Development and evaluation of the clinical utility of DNA-based detection methods for high-risk chronic lymphocytic leukaemia and lymphoma



Niamh M Appleby

Dept of Oncology

Keble College

Student number 1142150

Table of Contents

List of Figures 9
List of Tables 13
Acknowledgements 16
Abbreviations 17
Impact of COVID-19 pandemic on my D.Phil studies 22
Declaration 23
Summary of changes to the revised thesis 24
Abstract 26
Chapter 1: Introduction 28
1.Introduction 29
1.1 Clinical and biological heterogeneity characterises chronic lymphocytic leukaemia (CLL) 29
1.1.1 Natural history and clinical management of CLL 30
1.1.2 Pathobiology of CLL 31
1.1.3 Genome features currently used to inform prognosis in CLL 32
1.1.4. Challenges to delivery of precision diagnostics for CLL 34
1.2 High-grade transformation of CLL (Richter's transformation). 35
1.3 Introduction to Diffuse Large B-cell lymphoma 36
1.3.1 Clinical and biological heterogeneity in DLBCL 36
1.3.2 Challenges to the application of precision therapy for DLCBL 37
1.4. Circulating tumour DNA (ctDNA): key concepts and challenges 39
1.4.1 Pre-analytic variables impacting on cell-free DNA and ctDNA 39
1.4.2 Biological confounding factors for circulating tumour DNA 41
1.4.3 PCR-based methods for ctDNA analysis 41
1.4.4 Targeted next generation sequencing (NGS) 41
1.4.5 Limitations of targeted next generation sequencing for ctDNA 42
1.4.6 Whole genome approaches for ctDNA 42
1.4.7 Bioinformatics considerations for ctDNA analysis 45
1.5 Clinical applications of ctDNA in lymphoid malignancy 46
1.5.1 Assessment of disease burden, staging and prognostic scores 46
1.5.2 Minimally invasive molecular profiling of lymphoma 46
1.5.3 Minimally invasive assessment of response 47
1.5.4 Early identification of relapse, transformation and therapy resistance 47

1.6 Nanopore Sequencing in molecular diagnostics 48
1.6.1 Principle of Nanopore sequencing 48
1.6.2 Potential advantages of long-read sequencing using NS 50
1.6.3 Applications of nanopore sequencing to haematological cancers 50
1.6.4 Cell-free DNA (ctDNA & foetal DNA) using Nanopore sequencing 51
1.7: Study Objectives 52
Chapter 2: Methods 53
2.1 Patient samples and additional data sources 54
2.1.1 Existing cohorts of MBL and CLL 54
2.1.2 Samples and IGHV status for OXPLORED participants 54
2.1.3 Richter's transformation 55
2.1.4 Diffuse Large B-cell lymphoma 55
2.1.5 Cell lines and normal control samples 55
2.2. Establishment of the Oxford Premalignant Lymphoproliferative Disorders (OXPLORED) study ... 57
2.2.1 OXPLORED Study Objectives 57
2.2.2 Ethics application, procedures and approvals 58
2.2.3 Study design and procedure 58
2.2.4 Statistical considerations for OXPLORED study design 62
2.3. Establishment of the STELLAR Clinical Trial for Richter's Transformation (RT) 63
2.3.1 Objectives of the STELLAR trial 63
2.3.2 Ethics application, procedure and approval for the STELLAR study 64
2.3.3 STELLAR Study design and procedures 64
2.3.4 Statistical considerations for STELLAR 65
2.4. Sample processing and DNA extraction 70
2.4.1 Cell sorting for circulating clonal B-cells for the OXPLORED cohort 70
2.4.2 Plasma processing prior to cell-free DNA extraction 73
2.4.3 Extraction and Purification of genomic DNA (gDNA) 73
2.4.4 Extraction of cell-free DNA from plasma 74
2.5 DNA Quantification and quality assessment 76
2.5.1 DNA Quantification using Qubit fluorometry 76
2.5.2 DNA Quality control using Nanodrop spectrophotometry 76
2.5.3 Assessment of fragment length using automated electrophoresis 77
2.6 Capture hybridisation and sequencing-by-synthesis (SBS) 78

2.6.1 Introduction to capture hybridisation next generation sequencing (NGS) 78
2.6.2 Panel design and target selection 80
2.6.3 Fragmentation of genomic DNA 80
2.6.4 Library preparation using CAPP-Seq 81
2.6.5 Library preparation using Thruplex-IDT capture based hybridisation. 81
2.6.6 Quantification of DNA libraries by qPCR 83
2.6.7 Hybridisation capture of DNA libraries using xGen Lockdown Probes and Reagents 83
2.6.8 Library loading and Illumina sequencing-by-synthesis. 84
2.6.9 Post sequencing analysis of the capture hybridisation libraries 90
2.6.10 Variant filtering 91
2.6.11 Quantification of ctDNA burden (HGE/ml) 92
2.7 Nanopore sequencing of gDNA and cell-free DNA samples 93
2.7.1 IGHV and *TP53* primer design and PCR amplification 93
2.7.2. Preparation of amplicon libraries 95
2.7.3 Whole Genome Library Preparation using the SQK-RAD-004 rapid sequencing kit (genomic DNA) 97
2.7.4 Preparing cell-free DNA libraries for shallow WGS using nanopore 97
2.7.5 Sequencing on the MinION 98
2.7.6 Post nanopore sequencing analysis 98
2.8 Data analysis and statistical methods 101
2.8.1 General considerations 101
2.8.2 Data visualisation 102
2.8.3 Survival analysis 102
Chapter 3: Sequencing based detection of molecular features that inform prognosis in Monoclonal B-cell Lymphocytosis and early Chronic Lymphocytic Leukaemia 103
3.1 Introduction 104
3.1.1 Prognostic models of MBL and early CLL 104
3.1.2. Imprecise prognostic tools and toxicity of interventions limited the benefit of early intervention for CLL 105
3.1.3 Mutations in CLL-driver genes are detectable at early stages of disease and may refine prognosis 107
3.1.4 Hypothesis and objectives for Chapter 3 108
3.2 Methods 109
3.2.1 Study population and data sources 109

3.2.2 Capture hybridisation library preparation and sequencing 109
3.2.3 Variant filtering for the LYM1 and OXPLORED samples 109
3.2.4 Calculation of International Prognostic Score for Early CLL (IPS-E) 110
3.2.5 Statistical analysis and data visualisation 110
3.3 Results 111
3.3.1 The retrospective MBL and CLL cohort from the LYM1 study 111
Figure 3.3.1.3: Flow diagram of variant filtering for the LYM1 baseline samples 115
3.3.2 Mutations and patterns of cooccurrence in the LYM1 cohort 116
3.3.3 Relationship between IGHV, mutations, IPS-E and treatment-free survival 119
3.3.4 Clinical features of the OXPLORED cohort 123
3.3.5 Mutations and patterns of cooccurrence in the OXPLORED cohort 124
3.3.6 Sequential samples (LYM1 and OXPLORED cohorts) 127
3.4 Discussion 129
Chapter 4: Targeted sequencing of circulating tumour DNA as minimally invasive tool for the assessment of Diffuse Large B-cell lymphoma 132
4.1 Introduction 133
4.1.1 Targeted sequencing of ctDNA on the ACCEPT clinical trial 134
4.1.2 Hypothesis and objectives for Chapter 4 134
4.2 Methods: 135
4.2.1 Sample selection 135
4.2.2 DNA extraction, library preparation and sequencing 136
4.2.3 Post-sequencing analysis and variant filtering strategy 136
4.2.4 Integration of clinical information with sequencing data 139
4.2.5 Statistical tests and data visualisation 139
4.3 Results of the ACCEPT cohort 140
4.3.1. Clinical features of the ACCEPT cohort 140
4.3.2 Germline sequencing, non-cancer controls and artefacts 142
4.3.3. Pre-treatment ctDNA burden and baseline clinical parameters (n=31) 142
4.3.4 Molecular profiling using plasma ctDNA 145
4.3.5 Early eradication of plasma ctDNA in sequential samples 149
4.3.6 Plasma ctDNA changes and overall response 150
4.4 Discussion 154
Chapter 5: Targeted sequencing of circulating tumour DNA as minimally invasive tool for the assessment of Richter's Transformation 157

5.1 Introduction 158
5.1.2 Hypothesis and objectives for ctDNA in RT 159
5.2 Methods: 160
5.2.1 Ethical approval and study populations 160
5.2.2 Plasma cell-free DNA and tumour samples 160
5.2.3 Sequencing target design 160
5.2.4 Library preparation, target enrichment and sequencing 160
5.2.5 Post-sequencing data processing and variant filtering 161
5.2.6 Clinical parameters, response rates and patient outcomes 164
5.2.7 Statistical analysis and data visualisation 164
5.3 Results 165
5.3.1. Pre-treatment ctDNA burden and baseline clinical parameters (n=31) 169
5.3.2. Pre-treatment ctDNA burden and early radiological response (n=31) 171
5.3.3 Pre-treatment ctDNA burden, ORR, PFS and OS (O-CHOP, n=20) 172
5.3.4 Pre-treatment ctDNA detects additional mutations 175
5.3.5 Changes in ctDNA burden on treatment (STELLAR cohort only, n=11/14) 184
5.4 Discussion 187
5.4.1 Pre-treatment plasma ctDNA burden, clinical stage and interim response 187
5.4.2 Impact of pre-treatment plasma ctDNA burden on PFS and OS 187
5.4.3 Additional driver mutations are detectable in the ctDNA 188
5.4.4 Early reductions in plasma ctDNA burden on treatment are uncommon in RT 189
5.4.5 Limitations of this work 190
Chapter 6: Detection of clinically relevant biomarkers in chronic lymphocytic leukaemia by nanopore sequencing 192
6.1 Introduction 193
6.1.1 Detection of CLL biomarkers requires multiple diagnostic tests 193
6.1.2 Limitations of current laboratory techniques 194
6.1.3 Rationale for exploring Nanopore Sequencing 195
6.1.4 Hypothesis and objective for Chapter 6 195
6.2 Methods 196
6.2.1 Sample selection and sources of data for comparison 196
6.2.2 Primers, library preparation and sequencing 197
6.2.3 Identification of *TP53* mutations 197

6.2.4 Determining IGHV status 197
6.2.5 Copy Number calling using NS and QDNASeq 199
6.2.6 Statistical analysis 199
6.3 Results 200
6.3.1 Development of method for post-sequencing identification of IGHV status using control samples and synthetic fastq 200
6.3.2 Nanopore sequencing determines IGHV status in clinical CLL samples: Sequential application of IMGT/HighV-Quest, custom filtering and canu and nanopolish 203
6.3.3 Nanopore sequencing detects *TP53* SNVs and indels 212
6.3.4. Nanopore sequencing detects of Copy Number Alterations in gDNA 214
6.3.5 Proof of principle: combining *TP53*, *IGHV* and shallow WGS to detect clinically relevant molecular features in a single run 219
6.4 Discussion 220
6.4.1 Overview 220
6.4.2 IGHV status 220
6.4.3 Detection of *TP53* mutations 221
6.4.4 Detection of CNAs 222
6.4.5 Summary and future directions 222
Chapter 7: Development and application of a nanopore sequencing method for shallow Whole Genome Sequencing using cell-free DNA 223
7.1 Introduction 224
7.1.1 Detection of copy number alterations from plasma cell-free DNA 224
7.1.2 Genomic features differ between nodal tissue and circulating lymphocytes in Richter transformation (RT) 225
7.1.3 Hypothesis and objectives 225
7.2 Methods 226
7.2.1 Sample selection 226
7.2.2 Preparation of DNA library for nanopore sequencing 227
7.2.3 In-silico experiment evaluates the number of reads required to call CNA 227
7.2.4 Copy number calls using QDNASeq 227
7.2.5 Comparison of copy number alterations from nanopore sequencing with copy number alterations from the O-CHOP trial dataset 228
7.3 Results 229
7.3.1. Detection of copy number alterations in fragmented gDNA from the Burkitt cell line (TC413) using nanopore shallow Whole Genome Sequencing (sWGS) 229

7.3.2 Estimating the tumour fraction required to detect CNAs using synthetic fastq files and serial dilutions 233
7.3.3 Copy number changes in the cell-free DNA samples from the RT clinical trials 238
7.3.4 Comparison of O-CHOP cell-free DNA results with the FISH, karyotyping and previously published SBS WGS results performed as part of the clinical study 238
7.4 Discussion 244
Chapter 8: Discussion 248
8.1 Overview 249
8.2 Unmutated IGVH rearrangements co-exist with variants in key genes in Monoclonal B-cell lymphocytosis and early CLL 250
8.3 Plasma ctDNA provides a minimally invasive tool for with potential for disease assessment in high-grade B-cell lymphoma 251
8.4 Long-read nanopore sequencing can provide comprehensive molecular profiling 251
8.5 Large copy number alterations may be detected in the cell-free DNA using nanopore sequencing 252
8.6 Limitations of this work 252
8.7 Implications of this work 253
8.8 Conclusions 255
References: 256
Appendix A: OXPLORED synopsis and statistical considerations 290
Appendix B: STELLAR trial protocol synopsis 296
Appendix C: Panel targets for the DLBCL panel 304
Appendix D: Panel targets for the CLL and RT panel 305
Appendix E: Circulating tumour DNA variants detected in the ACCEPT (DLBCL) cohort 306
Appendix F: Circulating tumour DNA variants detected in the Richters Transformation cohorts 307
Appendix G: Examiners comments and actions taken for the revised thesis 308

List of Figures

Chapter 1: Introduction

Figure 1.4.1.1: Origins of plasma cell-free DNA 39

Figure 1.4.5.1: Schematic showing the differences between a targeted panel-based approach and shallow WGS 43

Figure 1.6.1.1: Illustration of the process of nanopore sequencing 48

Chapter 2: Materials and Methods

Figure 2.1.5.1: Workflow of patient samples, sequencing methods and thesis structure 55

Figure 2.2.3.1: Participant flow in the OXPLORED Study 59

Figure 2.2.3.2: Screenshot from OpenClinica showing data collection matrix. 60

Figure 2.3.3.1: Schematic for the STELLAR Randomised and Cohort 1 (acalabrutinib monotherapy) 65

Figure 2.3.3.2: Schematic for STELLAR Cohort 2 (RT arising on ibrutinib). 66

Figure 2.4.1.1: Sample flow for the MBL/CLL cohort of the OXPLORED study 70

Figure 2.4.4.1: Workflow of the QIAmp circulating cell-free DNA extraction kit 73

Figure 2.6.1.1: Molecular barcoding for detection of low frequency variants 77

Figure 2.6.8.1: Schematic of sequencing-by-synthesis (SBS) 84

Figure 2.6.9.1: Schematic of OMDC bioinformatics pipeline 89

Figure 2.7.2.2: Tapestation electrophenogram of BM2464 (66243) IGHV library 598bp 92

Chapter 3: Detecting low frequency variants in a retrospective cohort of Monoclonal B-cell Lymphocytosis and early Chronic Lymphocytic Leukaemia

Figure 3.3.1.1: CONSORT diagram of participants/samples 110

Figure 3.3.1.2: Boxplot of coverage of target genes 112

Figure 3.3.1.3: Flow diagram of variant filtering for the LYM1 baseline samples 113

Figure 3.3.2.1: Number of variants by IGHV status (n=17) 114

Figure 3.3.2.2: Co-occurrence heatmap of the baseline samples 116

Figure 3.3.3.1: Treatment-free survival in the LYM1 cohort by IGHV status (n=15) 118

Figure 3.3.3.2: Treatment-free survival by IGHV status and presence of mutations in the LYM1 cohort 119

Figure 3.3.3.3: Treatment-free survival by IPS-E score in the LYM1 cohort (n =14) 120
Figure 3.3.4.1: CONSORT diagram of OXPLORED participants/samples 122
Figure 3.3.5.1: Mutations detected at baseline in the OXPLORED samples 123

Chapter 4: Targeted sequencing of circulating tumour DNA as a minimally invasive tool for the assessment of Diffuse Large B-cell lymphoma

Figure 4.2.1.1: Schematic for ACCEPT trial 133
Figure 4.2.1.2: ACCEPT patients and sample flow-chart 136
Figure 4.2.3.1: Schematic illustrating the variant filtering strategy for the ACCEPT cell-free DNA libraries 136
Figure 4.3.1.1: Flow chart summarising outcomes for the ACCEPT cohort 139
Figure 4.3.3.1: Baseline ctDNA burden and IPI score (A) and response at EOT (B) 141
Figure 4.3.3.2: Baseline ctDNA burden and Progression Free Survival 142
Figure 4.3.3.3: Flow diagram showing patients by ctDNA status at baseline, radiological responses after treatment and outcome at 24 months 143
Figure 4.3.4.1: Baseline ctDNA variants by frequency (n=22) 144
Figure: 4.3.4.2: Baseline cDNA variants by patient 146
Figure 4.3.6.1: Dynamic changes in ctDNA by variant during treatment (n=31) 149
Figure 4.3.6.2: Dynamic changes in ctDNA by participant during treatment (n=31) 150

Chapter 5: Targeted sequencing of circulating tumour DNA as a minimally invasive tool for the assessment of Richter's Transformation (RT)

Figure 5.2.5.1: Source of Richter's Transformation tumour and plasma cell-free DNA samples 161
Figure 5.2.5.2: Schematic illustrating the variant filtering strategy for the RT cell-free DNA libraries 162
Figure 5.3.1: Boxplot of sequencing coverage of the Richter's tumours by gene 166
Figure 5.3.2: Boxplot of sequencing coverage of the Richter's cell-free DNA samples by gene 167
Figure 5.3.1.1: Relationship between pre-treatment ctDNA and Clinical Stage 169
Figure 5.3.2.1: Cox proportional hazards of factors impacting response rate at interim assessment 170
Figure 5.3.3.1: PFS and OS by ctDNA burden in the O-CHOP cohort (n=20) 172
Figure 5.3.3.2 Cox proportional hazards of factors impacting clinical outcomes 173
Figure 5.3.4.1: Genes affected by ctDNA mutations 174

Figure 5.3.4.2: Genes affected by mutations in the tumour (n=15) 176
Figure 5.3.4.3: Venn diagram of mutations in the paired RT tumour and ctDNA samples (n=8) 179
Figure 5.3.4.4: Frequency of variants detected in ctDNA by patient (n=28) 181
Figure 5.3.5.1: Sequential ctDNA burden by variant during treatment in the STELLAR samples (n=11) 184
Figure 5.3.5.2: Dynamic changes in ctDNA by variant during treatment in the STELLAR samples (n=11) 185

Chapter 6: Detection of clinically relevant biomarkers in chronic lymphocytic leukaemia by nanopore sequencing

Figure 6.2.4.1: Summary of the post-sequencing analysis for IGHV status 197
Figure 6.3.1.1: Number of reads supporting IGHV3-72*01 (red) in OXP10073, the three test files and the normal control 201
Figure 6.3.2.1: Flowchart of clinical samples and source of IGHV rearrangment data 203
Figure 6.3.2.2: Dominant VH clones by NS in the OXPLORED cohort 207
Figure 6.3.2.3: Comparison of the percentage germline homology for the IGHV rearrangment detected by SBS and by Nanopore Sequencing 209
Figure 6.3.2.4: Comparison of the percentage germline homology for the IGHV rearrangment detected by SS and by Nanopore Sequencing 210
Figure 6.3.3.1: Comparison of *TP53* variant allele frequency by Illumina SBS and by Nanopore Sequencing 212
Figure 6.3.4.1: Published Consensus Karyotype for the Burkitt lymphoma cell line CA46 (TC413) . 213
Figure 6.3.4.2: NS and QDNASeq analysis of the Burkitt cell line gDNA 214
Figure 6.3.4.3: NS and QDNASeq analysis of Burkitt cell line gDNA (TC413) (1.05x depth) 215
Figure 6.3.4.4: Visualisation of tri12 in early CLL case (BM-2483) 217

Chapter 7: Development and application of a nanopore sequencing method for shallow Whole Genome Sequencing using cell-free DNA

Figure 7.2.1.1: Summary of the samples and workflow for Chapter 7 225
Figure 7.3.1.1: Visualisation of Copy Number Alterations detected using NS and QDNAseq analysis of TC413 fragmented gDNA 230
Figure 7.3.2.1: Visualisation of Copy Number Alterations detected using NS and QDNAseq analysis of synthetic 'Tumour fraction 0.33' fastq file 233
Figure 7.3.2.2. Visualisation of Copy Number Alterations detected using NS and QDNASeq analysis of the 'Diluted 0.25 tumour fraction' ctDNA 235

Figure 7.3.4.1: Relationship between CNAs detected in the three DNA sources in RT 238

Figure 7.3.4.2: Visualisation of Copy Number Alterations detected using NS and QDNASeq analysis for CH1005 cell-free DNA 240

List of Tables

Chapter 2: Materials and Methods

Table 2.2.1.1: Objectives of the OXPLORED study 57

Table 2.2.3.1: Summary of the clinical and laboratory assessments for Cohort 1 (MBL/CLL) participants in OXPLORED 59

Table 2.3.1.1: Objectives of the STELLAR study 63

Table 2.3.3.1: Summary of schedule of events for the STELLAR Randomised cohort and Cohort 2 (R-CHOP with acalabrutinib) 68

Table 2.4.1.1: Guideline for leukocyte count and volume of CD19+ Selection Cocktail and Rapidspheres 70

Table 2.6.3.1: Fragmentation Conditions used on Covaris M200 instrument 79

Table 2.6.5.1: Thruplex Library Amplification Reaction 81

Table 2.6.5.2: Thruplex Tag Seq HV Library Amplification Reaction 81

Table 2.6.7.1: Post capture amplification 83

Table 2.6.8.1: Comparison of Illumina SBS instruments and flow cells used in the thesis 87

Table 2.6.8.2 Volumes for library preparation for NextSeq 550 88

Table 2.7.1.1: Primer Sequences to amplify the IGHV locus 92

Table 2.7.1.2: Primer Sequences to amplify *TP53* 94

Table 2.7.2.1: Preparation of the LSK-109 barcoded amplicon library for loading 95

