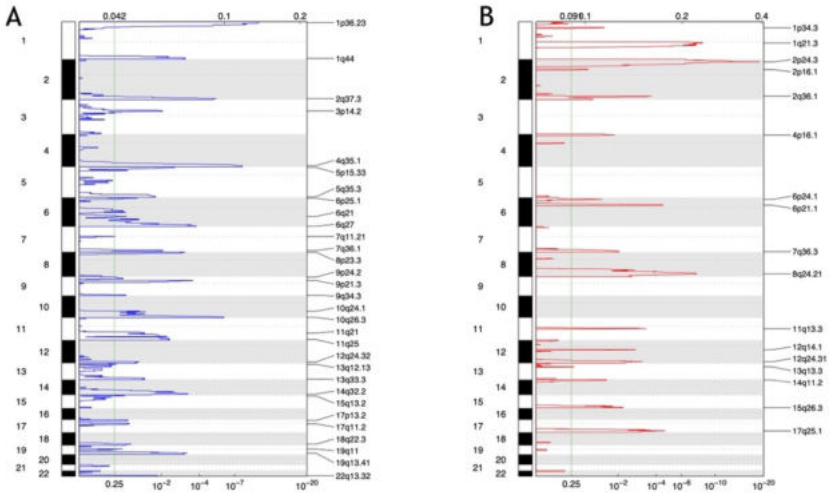


Figure 4-20 The identification of likely driver somatic copy-number alterations by evaluating the frequency and amplitude of observed events in high purity samples using GISTIC algorithm. (A) GISTIC deletion plot. (B) GISTIC amplification plot. The genome is oriented vertically, and GISTIC q-values at each locus are plotted on a log scale. The green line represents the significance threshold (q-value = 0.25). Bioinformatics analysis and figure by Dr. C. Lynn.

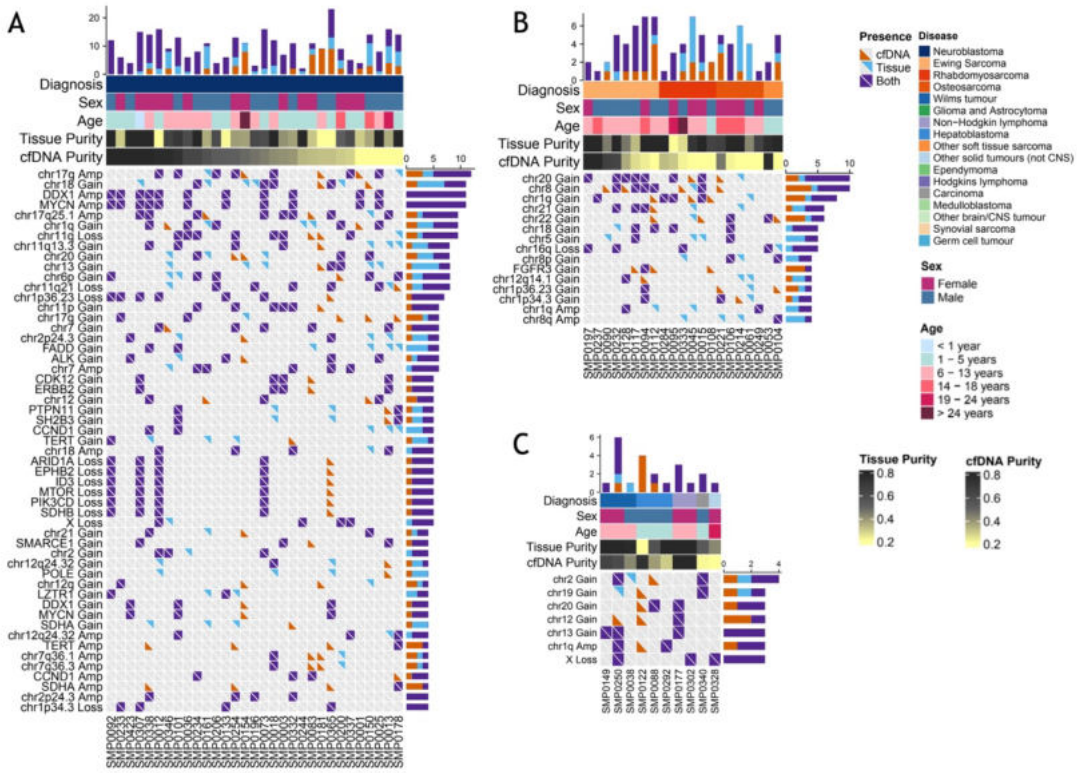


Figure 4-21 Comparison of CNV (chromosome, chromosome arm, significantly enriched cytoband and cancer-related gene deletion and amplification) detection using lcWGS sequencing in cfDNA and tissue biopsy. (A) In patients with neuroblastoma (only cases with >10% ctDNA and tissue purity; CNVs detected in at least 4 patients are shown). The histograms on the top indicate the total number of CNVs in each patient (blue – detected only in tissue, orange – detected only in cfDNA, purple - detected in both); histograms on the right indicate the number of patients with alterations in the indicated gene. (B) Detection of CNVs in patients with sarcoma (only cases with >10% ctDNA and tissue purity; CNVs detected in at least 3 patients are shown). (C) Detection of CNVs in patients with other cancer types (only cases with >10% ctDNA and tissue purity; CNVs detected in at least 3 patients are shown). Bioinformatics analysis and figure by Dr. C. Lynn.

In addition, cfDNA lcWGS profiles were different when compared to the tissue in 15 patients and therefore informative of alternative clones. The most extreme case was patient SMP0083 with neuroblastoma where the cfDNA copy number profile was much more complex, resembling the primary tumour more than the tissue biopsy taken at the primary site of relapse (Figure 4-22 A). A liver metastasis was also biopsied for this patient and the copy number profile of it closely resembled the cfDNA sample, highlighting the benefit of cfDNA profiling in patients with multiple sites. In addition, two SNVs had been detected in cfDNA sample and liver metastasis, but not the tissue sample from the primary site. A similar case of cfDNA being more informative than tissue biopsy was observed in patient SMP0349 with Hodgkin lymphoma (Figure 4-22 B), where cfDNA sample contained more CNVs that were also confirmed to be present at the tissue biopsy

at diagnosis. cfDNA testing might be of particular benefit for patients with Hodgkin lymphoma, as tissue biopsies are often of very low purity [182].

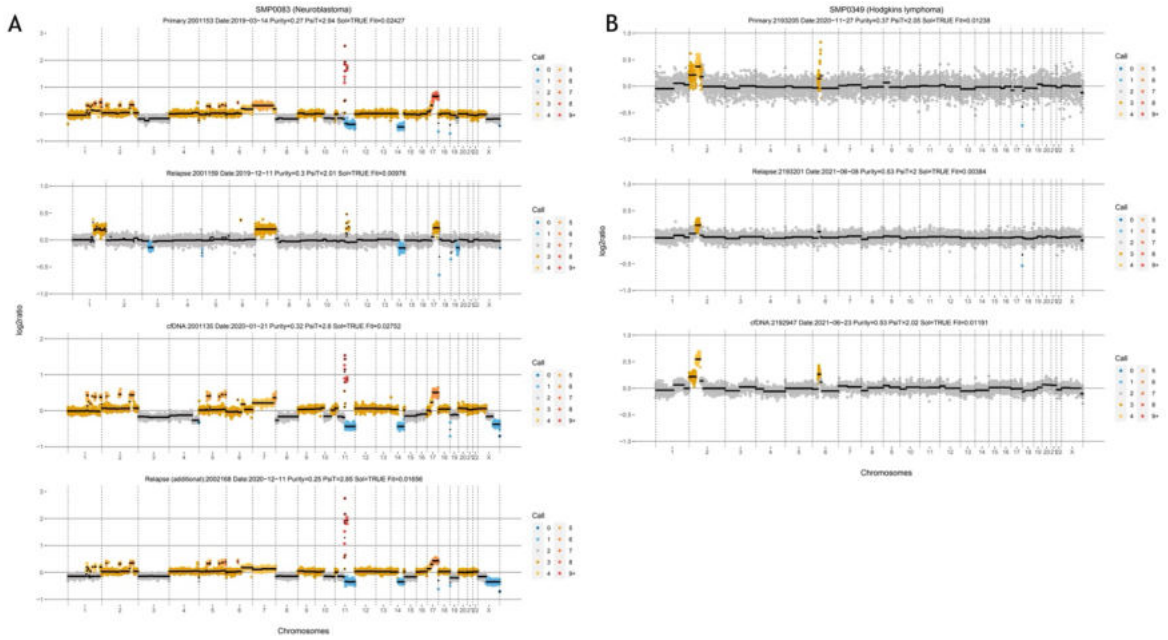


Figure 4-22 Differences between genome wide copy number profiles in multiple cfDNA and tissue biopsy samples. (A) lcWGS copy number profiles of SMP0083 with neuroblastoma. Primary and relapse tissue sample biopsies at primary thoracic site, cfDNA sample at relapse and additional relapse tissue biopsy sample collected from a metastatic lesion in the liver. (B) lcWGS copy number profiles of SMP0340 with Hodgkin lymphoma from primary and relapse tissue biopsies and cfDNA at relapse.

Overall, cfDNA and tissue CNV profiles from lcWGS are very similar in high purity samples with cfDNA showing better representation of tissue heterogeneity, as subclonal variants were missing or created an illusion of clonality in the tissue biopsy in some patients [183,184]. To fully validate this, multi region biopsies of tumour tissue would be ideal, but were not possible in this study.

4.3.4 Genomic changes acquired in paediatric solid tumour at relapse

The benefits of cfDNA profiling were demonstrated by showing good correlation between variant detection in high purity ctDNA samples and tissue and high numbers of cfDNA-unique variants at relapse described in the subchapter above and we wanted to assess the value of this in the context of tumour evolution. For a subset of patients on Stratified Medicine Paediatric Programme, a diagnostic tissue biopsy was available, and we were able to compare the genetic molecular profile of the same patient at diagnosis and relapse. Given that tissue biopsies are

not routinely performed at relapse for these patients and that cfDNA could provide a less invasive and adequate profiling approach we wanted to determine what changes drive relapse and relate these back to the cfDNA profiling, evaluating the role of tumour heterogeneity and evolution and the ability to detect these changes. Additionally, we wanted to evaluate if some of the cfDNA unique variants were present in the primary but not relapsed tissue biopsy.

For a subset of patients in this cohort, diagnostic tissue sample was available for analysis and panel (n=252) and lcWGS (n=169) profiling was performed to explore recurrent genetic changes between primary disease and relapse in matched samples. Small but significant increase in the number of mutations detected at relapse (Wilcoxon test, p = 0.0039) and the overall percentage of genome altered by copy number changes (Wilcoxon test, p = 0.0006) as well as slight but not significant increase in the number of aneuploid tumours at relapse was observed (Figure 4-23). The purity of the samples varied highly between the samples both at diagnosis and at relapse, thus not biasing the results, but reducing the numbers of good quality samples with matching primary and relapse pairs.

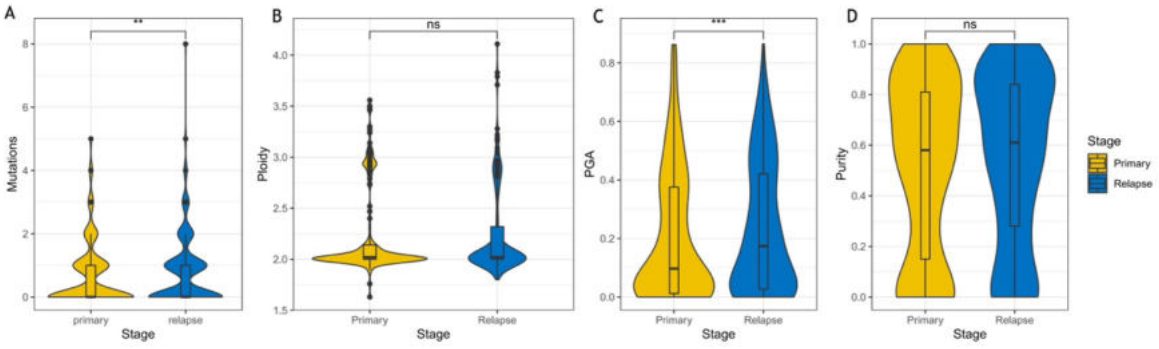


Figure 4-23 Characteristics of the tumours at diagnosis and relapse. (A) The number of mutations detected in the tumour using targeted panel sequencing. ** Wilcoxon test, p=0.0039 (B) The ploidy and (C) percentage of genome altered per sample in primary and relapse tumours as estimated by lcWGS. *** Wilcoxon test, p=0.00055 (D) The purity of samples as determined by lcWGS.

To describe the evolution of the landscape of copy number alterations, genome wide copy number profiles were obtained using lcWGS. In this cohort 99 pairs of primary and relapse tissue were successfully analysed and had copy number changes detected. Out of those, only 39% of the patients had identical profiles throughout the disease while the majority had differences in the copy number profile between primary and relapse sample to a different degree. The well-defined copy number changes such as gains in chromosome 1q, 2, 7, 8, 12 and 17q

and losses on 1p, 3p and 11q were frequently detected both in primary and relapse tumours (Figure 4-24 A, B). No relapse specific copy number signature was observed however, only a higher number of the key alterations, highlighting the universality of these oncogenic changes across different paediatric cancer types. It could be in part caused by the cohort, which is pre-selected for patients that experienced relapse and therefore potentially had more aggressive disease at presentation. Additionally, the diversity of the cohort obscures the changes seen in individual patients between diagnosis and relapse. To account for that relative copy number changes between primary and relapse tumours at individual patient level were analysed: copy number state at relapse was subtracted form copy number value for that genomic region in the matched primary tumour and normalized for tumour ploidy changes. Changes in the copy number state across the genome were observed (Figure 4-24 C), highlighting the universality of these oncogenic changes. To account for disease type specific differences, this analysis was repeated in the most abundant cancer types (Figure 4-24 D). Patients with neuroblastoma most often acquired gains in chromosome 1q, which was also common in patients with sarcomas. Highest levels of segmental aberrations were observed in patients with sarcoma, mostly in osteosarcoma patients. Relatively stable copy number profiles were observed in patients with CNS tumours (Figure 4-24 D).

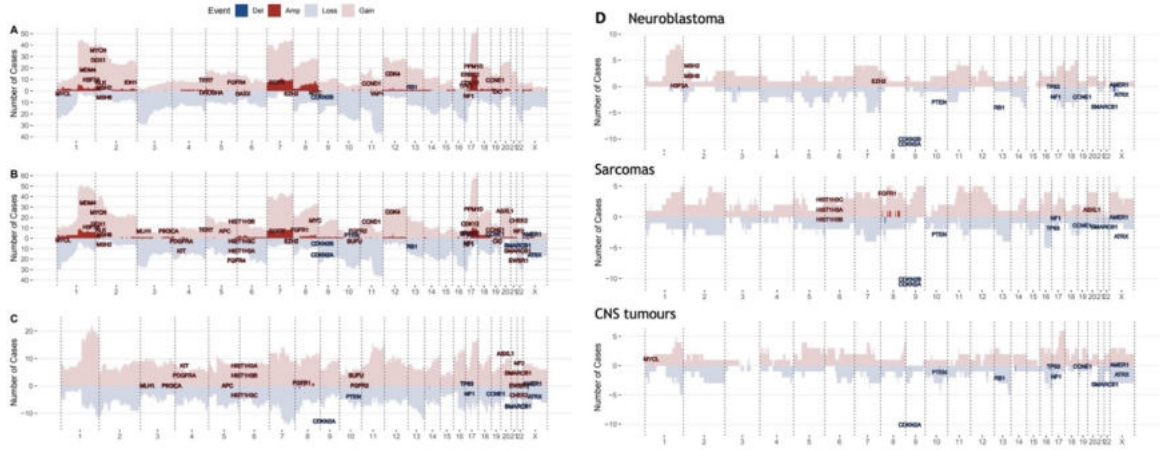


Figure 4-24 Genome wide copy number changes observed in diagnostic and relapse paediatric solid tumours. Histogram of copy number changes observed across the genome in primary tumours (A) and relapse tumours (B) with a number of cases in the y-axis, amplifications (red) and gains (lighter red) shown at the top, deletions and losses at the bottom in the cohort in primary (top) and relapse tumours (bottom). Location of key clinically relevant genes with common focal copy number changes annotated across the genome. (C) Relative differences between copy number changes between primary and relapse tumours in individual patients, normalized for amplification effect and tumour ploidy. Only samples with purity >0.15 included. (D) Relative differences between copy number changes between primary and relapse tumours in individual patients in different

cancer types, normalized for amplification effect and tumour ploidy. Only samples with purity >0.15 included. Bioinformatics analysis and figure by Dr. C. Lynn.

To detect cancer driver genes and measure the levels of selection observed in primary and relapsed paediatric solid tumours we used dN/dS method which calculates the ratio of the number of nonsynonymous substitutions to the number of synonymous substitutions as an indicator of selective pressures acting on protein coding genes [148]. A dN/dS ratio greater than one implies positive selection (the higher the ratio, the higher the selection), below one implies purifying selection and a ratio of one implies neutral selection. In this cohort, positive selection was observed for nonsynonymous and truncating variants in primary tumours and in all types of non-synonymous mutations in relapse tumours, with the strongest selection of nonsynonymous, splice site and truncating variants at relapse (Figure 4-25 A), which is not surprising, as the panel was designed to cover paediatric cancer driver genes. Higher selective pressures on mutations at relapse (cohort wise) and relapse-unique mutations (mutations unique to relapse in each patient) have been observed, indicating additional selection during the disease course as patients undergo treatment. *TP53* mutations, as well as nonsense and splice site mutation of *CDKN2A* were the most significantly positively selected both in primary tumours and to only slightly higher degree at relapse (Figure 4-25 B). In contrast, *ATRX, CHECK2, SETD2, ARID1A/B* and *CREBBP* nonsense and splice site mutations were strongly enriched only at relapse, mostly due to their low prevalence in the diagnostic tumours (Figure 4-25 C, E).

Truncating *TP53, NF1* and *CDKN2A* mutations were significantly positively selected for in both primary tumours and in relapse, highlighting their universal role in tumorigenesis (Figure 4-25 D). Interestingly, truncating *ATRX and NF1* mutations were positively selected for mutations occurring only in relapse. A higher number of missense mutations, including *TP53, CTNNB1, MYC, BRAF* and *PIK3CA* were selected for in primary tumours and relapse with missense *ALK, HRAS* and *PTEN* mutations specific to relapse tumours (Figure 4-25 D). Overall, *TP53* was by far the most commonly mutated gene both in primary and relapse tumours (Figure 4-25 D) with only slightly higher proportion in relapsed samples. However, when comparing patient matched primary and relapse pairs, *TP53* mutations were highly positively selected, indicating their role in driving the relapse (Figure 4-25 D). Different types of potentially inactivating mutations (mainly missense and

nonsense) have been detected in *TP53, NF1, SMARCA4, SETD2, AMER1* had *TSC2*, supporting their role as tumour suppressors in paediatric solid tumours (Figure 4-25 E). *ALK, PIK3CA, BRAF* and *MYC* were the most commonly reported oncogenes in this analysis, showing similar frequency of only missense mutations in primary and relapse tumours, except *ALK*, which was slightly more often reported at relapse (3 cases in primary, 6 in relapse) (Figure 4-25 E). This analysis could potentially be improved by looking into the changes in individual cancer types, however due to low patient numbers in most cancer types the analysis would suffer from low statistical power.

Differences between SNVs and indels detected in primary and relapse pairs have been observed, with accumulation of relapse specific alterations. Missense *ALK* and *TP53* mutations were the most common relapse specific alterations in patients with neuroblastoma (Figure 4-26 A) and missense *TP53* mutations were also the most common relapse specific alterations in patients with various sarcomas (Figure 4-26 B). Interestingly, in several patients different mutations in the same gene were observed between primary and relapse samples, most notably for in two patients with neuroblastoma, different *ALK* mutation was observed. As expected, most clonal mutations were retained between primary and relapse tumour pairs, however primary-specific alterations were detected as well. The loss of variants present at diagnosis indicates the eradication of these clones due to treatment and further exploration if any specific treatment strategies lead to this are upcoming. However, it is possible that a minor fraction of differences is caused by different analysis methods: the variants in tumours at relapse have been interpreted and reported on by two independent clinical scientists, while the variant detection in primary tumours was automated with minimal intervention and interpretation.

Overall, we showed that paediatric solid tumours undergo significant genomic changes from diagnosis to relapse, with accumulation of SNVs, indels and complex genome wide copy number changes. No overarching relapse signature was observed, but generally more complex genomes at relapse driving the disease were reported. Greater selective pressure on mutations at relapse was observed across the cohort with inactivating mutations in tumour suppressor genes such as *TP53* missense and *ATRX* and *NF1* truncating mutations most significantly positively selected for in relapse in these patients.

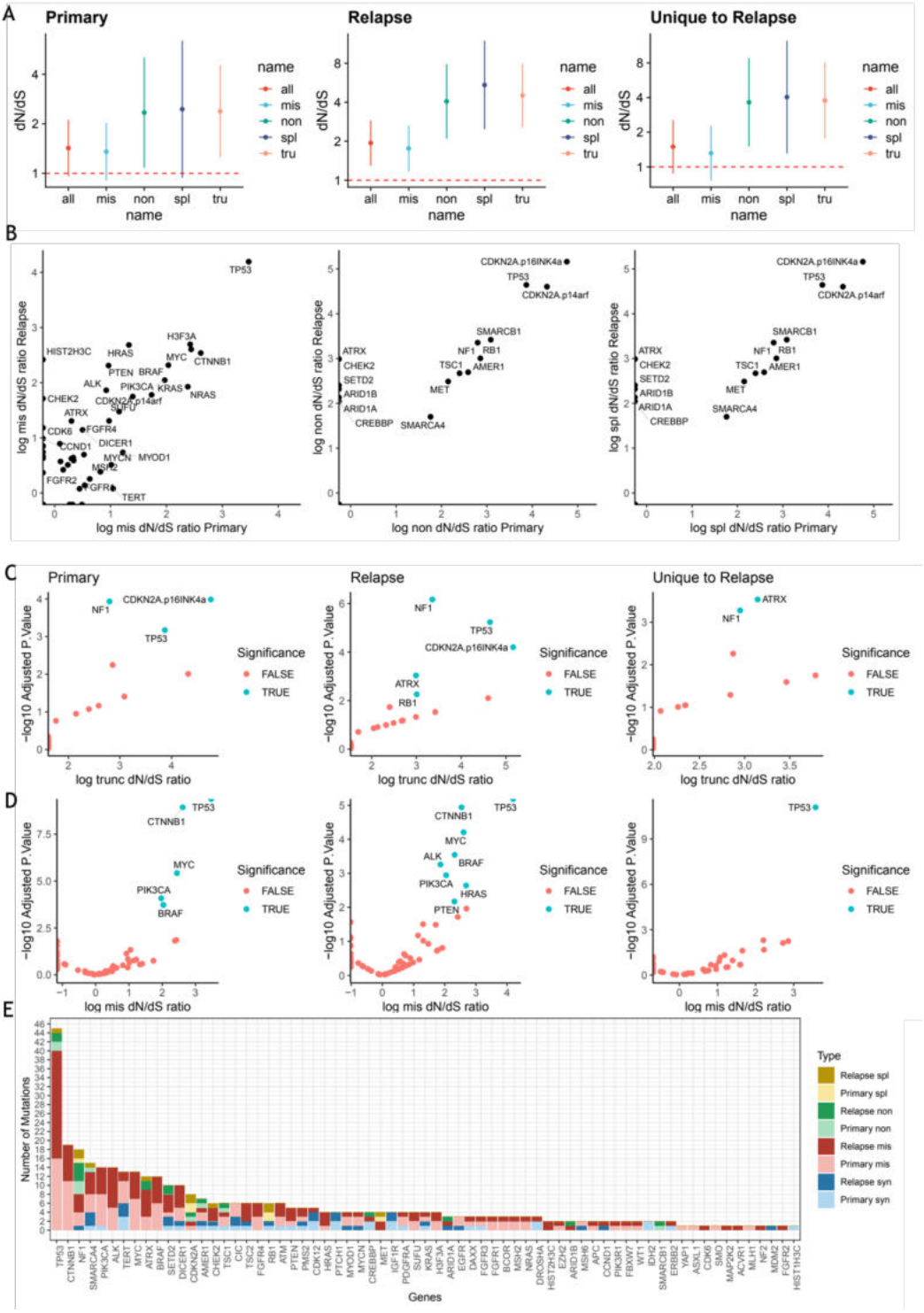


Figure 4-25 Selective pressures in primary and relapsed paediatric tumours and detection of driver genes. (A) dN/dS ratio (the number of nonsynonymous substitutions per non-synonymous site to the number of synonymous substitutions per synonymous site) of different types of mutation in primary, relapsed and unique to relapse (genes detected only in relapse in each patient). A ratio of one, above which positive selection is implied is shown as red dotted line; all – all mutations, mis – missense mutations, non - nonsense mutations, spl – splice site mutations, tru – truncating mutations. (B) Significantly positively selected genes (q value < 0.5 is highlighted in green) for truncating mutations in primary, relapsed and unique to relapse datasets. (C) Significantly positively selected genes (q value < 0.5 is highlighted in green) for missense mutations in primary, relapsed and unique to relapse datasets. (D) Total number of each type of mutation observed across the dataset in primary and relapsed tumours. Bioinformatics analysis and figure by Dr. C. Lynn.

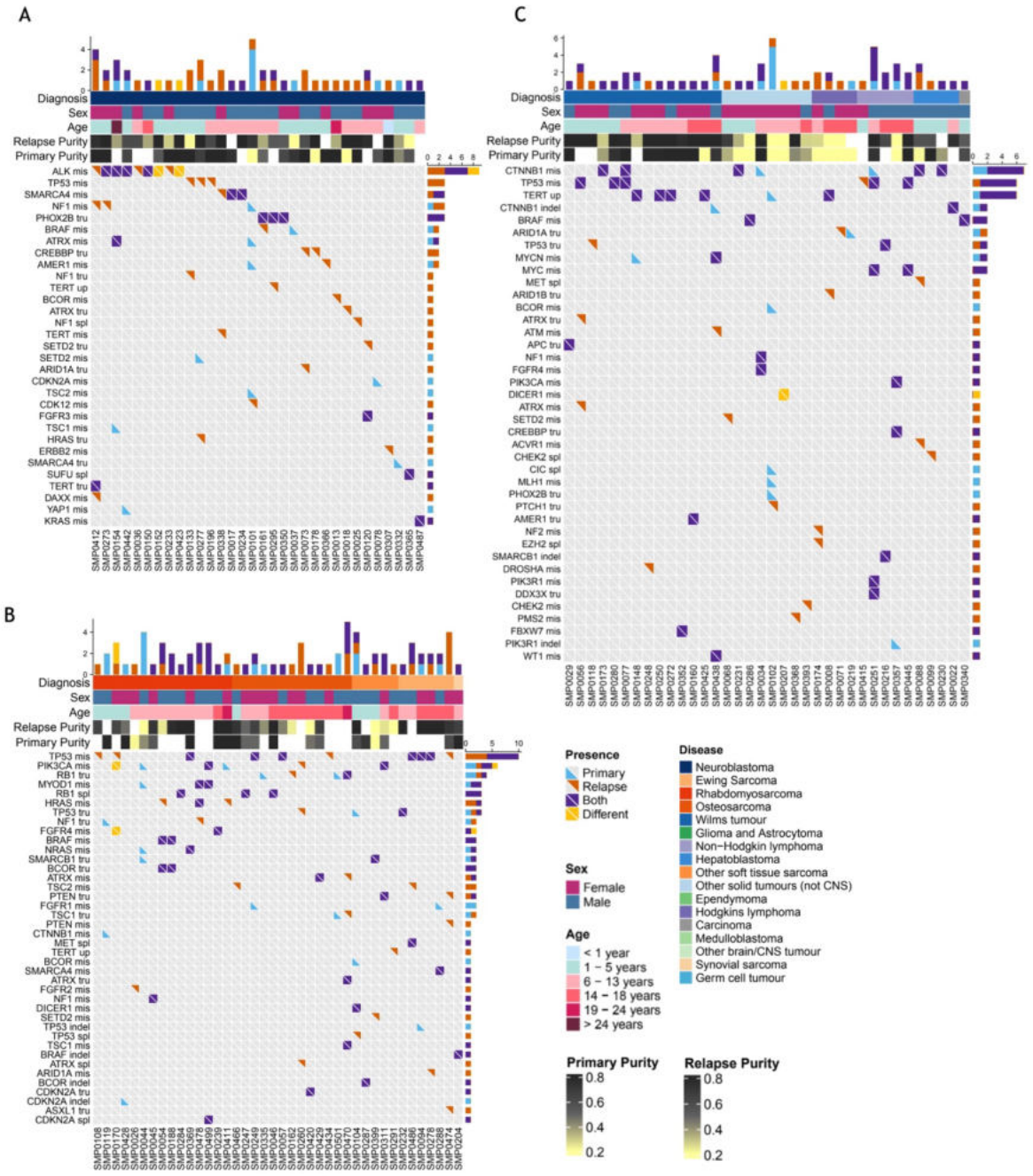


Figure 4-26 Comparison of SNV and indel detection between primary and relapse tissue biopsy in patients with neuroblastoma (A) sarcomas (B) and other extracranial solid tumours (C) (only cases with successful primary and relapse tissue analysis with at least one SNV or indel are shown). The histograms on the top indicate the total number of SNVs and indels in each patient (blue – detected only in primary tissue, orange – detected only in relapse tissue, purple - detected in both, yellow – different variant in the same gene between primary and relapse); histograms on the right indicate the number of patients with alterations in the indicated gene. Bioinformatics analysis and figure by Dr. C. Lynn.

4.4 Discussion

Throughout the studies discussed in this chapter I have shown that cfDNA profiling by NGS capture panel combined with lcWGS is informative for patients with extracranial solid tumours (patients with CNS tumours discussed in detail in Chapter 5) and yields comparable results to tissue biopsy molecular profiling utilizing the same platforms (Figure 4-27). Additionally, high number of cfDNA unique variants detected shows the potential to complement tissue biopsy testing in many clinical diagnostics situations by allowing detection of tumour heterogeneity and identifying variants missed by tissue biopsy profiling, some of which are potentially targetable or aiding enrolment to clinical trials. While panel-based cfDNA sequencing is unlikely to direct treatment at diagnosis, where conventional therapy is mostly effective, it could become of great importance after relapse.

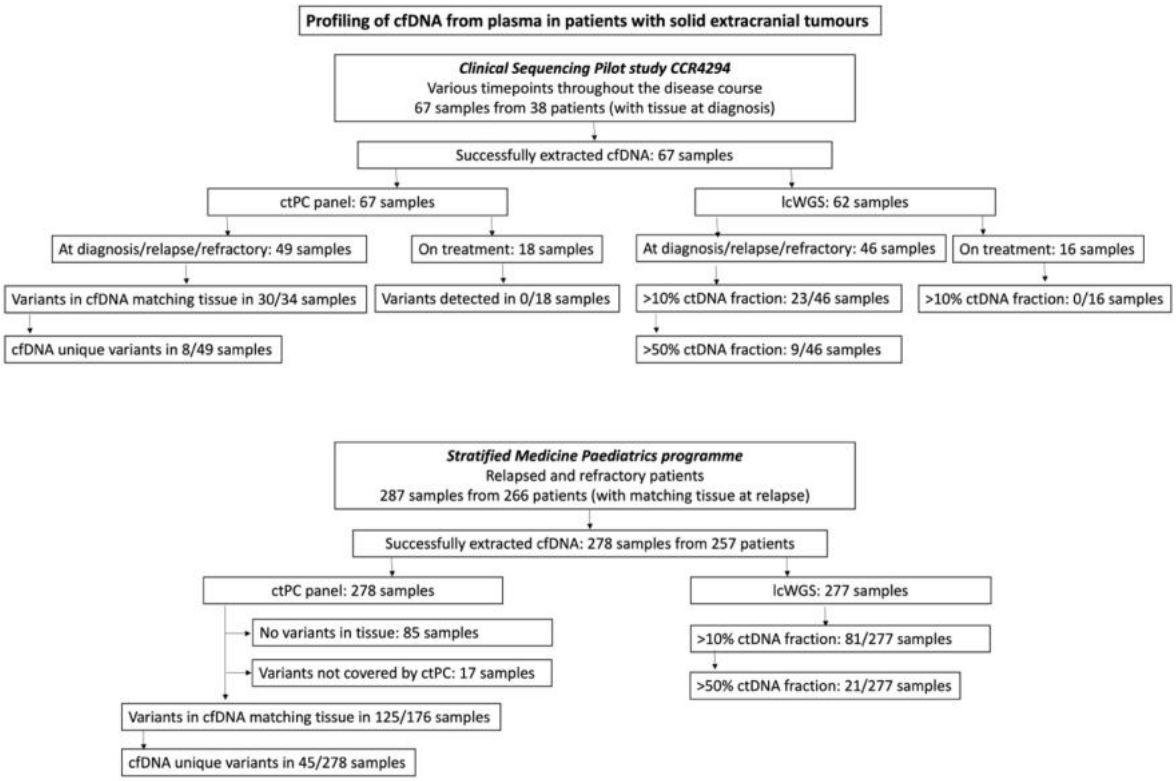


Figure 4-27 Summary flow chart of samples discussed in this chapter. Only patients with solid extracranial tumours included in this flow chart, The number of samples and number of patients included in different studies, highlighting the numbers of successfully processed samples and successful results of analysis in comparison to tissue profiling with comparable method (NGS capture panel sequencing and lcWGS of the same library). ctDNA fraction was evaluated using lcWGS, with 10% as a limit of detection. Due to differences in ctPC and Tissue panel designs, some variants were not mutually covered, as indicated by horizontal arrows.