Table 2.7.3.1: Preparation of the RAD004 library for loading 96

Table 2.8.2.1: Sample confusion matrix 101

Chapter 3: Detecting low frequency variants in a retrospective cohort of Monoclonal B-cell Lymphocytosis and early Chronic Lymphocytic Leukaemia

Table 3.1.1.2: CLL Prognostic Scoring Systems and Performance in Predicting Time to First Treatment (TTFT) 105

Table 3.3.1.1: Clinical characteristics of the LYM1 Cohort sequenced (n=17) 112

Table 3.3.2.1: Variants identified in the baseline LYM1 samples 116

Table 3.3.4.1: Clinical features of the OXPLORED participants sequenced 122

Table 3.3.5.1: Variants identified in the OXPLORED CD19+ B-cells at baseline 125

Table 3.3.6.1 Changes in variants detected in the sequential samples (n=8) 127

Chapter 4: Targeted sequencing of circulating tumour DNA as a minimally invasive tool for the assessment of Diffuse Large B-cell lymphoma

Table 4.3.1.1: Clinical features of the ACCEPT cohort (n=31) 139
Table 4.3.4.1: Plasma ctDNA *TP53* variants (n=8) 146
Table 4.3.5.1: Confusion matrix for dynamic ctDNA changes and EOT response 148
Table 4.3.6.1: Variants detected in ctDNA at end of treatment, Day 189 (n=6) 152

Chapter 5: Targeted sequencing of circulating tumour DNA as a minimally invasive tool for the assessment of Richter's Transformation (RT)

Table 5.3.1: Clinical features of the RT patients submitting baseline plasma samples 164
Table 5.3.4.1: Variants detected in baseline plasma samples from RT patients by frequency (top 20 variants displayed) 175
Table 5.3.4.2: Thirty-seven variants detected in the Richters tumours 177
Table 5.3.4.3: Fifteen TP53 variants detected in the ctDNA of Richter's Transformation 182
Table 6.3.1.1: Fastq files created to estimate limit of detection 194

Chapter 6: Detection of clinically relevant biomarkers in chronic lymphocytic leukaemia by nanopore sequencing

Table 6.2.1.1: Sources of samples and data for comparison 195
Table 6.3.1.1: Fastq files created to estimate limit of detection 199
Table 6.3.1.2: Method A: IGHV calls using canu and nanopolish in the test files 200
Table 6.3.1.3: Method B: IGHV calls using unpolished nanopore fastq and IGMT/HighVquest 200
Table 6.3.2.1: Comparison of nanopore sequencing IGHV results with other methods 204
Table 6.3.2.2: Secondary IGHV rearrangments detected by NS in five samples 206
Table 6.3.2.3: Discordant VH family results between Nanopore and Sanger Sequencing are predominantly seen in V4 family rearrangments 208
Table 6.3.2.5: Inter-run variability for NS to detect IGHV status 211
Table 6.3.3.1: *TP53* variants detected by nanopore and Illumina SBS 211
Table 6.3.4.1: Summary QDNASeq copy number calls from TC413 gDNA (long-reads) with declining coverage 216
Table 6.3.5.1: *TP53*, IGHV and CNA detected by NS in a single run in CLL374 218

Chapter 7: Development and application of a nanopore sequencing method for shallow Whole Genome Sequencing using cell-free DNA

Table 7.2.3.2: Variable proportions of the fragmented gDNA (Burkitts cell line TC413) and normal non-cancer control cell-free DNA 226

Table 7.3.1.1: Copy number calls detected by QDNASeq in TC413 gDNA compared to fragmented gDNA and as coverage declines 231

Table 7.3.2.1: Artefactual copy number calls in control cell-free DNA 232

Table 7.3.2.2: NS and QDNASeq analysis for synthetic Tumour Fraction 0.33 ctDNA 233

Table 7.3.2.3: Summary of Copy Number gains and losses detected as the proportion of tumour fragments decreases 236

Table 7.3.2.4: Recommendations for shallow WGS of cell-free DNA using NS based on the fragmented gDNA experiments 237

Table: 7.3.4.1: Comparison of clinically reported CNA ctDNA WGS (nanopore), PB-CLL WGS and tDNA WGS (Illumina) 241

Acknowledgements

Foremost, I would like to thank my supervisor Professor Anna Schuh for her guidance, her knowledge and her boundless enthusiasm for using molecular diagnostics to improve patient care. She has shown unwavering support (especially during the challenges of returning to studies after maternity leave and the disruption of COVID-19) and opened numerous doors for me in academic and clinical spheres.

My special thanks to Dr Laura López Pascua for her clear explanations and endless patience as I took my first steps in bioinformatics, and to Dr Kate Ridout for trouble-shooting data analysis queries and for tutorials on her kitchen table. And a special mention to Dr Dimitris Vavoulis for answering my questions and for his valuable feedback on my preliminary results. I would like to thank my post-doctoral colleagues: Dr Adam Burns for sharing his enthusiasm and expertise on nanopore sequencing; Dr Anthony Cutts for painstakingly teaching me each step laboratory workflow and for his help with troubleshooting; Dr Adele Timbs for bridging the divide between the research and diagnostics services and for commiserating with me on the trials of balancing parenting with academic work.

I am sincerely thankful to Hélène Dreau, Chief Scientist at the Oxford Molecular Diagnostics Service (OMDC) for her leadership, her expertise and her kindness. My sincere gratitude to Dr Frances Willenbrock for her painstaking eye for detail and her invaluable feedback on academic writing.

I am sincerely thankful to my clinical colleagues, especially Dr Toby Eyre for his enthusiasm for clinical trials, patient care and acting as a sounding board for challenging clinician and trial queries.

I would like to acknowledge the support of the Trials Units at the Churchill Hospital, the Oncology Clinical Trials Office (OCTO) at the University, patient support groups including CLL Support Association and Leukaemia Care and the many patients who contributed samples to this research.

On a personal note, I would like to thank my sister Caoileann and my good friends Nadri, Anna and Bernadette for their steadfast support. John, my husband for cheerfully packing up and following me to this Oxford adventure. John never doubted me. And Liam, who makes me smile every day and is so curious about the whole world.

Abbreviations

ABC	Activated B-cell
AE	Adverse Events
AF	allele frequency
AID	Activation induced deaminase
AI-REAL	Aggressive Infection-Related East Africa Lymphoma
AML	Acute myeloid leukaemia
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
ATM	ATM Serine/Threonine Kinase
ATR	ataxia telangiectasia and Rad3-related
AutoSCT	Autologous stem cell transplant
BAM	Binary Alignment map, Binary version of SAM file
BCR	B-cell receptor
β2M	beta2 microglobulin
BIRC3	Baculoviral IAP repeat containing 3
BM	Bone marrow
BMA	Bone marrow aspirate
BMAT	Bone marrow aspirate and trephine
BTK	Bruton's tyrosine kinase
BTKi	Bruton's tyrosine kinase inhibitors
BWA	Burrows-Wheeler Aligner
CAPP-seq	Cancer Personalised Profiling by Deep Sequencing
CBA	Chromosomal Banding Analysis
CD	Clusters of Differentiation
cDNA	complementary DNA
CDR	Complementary determining regions
CDKN2A	cyclin-dependent kinase Inhibitor 2A
cfDNA	cell-free DNA
CGH	Comparative Genomic Hybridisation
CK	Complex Karyotype
CHIP	Clonal haematopoiesis of indeterminant potential
CI	Chief Investigator
CLL	Chronic lymphocytic leukaemia
CMR	Complete Metabolic Response (by PET-CT)
CNA	Copy Number Alterations
CNS	Central Nervous System
cnLOH	copy neutral loss of heterozygosity
COO	cell of origin
COSMIC	Catalogue of Somatic Mutations in Cancer

CR	Complete Response
CTU	Clinical Trials Unit
CSV	comma separated values (file)
CT	Computer Tomography
ctDNA	Circulating tumour DNA
dbSNP	The Single Nucleotide Polymorphism Database
dsDNA	double stranded DNA
DDR	DNA Damage response
DLBCL	Diffuse large B-cell lymphoma
DMC	Data Monitoring Committee
DNA	Deoxyribonucleic acid
dNTPs	deoxyribonucleotide triphosphate
EB	Elution Buffer
EBV	Epstein Barr Virus
ECOG	Eastern Cooperative Oncology Group
EDTA	Ethylenediaminetetraacetic acid
EGR2	Early growth response 2
EOT	End of treatment
EORTC	European Organisation for Research and Treatment of Cancer
EPTU	Early Phase Trials Unit
ERIC	European Research Initiative of CLL
FBC	Full Blood Count
FBS	fetal bovine serum
FBXW7	F-box and WD repeat containing protein 7
FCR	fludarabine, cyclophosphamide and rituximab
FF	Fresh, frozen (tissue sample)
FFPE	Formalin fixed, paraffin embedded (biopsy)
FISH	Fluorescent in-situ hybridization
GC	Genomic Complexity
GCB	Germinal Centre B-cell
GCLLSG	German CLL Study Group
gDNA	Genomic DNA
GEP	Gene Expression Profiling
HGE	Haploid Genome Equivalents
HMDS	Haematological Malignancy Diagnostic Service, Leeds; collaborator on clinical trials
HR	Hazard ratios
ICGC	International Cancer Genome Consortium
IGH	Immunoglobulin heavy chain
IGHD	Immunoglobulin heavy chain diversity region

IGHJ	Immunoglobulin heavy chain joining region
IGHV	Immunoglobulin heavy chain variable region
IGL	Immunoglobulin light chain
IMGT	The International ImMunoGeneTics resource for immunoglobulin and T-cell receptor gene rearrangement data
indel	small insertion/deletion
IPI	International Prognostic Index
IPS-E	International Prognostic Score for early asymptomatic CLL
IRF4	Interferon Regulatory Factor 4
KMT2C	Lysine Methyltransferase 2C
KRAS	KRAS Proto-Oncogene, GTPase
LN	Lymph node
Log2R	Log2 Ratio
MBA	molecular barcode adaptors
MBL	Monoclonal B-cell lymphocytosis
MDR	Minimally Deleted Regions
MGUS	Monoclonal gammopathy of uncertain significance
m-IGHV	mutated immunoglobulin heavy chain rearrangement
MMR	Major Molecular Response
MRD	Measurable residual disease
MSI	microsatellite instability
MTA	Materials Transfer Agreement
MTV	Metabolic Tumour Volume
MYD88	Myeloid differentiation primary response 88
NFKB	Nuclear factor kappa beta
NaOH	sodium Hydroxide
NCRI	National Cancer Research Institute
NGRL	National Genome Reference Library
NGS	Next generation sequencing
NHL	Non- Hodgkin lymphoma
NS	Nanopore sequencing
NTC	No template control
OCTO	Oncology Clinical Trials Office
OMDC	Oxford Molecular Diagnostic Centre
ONT	Oxford nanopore technologies
ORB	Oxford Radcliffe Biobank
ORR	Overall response rate
OS	Overall Survival
PAX-5	Paired box 5
PB	Peripheral blood

PB-CLL	circulating CLL cells in the peripheral blood
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PD	Progressive disease
PET-CT	Position Emission Tomography
PFS	Progression Free Survival
Polyphen	Polymorphism Phenotyping
POT1	Protection of telomeres 1
PPV	Positive predictive value
PR	Partial Response
QoL	Quality of Life
qPCR	quantitative PCR
R/R	Relapsed and/or Refractory
RAG	Recombination activating gene
R-CHOP	cyclophosphamide, doxorubicin, vincristine, prednisolone, and rituximab
RIGHT	Research and Innovation for Global Health Transformation
rpm	revolutions per minute
RT	Richter Transformation, high-grade transformation of CLL
SAM	Sequence Alignment Mapping file
SAMHD1	SAM domain and HD domain 1
SBS	Sequencing-by-synthesis
SD	Stable disease
SF3B1	Splicing factor 3B subunit 1
SHM	Somatic hypermutation
SIFT	Sorting Intolerant from Tolerant
SLL	Small lymphocytic lymphoma
SOP	standard operating procedure
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SM	smouldering myeloma
SS	Sanger sequencing
TAP	Trials Acceleration Programme
TdT	Terminal deoxynucleotidyl transferase
TFS	Treatment free survival
TTFT	Time to first treatment
TGS	Targeted gene sequencing
TMG	Trial Management Group
TP53	Tumour protein 53

TSC	Trial Safety Committee
TTFT	Time to first treatment
UCSC Genome Browser	University of California, Santa Cruz Genome Browser
u-IGHV	unmutated immunoglobulin heavy chain variable region
UTR	untranslated region
UMI	unique molecular identifier
VAF	Variant allele frequency
VCF	Variant calling file
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WM	Waldenstrom's macroglobulinaemia
XPO1	Exportin 1
µg	Microgram
µl	Microlitre
µM	Micromolar

Impact of COVID-19 pandemic on my D.Phil studies

The pandemic impacted on my D.Phil studies in four main areas; redeployment to NHS clinical services; restricted access to laboratory; trial recruitment and via my caring responsibilities.

I suspended my studies from Trinity Term 2019 to Trinity Term 2020 for maternity leave.

The UK nationwide lockdown was introduced on 23 March 2020.

Redeployment to NHS clinical services

From March 2020 to July 2020, clinicians were redeployed from university work to direct clinical work. I supported the NHS lymphoma service during this time.

Restricted access to the laboratory

Staff numbers at the OMDC were restricted to facilitate social distancing from July 2020 until July 2021. When my NHS redeployment finished, I had a maximum of two days per week for benchwork.

As I had limited benchwork time, I focused on nanopore sequencing.

Clinical trial recruitment and sample availability

The OXPLORED study closed to recruitment 18 March 2020 to 7 August 2020. Eligible people were clinically vulnerable to COVID-19 and advised to 'shield' until July 2021, limiting recruitment.

The STELLAR trial opened on 23 Jul 2019 and was anticipated fully recruit within 3 years. The STELLAR trial closed to recruitment on 24 March 2020 until 21 July 2020. The study revised its recruitment goals, aiming to be fully recruited by 31 May 2024 and follow-up complete 31 May 2026.

As I had fewer prospective samples, I used retrospective CLL and RT samples for my D.Phil.

Caring responsibilities

Childcare facilities were closed from 23 March 2020 until 15 June 2020. In addition, the nursery closed on two further occasions (four weeks in total) due to COVID-19 outbreaks among staff.

Declaration

Experiments described in this thesis were designed, performed and analysed by myself.

Assistance of colleagues for specific work packages are acknowledged in the table below and in the relevant sections.

Work package	Impact on D.Phil	Colleague
Development of the bioinformatics pipeline for the ctDNA analysis	Chapter 4 Chapter 5	Dr Ed Wilson with Dr Dimitris Vavoulis, Dr Kate Ridout and Dr Laura Lopez-Pascua
Design of the capture hybridisation panel used for Diffuse Large B-cell lymphoma	Chapter 4	Dr Laura Lopez-Pascua
Design of the IGHV primers	Chapter 6	Dr Adam Burns
Analysis of Illumina WGS datasets from the O-CHOP study. (This data is previously published as *Blood* (2021) 137 (20): 2800–2816)	I refer to the published results in Chapter 5 and Chapter 7.	Dr Dimitris Vavoulis and Dr Jenny Klintman
Training in next-generation sequencing platforms and troubleshooting	Throughout	Dr Helene Dreau and Dr Anthony Cutts
Training in R-programming and bioinformatics analysis	Throughout	Dr Kate Ridout, Dr Laura Lopez-Pascua, Dr Pauline Robbe and Dr Dimitris Vavoulis

Summary of changes to the revised thesis

My thesis has been substantially revised and improved following examiners' criticism. A detailed description of all changes can be found in Appendix F.

Chapter 1: Introduction

The Introduction incorporates the recent key publications revealing the complex relationship between genetic and epigenetic features in chronic lymphocytic leukaemia (CLL) (Robbe et al. 2022; Knisbacher et al. 2022) and Richter's Transformation (RT) (Parry et al. 2023b; Nadeu et al. 2022). In response to examiners feedback, an outline of the role of microenvironment in CLL and the function of somatic hypermutation in normal B-cell development are included.

Chapter 2: Methods

The examiners commented that the flow of patients through each study was unclear. include flow charts visualising the source of clinical samples for each sequencing work package (Section 2.1.1, Figure 3.3.1.1, Figure 4.2.1.2, Figure 5.2.5.1, Table 6.2.1.1 and Figure 7.2.1.1). The description of DNA extraction and quality assessment is shortened, and the more detail is provided on library preparation. The data analysis and statistical methods section are expanded.

Chapter 3: Sequencing based detection of molecular features that inform prognosis in Monoclonal B-cell Lymphocytosis and early Chronic Lymphocytic Leukaemia

For the Monoclonal B-cell lymphocytosis (MBL) and CLL samples in Chapter 3, the sequencing results are completely reanalysed prioritising variants with a COSMIC ID previously reported in haematological cancers, in line with the European Research in CLL (ERIC) in Harmony strategy (Mansouri et al. 2023). Variants identified are described using HGVS nomenclature (Table 3.3.2.1, Table 3.3.5.1). As this strategy results in fewer patients with mutations, pathway-based analysis is omitted. The link between the pilot study presented and the OXPLORED study is more clearly articulated and a power calculation for this study is included. In the intervening months, OXPLORED WGS datasets became available and IGHV status determined by IgCaller has been included here and in Chapter 6.

Chapter 4: Targeted sequencing of circulating tumour DNA as minimally invasive tool for the assessment of Diffuse Large B-cell lymphoma

Chapters 4 reports circulating tumour DNA (ctDNA) sequencing using Diffuse Large B-cell Lymphoma (DLBCL) samples from the ACCEPT clinical trial. The introduction is more focused. As patients submitting plasma cell-free DNA samples are a subset of the full trial cohort, outcome data is provided for patients providing samples only. The figures are amended. As requested, the variants detected are reported in Appendix E and a table of variants detected at end of treatment has been included (Table 4.3.6.1). The discussion and conclusions map more closely with the data available from the study cohort.

Chapter 5: Targeted sequencing of circulating tumour DNA as minimally invasive tool for the assessment of Richter's Transformation

The revised introduction to Richter's Transformation in Chapter 5 includes two significant studies of RT (Parry et al. 2023b; Nadeu et al. 2022) and studies examining ctDNA in CLL (Furstenau et al. 2022; Trummer et al. 2022; Albitar et al. 2017). As recommended by the examiners, I have expanded on my role in the design of the STELLAR trial (Section 2.3.3).

The restructured results consider all baseline RT samples from both trials (O-CHOP and STELLAR) together in Section 5.3.1-5.3.2, followed by outcome data (O-CHOP only) and sequential samples (STELLAR only). Again, I reanalysed the sequencing results with post-sequencing variant filtering prioritising variants with a COSMIC ID previously reported in haematological cancer.

Chapter 6: Detection of clinically relevant biomarkers in chronic lymphocytic leukaemia by nanopore sequencing

Chapter 6 introduces nanopore sequencing to determine key biomarkers for CLL; *TP53* mutations, IGHV status, copy number changes and genomic complexity. The revised chapter summarises the results and displays comparison with Sanger Sequencing and Sequencing-By-Synthesis (Table 6.3.2.1, Figure 6.3.2.3, Figure 6.3.2.4), aiming to improve clarity for the reader. Further explanation on the bioinformatics packages and tools has been included (Section 2.1.2 and Section 6.2.2). I have expanded on the limitations of the work with respect to guidelines for detection of *TP53* and IGHV status.

Chapter 7: Development and application of a nanopore sequencing method for shallow Whole Genome Sequencing using cell-free DNA

Preliminary experiments using nanopore sequencing (NS) for cell-free DNA are presented in Chapter 7. One examiner suggested combining both nanopore sequencing chapters. I have chosen to retain two chapters as Chapter 6 examines a potential clinical application of NS, but Chapter 7 is more exploratory in approach. Regions of copy number gain and loss are systematically analysed, as described by Smolander et al. (Smolander et al. 2021) and comparison with different DNA sources provided in Table 7.3.1.1 and Table 7.3.2.3.

The results are summarised using fewer figures and tables that the original thesis submission. I have removed discussion on the detection of methylation profiles and detection of translocations using NS to focus on discussion of the results presented. I have expanded on the limitations of my results.

Chapter 8: Discussion

I have restructured the discussion to summarise my results in the order reported in the thesis and moved specific sections to the chapter specific discussions. I have clarified the relationship between my findings and published literature.

I have outlined how the work presented in my thesis influenced ongoing clinical studies (OXPLORED, STELLAR, ReMoDL-A) and the NIHR AI-Real programme grant.

General revisions

Throughout, the figures have been revised to better summarise the data. Most figures reproduced from published literature have been replaced by figures created in BioRender. I have retained the figures from my STELLAR trial manuscript (Appleby et al. 2019). All figures are followed by figure legends. Errors related to the censoring of data for Kaplan Meier curves have been corrected.

Overall, the revised thesis presents a clearer and more succinct manuscript, and I thank the examiners for their feedback.

Abstract

Clinical and biological heterogeneity characterise B-cell neoplasms. Chronic lymphocytic leukaemia (CLL) and the precursor Monoclonal B-cell Lymphocytosis (MBL) offer a model to study molecular features of progression to cancer. Despite advances in treatments targeting the B-cell receptor, CLL remains incurable. A minority (3-20%) experience transformation to high-grade lymphoma or Richter's transformation (RT) carrying a dismal prognosis and representing an unmet need for effective therapy.

Clinical prognostic scores developed for CLL on chemoimmunotherapy lack precision for early disease. I describe the establishment and objectives of an observational study, OXPLORED to collect sequential samples from 800 people MBL and CLL. The study aims to molecularly characterise progressors and those with long-term indolent disease to develop a predictive risk score. As a pilot study, I interrogated a retrospective early CLL cohort (n=17) using capture hybridisation targeted sequencing. Mutations were absent from 11/17 (64.7%) samples. Where present, mutations co-occurred with unmutated immunoglobulin heavy chain variable region (u-IGHV). However, targeted sequencing risks being uninformative in early disease, arguing for whole genome approaches for the main OXPLORED study.

Inadequate dissection of heterogeneity remains a barrier to precision therapy for high-grade B-cell lymphoma. Molecular tools capturing tumour heterogeneity and stratifying Diffuse Large B-cell lymphoma (DLBCL) and RT are needed, delivering genomic information within the diagnostic pathways. Highly sensitive sequencing of circulating tumour DNA (ctDNA) offers a tool to detect genomic features of nodal disease and can be repeated, enabling treatment to be tailored depending on early response.