By comparing diagnostic and relapse tissue biopsies from a subset of patients in the Stratified Medicine Paediatric Programme we showed that paediatric tumours evolve throughout their disease course. Positive selection and accumulation of mutations in *TP53, ATRX and NF1* at relapse were observed. Overall, TP53 was by far the most commonly mutated gene both in primary and relapse tumours, with enrichment in relapsed samples in this cohort, followed by *CTNNB1, NF1, ALK, PIK3CA, SMARCA4, BRAF, MYC* and *ATRX*, which is in agreement with other molecular profiling studies in paediatric solid tumours [15,17,18,42,44,45]. In agreement with previous studies, we showed higher number of mutations in paediatric solid tumours at relapse [14,47,48,50,51]. We did not evaluate TMB, because panel sequencing was used, but higher numbers of variants were reported using the same panel in relapse when compared to diagnostic biopsy from the same patients, in agreement with higher TMB reported at disease recurrence in paediatric patients [14,15].

In agreement to previous reports incorporating comparison of primary and relapse pairs [14,17,42], we observed substantial genetic divergence between primary tumours and relapse in some patients. Notably, in two patients with neuroblastoma (SMP0152 and SMP0423) different pathogenic *ALK* mutations were observed at diagnosis and relapse. Relapse specific mutations in *ALK* (F1174L, R1245V, R1275Q and I1170S) detected in our study have previously been reported to be associated with resistance to ALK inhibitors in neuroblastoma [185–188], even though the patients in this cohort did not receive targeted ALK therapies. In addition to emergence of new variants at relapse we also reported eradication of clones driving primary disease at relapse in some patients, which could be explained by negative selection due to treatment and analysis on which treatments are more prone to this are ongoing. No overarching relapse specific copy number signature was observed in this study, only higher number of the key alterations, highlighting the universality of these oncogenic changes and potentially suggesting that general genomic instability is more important than specific individual changes across different paediatric cancer types, as has been reported for accumulation of segmental alterations in neuroblastoma [189,190]. Good agreement between tissue and cfDNA molecular profiling (given good enough ctDNA purity) was shown by targeted panel and lcWGS analysis. cfDNA unique variants were detected in patients with various solid tumours, potentially showing better representation of tissue heterogeneity, as subclonal variants were missing or created an illusion of

clonality in the tissue biopsy in some patients. Therefore, cfDNA could potentially be used to monitor tumour evolution in a less invasive way.

The main drawback of liquid biopsies is low levels of ctDNA observed in some patients, particularly patients with CNS tumours, which will be discussed in more detail in the next chapter. The levels of ctDNA vary highly between different cancer types and between patients with the same cancer type, and methods to increase sensitivity without the loss of specificity are needed. Utilising biological features of ctDNA, such as selecting specific fragment sizes to enrich ctDNA fraction in vitro or in silico show potential [77,191]. Even though in our hands selection of shorter size fragments did not improve lcWGS profile resolution or detection of low VAF mutations, improved methods to recover single stranded, ultra-short or heavily damaged cfDNA could be explored in the future [191,192]. Combining genomic and fragmentomic analysis could aid detection of ctDNA especially in tumours with low mutation burden, as shown in paediatric Ewing and other sarcomas [160,193]. Novel cfDNA WGS methods with better error correction [194] or phased variant enrichment [195] are showing promise as well.

The biggest study described in this chapter was the Stratified Medicine Paediatric Programme, evaluating detection of SNVs, indels and genome wide copy number changes in paediatric cancer patients with solid tumours at relapse. In this study a total of 313 SNVs or indels were detected by panel sequencing: 31% of alterations were detected in both tissue and cfDNA, 41% detected only in tissue and 26% detected only in cfDNA. The vast majority of tissue specific variants were not detected in cfDNA due to low ctDNA purity, mainly in CNS tumours. The recently published European MAPPYACTS study, exploring cfDNA analysis to detected actionable alterations in paediatric and adolescent non-CNS tumours using WES instead of panel-based analysis showed higher levels of overlapping and lower levels of cfDNA unique variants (57% of SNVs were detected in both cfDNA and tumour, 31% and 11% were specific to the tumour and the cfDNA, respectively) [15]. Even if only non-CNS tumours were considered in our study, 29.4% of alterations were detected in both tissue and cfDNA, 38.5% were detected only in tissue and 31.5% were detected only in cfDNA. The differences might be influenced by different proportions of various cancer types included in the studies - a much higher proportion of cfDNA samples in MAPPYACTS study were from patients with sarcomas (65% of all patients in MAPPYACTS, compared to 29% in Stratified

Medicine Paediatric Programme), who had relatively good concordance between tissue and cfDNA and low number of cfDNA unique variants in both studies. Also, ctPC panel sensitivity is much higher than WES (0.125% vs 1% VAF), allowing detection of lower AF variants and hence potentially higher number of cfDNA unique variants. Although recent studies are showing the importance of tumour heterogeneity, the relevance of subclonal drivers to impact clinical outcomes remains to be explored [196,197].

In agreement to MAPPYACTS [15] and previous studies [90], we showed that ctDNA fraction in cfDNA is highest in patients with neuroblastoma when compared to other cancer types. In contrast to these studies, total cfDNA quantities did not show significant differences between cancer types in our cohort [15,90]. Variants expected form tissue sequencing were successfully detected in cfDNA in the majority of patients with neuroblastoma at relapse with at least one SNV expected from tissue profiling detected in cfDNA in 83.9% of patients and cfDNA unique variants were detected in 54% of patients. This is in agreement with the MAPPYACTS study where most SNVs expected form the tissue were detected in cfDNA in addition to several cfDNA unique variants [15]. The potential clinical value of cfDNA analysis was recently demonstrated high-risk neuroblastoma, where 68% of variants detected in cfDNA were not detected in the tissue and 50% of cfDNA-unique variants were potentially clinically targetable [187]. In agreement to the results described in this chapter, missense variants in *ALK* and *TP53* were the most common alterations found in ctDNA in patients in neuroblastoma [187]. Genetic alterations in *ALK* in high-risk patients are independent predictors of poor survival [26] and given good detection rates in cfDNA, the less invasive molecular profiling could be of great benefit in this patient population, potentially identifying patients for targeted ALK inhibitor treatment.

The presence of cfDNA unique variants has been reported in all adult and paediatric cfDNA studies [15,99,115,198] and potentially indicates the ability of cfDNA profiling to better represent tumour heterogeneity by detecting different (sub)clones present in the tumour. To fully validate the source of cfDNA unique alterations, multi region biopsies of tumour tissue would be ideal, but were not possible in our studies and are seldom achievable in real life practice. In several patients where primary site at diagnosis or at relapse and metastatic site was biopsied, cfDNA was able to represent a mix of both sites, as highlighted in

patients SMP0083, SMP0365 and SMP0154 discussed in detail. In our study, cfDNA-unique variants were detected in most cancer types, with the highest number of cfDNA unique variants in patients with neuroblastoma. The average VAF of cfDNA-unique variants was significantly lower than VAF of variants overlapping between tissue and cfDNA, even though higher threshold for reporting cfDNA-unique variants was applied, supporting the hypothesis of high presence of subclonal variants in cfDNA. The lower VAF of cfDNA unique variant in patients with neuroblastoma was also shown in recent study with median AF 20.4 vs. 0.7%, for overlapping and cfDNA unique variants, which is in good agreement to median AF of 19.7% and 3.5% observed in our study for patients with neuroblastoma [187]. The most commonly alerted gene in cfDNA was *TP53*, which was the most commonly mutated gene overall in the cohort and in similar molecular profiling studies in different paediatric cancers [15,17,18,42,44,45], adding to the evidence of tumour-origin of cfDNA unique variants. Importantly, for 6 patients in this study cfDNA-unique variants were present in the time-matched tissue, but below the limit of detection for tissue panel sequencing, showing added value of cfDNA profiling. Similarly, to the recent studies [15,187], potentially actionable alterations were detected in cfDNA only as well, such as pathogenic *ALK* mutations in neuroblastoma. However, the clinical value of detecting subclonal pathogenic, potentially actionable alterations is still not clear. For example, recent study of high-risk neuroblastoma reported that poorer overall survival was observed in patients with clonal *ALK* mutations, as opposed to subclonal based on tissue profiling [26]. The effectiveness of ALK inhibitors in the two subgroups remains to be evaluated as well as the potential to base treatment decisions on cfDNA unique variants.

A number of recent national and international studies have evaluated the ability to detected actionable alterations in paediatric cancer patients – The Zero Childhood Cancer Program [45], Genomes for kids [18], GAIN/iCat2 study [44], Paediatric MATCH trial [199], INFORM [42], iTHER [200] and MAPPYACTS [15] – showing high levels of potentially actionable alterations detectable by combination of DNA and RNA sequencing and methylation profiling. The Stratified Medicine Paediatric Programme also included multi-platform analysis in addition to targeted panel sequencing and lcWGS sequencing discussed in this thesis, such as RNAseq, WES and methylation profiling for CNS tumours. More information is coming that will enhance the evaluation of potentially actionable alterations in these patients.

However, the ability to detect SNVs by targeted sequencing in cfDNA was demonstrated with 66.7% of alteration deemed actionable in tissue by molecular tumour board also detected in cfDNA. This is in good agreement to MAPPYACTS study which also evaluated the potential of cfDNA to detected actionable alterations and showed that 76% of SNVs observed in the tumour could be identified in plasma cfDNA by WES [15]. Using lcWGS clinically relevant, poor prognosis markers such as loss of chromosomes 1p, 3p, 4p, 6q, and 11q and gains of chromosomes 1q, 2p, and 17q in patients with neuroblastoma [201], co-occurrence of chr1q gain and 16q loss in Ewing sarcoma [202], gain of 1q in Wilms tumours [203], have been successfully detected in cfDNA in this study. The main challenge now is better design of clinical trials to allow access to the targeted treatments for the patients with targets identified, as currently only a small fraction of alterations deemed actionable in these studies lead to change in treatment (4-32% of patients in the recent studies) [15,18,44,45,199,200,204].

Chapter 5 Liquid biopsy for paediatric cancer patients with CNS tumours

Parts of this chapter were published as part of [149]

Parts of this chapter were published as part of [205]

5.1 Introduction

Tumours of the central nervous system (CNS), are the second most common diagnostic group in paediatric oncology and the most common solid cancer group, making up 23% of all paediatric cancer cases [5]. The need for liquid biopsies is especially high in these patients because tissue biopsy carries high risk of complications and is not always possible [206]. The ability of neuroimaging to discriminate different brain tumour diagnoses is low, and current practice, as per World Health Organisation (WHO) criteria mandates both histopathological classification and advanced molecular characterisation, which is now considered standard of care [207]. A combination of magnetic resonance imaging (MRI) (sometimes with functional imaging) and clinical examination is used for diagnosis and to serially assess disease response to therapy, but these have limited sensitivity and specificity, despite international guidelines and consensus [208,209] and there remains a lack of consistency for defining tumour measurement and response for some tumour types [210,211]. Additionally, for certain CNS tumours, anatomical location precludes diagnostic biopsy, further highlighting the need for non-invasive molecular diagnostics [212]. Also, standard clinical imaging techniques do not facilitate the assessment of molecular changes during therapy or at relapse, which in turn limits treatment options available to these patients.

There is a growing number of genetic, epigenetic and protein expression biomarkers that can be evaluated in different paediatric brain tumour types using liquid biopsy tools on cfDNA, circulating tumour cells, miRNA and extracellular vesicles [213,214]. In this project I sought to explore the use of cfDNA molecular profiling, especially genomic molecular profiling identifying single nucleotide and copy number changes, in paediatric patients with a range of CNS tumours. In most

cancers, blood is a good source of circulating tumour DNA and can be used to obtain a molecular profile of the tumour in a minimally invasive way. However, in patients with CNS tumours the ability to detect ctDNA in the blood seems to be more limited compared to other solid tumours [212,215]. In the paediatric setting, most studies had been limited by small patient numbers and even though they were focused on using highly sensitive methods (mainly ddPCR) that allow detection of variants at very low levels (VAF of 0.01%–0.1%), the ability to detect pathogenic variants in cfDNA from plasma of patients with brain tumours has been low, as highlighted in the recent reviews [212,216,217].

The limited sensitivity of blood-based assays is mainly thought to be due to the blood brain barrier (BBB), that restricts the shedding of ctDNA into the bloodstream. This idea is supported by relatively higher ctDNA levels in high grade glioma (HGG), which is characterised by disrupted BBB, and meningioma which grows outside of BBB [218,219]. For example, a recent study showed a reasonable sensitivity of 62% (with 90% specificity) for detection of *TERT* promoter mutations in adult glioma patients, as well as the potential to track the disease course with serial blood samples in 5 patients [220]. Another study of adults with primary brain tumours at various clinical timepoints showed that half of the patients had detectable ctDNA alterations, albeit with average VAF of 0.33% and minimum VAF of 0.05%, highlighting the need for assays with very high sensitivity and specificity [219]. Additionally, some CNS tumour types, such as medulloblastoma or ependymoma, can sometimes metastasise outside of the CNS. In these cases, blood based cfDNA profiling may be useful, even before the metastasis occurs, as shown by the presence of circulating tumour cells (CTCs) and ctDNA in several therapy-naïve paediatric medulloblastoma patients [221]. However, this was not confirmed in a large longitudinal study (127 plasma samples from 41 patients) of paediatric non-brainstem HGG where no alterations were detected in any of the patients even though a highly sensitive ddPCR method was used on cfDNA from the blood [152]. This might have been in part caused by low blood volumes (mean of 0.5ml) collected in this study, which an inherent challenge in studies of paediatric cancer patients. The largest prospective study of cfDNA from children with CNS tumours, using both ultra-low pass WGS (ULP-WGS) and hybrid capture sequencing with UMIs has further highlighted that detection of ctDNA is limited by low ctDNA fraction and the low numbers of genetic events in these tumours [105].

The low sensitivity and inconsistency of cfDNA profiling from plasma in adult patients with brain tumours led researchers to investigate alternative body fluids as sources of ctDNA. A variety of different cfDNA profiling methods have been used with studies showing better detection rates in cfDNA derived from CSF (CSF-cfDNA) than from plasma in a range of CNS tumours, mostly High-Grade Gliomas and medulloblastomas [105,222–227]. Studies to date have generally been limited by small patient numbers and restricted availability of fully matched samples, with CSF and plasma often derived from different patients or from different time points, hindering statistical analysis. However, a recent study conducted a large prospective analysis of cfDNA obtained from plasma, CSF and urine in 564 specimens from 258 patients with paediatric brain tumours. This study showed best detection potential in CSF but highlighted low detection rates. Ultra low pass WGS detected copy number alterations in 20% of CSF, 1.3% of plasma, and 0% of urine samples, and deep capture panel sequencing detected alterations in 30% of CSF, 2.7% of plasma, and 0% of urine samples) with high-grade tumours showing the best detection for ctDNA in CSF and plasma [105]. Nevertheless, an overall trend for higher levels of ctDNA to be present in CSF than in plasma in patients with CNS tumours is emerging [215,228].

In the paediatric cancer literature, evidence is accumulating that profiling of CSF-cfDNA could be a feasible and efficient tool for the diagnosis and monitoring of paediatric diffuse midline gliomas and medulloblastomas [229,230] . Importantly, in paediatric diffuse midline gliomas, CSF-cfDNA profiling highlighted the possibility of detecting pathogenic variants and aiding the inclusion of patients into clinical trials that rely on H3K27M status as a stratification biomarker when a biopsy is not feasible [230]. In paediatric medulloblastoma, the largest study so far demonstrated the clinical utility of copy number variant (CNV) detection through low coverage WGS (lcWGS) on CSF-cfDNA and described it as a minimal residual disease (MRD) surrogate marker. The MRD tended to decline with treatment and persistent MRD correlated with a higher risk of relapse. Notably, MRD detection using CSF-cfDNA preceded radiographic progression in half of the patients who relapsed [231].

The potential to use cfDNA to identify the more aggressive subclones driving disease progression has been shown in cases where CSF-cfDNA at baseline was

more concordant with the relapsed tumour than with the corresponding primary tumour [231]. In addition, the ability of cfDNA to characterise intra-tumoural heterogeneity was shown in paediatric medulloblastoma patients where VAF in tissue and CSF-cfDNA had good correlation, indicating that CSF-cfDNA allows for the detection of small subclones present in the tumour) [227] and in paediatric brainstem glioma patients where CSF-cfDNA profiling detected variants not present in the primary tissue sequencing [232].

Despite encouraging signs of utility, it is important to acknowledge that CSF is not as easily obtainable as blood, and CSF sample processing is far from standardised. Inconsistent collection and suboptimal processing can lead to poor sample quality. For example, a recent study of paediatric medulloblastoma that failed to detect most of the mutations in CSF-cfDNA that were expected from the tissue sequencing, showed that only 15 out of 58 samples had detectable ctDNA (by fragment size analysis) [233]. The ability to compare the different studies in a meaningful way and issue guidelines is further hindered by the variability in time of collection, pre-extraction handling and in the collection methods used (CSF from lumbar puncture, ventricular shunt, external ventricular drain, and various CSF reservoirs have been tested). However, this is not possible to control for as it is dependent on the specific clinical situation.

Overall, the detection of ctDNA and clinically informative variants in the blood is only possible sporadically in patients with CNS tumours, with some tumours being potentially more suitable for blood based liquid biopsies than others. CSF is emerging as a more attractive source for cfDNA profiling, but it is more invasive, and it's collection and storage still needs to be standardised. In this chapter I will outline the results on the feasibility and clinical utility of cfDNA molecular profiling from plasma in patients with CNS tumours using targeted NGS panel. Additionally, I will evaluate the ability of our method to detect variants in CSF-cfDNA.

5.2 Blood based cfDNA profiling is of limited success in paediatric patients with CNS tumours

5.2.1 Early studies of blood based cfDNA profiling for patients with CNS tumours

In the pilot study, four patients with brain tumours (including ependymoma, astrocytoma and xanthoastrocytoma) had SNVs or CNVs detected in the tissue by targeted panel sequencing, but none of the cfDNA samples recapitulated these results (Appendix Table-2). Two of these patients however had the blood samples collected while on treatment, possibly influencing the results. Additionally, as part of a collaboration with Glioma Team led by Prof. C. Jones I have attempted to molecularly profile 29 blood-derived cfDNA samples from 13 patients with diffuse intrinsic pontine glioma (DIPG), glioblastoma multiforme (GBM), diffuse midline glioma (DMG) and high grade glioma (HGG), enrolled to various clinical trials or treated with standard therapy. None of the samples had any alterations confidently detected in plasma cfDNA (Table 5-1). It is important to highlight, that cfDNA yields were variable with the mean of 90.0ng of total cfDNA, ranging from 14.7ng to 1050ng (the ng/ml of plasma yield could not be calculated as the volume of plasma was not known for samples extracted in a different laboratory) but the ctDNA purity was low (<10% by lcWGS) in all the samples analysed. One of these patients had cyst fluid collected in addition to the blood sample and the pathogenic variants were detected at high VAF in cfDNA extracted from the cyst fluid (Table 5-1), indicating that molecular profiling of cfDNA is possible, but probably not from blood.

Table 5-1 cfDNA profiling of patients with CNS tumours on various trials. Disease, time point of blood sample collection in respect to clinical course, total depth of sequencing achieved for the cfDNA sample (UMIx) and molecular profiling results in tissue and cfDNA reported. DIPG - diffuse intrinsic pontine glioma, GMB - Glioblastoma multiforme, DMG - Diffuse Midline Glioma, HGG – High grade glioma, SOC – standard of care.

Sample ID	Trial ID	Trial	Disease	Timepoint	Depth (UMIx)	Variants known in tissue	Variants detected in cfDNA
16/02500	005	CCR 4294	Ependymoma of brain	6th relapse/progression	1292	PIK3CA, ATM and TP53 mutations	No variants detected
17/12648	132	CCR 4294	Posterior fossa astrocytoma	Post treatment, residual disease	562	ATRX mutation	No variants detected
16/10465	079	CCR 4294	Xanthoastrocytoma	After 2 cycles of chemotherapy	1657	ATM deletion	No variants detected
17/13083	115	CCR 4294	Xanthoastrocytoma	On treatment	812	BRAF mutation and CDKN2A mutation	No variants detected
20/01485	RMH_GT_001	BIOMEDE	DIPG	Taken at biopsy (treatment naïve)	947	HIST1H3B, ACVR1 and BCOR mutations	No variants detected
20/01486	RMH_GT_001	BIOMEDE	DIPG	pre-cycle 1	2257		No variants detected
20/01488	RMH_GT_001	BIOMEDE	DIPG	pre-cycle 2	1366		No variants detected
20/01490	RMH_GT_001	BIOMEDE	DIPG	pre-cycle 3	2398		No variants detected
20/05958	RMH_GT_001	BIOMEDE	DIPG	pre-cycle 4	1392		No variants detected
20/01491	RMH_GT_001	BIOMEDE	DIPG	pre-cycle 5	1943		No variants detected
20/02774	RMH_GT_001	BIOMEDE	DIPG	pre-cycle 6	2191		No variants detected
20/05959	RMH_GT_001	BIOMEDE	DIPG	pre-cycle 7	1233		No variants detected
20/05973	RMH_GT_001	BIOMEDE	DIPG	End of treatment	1125		No variants detected
21/92220	RMH_GT_002	SOC	GBM IDH WT (Hypermutator)	Pre cycle 1	1041	Hypermutator >1700 variants including TP53, MLH1 and MSH6	No variants detected
20/05960	RMH_GT_004	SOC	DIPG, DMGK27	NA	1064	HIST1H3B, PIK3CA and TP53 mutations	No variants detected
21/92222	RMH_GT_004	SOC	DIPG, DMGK27	pre cycle 4	1358		No variants detected
21/92223	RMH_GT_004	SOC	DIPG, DMGK27	post cycle 7	723		No variants detected
21/92224	RMH_GT_004	SOC	DIPG, DMGK27	cycle 12	1221		No variants detected
21/92225	RMH_GT_004	SOC	DIPG, DMGK27	Relapse	1347		No variants detected

20/06621	CXJ020_PLM_001	Larotrectinib SCOUT	spinal HGG	Post surgery	787	KCTD16:NTRK2 fusion	No variants detected
20/06622	CXJ020_PLM_002	Larotrectinib SCOUT	spinal HGG	On treatment	482		No variants detected
20/06623	CXJ020_PLM_003	Larotrectinib SCOUT	spinal HGG	On treatment	406		No variants detected
20/06624	CXJ020_PLM_004	Larotrectinib SCOUT	spinal HGG	On treatment	1086		No variants detected
20/06625	CXJ020_PLM_005	Larotrectinib SCOUT	spinal HGG	On treatment	953		No variants detected
20/06626	CXJ020_PLM_006	Larotrectinib SCOUT	spinal HGG	On treatment	1201		No variants detected
21/92258	CXJ042	SOC	DMG H3K27M	at the time of surgery	1452	HIST1H3B K27M mutation	No variants detected
21/92259	CXJ046	SOC	GBM	at the time of surgery	1285	No tissue information	No variants detected
21/92260	CXJ044	SOC	DIPG	at the time of surgery	1526	No tissue information	No variants detected
21/92262	CXJ051	SOC	DMG_H3K27M	at the time of surgery	1042	EGFR, PIK3CA and BCOR mutations	No variants detected
21/92263	CXJ052	SOC	GBM IDHmut	at the time of surgery	2174	IDH mutation	No variants detected
21/92264	CXJ057	SOC	HGG	at the time of surgery	1830	No tissue information	No variants detected
20/01493	HSJD_DIPG_025	SOC	DIPG	pre-cycle 1	2727	No tissue information	No variants detected
20/01495	DCEZ_DIPG_001	SOC	DIPG	NA	1774	ACVR1, HIST1H3B K27M, TP53 mutations	HIST at the limit of detection (VAF od 0.02%, 4 reads)
20/01498	DCEZ_DIPG_001	SOC	DIPG	FROM CYST FLUID	2205	ACVR1, HIST1H3B K27M, TP53 mutations	ACVR1 at 31%, TP53 at 34% and 29%, HIST1H3B at 32%

5.2.2 Detection of ctDNA in plasma from patients with CNS tumours at relapse

The largest cohort of patients with CNS tumours was part of The Stratified Medicine Paediatrics programme, which included 95 paediatric cancer patients with CNS tumours at relapse. In this cohort, an average of 6.1ml of plasma was collected per patient, resulting in the mean total cfDNA yield of 101ng per sample (Figure 5-1 A B). The mean cfDNA yield was 16ng/ml, with no significant differences (one way ANOVA test) between different tumour types (Figure 5-1 C D). Six samples with high HMW contamination were excluded from summary plots, but were included in further discussion, as size selection followed by successful panel sequencing was performed on these samples. The outlier sample with 616ng total yield had only minimal HMW DNA contamination (Figure 5-1 E). It is important to highlight however, that only for the high yield samples fragment size analysis was performed and there remains a possibility that some of the lower yield samples also had some HMW contamination.

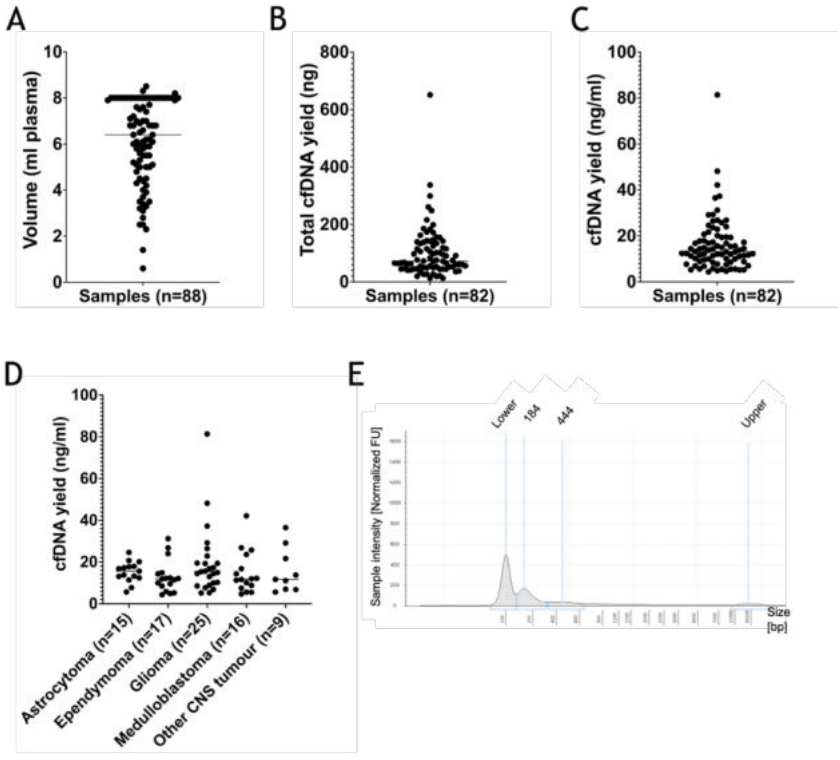


Figure 5-1 Sample characteristics from the patients with CNS tumours in Stratified Medicine Paediatrics programme. (A) Volume of plasma collected per patient (B) Total cfDNA yield (ng) per patient (C) cfDNA yield per ml of plasma (D) cfDNA yield per ml of plasma in different CNS tumour groups, one way ANOVA – no significant differences between the groups E. Fragment size analysis for high yield sample, showing very minimal HMW contamination.

In this study, 28 patients had SNVs detected in the tissue by panel sequencing. When ctPC panel was run on the time matched plasma cfDNA samples from these patients, only 6 patients had SNVs detected in the cfDNA that were expected from the tissue sequencing (Figure 5-2 A). In several cases, only the highest VAF variant was detected, and the lower ones were not. This can be explained by low ctDNA fraction in these samples - all the samples had low ctDNA purity (<10% by lcWGS and the highest VAF observed in the panel sequencing was 2%). This highlights the difficulty of molecular profiling from blood cfDNA for patients with CNS tumours.

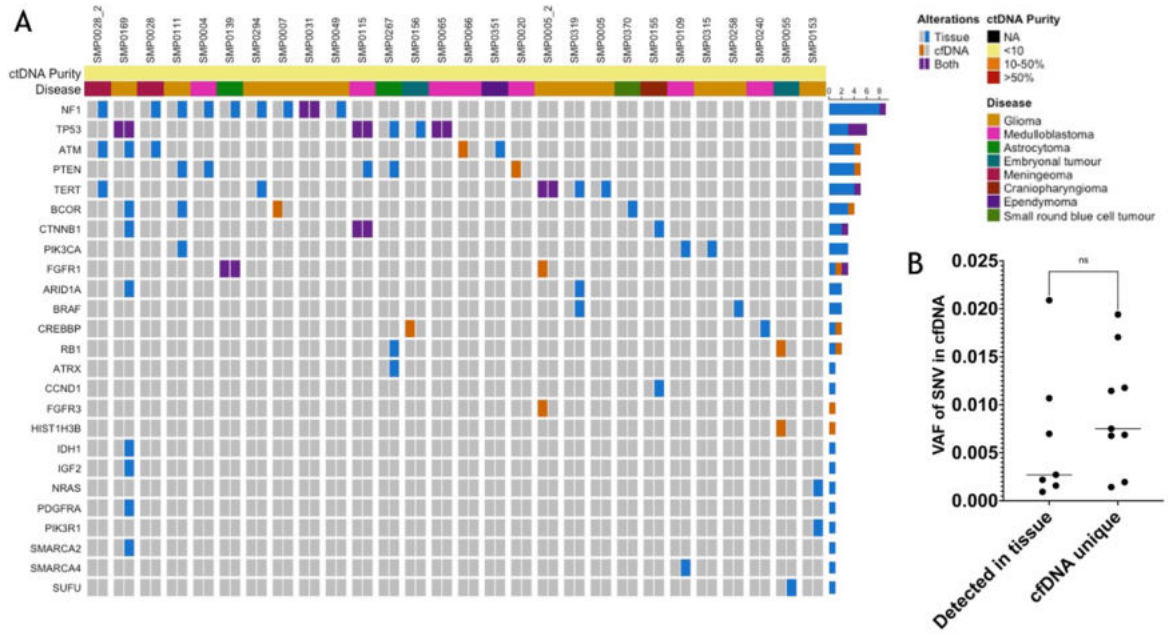


Figure 5-2 Comparison of panel sequencing results in tissue versus cfDNA from plasma in Stratified Medicine Paediatrics programme patients with CNS tumours (A) Comparison of SNVs detected in plasma vs tissue sequencing, ordered by the most commonly mutates gene. Each column represents single patient, colour coded for disease type at the top and ctDNA purity (by lcWGS). Alterations in purple have been detected in both tissue and cfDNA, blue indicates the SNVs detected in tissue but not cfDNA and orange highlights SNVs detected in cfDNA but not tissue. (B) VAF of SNVs detected in cfDNA, comparing SNVs detected both in tissue and cfDNA and cfDNA unique SNVs.

In this cohort, 16 SNVs have been reported in cfDNA – 7 matching tissue panel results and 9 unique to cfDNA (Table 5-2). The VAF was low in all samples, with no significant difference between cfDNA unique SNVs and SNVs that were detected in the tissue as well (Figure 5-2 B). If anything, the cfDNA unique variants had slightly higher VAFs, due to stricter filtering criteria for tissue-agnostic variant detection. For example, in patient SMP0005_2 (second relapse while on trial) *TERT* promoter mutation was detected both in tissue and cfDNA, but 3 cfDNA unique SNVs were detected, one of them at higher VAF than the tissue-confirmed variant.

In two patients, where no SNVs were detected in the tissue, cfDNA profiling identified an SNV, highlighting the added value of cfDNA profiling.

Table 5-2 SNVs detected in cfDNA in Stratified Medicine Paediatrics programme patients with CNS tumours. VAF in tissue and cfDNA; Total sequencing depth at the SNV location (UMIx) and the number of reads supporting the SNV in cfDNA. cfDNA unique mutations highlighted in orange. SNVs that were detected in tissue, but not cfDNA in otherwise cfDNA positive samples highlighted in blue.