Plasma samples from 61 high-grade B-cell lymphoma patients were sequenced. Baseline ctDNA burden correlated with International Prognostic Index in DLBCL (kruskal-wallis, p = 0.0017). Baseline ctDNA did not inform Progression Free Survival in DLBCL (p=0.18) nor RT (p= 0.65). Sequencing of ctDNA identified *TP53* mutations in 8/31 (25.5%) DLBCL and 13/31 (41.9%) RT. The positive predictive value of undetectable ctDNA (either at enrolment or after two treatment cycles) in identifying responding DLBCL patients is 92.3% (CI 67.6%-98.6%). By contrast, RT patients rarely cleared ctDNA on treatment.

The presence of *TP53* aberrations, the IGHV status and specific copy number changes is central to CLL prognostication, but burdensome laboratory methods limit universal availability of testing. Using a combination of amplicon sequencing and shallow whole genome sequencing (WGS), a nanopore sequencing based method can determine these key CLL biomarkers in clinical samples, which can be applied for simultaneous detection of all biomarkers in a single sequencing run. Using nanopore, germline homology of the rearranged IGHV can be determined with high concordance compared to Sanger Sequencing ($R=0.99$, $p = 6.5x10^{-13}$).

Next, I modified nanopore amplicon sequencing to detect copy number alterations in fragmented gDNA and cell-free DNA. Using nanopore-based shallow WGS, nodal and marrow-based copy number gains and losses from RT patients could be detected from plasma cell-free DNA (Chapter 7).

Although my conclusions are limited by small sample numbers, the exploratory work conducted in this thesis contributed to three national studies across the spectrum of B-cell neoplasms; the OXPLORED study for MBL and early CLL; the ctDNA exploratory objectives within the REMoDL-A trial for DLBCL and the STELLAR trial and RT.

.

Chapter 1: Introduction

1.Introduction

The development and application of genomic technology has advanced the understanding of haematological cancer pathophysiology. Sequencing-based data has found applications in the diagnosis, risk-stratification, treatment selection, detection of measurable residual disease (MRD) and informed the development of novel cancer therapies.

In my thesis, I applied sequencing technologies across the spectrum of B-cell malignancies from precursor clonal B-cells (Monoclonal B-cell lymphocytosis, (MBL)) to chronic lymphocytic leukaemia (CLL) to clinically aggressive lymphomas (*de novo* Diffuse Large B-cell lymphoma (DLBCL)) and Richter's transformation, (RT). I established a prospective observational study for early indolent B-cell lymphoproliferative disorders, OXPLORED and a randomised trial for RT, STELLAR. Both studies and laboratory programmes are further described in Section 2.2.

In this chapter, I discuss the challenges presented by clinical and biological heterogeneity in B-cell lymphomas. I introduce the potential role of deep targeted next-generation sequencing (NGS) of tumour and cell-free DNA to address knowledge gaps in prognostication and risk stratification. I introduce the hypothesis that deep targeted sequencing of genomic DNA (gDNA) and circulating tumour DNA (ctDNA) can identify clinically important mutations in B-cell malignancies.

Clinical application of molecular diagnostics is constrained by access to resource intensive sequencing-by-synthesis (SBS) platforms. Thus, I review the potential advantages and limitations of nanopore sequencing (NS) in molecular diagnostics. In the work presented in Chapters 6 and 7, I report the development of NS method for identifying biomarkers in CLL and RT.

1.1 Clinical and biological heterogeneity characterises chronic lymphocytic leukaemia (CLL)

CLL is the commonest adult leukaemia, with a prevalence of 44.1 per 100,000 UK adults (Li, Smith, et al. 2016). The diagnosis of CLL is established by the persistence of > 5 x 10^9/L clonal CD5+ CD23+ co-expressing B-lymphocytes for more than three months. The nodal subtype, small lymphocytic lymphoma (SLL) is characterised by tissue infiltration of clonal B-cells with this phenotype (Swerdlow et al. 2016; Alaggio et al. 2022; Hallek et al. 2018).

CLL derives from a mature B-cell, with weak expression of B-cell markers (surface membrane immunoglobulin, CD19 and CD20). Expression of CD5, CD19, CD23 and kappa or lambda light chains are regarded as essential markers to establish the diagnosis, with CD10, CD79b, CD81, CD200 and ROR1 helpful in distinguishing CLL from other small B-cell neoplasms (Rawstron et al. 2018; Alaggio et al. 2022).

CLL can be subdivided based on the mutational status of the immunoglobulin heavy chain variable region (IGHV) genes encoding the B-cell receptor (BCR)(Hamblin et al. 1999; Damle et al. 1999). CLL with mutated IGHV (m-IGHV) derives from antigen-experienced post germinal centre B-cells (GCB) (the site of somatic hypermutation). CLL with unmutated IGHV (u-IGHV) may arise from pre-GCB or germinal centre independent memory B-cells (Fabbri and Dalla-Favera 2016).

1.1.1 Natural history and clinical management of CLL

CLL follows a heterogeneous clinical course. In high-income countries, up to 80% CLL is diagnosed incidentally (Howell et al. 2013). Asymptomatic patients without bulky disease are managed with active surveillance, or a "watch-and-wait" approach. One third of patients have indolent disease and never fulfil International Workshop on Chronic Lymphocytic Leukaemia (IWCLL) criteria for treatment (Hallek et al. 2018). The remainder progress, following a relapsing, remitting course with progressive therapy-resistance. Clinical guidelines advocate deferral of CLL-directed therapy until the development of cytopenia, bulky lymphadenopathy and/or organomegaly and/or constitutional symptoms (Hallek et al. 2018; Schuh et al. 2018). Despite therapeutic advances exploiting the neoplastic cells dependence on BCR signalling, the disease remains incurable.

Monoclonal B-cell lymphocytosis (MBL)

MBL precedes all CLL, yet only a minority of individuals with MBL develop CLL (Shanafelt et al. 2010). Phenotypic criteria, clonal persistence and absence of features of lymphoproliferative or autoimmune disease define MBL (Rawstron and Hillmen 2010). The threshold of $5x10^9$/L to distinguish high-count MBL from early stage CLL is pragmatic.

MBL is divided into low-count MBL and high-count MBL based on a threshold of 0.5 $x10^9$/L clonal B-cells (Scarfo and Ghia 2016). Low-count MBL requires sensitive flow cytometry to detect and is typically found only in academic research screening populations. High-count MBL can be an incidental finding and is detectable using techniques available in a routine diagnostic laboratory (Rawstron et al. 2008). MBL progresses to CLL at a rate of 1-4% per annum (Shanafelt et al. 2010; Rossi et al. 2009). For the remainder of my thesis, I focus on high-count MBL.

Re-evaluating the role of active surveillance for early asymptomatic disease

Active surveillance, or "watch-and-wait" is recommended for asymptomatic early CLL (Hallek et al. 2018; Schuh et al. 2018). However, data supporting this approach arose prior to targeted therapy. Single-agent chemotherapy and monoclonal antibodies were tested in early asymptomatic CLL. All randomised studies report a period of disease control, but no OS benefit (Dighiero et al. 1998; Hoechstetter et al. 2017; Herling et al. 2020). Each study defined risk differently, resulting in mixed early CLL cohorts and limiting the impact of intervention for the truly high-risk patients.

Intervention in early asymptomatic CLL merits re-evaluation with modern, effective therapies. The CLL12 trial randomised high-risk CLL (defined by German CLL Study Group score) to ibrutinib or placebo (Langerbeins et al. 2022). After four years, 47% of high-risk patients randomised to placebo had not progressed, a rate of progression lower than predicted and underlining the need for more precise tools.

Delivering combinations of targeted therapy to people with low disease bulk might allow long-term MRD-negative remissions and is the subject of ongoing clinical trials (Da Cunha-Bang et al. 2019; Stephens 2021). Early intervention with targeted therapies continues to employ variable high-risk definitions. Various scoring systems which aim to predict progression in newly diagnosed patients have limited precision (detailed in Section 3.1)(Molica et al. 2017). Refined tools incorporating biological and genomic parameters are

needed to define future progressors (Muchtar, Kay, and Parikh 2021) in order to prevent studies of intervention in early asymptomatic CLL treating some patients unnecessarily. The need for a truly predictive marker of progression underpins the rationale behind the Oxford Pre-malignant Lympho-proliferative Disorders (OXPLORED) study, which is discussed in Section 2.2 and Appendix A.

1.1.2 Pathobiology of CLL

Secondary lymphoid organs host the expansion and maturation of B-cells, resulting in antigen specific B-cells and antibody producing plasma cells. Normal B-cell development in the germinal centre of the lymph node occurs via antigen selection and clonal expansion, modulated by signals from T-cells and antigen presenting cells. CLL shares this dependence on neighbouring cells in the node and bone marrow environments, interacting with stromal cells (monocyte derived nurse-like cells and mesenchymal stromal cells) and T-cells as they proliferate. CLL cells co-operate with non-tumour cells in the microenvironment of secondary lymphoid organs to ensure their survival (Burger and Gribben 2014).

Bruton's tyrosine kinase (BTK) plays a central role in intra-cellular signalling via the BCR. Activation of BCR signalling via BTK results in activation of transcription factors required for B-cell proliferation, adhesion, and migration. Pharmacological inhibition of BTK inhibits CLL cell adhesion and migration, triggering the characteristic redistribution of CLL cells from the tissues in the circulation in the early weeks of Bruton's tyrosine kinase inhibitor (BTKi) therapy.

Insights from epigenetic studies report a complex inter-relationship between genetic and epigenetic factors influencing CLL pathobiology. DNA methylation studies describe three epigenetic CLL subgroups (epitypes) based on their similarity to pre- and post-germinal centre B-cells called naïve-like CLL, intermediate CLL and memory-like CLL (Kulis et al. 2012; Klein et al. 2001). CLL with the naïve-like epitype is predominantly u-IGHV whereas memory-like CLL is predominantly m-IGHV (Kulis et al. 2012). CLL with u-IGHV has more active and open chromatin then m-IGHV (Beekman, Chapaprieta et al. 2018). CLL epitypes, based on DNA methylation profiles informs time to first treatment (TTFT) and overall survival (OS) (Beekman et al. 2018; Queiros et al. 2015). These same three epitypes are found in high-count MBL, which was enriched for the high-programmed subtype associated with an indolent clinical course (Giacopelli et al. 2019). MBL epitype predicted who would progress to CLL requiring treatment. Using archived buffy coats, CpG methylation patterns described in CLL were found in healthy subjects' years prior to diagnosis (Georgiadis et al. 2017). A recent study used RNA-seq and gene expression clustering to further subtype CLL. This described segregation of naïve-like CLL into two subtypes and memory-like CLL into four subtypes, which correlate with clinical outcomes (Knisbacher et al. 2022).

1.1.3 Genome features currently used to inform prognosis in CLL

Accessible and reproducible clinical parameters remain the foundation of staging and prognostication. Both the Binet (Binet A-C) and the modified Rai (Stages 0-IV) classifications, describe low, intermediate, and high-risk disease on the basis of clinically palpable lymphadenopathy and presence of cytopenia (Binet et al. 1981; Rai et al. 1975).

Accurate prognostication enables identification of progressors and long-term stable disease. The Cochrane review of prognostic models for early CLL criticised the paucity of external validation studies (Kreuzberger et al. 2020). Of the 12 models included, three had undergone external validation for prediction of OS but none had undergone validation for TFS or TTFT. Two recent prognostic scores optimised for early CLL, the International Prognostic Score for early asymptomatic CLL (IPS-E) and CLL-WONT scores were not included (Condoluci et al. 2020; Brieghel et al. 2022). Both scores predict 8-10% of low-risk patients progress at 5 years, although the IPS-E is simpler to apply.

Copy number alterations detected by FISH

Chromosomal structural abnormalities detectable by fluorescent-in-situ hybridisation (FISH) delineate five CLL prognostic categories (Dohner et al. 2000). Deletion of 17p (the *TP53* locus) predicts a poor chemotherapy response (Zenz et al. 2010). Resistance observed with del17p CLL is attributed to the co-existence of *TP53* mutation on the contralateral allele and loss of p53 function (Dicker et al. 2009; Zenz et al. 2008; Gaidano et al. 1991; Fenaux et al. 1992; el Rouby et al. 1993). Isolated *TP53* mutations (in the absence of del17p) confer inferior outcomes following chemoimmunotherapy (Dicker et al. 2009).

Immunoglobulin heavy chain variable region rearrangement status

During normal B-cell development, the variable (V), diversity (D) and joining (J) segments of the immunoglobulin genes are combined in an ordered manner, generating the primary Immunoglobulin (Ig) repertoire. Ig heavy chain gene (IGH) rearrangement precedes light chain gene rearrangement, D to J precedes the joining of V to DJ. Together these genes can be combined in 2×10^6 distinct rearrangements (Roth 2014). Additional modifications are introduced into the rearranged genes during the recombination process through the addition and removal of nucleotides. If the VDJ recombination is functional, the B-cell enters the germinal centre of the lymph node where it undergoes somatic hypermutation and class switching. Diversity is further increased by somatic hypermutation (SHM), a process of extremely high mutation rate that is at least 10^5 times greater than the normal rate of mutation. The alternate IGH allele is removed from further attempts at recombination.

The degree of SHM (with reference to the germline) of the rearranged IGHV locus subdivides CLL into two subgroups. The subgroups are designated mutated IGHV (m-IGHV) and unmutated IGHV (u-IGHV) conveying favourable and unfavourable prognosis respectively (Hamblin et al. 1999; Damle et al. 1999). The IGHV mutation status emerged as the single most important molecular predictor of TTFT across almost 5000 people with early CLL (Condoluci et al. 2020).

Recent publications using NGS challenged the mutated vs unmutated binary, identifying multiple IGHV subclones in one quarter of CLL patients (Fuentes et al. 2017), a finding later confirmed in an independent cohort (Stamatopoulos et al. 2017). The number of CLL subclones correlates with shorter TTFT (Puente et al. 2011; Stamatopoulos et al. 2017;

Nadeu et al. 2018). Subclones were defined by the distribution of driver mutations or the number of productive IGHV rearrangements.

CLL has restricted immunoglobulin gene repertoires and displays biased and stereotyped patterns. Approximately one third of CLL patients carry BCRs with structurally similar Complementarity-Determining Regions (CDRs), or stereotyped immunoglobulin genes (Agathangelidis et al. 2012). Just 19 immunoglobulin gene subsets account for 10% of all CLL (Vardi et al. 2014). The detection of restricted CDR3 sequences is accepted as evidence for antigen selection of the neoplastic B-cell clone although the precise antigen remains elusive. Stereotypy identifies CLL subgroups with shared clinical features. For example, subset#2 delineates a m-IGHV CLL with short TTFT whereas subset#8 is associated with trisomy 12 and risk of high-grade transformation (Jaramillo et al. 2020; Agathangelidis et al. 2021).

Somatic mutations identified by NGS in CLL

Landmark studies characterising the genomic landscape of CLL identified recurrent mutations of genes impacting key pathways in CLL biology including RNA and ribosomal processing (*SF3B1*), DNA damage and cell cycle control *(ATM* and *TP53*), *NOTCH* signalling (*NOTCH1),* inflammatory pathways (*BIRC3* and *MYD88*) and RAS pathways (*KRAS*) (Landau et al. 2013; Puente et al. 2015; Puente et al. 2011; Quesada et al. 2011; Rossi et al. 2012; Wang et al. 2011). NGS, whole exome sequencing (WES) and whole genome sequencing (WGS) identified recurrent mutations in *NOTCH1, XPO1, SF3B1, RSP15* and *SAMHD1* predicting for shorter TTFT, poor responses after chemoimmunotherapy and shorter OS (Puente et al. 2011; Quesada, Ramsay, and Lopez-Otin 2012; Clifford et al. 2014; Knight et al. 2012).

Of these, only *TP53* is mandated prior to treatment (Hallek et al. 2018). Guidelines require examination of *TP53* exons 4-10 at minimum, reporting variants at > 10% by SS and > 5% by NGS (Malcikova et al. 2018). The use of targeted gene panels is transitioning from research into routine clinical care, prompting the European Research Initiative on CLL (ERIC) to assess the performance and variability of gene panels in a multi-centre study (Sutton et al. 2021).

Complex karyotype and Genomic Complexity

Complex karyotype (CK) and/or genomic complexity (GC) is an emerging adverse prognostic marker. CK confers shorter TTFT, shorter progression free survival (PFS) following treatment and higher risk of Richter's Transformation (RT) (Visentin et al. 2022). CK is incorporated into combination targeted therapy trials for high-risk disease (Huber et al. 2022).

CK overlaps with other adverse prognostic biomarkers including u-IGHV, *TP53* aberrations and telomerase dysfunction. Deletions and mutations in *TP53* promote genomic instability and enable the neoplastic cell to tolerate the metabolic effects of aneuploidy (Kastenhuber and Lowe 2017). However, up to 20% of CLL with CK has no detectable *TP53* aberration (Baliakas et al. 2019).

The definition of GC is evolving. Conventionally, CK was defined as ≥ 3 or more structural or numerical abnormalities detected by Chromosomal Banding Analysis (CBA). This has been refined in low-CK, intermediate-CK and high-CK based on 3, 4, or ≥ 5 abnormalities (Baliakas et al. 2019). This definition limited CK to structural changes at 10-20Mb size, the resolution of CBA.

A more recent technique, chromosomal microarray analysis, based on Comparative Genomic Hybridization (CGH) or Single Nucleotide Polymorphisms (SNPs) allows for considerably smaller structural changes to be detected and facilitates the detection of multiple changes in a single experiment. Importantly, cell culture is not required. High GC defined as ≥ 5 CNA with a size cutoff of ≥ 5Mb detected by arrays delineated CLL with shorter TTFT and OS (Leeksma et al. 2021). Lowering the size cutoff to ≥1Mb did not alter the relationship observed between GC and clinical outcomes.

There are no universally accepted criteria for GC by WES or WGS. Focal CNAs measuring 10Mb were included in assessment for GC using International Cancer Genome Consortium (ICGC) WES datasets (Mosquera Orgueira et al. 2019). A length of 100kb defined Copy Number Alterations (CNA) in the Genomics England WGS CLL pilot (Robbe et al. 2022).

Genomic features influence MBL progression to CLL.

Progression from MBL to CLL is preceded by acquisition of mutations and subclonal expansion (Barrio et al. 2017; Agathangelidis, Ljungstrom, et al. 2018; Ojha et al. 2014). Targeted NGS performed on paired MBL samples identified mutations in *NOTCH1, BIRC3* and *SF3B1* associated with shortened TTFT. *TP53* and *XPO1* mutations were undetected, implying these are secondary events (Barrio et al. 2017). Longitudinal analysis of MBL by WES reported that evolution of mutational profiles anticipated clinical progression (Ojha et al. 2014). Two WES studies describe similar profiles between MBL and non-progressive CLL but restrict analysis to variants at allele frequencies >20% and >10% respectively (Ojha et al. 2014; Agathangelidis, Ljungstrom, et al. 2018).

Thus, molecular heterogeneity in CLL is present from the earliest stage of disease. Detection of adverse biomarkers in MBL and early CLL may offer a tool for personalised outcome prediction and be a prerequisite for early intervention studies.

1.1.4. Challenges to delivery of precision diagnostics for CLL

Clinical guidelines for CLL recommend testing for *TP53* dysfunction (*TP53* deletion and/or mutation), IGHV status to correctly assign prognosis using the CLL International Prognostic Index (CLL-IPI) (Hallek et al. 2018; International 2016). Assessment of both factors requires a diagnostic service equipped for both FISH and sequencing.

However, these techniques are labour intensive, have limited sensitivity and fail to detect emerging predictive markers, including GC. Conventional karyotyping and CBA relies on cells in culture generating sufficient metaphases for analysis and stimulation of CLL cells with B-cell mitogens to yield sufficient metaphases is required (Baliakas et al. 2022). Other drawbacks of CBA include the need for invasive bone marrow aspirate (BMA) and the availability of appropriate laboratory facilities and skilled cytogeneticists. CBA analysis is labour intensive and thus not readily scalable.

Real-world registry data indicated that testing is not universally applied (Mato et al. 2020). Low-and middle-income countries do not have routine access to precision diagnostics (Wilson et al. 2018; Gopal et al. 2012). Taken together, these factors mean many CLL patients are not benefiting from precision diagnostics and alternative laboratory techniques merit exploration.

1.2 High-grade transformation of CLL (Richter's transformation).

Transformation to high-grade malignancy characterised by therapy resistance and rapid progression is observed across cancers (Costa, Altemani, and Hermsen 2011; Friedmann-Morvinski and Verma 2014). Richter's Transformation (RT, formerly Richter's syndrome) is an aggressive, chemotherapy resistant high-grade B-cell lymphoma arising in CLL (Alaggio et al. 2022). Most RT shares morphological features with activated B-cell type diffuse large B-cell lymphoma (DLBCL) (84–95%), with a minority developing Hodgkin-like lymphoma or prolymphocytic transformation (Swerdlow et al. 2016; Soilleux et al. 2016; Lenartova et al. 2019; Alaggio et al. 2022).

Although RT shares histological features with DLBCL, its molecular profile is distinct (Parry et al. 2023b; Nadeu et al. 2022). Academic studies report 90% of RT shares molecular lesions with the founding CLL-related clone, with mutations in *TP53* (60% to 80%), *CDKN2A* (30%), *MYC* (30%), *NOTCH1* (30%), or *MGA* (7%) (Tsimberidou and Keating 2005; Fabbri et al. 2013; Chigrinova et al. 2013; Rossi, Spina, et al. 2011). Whole genome studies revealed RT mutational signatures associated with prior therapies and complex chromosomal changes including whole genome duplication and chromothripsis (Parry et al. 2023b; Nadeu et al. 2022). Intriguingly, analysis of circulating neoplastic lymphocytes (PB-CLL) and plasma cell-free DNA samples taken years prior to RT diagnosis identified small populations carrying molecular features of the subsequent tumour suggesting that it may be possible to identify in advance CLL patients at risk of RT.