Sample ID	Trial ID	SNV	VAF in tissue	VAF in cfDNA	cfDNA depth at SNV	cfDNA supporting reads
20/03874	SMP0005_2	FGFR3 c.1187C>A p.Thr396Lys	0.00%	0.14%	3469	5
20/03874	SMP0005_2	FGFR1 c.1682C>A p.Ser561Ter	0.00%	0.20%	2546	5
20/03874	SMP0005_2	TERT g.1295250G>A Promoter	38.41%	1.07%	1309	14
20/03874	SMP0005_2	FGFR1 c.1731C>G p.Asn577Lys	0.00%	1.71%	2405	41
19/06478	SMP0007	BCOR c.4199G>A p.Arg1400Gln	0.00%	1.94%	1390	27
19/06478	SMP0007	NF1 c.4872dupA p.Tyr1625fs	94.79%	0.00%	2028	0
19/08746	SMP0020	PTEN c.414T>G p.Tyr138Ter	0.00%	0.68%	1325	9
19/09802	SMP0031	NF1 c.5305C>T p.Arg1769Ter	68.35%	2.09%	2009	42
19/13191	SMP0055	RB1 c.1852dup p.Ser618PhefsTer35	0.00%	1.15%	3140	36
19/13191	SMP0055	HIST1H3B c.83A>G p.Lys28Arg	0.00%	1.18%	3646	43
19/13191	SMP0055	SUFU c.151C>T p.Pro51Ser	80.55%	0.00%	1437	0
19/14136	SMP0065	TP53 c.375+1G>A p.T125T	82.73%	0.22%	2730	6
20/01483	SMP0066	ATM c.8986A>C p.Ser2996Arg	0.00%	0.69%	872	6
20/03775	SMP0115	TP53 c.699_706del p.His233GlnfsTer4	92.49%	0.16%	3147	5
20/03775	SMP0115	CTNNB1 c.94G>T p.Asp32Tyr	52.09%	0.70%	4149	29
20/08626	SMP0139	FGFR1 c.1609A>C p.Asn537His	3.00%	0.27%	3682	10
20/10156	SMP0156	CREBBP c.2265C>A p.Ser755Arg	0.00%	0.75%	1328	10
20/10156	SMP0156	TP53 c.714dup p.Asn239Ter	8.00%	0.00%	1255	0
20/11503	SMP0169	TP53 c.817C>T p.Arg273Cys	47.36%	0.09%	3176	3

However, in three patients, cfDNA unique variants were detected when the most dominant SNV in the tissue profiling were not detected (Table 5-2 highlighted in blue). Given that blood and tissue samples were collected at the same time it raises the question about the reliability of cfDNA unique variants. In these cases, however, all 4 cfDNA unique SNVs were detected at VAF above 0.5% which our validation had shown to be very reliable. Therefore, we are confident that the SNVs are present in cfDNA, but their biological origin of them is not clear without multi-region and multi-site (for metastatic patients) biopsies. Only one SNV (*HIST1H3B* c.83A>G p.K28R) was reported as potentially pathogenic in COSMIC database, with one case only. The other three SNVs were not reported as

pathogenic on cancer databases (COSMIC, VarSome) raising the possibility that they may have originated not from tumour cells, but from other organs in the body (not blood cells though, as these SNVs were not detected in matched blood cell pellet samples).

In all patients with CNS tumours, lcWGS analysis on cfDNA showed silent profiles, even for the patients where SNVs were detected by panel sequencing (representative case - SMP0007 in Figure 5-3 A). It is known that CNS tumours tend to have less aberrant lcWGS profiles, but corresponding tissue profiling showed CNVs, supporting the low ctDNA purity hypothesis. This is in agreement with the low VAF of variants detected in cfDNA - the highest VAF observed in panel sequencing was at 2%. However, low purity samples should not be dismissed as indicated by patient SMP0031 where a focal high MYCN amplification could be seen both in panel and lcWGS (Figure 5-3 B).

Panel sequencing approach is economically viable, but inherently limited in patients with low mutation burden. In this cohort, 44 patients with CNS tumours had no SNV, CNV or SV detected in tissue and cfDNA by panel sequencing. Additionally, 16 patients had only SVs and/or CNVs detected in tissue by panel sequencing and there were 7 samples where variants detected in tissue were not covered by ctPC panel. This, combined with very low detection rate of variants covered by the panel (only 13% (8/60) of SNVs detected in cfDNA that were expected from the tissue sequencing in patients at relapse), highlights the need to look for different methods to use liquid biopsy in these patients. This could be either by higher sensitivity techniques for blood cfDNA or by looking into alternative biofluids such as CSF.

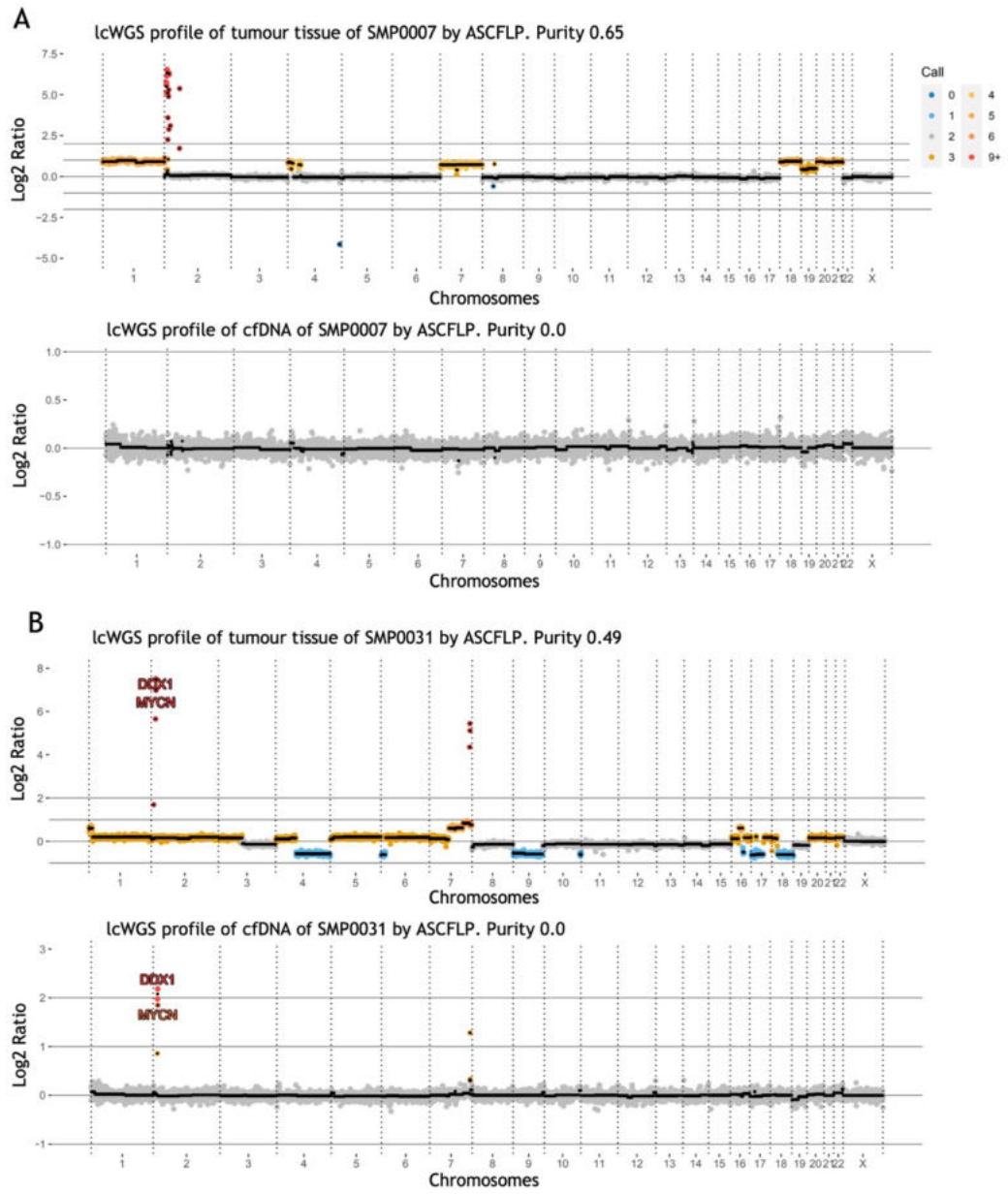


Figure 5-3 Representative lcWGS profiles of tissue and cfDNA from patients with CNS tumours in Stratified Medicine Paediatrics programme. (A) Patient SMP0007 showing CNV changes in tissue, but no CNVs detectable in cfDNA (B) Patient SMP0031 showing numerous CNV changes in tissue, but only high level MYCN and DDX1 amplification detectable in cfDNA with otherwise silent CNV profile.

5.3 CSF as an alternative source of cfDNA for patients with CNS tumours

In agreement with published literature, blood does not seem to be a very good source of ctDNA in patients with CNS tumours, therefore I have explored the molecular profiling of CSF-cfDNA. Several studies had shown that for patients with

CNS tumours, CSF is a much better source of ctDNA than blood [105,222–227]. However, obtaining CSF samples is more invasive and demanding than collecting peripheral blood (e.g. requirement for sedation/anaesthesia, risk of infection when intraventricular devices are used). Therefore, studies so far mostly utilised existing clinical practices and collected surplus CSF when routine lumbar punctures (LPs) were performed [227,231]. We followed the same practice in our proof-of-concept studies.

5.3.1 Preliminary results of CSF-cfDNA analysis

The collection of CSF samples is much more invasive and at the beginning of this project, there were no clinical trials allowing the collection of CSF for research purposes in our partner hospitals. However, though one of the trials we had consent to collect surplus CSF after routine diagnostic tests. In practice that meant collecting surplus CSF samples from the routine diagnostic microbiology laboratory, which led to suboptimal sample processing (samples kept at room temperature for extended periods of time) and small sample volumes (mean of 1.2ml, min 0.15ml, max 3.3ml) (Figure 5-4 A). In total, 22 CSF samples from 13 patients were collected (3 failed extraction) and all of them had extremely low cfDNA yields (Figure 5-4 B). Given such low cfDNA content and suboptimal pre-analytical handling, I have analysed the fragment sizes of the samples and confirmed that most of them did not contain the cfDNA specific fragment profile (Figure 5-4 C-E). Some samples seemed to be over fragmented, such as Figure 5-4 E, while others seemed to have most of the fragments at the high molecular weight fraction Figure 5-4 C. Out of these I have selected the three highest quality samples for full molecular profiling (ctPC panel and lcWGS), but all of them produced poor quality NGS libraries and resulted in very low sequencing depth (<200xUMI) and no oncogenic variants were detected.

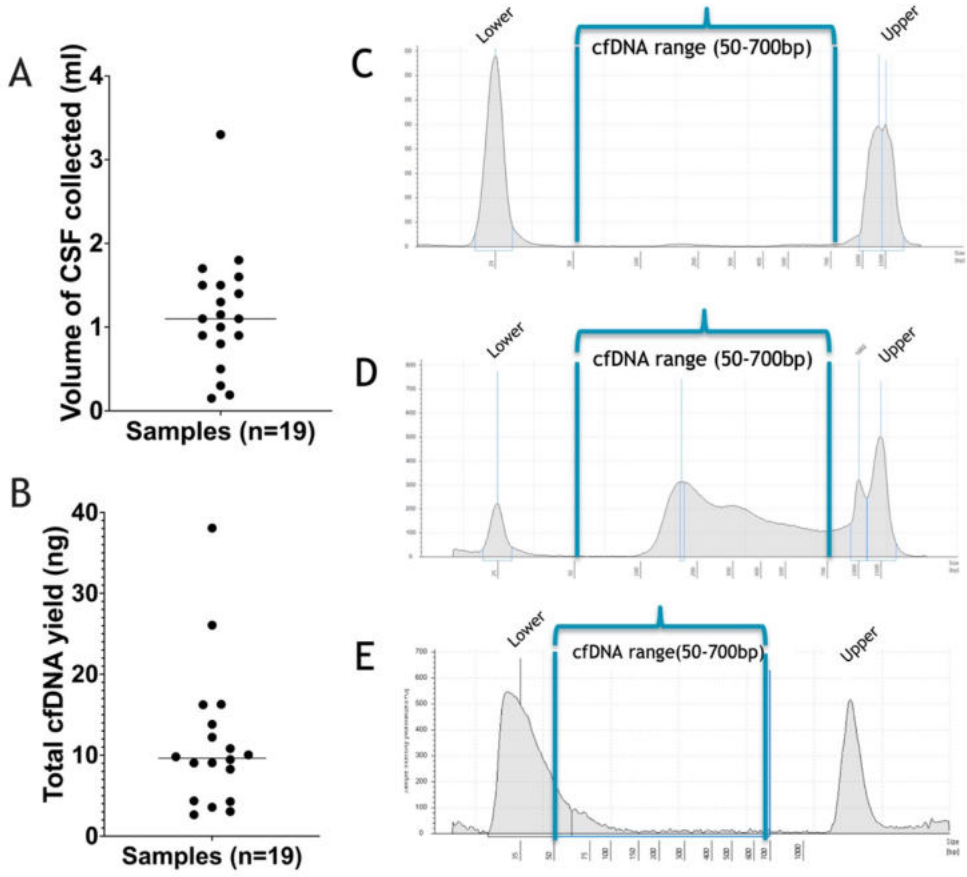


Figure 5-4 CSF samples from NGS-4294 study. (A) The volume of CSF samples collected per patient in this trial (ml of CSF) (B) The total cfDNA yield from these CSF samples (total ng). C-E Examples of typical fragment size analysis of CSF-cfDNA by Tapestation with the High Sensitivity (C and D) and cfDNA (E) DNA ScreenTape assay. The y-axis showing the signal intensity (FU) and the x-axis showing the DNA fragment size (base pairs). cfDNA range shown in blue, assay markers at (50bp and 1500bp marked as lower and upper marks). (C) CSF-cfDNA sample with most of DNA in the HMW fraction (D) CSF-cfDNA sample with a peak of cfDNA at 173bp and some HMW contamination E. CSF-cfDNA sample with most of DNA in the ultra short fragment fraction (overlapping with lower marker).

Later, a clinical trial focusing on liquid biopsy profiling for paediatric cancer patients was opened in our partner hospital and several CSF samples have been collected specifically for research purposes. All CSF samples were from lumbar punctures and processed by spinning down and freezing the CSF supernatant straight after the collection with an average volume of 2.8ml collected (min.1.1 max 5ml of CSF) (Table 5-2). However, from 8 samples collected, only 3 had good enough quality cfDNA to perform panel sequencing (Figure 5-5). The mutation expected from tissue sequencing was detected only in one of the three CSF-cfDNA samples, R0014, which had the highest total cfDNA yield. In addition, lcWGS profiling was able to provide a distinct CNV profile for this sample as well, reporting relatively high ctDNA fraction (Figure 5-5 D). Unfortunately, blood

sample was not available for this patient for comparison of detection between CSF and plasma cfDNA.

This preliminary data supports the exploration of CSF cfDNA profiling, however much more data is needed to make any conclusions. We still need to explore the factors influencing the quality of cfDNA which seems to be a hurdle for successful CSF-cfDNA profiling.

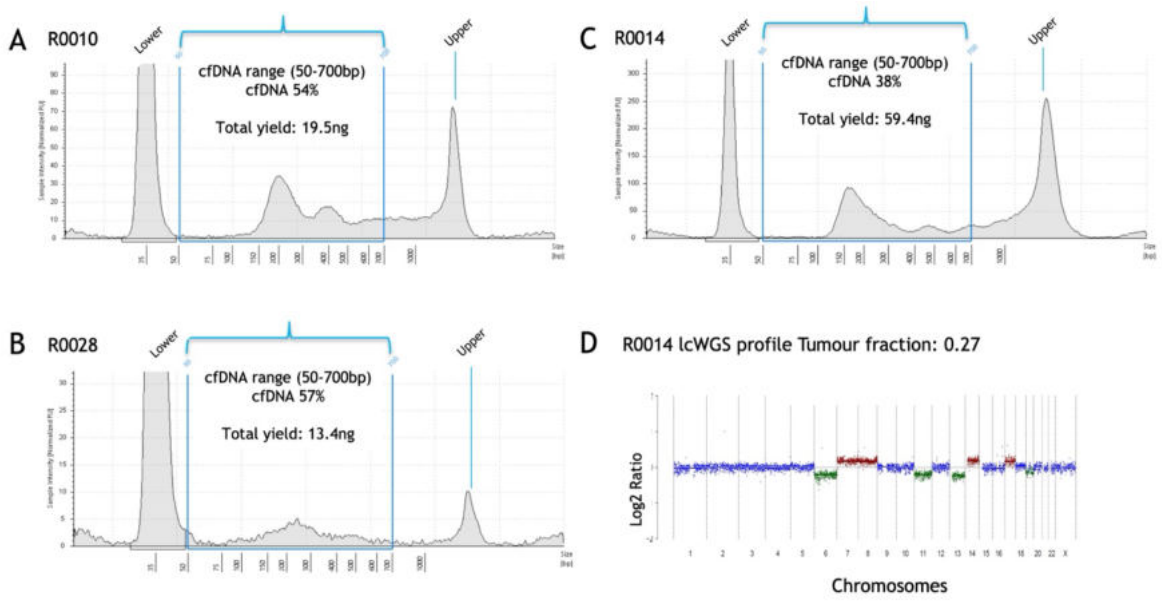


Figure 5-5 The three CSF-cfDNA samples moved forward to sequencing using ctPC panel and lcWGS. A-C The fragment size analysis of CSF-cfDNA by Tapestation with the High Sensitivity DNA ScreenTape assay. The y-axis showing the signal intensity (FU) and the x-axis showing the DNA fragment size (base pairs). cfDNA range shown in blue, assay markers at (50bp and 1500bp marked as lower and upper marks). The percentage of DNA in the cfDNA range highlighted for each sample next to total yield. (D) lcWGS profile of sample R0014 by ichorCNA – Log2 ratio of bins in each chromosome plotted on the Y axis with gains highlighted in red and losses in green.

Table 5-2 CSF samples collected for the Liquid biopsy study. Quality of cfDNA was assessed for all samples by qubit and fragment size analysis where enough DNA was available. NA indicates the analysis was not performed. Only 3 samples were moved froward for sequencing with ctPC panel and lcWGS (R0028 all NGS library used up for panel sequencing and lcWGS not preformed).

			Quality of cfDNA			ctPC and lcWGS results			
Trial ID	**Diagnosis**	**Timepoint**	**Volume of plasma (ml)**	**Total cfDNA yield (ng)**	**ctDNA fraction (by fragment size analysis)**	**Total reads**	**Depth UMIx**	**ctDNA purity (by lcWGS)**	**cfDNA sequencing results**
R0010	Pilocytic astrocytoma	Diagnosis	4.00	19.49	54%	4.47E+07	480	0	Potentially PTPN11 variant at 1.5% (tissue not sequenced at the moment)
R0014	Medulloblastoma	Diagnosis	5.00	59.4	38%	3.74E+07	1152	0.27	CTNNB1 variant known from tissue detected at 29.8%
R0028	Ependymoma	Relapse	2.60	13.42	52%	2.71E+07	92	Not enough library	No variants in cfDNA (tissue not sequenced at the moment)
R0003	Medulloblastoma	Diagnosis	4.00	58.08	16%	NA	NA	NA	NA
R0005	Diffuse Leptomeningeal Glioneuronal Tumour	Diagnosis	0.10	2.42	NA	NA	NA	NA	NA
R0006	Left Thalamic Low Grade Glioma	On treatment	4.10	33.26	6%	NA	NA	NA	NA
R0011	Low grade glioneuronal tumour	Diagnosis	1.60	13.81	17%	NA	NA	NA	NA
R0024	Ependymoma	Relapse	1.10	2.86	NA	NA	NA	NA	NA

5.3.2 Practical and technical considerations for liquid biopsy implementation for children with CNS tumours

Success of molecular profiling of cfDNA depends upon maintaining cfDNA integrity and minimising the contamination from genomic DNA from non-cancer cells. This has been highlighted by the poor results and poor sample quality in our preliminary studies. Rigorous studies exploring the pre-analytical conditions of CSF sample processing for liquid biopsies are lacking, mainly due to the difficulty of acquiring CSF samples from patients. Therefore, the best practices must be inferred from studies in blood and the limited studies we have on CSF. Therefore, based on the methodologies in the most recent literature [142,227,232,234-236], I aimed to outline the current best practices of pre-analytical CSF cfDNA handling, which is shown in Figure 5-6:

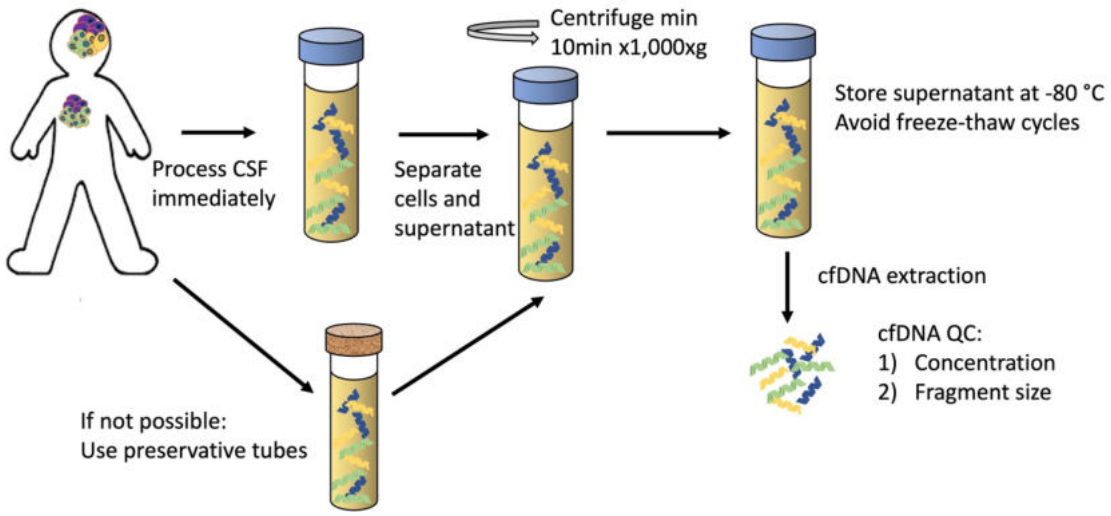


Figure 5-6 The proposed guidelines for pre-analytical CSF handling for cfDNA analysis.

Sample needs to be processed immediately:

- Separate supernatant from the cell pellet (min 10 min x 1000xg)
- Store at -80°C prior to processing; multiple freeze-thaw cycles should be avoided
- Extraction methods have not been systematically evaluated, most commonly used – Qiagen® "QIAamp Circulating Nucleic Acid Kit"

- Quality Control to assess DNA concentration (Qubit® assay or qPCR-based measurement) and fragment size (by automated electrophoresis systems) to confirm the presence of cfDNA

If immediate processing is not possible, collection tubes with preservative should be used.

Even though CSF has low cellular content, the separation of CSF into the supernatant and the cell pellet is necessary. The cfDNA in the supernatant of CSF has been shown to have higher VAF of cancer-associated variants when compared to the pellet from the same sample, indicating that supernatant often contains sufficient cfDNA and provides more reliable results than the cellular fraction [234,237]. It is partly explained by the presence of contaminating cells from CSF collection procedure and/or infiltrating lymphocytes, diluting the tumour variant signal in the pellet [237]. However, a high proportion of the patients in these studies harboured solid tumours with CNS metastases, and a more extensive confirmation in localised CNS primary tumours would be helpful. The preliminary data points to higher numbers of SNVs and structural rearrangements detected in CSF supernatant in primary CNS tumours [237], therefore, it is routine to spin down the CSF to remove contamination from non-cancer cells, but the exact protocol differs between laboratories - speeds ranging from 500xg for 5 minutes [153] to 3,000xg for 5 minutes [227] to 1,900g for 10 minutes followed by a further 16,000g for 10minutes [153] have been used. Huang and colleagues [235] have investigated this more systematically (albeit only in one patient) and showed that after 1,000xg x10min centrifugation, extracted DNA fragments were exclusively around 150bp, consistent with cfDNA and at a lower speed with a shorter time, larger fragments were also observed. A systematic study evaluating different CSF centrifugation protocols and their effect on the cfDNA quality and detection of cancer specific alterations is needed, but until then, the standard cell separation protocol of ≥ 10min centrifugation at 1,000xg should be used for future studies.

To minimise the contamination from lysing non-cancer cells when samples cannot be processed immediately, specialist cell-stabilising preservation tubes have been used for blood collection for liquid biopsies. There is a range of tubes on the market, but none are specifically designed for CSF. The only study to date comparing different conservation tubes for CSF samples suffered from the

limitations of sample availability but showed preliminary evidence that Norgen® tubes with phosphate-buffered saline to top up low volume samples have the highest cfDNA yields of the different tubes tested [142].

5.4 Discussion

Liquid biopsy based molecular profiling is more challenging in patients with CNS tumours compared to other solid tumours. Throughout the course of this project, 128 blood samples from patients with CNS tumours have been molecularly profiled to detect SNVs and CNVs using ctPC panel and lcWGS (Figure 5-7). Blood based cfDNA profiling showed limited results, with only 6 out of 28 patients showing detectable levels of ctDNA at relapse (with 13% of SNVs expected from the tissue sequencing detectable in cfDNA). Several patients had longitudinal samples collected, but variants expected from the tissue profiling were not detectable in any of these patients, even at late disease stages. All blood derived cfDNA samples showed silent lcWGS profiles (except a few very high focal amplifications), highlighting very low ctDNA levels even in the samples where mutations were detected by tissue sequencing. The highest VAF detected in blood cfDNA was 2% in a patient with high-grade glioma at relapse. These findings agree with most adult and paediatric studies, showing poor variant detection in blood in patients with CNS tumours [105,222–227].

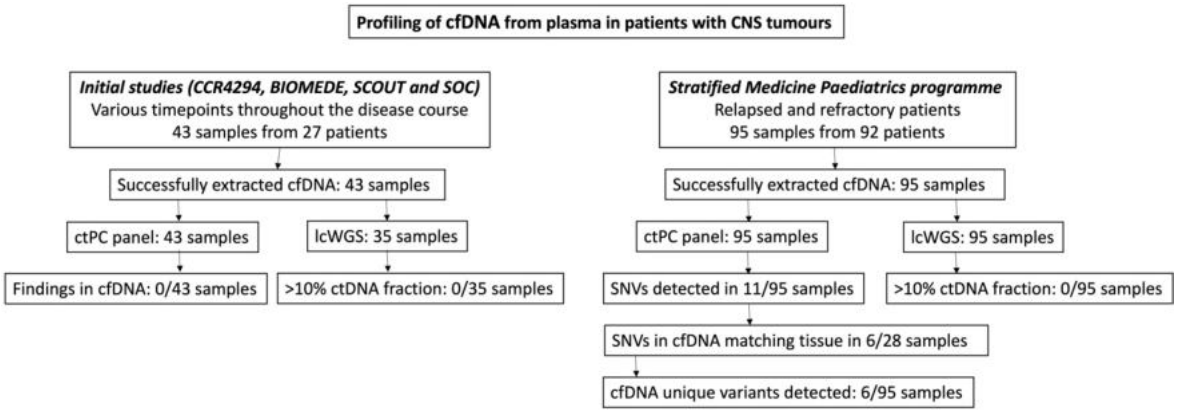


Figure 5-7 Summary flow chart of blood derived plasma cfDNA samples discussed in this chapter. Only patients with CNS tumours included. The number of samples and number of patients included in different studies, highlighting the numbers of successfully processed samples and successful results of analysis in comparison to tissue profiling with comparable method (NGS capture panel sequencing and lcWGS of the same library). ctDNA fraction was evaluated using lcWGS, with 10% as a limit of detection.

Several ways to improve the ctDNA detection from the blood of patients with brain tumours have been explored. For example, pre-amplification of cfDNA after extraction has been described as a way to improve the detection of low-level ctDNA variants. This led to the detection of ctDNA in 80% of newly-diagnosed patients with glioma from 1ml of plasma and in 87% of patients from 0.5ml of cerebrospinal fluid (CSF) [96]. In a different study of newly diagnosed patients with diffuse intrinsic pontine glioma, H3K27M mutations were detected in 92% of the patients using the same pre-amplification method [97]. The method was further optimised on different ddPCR platforms and tested in multiple laboratories, setting the first steps towards standardisation [153]. Mutations were detected in all plasma specimens in this study, but the VAFs were lower in blood than in CSF. However, the low number of samples tested limits the conclusions that can be drawn about the broader applicability of the method. Additionally, it is important to note that while the chances of errors occurring at the specific site during pre-amplification are low, they are not nil. Therefore, while pre-amplification is appropriate for ddPCR based methods, it is likely to introduce higher numbers of false positives if broader NGS based profiling methods (such as NGS panels or WES) are to be used. Assessing individual reads without error correction at pre-defined loci can result in high false-positive rates, highlighting the need to carefully assess the sensitivity and specificity of each assay [105]. Therefore, we did not attempt to pre-amplify our cfDNA samples that were to be analysed by NGS panels. In the future, when better error correction methods are established, this could be tested to try and increase the detection rates in these patients. However, ultimately sensitivity depends on absolute detection – there must be enough ctDNA molecules for detection.

Another way to increase the sensitivity of detecting variants in cfDNA would be to increase the permeability of the BBB to allow more ctDNA to enter the bloodstream. *In-vivo* and patient studies have both shown that higher ctDNA levels are present in the blood after radiotherapy [153,232]. Given that radiotherapy is a mainstay of therapy for most paediatric malignant CNS tumours (with the exception of the youngest patients, for reasons of neurocognitive toxicity sparing) this opens the possibility of planning blood collection for a time when maximal

tumour DNA shedding and BBB disruption is predicted. Although this would only provide a snapshot of the molecular profile of a tumour at one specific timepoint during therapy, this may still provide valuable information, particularly if a patient subsequently experiences relapse. This requires a change in the protocols of blood collection, and clinical trials specifically designed to collect blood at the relevant timepoints are needed to test this hypothesis and evaluate if this approach is feasible in the clinic.

Most literature shows that CSF is a better source of cfDNA for CNS tumours [105,222-227]. I did not have many samples to test this hypothesis, but in one case the blood and cyst fluid were collected from a patient with DIPG and only the cyst fluid had ctDNA and variants expected from tissue sequencing detected. In addition, the preliminary data showed that poor quality of CSF can be a major roadblock to the successful implementation of these assays (Figure 5-8), therefore I aimed to outline recommendations for pre-analytical sample processing. Moving forward, standardisation of pre-analytical sample processing and thorough validation of assays is needed.

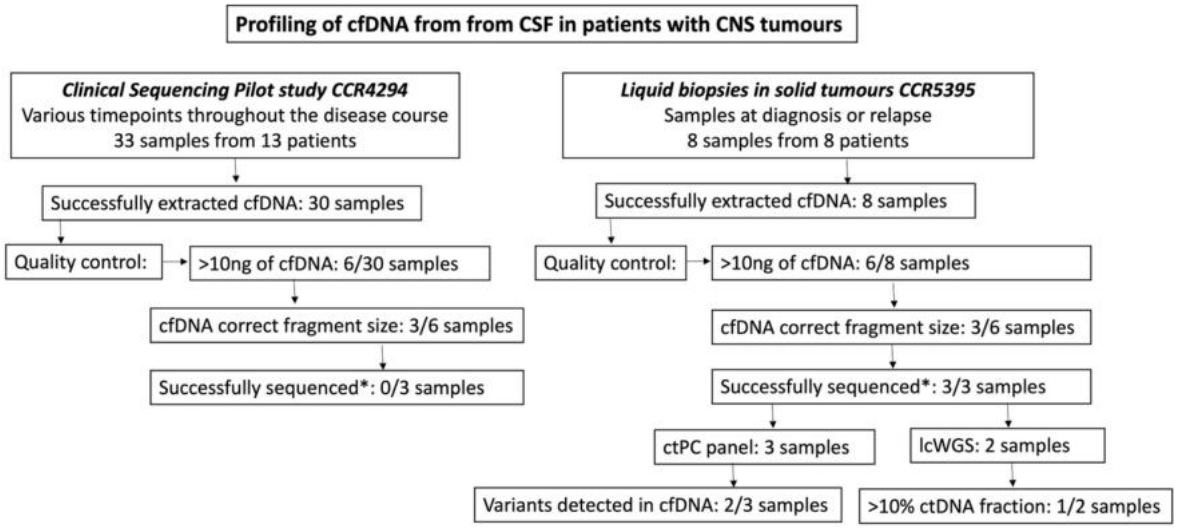


Figure 5-8 Summary flow chart of CSF derived cfDNA samples discussed in this chapter. The number of samples and number of patients included in different studies, highlighting the numbers of successfully processed samples and samples passing quality control. Fragment size of cfDNA was assessed using TapeStation analyser. Successful sequencing* defined as UMIx depth of at least 250x.

In our study, from the three patients with the best quality of CSF-cfDNA (Figure 5-8), variants expected from tissue sequencing were detected only in one of the patients. The other two however did not have tissue profiled at the time of this analysis and it is possible that no SNVs were present in the tissue as well. A similar issue has been reported by Sun et al., where most of the CSF-cfDNA samples showed poor ctDNA detection in paediatric medulloblastoma patients most likely due to poor sample quality [233]. Additionally, in a study exploring methylation and hydroxymethylation patterns on CSF cfDNA in patients with medulloblastoma low genomic coverage or low bisulfite conversion efficiency was observed in some samples [173]. However, several studies had shown detection of good quality ctDNA in CSF in paediatric cancer patients with CNS tumours [227,231,232]. Therefore, more evidence is still required before LPs can be routinely recommended for follow-up or early diagnosis of relapse or disease response monitoring in patients with CNS tumours, particularly the factors influencing the cfDNA quality. However, as the weight of evidence increases to support the use of CSF-cfDNA, there is scope for evaluation in prospective studies comparing MRI with ctDNA for response assessment and monitoring of relapse. Key questions include how confidently and how much earlier than MRI, cfDNA can detect emerging relapse, and whether this indeed makes a difference to patient outcomes if alteration to treatment is considered earlier.

In addition, CNS tumours often have less complex SNV and CNV profiles and few recurrent mutations to track [14], which limits the number of alterations to assess and hence the sensitivity of the assays. Other forms of molecular analysis such as epigenetic, metabolomic and transcriptomic profiling have also been performed in cfDNA from patients with brain tumours [213,214]. Epigenetic profiling is the most advanced of these techniques at the moment and includes methylation, fragmentation, and nucleosome occupation analyses [238]. Methylation profiling in particular, is now being routinely used as a diagnostic tool in tissue biopsy samples, and the evidence of its utility on cfDNA is accumulating [160,172,173,239]. In paediatric sarcoma patients, ctDNA detection and classification based on cancer-specific chromatin signatures and independent of genomic alterations has been successful [160]. This can also potentially be applied to CSF-cfDNA and has already been demonstrated in paediatric medulloblastoma where epigenetic signatures were similar in tissue and CSF-cfDNA and retained subtype specificity in patients

with good quality samples [173]. Dynamic changes in methylation of signature clusters were reported with reduction of methylation levels of cfDNA in patients responding to treatment [173]. A proof-of-concept study using reduced representation bisulphite sequencing on cfDNA from plasma and CSF successfully classified 81% of samples from a range of paediatric tumours using less than 10ng of DNA [172]. The study included only 4 CSF samples from patients with CNS tumours but was able to distinguish medulloblastoma from ATRT. If validated in larger cohorts of patients with CNS tumours, these technologies may be of particular relevance in certain paediatric brain tumours such as medulloblastoma and ependymoma which have a low mutational burden. We have explored CSF-cfDNA methylation analysis by nanopore sequencing to some extent, but due to a limited number of samples and technical issues we did not get any meaningful results. However, this is one direction for future studies, both for profiling and ctDNA detection and monitoring for paediatric patients with CNS tumours.

Overall, this chapter adds to the evidence that blood based cfDNA profiling is challenging in patients with CNS tumours and CSF-cfDNA is emerging as a biomarker worthy of further evaluation for molecular profiling, especially in specific tumour types such as medulloblastomas or paediatric high-grade gliomas, in patients with unresectable tumours (such as diffuse midline gliomas), or where repeated tissue biopsies are required. However, to realise this potential for clinical benefit in children with CNS tumours, CSF-cfDNA tests must be robustly evaluated. Firstly, standardisation of sample processing is needed. Subsequently for each methodology, robust QC measures, normal ranges and diagnostic cut-offs must be defined. This will require larger scale studies comparing the different methodologies for different indications. Standardisation of methodologies across international consortia will be crucial with large-scale implementation and harmonisation best placed within the context of international disease-specific clinical trials. Initiatives like this are already underway, such as the SIOP High Risk Medulloblastoma [240] and SIOP II Ependymoma (NCT02265770) trials.

As the weight of evidence for the technical utility of liquid biopsy assays increases, we must next address where the greatest potential is for added benefit in patients. It is clear for example, that the use of CSF-cfDNA profiling to aid inclusion into clinical trials when biopsy is not feasible is beneficial in the proportion of patients where cfDNA profiling is informative. It will be challenging

to change treatment based on a non-imaging finding, which is not accommodated in current clinical trial designs. At present, inclusion criteria and response assessment endpoints are imaging based, and LPs (for CSF-cfDNA collection) are generally not mandated on clinical trials. However, as clinical trials (such as SIOP Ependymoma II trial (NCT02265770)) are starting to collect CSF at the time of the initial biopsy and after surgery this offers the potential for further validation of these assays, and the opportunity to drive a paradigm change in clinical trial design.

Chapter 6 Monitoring disease progress using cfDNA

Parts of this chapter were published as part of [149]

6.1 Introduction

Tissue biopsy samples represent a snapshot of a single point in time and are subject to spatial selection bias, especially in heterogenous tumours. However, liquid biopsy techniques offer the opportunity to overcome sampling limitations inherent to tissue biopsy. Genomic analysis of single-lesion tumour biopsies upon disease progression under-represents tumour heterogeneity and risks missing clinically relevant resistance mechanisms in most adult solid tumours [196,241–244]. Evidence is accumulating in patients with adult cancers that cfDNA can overcome this issue by better representing heterogeneity present in the tumour and most importantly identify multiple concurrent resistance mechanisms [128,245–248]. Repeated tissue samples are difficult to obtain and cannot be used for dynamic monitoring of disease progression and response to therapy. For most cancer types, monitoring is performed using scans, but this does not allow interrogation of the genomic changes that occur in the tumour as the disease progresses. The minimally invasive nature of liquid biopsies from blood allows serial sample collection and opens up opportunities to track disease progress and potentially study tumour evolution.

Early paediatric liquid biopsy studies explored whether cfDNA levels could be informative in monitoring the disease progression – total cfDNA levels were shown to correlate with tumour burden in patients with neuroblastoma [102]. More specific methods, allowing quantification of SNV, CNV and SV, showed that ctDNA levels correspond to treatment response in patients with neuroblastoma [99,101], Ewing sarcoma [90,92,130], osteosarcoma [90,94] and rhabdomyosarcoma [90,110]. In patients harbouring fusion genes, highly sensitive monitoring of patient-specific alterations allowed monitoring tumour burden, response to therapy and disease relapse [92,110], however patient-specific assays are required for this approach. Serial sample collection might be of particularly high relevance in children where tumours are

likely to acquire changes between initial diagnosis and relapse [48–50,125–127] and serial profiling of cfDNA could identify the emergence of clinically relevant resistance alterations, as reported in adult cancers [128,129].

In patients with high ctDNA levels, serial cfDNA profiling opens the door to studying cancer evolution which allows for the prediction of the changes in intra-tumour subclone dynamics and potentially tailoring the treatment. For example, in adult colorectal cancer patients treated with EGFR specific antibodies, cfDNA molecular profiling identified the emergence of resistance mutations and dynamics of resistant clones throughout the treatment [245,249]. This then allowed dynamic treatment strategies, involving rechallenge with targeted inhibitors in patients based on cfDNA profiling [245,250]. Additionally, as shown in another study of adult colorectal cancer patients, evolutionary mathematical modelling can be utilized to construct predictive models for individual patients [159]. Frequent serial sampling potentially allows for response and progression rates to be determined for each subclone in the patients' tumour that in turn allows quantitative prediction of the time to progression in the patients who initially respond to a targeted treatment as well as predicting which subclone will dominate the dynamics of progression [159]. Similarly, in patients with neuroblastoma, quantification of the levels of different subclones (by targeted panel based on tissue-detected alterations) allowed for tumour evolution modelling, identifying several different evolutionary patterns [99].

Overall, the studies in paediatric cancer patients have shown the potential for disease monitoring using cfDNA based assays, but most of the studies required patient-specific assays. In this project I wanted to build on this knowledge and evaluate if we could use a tumour-agnostic panel to monitor the patients throughout their disease course and potentially study tumour evolution in cfDNA.

6.2 Serial cfDNA profiling can be used to monitor the course of disease

6.2.1 cfDNA can monitor disease progress

The minimally invasive nature of liquid biopsies using blood offers the possibility of serial sample collection and therefore possibility of monitoring disease progress. In this project, I aimed to evaluate the benefits of serial cfDNA profiling in patients with paediatric solid tumours. I identified 14 patients (10 with neuroblastoma and 4 with rhabdomyosarcoma) who had blood or tissue biopsies taken at more than one time-point throughout their disease course (1-3 cfDNA samples per patient, Appendix Table 3. Of these, 2 patients were responding to therapy, and ctDNA findings were consistent with clinical response.

Patient 11 was a 7-month-old girl diagnosed with Stage M, *MYCN* amplified neuroblastoma who was commenced on rapid COJEC induction chemotherapy. *CDKN2A* deletion and a pathogenic *ALK* F1174L mutation were also detected in both tumour tissue and cfDNA. The patient had a very good partial response to induction treatment and coincident with this, the *ALK* mutation, *MYCN* amplification and *CDKN2A* deletion were undetectable at the end of induction therapy in cfDNA (Figure 6-1 A). In agreement with targeted panel sequencing results, lcWGS revealed multiple genome-wide copy number changes in the cfDNA sample at diagnosis, which became undetectable during the treatment as the ctDNA fraction fell. In contrast, in patient 3 (a 13-year-old girl diagnosed with neuroblastoma), persistence of stable *MAP2K1* K57N mutation in ctDNA was shown with treatment refractory disease (Figure 6-1 B). In this patient no copy number changes were detected by lcWGS at any stage, most likely due to relatively low ctDNA levels (as indicated by *MAP2K1* mutation VAF=3.1-6.0%).

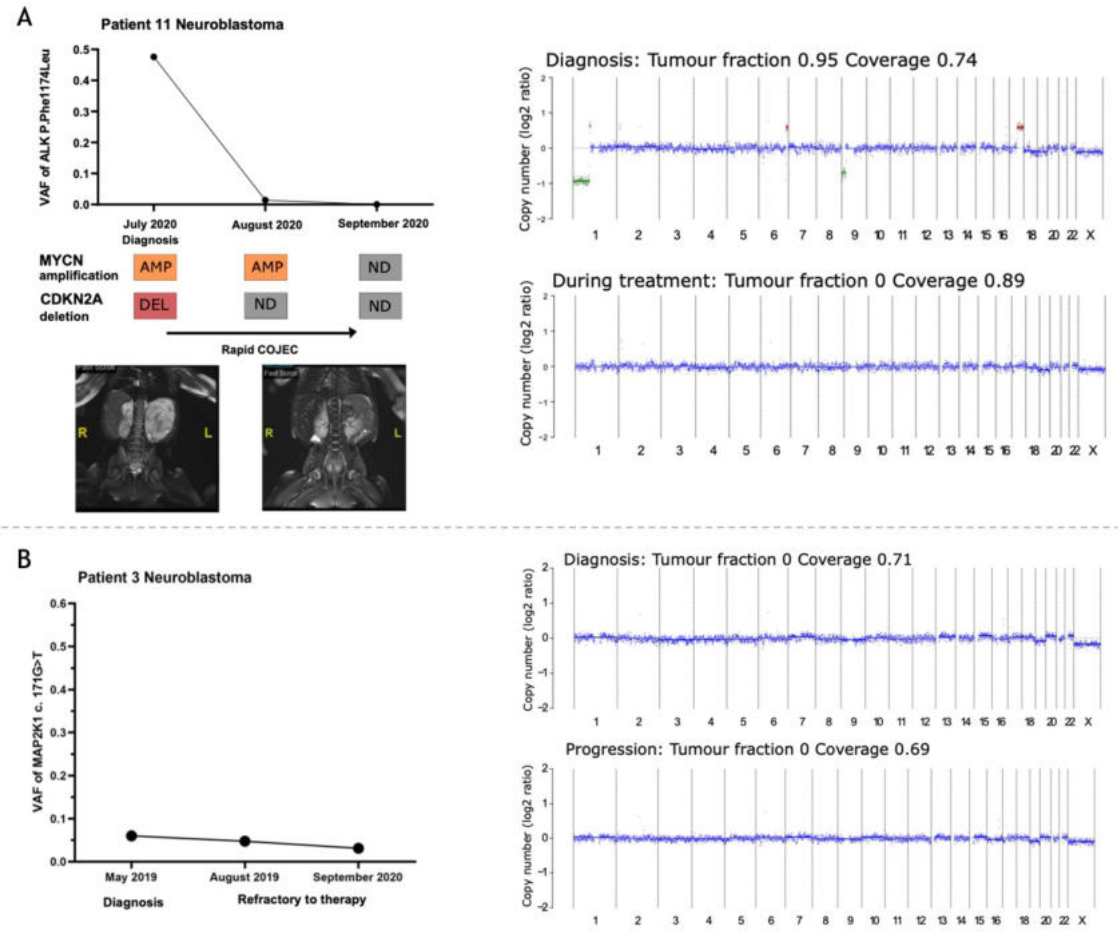


Figure 6-1 Monitoring response to treatment using cfDNA in patients with neuroblastoma. Levels of mutations detected in cfDNA (VAF) shown in the top graph for each patient. Presence of copy number changes indicated below as well as time matched lcWGS plots of cfDNA for each patient at different treatment timepoints. (A) Levels of ALK mutation and CNVs detected in cfDNA of patient 11. Below the mutations detected in cfDNA, diagnostic and end of treatment MRI scans are shown, showing reduction in size of all the lesions, consistent with decreasing variant levels in cfDNA. The lcWGS profile at diagnosis showing multiple CNVs is on the right. An example of lcWGS profile of the samples taken on treatment is shown below (silent profile in both on-treatment timepoints). (B) Levels of MAP2K1 mutation detected in cfDNA in patient 3. The lcWGS profile at diagnosis and at progression shown in the right of the graph, all timepoints silent profile.

In a different clinical trial, serial blood samples were collected for 3 patients with neuroblastoma throughout their disease course (6-9 blood samples per patient). Patient RMH027 (Figure 6-2 A) was a 2.5-year-old girl diagnosed with stage M neuroblastoma with a primary tumour in the abdomen and bone, bone marrow and lymph node metastases. A blood sample collected prior to the start of treatment showed high ctDNA levels with a pathogenic *ALK* L1249V (at VAF of 0.59) and a pathogenic *TERT* promoter (g.1295228G>A) mutation (at VAF of 0.12). Both mutations were also detected in the diagnostic tissue biopsy. The copy number profile of cfDNA prior to the start of the treatment was also consistent with the results of SNP array of tissue biopsy (performed as part of standard of care

molecular profiling), showing gains of chromosome 1q, 2, 3, 4, 5p, 7, 9, 13, 17, 18, 19, 20, 22 and loss of 11q. The patient was commenced on rapid COJEC induction chemotherapy and the blood sample collected post day 70 of the treatment showed highly reduced ctDNA levels (*ALK* at VAF of 0.004 and *TERT* at VAF of 0.007). The patient then underwent surgery and achieved complete remission which was reflected by undetectable ctDNA levels in the blood samples collected after the surgery and throughout further radiotherapy and immunotherapy.

Patient RMH035 (Figure 6-2 B) was a 12-year-old girl diagnosed with stage M neuroblastoma with a thoracic primary tumour, bone and bone marrow metastases and large sphenoidal mass with involvement of both orbits. At diagnosis, three variants were detected in ctDNA (pathogenic *ATRX* G1567VfsTer at VAF of 0.07, *FGFR1* N577K at VAF of 0.12 and variant of unknown significance in *BCOR* N1495K at VAF of 0.03). The pathogenic *ATRX* and *FGFR1* mutations were also detected in the diagnostic tissue biopsy. The variant of unknown significance in *BCOR* was not detected in the tissue biopsy, however it was not present in the matched blood cell pellet sample as well. The copy number profile of cfDNA prior to the start of the treatment was only partially consistent with the results of SNP array (performed as part of standard of care molecular profiling), showing gains of chromosome 7, 13, 18, 20 and loss of 3, 4, 8, 9, 11, 14q and 19, while the SNP array did not show the gain of chromosome 13 and loss of 14q, but showed gain of chromosome 17 that was not detected in cfDNA. The patient was commenced on rapid COJEC induction chemotherapy and blood samples collected during the treatment showed reduced ctDNA levels (*ATRX* and *FGFR1* mutations at VAF <0.015 while on treatment, *BCOR* not detectable). The patient had heterogenous response to the treatment with primary tumour (differentiating histology) responding well, but metastatic (poorly differentiated histology) sites progressing and leading to relapse. Upon relapse the patient was commenced on PARC which led to stable disease for several months, at which point patient was taken off the study due to spinal cord compression that required radiotherapy. The ctDNA levels remained stable and relatively low at relapse and during treatment on PARC (*ATRX* and *FGFR1* at VAF range of 0.015-0.026), consistent with stable disease in the patient.

Patient RMH015 (Figure 6-2 C) was a 13-year-old girl with primary abdominal tumour and extensive metastases in bone, bone marrow and lymph nodes at diagnosis. The patient was treated with rapid COJEC, followed by 4 cycles of topotecan and temozolomide after refraction. The patient was then enrolled into BEACON study and received 6 cycles of topotecan, temozolomide and dinutuximab. The first blood sample for liquid biopsy study was collected while on this treatment, when patient had stable disease. *MAP2K1* K57N pathogenic mutation was detected at VAF of 0.01 in cfDNA at this timepoint. The patient remained stable by MRI and bone marrow testing, had surgery in December 2021 and relapsed in February 2022. A cfDNA sample two weeks post-surgery showed high levels of ctDNA (*MAP2K1* at VAF of 0.36), indicative of high levels of active disease. Blood samples were not collected prior to surgery or any other timepoints prior to relapse, therefore it is hard to evaluate if the progression and relapse could have been anticipated by cfDNA testing. After the relapse, the patent was commenced on chemotherapy where some response was observed followed by progression in July 2022. This was reflected in ctDNA levels, with reduced levels (*MAP2K1* at VAF of 0.19) after chemotherapy and very high levels (*MAP2K1* at VAF of 0.5) just prior and during progression. The limited number of blood samples during chemotherapy (the only intermediate blood sample during the treatment failed cfDNA extraction) limits the conclusions on how closely the response could be monitored and how much earlier the progression could be anticipated, but generally the timepoints with cfDNA available for analysis correspond closely with the clinical state. Interestingly, this patient had multi-region biopsy of the tumour where *MAP2K1* mutation was detected in all samples, and a subclonal pathogenic *ATRX* R250Ter mutation was detected in one of the samples. Consistent with this, *MAP2K1* mutation could be tracked in all cfDNA samples of the patient and *ATRX* mutation was only detectable at very low levels (VAF of 0.002) in the ctDNA sample with highest purity, highlighting that ctDNA can mirror tissue heterogeneity in patients with high ctDNA levels. In this patient total cfDNA yield (as measured by cfDNA yield per ml of plasma) corresponded quite well with the disease course and lcWGS identified multiple copy number alterations.

In all 3 patients ctDNA levels as measured by mutations tracked using ctPC panel and lcWGS reflected disease course accurately. Additionally, lcWGS provided informative copy number profiles, informing of clinically relevant alterations,

such as chromosome 11q deletion and 17q gain. Total cfDNA levels (as measured by cfDNA yield per ml of plasma) were variable and did not correspond with the disease course in 2 of the 3 patients.

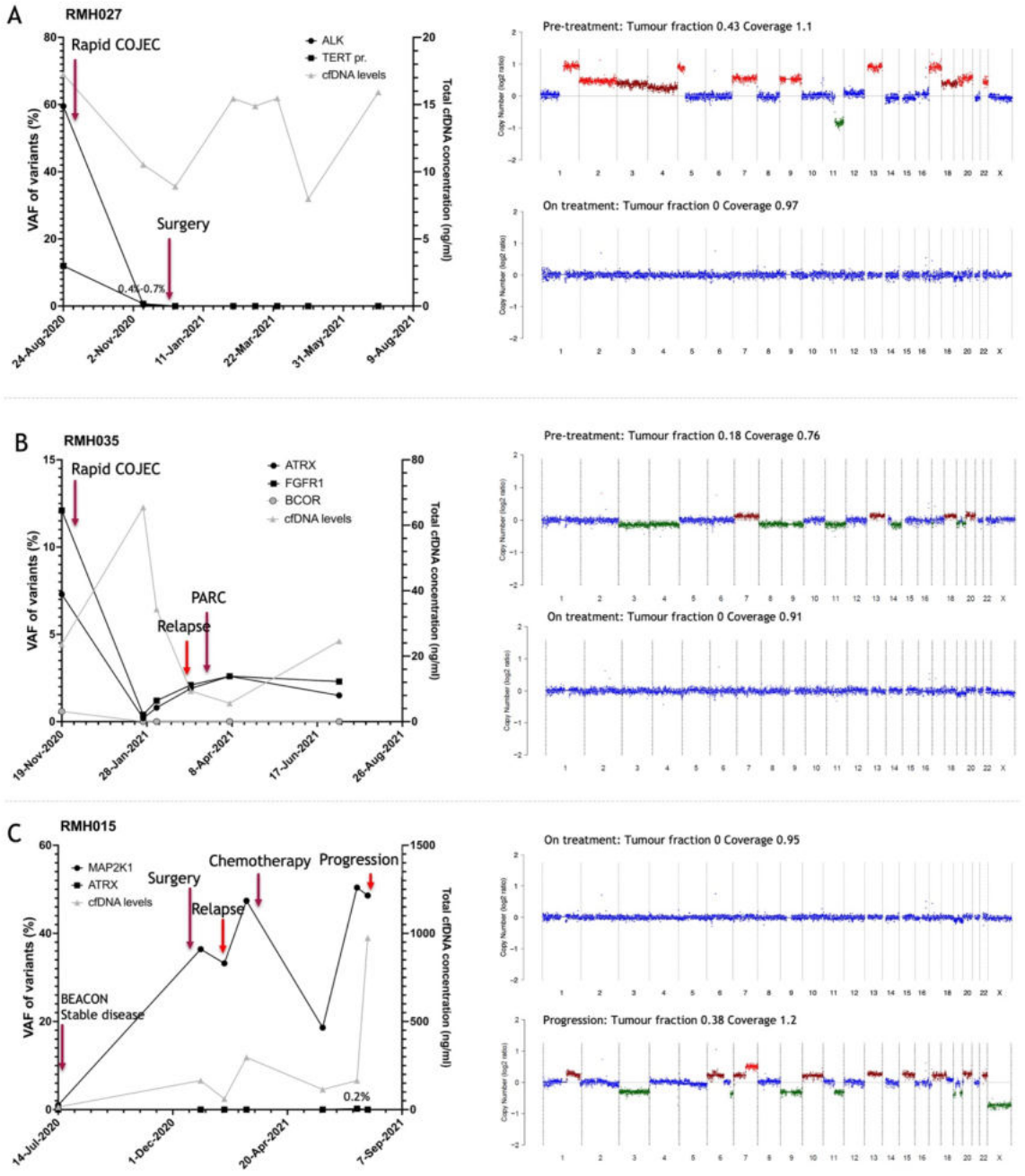


Figure 6-2 Monitoring response to treatment using cfDNA in patients with neuroblastoma. Levels of mutations detected in cfDNA (VAF) by ctPC panel shown in black for different mutations tracked, with the y-axis on the left. Total cfDNA levels (ng of cfDNA per ml of plasma) shown in grey with y-axis on the right. Treatment starting points are indicated by burgundy arrows and changes in clinical status (relapse or progression) are highlighted by red arrows. Representative plots of lcWGS of cfDNA throughout different treatment points are shown on the right for each patient. (A) Levels of cfDNA, VAF of mutations detected and representative lcWGS profiles of patient RMH0027. (B) Levels of cfDNA, VAF of mutations detected and representative lcWGS profiles of patient RMH0035. (C) Levels of cfDNA, VAF of mutations detected and representative lcWGS profiles of patient RMH0015.

Serial samples were also collected form several patients with CNS tumours to attempt disease monitoring. However, as highlighted in the previous chapter, the detection of ctDNA in the plasma of patients with CNS tumours is possible only sporadically and, in our cohort, no variants in plasma samples even at diagnosis or relapse were detected, therefore disease monitoring was not possible using liquid biopsy techniques.

Taken together, these case studies demonstrate the potential utility of serial ctDNA analyses for response assessment in extra-cranial solid tumours using pan-cancer tissue-agnostic capture panel and highlight the importance of future prospective evaluation of ctDNA analysis alongside standard of care investigations.

6.2.2 Monitoring response to treatment in patients on targeted therapies using cfDNA

Our studies included two patients with neuroblastoma, where a pathogenic ALK variant was identified by tissue and/or cfDNA profiling and based on that, patients received Lorlatinib on compassionate use access. For these two patients, serial blood samples (on average one sample each month) were collected throughout the treatment to evaluate the ability of cfDNA to monitor the disease progress. Lorlatinib is a selective adenosine triphosphate (ATP)-competitive inhibitor of ALK and ROS1 tyrosine kinases which is approved as a monotherapy for the treatment of adult patients with ALK-positive advanced non-small cell lung cancer (NSCLC)[251]. The safety and efficacy of this drug in paediatric patients is not established yet, but the current NANT 2015-02 Phase 1 Study of Lorlatinib (NCT03107988) is in progress to evaluate it for patients with ALK driven relapsed or refractory neuroblastoma.

One of the patients was a girl with relapsed/refractory neuroblastoma with a pathogenic *ALK* F1174L mutation detected in tissue and cfDNA at relapse (timepoint 1 in Figure 6-3). Blood samples were collected monthly throughout her treatment and analysed using ctPC panel. The levels of pathogenic *ALK* F1174L mutation detectable in the blood showed good agreement with the disease course – *ALK* was detected at 30% VAF in a sample prior to the treatment and was not detected in the cfDNA sample collected a month later, following the

commencement of treatment with Lorlatinib. For the next 16 months no mutations by ctPC panel or CNVs by lcWGS were detectable in the monthly cfDNA samples, which was consistent with good clinical response in this patient. Most importantly, the same pathogenic *ALK* mutation remerged at low levels (0.4%-2.7% VAF) and remained detectable for a year before the clinical relapse was detected by magnetic resonance imagining (MRI). A tissue biopsy at relapse again showed the *ALK* F1174L as the dominant genetic change with emergence of *RIT1* and *IGFR1* gains. A time matched cfDNA sample showed the presence of *ALK* mutation, but the focal copy number gains were not detected (*RIT1* is not covered by ctPC panel, *IGFR1* is covered, but was not detected), likely due to low ctDNA purity. After the relapse, the patient received chemotherapy (topotecan/temozolomide) with anti-GD2 immunotherapy and Lorlatinib. Following the treatment, the levels of *ALK* in cfDNA returned to undetectable levels, in agreement with good response to the treatment. No new emerging variants were detected in any of the *ALK* positive cfDNA samples.

To evaluate ctDNA levels and assess the genome wide copy number changes, lcWGS sequencing was performed on cfDNA samples from this patient as well. Multiple copy number changes were detectable in the pre-treatment sample (Figure 6-3 B) and the tumour fraction was estimated at 38% which was in good agreement with pathogenic *ALK* mutation detectable at 30% in the same sample. Throughout the treatment (while the patient was responding well), lcWGS profiles were silent (diploid with no CNVs observed), with an example shown in Figure 6-3 B 2. However, even when *ALK* mutation was detectable by the ctPC panel sequencing, the lcWGS profiles remained silent, highlighting different sensitivities of these methods. In the sample just prior to the relapse, showing the highest ctDNA levels in panel sequencing (*ALK* detected at VAF of 5%), borderline detectability of ctDNA by lcWGS could be seen Figure 6-3 B 3.

It is also important to note that total cfDNA levels (ng cfDNA per ml of plasma) fluctuated highly, irrespective of the disease course (Figure 6-3 A). One notable example, highlighting the importance of quality control metrics and the need to take cfDNA quality into account when interpreting the results is timepoint 4 in Figure 6-3. The extremely high total cfDNA value caused by high molecular weight DNA fragments in the sample diluted the signal of ctDNA and the pathogenic *ALK*

variant was not detectable, even though it was present in cfDNA samples a month earlier and month after.

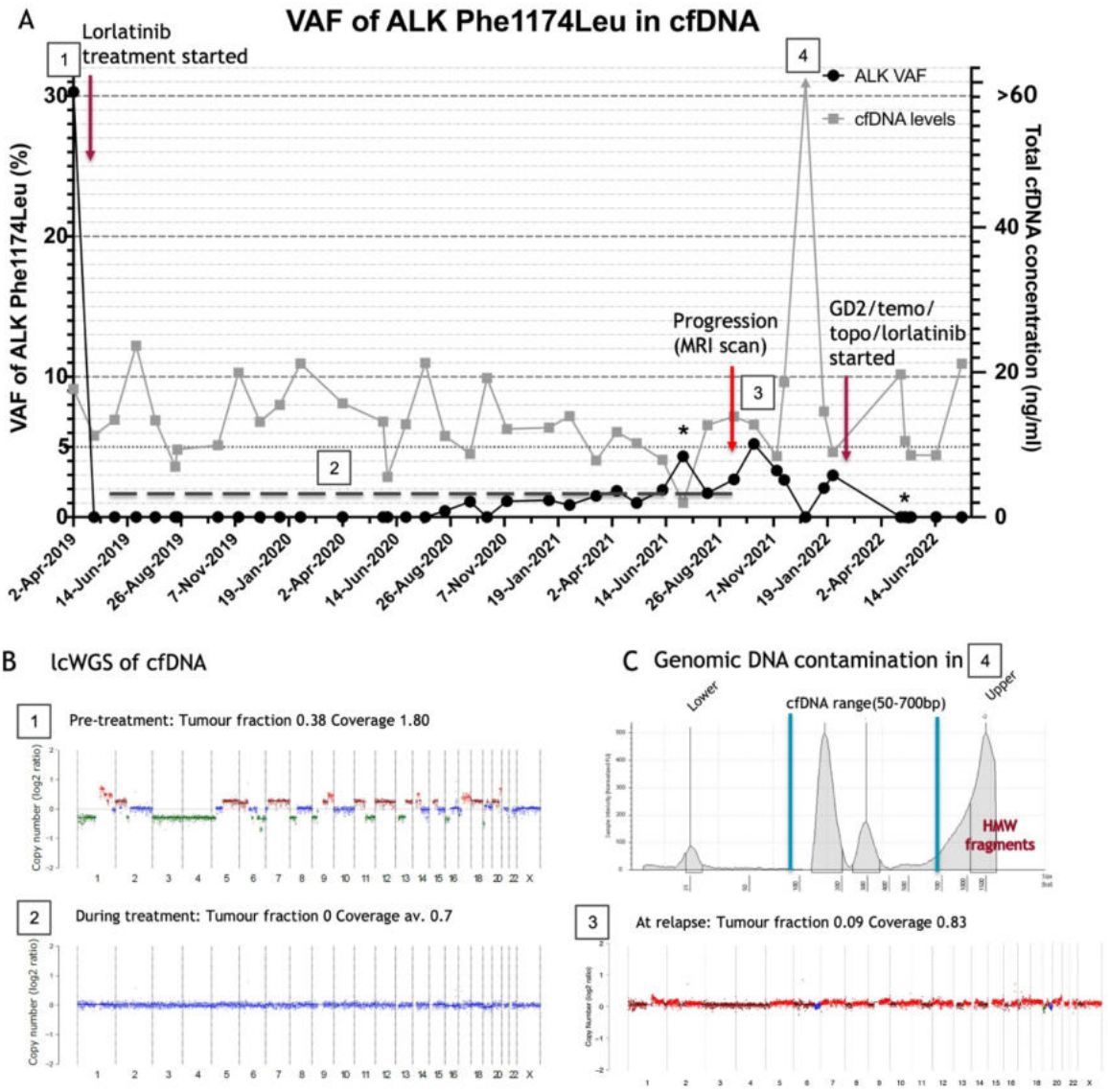


Figure 6-3 Monitoring the response to treatment in a patient with neuroblastoma. (A) Levels of ALK mutation detected by ctPC panel and total cfDNA levels (ng/ml of plasma) in a patient on treatment. The % VAF of ALK mutation detected in cfDNA are shown in back circles with the y-axis on the left. At two timepoints, indicated by asterisks, cfDNA sequencing was of suboptimal depth (<250x), therefore VAF should be interpreted with caution. Total cfDNA levels shown in grey with y-axis on the right. Treatment starting points are indicated by burgundy arrows (Lorlatinib in April 2019 and GD2/temo/topo/lorlatinib in January 2022) and progression as indicated by MRI scan is highlighted in red arrow. (B) Representative plots of lcWGS of cfDNA throughout different treatment points. The timepoints corresponding to the lcWGS profiles are highlighted by numbers 1-3 in the timeline above. During the treatment samples (2) no genome wide copy number changes were observed, most likely due to low ctDNA purity, the average coverage for these silent lcWGS profile samples is indicated by the dashed line in plot (A) (C) Genomic contamination was observed in sample at timepoint (4) as shown in the fragment size profile of Tapestation with the High Sensitivity DNA ScreenTape assay. cfDNA range is indicated in blue, lower and upper assay markers highlighted and high molecular weight (HMW) contamination is indicated in red.