Despite progress in CLL, the prognosis remains dismal for RT, with median OS of 8-12 months (Eyre et al. 2016; Tsimberidou et al. 2013; Rogers et al. 2018). One multi-institutional series including RT patient's ineligible for trials reported over half of RT patients died within three months of diagnosis (Parry et al. 2023b). Norwegian population data reports a median survival of 27 months (Lenartova et al. 2019). Conversely, *de novo* RT (without prior CLL treatment) where the neoplastic RT clone typically shows a distinct IGHV rearrangement to the co-existent CLL clone has a more favourable outlook, with a median OS of 6.1-6.4 years (Ben-Dali et al. 2020; Lenartova et al. 2019).

RT therapy mirrors that for DLCBL but outcomes are unsatisfactory (Rogers et al. 2018; Younes, Brody, et al. 2019; Tsimberidou et al. 2008). Addition of the monoclonal antibody ofatumumab, to cyclophosphamide, doxorubicin, vincristine and prednisolone (O-CHOP) chemotherapy did not improve outcomes in our groups' Phase II study of RT (Eyre et al. 2016). Small cohort studies demonstrated activity for BTKi monotherapy in RT (Tsang et al. 2015; Hillmen et al. 2016) and more recently, in combination with nivolumab (Younes, Brody, et al. 2019; Jain, Ferrajoli, et al. 2018; Bruscaggin et al. 2021). The combination of acalabrutinib with R-CHOP in the STELLAR trial, described in Section 2.3 is designed to obtain rapid disease control in this clinically aggressive lymphoma.

Mutations and CNAs found in the RT tumour have been detected in plasma cell-free DNA. In individual patients, the ctDNA changes anticipated the histological diagnosis of RT (Parry et al. 2023b; Yeh, Hunter, et al. 2017). Evaluation of tumour and cell-free DNA is incorporated into the analysis plan to the STELLAR trial.

1.3 Introduction to Diffuse Large B-cell lymphoma

1.3.1 Clinical and biological heterogeneity in DLBCL

DLBCL is the commonest adult high-grade lymphoma. Approximately 60% of DLBCL is cured using standard chemoimmunotherapy with rituximab, cyclophosphamide, doxorubicin, vincristine and prednisolone (R-CHOP). The remainder develop progressive or recurrent disease, with poor outcomes. Early identification of patients who benefit from additional therapies is challenging, in part due to the heterogeneity of the disease.

Morphological and clinical features subtype DLBCL into subgroups with distinct clinical courses (Swerdlow et al. 2016). Gene expression profiling (GEP) subdivides into activated B-cell (ABC), germinal-centre B-cell like (GCB) and unclassified subgroups according to the proposed cell of origin (COO). These COO subgroups are associated with differing clinical behaviours and treatment responses (Alizadeh et al. 2000; Chapuy et al. 2018; Schmitz et al. 2018). GEP is not routinely available for clinical diagnostics and immunohistochemical algorithms are used as a surrogate for the GEP subgroups (Hans et al. 2004; Choi et al. 2009) However there remain concerns about limited concordance with GEP, reproducibility and the ability of immunohistochemically defined subgroups to predict outcome (Ott et al. 2010; Gutierrez-Garcia et al. 2011; Gleeson et al. 2015).

Alterations at specific loci delineates DLBCL subtypes with adverse clinical outcomes. The *MYC* gene, via the transcription factor MYC plays a key role in apoptosis, proliferation, differentiation and metabolism (Ott, Rosenwald, and Campo 2013). Overexpression of MYC protein, additional copies of *MYC* and rearrangements of *MYC* are associated with inferior OS (Wight et al. 2018; Barrans et al. 2010). Morphological and immunohistochemical testing are insufficient to capture all cases, prompting the recommendation for FISH testing for *MYC*, *BCL2* and *BCL6* rearrangements for all DLCBL (Tilly et al. 2015; Chaganti et al. 2016).

Integration of GEP with CNA and targeted sequencing data delineates four molecular subtypes of DLCBL; MCD (based on co-occurrence of *MYD88*^{L265P} and *CD79B* mutations), BN2 (*BCL6* fusions and *NOTCH2* mutations), N1 (*NOTCH1* mutations) and EZB (*BCL2* translocations and *EZH2* mutations) (Schmitz et al. 2018). BN2 and EZB subtypes experience favourable outcomes following R-CHOP (Schmitz et al. 2018). MCD and BN2 subtypes are dependent on chronic active B-cell receptor signalling, offering a target for therapeutic inhibition.

Whole Exome Sequencing offers further insights into the complexity of DLBCL. Somatic mutations, structural variants (SV) and mutational signatures (notably AID-driven signatures) clustered DLBCL into five subgroups (Chapuy et al. 2018). A subgroup with chronic active B-cell receptor signalling was again identified, along with a new cohort characterised by *TP53* disruption and genomic instability (Chapuy et al. 2018).

Deletions and mutations in *TP53* are an emerging biomarker for DLBCL. *TP53* mutations affecting the DNA binding domains were associated with inferior PFS and OS following R-CHOP. *TP53* mutations confer adverse prognosis, irrespective of GCB or non-GCB subtype (Xu-Monette et al. 2012). Including *TP53* variants detected in ctDNA (and thus all sites of disease), *TP53* mutations were identified in one quarter of relapsed DLBCL patients, persisted during treatment and conferred a shorter PFS (Bruscaggin et al. 2021).

1.3.2 Challenges to the application of precision therapy for DLCBL

Many well-designed trials have attempted to improve R-CHOP outcomes through additional drugs (Nowakowski, Chiappella, et al. 2021; Nowakowski, Hong, et al. 2021), dose intensification (Howlett et al. 2015) and post treatment consolidation (Davies et al. 2020). Intensification of DLBCL therapy using da-R-EPOCH for MYC-translocated DLCBL improved ORR but not OS (Howlett et al. 2015). The POLARIX randomised trial reported ORR and PFS favouring polatuzumab vedotin with R-CHP over R-CHOP. No OS difference has been observed at 28 months, possibly due to immature follow-up and higher proportion of patients progressing after R-CHOP receiving salvage cellular therapies (Tilly et al. 2022).

Attempts to individualise DLBCL therapy based on GEP have been disappointing (Davies et al. 2019; Nowakowski, Chiappella, et al. 2021). Initial results of the REMoDL-B randomised trial which added bortezomib based on real-time COO classification reported no difference in outcomes at 30 months follow-up. Retrospective reanalysis of the GEP data to distinguish molecular high-grade DLBCL (MHG) (Sha et al. 2019) and longer follow-up reported improved PFS in the bortezomib-treated ABC and MHG subgroups. However, 26% (276/1077) biopsies were unable to be classified (Davies et al. 2023). The Phoenix trial targeted BCR signalling by adding ibrutinib to R-CHOP in non-GCB DLBCL, reported improved outcomes in younger (< 60 years) patients but increased adverse events (AE), worsened PFS and OS in older patients (Younes, Sehn, et al. 2019).

Until recently, identification of diagnostic, prognostic and predictive biomarkers in DLBCL relied on immunohistochemistry and/or FISH. NGS technologies have expanded our understanding of DLBCL pathobiology. However, the translation of NGS into clinical diagnostics is challenged by the emergent presentation of DLBCL, access to fresh tissue biopsies, availability and standardisation of NGS in diagnostics laboratories and conflicting reports on how to implement NGS results into clinical practice (Vermaat et al. 2015).

Inability to characterise tumour heterogeneity remains a barrier to delivery of precision medicine for DLBCL. Molecular tools to stratify DLBCL patients are needed, delivering genomic information within the diagnostic pathways. The availability of equipment, the skilled personnel and the long turnaround times (TAT) involved in defining molecular subgroups inhibit their use in clinical decision making for DLBCL treatments. Quantification of baseline ctDNA may have a role in refining pre-treatment prognosis and dynamic changes in ctDNA burden have been proposed to delineate DLBCL patients with favourable and unfavourable outcomes (Kurtz et al. 2019).

PET-adapted treatment, which is established in Hodgkin lymphoma (HL) adjusts chemotherapy based on response after two cycles of treatment (Cheson et al. 2014; Trotman and Barrington 2021). In DLBCL, Complete Metabolic Response (CMR) by PET-CT after R-CHOP may select patients who can safely omit consolidation radiotherapy (Freeman et al. 2021). Early interim PET-CT identifies poor responders (Eertink et al. 2021). However, escalation of treatment based on an interim positive PET-CT has not improved clinical outcomes for DLBCL (Duhrsen et al. 2018).

An alternative strategy is to tailor treatment depending on early response, potentially escalating treatment in patients with early indications of poor response and de-escalating in patients with good responses. Learning from 'win-probability' models, Kurtz et al developed a framework for dynamic prognostication, the Continuous Individualised Risk Index (CIRI) for DLCBL combining baseline data (International Prognostic Index (IPI), COO and ctDNA burden) with on-treatment assessments (ctDNA quantity after 2 and 3 cycles)(Kurtz et al. 2019).

1.4. Circulating tumour DNA (ctDNA): key concepts and challenges

The identification of foetal DNA in maternal plasma prompted the use of cell-free DNA for the minimally invasive diagnosis of foetal monogenic disorders (Saito et al. 2000) and aneuploidy (Lo et al. 2007). Recognition that cell-free DNA from cancer patients contained tumour-derived fragments and development of NGS sequencing methods capable of detecting variants occurring at low allele frequencies encouraged interest in the clinical application of ctDNA for minimally invasive diagnosis, molecular profiling and response assessment (Wan et al. 2017; Corcoran and Chabner 2018; Schwarzenbach, Hoon, and Pantel 2011).

Plasma cell-free DNA circulates as low concentrations of short double-stranded DNA fragments (Fleischhacker and Schmidt 2007). The cell-free DNA fragment size corresponds to nucleosomes (~147 bp) and chromatosomes (nucleosome + linker histone; ~167 bp). Subtle differences in the fragment size have been observed between foetal and maternal cell-free DNA and between tumour-derived ctDNA and non-tumour cell-free DNA (Mouliere et al. 2011; Snyder et al. 2016; Meriranta et al. 2022), prompting the exploration of cell-free DNA analysis in cancer screening (Cohen et al. 2018; Jamshidi et al. 2022)and non-invasive prenatal diagnosis (Lo et al. 2007; Saito et al. 2000; Jiang et al. 2021). Circulating cell-free DNA retains the methylation profile of the cell of origin (Lo et al. 2021; Chabon et al. 2020).

1.4.1 Pre-analytic variables impacting on cell-free DNA and ctDNA

Quantification of the plasma cell-free DNA can be performed with fluorescent, electrophoretic methods or quantitative PCR. A range of metrics are used to quantify the ctDNA fraction.

The detection limit of many technologies depends on the number of cell-free DNA molecules interrogated. In lymphoid malignancies, the cell-free DNA concentration is variable, related to tumour stage with a mean concentration of 30ng/ml of plasma (Kurtz et al. 2018; Roschewski et al. 2015; Armand et al. 2013). A minimum of 10ml whole blood (4-6 ml plasma) is recommended to obtained sufficient cell-free DNA molecules for sequencing (Rossi et al. 2020; Lauer, Mutter, and Scherer 2022).

The total cell-free DNA obtained from plasma is influenced by the collection tube and storage parameters. Leakage of cellular gDNA from non-tumour leukocytes during the interval between phlebotomy and plasma processing may overwhelm the capacity to detect ctDNA variants. Contamination of the plasma cell-free DNA is problematic for blood collected in EDTA tubes, prompting the recommendation to process EDTA tubes within six hours of phlebotomy (Rossi et al. 2020). Specialised analysis tubes containing leukocyte stabilising preservatives such as Streck®, Roche® or PAX-gene® have been developed, with reported stability for 7-14 days. However, the consensus recommendation is processing within 48 hours (Rossi et al. 2020).

Figure 1.4.1.1 Plasma cell-free DNA in cancer

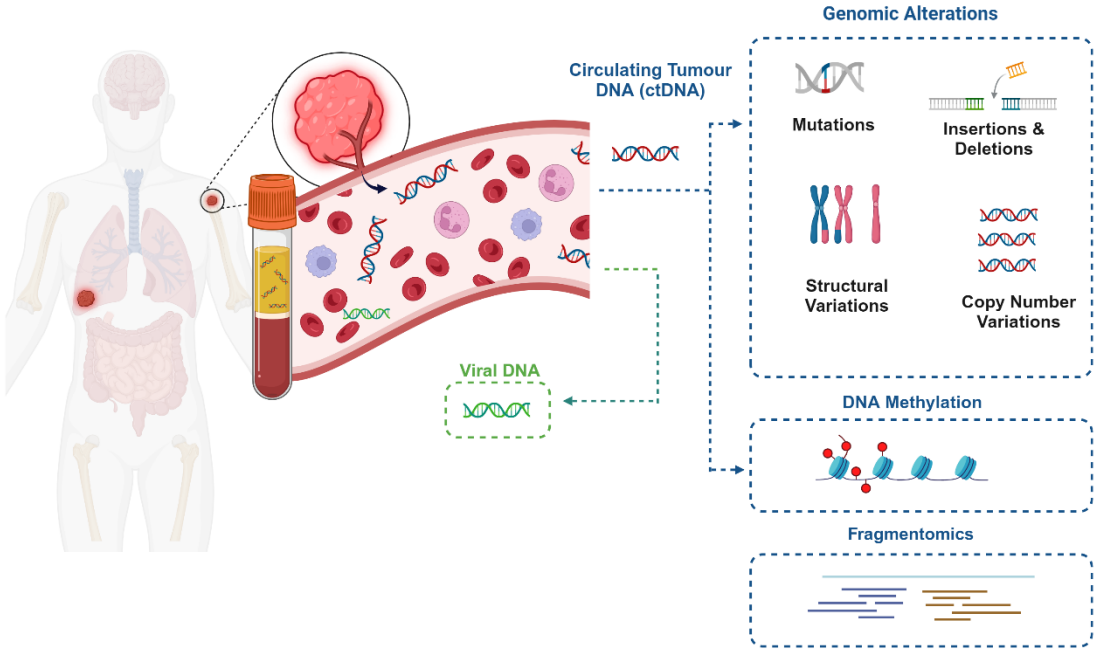


Figure 1.4.1.1: Tumour-related cell-free DNA is released from tumour cells via apoptosis, necrosis and secretion. which circulates in the blood of cancer patients, is released by tumour cells in different forms and at different levels. DNA can be shed as both single-stranded and double-stranded DNA. Somatic mutations, copy number changes, changes in methylation and fragmentation patterns can be detected from circulating tumour DNA in plasma. For selected cancer subtypes, fragments of viral DNA can be found in plasma. Figure made in BioRender.com

1.4.2 Biological confounding factors for circulating tumour DNA

The accurate detection of tumour-derived variants in ctDNA is hampered by the presence of somatic variants outside the tumour of interest. This is particularly challenging when the tumour mutational profile is unknown or when ctDNA is studied for cancer screening.

Somatic mutations detectable in myeloid cells, or clonal haematopoiesis of indeterminate potential (CHIP) are detectable in 10% of adults over 65 years old and confer a 11-12 fold increased risk of haematological cancer (Genovese et al. 2014; Jaiswal et al. 2014). CNA and somatic mutations are detected in pre-malignant tissues including Barrett's oesophagus, (Killcoyne et al. 2020) endometrial hyperplasia (Kiechle et al. 2000) and colonic polyps (Juarez et al. 2017). Paired germline DNA (peripheral blood or salivary DNA) samples are recommended to prevent misclassification of rare polymorphisms detected in the cell-free DNA as somatic variants (Rossi et al. 2020). Additional samples and sequencing costs are a barrier to translation of ctDNA technology to clinical use.

1.4.3 PCR-based methods for ctDNA analysis

Broadly, methods for ctDNA analysis can be divided into PCR-based and NGS-based approaches. PCR-based approaches include allele-specific oligonucleotide PCR (ASO-PCR) and digital droplet PCR (ddPCR).

Both ASO-PCR and ddPCR assays require primers specific to a pre-defined mutation, restricting their application to single mutation detection. Most studies identify the specific target in the tissue biopsy before examining cell-free DNA. Scherer et al calculates that a maximum of ~20,000 haploid genome equivalents (HGE) can be obtained at venepuncture from a single plasma sample, setting a hypothetical limit of detection for single gene ctDNA tests at 1 in 20,000 or 0.005% (Scherer et al. 2017).

Droplet digital PCR facilitates the absolute quantification of target nucleic acids. The sample is partitioned or sub-divided into many independent sub-reactions. Each sub-reaction contains either a few target sequences in relative concentration (due to the small volume) or no target sequences. Sub-dividing the sample into small reactions isolates the target mutant sequence in its own independent micro reaction and reduces the template competition from wild-type sequences, enabling detection of rare mutations. After ddPCR, the proportion of amplification positive partitions is used to quantify the concentration of the target sequence in the original sample using probability statistics.

Both allele-specific PCR and ddPCR have found clinical applications in MRD monitoring for lymphomas with recurrent translocations, or specific, conserved single nucleotide variants (SNV), such as $EZH2^{Y641N}$ in follicular lymphoma or $MYD88^{L265P}$ in primary CNS lymphoma (Watanabe et al. 2019; Schmitz et al. 2018; Chapuy et al. 2018; Lakhotia et al. 2022). Droplet digital PCR has established a role in detecting emerging treatment resistance $EGFR^{T790M}$ mutations in non-small cell lung cancer (Spence et al. 2021; Aggarwal et al. 2021) and is being explored for *BTK* and *PLYG2* mutations associated with BTKi resistance (Bodor et al. 2021).

1.4.4 Targeted next generation sequencing (NGS)

While PCR and ddPCR are limited to specific mutation detection, targeted sequencing permits examination of multiple genomic loci simultaneously. Amplicon-based

methodologies target up to hundreds of amplicons with high sensitivity. Hybrid capture-based assays expand the genomic region studied (Samorodnitsky et al. 2015). Both rely on prior knowledge of the molecular features of the disease to select the appropriate panel targets however patient-specific assays are not required. NGS based methods, with UMI and error correction bioinformatics can sequence cell-free DNA fragments to depths of 40,000 reads per base (Cohen et al. 2018). Methods that incorporate suppression of background noise improve sensitivity (Newman et al. 2016a).

1.4.5 Limitations of targeted next generation sequencing for ctDNA

The power of cell-free DNA analysis to detect variants from small volume tumours (in the context of cancer screening or after treatment) is limited. Asymptomatic people have lower total cell-free DNA quantities available for sequencing, typically 10ng/µl. In one study of 117 healthy women, just 47% (55/117) yielded sufficient quality cell-free DNA to proceed to library preparation (Alborelli et al. 2019). Lower input cell-free DNA levels are inversely correlated with the limit of detection and inhibits the capacity to detect ctDNA variants (Alborelli et al. 2019). A minimum of 32ng input DNA is recommended for capture hybridisation library preparation using CAPP-Seq (Rossi et al. 2019). Additional plasma volumes at phlebotomy may compensate for low quantities and this would need to be incorporated into clinical studies of cell-free DNA for early detection.

The proportion of cell-free DNA arising from the tumour, or ctDNA is proportional to the tumour bulk. However, when the proportion of tumour fragments in cell-free DNA is sufficient to allow detection of variants on a targeted panel, the tumour is also detectable radiologically (Abbosh et al. 2017; Phallen et al. 2019; Cohen et al. 2018). In this context, targeted ultra-deep sequencing of cell-free DNA offers little advantage over radiological methods for cancer screening or monitoring of MRD.

Since ctDNA requires deep sequencing, smaller targeted panels strike a balance between the cost of sequencing and the likelihood of detecting a genuine variant. Disadvantages of small panels include failure to detect variants which are outside the target range and limited capacity to call CNA. In addition, if the mutated DNA fragment carrying the target of interest is rare in the circulating plasma and/or is not sampled at phlebotomy, it will not be detectable by hypothetical error-free sequencing. These limitations prompt the exploration of whole genome approaches for ctDNA analysis (Figure 1.4.5.1).

1.4.6 Whole genome approaches for ctDNA

Chromosomal structural abnormalities are detectable by low-coverage WGS and comparing the relative read number in the genomic region of interest to controls, or to across samples or with regions of equal size from the same sample. For cell-free DNA shallow WGS, each DNA fragment contributes to the dataset. Tumour derived CNA have been detected in cell-free DNA for neuroblastoma (Van Roy et al. 2017), lung carcinoma (Raman et al. 2020), and non-Hodgkin lymphoma (Bouzidi et al. 2021; Yeh, Hunter, et al. 2017). Using 35x depth WGS to identify CNA in cell-free DNA in combination with the tumour profile, Zviran et al developed MRDetect, a bioinformatics package to identify tumour associated profiles in cell-free DNA for monitoring of MRD down to 0.001% (Zviran et al. 2020). Shallow WGS has been used to identify RT-specific CNA in the cell-free DNA (Parry et al. 2023a; Yeh, Hunter, et al. 2017).

Detecting tumour associated CNA using plasma ctDNA is particularly useful when tumours are inaccessible and in tumours characterised by aneuploidy or GC. Shallow WGS is a time and cost-effective approach, making it attractive for incorporation into clinical trials.

Another alternative approach to ctDNA sequencing is based on the fragmentation patterns of the cell-free DNA. When released from apoptotic cells, nucleosomes, transcription factors and other DNA-binding proteins produce specific patterns of fragmentation in the cell-free DNA. The fragmentation pattern can be linked to the cell of origin (Snyder et al. 2016) and distinguishes CHIP cell-free DNA fragments from lung carcinoma fragments (Chabon et al. 2020). Cell-free DNA fragment size distribution identified shorter fragments carrying somatic mutations and the cell-free DNA fragment profile anticipated treatment failure in high-risk DLCBL (Meriranta et al. 2022). More recently, machine learning approaches identified mutational signatures including aging, smoking, APOBEC and MSI signatures in very shallow WGS (0.3-1.5x) cell-free DNA datasets from healthy controls and solid tumours (Wan et al. 2022).