The second patient was a boy with stage M neuroblastoma. No tissue biopsies were available for sequencing, but three separate blood samples collected 3 months apart while patient was experiencing relapse showed the presence of pathogenic *ALK* F1174L mutation in cfDNA (Figure 6-4 A timepoints 1 and 2). The patient was therefore considered for and received Lorlatinib on compassionate access. Similarly, as in the patient discussed above the *ALK* levels were very high in the cfDNA sample collected prior to the start of the treatment (25% VAF) and reduced in the samples collected throughout the treatment. The levels of ctDNA never dropped to undetectable levels in this patient but reduced initially when the patient was responding to treatment. However, 4-5 months after the start of the treatment the patient started progressing with multiple symptoms, including low platelet counts and pain. Increasing *ALK* levels in the blood were consistent with eventual relapse. Unfortunately, blood samples were not available at these later timepoints for analysis. To evaluate ctDNA levels and assess the genome wide copy number changes, lcWGS sequencing was performed on cfDNA samples. Multiple copy number changes were detectable, including several segmental chromosomal alterations, such as partial 1p loss and 17q gain (Figure 6-4 C). The CNV profiles remained very stable throughout the treatment and during progression.

In addition to the pathogenic *ALK* F1174L mutation, several other variants at much lower levels (<1% VAF) were detected (*CDKN2A* S12L, *TP53* R175H, *PPM1D* L484Ter, *BCOR* E1415Ter and a deletion in *ARID1A*) in some but not all cfDNA samples. These variants were also detected in several blood cell pellet DNA samples from the same blood draws at similar levels (0.2-0.8% VAF). This, together with the fact that the levels of these variants did not follow the same pattern as the pathogenic *ALK* variant indicate that these are most likely mutations present in the blood cells of the patient and not tumour-derived.

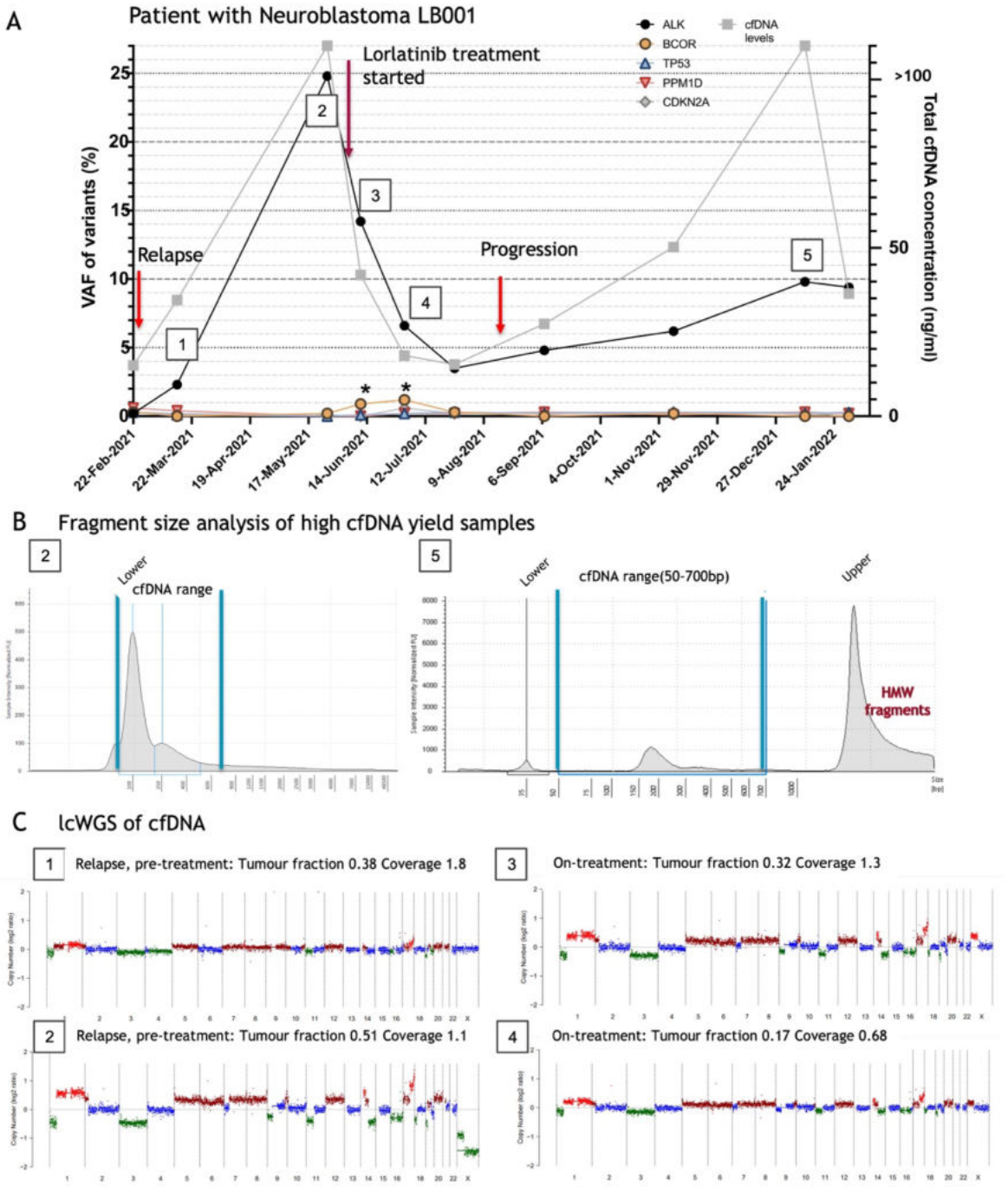


Figure 6-4 Monitoring the response to treatment in a patient with neuroblastoma. (A) Levels of mutations detected by ctPC panel and total cfDNA levels (ng/ml of plasma) in a patient on treatment. The % VAF of ALK mutation detected in cfDNA are shown in back circles with the y-axis on the left. The other low-level mutations in BCOR, TP53, PPM1D and CDKN2A were also detected at similar VAF in the matched blood cell pellet samples from the timepoints indicated by asterisks. Treatment starting point indicated by burgundy arrow. Total cfDNA levels shown in grey with y-axis on the right with two samples showing >100ng/ml cfDNA yield. (B) Fragment size anlaysis by Tapestation in the high cfDNA yield samples. Sample at timepoint (2) showing minimal HMW contamination on the genomic ScreenTape assay. Sample at timepoint (5) showing high levels of HMW conatamination by analysis on cfDNA ScreenTape. C Representative plots of lcWGS of cfDNA throughout different treatment points. The timepoints corresponding to the lcWGS profiles are highlighted by numbers 1-4 in the timeline above.

The total cfDNA levels, as measured by cfDNA amount (ng cfDNA per ml of plasma) were quite consistent with the disease course in this patient (Figure 6-4 A). Of note are two timepoints with extremely high cfDNA yield (>100ng/ml of plasma) – timepoint 2 in Figure 6-4 with extremely high cfDNA levels, and most of the DNA falling in the cfDNA range of 50-770bp (Figure 6-4 B), consistent with high disease burden at that timepoint. In contrast, high cfDNA levels observed in timepoint 5 were mainly due to significant HMW DNA contamination (Figure 6-4 B), consistent with the fact that the patient was experiencing a respiratory infection at the time of blood sample collection. Despite the high levels of HMW DNA in the sample, the pathogenic ALK mutation was detected at VAF of 9%, indicating that in patients with high ctDNA shedding even sub-optimal quality samples can yield useful information.

6.2.3 cfDNA unique variants detected in patients at progression

Twelve ctDNA samples were analysed from six patients (five with neuroblastoma and one with rhabdomyosarcoma) who experienced disease relapse following first line therapy (Figure 6-5). Matched tumour samples were available at 12 ctDNA time points and ctDNA analysis was able to detect all variants detected in tumour where ctDNA purity was >10%. Paired diagnostic and relapse ctDNA analyses showed the emergence of novel ctDNA variants in 5/6 patients, which is consistent with our analysis of Stratified Medicine Paediatric Programme cohort where patients with neuroblastoma tended to have high number of cfDNA unique SNVs detected at relapse (discussed in detail in Chapter 4).

The cfDNA profiling could provide clinically useful information, complementary to the tissue biopsies. In patient 19 (neuroblastoma), an *ATM* mutation at VAF of 7% was detected in the adrenal primary at diagnosis, but not in the post-surgical ctDNA sample. In a subsequent isolated CNS relapse a *SETD2* mutation and the same *ATM* mutation were detected in tumour tissue (12% and 6% VAF). The *SETD2* mutation is not covered by ctDNA panel, but the *ATM* mutation was detected at low levels (0.1%) in time-matched ctDNA sample. Interestingly, additional higher frequency variants were found in ctDNA (*TP53, RB1, FGFR3* and *ARID1A*) that were not detected in the CNS metastasis. The patient subsequently experienced a systemic relapse. In patient 6, *NF1* mutation was detected in the cfDNA at 2nd

relapse at 0.3% and at increasing levels with each subsequent relapse. In tissue profiling, this variant was reported only at 3rd relapse, but when retrospective analysis was done on the tissue sample from 2nd relapse, the variant was detectable, but below the limit of detection for clinical reporting. This highlights the limitations of tissue biopsy molecular profiling and the added value of cfDNA analysis.

More studies are needed to understand the relevance of variants detected at low levels in cfDNA, but when found consistently, might lead to identification of clones resistant to treatment and therefore have the potential to guide further treatment strategies. The potential of ctDNA to aid clinical decision making is illustrated by patient 20 with neuroblastoma, where a *CCND1* amplification was detected in the diagnostic tissue. This was not detected in two relapse tissue samples (from the primary tumour site and a liver metastasis). However, cfDNA from the time of relapse clearly showed the *CCND1* amplification, consistent with the molecular findings in the primary tissue, in addition to a novel *ARID1A* mutation (Figure 6-5 A Patient 20). Both alterations are potentially targetable [252–254].

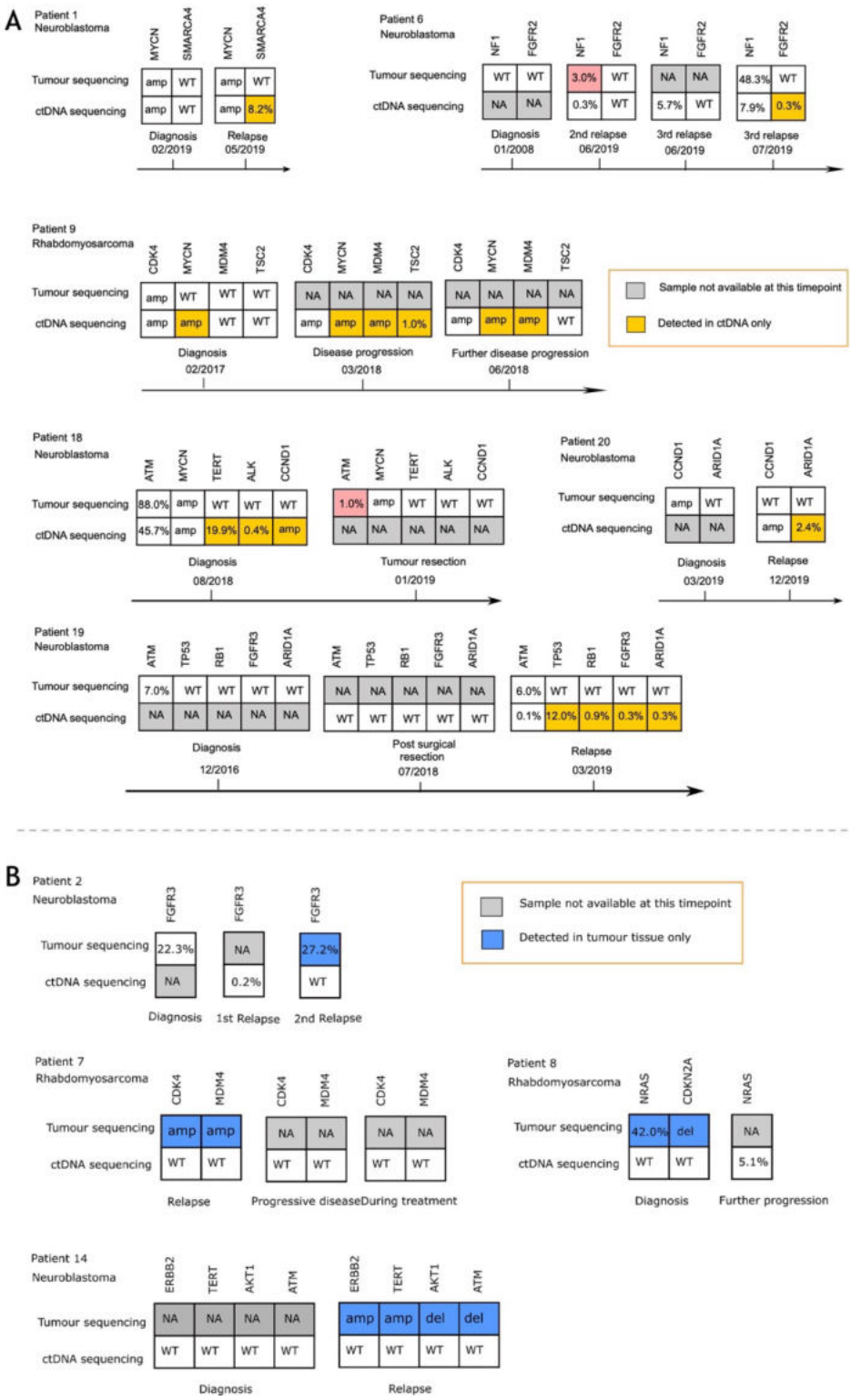


Figure 6-5 Serial liquid biopsy sampling results compared to tissue biopsy at time points where tissue biopsy was available. The top boxes represent tissue sequencing results at a given time point, the bottom ones – cfDNA sequencing. NA – sample was not available. The differences are highlighted in yellow. The samples that had VAF below the limit of detection in tissue sequencing (<5% VAF) are highlighted in red. The variants only detected in the tissue highlighted in blue (A) Serial liquid biopsy sampling results compared to tissue biopsy in high purity samples (estimated ctDNA fraction >10%). For patient 20, two tissue biopsies at relapse were analysed, both showing no variants, therefore merged together. (B) Serial liquid biopsy sampling results compared to tissue biopsy in low purity samples (estimated ctDNA fraction 0%).

6.3 Discussion

One of the key benefits of cfDNA profiling is the potential to monitor disease course in a minimally invasive way. Case studies outlined in this chapter add to the growing evidence for cfDNA profiling potential to be utilised in this way for paediatric cancer patients [90,92,94,99,102,130,132]. This chapter outlined case studies of patients with serial blood sample collection enrolled in several different studies and the ability of cfDNA profiling by ctPC targeted panel to track mutations throughout the treatment was shown. Reduced or disappearing ctDNA levels (as tracked by the detection of mutations using ctPC panel and lcWGS) in patients responding to treatment was shown, in agreement with multiple previous studies [90,92,99,132]. In several patients, increasing ctDNA levels and increasing VAF of the tracked variants were observed at disease progression and/or relapse. Unfortunately, these studies were not designed specifically to test the hypothesis if ctDNA levels track with tumour burden, so detailed clinical data (such as tumour measurements or scans) for comparison was not available, but VAF levels of mutations agreed with clinical disease course as interpreted by the oncologists involved in the studies.

Due to nature of our studies, most of the patients were patients with neuroblastoma, which is known to shed large quantities of ctDNA and therefore simpler to monitor using liquid biopsies. The ability to monitor disease progress in patients with neuroblastoma has already been shown [99,102,187,255], with deep sequencing methods allowing modelling of the tumour evolution [99]. However, previous studies relied on the knowledge of variants present in the tumour to track variants in the cfDNA or on WGS/WES which bares significant costs. By designing our ctPC panel to cover most common drivers and clinically relevant mutations, including known resistance hotspot mutations, we wanted to assess if this can be done in a tissue-agnostic way, and we showed that monitoring is possible using targeted pan-cancer tissue-agnostic panel.

Importantly, we showed the added benefit of cfDNA profiling by detecting several potentially clinically relevant alterations in cfDNA that were either missed by tissue biopsy or the patient didn't have a tissue biopsy at that timepoint (again highlighting the benefit and ease of blood sample collection in comparison to tissue biopsy). Emergence of new cfDNA-unique variants was observed in a patient

with metastatic neuroblastoma who subsequently experienced a systemic relapse (Patient 19 Figure 6-6 A). Interestingly in this patient cfDNA unique variants were detected at higher VAFs than the variant also found in the matched tissue sample. In another patient with neuroblastoma *CCND1* amplification (which was present in tissue at diagnosis) and *ARID1A* mutation were detected in cfDNA, but not in a matched tissue sample at relapse (Patient 20 Figure 6-5). Both alterations are potentially targetable [252–254], highlighting the added value of cfDNA analysis.

In two patients treated with Lorlatinib, reduced ALK mutation levels were observed in patients responding to therapy and increased at disease progression. These results are consistent with a small-scale report of 3 patients with neuroblastoma treated with Lorlatinib, where the pathogenic *ALK* levels in cfDNA were lower (but did not go to 0%) in two patients with good response to the treatment and a 5-fold increase in relative *ALK* levels in a patient with progressive disease in the blood sample 6-8 weeks into the treatment [132]. Interestingly no new mutations were detected at relapse/progression in the cfDNA of the two patients studied closely while on targeted ALK inhibitor treatment in our study. We saw an increase in the VAF of the known pathogenic *ALK* mutation, indicating the increase of ctDNA shedding and presumably the expansion of the pre-existing clone in the tumour. This is in contrast to adult NSCLC where acquired resistance mutations, especially single (G1202R and I1171X) and compound *ALK* mutations (such as C1156Y/L1198F, G1202R/L1196M, I1171N/D1203N) and *NF2* loss of function mutations have been reported [256–258]. For neuroblastoma, *in vitro* and *in vivo* studies revealed *NF1* loss and RAS pathway activating mutations and hyperactivation of *EGFR* and *ERBB4* as resistance mechanisms to Lorlatinib [259,260]. A case study of patient with neuroblastoma relapsing on Lorlatinib treatment showed the emergence of *CDK4* and *FGFR1* amplification and *NRAS* Q61K mutation in cfDNA in addition to the *ALK* F1174L and *PIK3CA* C692Ffs mutations that were present at the first relapse, prior to Lorlatinib treatment [188]. Another study reported emergence of activating RAS pathway mutations (NRAS and HRAS Q61K) in 2 patients progressing after Lorlatinib treatment [260]. A recent study of longitudinal cfDNA profiling in high-risk neuroblastoma patients also included 5 patients on Lorlatinib and mutations potentially bypassing ALK inhibition were detected in cfDNA in multiple patients, including additional mutations in ALK tyrosine kinase domain (L1196M, F1245L, R1275Q, F1174I) and variants in RAS-MAPK pathway (NRAS Q61R, NF1 S641fs*47, CDKN2A E88*) [187]. All these variants

were covered by ctPC panel but were not detected in cfDNA of the patients in our study. A tissue biopsy in one of our patients at the time of relapse showed *RIT1* and *IGFR1* gains, which were not detected in cfDNA (*RIT1* was not covered by the ctPC panel design), most likely due to worse sensitivity of CNV detection in low purity samples. Gains of these genes have not been reported as resistance mechanisms to Lorlatinib, but RIT1 encodes a Ras family GTPase, reported to have a role in various cancers through regulation of RAF/MEK/ERK pathway which is key in ALK signalling [261]. Therefore, gain of *RIT1*, if resulting in significant increase of RAS signalling might confer ALK inhibitor resistance. In this study we did not have the means to investigate this further, but the lack of resolution of CNVs in low purity ctDNA samples highlights the complementary nature of tissue and cfDNA profiling.

The inability to detect the cause of the resistance to treatment in two patients treated with Lorlatinib raise the question if there are unknown or non-genetic resistance mechanisms at play. Currently ongoing NANT 2015-02 Phase 1 Study of Lorlatinib (NCT03107988) for patients with ALK-Driven relapsed or refractory neuroblastoma will provide data on the clinical activity in this patient population and will be able to shed more light on the resistance mechanisms specific to paediatric patients. NANT 2015-02 study is also allowing collection of blood samples for liquid biopsy approaches, therefore the utility of cfDNA ability to monitor the response and potentially detect resistance alterations will be evaluated. If new genetic resistance mechanisms will be discovered, the ctPC panel is easily amendable and new targets can be added to future versions if needed.

Importantly, in one of the patients treated with Lorlatinib the re-emergence of the pathogenic *ALK* mutation in cfDNA preceded the clinical relapse (as indicated by tumour growth on MRI scan) by 14 months (Figure 6-3). In another patient with neuroblastoma, closely monitored while on treatment (Figure 6-2 C) high ctDNA levels were observed after surgery and prior to progression, indicating the ability of cfDNA to detect progression earlier. The limited number of blood samples in intermediate timepoints however limit the conclusions on how much earlier the progression could have been anticipated in this patient. Early molecular relapse by patient-specific or broad commercial cfDNA assays had already been reported, in several case studies [90,92,110,187] but structured studies are needed to address the

clinical utility of early detection of molecular relapse in paediatric patients. Randomised interventional studies are needed to evaluate if changes to treatment based on cfDNA dynamics can improve outcomes [262]. In the two patients studied on targeted inhibitors, we did not detect the emergence of resistance, but if detected, it could potentially inform choice of therapy. However, there is still insufficient evidence that acting on cfDNA findings in these contexts improves outcome in adult or paediatric cancer patients [262].

Additionally, we confirm that cfDNA levels on their own are too variable to allow monitoring of disease progress and depend on other general health factors [263]. In several patients an extreme increase in cfDNA levels (with significant HMW DNA contamination) was observed when patients were known to have infection, in agreement with earlier reports [73,263]. It is important to take this into account when analysing serial samples from a patient, because as shown in one timepoint in patient 110, extremely high HMW levels can dilute the levels of the pathogenic variant (Figure 6-3). This is supported by a recent pan-paediatric cancer study, showing better detection of CNVs in cfDNA samples when quality is high (as determined by the ratio between cfDNA versus HMW in the sample) [106].

Depending on the clinical situation, it is likely that cfDNA analysis will be performed without a time-matched tissue sample. It is well-documented in adult cancers that haematopoietic cells also acquire mutations that could mistakenly be profiled as coming from the tumour tissue in blood derived cfDNA analysis [264]. The accumulation of these mutations increases with age and smoking [265,266] and therefore are less likely to be a significant concern in paediatric cancer patients. Nevertheless, accumulation of mutations driving clonal haematopoiesis during cancer therapy have been reported, including paediatric cancer patients [267–270]. Clonal haematopoiesis mutations associated with previous exposure to cancer therapy were most often observed in the DNA damage response and repair genes (*TP53, PPM1D* and *CHEK2*) [268,269], which are also often mutated in paediatric tumours. Therefore, caution is needed when interpreting low level variants in cfDNA without matched tissue analysis. This can be done by either analysing the blood cell pellet in parallel to cfDNA to exclude CHIP (clonal haematopoiesis of indeterminate potential) derived variants or more rigorous filters on cfDNA at genomic locations commonly associated with CHIP to reduce the risk of interpreting CHIP variants as real tumour-related variants in cfDNA analyses [143,271].

This issue has been highlighted in a recent study monitoring high-risk neuroblastoma patients, where ctDNA surveillance of the levels of three TP53 variants did not correlate with clinical disease evaluation in one of the patients monitored [187]. Given that *TP53* variants are common in clonal haematopoiesis, the authors hypothesised that these *TP53* variants might not been derived from the tumour, but the lack of matching blood cell pellet analysis hindered the ability to prove this. In our study, in one of the patients treated with Lorlatinib, several low VAF variants (*CDKN2A S12L, TP53 R175H, PPM1D L484Ter, BCOR E1415Ter and a deletion in ARID1A*) were detected in cfDNA and after matched blood cell pellet DNA analysis they were deemed to be coming from the haematopoietic cells. *TP53* and *PPM1D* are well-known CHIP genes, highlighting the possibility of clonal haematopoiesis in paediatric cancer patients. The lack of pre-treatment/diagnostic blood sample for this patient however prevented us from knowing if the variants were induced by intense chemotherapy or if they were present in the patient before any cancer therapy.

Overall, in this chapter I have outlined cases studies showing the benefit of serial cfDNA profiling in patients with paediatric solid tumours (mainly neuroblastomas in this cohort) using ctPC capture panel and lcWGS. Longitudinal sample collection and cfDNA molecular profiling allowed close disease course monitoring, with ctDNA levels dropping in patients responding to treatment, remaining relatively unchanged in patients with stable disease and increasing during progression and relapse. Importantly, in some patients increasing ctDNA levels were observed before the clinical signs of progression, indicating the potential for surveillance and future studies evaluating the clinical utility of early detection of molecular relapse are now needed.

Chapter 7 Discussion

In this thesis I have described the development, validation, and the results from application of pan-cancer paediatric solid tumour cfDNA specific NGS capture panel (ctPC) in conjunction with lcWGS on cfDNA from plasma in paediatric patients with solid tumours at diagnosis, relapse and throughout treatment and remission. I have shown that this method can be successfully used to detect SNVs and indels and genome wide CNVs in clinical samples from paediatric patients with solid tumours, monitor the disease progress through serial sample collection and identify genetic variants not detected by tissue sequencing. I have shown good detection of SNVs and indels in cfDNA from plasma in patients with neuroblastoma, rhabdomyosarcoma, Ewing sarcoma, non-Hodgkin lymphoma, hepatoblastoma and other rare solid tumours at relapse. Plasma derived cfDNA profiling showed to be of limited benefit for patients with intracranial tumours therefore I explored and showed the possibility of using ctPC panel and lcWGS on CSF-cfDNA for patients with CNS tumours. By comparing primary and relapse pairs of tissue biopsies from paediatric patients with solid tumours we showed that these tumours undergo significant changes with accumulation of SNVs and CNVs at relapse. The ability to detect these changes in cfDNA was shown in patients with good purity ctDNA, thus cfDNA profiling could potentially be used to monitor tumour evolution in a less invasive way. Case studies of successful monitoring of patients throughout standard and targeted treatment are outlined in this thesis.

7.1 Development and validation of the method

The implementation of liquid biopsies into clinical practice is well underway for adult patients, with multiple commercial assays available for liquid biopsy molecular profiling, such as FDA approved Guardant360CDx and FoundationOne® Liquid CDx, MSK- ACCESS and PredicineCARE™. Given the numerous differences between the most common genetic alterations detected in paediatric and adult cancers [14,16,18,21,44], the implementation of liquid biopsies into clinical management for paediatric cancer patients requires custom approaches. To that end cancer type specific or patient-specific assays, pan-cancer adult focused panels or broad profiling using WES have been used for neuroblastoma [132,150,151,162], glioma[96,152,153] and paediatric sarcomas [89,90,94,160,171]. A sarcoma specific ctDNA panel aiming to

detect translocations and CNVs in a tissue agnostic manner showed 81% sensitivity (detection in 13/16 patients) [171]. Foundation One® liquid CDx panel was used to successfully profile paediatric patients with relapsing/refractory disease and detect molecular alterations in 76% of successfully analysed cfDNA samples [59]. FoundationACT™ assay has been successfully used to profile serial cfDNA samples from patients with neuroblastoma, detecting at least one alteration in 73% of patients studied [187]. The biggest study so far evaluating cfDNA profiling (MAPPYACTS, NCT02613962) in recurrent and refractory extra-cerebral tumours using WES showed successful sequencing results in 62% of patients [15]. However, to our knowledge no pan-cancer tissue-agnostic panels designed specifically for liquid biopsies in paediatric patients have been developed and this project aims to fill this gap. In chapter 3 I described the development and validation of ctPC panel and showed it is highly sensitive, specific, and reproducible for SNV and indel detection at low variant allele fractions (0.125% VAF for tissue-informed and 0.2% VAF for tissue-agnostic variant detection) that are expected in cfDNA analysis. The method was validated to the clinical assay validation standards [157] for SNV and indel detection and validation for CNV and selected fusion detection is ongoing. The current panel is designed to be used as a tool to identify therapeutic targets and help diagnosis and prognostication. The relatively small size of the panel allows cost-effective, high resolution molecular profiling with manageable analysis time, paving the way for implementation into the clinical practice.

The lcWGS analysis of cfDNA is quite well-established technique, allowing detection of ctDNA and showing good concordance between tissue and cfDNA in several paediatric cancer types including neuroblastoma, osteosarcoma, Ewing sarcoma, rhabdomyosarcoma, soft tissue sarcoma and Wilms tumour [89,90,100,106,272]. In this thesis I have shown that lcWGS profiling can be performed together with high depth panel sequencing, without the requirement of additional cfDNA and provide information on genome wide copy number alterations as well as estimation of ctDNA purity. Great concordance between tissue and cfDNA copy number profiles was observed and, in several patients, copy number alterations were detected in cfDNA that were missing in the tissue and vice versa, highlighting that the two bioanalytes can complement each other, as shown in a recent study of patients with neuroblastoma as well [106]. In several patients, cfDNA profiling was

more informative and better represented tissue heterogeneity than single tissue biopsy, highlighting clinical value of this method.

However a big limitation for cfDNA profiling implementation is poor quality and/or low ctDNA levels in some patients. In the studies described in this thesis, ~10% of cfDNA samples suffered from low quality due to the presence of HMW DNA contamination. Even higher numbers of poor quality (as defined by cfDNA/HMW ratio) samples has been reported in a recent study using routinely collected blood samples [106], highlighting the improved results when DNA stabilising tubes are used for sample collection in our studies. In a different study of patients with neuroblastoma at diagnosis (which is one of the most highly ctDNA shedding tumour types in paediatric patients) 9/20 patients were excluded from targeted panel analysis due to low quantity or quality of cfDNA [162]. In MAPPYACTS, 38% of patients did not achieve successful WES of cfDNA, mostly due to low ctDNA levels and technical limitations [15]. However, as highlighted in this thesis, even when sample handling guidelines are adhered to, biological factors might lead to suboptimal cfDNA sample quality. This is not a cfDNA-unique problem, as tissue biopsies also often suffer of insufficient material and low tumour purity: 5-25% of tumour tissue samples were of too low tumour content for analysis in the recent paediatric precision programs [15,44,199,200]. Therefore, it is important to take the sample quality into account when interpreting the results, especially negative results and combine the two modalities – cfDNA and tissue – for better results. Practically, in Chapter 3 I proposed guidelines for streamlining the sample processing based on the quality of cfDNA sample to achieve optimal results (Figure 3-24) in challenging samples.

7.2 Clinically relevant and actionable genomic alterations can be detected in cfDNA

The performance of ctPC panel and lcWGS on cfDNA samples from patients with various solid tumours at diagnosis, throughout the treatment and at relapse was evaluated by comparing the results with tumour tissue profiling. Good concordance between tissue and cfDNA was observed in patients with active disease in patients with high ctDNA purity (>10% ctDNA fraction in cfDNA as determined by lcWGS). Different ctDNA levels were observed in various solid tumours with neuroblastoma shedding highest ctDNA levels into the blood and CNS

tumours having mostly undetectable levels. In high ctDNA purity samples 91% of SNVs and indels detected in tissue were also detected in cfDNA, with high levels of cfDNA unique variants reported across different cancer types. Genome wide copy number alterations were consistent between tissue and cfDNA in the majority of patients with detectable ctDNA.

To study the evolution of paediatric solid tumours, targeted panel and lcWGS analysis of diagnostic and relapse biopsy samples was performed. In agreement to previous studies [14,47,48,50,51], accumulation of mutations and copy number alterations at relapse was observed. No overarching relapse specific copy number signature was observed, only higher incidence of the key alterations, highlighting the universality of these oncogenic changes across different paediatric cancer types. *TP53* was the most commonly mutated gene both in primary and relapse tumours, with enrichment in relapsed samples in this cohort. This could be in part explained by the cohort, which is pre-selected for patients that experienced relapse and therefore potentially had more aggressive disease at presentation. In agreement to previous reports including comparison of primary and relapse pairs [17,42,273], we observed substantial genetic divergence between primary tumours and relapse in some patients. Most importantly, since good agreement between tissue and cfDNA molecular profiling (in samples with detectable ctDNA) was shown, cfDNA could potentially be used to detect and monitor tumour evolution in a less invasive way.

Patients with neuroblastoma had the highest levels of ctDNA, in agreement with other studies [15,90]. Variants expected from tissue sequencing were successfully detected in cfDNA in the majority of patients with active disease and more than a half of all patients studied harboured cfDNA unique variants. At relapse we showed that 68% of variants detected were not detected in time-matched tissue, including pathogenic alterations in *ALK* or RAS-MAPK pathway genes such as *NF1* and *PTPN11*, further validating the importance of these pathways in neuroblastoma and ability to detect clinically relevant alterations in cfDNA. These observations are in perfect agreement with a recent study in high-risk neuroblastoma showing 67% of variants to be cfDNA unique [187]. Similar to that study we also showed cfDNA unique variants to be of lower VAF when compared to variants overlapping between tissue and cfDNA [187], pointing towards subclonal origin of these variants. Multiregional biopsies of patients with neuroblastoma

showed significant spatial and temporal heterogeneity [125,176], which is most likely represented in cfDNA unique variants. Pathogenic missense variants in *ALK* were the most common alteration, detected in both tissue and cfDNA, closely followed by *TP53* (which was cfDNA unique in half of the cases) in patients with neuroblastoma at relapse, which is again in perfect agreement with high risk neuroblastoma cfDNA study[187]. In MAPPYACTS however *ALK* was the most commonly mutated gene in patients with neuroblastoma but no patients carrying *TP53* mutation in cfDNA were reported [15]. In tissue profiling studies, however *ALK* and *TP53* mutations were shown to be common at relapse in patients with neuroblastoma [49–51], supporting their relevance in cfDNA. Copy number alterations are more difficult to detect in cfDNA but poor prognosis markers, *MYCN* amplification, chr1p loss and chr 11q loss, expected from tissue sequencing, were successfully detected in cfDNA in 71%, 100% and 88% of patients at relapse in our study. Overall, in agreement with the literature, we showed high levels of ctDNA, good concordance between tissue and cfDNA and potential added value of detecting cfDNA unique variants in patients with neuroblastoma.

Even in cancer types with lower ctDNA levels than neuroblastoma, we showed good agreement between tissue and cfDNA, ranging from 87% of patients having at least one SNV detected in tissue confirmed in cfDNA in patients with rhabdomyosarcoma, to only 39% in patients with osteosarcoma. The ability to detect SNVs was similar to other studies, except for patients with osteosarcoma, which was reported to have high detection rates in other studies: 57% in newly diagnosed osteosarcoma patients [89], 71% in patients with recurrent/refractory disease [15], 44% in small osteosarcoma study, which included samples on treatment as well [94], compared to 39% in our cohort at relapse. However, we also reported the ability to successfully detect highly fragmented copy number profiles in patients with osteosarcoma by lcWGS, which was previously reported [14]. Additionally, in our cohort Wilms tumours was the only cancer type with no cfDNA unique variants, except chr12 gain in one patient. Wilms tumours are known to be highly heterogenous [274,275] and other studies reported cfDNA unique SNVs and CNVs in patients with Wilms tumours [95,276]. Finally, we showed detection of potentially clinically actionable alterations in very rare cancer types, such as *BRAF* V600E mutation in a patient with Langerhans cell histiocytosis, which has been successfully targeted by selective BRAF inhibitors [277–279], even though it is not yet

approved for this cancer type by MHRA, EMA or FDA. Similarly, ARID1A mutant solid tumours are considered for targeted treatment with EZH2 inhibitor tazemetostat in adults (NCT05023655) and we have reported *ARID1A* mutations in patients with Hodgkin lymphoma and hepatoblastoma. Overall, the results of cfDNA profiling are robust between the different studies, with differences likely due to sampling bias of patient cohorts and future studies should explore this further.

The detection of ctDNA in plasma from patients with CNS tumours is much worse with virtually all plasma cfDNA samples showing silent lcWGS profiles (except few very high focal amplifications) and low levels of mutations detected by panel sequencing (the highest VAF detected in plasma cfDNA was 2% in a patient with high grade glioma at relapse). In the Stratified Medicine Paediatric Programme study, evaluating the feasibility of cfDNA profiling at relapse, only 9 out of 30 patients who had SNVs detected in tissue showed detectable ctDNA levels (with 6% of SNVs expected from the tissue sequencing detectable in cfDNA). This agrees with numerous adult and paediatric studies showing poor variant detection in cfDNA from plasma in patients with CNS tumours [105,222–227], pointing to the fact that there are inherent biological limitations on plasma based cfDNA analysis in these patients, such as limited permeability of the blood brain barrier (BBB). Therefore, alternative sources of cfDNA for these patients are being explored with evidence accumulating that CSF is a good source of cfDNA for adult [105,222–227] and paediatric patients [227,231,232]. In studies described in this thesis, low numbers of CSF samples available limited the extent of evaluation of ctPC panel on CSF-cfDNA, but I showed the ability to detect SNVs and informative lcWGS copy number profile from CSF-cfDNA. Sample quality was an important issue in CSF-cfDNA profiling with low cfDNA yield and uncharacteristic fragment size profiles seen in a high proportion of the patients. This seems to be a common problem with several other CSF-cfDNA profiling studies attributing poor variant detection in some samples to poor sample quality [173,233]. Rigorous studies exploring the pre-analytical conditions of CSF sample processing for liquid biopsies are lacking, mainly due to the difficulty of acquiring CSF samples from patients and the best practices must be inferred from studies in the blood and the limited studies we have in CSF. However, patients with CNS tumours could benefit greatly from less invasive profiling methods and future efforts should be focused on CSF-cfDNA.

Multiple large scale precision medicine programs report the feasibility of NGS molecular profiling in a clinical setting to justify the development of biomarker driven trials for paediatric cancers. These include the European INFORM [204], European MAPPYACTS [15], USA Paediatric MATCH trial [199], USA Genomes for Kids [18], USA iCAT/GAIN [44], Australian Zero Childhood Cancer Program [45] and Dutch iTHER [200]. Extensive tumour tissue profiling, including WGS or WES, RNAseq, methylation and transcriptomic profiling identified potentially clinically impactful variants in 32-86% of the patients included in these studies [15,18,44,45,199,200,204]. The range of results is in part caused by the lack of standardisation of the definition of "an actionable variant", in most studies meaning that the detection of a particular variant can change patient management by informing diagnosis, prognosis or identifying targets for treatment, while other studies only considered variants that would allow enrolment into open clinical trials. Only 4-32% of patients harboured alterations that were acted on [15,18,44,45,199,200,204], with high-impact or "ready for use" alterations detected in only 4-8% of the patients [15,204], highlighting the need for better access to treatments for these patients. Clinical trials, such as the Secured Access–European proof-of-concept therapeutic Stratification trial of Molecular Anomalies in Relapsed or refractory Tumors (AcSé-ESMART) (NCT02813135), exploring targeted treatments and combinations in molecularly enriched paediatric patient populations will be key in advancing the much-needed precision medicine approaches for paediatric cancer patients. The Stratified Medicine Paediatric Programme described in this thesis also included multi-platform analysis in addition to targeted panel and lcWGS sequencing discussed in this thesis, such as RNAseq, WES and methylation profiling for CNS tumours, and full evaluation of potentially actionable alterations is ongoing. However, the ability to detect SNVs by targeted sequencing in cfDNA was demonstrated with 66.7% of alteration deemed actionable in tissue by molecular tumour board also detected in cfDNA. This is in line with MAPPYACTS, where 76% of alterations found to be actionable in tissue profiling were also detected in cfDNA [15].

7.3 Longitudinal cfDNA profiling can successfully monitor disease progress

The minimally invasive nature of blood based cfDNA profiling allows longitudinal sample collection and therefore opens the possibilities for disease monitoring. A number of studies highlighted case studies where tracking a known genetic

variant, or a number of variants, in cfDNA correlated well with clinical disease course in patients with neuroblastoma [99,101,102,132,187,255], Ewing sarcoma [90,92,130,280], rhabdomyosarcoma [90,110,281], osteosarcoma [90,94], malignant peripheral nerve sheath tumour [272], hepatoblastoma [282] and medulloblastoma using CSF-cfDNA [231]. In chapter 6 I have outlined several case studies from our cohort adding to this evidence. We showed that cfDNA levels are variable and depend on various physiological and biological conditions, but tracking variants specific to ctDNA using tissue-agnostic panel can be very informative. Reducing or disappearing AF levels of mutations in cfDNA in patients responding to treatment and increasing VAF of the variants in cfDNA in patients who progress and/or relapse were reported. Importantly, I show the ability to detect and track clinically relevant variants in a tissue-agnostic way using ctPC panel, as highlighted by patient LB001 (Figure **6-4**, chapter 6) where detection of *ALK* F1174L resulted in access to targeted inhibitor treatment and good, albeit short, response. Importantly, in another patient the re-emergence of *ALK* mutation in cfDNA preceded clinical relapse (as indicated by tumour growth on MRI) by 14 months. Structured studies are needed to determine the clinical utility of early molecular relapse in paediatric patients, but the ability to detect relapse using cfDNA earlier than conventional approaches has been reported in various paediatric cancers [90,92,94,110,171,187,272]. As the evidence accumulates, randomised interventional studies are needed to evaluate if interventions based on cfDNA dynamics can improve outcomes [262].

Chapter 6 outlined the results of extensive monitoring of two patients with neuroblastoma who carried pathogenic *ALK* mutation and therefore were treated with targeted ALK inhibitor Lorlatinib. We hypothesised that cfDNA profiling using ctPC panel would allow detection of the emergence of resistance alterations (mutations in RAS-MAPK pathway, *CDK4* and/or *FGFR1* amplifications) if these patients would experience relapse, as reported in several case studies [188,260]. However, no new mutations were detected at relapse/progression in the cfDNA in two patients in our study. Instead, an increase of the VAF of the tracked *ALK* mutation was observed prior to relapse, indicating the increase of ctDNA shedding and presumably the expansion of the pre-existing clone in the tumour. It is important to acknowledge, that a targeted tissue-agnostic capture panel was used to monitor the known variants in cfDNA and evaluate the potential to detect novel

variants. Therefore, the absence of new resistance mutations in relapsing patients on targeted treatment could also be a result to the limited scope of the panel. WES could be considered in these cases, however it would only be feasible and informative in patients with high ctDNA fraction. Bigger studies are needed to understand all possible resistance mechanisms in these patients and the current NANT 2015-02 Phase 1 Study of Lorlatinib (NCT03107988) for patients with ALK-driven relapsed/refractory neuroblastoma will provide this information.

Generally, two cfDNA analysis approaches are available for the monitoring of disease course – tumour-informed and tumour-agnostic, each with their own strengths and limitations. Tumour-informed panels (constructed by sequencing tumour biopsy and selecting the variants to follow) suffer from a significant lag time between tumour testing and ctDNA panel creation and therefore are logistically difficult to apply. Tumour agnostic panels might end up missing some rare variants, but they offer easier implementation and economy of scale. WES is a broad and tumour agnostic method which could be an attractive alternative in high ctDNA purity samples, but of limited benefit for minimal residual disease monitoring at the moment. However, new technologies are emerging, showing the potential of lower-cost WGS (Ultima Genomics) for tumour burden tracking and residual disease detection by genome wide mutational integration of thousands of somatic mutations [194,283]. In the future studies, a combination of these approaches might be most useful. For example, broad profiling at diagnosis (or relapse if it is when the monitoring would be started), by deep sequencing with targeted panel or WES/WGS (if high ctDNA purity is indicated by lcWGS) could be followed by either panel or individual assays to track the variants detected. At the next relapse (or increase of ctDNA levels) a broader scope profiling could be performed again to identify any novel variants indicating possible emergence of resistance. WGS/WES is an excellent approach in research setting allowing better understanding of the tumour biology and potentially identifying new drivers. In the clinical setting however, where clinically relevant mutations are predefined by the treatments available, tumour agnostic panel sequencing (such as ctPC described in this thesis) is fit for purpose.

7.4 Implementation into clinical practice and future directions

The clinical implementation of a new assay into clinical practice requires multiple steps (Figure 7-1) [60]. Firstly, the ability of an assay to measure the analyte of interest reliably and accurately as evaluated according to the assay's sensitivity, specificity, reliability, and robustness (analytical validity) and the ability to accurately measure the clinical feature of interest as evaluated by clinical sensitivity, specificity, and positive and negative predictive values (clinical validity) needs to be shown [60]. Then the evidence of improved outcomes (clinical utility) compared with standard methods needs to be collected to obtain evidence for regulatory approval [60]. Once the evidence is available, the cost-effectiveness of the assay needs to be evaluated and once incorporated into the guidelines, the assay can be incorporated into clinical workflows [60]. Accreditation and regulatory organisations like the FDA in the US, the MHRA in the UK and the EMA in Europe will require standardisation of sampling and robust clinical validation which is likely to limit the number of laboratories offering cfDNA sequencing as a clinical assay. Commercial assays are beginning to make the transition from clinical trials into standard of care testing, either through send-away testing or through large regional accredited laboratories like the Genomic Laboratory Hubs in England. However, the commercial market for paediatric tumour testing is limited, therefore focussed tests are likely to continue to depend upon translation of academic research assays.

Figure 7-1 Steps needed for integration of a liquid biopsy assay into the clinical practice. Figure adapted from [60].

Liquid biopsy assays have multiple potential clinical applications - screening, characterisation of disease, identification of targets for treatment, detection of molecular residual disease, prediction of relapse, assessment of response to

treatment and detection of resistance at progression. The FDA draft guidance for the use of cfDNA in drug design and clinical trial design for early-stage solid cancers supports the use of cfDNA to detect targetable alterations, to select patient population harbouring genetic or epigenetic alterations that could guide eligibility and act as a proxy endpoint shortening the length of clinical trials, by reflecting response to treatment or to evaluate treatment efficacy [284]. Each of these applications have different limitations and require appropriate assays that fit the specific clinical situation best. For example, early disease detection, monitoring of patients on treatment and detection of molecular residual disease requires high sensitivity assays that work reliably with limited amount of cfDNA, as cfDNA levels are expected to be low in these situations. Patient specific ddPCR assays or methylation profiling are likely candidate assays in these cases. On the other hand, in patients with advanced disease broader cfDNA profiling methods capable of novel variant detection to identify emergence of new drivers and resistance mechanisms are needed. Targeted panel sequencing falls in between the two extremes, providing high sensitivity of variant detection (excellent for SNVs, lower sensitivity for CNV and SV detection in low purity samples) and allowing detection of multiple tissue-agnostic alterations.

To this end, in this thesis I have outlined the development and validation of the assay showing that analytical validity of ctPC capture panel is up to the clinical standards. I showed the clinical validity for the detection of SNVs expected from tissue genetic profiling at relapse (if tissue-based profiling is considered as the gold standard to compare to) in patients with neuroblastoma, rhabdomyosarcoma, and Ewing sarcoma. For rare cancer types more evidence still needs to be collected, but cfDNA profiling showed good indication of clinical validity for patients with non-Hodgkin lymphoma, some other soft tissue sarcomas and hepatoblastoma in our study. This is supported by other studies, where ability to detect *CTNNB1* driver mutations in cfDNA has been demonstrated in patients with hepatoblastoma [282], and tumour specific alterations were detected successfully in cfDNA from patients with rare types of sarcomas [15]. As the weight of evidence for the technical utility of liquid biopsy assays increases, we next step is to address where the greatest potential is for added benefit in patients.

Additionally, in chapter 6 I outlined case studies showing the evidence of clinical validity of our method for detection of molecular residual disease, early prediction

of relapse, assessment of response to treatment and detection of resistance at progression. We still need more information to understand how much earlier and with what confidence emerging relapse can be detected in cfDNA when compared to standard surveillance methods and interventional clinical trials, acting on cfDNA profiling results are needed to evaluate the clinical utility in the context of different disease types. For example, for patients with sarcomas, which are often driven by fusions, a combination of SNV and pan-sarcoma fusion detection panel possibly would be most appropriate [171]. A combination of fragmentation and cancer-specific chromatin signatures [160] or methylation [173,239] analysis could be of more benefit in patients with low mutation burden. Patients with CNS tumours face particular challenges for liquid biopsy implementation, but CSF-cfDNA is emerging as a biomarker worthy of further evaluation with initiatives such as the SIOP High Risk Medulloblastoma (89) and SIOP II Ependymoma (NCT02265770) underway.

The reporting and variant interpretation is another key aspect of translation of cfDNA assays to the clinic. I have outlined the guidelines for SNV reporting, but quite high background noise in the cfDNA analyses is still a hurdle, which leads to the need of extensive manual variant curation. It is manageable in the research setting, but to implement this test to the clinic, improved background suppression methods would be needed to reduce the number of variants that need to be reviewed and manually interpreted. To this point, even though WES/WGS provide more information, the clinical scientist and treating clinicians would need to be trained to interpret this additional information. Therefore, a panel-based approach, focused on clinically relevant variant detection would be easier and more realistic to implement.

An inherent limitation of the sequencing approaches discussed in this thesis is low mutational burden in some paediatric tumours, especially CNS tumours, which limits the number of alterations to assess and hence the sensitivity of the assays. One way to increase the sensitivity of molecular profiling tests is to increase the number of targets analysed. To that end, other forms of molecular analysis such as epigenetic (methylation, fragmentation, and nucleosome occupation analyses), metabolomic and transcriptomic profiling have the potential to provide useful information and has been successfully performed in cfDNA from paediatric cancer patients [106,160,172,173,213,214,238,239]. In paediatric sarcoma patients, ctDNA detection

and classification based on cancer-specific chromatin signatures and independent of genomic alterations has been successful [160]. Similar approach has been successful in CSF-cfDNA in paediatric medulloblastoma patients where epigenetic signatures were similar in tissue and CSF-cfDNA, and dynamic changes in methylation of signature clusters were reported with reduction of methylation levels of cfDNA in patients responding to treatment [173]. A proof-of-concept study using reduced representation bisulphite sequencing on cfDNA from plasma and CSF successfully classified 81% of samples from a range of paediatric tumours [172]. Another approach would be to combine cfDNA analysis with protein biomarkers as done by CancerSEEK [285] or combine targeted panel sequencing with methylation and WGS assays as done by GRAIL early detection assays [286]. If validated in larger cohorts of patients, these technologies may be of relevance in certain paediatric brain tumours such as medulloblastoma and ependymoma which have a low mutational burden.

Overall, in this thesis I have shown that ctDNA is a promising biomarker in paediatric cancer care and can be reliably detected using a targeted NGS panel (ctPC panel). In conjunction with lcWGS the approach is suitable for genetic molecular profiling and detection of clinically relevant alterations in many patients with paediatric extra-cranial solid tumours. I have outlined case studies showing the potential to monitor disease course and study tumour evolution in a minimally invasive way. Moving forward, molecularly enriched/predictive biomarker-driven interventional clinical trials are needed. Integration of comprehensive tissue and cfDNA based molecular profiling with better access to treatment hopefully will improve outcomes for paediatric cancer patients.

Appendix

Table-Appendix-1 Regions showing consistently poor capture in validation runs in HD and Promega control samples. Probe name indicates the gene, exon number or in case of copy number probes, rs number of the SNP that the probe covers. Only probes failing in more than 1 sample out of 10 are shown

Probe name	Frequency of failure	Probe name	Frequency of failure
ATM_CN_rs583725	**10**	ATRX_mut_22	**5**
ATRX_mut_13	**10**	PIK3CA_mut_13	**5**
ATRX_mut_8	**10**	PTEN_mut_9	**5**
BCOR_mut_13	**10**	STAG3_mut_10	**5**
FGFR1_CN_rs2932005	**10**	ATRX_CN_rs3027544	**4**
HIST2H3C_mut_1	**10**	RB1_mut_6	**4**
KIT_CN_rs4864919	**10**	ATRX_CN_rs45567038	**3**
MDM4_CN_rs4252685	**10**	ATRX_mut_21	**3**
RB1_CN_rs4600372	**10**	ATRX_mut_33	**3**
STAG2_mut_23	**10**	PIK3CA_mut_12	**3**
ATM_CN_rs227053	**9**	ATRX_mut_10	**2**
ATRX_CN_rs3027546	**9**	ATRX_mut_18	**2**
ATRX_mut_15	**9**	ATRX_mut_20	**2**
ATRX_mut_4	**9**	BCOR_mut_3	2
STAG2_mut_22	**9**	RB1_CN_rs1951775	2
STAG2_mut_35	**9**	STAG2_mut_1a	2
STAG2_mut_4	**9**	STAG2_mut_3	2
STAG2_mut_7	**9**		
ATRX_mut_24	**8**		
BCOR_mut_5	**8**		
STAG2_mut_25	**8**		
STAG2_mut_34	**8**		
ATM_CN_rs650128	**7**		
MET_CN_rs38848	**7**		
STAG2_mut_2	**7**		
ATRX_mut_29	**6**		

Table Appendix-2 Comparison of single nucleotide variants expected from the tumour sequencing and detected in the plasma sequencing using ctPC panel. Only the genes present in both tissue and ctDNA panels are shown. The cohort is not perfectly time matched and the number of days between the tissue biopsy and the blood draw for liquid biopsy is shown in the table. Treatment position of each patient at each sample collection time point is indicated and the allelic frequency of each variant is shown. ctDNA purity as estimated by lcWGS for each sample is shown in the last column. Variants detected only in ctDNA that were not present in the tissue are show separately with corresponding AFs.

Patient ID	Cancer type	Days between tissue and liquid biopsy	cfDNA yield (ng/ml plasma)	Treatment position of tumour tissue sample	Gene	Variant detected (amino acid change/ copy number change)	Variant detected (mutation c.)	AF in tumour	Treatment position of plasma sample	AF in plasma	Additional variants in plasma (VAF >0.003)	AF of additional variants in plasma	ctDNA purity
2	Neuroblastoma	16	11.33	Relapse	FGFR3	p.Glu140Lys	c.418G>A	0.27	Relapse	0.002			0.00
3	Neuroblastoma	19	44.36	Diagnostic	MAP2K1	p.Lys57Asn	c.171G>T	0.508	On induction therapy	0.031			0.00
4	Rhabdomyosarcoma	37	8.89	Diagnostic	NF1	p.Leu1411Val	c.4231C>G	0.61	Diagnostic	0.254			0.16
4	Rhabdomyosarcoma	37	8.89	Diagnostic	PIK3CA	p.Glu542Lys	c.1624G>A	0.3	Diagnostic	0.226			0.16
4	Rhabdomyosarcoma	37	8.89	Diagnostic	TP53	p.Arg196Ter	c.586C>T	0.56	Diagnostic	0.199			0.16
6	Neuroblastoma	10	10.32	Relapse	NF1	NA	c.2002-1G>A	0.48	Relapse	0.057	FGFR2 p.Ala53Val c.158C>T	0.003	0.13
8	Rhabdomyosarcoma	NA	19.00	4th relapse/ progression	NRAS	p.Gln61Arg	c.182A>G	0.42	5th episode progression	0.051			0.00
13	Ewing sarcoma	84	468.95	1st relapse/ progression	TP53	p.Cys176Tyr	c.527G>A	0.87	Disease progression	0.571			0.61
13	Ewing sarcoma	84	468.95	1st relapse/ progression	ATM	p.Lys1114Glu	c.3340A>G	0.06	Disease progression	0.000			0.61

15	Adrenocortical carcinoma	138	19.46	Diagnostic	NF1	p.Ser636Cys	c.1907C>G	0.63	Disease progression	0.005			0.00
16	Rhabdomyosarcoma	NA	24.49	Diagnostic	FGFR4	p.Asn535Lys	c.1605C>A	0.32	Relapse after 2 lines of therapy	0.109	ARID1A p.Ser1352Gly c.405A>G	0.111	0.16
17	Neuroblastoma	1139	17.69	1st relapse	ALK	p.Phe1174Leu	c.3522C>A	0.17	5th relapse	0.303			0.37
18	Neuroblastoma	9	264.20	Diagnostic	ATM	p.Cys1286Phe	c.3857G>T	0.88	Diagnostic	0.457			0.62
18	Neuroblastoma	9	264.20	Diagnostic	ALK	p.Arg1275Gln	c.3824G>A	0 in matched; 4% in later sample	Diagnostic	0.004	TERT g.1295250G>A	0.199	0.62
21	Malignant neoplasm of kidney	33	17.22	Diagnostic	CTNNB1	p.Thr41Ala	c.121A>G	0.42	Diagnostic	0.013			0.00
22	Wilms tumour	5182	12.31	Diagnostic	FBXW7	p.Ser476ArgfsTer11	c.1427_1428insGTTCTCG	0.18	5th relapse/progression	0.006			0.00
23	Carcinoma (other)	575	22.88	Diagnostic	SMARCB1	p.Arg374Gln	c.1121G>A	0.13	3rd relapse	0.113			0.12
24	Neuroblastoma	620	5568.75	Relapse	ALK	p.Arg1275Gln	c.3824G>A	0.08	Relapse	0.001	ATM p.Val2906Ile c.8716G>A and SUFU p.Gly112Glu c.335G>A	0.083 and 0.014	0.59
27	Wilms tumour	1260	15.08	Diagnostic	TP53	p.Arg181Cys	c.541C>T	0.86	4th relapse	0.057			0.00
28	Neuroblastoma	NA	180.75	Diagnostic	ALK	p.Arg1275Gln	c.3824G>A	positive	Diagnostic	0.055			0.64
29	Rhabdomyosarcoma	308	28.77	Diagnostic	BCOR	p.Phe1419fs	c.4257delC	0.16	Disease progression	0.106			0.27
29	Rhabdomyosarcoma	308	28.77	Diagnostic	ARID1A	p.Trp1545Ter	c.4634G>A	0.05	Disease progression	0.000			0.27
31	Wilms Tumour	27	7.56	Recurrence	AMER1	p.Arg353Ter	c.1057C>T	0.49	Recurrence	0.000			0.00

31	Wilms Tumour	27	7.56	Recurrence	BCOR	p.Ile1324fs	c.3972delT	0.61	Recurrence	0.000			0.00
5	Hepatoblastoma	160	24.86	Diagnostic	BCOR	p.Asn1459Ser	c.4376A>T	0.4	During relapse therapy	0			0
5	Hepatoblastoma	160	24.86	Diagnostic	CREBBP	p.Arg75*	c.223C>T	0.35	During relapse therapy	0			0
19	Neuroblastoma	573	14.31	Diagnostic	ATM	p.Ile1407Thr	c.4220T>C	0.07	Post surgical resection	0			0
30	Ependymoma of brain	1126	14.99	Diagnostic	PIK3CA	p.Gly106Arg	c.316G>C	0.37	6th relapse/progression	0			0
30	Ependymoma of brain	1126	14.99	Diagnostic	ATM	p.Asp2650Glu	c.7950C>A	0.37	6th relapse/progression	0			0
30	Ependymoma of brain	1126	14.99	Diagnostic	TP53	p.Phe134Leu	c.400T>C	0.05	6th relapse/progression	0			0
32	Posterior fossa astrocytoma	1498	20.39	Diagnostic	ATRX	p.Pro694fs	c.2081delC	0.24	Post drug, residual disease (after progression)	0			0
33	Rhabdomyosarcoma	682	13.13	Diagnostic	PIK3CA	p.Asn1044Lys	c.3132T>A	0.3	Remission plasma sample	0			0
33	Rhabdomyosarcoma	682	13.13	Diagnostic	MYOD1	p.Leu122Arg	c.365T>G	0.51	Remission plasma sample	0			0

34	Xanthoastrocytoma	148	15.63	Diagnostic	BRAF	p.Val600Glu	c.1799T>A	0.31	On treatment plasma sample	0			0
37	Germinoma	1455	6.39	Diagnostic	PDGFRA	Amplification		Amplification	2nd relapse (lateral ventricle and 4th ventricle)	0			0
37	Germinoma	1455	6.39	Diagnostic	KIT	Amplification		Amplification	2nd relapse (lateral ventricle and 4th ventricle)	0			0
39	Xanthoastrocytoma	105	12.57	Diagnostic	ATM	Deletion		Deletion	After 2 cycles of chemo (disease responding)	0			0
40	Neuroblastoma	245	173.28	Diagnostic	MYCN	Amplification		Amplification	End of induction chemotherapy	0			0

Table Appendix-3 Molecular profiling results for patients with ctDNA and/or tissue biopsies taken at more than one time-point. For cfDNA samples, purity was estimated from lcWGS using ichorCNA). Each sample from the same patient separated by shading.

Patient ID	Sample type	Date taken	Timepoint	Type of cancer	Gene	Variant detected (amino acid change/ copy number change)	Variant detected (mutation c.)	Variant allele frequency	ctDNA purity estimate (lcWGS)
1	DNA from tissue - oral lesion	2019	Diagnostic	Neuroblastoma	MYCN	amplification		amplification	NA
1	cfDNA	17/02/2019	Diagnostic	Neuroblastoma	MYCN	amplification		amplification	0
1	cfDNA	04/04/2019	On induction chemotherapy	Neuroblastoma	MYCN	amplification		amplification	0
1	cfDNA	14/05/2019	Refractory to induction chemotherapy	Neuroblastoma	SMARCA4	p.(Met202Ile)	c.606G>A	0.082	0.172
1	cfDNA	14/05/2019	Refractory to induction chemotherapy	Neuroblastoma	MYCN	amplification		amplification	0.172
1	DNA from tissue - supra-renal mass	14/05/2019	Refractory to induction chemotherapy	Neuroblastoma	SMARCA4	p.(Met202Ile)	c.606G>A	0.000	0.266
1	DNA from tissue - supra-renal mass	14/05/2019	Refractory to induction chemotherapy	Neuroblastoma	MYCN	amplification		amplification	0.266
2	DNA from tissue	06/12/2013	Diagnostic	Neuroblastoma	FGFR3	p.(Glu140Lys)	c.418G>A	0.223	NA
2	cfDNA	19/02/2018	First relapse	Neuroblastoma	FGFR3	p.(Glu140Lys)	c.418G>A	0.002	0
2	DNA from tissue	05/05/2020	Relapse	Neuroblastoma	FGFR3	p.(Glu140Lys)	c.418G>A	0.272	0.548

2	cfDNA	21/05/2020	Relapse	Neuroblastoma	FGFR3	p.(Glu140Lys)	c.418G>A	0.000	0
3	cfDNA	22/05/2019	Diagnostic	Neuroblastoma	MAP2K1	p.(Lys57Asn)	c.171G>T	0.060	0
3	DNA from tissue	10/06/2019	Diagnostic	Neuroblastoma	MAP2K1	p.(Lys57Asn)	c.171G>T	0.508	NA
3	cfDNA	28/08/2019	Refractory to first and second line therapy	Neuroblastoma	MAP2K1	p.(Lys57Asn)	c.171G>T	0.048	0
3	cfDNA	02/09/2020	Refractory to third line therapy	Neuroblastoma	MAP2K1	p.(Lys57Asn)	c.171G>T	0.031	0
4	DNA from tissue	14/06/2017	Diagnostic	Rhabdomyosarcoma	NF1	p.(Leu1411Val)	c.4231C>G	0.610	NA
4	DNA from tissue	14/06/2017	Diagnostic	Rhabdomyosarcoma	PIK3CA	p.(Glu542Lys)	c.1624G>A	0.300	NA
4	DNA from tissue	14/06/2017	Diagnostic	Rhabdomyosarcoma	TP53	p.(Arg196Ter)	c.586C>T	0.560	NA
4	cfDNA	21/07/2017	Diagnostic	Rhabdomyosarcoma	NF1	p.(Leu1411Val)	c.4231C>G	0.254	0.162
4	cfDNA	21/07/2017	Diagnostic	Rhabdomyosarcoma	PIK3CA	p.(Glu542Lys)	c.1624G>A	0.226	0.162
4	cfDNA	21/07/2017	Diagnostic	Rhabdomyosarcoma	TP53	p.(Arg196Ter)	c.586C>T	0.199	0.162
4	cfDNA	24/11/2017	Post surgery	Rhabdomyosarcoma	NF1	p.(Leu1411Val)	c.4231C>G	0.000	0
4	cfDNA	24/11/2017	Post surgery	Rhabdomyosarcoma	PIK3CA	p.(Glu542Lys)	c.1624G>A	0.000	0
4	cfDNA	24/11/2017	Post surgery	Rhabdomyosarcoma	TP53	p.(Arg196Ter)	c.586C>T	0.000	0
6	DNA from tissue	09/01/2008	Diagnostic	Neuroblastoma	NF1	NA	c.2002-1G>A	0.000	0.807
6	DNA from tissue	05/05/2016	2nd relapse	Neuroblastoma	NF1	NA	c.2002-1G>A	0.030	NA

6	cfDNA	27/06/2016	On treatment for 2nd relapse	Neuroblastoma	NF1	NA	c.2002-1G>A	0.003	NA
6	DNA from tissue	09/07/2019	3rd relapse	Neuroblastoma	NF1	NA	c.2002-1G>A	0.484	0.894
6	DNA from tissue	09/07/2019	3rd relapse	Neuroblastoma	FGFR2	p.(Ala53Val)	c.158C>T	0	0.894
6	cfDNA	19/06/2019	3rd relapse	Neuroblastoma	NF1	NA	c.2002-1G>A	0.057	0.129
6	cfDNA	06/08/2019	3rd relapse	Neuroblastoma	NF1	NA	c.2002-1G>A	0.080	0.169
6	cfDNA	06/08/2019	3rd relapse	Neuroblastoma	FGFR2	p.(Ala53Val)	c.158C>T	0.003	0.169
7	cfDNA	27/07/2017	On treatment progression	Rhabdomyosarcoma	CDK4	amplification		Not detected	0
7	cfDNA	27/07/2017	On treatment progression	Rhabdomyosarcoma	MDM4	amplification		Not detected	0
7	DNA from tissue	07/08/2017	Relapse	Rhabdomyosarcoma	CDK4	amplification		amplification	NA
7	DNA from tissue	07/08/2017	Relapse	Rhabdomyosarcoma	MDM4	amplification		amplification	NA
7	cfDNA	17/01/2018	Post drug, reassessment after 6 cycles of chemo - progressive disease	Rhabdomyosarcoma	CDK4	amplification		amplification	0
7	cfDNA	17/01/2018	Post drug, reassessment after 6 cycles of chemo - progressive disease	Rhabdomyosarcoma	MDM4	amplification		amplification	0
7	cfDNA	27/04/2018	Post drug, partial response to relapse therapy	Rhabdomyosarcoma	CDK4	amplification		amplification	0
7	cfDNA	27/04/2018	Post drug, partial response to relapse therapy	Rhabdomyosarcoma	MDM4	amplification		Not detected	0
8	DNA from tissue	01/11/2016	Diagnostic	Rhabdomyosarcoma	NRAS	p.(Gln61Arg)	c.182A>G	0.420	NA