Figure 1.4.5.1: Schematic showing the differences between a targeted panel-based approach and shallow WGS

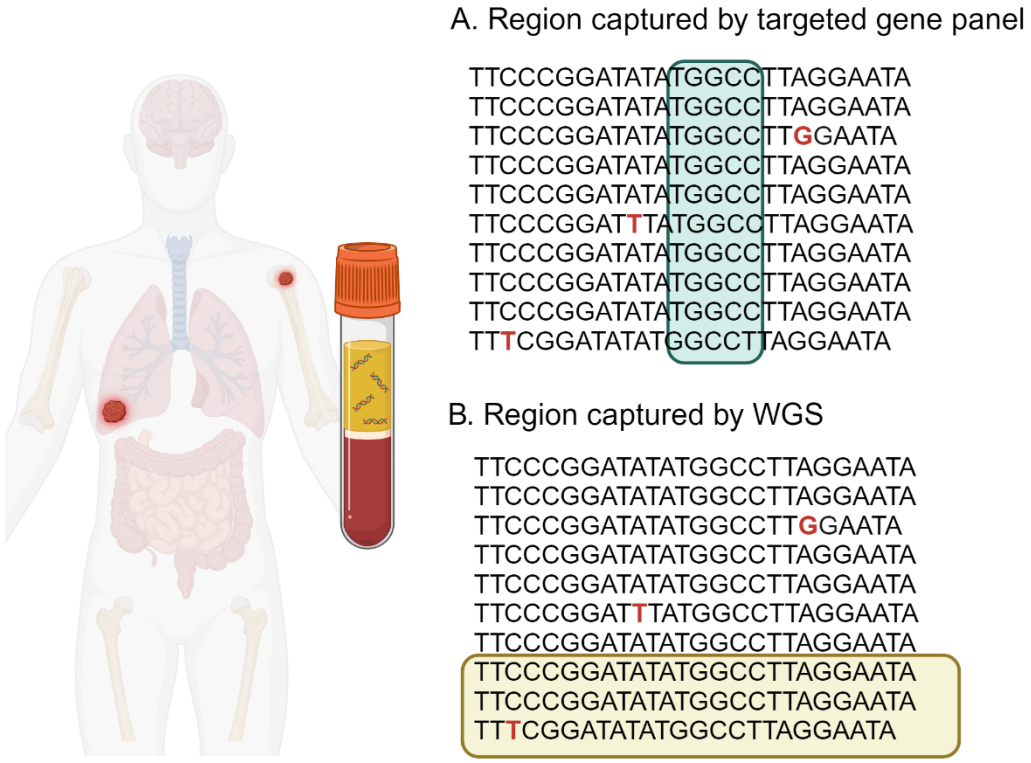


Figure 1.4.5.1 Illustration showing mutations captured by targeted panel deep sequencing are only a small fraction of tumour mutations. Blood samples collect a small sample of cfDNA, resulting in higher probability for mutation capture by WGS than by targeted panels. Figure made in BioRender.com

1.4.7 Bioinformatics considerations for ctDNA analysis

Tumour-derived DNA fragments carry variants at low abundance against a background of normal (non-clonal leukocyte derived) cell-free DNA. Tumour-derived DNA fragments cannot be separated from non-clonal cell-free DNA, obliging bulk sequencing of the entire circulating cell-free DNA pool.

The quantity of ctDNA overlaps with the limit of detection on sequencing-by-synthesis (SBS) platforms. Polymerase errors and sequencing errors, and artefactual changes introduced by software can produce high levels of noise and false positives. In addition, detection of gene fusion and translocations is difficult using short cell-free DNA fragments. Tumour aneuploidy results in subtle differences in the ctDNA fraction, making CNA alterations challenging.

Specialist software tools have been developed to analyse ctDNA. As cell-free DNA reads are short, 2x150bp paired-end sequencing will produce overlapping read pairs. There exists a range of alignment tools to map DNA sequencing reads to reference genome, including bowtie2 and Burrows-Wheeler Aligner (BWA)(Li and Durbin 2009). The BWA software offers three software packages. BWA-MEM is generally recommended for high-quality queries, as it is faster and more accurate. When working with cell-free DNA fragments, the alignment should be initiated from the centre of the DNA read. Repetitive or homologous sequences are vulnerable to misalignment with short cell-free DNA fragments.

Variant callers for identifying polymorphisms in germline datasets are unsuitable for cell-free DNA data. Platforms for somatic mutations capable of detecting low frequency variants are preferred. The choice of software tool requires balance between detection of low-frequency variants and reporting an abundance of false positives. The authors of one review advise VarScan2 combined with strict variant filtering (Chen, Liu, and Zhou 2018). The variant calling file is then annotated using tools like Annovar to compare sequence data with databases like dbSNP, ClinVar and COSMIC (Wang, Li, and Hakonarson 2010).

1.5 Clinical applications of ctDNA in lymphoid malignancy

Circulating tumour DNA offers a tool to address unmet needs in the diagnosis, genotyping and post treatment monitoring of lymphomas. Unlike BCR-ABL transcripts quantification for Chronic Myeloid Leukaemia, no single mutation is pathognomonic of B-cell lymphomas, making single mutation detection tools unreliable for MRD monitoring. Tumours display heterogeneity across sites of disease (Araf et al. 2018; Del Giudice et al. 2016) and the disease course (Schuh et al. 2012; Landau et al. 2015). As NGS-based ctDNA detects multiple tumour-specific mutations, it can potentially provide MRD-monitoring based on personalised tumour profiles.

1.5.1 Assessment of disease burden, staging and prognostic scores

Accurate measurements of disease bulk inform clinical prognosis, response assessment and treatment decisions. Quantification of ctDNA burden (measured as human genome equivalents, HGE) correlates with metabolic tumour volume by PET-CT in primary CNS lymphoma (Fontanilles et al. 2017), classical HL (Spina et al. 2018), follicular lymphoma (Delfau-Larue et al. 2018), mantle cell lymphoma (Agarwal et al. 2019) and DLCBL (Scherer et al. 2016; Kurtz et al. 2018; Meriranta et al. 2022). Pre-treatment ctDNA burden correlates with International Prognostic Index (IPI) in DLBCL (Scherer et al. 2016; Meriranta et al. 2022), M-IPI in mantle cell lymphoma (Lakhotia et al. 2022) and the German Hodgkin Study Group score in HL (Spina et al. 2018). Significantly, pre-treatment ctDNA burden predicts clinical response to chemoimmunotherapy in DLBCL and HL (Scherer et al. 2016; Kurtz et al. 2018; Meriranta et al. 2022; Spina et al. 2018).

1.5.2 Minimally invasive molecular profiling of lymphoma

Integrated haematopathological assessment of lymphoma tissue acquired from invasive surgical biopsy remains the gold standard for diagnosis and molecular subclassification. The trend towards core biopsies means less material is available for molecular studies. Artefactual changes induced by FFPE hampers NGS (Do and Dobrovic 2015).

The time taken to deliver molecular results on diagnostic tissue samples hinders the rollout of trials based on molecular subgroups of lymphoma. In the DIRECT study for high-grade B-cell lymphoma, the median time taken to receive biopsy material from external trial sites was 67 days (Dan Hodson, oral presentation 2022). Patients with high-grade lymphoma cannot wait on molecular subclassification to start treatment. By contrast, most plasma samples were received within 3 days of phlebotomy. Plasma ctDNA profiling, rather than tissue-based analysis might be preferable to delivering molecularly targeted clinical trials in a timely fashion.

Concordance between tumour and ctDNA derived molecular profiles is anticipated to be imperfect, due to intra-tumour heterogeneity. Numerous studies using panel based NGS show that concordance between the tissue biopsy and ctDNA genotype is 70-80% in HL (Spina et al. 2018) and DLBCL (Rossi et al. 2017).

Mutations detected in the ctDNA but absent from the tumour tissue are frequently encountered, underlining that biopsy of a single site does not capture the complexity of the

entire tumour. In one illustrative example, *CARD11* and *PIM1* mutations were detected in plasma ctDNA alone at diagnosis. However, at the time of histological transformation, the *CARD11* and *PIM1* mutations were seen in both plasma ctDNA and the new retroperitoneal mass (Scherer et al. 2016).

1.5.3 Minimally invasive assessment of response

Responses to chemoimmunotherapy are variable, even within B-cell lymphomas of similar clinical stage. Identification of non-responders at an early timepoint would improve personalisation of lymphoma treatment strategies. Response adapted treatment, based on interim PET-CT results has become standardised in HL (Trotman and Barrington 2021), sparing responders additional bleomycin and facilitating early planning of consolidation and salvage strategies for non-responders.

MRD assessment by multi-parameter flow cytometry or PCR-based methodologies is established in the management of leukaemia but generally requires BM biopsies. Circulating tumour DNA expands the role of serial MRD assessment to nodal malignancies. In an early study, eradication of the IGH clonotype from ctDNA after two cycles was associated with favourable PFS in DLCBL (Roschewski et al. 2015). Using the CAPP-Seq platform, changes in the ctDNA quantity after two chemotherapy cycles predicts PFS in DLCBL and HL (Kurtz et al. 2018; Spina et al. 2018).

Circulating tumour DNA has been compared with flow-cytometric MRD assessment in CLL and myelodysplastic syndrome (MDS) (Yeh, Dickinson, et al. 2017; Yeh, Hunter, et al. 2017; Furstenau et al. 2022). In MDS, all MRD positive patients by flow-cytometry had detectable ctDNA but ctDNA detected persistent somatic mutations arising from disease in two additional patients with negative flow-cytometry results (Yeh, Hunter, et al. 2017). In CLL, ddPCR ctDNA detected MRD in 29% reported as undetectable MRD by flow-cytometry, suggesting persistence of nodal disease (Furstenau et al. 2022).

1.5.4 Early identification of relapse, transformation and therapy resistance

A range of modalities monitor treatment response in lymphoma including invasive biopsies and radiological examinations. Bone marrow biopsies may be painful, while imaging necessitates radiation exposure. Residual masses on imaging require re-biopsy to determine histology.

Several authors report emerging ctDNA mutations forecast treatment resistance or transformation. Deep targeted sequencing of ctDNA revealed emerging treatment resistance clones and anticipates DLBCL relapse. BTK drug-binding site mutations were detectable prior to relapse (Scherer et al. 2016). In CLL, cell-free DNA may be more reliable than cellular DNA at detecting BTKi resistance mutations (Albitar et al. 2017), although a more recent study failed to replicate this finding (Trummer et al. 2022). In MDS, ctDNA demonstrated the emergence of novel clones anticipating progression to acute myeloid leukaemia (AML) (Yeh, Dickinson, et al. 2017).

Somatic mutations and CNA found in the tumour DNA of RT are mirrored in the ctDNA (Yeh, Hunter, et al. 2017). RT cells precede the development of the clinical syndrome by many years (Nadeu et al. 2022). No studies examined the interval between development of RT-associated ctDNA profiles and subsequent clinical presentation with RT.

1.6 Nanopore Sequencing in molecular diagnostics

Nanopore Sequencing (NS), most familiar as the commercial MinION from Oxford Nanopore Technologies (ONT) represents a distinct sequencing modality due to ability to sequence long DNA strands, its portability and its relatively low cost. The USB sized device connects to a laptop, resulting in a smaller laboratory footprint.

1.6.1 Principle of Nanopore sequencing

The nanopores sit in a membrane that separates two ionic solutions, allowing an electrical current to flow through the nanopores. Individual DNA strands move through biological nanopores under the influence of an electrical field. Double stranded DNA (dsDNA) is bound to a lead adapter and a hairpin adapter. The lead adaptor passes through the nanopore, followed by the template DNA strand and the hairpin adaptor and then the complementary strand. DNA molecules pass through the protein pore, producing an ionic current change as they pass through, which is detected by a sensor with a constant sampling frequency. The rate of DNA strands passing though the nanopore is controlled by a processive enzyme bound to DNA at the pore orifice. A maximum of 512 DNA molecules can be determined simultaneously through each independent nanopore. The change in electrical current is converted into a nucleotide sequence using the base calling algorithms (Goodwin et al., 2016; Magi et al., 2017). Each change corresponds to a five-nucleotide DNA ƙ-mer. Metrichor, an ONT developed cloud-based software employing Hidden Markov Model statistics calls DNA bases. Unlike second-generation sequencing methods, no new strands are created.

Figure 1.6.1.1 Illustration of the process of nanopore sequencing

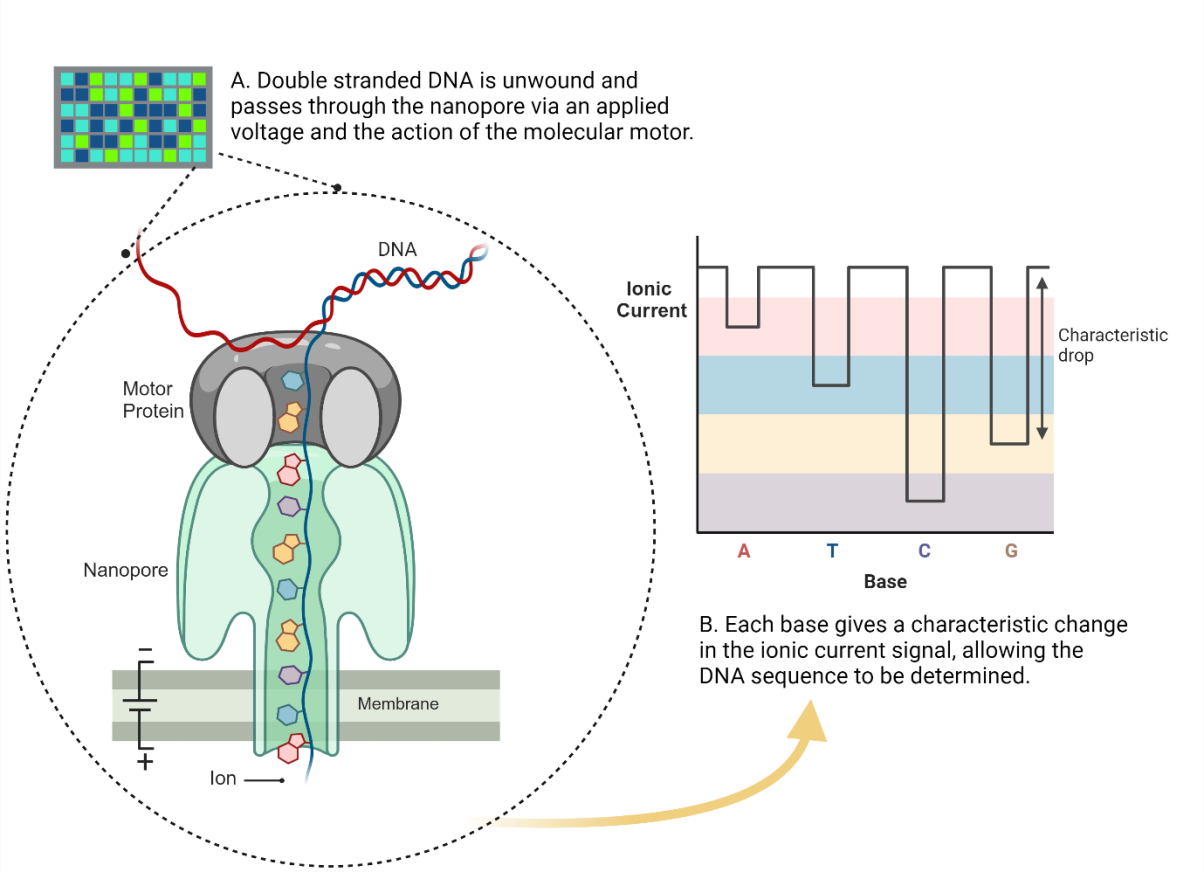


Figure 1.6.1.1: Visual representation of the nanopore sequencing. Within the nanopore flow cell, active nanopores are represented in with a green square. The DNA strand, (combined with lead adaptor and molecular motor enters the nanopore (A) via an applied voltage and the action of the molecular motor. Using the hairpin adaptor, the complementary strand and trailing adaptor pass through the nanopore in turn (not pictured). Each DNA base generates a unique current signal used to aid base calling. Figure made in BioRender.com

1.6.2 Potential advantages of long-read sequencing using NS

Long-read sequencing enables the determination of the allelic context of mutations affecting the same gene, when the mutations are too distant to be captured on the same read by short-read sequencing-by-synthesis (SBS) methods. NS facilitates the analysis of GC-rich regions or repeat regions difficult to examine with SBS approaches. NS facilitates sequencing of longer reads, a higher throughput, and is more competitively priced than an alternative long-read sequencing technology, PacBio (Minervini et al. 2020).

Initially, the main limitation of NS was the high base error rate compared with SBS. NS is relatively unaffected by GC-rich regions and the error rate in NS is random. Homopolymer regions remain a source of errors in NS as the electric signal is unable to correctly determine the length of the region. However, refinements in sequencing chemistry, improvement of base-calling algorithms, development of post-sequencing correction tools and tools to determine SNV/indels have reduced the error rate with MinION (Volden et al. 2018; Sereika et al. 2022).

1.6.3 Applications of nanopore sequencing to haematological cancers

Genomic data permits insights into the pathogenesis of haematological cancers and informs diagnosis, prognostication, treatment decisions and monitoring for MRD. NGS technologies, predominantly SBS have expanded our knowledge base but remain expensive and time-consuming, limiting patient access to the genomics driven cancer care. Short reads obtained using SBS have limited capacity to detect SV, especially translocations and are unable to identify biallelic mutations. The economical cost of investing in SBS technology remains a barrier to widespread access to genomic diagnostics.

Targeted sequencing of pre-specified somatic mutations is the most widespread clinical application of SBS. Results of panel based targeted sequencing is used to aid clinical decision establishing a diagnosis, prognostication, selection of therapy and monitoring following treatment. NS can detect somatic mutations in clinical samples. Using NS, *TP53* mutations are detectable as a single test (Minervini et al. 2016) or with a targeted panel (Orsini et al. 2018). In the case of multiple *TP53* mutations, NS determines the allelic context.

SV defined as regions of genomic change > 50bp include translocations, inversions, insertions, deletions and combinations of these resulting in location shift or change in copy number. In clinical diagnostics, SV are commonly identified using FISH alone or in combination with conventional karyotyping. These techniques are arduous and limited by low sensitivity. To optimally characterise many haematological cancers, information on both SV and the presence or absence of specific mutations is required.

The long strands of NS facilitate the accurate identification of SV and is not restricted by the relatively high error of NS (De Coster et al. 2019). In Philadelphia chromosome positive leukaemia, relapsed disease is mediated by mutations in the kinase domain of *BCR-ABL* conferring insensitivity to tyrosine kinase inhibitor therapy. Long-read sequencing simplifies the characterisation of genomic junctions and has been used to facilitate personalised monitoring of *BCR-ABL* with digital PCR (Cumbo et al. 2018). NS can detect *BCR-ABL* kinase domain mutations (Minervini et al. 2017). Unlike SBS, NS distinguishes distinct mutations in different disease clones. Identifying patients with multiple mutations in the same clone is

clinically relevant as this subgroup experience an increased likelihood of resistance to second generation tyrosine kinase inhibitors (Soverini, Baccarani, and Martinelli 2016). Nanopore based WGS has detected of SV in AML and CLL (Au et al. 2019; Burns 2018). Detection of multiple translocations and fusion genes using NS has been tested in cell lines and clinical samples and offers a more rapid turnaround compared with conventional karyotyping (Jeck et al. 2019).

1.6.4 Cell-free DNA (ctDNA & foetal DNA) using Nanopore sequencing

Given many of the advantages of NS derive from the ability to capture longer DNA fragments, it seems counter intuitive to explore NS in short fragment cell-free DNA. However, the potential advantages of economical set-up, relatively short laboratory workflows and preservation of methylation status merit exploration.

Current NS methodologies have a higher error rate than Illumina SBS methods (Lu, Giordano, and Ning 2016) and the quality of the base calling declines with read lengths of <200 nucleotides, possibly due to premature termination of sequencing on the 1D strand or steric interactions at the hairpin adapter (Krishnakumar et al. 2018).

One approach to overcoming the limits of NS for short fragments is to circularise each short DNA fragment followed by amplification of each circular DNA piece to generate linear DNA consisting of multiple tandem repeats (concatemers). (Li, Chng, et al. 2016). However, the technique requires blunt end ligation using minimum length of initial template DNA of 600-800 bp, beyond the typical 170 bp fragments of cell-free DNA.

Application of NS to cell-free DNA offers the potential advantage of sequencing outside major centres and providing real-time results as data can be processed while sequencing is ongoing. The potential advantages of rapid, near-patient testing with NS prompted its successful exploration in non-invasive prenatal testing (Cheng et al. 2015; Jiang et al. 2021) and detection of CNA from cell-free DNA in cancer (Martignano et al. 2021).

1.7: Study Objectives

The heterogenous clinical behaviour of CLL offers a model to study the progression from the pre-cursors clonal disorder to asymptomatic cancer to symptomatic disease.

Key questions my thesis aims to address:

1. Whether molecular features detected by targeted deep sequencing inform prognosis in MBL and early CLL
2. Whether plasma ctDNA provides a tool for disease assessment for
    a. Minimally invasive detection of clinically relevant mutations and CNA in nodal B-cell malignancies (RT and DLBCL) at baseline
    b. Minimally invasive detection of persistent disease after treatment
3. Whether nanopore sequencing can provide an alternative tool for molecular profiling of gDNA and ctDNA
4. Whether the mutational profile (SNV, indels and SV) observed in plasma ctDNA provides a comprehensive assessment of PB, LN and BM disease.

To achieve these objectives, I will

1. Establish two prospective clinical studies (OXPLORED & STELLAR) spanning the clinical spectrum of MBL, CLL and RT encompassing sequential PB, BM and plasma sampling.
2. Test different sequencing approaches for ctDNA mutation detection using plasma samples from the high-grade B-cell lymphoma trials (Chapters 4 & 5)
3. Develop a NS based methods to determine
    - IGHV and *TP53* mutation status
    - Copy number alterations (CNA) in gDNA and cell-free DNA

Chapter 2: Methods

2.1 Patient samples and additional data sources

2.1.1 Existing cohorts of MBL and CLL

Monoclonal B-cell lymphocytosis (MBL) and Rai 0 CLL samples were collected under LYM1 Study (REC06/Q2202/30) and obtained via Material Transfer Agreement (MTA). PBMC were stored at $5x10^7$ cells/ml in 10% DMSO 90% foetal calf serum in liquid nitrogen. Hereafter, these patient samples are referred to as the LYM1 cohort.