8	DNA from tissue	01/11/2016	Diagnostic	Rhabdomyosarcoma	CDKN2A	deletion		deletion	NA
8	cfDNA	20/10/2017	4th episode disease progression	Rhabdomyosarcoma	NRAS	p.(Gln61Arg)	c.182A>G	0.000	0
8	cfDNA	20/10/2017	4th episode disease progression	Rhabdomyosarcoma	CDKN2A	deletion		Not detected	0
8	cfDNA	18/01/2018	Post drug, further (5th episode) progression	Rhabdomyosarcoma	NRAS	p.(Gln61Arg)	c.182A>G	0.051	0
8	cfDNA	18/01/2018	Post drug, further (5th episode) progression	Rhabdomyosarcoma	CDKN2A	deletion		Not detected	0
9	DNA from tissue	25/01/2017	Diagnostic	Rhabdomyosarcoma	CDK4	amplification		amplification	NA
9	cfDNA	10/02/2017	Diagnostic	Rhabdomyosarcoma	CDK4	amplification		amplification	0
9	cfDNA	10/02/2017	Diagnostic	Rhabdomyosarcoma	MYCN	amplification		amplification	0
9	cfDNA	26/03/2018	Disease progression - commenced on VIT trial	Rhabdomyosarcoma	CDK4	amplification		amplification	0
9	cfDNA	26/03/2018	Disease progression - commenced on VIT trial	Rhabdomyosarcoma	MDM4	amplification		amplification	0
9	cfDNA	26/03/2018	Disease progression - commenced on VIT trial	Rhabdomyosarcoma	MYCN	amplification		amplification	0
9	cfDNA	26/03/2018	Disease progression - commenced on VIT trial	Rhabdomyosarcoma	TSC2	p.(Arg1409Trp)	c.425C>T	0.010	0
9	cfDNA	11/06/2018	Further progression while on VIT trial	Rhabdomyosarcoma	CDK4	amplification		amplification	0.1304
9	cfDNA	11/06/2018	Further progression while on VIT trial	Rhabdomyosarcoma	MDM4	amplification		amplification	0.1304

9	cfDNA	11/06/2018	Further progression while on VIT trial	Rhabdomyosarcoma	MYCN	amplification		amplification	0.1304
9	cfDNA	11/06/2018	Further progression while on VIT trial	Rhabdomyosarcoma	TSC2	p.(Arg1409Trp)	c.425C>T	0.000	0.1304
10	cfDNA	15/04/2015	3rd relapse	Neuroblastoma	ATM	p.(Trp3055Ser)	c.9164G>C	0.016	0
10	cfDNA	15/04/2015	3rd relapse	Neuroblastoma	EGFR	p.(Thr446Pro)	c.1336A>C	0.010	0
10	cfDNA	15/04/2015	3rd relapse	Neuroblastoma	MYCN	amplification		amplification	0
10	cfDNA	22/12/2015	4th relapse	Neuroblastoma	ATM	p.(Trp3055Ser)	c.9164G>C	0.494	0.499
10	cfDNA	22/12/2015	4th relapse	Neuroblastoma	EGFR	p.(Thr446Pro)	c.1336A>C	0.165	0.499
10	cfDNA	22/12/2015	4th relapse	Neuroblastoma	MYCN	amplification		amplification	0.499
11	cfDNA	09/07/2020	Diagnostic	Neuroblastoma	ALK	p.(Phe1174Leu)	c.3522C>A	0.476	0.95
11	cfDNA	09/07/2020	Diagnostic	Neuroblastoma	MYCN	amplification		amplification	0.95
11	cfDNA	09/07/2020	Diagnostic	Neuroblastoma	CDKN2A	deletion		deletion	0.95
11	cfDNA	17/08/2020	D40 disease assessments	Neuroblastoma	ALK	p.(Phe1174Leu)	c.3522C>A	0.014	0
11	cfDNA	17/08/2020	D40 disease assessments	Neuroblastoma	MYCN	amplification		amplification	0
11	cfDNA	23/09/2020	Disease assessment while on rapid COJEC	Neuroblastoma	ALK	p.(Phe1174Leu)	c.3522C>A	0.001	0
11	cfDNA	23/09/2020	Disease assessment while on rapid COJEC	Neuroblastoma	MYCN	amplification		Not detected	0
14	cfDNA	01/06/2020	Diagnostic	Neuroblastoma		No variants detected			0

14	cfDNA CSF	17/07/2020	Diagnostic	Neuroblastoma		No variants detected			NA
14	DNA from tissue	15/07/2020	Relapse	Neuroblastoma	AKT1	deletion		deletion	NA
14	DNA from tissue	15/07/2020	Relapse	Neuroblastoma	ATM	deletion		deletion	NA
14	DNA from tissue	15/07/2020	Relapse	Neuroblastoma	ERBB2	amplification		amplification	NA
14	DNA from tissue	15/07/2020	Relapse	Neuroblastoma	TERT	amplification		amplification	NA
14	cfDNA	05/08/2020	Relapse	Neuroblastoma	AKT1	deletion		not detected	0
14	cfDNA	05/08/2020	Relapse	Neuroblastoma	ATM	deletion		not detected	0
14	cfDNA	05/08/2020	Relapse	Neuroblastoma	ERBB2	amplification		not detected	0
14	cfDNA	05/08/2020	Relapse	Neuroblastoma	TERT	amplification		not detected	0
18	cfDNA	12/07/2018	Diagnostic	Neuroblastoma	ATM	p.(Cys1286Phe)	c.3857G>T	0.457	0.6616
18	cfDNA	12/07/2018	Diagnostic	Neuroblastoma	TERT	NA	g.1295250G>A	0.199	0.6616
18	cfDNA	12/07/2018	Diagnostic	Neuroblastoma	ALK	p.(Arg1275Gln)	c.3824G>A	0.004	0.6616
18	cfDNA	12/07/2018	Diagnostic	Neuroblastoma	MYCN	amplification		amplification	0.6616
18	cfDNA	12/07/2018	Diagnostic	Neuroblastoma	CCND1	amplification		amplification	0.6616
18	DNA from tissue	21/07/2018	Diagnostic	Neuroblastoma	ATM	p.(Cys1286Phe)	c.3857G>T	0.880	NA
18	DNA from tissue	21/07/2018	Diagnostic	Neuroblastoma	TERT	NA	g.1295250G>A	0.000	NA
18	DNA from tissue	21/07/2018	Diagnostic	Neuroblastoma	ALK	p.(Arg1275Gln)	c.3824G>A	0.000	NA

18	DNA from tissue	21/07/2018	Diagnostic	Neuroblastoma	MYCN	amplification		amplification	NA
18	DNA from tissue	21/07/2018	Diagnostic	Neuroblastoma	CCND1	amplification		Not detected	NA
18	DNA from tissue	08/01/2019	After surgery (had chemo before surgery)	Neuroblastoma	MYCN	amplification		amplification	NA
18	DNA from tissue	08/01/2019	After surgery (had chemo before surgery)	Neuroblastoma	CCND1	amplification		Not detected	NA
18	DNA from tissue	08/01/2019	After surgery (had chemo before surgery)	Neuroblastoma	ATM	p.(Cys1286Phe)	c.3857G>T	0.010	NA
18	DNA from tissue	08/01/2019	After surgery (had chemo before surgery)	Neuroblastoma	ALK	p.(Arg1275Gln)	c.3824G>A	0.000	NA
18	DNA from tissue	08/01/2019	After surgery (had chemo before surgery)	Neuroblastoma	TERT	NA	g.1295250G>A	0.000	NA
19	DNA from tissue	13/12/2016	Surgical resection	Neuroblastoma	ATM	p.(Ile1407Thr)	c.4220T>C	0.070	NA
19	cfDNA	09/07/2018	Post surgical resection	Neuroblastoma	ATM	p.(Ile1407Thr)	c.4220T>C	0.000	0
19	cfDNA	22/03/2019	Metastatic relapse	Neuroblastoma	ATM	p.(Ile1407Thr)	c.4220T>C	0.010	0.397
19	cfDNA	22/03/2019	Metastatic relapse	Neuroblastoma	TP53	p.(His179Gln)	c.537T>A	0.120	0.397
19	cfDNA	22/03/2019	Metastatic relapse	Neuroblastoma	RB1	NA	c.1960+2T>C	0.010	0.397
19	cfDNA	22/03/2019	Metastatic relapse	Neuroblastoma	FGFR3	p.(Ala719Val)	c.2156C>T	0.003	0.397
19	cfDNA	22/03/2019	Metastatic relapse	Neuroblastoma	ARID1A	p.(Pro650Leu)	c.1949C>T	0.003	0.397

19	cfDNA	22/03/2019	Metastatic relapse	Neuroblastoma	ARID1A	p.(Ala2234Thr)	c.6700G>A	0.004	0.397
19	DNA from tissue	22/03/2019	Metastatic relapse	Neuroblastoma	ATM	p.(Ile1407Thr)	c.4220T>C	0.060	NA
19	DNA from tissue	22/03/2019	Metastatic relapse	Neuroblastoma	SETD2	p.(Asp1321Asn)	c.3961G>A	0.119	NA
20	DNA from tissue	11/12/2019	Relapse - thoracic mass	Neuroblastoma	ARID1A	p.(Gly1158Ter)	c.3471_3472delinsGT	0.000	NA
20	DNA from tissue	11/12/2019	Relapse - thoracic mass	Neuroblastoma	CCND1	amplification		Not detected	#N/A
20	cfDNA	21/01/2020	Relapse	Neuroblastoma	ARID1A	p.(Gly1158Ter)	c.3471_3472delinsGT	0.024	0.264
20	cfDNA	21/01/2020	Relapse	Neuroblastoma	CCND1	amplification		amplification	0.264
20	DNA from primary tissue	14/03/2019	Primary tumour - thoracic mass	Neuroblastoma	SETD2	p.(Asn1943Lys)	c.5829C>G	0.452	#N/A
20	DNA from primary tissue	14/03/2019	Primary tumour - thoracic mass	Neuroblastoma	CCND1	amplification		amplification	#N/A
20	DNA from liver mets tissue	11/12/2019	Liver met	Neuroblastoma	ARID1A	p.(Gly1158Ter)	c.3471_3472delinsGT	0.000	#N/A
20	DNA from liver mets tissue	11/12/2019	Liver met	Neuroblastoma	CCND1	amplification		Not detected	#N/A

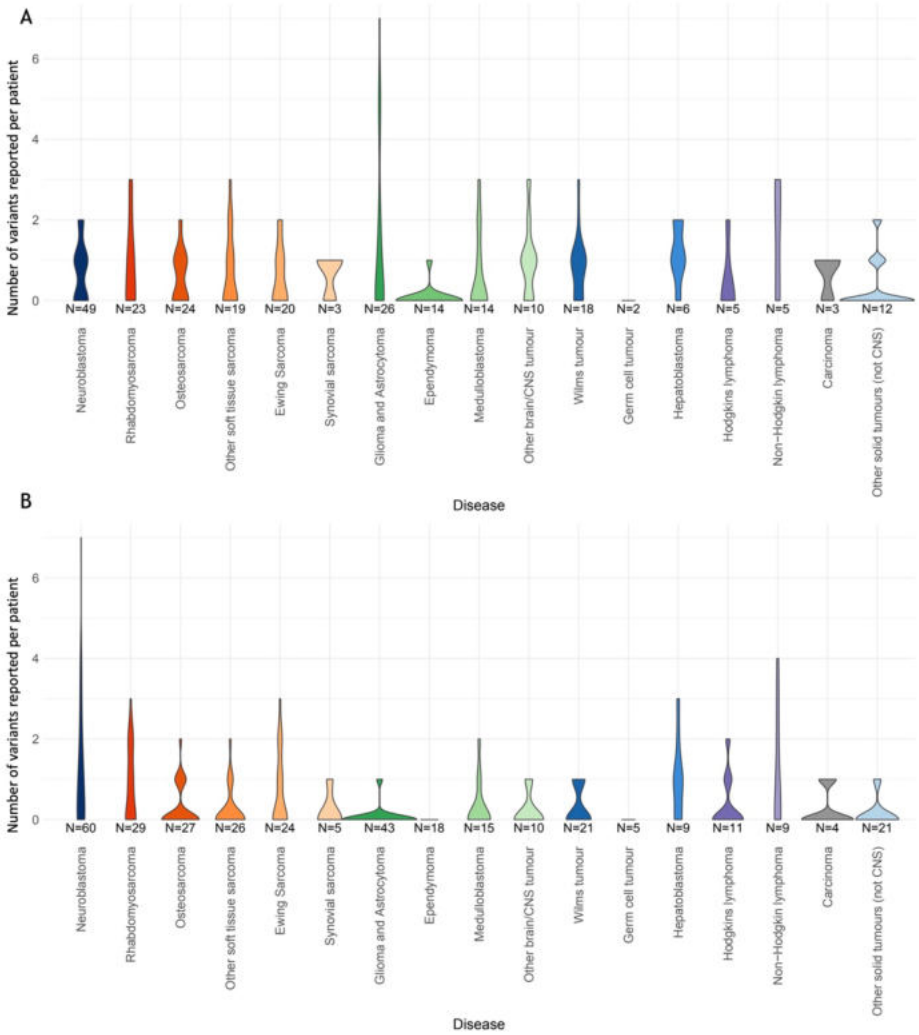


Figure Appndix-0-1 The number of SNVs and indels reported by targeted panel sequencing in tissue (A) and cfDNA (B) in patients with different cancer types at relapse in Stratified Medicine Paediatric Programme. For tissue samples, only samples with purity >0.1 shown.

Bibliography

1. https://www.cancerresearchuk.org/health-professional/cancer-statistics/childrens-cancers. Cancer Research UK. *Accessed 07/2022*.
2. Lam, C. G., Howard, S. C., Bouffet, E. & Pritchard-Jones, K. Science and health for all children with cancer. *Science* **363**, 1182–1186 (2019).
3. Steliarova-Foucher, E. *et al.* International incidence of childhood cancer, 2001–10: a population-based registry study. *Lancet Oncol* **18**, 719–731 (2017).
4. Miller, K. D. *et al.* Cancer statistics for adolescents and young adults, 2020. *CA Cancer J Clin* **70**, 443–459 (2020).
5. Kaatsch, P. Epidemiology of childhood cancer. *Cancer Treat Rev* **36**, 277–285 (2010).
6. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J Clin* **66**, 7–30 (2016).
7. Kopp, L. M., Gupta, P., Pelayo-Katsanis, L., Wittman, B. & Katsanis, E. Late effects in adult survivors of pediatric cancer: a guide for the primary care physician. *Am J Med* **125**, 636–641 (2012).
8. Henderson, T. O. *et al.* Risk Factors Associated With Secondary Sarcomas in Childhood Cancer Survivors: A Report From the Childhood Cancer Survivor Study. *International Journal of Radiation Oncology*Biology*Physics* **84**, 224–230 (2012).
9. Neglia, J. P. *et al.* Second Malignant Neoplasms in Five-Year Survivors of Childhood Cancer: Childhood Cancer Survivor Study. *JNCI: Journal of the National Cancer Institute* **93**, 618–629 (2001).
10. Meadows, A. T. *et al.* Second Neoplasms in Survivors of Childhood Cancer: Findings From the Childhood Cancer Survivor Study Cohort. *Journal of Clinical Oncology* **27**, 2356 (2009).
11. Kentsis, A. Why do young people get cancer? *Pediatric Blood and Cancer* vol. 67 Preprint at https://doi.org/10.1002/pbc.28335 (2020).
12. Hendrikse, L. D. *et al.* Failure of human rhombic lip differentiation underlies medulloblastoma formation. *Nature 2022 609:7929* **609**, 1021–1028 (2022).
13. Downing, J. R. *et al.* The Pediatric Cancer Genome Project. *Nature Genetics 2012 44:6* **44**, 619–622 (2012).
14. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature 2018 555:7696* **555**, 321–327 (2018).
15. Berlanga, P. *et al.* The European MAPPYACTS trial: Precision Medicine Program in Pediatric and Adolescent Patients with Recurrent MalignanciesThe MAPPYACTS pediatric molecular profiling trial. *Cancer Discov* (2022) doi:10.1158/2159-8290.CD-21-1136.
16. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature 2018 555:7696* **555**, 371–376 (2018).
17. Harttrampf, A. C. *et al.* Molecular Screening for Cancer Treatment Optimization (MOSCATO-01) in pediatric patients: A single-institutional prospective molecular stratification trial. *Clinical Cancer Research* **23**, 6101–6112 (2017).
18. Newman, S. *et al.* Genomes for kids: The scope of pathogenic mutations in pediatric cancer revealed by comprehensive dna and rna sequencing. *Cancer Discov* **11**, 3008–3027 (2021).
19. Khater, F. *et al.* Molecular Profiling of Hard-to-Treat Childhood and Adolescent Cancers. *JAMA Netw Open* **2**, e192906–e192906 (2019).
20. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature 2013 499:7457* **499**, 214–218 (2013).

21. Alejandro Sweet-Cordero, E. & Biegel, J. A. The genomic landscape of pediatric cancers: Implications for diagnosis and treatment. *Science* **363**, 1170 (2019).
22. Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226–231 (2012).
23. Wu, G. *et al.* Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat Genet* **44**, 251–253 (2012).
24. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
25. George, S. L. *et al.* A tailored molecular profiling programme for children with cancer to identify clinically actionable genetic alterations. *Eur J Cancer* **121**, 224–235 (2019).
26. Bellini, A. *et al.* Frequency and Prognostic Impact of ALK Amplifications and Mutations in the European Neuroblastoma Study Group (SIOPEN) High-Risk Neuroblastoma Trial (HR-NBL1). *J Clin Oncol* **39**, 3377–3390 (2021).
27. Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nature Genetics 2013 45:3* **45**, 279–284 (2013).
28. Decaesteker, B. *et al.* From DNA Copy Number Gains and Tumor Dependencies to Novel Therapeutic Targets for High-Risk Neuroblastoma. *J Pers Med* **11**, (2021).
29. Depuydt, P. *et al.* Genomic Amplifications and Distal 6q Loss: Novel Markers for Poor Survival in High-risk Neuroblastoma Patients. *J Natl Cancer Inst* **110**, (2018).
30. Mackay, A. *et al.* Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma. *Cancer Cell* **32**, 520-537.e5 (2017).
31. Ozer, E., Sevinc, A., Ince, D., Yuzuguldu, R. & Olgun, N. BRAF V600E Mutation: A Significant Biomarker for Prediction of Disease Relapse in Pediatric Langerhans Cell Histiocytosis. *Pediatr Dev Pathol* **22**, 449–455 (2019).
32. Parsons, D. W. *et al.* Diagnostic Yield of Clinical Tumor and Germline Whole-Exome Sequencing for Children With Solid Tumors. *JAMA Oncol* **2**, 616 (2016).
33. Zhang, J. *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med* **373**, 2336 (2015).
34. Ripperger, T. *et al.* Childhood cancer predisposition syndromes-A concise review and recommendations by the Cancer Predisposition Working Group of the Society for Pediatric Oncology and Hematology. *Am J Med Genet A* **173**, 1017–1037 (2017).
35. Lee, V., Murphy, A., Le, D. T. & Luis A. Diaz, Jr. Mismatch Repair Deficiency and Response to Immune Checkpoint Blockade. *Oncologist* **21**, 1200 (2016).
36. Long, A. H., Morgenstern, D. A., Leruste, A., Bourdeaut, F. & Davis, K. L. Checkpoint Immunotherapy in Pediatrics: Here, Gone, and Back Again. *Am Soc Clin Oncol Educ Book* **42**, 1–14 (2022).
37. Nigro, O. *et al.* Controversies on the possible role of immune checkpoint inhibitors in pediatric cancers: balancing irAEs and efficacy. *https://doi.org/10.1177/03008916211010214* **107**, 276–281 (2021).
38. Que, Y., Hu, Y., Hong, D. & Zhang, Y. Trends in clinical development of pediatric cancer for PD-1 and PD-L1 inhibitors: an analysis of ClinicalTrials.gov. *J Immunother Cancer* **9**, e002920 (2021).
39. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature 2013 502:7471* **502**, 333–339 (2013).
40. Erdmann, F. *et al.* Childhood cancer: Survival, treatment modalities, late effects and improvements over time. *Cancer Epidemiol* **71**, 101733 (2021).
41. Harris, M. H. *et al.* Multicenter Feasibility Study of Tumor Molecular Profiling to Inform Therapeutic Decisions in Advanced Pediatric Solid Tumors: The Individualized Cancer Therapy (iCat) Study. *JAMA Oncol* **2**, 608–615 (2016).
42. Worst, B. C. *et al.* Next-generation personalised medicine for high-risk paediatric cancer patients - The INFORM pilot study. *Eur J Cancer* **65**, 91–101 (2016).

43. Mody, R. J. *et al.* Integrative clinical sequencing in the management of refractory or relapsed cancer in youth. *JAMA - Journal of the American Medical Association* **314**, 913–925 (2015).
44. Church, A. J. *et al.* Molecular profiling identifies targeted therapy opportunities in pediatric solid cancer. *Nature Medicine 2022* 1–9 (2022) doi:10.1038/s41591-022-01856-6.
45. Wong, M. *et al.* Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. *Nature Medicine 2020 26:11* **26**, 1742–1753 (2020).
46. Lambo, S. *et al.* The molecular landscape of ETMR at diagnosis and relapse. *Nature* **576**, 274–280 (2019).
47. Crompton, B. D. *et al.* The Genomic Landscape of Pediatric Ewing Sarcoma. (2014) doi:10.1158/2159-8290.CD-13-1037.
48. Morrissy, A. S. *et al.* Divergent clonal selection dominates medulloblastoma at recurrence. *Nature* **529**, 351–357 (2016).
49. Schleiermacher, G. *et al.* Emergence of new ALK mutations at relapse of neuroblastoma. *Journal of Clinical Oncology* **32**, 2727–2734 (2014).
50. Padovan-Merhar, O. M. *et al.* Enrichment of Targetable Mutations in the Relapsed Neuroblastoma Genome. *PLoS Genet* **12**, (2016).
51. Eleveld, T. F. *et al.* Relapsed neuroblastomas show frequent RAS-MAPK pathway mutations. *Nature Genetics 2015 47:8* **47**, 864–871 (2015).
52. Zöllner, S. K. *et al.* Ewing Sarcoma—Diagnosis, Treatment, Clinical Challenges and Future Perspectives. *Journal of Clinical Medicine 2021, Vol. 10, Page 1685* **10**, 1685 (2021).
53. Tirode, F. *et al.* Genomic landscape of ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov* **4**, 1342–1353 (2014).
54. Brohl, A. S. *et al.* The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genet* **10**, (2014).
55. Negri, G. L. *et al.* Integrative genomic analysis of matched primary and metastatic pediatric osteosarcoma. *J Pathol* **249**, 319 (2019).
56. Ramaswamy, V. *et al.* Recurrence patterns across medulloblastoma subgroups: An integrated clinical and molecular analysis. *Lancet Oncol* **14**, 1200–1207 (2013).
57. Wang, X. *et al.* Medulloblastoma subgroups remain stable across primary and metastatic compartments. *Acta Neuropathol* **129**, 449–457 (2015).
58. Cohen, B. *et al.* Pediatric Oncology Provider Views on Performing a Biopsy of Solid Tumors in Children with Relapsed or Refractory Disease for the Purpose of Genomic Profiling. *Ann Surg Oncol* **23**, 990–997 (2016).
59. Cahn, F. *et al.* Blood-Derived Liquid Biopsies Using Foundation One® Liquid CDx for Children and Adolescents with High-Risk Malignancies: A Monocentric Experience. *Cancers (Basel)* **14**, 2774 (2022).
60. Ignatiadis, M., Sledge, G. W. & Jeffrey, S. S. Liquid biopsy enters the clinic — implementation issues and future challenges. *Nature Reviews Clinical Oncology 2021 18:5* **18**, 297–312 (2021).
61. Mandel P, M. P. Les Acides Nucléiques Du Plasma Sanguin Chez l'Homme. *C R Seances Soc Biol Fil* **142**, 3–4 (1948).
62. Dennis Lo, Y. M. *et al.* Presence of fetal DNA in maternal plasma and serum. *Lancet* **350**, 485–487 (1997).
63. Leon, S. A., Shapiro, B., Sklaroff, D. M. & Yaros, M. J. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res* **37**, 646–50 (1977).
64. Caldas, C. *et al.* Detection of K-ras mutations in the stool of patients with pancreatic adenocarcinoma and pancreatic ductal hyperplasia. *Cancer Res* **54**, 3568–73 (1994).

65. Vasioukhin, V. *et al.* Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia. *Br J Haematol* **86**, 774–779 (1994).
66. Stroun, M. *et al.* Neoplastic characteristics of the DNA found in the plasma of cancer patients. *Oncology* **46**, 318–322 (1989).
67. Sun, K. *et al.* Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc Natl Acad Sci U S A* **112**, E5503–E5512 (2015).
68. Tug, Tug, S. et al. "Exercise-induced increases in cell free DNA in human plasma originate predominantly from cells of the haematopoietic lineage." Exercise immunology review 21 (2015): 164-73 .
69. Thierry, A. R., el Messaoudi, S., Gahan, P. B., Anker, P. & Stroun, M. Origins, structures, and functions of circulating DNA in oncology. *Cancer and Metastasis Reviews* **35**, 347–376 (2016).
70. Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**, 57–68 (2016).
71. *Moss, J. et al..* Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun* **9**, (2018).
72. Rodrigues Filho, E. M. *et al.* Elevated Cell-Free Plasma DNA Level as an Independent Predictor of Mortality in Patients with Severe Traumatic Brain Injury. *J Neurotrauma* **31**, 1639 (2014).
73. de Vlaminck, I. *et al.* Noninvasive monitoring of infection and rejection after lung transplantation. *Proc Natl Acad Sci U S A* **112**, 13336–13341 (2015).
74. de Vlaminck, I. *et al.* Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. *Sci Transl Med* **6**, (2014).
75. Thierry, A. R. *et al.* Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Res* **38**, 6159–6175 (2010).
76. Sanchez, C. *et al.* Circulating nuclear DNA structural features, origins, and complete size profile revealed by fragmentomics. *JCI Insight* **6**, (2021).
77. Mouliere, F. *et al.* Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* **10**, (2018).
78. Wan, J. C. M. *et al.* Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nature Reviews Cancer 2017 17:4* **17**, 223–238 (2017).
79. Hudecova, I. *et al.* Characteristics, origin, and potential for cancer diagnostics of ultrashort plasma cell-free DNA. *Genome Res* gr.275691.121 (2021) doi:10.1101/GR.275691.121.
80. Chan, K. C. A., Leung, S. F., Yeung, S. W., Chan, A. T. C. & Lo, Y. M. D. Persistent aberrations in circulating DNA integrity after radiotherapy are associated with poor prognosis in nasopharyngeal carcinoma patients. *Clin Cancer Res* **14**, 4141–4145 (2008).
81. Cristiano, S. *et al.* Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385–389 (2019).
82. Crowley, E., di Nicolantonio, F., Loupakis, F. & Bardelli, A. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol* **10**, 472–484 (2013).
83. Paemel, R. van *et al.* The pitfalls and promise of liquid biopsies for diagnosing and treating solid tumors in children: a review. *Eur J Pediatr* **179**, 191–202 (2020).
84. Adashek, J. J., Janku, F. & Kurzrock, R. Signed in Blood: Circulating Tumor DNA in Cancer Diagnosis, Treatment and Screening. *Cancers (Basel)* **13**, (2021).
85. *Bettegowda, C. et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, (2014).

86. Zill, O. A. *et al.* The Landscape of Actionable Genomic Alterations in Cell-Free Circulating Tumor DNA from 21,807 Advanced Cancer Patients. *Clin Cancer Res* **24**, 3528–3538 (2018).
87. Stewart, C. M. *et al.* The value of cell-free DNA for molecular pathology. *J Pathol* **244**, 616 (2018).
88. Diaz, L. A. & Bardelli, A. Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol* **32**, 579–586 (2014).
89. Shulman, D. S. *et al.* Detection of circulating tumour DNA is associated with inferior outcomes in Ewing sarcoma and osteosarcoma: a report from the Children's Oncology Group. *British Journal of Cancer 2018 119:5* **119**, 615–621 (2018).
90. Klega, K. *et al.* Detection of Somatic Structural Variants Enables Quantification and Characterization of Circulating Tumor DNA in Children With Solid Tumors. *https://doi.org/10.1200/PO.17.00285* 1–13 (2018) doi:10.1200/PO.17.00285.
91. Shukla, N. N. *et al.* Plasma DNA-based molecular diagnosis, prognostication, and monitoring of patients with EWSR1 fusion-positive sarcomas. *JCO Precis Oncol* **2017**, 1–11 (2017).
92. Krumbholz, M. *et al.* Genomic EWSR1 Fusion Sequence as Highly Sensitive and Dynamic Plasma Tumor Marker in Ewing Sarcoma. *Clin Cancer Res* **22**, 4356–4365 (2016).
93. Mussolin, L. *et al.* Plasma cell-free DNA in paediatric lymphomas. *J Cancer* **4**, 323–329 (2013).
94. Barris, D. M. *et al.* Detection of circulating tumor DNA in patients with osteosarcoma. *Oncotarget* **9**, 12695–12704 (2018).
95. Jiménez, I. *et al.* Circulating tumor DNA analysis enables molecular characterization of pediatric renal tumors at diagnosis. *Int J Cancer* **144**, 68–79 (2019).
96. Panditharatna, E. *et al.* Clinically relevant and minimally invasive tumor surveillance of pediatric diffuse midline gliomas using patient derived liquid biopsy. *Clin Cancer Res* **24**, 5850 (2018).
97. Mueller, S. *et al.* A pilot precision medicine trial for children with diffuse intrinsic pontine glioma—PNOC003: A report from the Pacific Pediatric Neuro-Oncology Consortium. *Int J Cancer* **145**, 1889–1901 (2019).
98. *Chicard, M. et al.* Genomic Copy Number Profiling Using Circulating Free Tumor DNA Highlights Heterogeneity in Neuroblastoma. *Clin Cancer Res* **22**, 5564–5573 (2016).
99. Chicard, M. *et al.* Whole-Exome Sequencing of Cell-Free DNA Reveals Temporo-spatial Heterogeneity and Identifies Treatment-Resistant Clones in Neuroblastoma. *Clinical Cancer Research* **24**, 939–949 (2018).
100. Roy, N. van *et al.* Shallow Whole Genome Sequencing on Circulating Cell-Free DNA Allows Reliable Noninvasive Copy-Number Profiling in Neuroblastoma Patients. **23**, (2017).
101. Wang, X. *et al.* Plasma cell-free DNA quantification is highly correlated to tumor burden in children with neuroblastoma. *Cancer Med* **7**, 3022–3030 (2018).
102. Su, Y. *et al.* Dynamic alterations of plasma cell free DNA in response to chemotherapy in children with neuroblastoma. *Cancer Med* **8**, 1558–1566 (2019).
103. Combaret, V. *et al. Circulating MYCN DNA as a Tumor-specific Marker in Neuroblastoma Patients*. *CANCER RESEARCH* vol. 62 http://aacrjournals.org/cancerres/article-pdf/62/13/3646/2495205/ch1302003646.pdf (2002).
104. *Kurihara, S. et al.* Circulating free DNA as non-invasive diagnostic biomarker for childhood solid tumors. *J Pediatr Surg* **50**, 2094–2097 (2015).

105. Pagès, M. *et al.* Liquid biopsy detection of genomic alterations in pediatric brain tumors from cell-free DNA in peripheral blood, CSF, and urine. *Neuro Oncol* (2022) doi:10.1093/NEUONC/NOAB299.
106. Paemel, R. van *et al.* The feasibility of using liquid biopsies as a complementary assay for copy number aberration profiling in routinely collected paediatric cancer patient samples. *Eur J Cancer* **0**, (2021).
107. Mok, T. *et al.* Detection and Dynamic Changes of EGFR Mutations from Circulating Tumor DNA as a Predictor of Survival Outcomes in NSCLC Patients Treated with First-line Intercalated Erlotinib and Chemotherapy. *Clin Cancer Res* **21**, 3196–3203 (2015).
108. Andersson, D., Andersson, D., MG, D. & A, S. Circulating cell-free tumor DNA analysis in pediatric cancers. *Mol Aspects Med* **72**, (2020).
109. Abbou, S. D., Shulman, D. S., DuBois, S. G. & Crompton, B. D. Assessment of circulating tumor DNA in pediatric solid tumors: The promise of liquid biopsies. *Pediatr Blood Cancer* **66**, e27595 (2019).
110. Eguchi-Ishimae, M. *et al.* Early detection of the PAX3-FOXO1 fusion gene in circulating tumor-derived DNA in a case of alveolar rhabdomyosarcoma. *Genes Chromosomes Cancer* **58**, 521–529 (2019).
111. Ueno-Yokohata, H. *et al.* Preoperative diagnosis of clear cell sarcoma of the kidney by detection of BCOR internal tandem duplication in circulating tumor DNA. *Genes Chromosomes Cancer* **57**, 525–529 (2018).
112. Voigt, W. *et al.* Beyond tissue biopsy: a diagnostic framework to address tumor heterogeneity in lung cancer. *Curr Opin Oncol* **32**, 68–77 (2020).
113. Xu, L. X. *et al.* Genomic and transcriptional heterogeneity of multifocal hepatocellular carcinoma. *Ann Oncol* **30**, 990–997 (2019).
114. Wu, B. *et al.* Intratumoral heterogeneity and genetic characteristics of prostate cancer. *Int J Cancer* **146**, 3369–3378 (2020).
115. Pereira, B. *et al.* Cell-free DNA captures tumor heterogeneity and driver alterations in rapid autopsies with pre-treated metastatic cancer. *Nature Communications 2021 12:1* **12**, 1–13 (2021).
116. Takai, E. *et al.* Post-mortem Plasma Cell-Free DNA Sequencing: Proof-of-Concept Study for the 'Liquid Autopsy'. *Sci Rep* **10**, (2020).
117. *Cresswell, G. D. et al.* Mapping the breast cancer metastatic cascade onto ctDNA using genetic and epigenetic clonal tracking. *Nat Commun* **11**, (2020).
118. Abbosh, C. *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature 2017 545:7655* **545**, 446–451 (2017).
119. Murtaza, M. *et al.* Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun* **6**, (2015).
120. Huang, A. *et al.* Circumventing intratumoral heterogeneity to identify potential therapeutic targets in hepatocellular carcinoma. *J Hepatol* **67**, 293–301 (2017).
121. Lin, J. *et al.* Targeted Next-Generation Sequencing Combined With Circulating-Free DNA Deciphers Spatial Heterogeneity of Resected Multifocal Hepatocellular Carcinoma. *Front Immunol* **12**, 2286 (2021).
122. Coto-Llerena, M. *et al.* Circulating Cell-Free DNA Captures the Intratumor Heterogeneity in Multinodular Hepatocellular Carcinoma. *JCO Precis Oncol* **6**, e2100335 (2022).
123. Passiglia, F. *et al.* Metastatic Site Location Influences the Diagnostic Accuracy of ctDNA EGFR- Mutation Testing in NSCLC Patients: a Pooled Analysis. *Curr Cancer Drug Targets* **18**, 697–705 (2018).
124. Bando, H. *et al.* Effects of Metastatic Sites on Circulating Tumor DNA in Patients With Metastatic Colorectal Cancer. *JCO Precis Oncol* **6**, e2100535 (2022).
125. López-Carrasco, A. *et al.* Intra-Tumour Genetic Heterogeneity and Prognosis in High-Risk Neuroblastoma. *Cancers (Basel)* **13**, 5173 (2021).

126. Pincez, T. *et al.* Feasibility and clinical integration of molecular profiling for target identification in pediatric solid tumors. *Pediatr Blood Cancer* **64**, e26365 (2017).
127. López, G. Y. *et al.* The genetic landscape of gliomas arising after therapeutic radiation. *Acta Neuropathologica 2018 137:1* **137**, 139–150 (2018).
128. Parikh, A. R. *et al.* Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med* **25**, 1415 (2019).
129. Abbosh, C. *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature 2017 545:7655* **545**, 446–451 (2017).
130. Hayashi, M. *et al.* Highly Personalized Detection of Minimal Ewing Sarcoma Disease Burden from Plasma Tumor DNA. *Cancer* **122**, 3015 (2016).
131. Starza, I. della *et al.* Minimal Residual Disease in Acute Lymphoblastic Leukemia: Technical and Clinical Advances. *Front Oncol* **9**, (2019).
132. Kahana-Edwin, S. *et al.* Neuroblastoma Molecular Risk-Stratification of DNA Copy Number and ALK Genotyping via Cell-Free Circulating Tumor DNA Profiling. *Cancers (Basel)* **13**, 3365 (2021).
133. Bohers, E., Viailly, P. J. & Jardin, F. Cfdna sequencing: Technological approaches and bioinformatic issues. *Pharmaceuticals* vol. 14 Preprint at https://doi.org/10.3390/ph14060596 (2021).
134. Heinrichs, S. & Look, A. T. Identification of structural aberrations in cancer by SNP array analysis. *Genome Biol* **8**, 1–5 (2007).
135. Salk, J. J., Schmitt, M. W. & Loeb, L. A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nature Reviews Genetics 2018 19:5* **19**, 269–285 (2018).
136. Mitchell, K. *et al.* Benchmarking of computational error-correction methods for next-generation sequencing data. *Genome Biol* **21**, 1–13 (2020).
137. Laehnemann, D., Borkhardt, A. & McHardy, A. C. Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Brief Bioinform* **17**, 154–179 (2016).
138. Newman, A. M. *et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. *Nature Biotechnology 2016 34:5* **34**, 547–555 (2016).
139. https://eu.idtdna.com/pages/support/faqs/how-do-i-use-umis-for-error-correction. IDT. *Accessed 09/2022*.
140. Izquierdo, E. *et al.* Development of a targeted sequencing approach to identify prognostic, predictive and diagnostic markers in paediatric solid tumours. *Oncotarget* **8**, 112036–112050 (2017).
141. Zhao, Y. *et al.* Performance comparison of blood collection tubes as liquid biopsy storage system for minimizing cfDNA contamination from genomic DNA. *J Clin Lab Anal* **33**, (2019).
142. Maass KK *et al.* From Sampling to Sequencing: A Liquid Biopsy Pre-Analytic Workflow to Maximize Multi-Layer Genomic Information from a Single Tube. *Cancers (Basel)* **13**, (2021).
143. Gerstung, M. *et al.* Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nature Communications 2012 3:1* **3**, 1–8 (2012).
144. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
145. Adalsteinsson, V. A. *et al.* Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nature Communications 2017 8:1* **8**, 1–13 (2017).
146. van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A* **107**, 16910–16915 (2010).
147. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology 2011 29:1* **29**, 24–26 (2011).

148. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029-1041.e21 (2017).
149. Stankunaite, R. *et al.* Circulating tumour DNA sequencing to determine therapeutic response and identify tumour heterogeneity in patients with paediatric solid tumours. *Eur J Cancer* **162**, 209–220 (2022).
150. Lodrini, M. *et al.* Using droplet digital PCR to analyze MYCN and ALK copy number in plasma from patients with neuroblastoma. *Oncotarget* **8**, 85234–85251 (2017).
151. Peitz, C. *et al.* Multiplexed Quantification of Four Neuroblastoma DNA Targets in a Single Droplet Digital PCR Reaction. *Journal of Molecular Diagnostics* **22**, 1309–1323 (2020).
152. Izquierdo, E. *et al.* Droplet digital PCR-based detection of circulating tumor DNA from pediatric high grade and diffuse midline glioma patients. *Neurooncol Adv* **3**, 1–12 (2021).
153. Li, D. *et al.* Standardization of the liquid biopsy for pediatric diffuse midline glioma using ddPCR. *Scientific Reports 2021 11:1* **11**, 1–10 (2021).
154. Diefenbach, R. J., Lee, J. H., Kefford, R. F. & Rizos, H. Evaluation of commercial kits for purification of circulating free DNA. *Cancer Genet* **228–229**, 21–27 (2018).
155. Sorber, L. *et al.* A Comparison of Cell-Free DNA Isolation Kits: Isolation and Quantification of Cell-Free DNA in Plasma. *J Mol Diagn* **19**, 162–168 (2017).
156. Warton, K., Graham, L. J., Yuwono, N. & Samimi, G. Comparison of 4 commercial kits for the extraction of circulating DNA from plasma. *Cancer Genet* **228–229**, 143–150 (2018).
157. Mattocks, C. J. *et al.* A standardized framework for the validation and verification of clinical molecular genetic tests. *European Journal of Human Genetics 2010 18:12* **18**, 1276–1288 (2010).
158. Smith, T., Heger, A. & Sudbery, I. UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491–499 (2017).
159. Khan, K. H. *et al.* Longitudinal liquid biopsy and mathematical modeling of clonal evolution forecast time to treatment failure in the prospect-c phase ii colorectal cancer clinical trial. *Cancer Discov* **8**, 1270–1285 (2018).
160. Peneder, P. *et al.* Multimodal analysis of cell-free DNA whole-genome sequencing for pediatric cancers with low mutational burden. *Nat Commun* **12**, 1–16 (2021).
161. Edmonson, M. N. *et al.* Pediatric cancer variant pathogenicity information exchange (PECANPIE): A cloud-based platform for curating and classifying germline variants. *Genome Res* **29**, 1555–1565 (2019).
162. Cimmino, F., Lasorsa, V. A., Vetrella, S., Iolascon, A. & Capasso, M. A Targeted Gene Panel for Circulating Tumor DNA Sequencing in Neuroblastoma. *Front Oncol* **10**, 2622 (2020).
163. Dupain, C., Harttrampf, A. C., Urbinati, G., Geoerger, B. & Massaad-Massade, L. Relevance of Fusion Genes in Pediatric Cancers: Toward Precision Medicine. *Mol Ther Nucleic Acids* **6**, 315–326 (2017).
164. Zhao, X. *et al.* NTRK Fusions Identified in Pediatric Tumors: The Frequency, Fusion Partners, and Clinical Outcome. *https://doi.org/10.1200/PO.20.00250* 204–214 (2021) doi:10.1200/PO.20.00250.
165. Vellichirammal, N. N., Chaturvedi, N. K., Joshi, S. S., Coulter, D. W. & Guda, C. Fusion genes as biomarkers in pediatric cancers: A review of the current state and applicability in diagnostics and personalized therapy. *Cancer Letters* vol. 499 24–38 Preprint at https://doi.org/10.1016/j.canlet.2020.11.015 (2021).
166. Diehl, F. *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* **102**, 16368–16373 (2005).

167. Xu, R. H. *et al.* Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nature Materials 2017 16:11* **16**, 1155–1161 (2017).
168. Newman, A. M. *et al.* An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nature Medicine 2014 20:5* **20**, 548–554 (2014).
169. Sun, M. Y. *et al.* Targeted Next-Generation Sequencing of Circulating Tumor DNA Mutations among Metastatic Breast Cancer Patients. *Curr Oncol* **28**, 2326–2336 (2021).
170. Choudhury, A. D. *et al.* Tumor fraction in cell-free DNA as a biomarker in prostate cancer. *JCI Insight* **3**, (2018).
171. Shah, A. T. *et al.* A Comprehensive Circulating Tumor DNA Assay for Detection of Translocation and Copy-Number Changes in Pediatric Sarcomas. *Mol Cancer Ther* **20**, 2016–2025 (2021).
172. van Paemel, R. *et al.* Minimally invasive classification of paediatric solid tumours using reduced representation bisulphite sequencing of cell-free DNA: a proof-of-principle study. *Epigenetics* **16**, 196 (2021).
173. Li, J. *et al.* Reliable tumor detection by whole-genome methylation sequencing of cell-free DNA in cerebrospinal fluid of pediatric medulloblastoma. *Sci Adv* **6**, (2020).
174. Chen, X. *et al.* Low-pass Whole-genome Sequencing of Circulating Cell-free DNA Demonstrates Dynamic Changes in Genomic Copy Number in a Squamous Lung Cancer Clinical Cohort. *Clinical Cancer Research* **25**, 2254–2263 (2019).
175. Hovelson, D. H. *et al.* Rapid, ultra low coverage copy number profiling of cell-free DNA as a precision oncology screening strategy. *Oncotarget* **8**, 89848–89866 (2017).
176. Schmelz, K. *et al.* Spatial and temporal intratumour heterogeneity has potential consequences for single biopsy-based neuroblastoma treatment decisions. *Nat Commun* **12**, (2021).
177. Gomez, R. L., Ibragimova, S., Ramachandran, R., Philpott, A. & Ali, F. R. Tumoral heterogeneity in neuroblastoma. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 188805 (2022) doi:10.1016/J.BBCAN.2022.188805.
178. Koschmann, C. *et al.* ATRX loss promotes tumor growth and impairs nonhomologous end joining DNA repair in glioma. *Sci Transl Med* **8**, (2016).
179. Campbell, B. B. *et al.* Comprehensive Analysis of Hypermutation in Human Cancer. *Cell* **171**, 1042-1056.e10 (2017).
180. Yu, Y. *et al.* Temozolomide-induced hypermutation is associated with distant recurrence and reduced survival after high-grade transformation of low-grade IDH-mutant gliomas. *Neuro Oncol* **23**, 1872–1884 (2021).
181. Nassif, E. F. *et al.* TP53 Mutation as a Prognostic and Predictive Marker in Sarcoma: Pooled Analysis of MOSCATO and ProfiLER Precision Medicine Trials. *Cancers 2021, Vol. 13, Page 3362* **13**, 3362 (2021).
182. Küppers, R., Engert, A. & Hansmann, M. L. Hodgkin lymphoma. *J Clin Invest* **122**, 3439–3447 (2012).
183. Turajlic, S., Sottoriva, A., Graham, T. & Swanton, C. Resolving genetic heterogeneity in cancer. *Nature Reviews Genetics 2019 20:7* **20**, 404–416 (2019).
184. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613–628 (2017).
185. Heuckmann, J. M. *et al.* ALK mutations conferring differential resistance to structurally diverse ALK inhibitors. *Clin Cancer Res* **17**, 7394–7401 (2011).
186. Sasaki, T. *et al.* The neuroblastoma associated F1174L ALK mutation causes resistance to an ALK kinase inhibitor in ALK translocated cancers. *Cancer Res* **70**, 10038 (2010).

187. Bosse, K. R. *et al.* Serial profiling of circulating tumor DNA identifies dynamic evolution of clinically actionable genomic alterations in high-risk neuroblastoma. *Cancer Discov* (2022) doi:10.1158/2159-8290.CD-22-0287.
188. Liu, T. *et al.* Exceptional response to the ALK and ROS1 inhibitor lorlatinib and subsequent mechanism of resistance in relapsed ALK F1174L-mutated neuroblastoma. *Cold Spring Harb Mol Case Stud* **7**, (2021).
189. Schleiermacher, G. *et al.* Accumulation of segmental alterations determines progression in neuroblastoma. *J Clin Oncol* **28**, 3122–3130 (2010).
190. Schleiermacher, G. *et al.* Segmental chromosomal alterations have prognostic impact in neuroblastoma: a report from the INRG project. *British Journal of Cancer 2012 107:8* **107**, 1418–1422 (2012).
191. Hellwig, S. *et al.* Automated size selection for short cell-free DNA fragments enriches for circulating tumor DNA and improves error correction during next generation sequencing. *PLoS One* **13**, (2018).
192. Marass, F. *et al.* Fragment Size Analysis May Distinguish Clonal Hematopoiesis from Tumor-Derived Mutations in Cell-Free DNA. *Clin Chem* **66**, 616 (2020).
193. Peneder, P., Bock, C. & Tomazou, E. M. LIQUORICE: Detection of epigenetic signatures in liquid biopsies based on whole-genome sequencing data. *Bioinformatics Advances, YYYY* 0–0 doi:10.1093/Bioinformatics.
194. Cheng, A. P. *et al.* Whole genome error-corrected sequencing for sensitive circulating tumor DNA cancer monitoring. *bioRxiv* 2022.11.17.516904 (2022) doi:10.1101/2022.11.17.516904.
195. Kurtz, D. M. *et al.* Enhanced detection of minimal residual disease by targeted sequencing of phased variants in circulating tumor DNA. *Nature Biotechnology 2021 39:12* **39**, 1537–1547 (2021).
196. McGranahan, N. & Swanton, C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* **27**, 15–26 (2015).
197. Gilson, P., Merlin, J.-L. & Harlé, A. Deciphering Tumour Heterogeneity: From Tissue to Liquid Biopsy. *Cancers (Basel)* **14**, 1384 (2022).
198. Bayle, A. *et al.* Liquid versus tissue biopsy for detecting actionable alterations according to ESCAT in patients with advanced cancer: A study from the French National Center for Precision Medicine (PRISM). *Annals of Oncology* (2022) doi:10.1016/J.ANNONC.2022.08.089.
199. Parsons, D. W. *et al.* Actionable Tumor Alterations and Treatment Protocol Enrollment of Pediatric and Young Adult Patients With Refractory Cancers in the National Cancer Institute-Children's Oncology Group Pediatric MATCH Trial. *J Clin Oncol* **40**, 2224–2234 (2022).
200. Langenberg, K. P. S. *et al.* Implementation of paediatric precision oncology into clinical practice: The Individualized Therapies for Children with cancer program 'iTHER'. *Eur J Cancer* **175**, 311–325 (2022).
201. Ackermann, S. *et al.* A mechanistic classification of clinical phenotypes in neuroblastoma. *Science* **362**, 1165–1170 (2018).
202. Shulman, D. S. *et al.* An international working group consensus report for the prioritization of molecular biomarkers for Ewing sarcoma. *npj Precision Oncology 2022 6:1* **6**, 1–11 (2022).
203. Hing, S. *et al.* Gain of 1q Is Associated with Adverse Outcome in Favorable Histology Wilms' Tumors. *Am J Pathol* **158**, 393 (2001).
204. van Tilburg, C. M. *et al.* The pediatric precision oncology inform registry: Clinical outcome and benefit for patients with very high-evidence targets. *Cancer Discov* **11**, 2764–2779 (2021).
205. Stankunaite, R. *et al.* Liquid biopsy for children with central nervous system tumours: Clinical integration and technical considerations. *Front Pediatr* **0**, 2057 (2022).

206. Malone, H. *et al.* Complications Following Stereotactic Needle Biopsy of Intracranial Tumors. *World Neurosurg* **84**, 1084–1089 (2015).
207. Louis, D. N. *et al.* The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol* **23**, 1231–1251 (2021).
208. Chukwueke, U. N. & Wen, P. Y. Use of the Response Assessment in Neuro-Oncology (RANO) criteria in clinical trials and clinical practice. *CNS Oncol* **8**, CNS28 (2019).
209. Erker, C. *et al.* Response assessment in paediatric high-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group. *Lancet Oncol* **21**, e317–e329 (2020).
210. Hargrave, D., Chuang, N., & Bouffet, E. Conventional MRI cannot predict survival in childhood diffuse intrinsic pontine glioma. *J Neurooncol* **86**, 313–319 (2008).
211. Fangusaro, J. *et al.* Response assessment in paediatric low-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group. *Lancet Oncol* **21**, e305–e316 (2020).
212. Bunda, S. *et al.* Liquid biomarkers for improved diagnosis and classification of cns tumors. *Int J Mol Sci* **22**, (2021).
213. Chakraborty, R. *et al.* Role of liquid biopsy in central nervous system tumors. (2021) doi:10.4103/IJNO.IJNO_425_21.
214. Madlener, S. *et al.* Liquid Biomarkers for Pediatric Brain Tumors: Biological Features, Advantages and Perspectives. *J Pers Med* **10**, 1–15 (2020).
215. Ramkissoon, L. A. *et al.* Genomic Profiling of Circulating Tumor DNA From Cerebrospinal Fluid to Guide Clinical Decision Making for Patients With Primary and Metastatic Brain Tumors. *Front Neurol* **0**, 1233 (2020).
216. McEwen, A. E. *et al.* Beyond the Blood: CSF-Derived cfDNA for Diagnosis and Characterization of CNS Tumors. *Front Cell Dev Biol* **8**, (2020).
217. Yan, W., Xu, T., Zhu, H. & Yu, J. Clinical applications of cerebrospinal fluid circulating tumor DNA as a liquid biopsy for central nervous system tumors. *Onco Targets Ther* **13**, 719–731 (2020).
218. Gatto, L. *et al.* Liquid Biopsy in Glioblastoma Management: From Current Research to Future Perspectives. *Oncologist* **26**, 865–878 (2021).
219. Piccioni, D. E. *et al.* Analysis of cell-free circulating tumor DNA in 419 patients with glioblastoma and other primary brain tumors. *CNS Oncol* **8**, CNS34 (2019).
220. Muralidharan, K. *et al.* TERT Promoter Mutation Analysis for Blood-Based Diagnosis and Monitoring of Gliomas. *Clinical Cancer Research* **27**, 169–178 (2021).
221. Garzia, L. *et al.* A Hematogenous Route for Medulloblastoma Leptomeningeal Metastases. *Cell* **172**, 1050-1062.e14 (2018).
222. Pan W. *et al.* Brain tumor mutations detected in cerebral spinal fluid. *Clin Chem* **61**, 514–522 (2015).
223. Juratli, T. A.*et al.* TERT Promoter Mutation Detection in Cell-Free Tumor-Derived DNA in Patients with IDH Wild-Type Glioblastomas: A Pilot Prospective Study. *Clin Cancer Res* **24**, 5282–5291 (2018).
224. Pan, Y., Long, W. & Liu, Q. Current Advances and Future Perspectives of Cerebrospinal Fluid Biopsy in Midline Brain Malignancies. *Curr Treat Options Oncol* **20**, (2019).
225. De Mattos-Arruda, L. et al.*et al.* Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun* **6**, (2015).
226. Li, J. H.*et al.* Assessment of ctDNA in CSF may be a more rapid means of assessing surgical outcomes than plasma ctDNA in glioblastoma. *Mol Cell Probes* **46**, (2019).
227. Escudero, L. *et al.* Circulating tumour DNA from the cerebrospinal fluid allows the characterisation and monitoring of medulloblastoma. *Nat Commun* **11**, 1–11 (2020).

228. McEwen, A. E., Leary, S. E. S. & Lockwood, C. M. Beyond the Blood: CSF-Derived cfDNA for Diagnosis and Characterization of CNS Tumors. *Front Cell Dev Biol* **8**, 45 (2020).
229. Bounajem, M. T. *et al.* Liquid biopsies for the diagnosis and surveillance of primary pediatric central nervous system tumors: A review for practicing neurosurgeons. *Neurosurg Focus* **48**, (2020).
230. Azad, T.D. *et al.* Liquid biopsy for pediatric diffuse midline glioma: a review of circulating tumor DNA and cerebrospinal fluid tumor DNA. *Neurosurg Focus* **48**, (2020).
231. Liu, A. P. Y. *et al.* Serial assessment of measurable residual disease in medulloblastoma liquid biopsies. *Cancer Cell* **39**, 1519-1530.e4 (2021).
232. Pan, C. *et al.* Molecular profiling of tumors of the brainstem by sequencing of CSF-derived circulating tumor DNA. *Acta Neuropathol* **137**, 297–306 (2019).
233. Sun, Y. *et al.* Exploring genetic alterations in circulating tumor DNA from cerebrospinal fluid of pediatric medulloblastoma. *Scientific Reports 2021 11:1* **11**, 1–8 (2021).
234. Pentsova, E. I. *et al.* Evaluating Cancer of the Central Nervous System Through Next-Generation Sequencing of Cerebrospinal Fluid. *J Clin Oncol* **34**, 2404–2415 (2016).
235. Huang, T. Y.*et al.* Detection of Histone H3 mutations in cerebrospinal fluid-derived tumor DNA from children with diffuse midline glioma. *Acta Neuropathol Commun* **5**, 28 (2017).
236. Li, D. *et al.* Standardization of the liquid biopsy for pediatric diffuse midline glioma using ddPCR. *Scientific Reports 2021 11:1* **11**, 1–10 (2021).
237. Bale, T. A. *et al.* Clinical Experience of Cerebrospinal Fluid–Based Liquid Biopsy Demonstrates Superiority of Cell-Free DNA over Cell Pellet Genomic DNA for Molecular Profiling. *Journal of Molecular Diagnostics* **23**, 742–752 (2021).
238. Liu, Y. At the dawn: cell-free DNA fragmentomics and gene regulation. *British Journal of Cancer 2021 126:3* **126**, 379–390 (2021).
239. Sabedot, T. S. *et al.* A serum-based DNA methylation assay provides accurate detection of glioma. *Neuro Oncol* **23**, 1494–1508 (2021).
240. Bailey, S. *et al.* Clinical Trials in High-Risk Medulloblastoma: Evolution of the SIOP-Europe HR-MB Trial. *Cancers (Basel)* **14**, 374 (2022).
241. Piotrowska, Z. *et al.* Heterogeneity and Coexistence of T790M and T790 Wild-Type Resistant Subclones Drive Mixed Response to Third-Generation Epidermal Growth Factor Receptor Inhibitors in Lung Cancer. *JCO Precis Oncol* **2018**, 1–15 (2018).
242. Russo, M. *et al.* Tumor Heterogeneity and Lesion-Specific Response to Targeted Therapy in Colorectal Cancer. *Cancer Discov* **6**, 147–153 (2016).
243. Goyal, L. *et al.* Polyclonal Secondary FGFR2 Mutations Drive Acquired Resistance to FGFR Inhibition in Patients with FGFR2 Fusion-Positive Cholangiocarcinoma. *Cancer Discov* **7**, 252–263 (2017).
244. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883–892 (2012).
245. Siravegna, G. *et al.* Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med* **21**, 795–801 (2015).
246. Strickler, J. H. *et al.* Genomic Landscape of Cell-Free DNA in Patients with Colorectal Cancer. *Cancer Discov* **8**, 164–173 (2018).
247. Diaz, L. A. *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540 (2012).
248. Chabon, J. J. *et al.* Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. *Nat Commun* **7**, (2016).

249. Misale, S. *et al.* Blockade of EGFR and MEK intercepts heterogeneous mechanisms of acquired resistance to Anti-EGFR therapies in colorectal cancer. *Sci Transl Med* **6**, (2014).
250. Cremolini, C. *et al.* Rechallenge for Patients with RAS and BRAF Wild-Type Metastatic Colorectal Cancer with Acquired Resistance to First-line Cetuximab and Irinotecan: A Phase 2 Single-Arm Clinical Trial. *JAMA Oncol* **5**, 343–350 (2019).
251. Shaw, A. T. *et al.* First-Line Lorlatinib or Crizotinib in Advanced ALK-Positive Lung Cancer. *N Engl J Med* **383**, 2018–2029 (2020).
252. Liu, Z., Rader, J., He, S., Phung, T. & Thiele, C. J. CASZ1 inhibits cell cycle progression in neuroblastoma by restoring pRb activity. *Cell Cycle* **12**, 2210–2218 (2013).
253. Williamson, C. T. *et al.* ATR inhibitors as a synthetic lethal therapy for tumours deficient in ARID1A. *Nat Commun* **7**, 1–13 (2016).
254. Yap, T. A. *et al.* Phase I Trial of First-in-Class ATR Inhibitor M6620 (VX-970) as Monotherapy or in Combination With Carboplatin in Patients With Advanced Solid Tumors. *J Clin Oncol* **38**, 3195–3204 (2020).
255. Lodrini, M. *et al.* Circulating Cell-Free DNA Assessment in Biofluids from Children with Neuroblastoma Demonstrates Feasibility and Potential for Minimally Invasive Molecular Diagnostics. *Cancers (Basel)* **14**, (2022).
256. Cooper, A. J., Sequist, L. v. & Lin, J. J. Third-generation EGFR and ALK inhibitors: mechanisms of resistance and management. *Nature Reviews Clinical Oncology 2022 19:8* **19**, 499–514 (2022).
257. Pailler, E. *et al.* Acquired resistance mutations to ALK inhibitors identified by single circulating tumor cell sequencing in ALK-rearranged non–small-cell lung cancer. *Clinical Cancer Research* **25**, 6671–6682 (2019).
258. Recondo, G. *et al.* Diverse resistance mechanisms to the third-generation ALK inhibitor lorlatinib in ALK-rearranged lung cancer. *Clinical Cancer Research* **26**, 242–255 (2020).
259. Redaelli, S. *et al.* Lorlatinib treatment elicits multiple on- and off-target mechanisms of resistance in ALK-driven cancer. *Cancer Res* **78**, 6866–6880 (2018).
260. Berlak, M. *et al.* Mutations in ALK signaling pathways conferring resistance to ALK inhibitor treatment lead to collateral vulnerabilities in neuroblastoma cells. *Molecular Cancer 2022 21:1* **21**, 1–19 (2022).
261. Van, R., Cuevas-Navarro, A., Castel, P. & McCormick, F. The molecular functions of RIT1 and its contribution to human disease. *Biochemical Journal* vol. 477 2755–2770 Preprint at https://doi.org/10.1042/BCJ20200442 (2020).
262. Pascual, J. *et al.* ESMO recommendations on the use of circulating tumour DNA assays for patients with cancer: a report from the ESMO Precision Medicine Working Group. *Ann Oncol* **33**, 750–768 (2022).
263. Bronkhorst, A. J. *et al.* New Perspectives on the Importance of Cell-Free DNA Biology. *Diagnostics* **12**, 2147 (2022).
264. Abbosh, C., Swanton, C. & Birkbak, N. J. Clonal haematopoiesis: a source of biological noise in cell-free DNA analyses. *Annals of Oncology* **30**, 358 (2019).
265. Bick, A. G. *et al.* Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763–768 (2020).
266. Watson, C. J. *et al.* The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science (1979)* **367**, 1449–1454 (2020).
267. Hsu, J. I. *et al.* PPM1D Mutations Drive Clonal Hematopoiesis in Response to Cytotoxic Chemotherapy. *Cell Stem Cell* **23**, 700-713.e6 (2018).
268. Coombs, C. C. *et al.* Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell* **21**, 374-382.e4 (2017).

269. Bolton, K. L. *et al.* Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nature Genetics 2020 52:11* **52**, 1219–1226 (2020).
270. Coorens, T. H. H. *et al.* Clonal hematopoiesis and therapy-related myeloid neoplasms following neuroblastoma treatment. *Blood* **137**, 2992–2997 (2021).
271. Kleftogiannis, D. *et al.* Identification of single nucleotide variants using position-specific error estimation in deep sequencing data. *BMC Med Genomics* **12**, (2019).
272. Szymanski, J. J. *et al.* Cell-free DNA ultra-low-pass whole genome sequencing to distinguish malignant peripheral nerve sheath tumor (MPNST) from its benign precursor lesion: A cross-sectional study. *PLoS Med* **18**, e1003734 (2021).
273. Lee, B. *et al.* Clinical Relevance of Genomic Changes in Recurrent Pediatric Solid Tumors. *Transl Oncol* **11**, 1390–1397 (2018).
274. de Sá Pereira, B. M. *et al.* Intra-tumor genetic heterogeneity in Wilms tumor samples. *Rev Assoc Med Bras (1992)* **65**, 1496–1501 (2019).
275. Cresswell, G. D. *et al.* Intra-Tumor Genetic Heterogeneity in Wilms Tumor: Clonal Evolution and Clinical Implications. *EBioMedicine* **9**, 120–129 (2016).
276. Madanat-Harjuoja, L. M. *et al.* Circulating Tumor DNA as a Biomarker in Patients with Stage III and IV Wilms Tumor: Analysis from a Children's Oncology Group Trial, AREN0533. *Journal of Clinical Oncology* **50**, (2022).
277. Kolenová, A. *et al.* Targeted inhibition of the MAPK pathway: emerging salvage option for progressive life-threatening multisystem LCH. *Blood Adv* **1**, 352 (2017).
278. Héritier, S. *et al.* Vemurafenib Use in an Infant for High-Risk Langerhans Cell Histiocytosis. *JAMA Oncol* **1**, 836–838 (2015).
279. Heisig, A. *et al.* Vemurafenib in Langerhans cell histiocytosis: report of a pediatric patient and review of the literature. *Oncotarget* **9**, 22236–22240 (2018).
280. Seidel, M. G. *et al.* Clinical implementation of plasma cell-free circulating tumor DNA quantification by digital droplet PCR for the monitoring of Ewing sarcoma in children and adolescents. *Front Pediatr* **10**, (2022).
281. Ruhen, O. *et al.* Molecular Characterization of Circulating Tumor DNA in Pediatric Rhabdomyosarcoma: A Feasibility Study. *JCO Precis Oncol* **6**, (2022).
282. Kahana-Edwin, S. *et al.* Exploration of CTNNB1 ctDNA as a putative biomarker for hepatoblastoma. *Pediatr Blood Cancer* **67**, e28594 (2020).
283. Zviran, A. *et al.* Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring. *Nature Medicine 2020 26:7* **26**, 1114–1124 (2020).
284. *Use of Circulating Tumor DNA for Early-Stage Solid Tumor Drug Development Guidance for Industry DRAFT GUIDANCE*. https://www.fda.gov/vaccines-blood-biologics/guidance-.
285. Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).
286. Klein, E. A. *et al.* Development of a comprehensive cell-free DNA (cfDNA) assay for early detection of multiple tumor types: The Circulating Cell-free Genome Atlas (CCGA) study. *https://doi.org/10.1200/JCO.2018.36.15_suppl.12021* **36**, 12021–12021 (2018).