Data on chromosomal deletions by FISH and IGHV status determined by Sanger sequencing (SS) were provided by Bournemouth NHS Trust. These results were compared with my results in Chapter 6.

Advanced CLL samples for nanopore sequencing (NS) were obtained with consent and the approval of the Oxford Radcliffe Biobank (ORB) (09/H0606/5). Samples were previously characterised by SBS and/or SNP arrays (Stamatopoulos et al. 2017; Burns et al. 2018; Stamatopoulos et al. 2018) and selected for known CNA, *TP53* aberrations and IGHV rearrangements.

2.1.2 Samples and IGHV status for OXPLORED participants

Additional MBL and early CLL samples came from the OXPLORED study (REC 19/SC/0065), detailed in Section 2.2. At the time of submission, baseline clinical characteristics and status at 24-months was available. OXPLORED participants provided optional informed consent to the National Genome Research Library (NGRL; REC: 20/EE/0035). Sorted CD19+ B-cell (tumour) and germline (salivary) DNA samples were submitted to Genomics England's Illumina facility for WGS.

Data on IGHV status was determined from WGS data using IgCaller, a python3 programme designed to identify immunoglobulin gene rearrangements, translocations and mutation status from WGS data (Nadeu et al. 2020). WGS data (CRAM files and vcf) was provided in the National Genome Reference library (NGRL) via the Haematological Malignancy Genomics England Clinical Interpretation Partnership (GECIP). Indexed bam files were input into IgCaller v1.2.1, with assistance from Dr Dimitris Vavoulis. To assign CLL stereotypes, the IgCaller output was used as input for AssignSubsets online tool (http://bat.infspire.org/arrest/assignsubsets/), which annotates the 19 major subsets, including subsets #1, #2, #4 and #8 (Bystry et al. 2015). Where more than one rearrangement was detected, the highest scoring rearrangement was considered as the principal rearrangement.

No data on mutations or copy number alterations (CNA) was available at the time of submission.

2.1.3 Richter's transformation

Richter's transformation (RT) samples from the O-CHOP trial (Eyre et al. 2016) were collected under (REC10/H0604/85, NCT01171378) and accessed under ORB 18/A002. Prospective RT samples came from the STELLAR clinical trial (EudraCT: 2017-004401-40, NCT03899337), described in Section 2.3.

Illumina WGS datasets prepared from FFPE tumour DNA (tissue-RT) and circulating leukocytes (PB-CLL) from 17 O-CHOP participants were published (Klintman et al. 2021). PCR-free libraries were sequenced on the HiSeq 2500/HiSeq 4000 to a mean depth of 87x for PB-CLL (range, 34× to 129×), 88× for tissue-RT (range, 42× to 132×), and 44× for germline (range, 25× to 67×) samples. For copy-number analyses, paired and unpaired RT-tissue and germline, and PB-CLL and germline WGS dataset and were analysed and events flagged and visualised using Nexus Discovery Edition 7.5 (BioDiscovery, Inc., El Segundo, CA). All candidate CNA were manually curated. These WGS results were considered the 'ground truth' for comparison with the nanopore copy number calls reported in Chapter 7.

Baseline parameters were obtained from each trial dataset. Response rates and clinical outcome data was obtained from the Oncology Clinical Trials Office for the O-CHOP cohort. As STELLAR is an ongoing trial, data on response at interim assessment only was provided by the Clinical Trials Research Unit at Birmingham.

2.1.4 Diffuse Large B-cell lymphoma

Paired PBMC (germline) and plasma cell-free DNA samples were collected at baseline and further cell-free DNA samples at C2D1, C3D1 and EOT from the National Cancer Research Institute (NCRI) phase Ib/phase II ACCEPT trial for DLBCL (REC 16/SC/0657, EudraCT: 2015-003213-18, NCT03571308). The primary outcome was dose limiting toxicity of acalabrutinib combined to R-CHOP (phase Ib). In the phase II cohort, the primary outcome was the overall response rate (ORR) and safety of the combination therapy. As the study had completed recruitment, baseline data and clinical outcome data was obtained from the Southampton Clinical Trials Unit (CTU).

2.1.5 Cell lines and normal control samples

Two non-cancer cell-free DNA control samples were obtained from the Suspected CANcer (SCAN) pathway. SCAN is an investigative pathway for adults > 40 years old with non-specific symptoms (weight loss or fatigue), who do not fulfil urgent cancer referral criteria or for whom symptom causation remains unclear after investigation (Nicholson et al. 2018). Selected participants had no evidence of cancer after senior medical evaluation and CT Thorax/Abdomen/Pelvis imaging. Participants provided consent for use of blood for research purposes (REC 19/SC/0173). A normal tonsillar tissue biopsy was used as a normal control for gDNA. (REC09/H0606/5+5; OCHRe 18/A135).

As clinical cell-free DNA samples were limited, the initial sWGS experiments used cell line DNA. TC413 (CA46), an EBV-negative Burkitt lymphoma cell line was chosen as a karyotype is published (Zimonjic, Keck-Waggoner, and Popescu 2001). TC413 was obtained from Dr Sam

Mbulaiteye, an external collaborator on a NIHR-RIGHT grant to investigate liquid biopsy for paediatric lymphoma diagnosis in East Africa (Legason et al. 2022) Non-cancer control plasma from EBV-positive infectious mononucleosis samples were shared from the same project.

Figure 2.1.5.1 Workflow of patient samples, sequencing methods and thesis structure

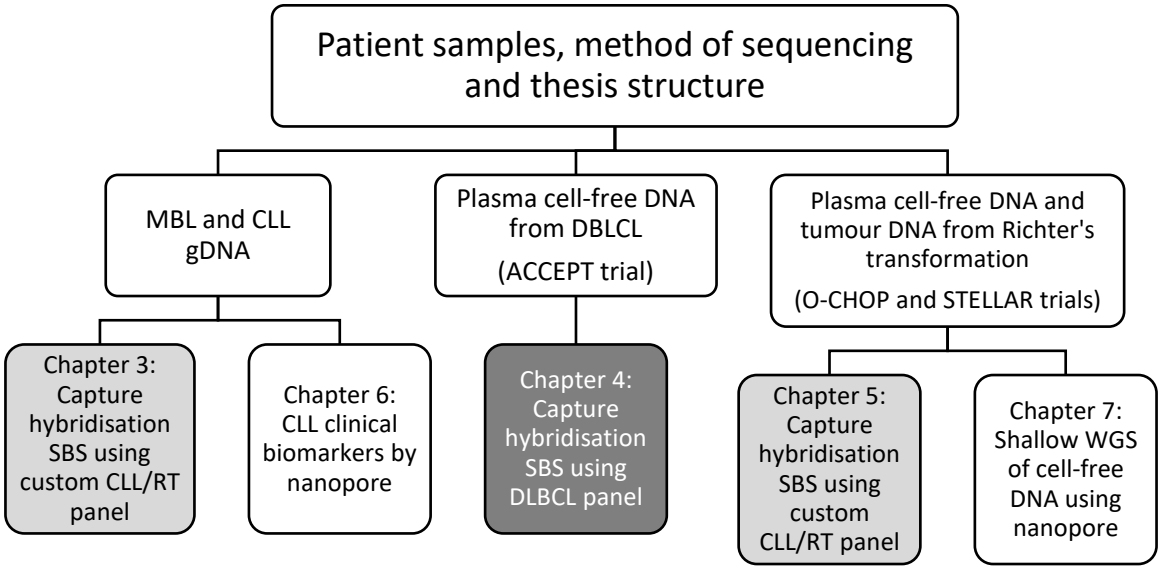


Figure 2.1.5.1: CLL Chronic Lymphocytic Leukaemia; DLBLC Diffuse Large B-cell Lymphoma; MBL Monoclonal B-cell lymphocytosis; SBS sequencing-by-synthesis; RT Richter's transformation; WGS whole genome sequencing. The CLL/RT panel (light grey) was used for gDNA MBL/CLL and cell-free DNA from RT. The DLBCL panel (dark grey) was used for DLBCL.

2.2. Establishment of the Oxford Premalignant Lymphoproliferative Disorders (OXPLORED) study

OXPLORED is an observational cohort study for precursor mature clonal B-cell conditions. Adults with MBL, early asymptomatic CLL, monoclonal gammopathy of uncertain significance (MGUS), asymptomatic smouldering myeloma (SM) and asymptomatic WM diagnosed in the preceding three years are eligible. Up to 31st August 2022, the study was open at eleven sites and recruited 274 participants. My thesis focuses on MBL and early asymptomatic CLL.

2.2.1 OXPLORED Study Objectives

The primary objective is to identify the clinical, genomic and immunological predictive markers of progression to malignant disease (Table 2.2.1.1). These primary outcome measures will be combined into a single probability risk score. The choice of relevant markers will be guided by emerging evidence and techniques, under the guidance of the Study Scientific Advisory Board.

Table 2.2.1.1 Objectives of the OXPLORED study

Primary objectives	Identify clinical and genomic biomarkers predictive of disease progression. Develop a probability risk score
Secondary objectives	Assess Quality of Life (QoL) in early indolent lymphoproliferative disorders. Recording of non-progression clinically significant events (hospitalisation with infection, thrombosis, non-lymphoid cancers, and deaths not due to progression to lymphoid cancers)
Exploratory objectives	Gather feasibility data for future early intervention trial on high-risk subgroup. Identification of novel surrogate markers of cure Establishment of evidence-based guidance for monitoring and follow-up of early lymphoproliferative disorders

The OXPLORED study continues beyond the duration of my D.Phil. My thesis focuses on MBL/CLL participants where I examine

- Sensitive sequencing techniques to identify somatic variants linked to progression and long-term stable disease (Chapter 3).
- Nanopore sequencing (NS) to determine clinically relevant biomarkers in MBL/CLL (Chapters 6).

2.2.2 Ethics application, procedures and approvals

The study protocol, patient information documentation, case report and consent forms were designed by myself and approved by the study sponsor, the University of Oxford. Patient and public involvement (PPI) input from the Oxford Blood Group at the Oxford Biomedical Research Centre informed study design and documentation. Participants consent for genomic analyses of tumour, blood and germline samples, with optional consent for sample and data sharing with Genomics England and the National Genome Reference Library (NGRL).

This study is undertaken in accordance with the ethical principles from the Declaration of Helsinki and Good Clinical Practice. The study was approved by the South Central (Oxford) Research Ethics Committee (REC 19/SC/0065), with the current v5.2 approved on 27 September 2022. OXPLORED was registered on the NIHR Central Portfolio Management System (NCT 242634) and clinicaltrials.gov (NCT04023747).

Protocol synopsis can be found in Appendix A.

2.2.3 Study design and procedure

Participants provide samples for germline DNA analysis, tumour and non-tumour leukocyte DNA sequencing and cell-free DNA profiling (see Section 2.4). Initially, all participants underwent bone marrow aspirate (BMA) and imaging. The protocol was later amended to make BMA and CT imagining optional for MBL/early CLL participants. BMA and cross-sectional imaging were retained for MGUS as these investigations represent standard-of-care. Clinical and laboratory assessments for Cohort 1 (MBL/CLL) and study schedule are summarised in Table 2.2.3.1. Further details on cell sorting, plasma processing, DNA extraction and QC metrics can be found in Section 2.4.

Participants continue their NHS care pathway throughout (Figure 2.2.3.1). Telephone follow-up at 12 months and in-person follow-up at 24-months was chosen to balance identifying early progressors with the burden of study participation. As NHS hospitals use a mixture of paper based and digital healthcare records, study data is captured using electronic case report forms (Figure 2.2.3.2) to maximise uniformity and completeness of the dataset. Many NHS sites (including OUH NHS Trust at the time of study establishment) record clinical examination findings and outpatient visits on paper, limiting the capacity of NHS data streams to capture key study parameters. Participant data is entered into OpenClinica, the University of Oxford clinical trial data capture system.

Table 2.2.3.1: Summary of clinical and laboratory assessments for Cohort 1 (MBL/CLL) participants in OXPLORED

Assessment		Enrolment	12 Month	24 Month	60 Month	Progression [a]
Informed consent		X				
Demographics and Relevant Medical History		X				
Clinical and Physical Assessment		X		X	X	X
ECOG performance status		X				X
Height and weight		X				X
Pregnancy Test (if relevant)		X				X
Biochemistry/Haematology/Immunology		X		X	X	X
CT Thorax/Abdomen/Pelvis [b]		X [b]				X
Research Samples	Blood for cell-sorts, DNA/RNA, plasma	X		X	X	X
	Blood sample for biobanking (optional)	X		X	X	X
	Salivary sample (germline DNA)	X				
	Bone marrow aspirate [b]	X [b]				X [b]
	Aspirate samples for biobanking (optional)	X				X
Quality of Life Questionnaires		X		X	X	X
Concomitant Medications		X		X	X	X
Clinically Significant Events			X	X	X	X
Telephone follow up			X			

a Participants investigated for progression but found not to require treatment remain on study.
b Optional for MBL/CLL participants

Figure 2.2.3.1: Participant flow in the OXPLORED Study

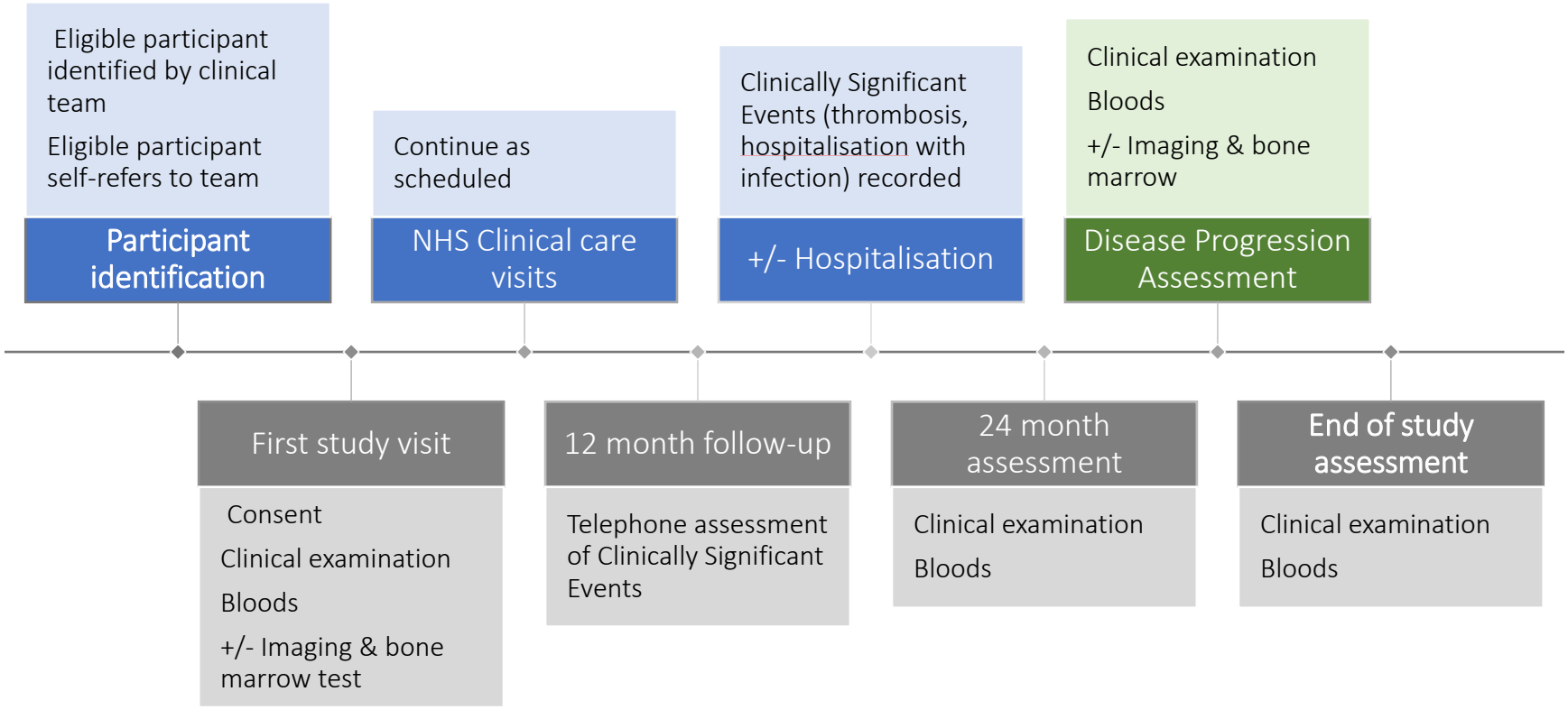
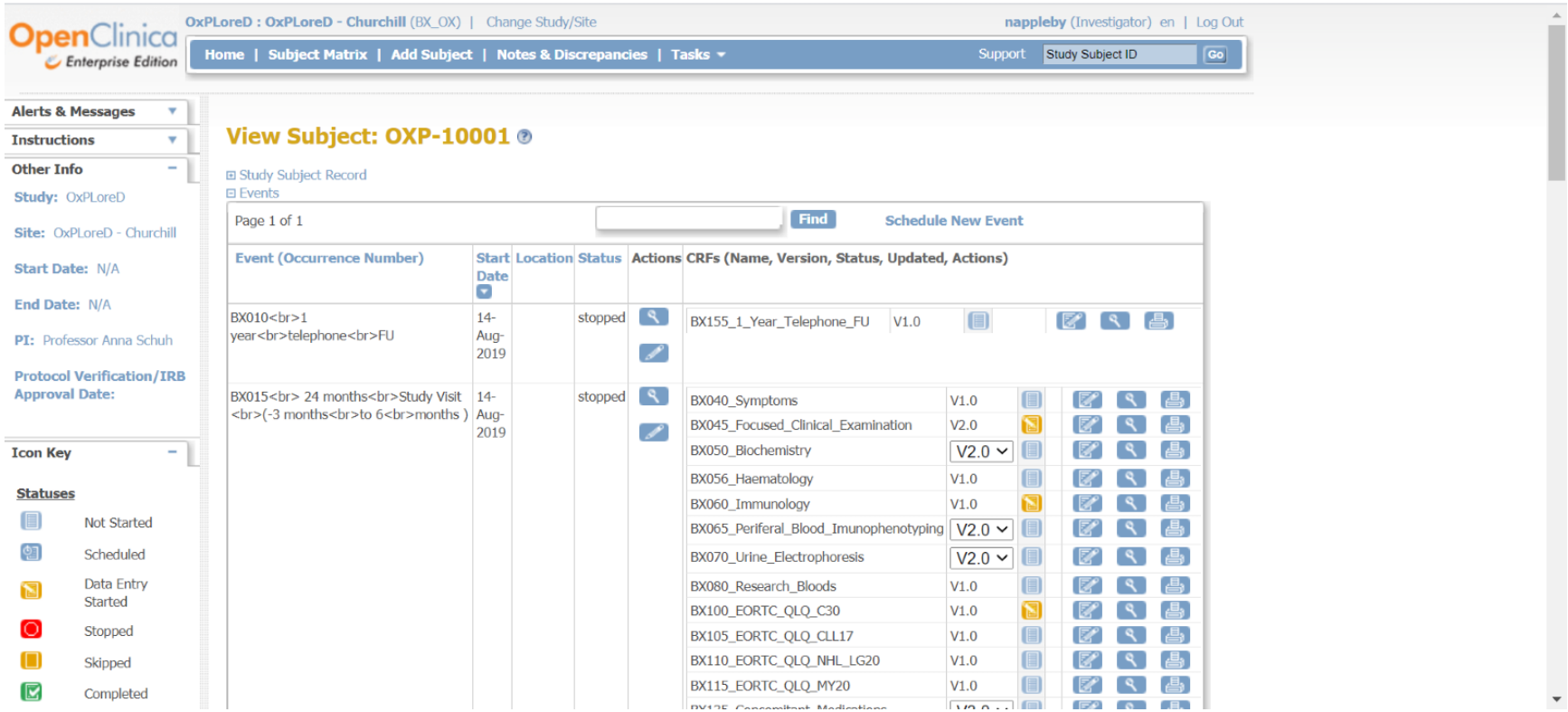


Figure 2.2.3.1: Schematic illustrating the relationship between the OXPLORED study visits (grey boxes) and the NHS standard-of-care visits (blue boxes). At disease progression (green boxes) a study visit coincides with NHS care at progression.

Figure 2.2.3.2: Screenshot from OpenClinica showing data collection matrix for OXP-10001



2.2.4 Statistical considerations for OXPLORED study design

Sample size and power calculations were conducted with Dr Dimitris Vavoulis, computational statistician with the Oxford Molecular Diagnostics Centre (OMDC).

The sample size calculation assumed that 10% of OXPLORED participants would progress to require treatment and up to 20% participants would be lost to follow-up. In order to recruit 100 progressors, 1650 participants were needed.

OXPLORED will examine known candidate predictors of early progression. It is likely that no candidate predictor will have sufficient predictive value on its own, therefore the study will combine multiple predictors into a risk score for each group. The simplest approach will be to apply a χ^2 test on a 2×K contingency matrix, where 2 stands for the progressor/non-progressor phenotype, and K is the number of levels in each categorical predictor, (a significance level of 5% and two-sided tests are assumed). For example, using number of genes on a targeted panel affected by mutations (grouped as 0, 1 and ≥ 2) a χ^2 test with two degrees of freedom can be applied on a 2x3 contingency table.

Further statistical considerations are detailed in Appendix A.

2.3. Establishment of the STELLAR Clinical Trial for Richter's Transformation (RT)

STELLAR is a prospective randomised trial for newly diagnosed DLBCL-type RT. I made significant contributions to the trial design (Appleby et al. 2019). The study includes two additional platforms for studying RT after chemoimmunotherapy (Cohort 1) and RT arising on BTKi treatment (Cohort 2). As of 1 September 2022, 14/17, potential UK Trial Acceleration Programme (TAP) sites are open. 23 participants enrolled in the Randomised cohort, eight in Cohort 1 and four in Cohort 2. My thesis evaluates the potential of ctDNA sequencing in RT (Chapter 5).

2.3.1 Objectives of the STELLAR trial

The primary objective of the randomised study is comparison of progression free survival (PFS) between R-CHOP and R-CHOP plus acalabrutinib groups. For both Platform Cohort 1 (Progressive RT following chemoimmunotherapy) and Platform Cohort 2 (de novo RT arising on ibrutinib), the primary outcome is ORR (complete response (CR) or partial response (PR)) after 24 weeks of treatment (Platform Cohort 1) or 6 cycles of R-CHOP plus acalabrutinib (Platform Cohort 2). Secondary objectives in all cohorts are Quality of Life (QoL), OS, treatment toxicity, and the proportion proceeding to cellular therapies (Table 2.3.1.1). Response assessment, including PET-CT based criteria, will use the categories defined by Cheson et al (Cheson et al. 2014). QoL will be evaluated using the ECOG performance status, European Organisation for Research and Treatment of Cancer (EORTC) CLL17 and NHL-HG29 questionnaires.

Table 2.3.1.1 Objectives of the STELLAR study

	Randomised cohort	**Platform studies**
Primary objectives	PFS between the R-CHOP and R-CHOP plus acalabrutinib	Overall Response Rate (ORR)
Secondary objectives	ORR after 6 cycles Overall Survival (OS) Treatment related toxicities Quality of Life (QoL) Proportion of RT patients receiving cellular therapy	OS Treatment related toxicities QoL Proportion of RT patients receiving cellular therapy
Exploratory objectives	MRD by flow-cytometry (10^{-4}/L) Relationship between ctDNA parameters and clinical outcomes	

MRD measurable residual disease; ORR Overall Response rate (Complete and Partial Response); PFS Progression Free Survival; QoL quality of life

Exploratory objectives include the proportion of patients attaining measurable residual disease (MRD) negative remission (defined as CLL cells < 1 in 10^4 by multi-colour flow cytometry), the relationship between ctDNA parameters and ORR, durable remissions or clinical relapse.

MRD-negativity is uncommon in ibrutinib studies (Ahn et al. 2018; Jain et al. 2017) and acalabrutinib trials have not reported MRD data to date (Byrd et al. 2016; Byrd et al. 2021). Sequential assessment of PB-MRD with paired ctDNA samples facilitates insights into dynamic changes in the circulating and nodal compartments of RT during treatment. My thesis focuses on ctDNA analysis alone (Chapter 5).

The STELLAR trial synopsis is included in Appendix B.

2.3.2 Ethics application, procedure and approval for the STELLAR study

The laboratory analysis plan, patient information documentation, case report forms and consent forms were designed by myself and approved by the study sponsor, the University of Birmingham.

The REC South Central Oxford B. (REC: 18/SC/0634) approved the trial on the 31 Jan 2019 (protocol v2.0), the current protocol v5.0 was approved 21 Mar 2021. The study was amended to facilitate participants with diagnostic FFPE biopsies to minimise repeat biopsies and delays to enrolment. As of 1 September 2022, 23 participants enrolled in the Randomised cohort, eight in Cohort 1 and four in Cohort 2.

2.3.3 STELLAR Study design and procedures

The STELLAR study is designed in two elements, consisting of a randomised study to evaluate the safety and activity of R-CHOP (cyclophosphamide, doxorubicin, vincristine, prednisolone and rituximab) chemoimmunotherapy alone or with acalabrutinib in newly diagnosed RT (Figure 2.3.3.1) and single-arm studies for other RT patient cohorts (Figure 2.3.3.2).

A histological diagnosis established by local pathology review is acceptable for trial inclusion, mirroring “real-world” clinical practice. Histological criteria for fulfilment of the diagnosis of RT (DLBCL-subtype) include: (i) large B-lymphoid cells with a nuclear size equal to or exceeding that of normal macrophage nuclei or more than twice the size of a normal lymphocyte; and (ii) diffuse growth pattern (Swerdlow et al. 2016; Soilleux et al. 2016). Although specialist haematopathologist review alters the diagnosis in 17.5% of suspected RT cases (Soilleux et al. 2016), the aggressive clinical course of RT means patients often need to start treatment urgently. In the context of the STELLAR trial, a requirement for centralised haematopathology review would delay recruitment and potential bias the study towards participants with more indolent disease course making the study population unrepresentative of the general RT population.

At the time of study conception, there was a paucity of data on ctDNA in RT and in BTKi treated lymphoid malignancies and therefore the optimal timepoint for ctDNA sampling was

unknown. As RT is clinically distinct from DLBCL, I decided it would be inappropriate to assume that the Cycle 3 Day 1 timepoint reported as predictive in DLBCL (Kurtz et al. 2018) would be informative in RT. In addition, the mode of action of BTKi with redistribution of tissue based neoplastic lymphocytes into the circulation was anticipated to result in a differing dynamic profile of ctDNA in the study. I opted to collect plasma cell-free DNA samples with each treatment cycle. STELLAR samples collection is summarised in Table 2.3.3.1. Results of clinical evaluations, diagnostic investigations and response assessments are collected using a standardised clinical record, pseudo-anonymised and recorded on an online secure access database at the TAP central site. Full eligibility criteria are summarised in Appendix B.

2.3.4 Statistical considerations for STELLAR

Sample size calculations and statistical analysis plans were designed by trial statisticians at the Birmingham CTU using SAS (v9.4) and R (v3.4.2). The primary analysis for the randomised component of the trial is powered. Data from the single-arm cohorts will provide initial activity data only with the primary analyses of these being Bayesian in nature. Platform cohorts will use a Bayesian framework utilising a weakly-informative prior of Beta to establish the probability that the true event rate exceeds the level designated to represent promising activity (see also Appendix B).

Sample size determination and power in the randomised cohort

PFS is the primary outcome for the randomised cohort. Based on previous studies, the PFS at 2 years is anticipated to be 20% in the standard of care R-CHOP arm (Eyre et al. 2016; Langerbeins et al. 2014). An increase to 40% PFS in the intervention arm (R-CHOP plus acalabrutinib) is considered clinically significant.

54 randomised participants are anticipated to generate sufficient progression events to compare the trial arms at one-sided significance level of 0.15 to achieve power of 80%. The sample size is based on the log-rank test and based on the assumption of exponential survival times with a recruitment over 3 years and follow up for a minimum of 2 years. Allowing for 10% drop-out rate, 60 participants (30 per arm) is the target recruitment population for the randomised component of the trial.

Sample size determination and power in the platform cohorts

For Platform Cohort 1 (Progressive RT following chemoimmunotherapy) an ORR > 20% will represent promising clinical activity for acalabrutinib monotherapy. If responses are observed in 14–29% of 21 participants, the posterior probability that the true response rate exceeds 20% is 33–87%.

For Platform Cohort 2 (anthracycline-naïve RS patients, diagnosed while on ibrutinib), R-CHOP with acalabrutinib will represent promising clinical activity if the ORR > 50%. As with cohort 1 a Bayesian approach utilising a Beta [1] prior. If responses are observed in 8–11 patients, the posterior probability that the true response rate is greater than 50% is 60–96%.

Analysis of STELLAR Outcomes

The intention to treat population, consisting of all registered participants will be used for efficacy analysis, irrespective of treatment received. For analysis by treatment arm within the randomised component, subjects will be analysed based on randomisation. The safety population consists of all subjects who received at least one dose of treatment on study.

There is no formal interim analysis planned. Final analysis for the primary and secondary outcomes will take place once all participants have been followed for 2 years. There is a planned subgroup analysis for the randomised cohort based on the presence or absence of *TP53* disruption.

Figure 2.3.3.1: Schematic for the STELLAR Randomised and Cohort 1 (acalabrutinib monotherapy)

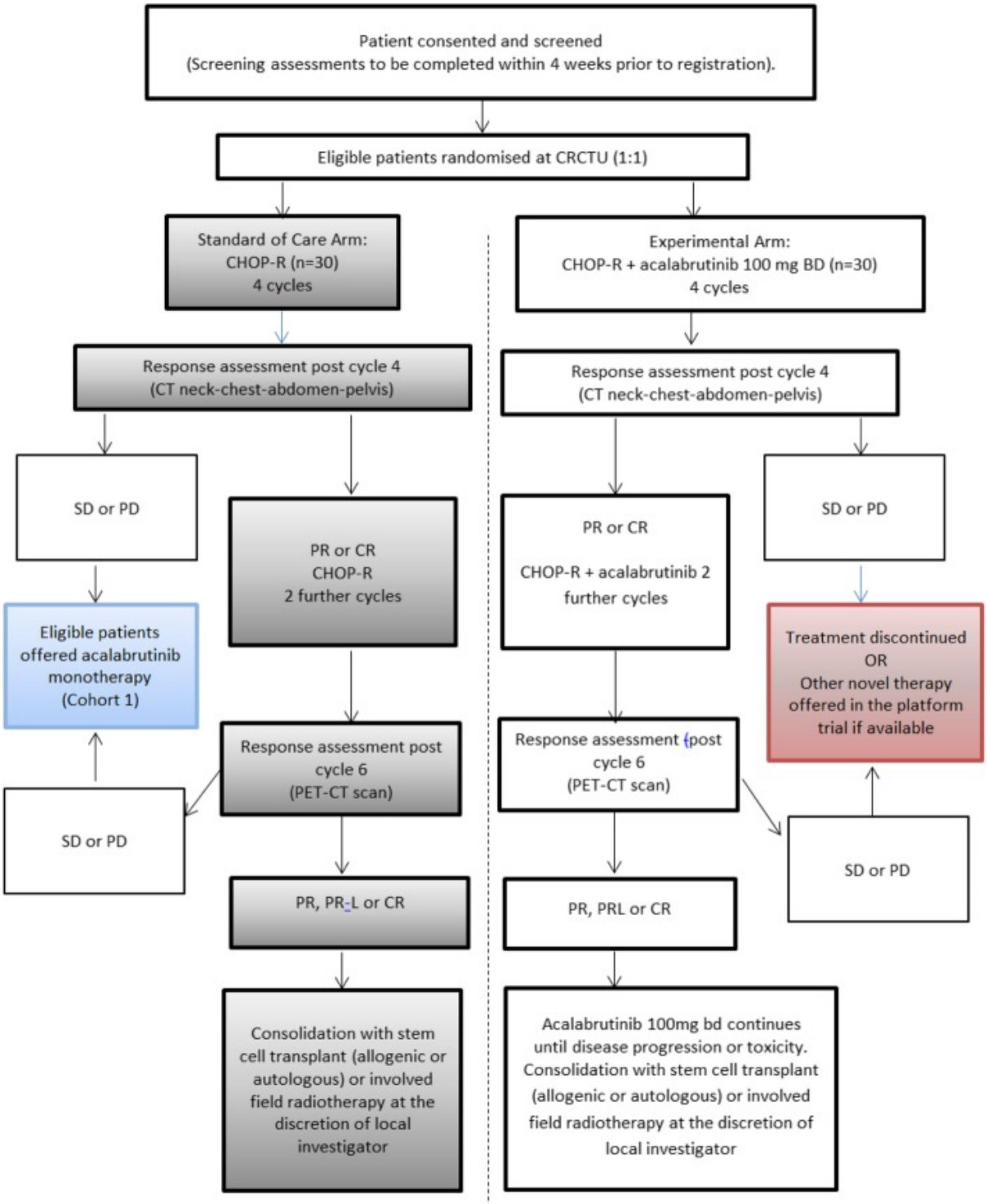


Figure 2.3.3.1: R-CHOP cyclophosphamide, doxorubicin, vincristine, prednisolone and rituximab; CRCTU Clinical Research Trial Unit; CR complete response; PR partial response; SD stable disease; PD progressive disease. Reproduced from (Appleby et al. 2019).

Figure 2.3.3.2: Schematic for STELLAR Cohort 2 (RT arising on ibrutinib).

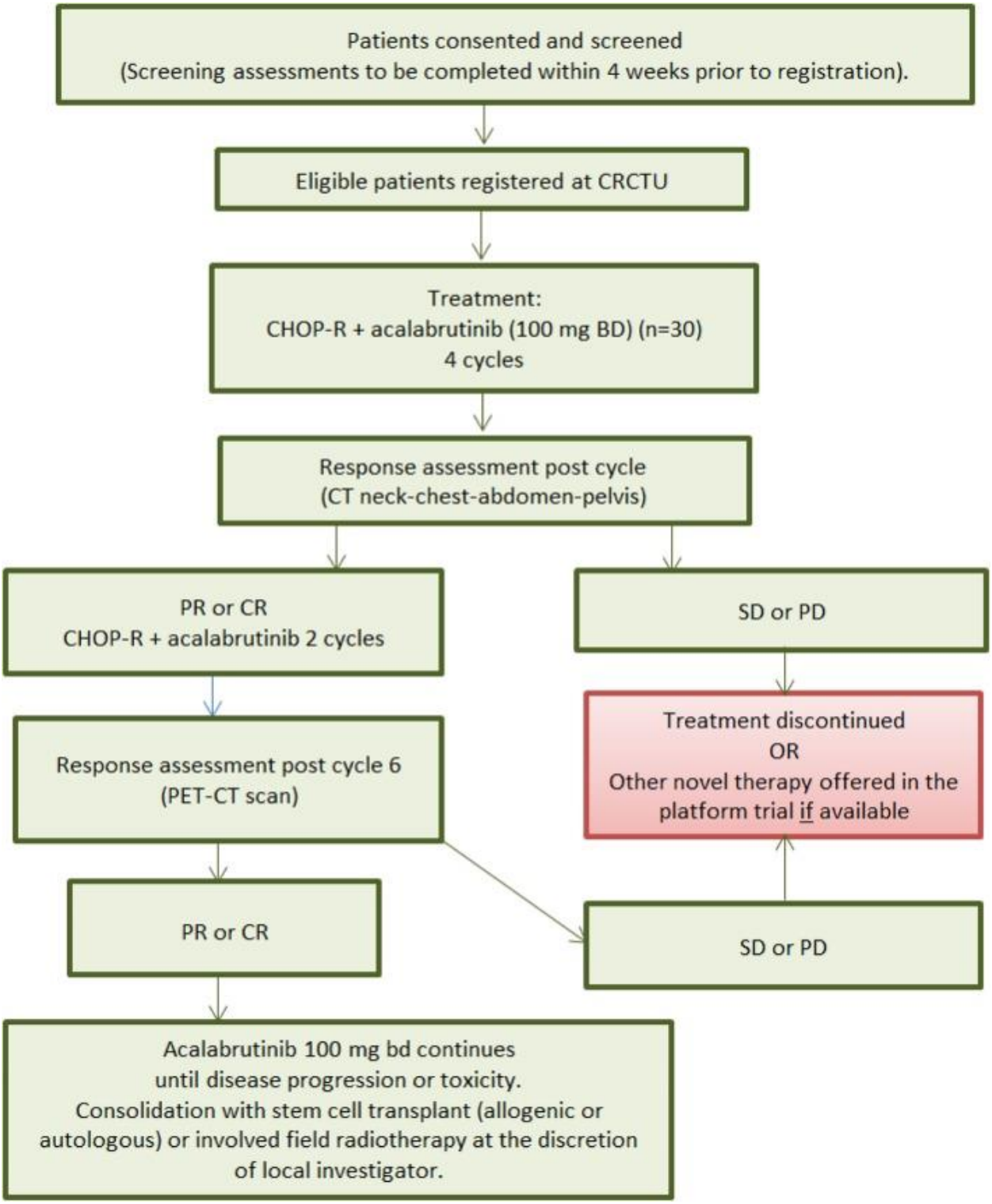


Figure 2.3.3.2.: R-CHOP cyclophosphamide, doxorubicin, vincristine, prednisolone and rituximab; CR complete response; PR partial response; SD stable disease; PD progressive disease. Reproduced from (Appleby et al. 2019).

Table 2.3.3.1. Summary of schedule for the STELLAR Randomised cohort and Cohort 2 (R-CHOP with acalabrutinib)

Assessments[1] *Unless otherwise indicated, all assessments are for all patients in the randomised arms and Cohort 2*		**Screening** within 4 weeks prior to registration	**During R-CHOP Treatment**				**Post-R-CHOP treatment**					**Treatment discontinuation**	**Disease Progression**	**Follow up**
			Cycles 1-4 (D1)	**Post Cycle 4 (C4D14-21)**	**Cycles 5-6 (D1)**	**Post Cycle6 (C6D14-21)**	**Week 24**	**Week 36**	**Week 48**	**Week 60**	**Week 72**			
Pregnancy test (patients of child-bearing potential)		X[3]												
Local laboratory assessments	Biochemistry & haematology	X	X		X		X	X	X	X	X	X	X	X
	Immunology & virology	X												
	BMAT	X[5]				X							X	
RESEARCH HMDS	BMA	X[7]				X							X	
	PB	X	X	X	X	X	X		X		X		X	
RESEARCH OMDC	PB and plasma	X	X	X		X	X		X		X		X	
	Lymph node biopsy (DNA)	X											X	
RESEARCH UKCLL Biobank	PB	X											X	
	Saliva	X												

BMAT bone marrow aspirate and trephine, PB peripheral blood,
HMDS Haematological Malignancy Diagnostic Service, OMDC, Oxford Molecular Diagnostic Service

2.4. Sample processing and DNA extraction

2.4.1 Cell sorting for circulating clonal B-cells for the OXPLORED cohort

Circulating clonal B-cell populations in MBL and early CLL are sorted to optimise the detection of low frequency DNA variants. I use OXPLORED sorted B-cells for targeted NGS (Chapter 3) and unsorted PB-cells for IGHV status (Chapter 6). Samples from the LYM1 cohort were supplied as PBMC in DMSO and were not sorted after thawing.

Different cell lineages express cell-surface markers known as Clusters of Differentiation (CD), which may function as receptors, ligands or play a role of cell adhesion. Labelled antibodies can be used to bind CD to select for different cell lineages. Separation of the T-cells facilitates greater accuracy in determining oligoclonality and T-cell receptor rearrangements, which will be evaluated in future OXPLORED work. Sample processing for OXPLORED MBL/CLL cohort is summarised in Figure 2.4.1.1.

Sorting Buffer (SB) was prepared from Phosphate Buffered Saline (PBS), Foetal Bovine Serum (FBS) and EDTA (OMDC standard operating procedure (SOP) HC2549). Briefly, 2ml of pH8 0.5M EDTA and 2.52ml FBS was combined with 500ml of PBS to obtain a solution of 2mM EDTA 0.5% FBS.

For each sample, 7ml EDTA whole blood was combined with 7ml SB in a labelled 15ml falcon tube and mixed slowly by pipetting until the solution is homogeneous. The mixture was centrifuged at 800x g for 10 minutes to separate blood into plasma, buffy coat and erythrocytes. Keeping the Pasteur pipette just below the meniscus, most of the plasma layer was removed. Up to 4.5ml of the buffy coat was transferred into a new 14ml round-bottomed tube and combined with an equal volume of 1X EasySEP™ RBC lysis buffer to remove the erythrocytes. To sort to T-cells, 25µL of CD3+ Selection Cocktail (Stem cell®) per mL of the combination was added, mixed by pipetting and incubated at room temperature for three minutes.

The RapidSpheres (EasySep™) were vortexed for 30 seconds to disperse the beads evenly. 25µL RapidSpheres of per mL was added to the cell-buffer mixture, mixed using a glass pipette and incubated at room temperature for three minutes. The bead-cell mixture was incubated on the magnet for six minutes.

After incubation, the supernatant was transferred to a labelled 50ml Falcon tubes using the lever on the magnetic stand. This CD3- supernatant was brought forward for CD19+ selection. The CD3+ cells remain in the 14ml round-bottomed tube. Using SB and a mechanical pipette, the cells underwent sequential washing steps before being resuspended in 5ml of SB. The supernatant (containing unsorted B-cells) was transferred to a new labelled 14ml round-bottomed tube. The cell solution was combined with CD19+ Selection Cocktail in a volume proportionate to the leukocyte count (Table 2.4.1.1), mixed and incubated at room temperature for five minutes.

Table 2.4.1.1: Guideline for leukocyte count and volume of CD19+ Selection Cocktail and RapidSpheres

Leukocytes *(x 10^9/L)*	***Volume/mL required*** *(µl)*
1.00 to ≤ 10.00	25
10.01 to ≤ 25.00	50
25.01 to ≤ 30.00	75
> 30	100

An equal volume of vortexed RapidSpheres was added, mixed by pipetting and incubated at room temperature for five minutes. SB was added to bring the sample up to 5ml or 10ml and incubated on the magnet for five minutes. Using the lever on the side of the magnetic stand, the supernatant was decanted into a 50ml falcon tube. This is the CD3-/CD19- non-lymphoid fraction.

Taking the 14ml round bottomed tube containing the CD19+ cells, the appropriate volume of SB was added and mixed gently. A cell count was performed on a 50ul aliquot of the CD3+ and CD19+ cell fractions. The cell fractions were suspended in 800ul FBS and 200ul DMSO and stored in 2ml FluidX tubes at -80°C. Batches of cells are transferred to liquid nitrogen periodically.

Figure 2.4.1.1 Sample flow for the MBL/CLL Cohort of the OXPLORED study

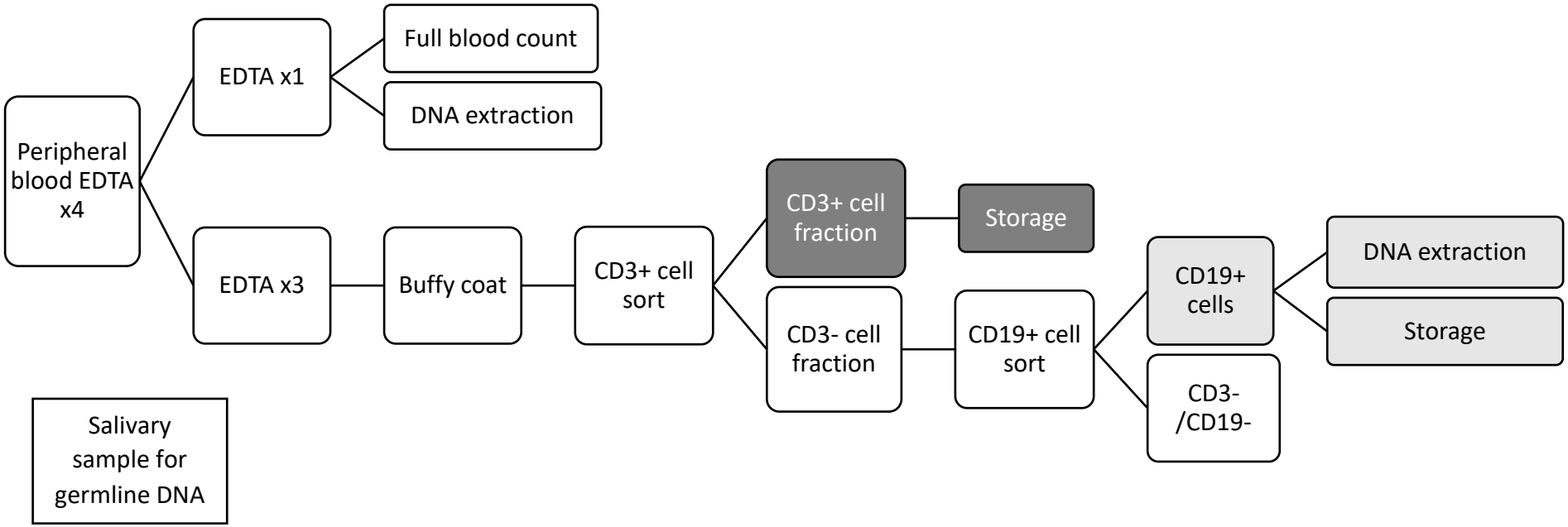


Figure 2.4.1.1: The CD3+ selection for T-cells (dark grey) was performed first, as unlike in healthy controls, circulating B-cells are in excess in MBL/CLL. CD19+ selection for B-cells (light grey) was performed on the CD3-negative fraction to maximise the purity of tumour samples for downstream analysis. Tumour purity was assessed by CD19+ flow cytometry during validation, training new staff and periodically thereafter to ensure tumour purity was > 90%.

2.4.2 Plasma processing prior to cell-free DNA extraction

Plasma samples from the O-CHOP cohort were collected in EDTA, centrifuged at 1800rpm for 20 minutes within 24 hours of phlebotomy and stored at -80C in accordance with SOP ORB-LAB015 v2.0.

STELLAR and ACCEPT plasma samples were collected in cell-free DNA preservative tubes and processed within 72 hours of phlebotomy (SOP OM0010). All samples were centrifuged at 1600g for ten minutes. The plasma layer was removed using a Pasteur pipette and transferred to a 1.5ml tube for a second centrifugated at 4500g for ten minutes. The plasma was transferred into labelled tubes and stored at -80°C.

2.4.3 Extraction and Purification of genomic DNA (gDNA)

The QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) was used to extract gDNA (gDNA). The extraction was performed according to the manufacturer's protocol, with minor modifications as described below. This method was used for ACCEPT germline DNA and extraction of gDNA from LYM1 and OXPLORED samples.

Salivary DNA (germline for OXPLORED and STELLAR) was extracted using the Oragene DNA kit according to the manufacturer's instructions (DNA Genotek, Ottawa, Ontario, Canada).

Extraction of gDNA from a cell suspension

The cell suspension sample was made up to 200µl with PBS and combined with 20µl of QIAGEN proteinase K to digest nucleases and contaminating proteins. The sample was mixed by vortexing with 200µl Buffer AL before incubating at 56°C with 500RPM shaking for 10 minutes. Following incubation, the sample was mixed by vortexing with 96-100% ethanol and centrifuged at 5900g for 15 seconds. The entire volume was transferred to a QIAamp Mini spin column and centrifuged at 16,100 x g for one minute. The QIAamp Mini spin column was then transferred to a new collection tube and centrifuged at 6000x g for one minute.

500µl Buffer AW1 was added to the QIAamp Mini spin column, and the column centrifuged at 6000x g for one minute. The collection tube was discarded, and the spin column transferred to a second 2ml collection tube. 500µl Buffer AW2 was added to the QIAamp Mini spin column, and the column centrifuged at 6000x g for one minute. The second collection tube was discarded. The QIAamp spin column was placed in a clean 1.5ml Eppendorf and 200µl Buffer AE added, incubated at room temperature for one minute and spun at 6000x g for one minute.

Salivary germline DNA (O-CHOP, OXPLORED and STELLAR cohorts)

Circulating neoplastic cells are seen in CLL and RT, necessitating an alternative to PB as a source of germline DNA. Salivary DNA was chosen as a pragmatic compromise between the desire for an uncontaminated sample and the feasibility of collection. While salivary DNA may be contaminated by neutrophils, this was less significant for lymphoid malignancies. Germline DNA is sequenced to target depth of 30x, limiting the risk of detecting low level variants arising from contaminating clonal B-cells. Salivary DNA germline was used as a source of germline DNA for

CLL in the Genomics England 100,000 Genomes Programme, facilitating comparison of OXPLORED and STELLAR results with this dataset (Robbe et al. 2022).

Salivary DNA was extracted using the Oragene DNA kit according to the manufacturer's protocol (DNA Genotek, Ottawa, Ontario, Canada). Briefly, the saliva collected was mixed by inverting the sample ten times and incubated at 50°C for 1 hour. The sample was transferred to a 15 ml Falcon tube containing a 1/25th volume of PT-L2P solution, vortexed for 10 seconds and incubated at 4°C for 10 minutes to precipitate impurities. The samples were then centrifuged at 3500xg for 10 minutes. The supernatant was transferred to a new tube and 1.2x volume of 100% ethanol precipitated DNA. The tube was mixed by inversions and incubated at room temperature for 10 minutes before centrifugation at 3500xg for 10 minutes. The supernatant was removed, discarded and the pellet washed with 1 ml of 70% ethanol. The sample was incubated at room temperature for 1 minute. The supernatant was removed. DNA (contained in the pellet) was rehydrated by adding 0.5 ml of TE buffer (10 mM Tris-HCl, 1 mM disodium EDTA), pH 8.0) (Sigma-Aldrich, St Louis, MO, USA) and vortexed for 30 seconds. The sample was incubated at room temperature overnight, vortexed for 30 seconds and transferred to a labelled tube.

O-CHOP cohort (FFPE tumour samples)

Paraffin was removed from O-CHOP FFPE 10µm sections scraped from 5-10 slides using the Adaptive Focused Acoustics (AFA™) on the M220 Focused-ultrasonicator. DNA was thereafter purified with the truXTRAC FFPE DNA Kit according to the manufacturer's protocol (Covaris Inc., Woburn, MA, USA), but using optimized reverse cross-linking conditions as previously described (Robbe et al. 2018).

STELLAR cohort (FF tumour samples)

DNA was extracted from STELLAR FF tissue biopsies using the All-Prep Mini Kit (Qiagen), as modified in OMDC SOP HC2523. 40ml Buffer RLT Plus and 400ml βME were mixed in the fume hood. 20µl buffer per mg of thawed tissue was used. The tissue was disrupted and homogenised using the rotor-stator homogeniser for 30 seconds. The tissue lysate was centrifuged at 1500x g for three mins. The supernatant was transferred to the AllPrep DNA spin column and spun for 30 seconds at 9600x g.

To extract gDNA, 500µl of Buffer AW1 was added to the spin column and centrifuged at 9600x g for 15 seconds. The flow-through was discarded. 500µl of Buffer AW2 was added and the tube spun for two minutes. The spin column was transferred to a new elution tube. 50µl of Buffer EB was added, incubated for one minute and centrifuged at 9600x g for one minute to elute DNA. This was repeated for a second elution.

2.4.4 Extraction of cell-free DNA from plasma

Cell-free DNA was extracted using the QiaAMP circulating nucleic acid extraction kit (Qiagen, Hilden, Germany) in accordance with manufacturer's instructions and OMDC SOP Haem2115 (Figure 2.4.4.1).

Figure 2.4.4.1: Workflow of the QIAamp circulating nucleic acid extraction kit procedure
(image from QIAamp DSP Circulating NA Kit Handbook Version 1, September 2019)

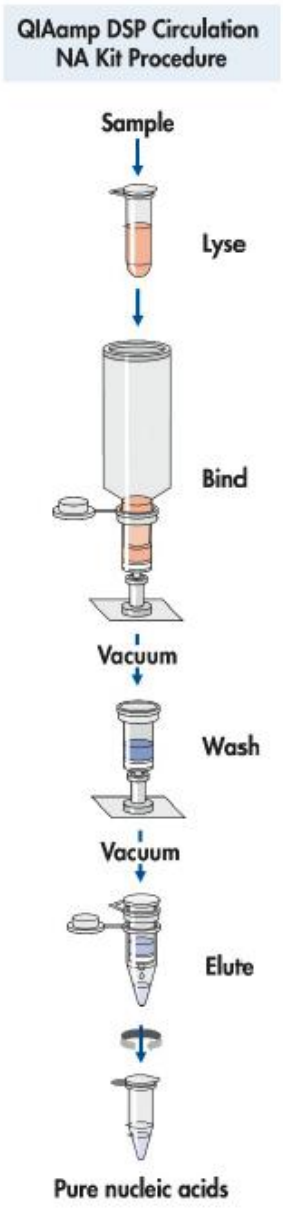


Figure 2.4.4.1: Schematic for extraction of plasma cell-free DNA. 1-5ml of plasma was mixed with the appropriate volume of Qiagen Proteinase K, carrier RNA and Buffer ACL and incubated at 60°C for 30mins for EDTA samples and 60mins for Streck®/Roche® samples. Buffer ACB was added to the lysate and incubated on ice for 5 minutes. Each sample was inserted on a spin column using the Qiagen vacuum manifold and sequentially washed with 600µl ACW1 buffer, 750µl ACW2 buffer and 750µl 100% ethanol. The spin columns were centrifuged at 20,000xg for 3 mins before being dried at 56°C for 10 minutes. Cell-free DNA were eluted in 30µl buffer AVE for O-CHOP samples and 50µl buffer AVE was eluted for STELLAR and ACCEPT samples, corresponding to starting volumes of each library preparation kit.

2.5 DNA Quantification and quality assessment

2.5.1 DNA Quantification using Qubit fluorometry

DNA was quantified using the Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). DNA quantity and fragment length were analysed using the Nanodrop 2000c UV-Vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), Agilent Bioanalyzer (Agilent, Santa Clara, CA, USA) or Agilent 4200 Tapestation System (Agilent, Santa Clara, CA, USA) depending on DNA source and quantity.

DNA Quantification by fluorimetry using Qubit

Fluorometric dye is added to the sample, binding dsDNA by a base intercalation process. The quantity of dye bound is proportional to fluorescence intensity and dsDNA quantity. The concentration of gDNA was measured using the dsDNA broad range assay kit (Q32850, Qiagen, Hilden, Germany). Cell-free DNA samples and DNA libraries were quantified using the dsDNA high sensitivity kit (Q32854, Qiagen, Hilden, Germany).

2.5.2 DNA Quality control using Nanodrop spectrophotometry

For the gDNA samples, nucleic acid concentration and purity was assessed by spectrophotometry using a Nanodrop 2000c UV-Vis spectrophotometer.

Principle of spectrophotometry and measurement

Nucleic acids absorb UV light at a wavelength of 260 nanometres (nm). A photodetector measures the light going through the sample and gives an optical density result. Light absorption is proportional to the nucleic acid concentration. The optical density result is converted to a concentration in nucleic acids using the Beer-Lambert law, which describes the linear relationship between a concentration and absorbance.

First, 1µl elution buffer used for DNA extraction was measured as a blank. The instrument pedestal was cleaned with distilled water between measurements. 1.5µl of sample was pipetted on the pedestal and absorbance measured. The software Nanodrop 2000 software package (v1.6.198) was used to calculate concentrations and A260/A280 ratios.

Spectrophotometry quantifies all nucleic acids in the test sample, not only DNA. Conversely, proteins absorb UV light at a wavelength of 280 nm. A ratio of absorbance at 260 and 280 (A260/A280) is used to determine the degree of protein contamination in the sample. An acceptable A260/A280 ratio for downstream sequencing is between 1.8 and 2.

2.5.3 Assessment of fragment length using automated electrophoresis

Principle of the electrophoresis

Automated electrophoresis enables benchtop gel electrophoretic separation of nucleic acids on a microfluidic chip. Each chip contains an interconnected set of gel-filled microchannels laid out in an interconnected network of fluid reservoirs filled with sieving polymer and fluorescent dye. The electrodes contact the DNA sample when the instrument lid is closed. As the electrical voltage is applied, DNA fragments intercalating with dye are separated according to their mass and detected by a laser-induced detector. Raw data is read from each well. A ladder, made of DNA fragments of known sizes, is included in every run to assess the fragment sizes and confirm that the assay has been run successfully. In addition, two marker fragments are run with each of the samples bracketing the overall sizing range. The "lower" and "upper" markers are internal standards used to align the ladder data with data from the sample wells.

Electrophoresis ensured DNA was a suitable size for library preparation i.e., cell-free DNA recorded a peak at ~170bp and fragmentation of gDNA was successful. It shows DNA libraries were suitable for sequencing and free from additional peaks suggestive of primer dimers or other contaminants.

Fragment assessment using Agilent 2100 Bioanalyser

The Agilent 2100 Bioanalyser and D1000 Chip was used for fragment analysis of DNA and libraries for the LYM1 and O-CHOP cohorts. It contains an electrode cartridge of 16 electrodes, each corresponding to the well of a chip. Samples were prepared according to manufacturer's instructions. The Bioanalyzer 2100 Expert Software filters the data and plots an electrophenogram of each well. The Bioanalyser was superseded by the Aligent Tapestation.

Fragment assessment using Agilent4200 Tapestation

The Tapestation is an automatic instrument measuring DNA fragment and DNA library sizes from 35bp – 1000bp. Samples were prepared for analysis in accordance with manufacturer's instructions and OMDC SOP OM0042. I ensured that the sample was within range (0.1 - 50ng/µl for D5000 and 0.1-4ng/µl for the cell-free DNA assay) range. Samples were diluted if outside of range. Results are displayed using the 2200 TapeStation Controller Software.

The D5000 kit was used for fragmented gDNA, Thruplex-IDT libraries (Chapter 3 and Chapter 5) and IGHV and *TP53* PCR products (Chapter 7). The cell-free DNA kit was used for cell-free DNA analysis.

2.6 Capture hybridisation and sequencing-by-synthesis (SBS)

2.6.1 Introduction to capture hybridisation next generation sequencing (NGS)

Targeted NGS facilitates the interrogation of numerous genomic targets simultaneously. Amplicon based gene panels were designed for recurrent mutations in CLL, providing higher sequencing depth and good correlation with mutations detected by SS (Mollstedt, Mansouri, and Rosenquist 2023). For clinical reporting of NGS results, a threshold of VAF ≥5% is used to distinguish variants from sequencing artefact (Mansouri et al. 2023). In a multi-centre evaluation of amplicon-based panels for CLL somatic mutations, high reproducibility at VAF >5% was reported which declined for variants detected <5% (Sutton et al. 2021).

To detect low frequency variants, capture hybridisation followed by SBS approaches have been developed. Capture-based gene panels can sequence a higher number of targets with more uniform depth, including across regions with high GC content. With appropriate target selection, this technique also offers the potential to detect CNAs and structural variants and has been applied to CLL and MM (Stewart et al. 2021; Navrkalova et al. 2021; Sudha et al. 2022).

Overview of capture hybridisation library preparation

DNA libraries are denatured by heating and hybridised with DNA single stranded oligonucleotide probes specific to the regions of interest. Unbound molecular are washed away and the DNA library, enriched for the targets of interest is sequenced. Post sequencing bioinformatics analysis includes trimming of adaptor sequencings, removal of low-quality reads and mapping of reads to a reference genome.

The introduction of molecular barcoding enhanced the sensitivity of sequencing methods, with the additional costs arising from the higher depth of sequencing coverage required. Integration of unique molecular identifiers (molecular barcode adaptors or MBA) facilitates the distinction between sequencing artefact and genuine low frequency variants (Figure 2.6.1.1.).

This approach is particularly useful for cell-free DNA, where variants of interest are anticipated to occur at low frequency. I used the CAPP-Seq library preparation method (Rossi et al. 2019; Rossi et al. 2020) for the ACCEPT libraries (Chapter 4). I used the Thruplex-IDT library preparation method which encompasses both UMIs and capture hybridisation for CLL (Chapter 3) and RT targeted NGS (Chapter 5).

Figure 2.6.1.1. Molecular barcoding for detection of low frequency variants

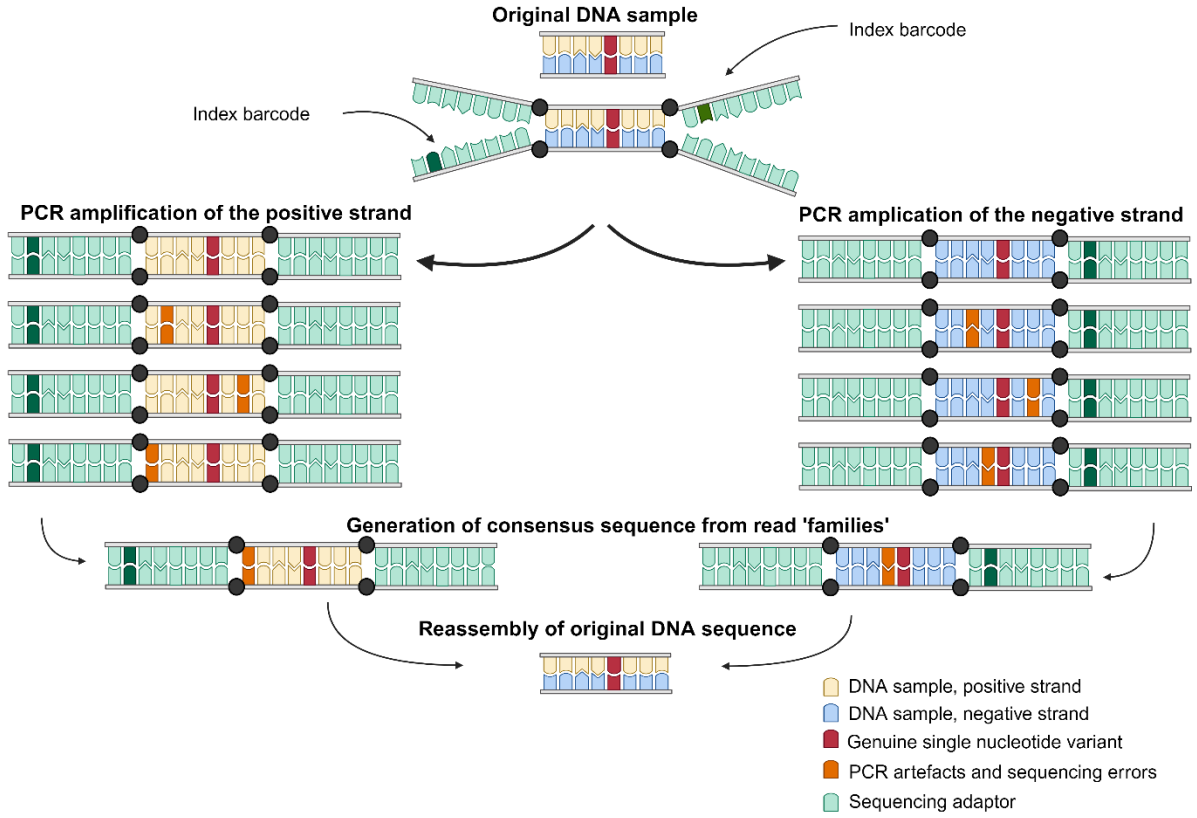


Figure 2.6.1.1: Illustration of molecular barcoding. In molecular barcoding, adaptors (light green) and unique barcoding sequences (dark green) are added to each starting DNA fragment prior to library amplification. The original DNA fragment and the barcoding sequence are amplified. Following sequencing, all reads sharing the same molecular identifier (and arising from the same DNA fragment) are grouped together in the same "family" and compared to generate a consensus sequence. Thus, any true variants (red) are shared by all the reads within the same read family and artefacts (orange) are limited to a minority of reads sharing the same barcode. Figure created with BioRender.com

2.6.2 Panel design and target selection

DLBCL panel

The panel used for the ACCEPT libraries was designed with Dr. Laura Lopez-Pascua and aimed to capture drivers and somatic mutations in the greatest proportion of patients. The targets were selected based on variants previously reported in DLBCL and previously sequenced in lymphoma ctDNA (Scherer et al. 2016; Roschewski et al. 2015; Rossi et al. 2017). In total the genome target for capture comprised 799 regions over 50 genes (>200Kb) known to be involved in lymphomagenesis. Panel targets are summarised in Appendix C.

Richter's Transformation and CLL panel

I designed a Thruplex-IDT customised targeted panel, using the IDT (Integrated DNA Technologies, Inc., Skokie, IL, USA) xGen Lockdown Probe Pools online design tool (https://eu.idtdna.com/site/order/ngs). Panel targets were chosen to reflect variants contributing to the pathogenesis of CLL (Landau et al. 2013; Nadeu et al. 2018; Ljungstrom et al. 2016; De Paoli et al. 2013; Fabbri et al. 2013) or variants amenable to targeted therapy (Woyach et al. 2017; Jones et al. 2017; Xu et al. 2017; Liu et al. 2015; Mohanty et al. 2016). Potential targets were supplemented from the O-CHOP WGS results (Klintman et al. 2021). UTRs were included for targets where non-coding mutations impacted clinical outcomes and/or highlighted regions of kataegis in B-cell regulators (Puente et al. 2015; Burns et al. 2018; Larrayoz et al. 2017). Four genes (*ASXL1, DNMT3A, JAK2* and *TET2*) with variants reported in CHIP (Xie et al. 2014; McKerrell et al. 2015) were included to facilitate distinction between CHIP and tumour variants.

Coverage for each target was predicted by IDT proprietary software, which allowed adjustments to the panel design to attain maximal coverage. Panel targets (56 genes) are included in Appendix D.

2.6.3 Fragmentation of genomic DNA

PMBC germline (ACCEPT cohort), gDNA from PMBC and sorted B-cells (OXPLORED and LYM1 cohorts) and tumour gDNA (CLL and RT cohorts) samples were sheared with a Covaris M220 instrument to achieve 200-bp fragments. Post-fragmentation analysis was performed as described (Section 2.5).

Table 2.6.3.1: Fragmentation Conditions used on Covaris M200 instrument.

	Peak	**Duty Factor**	**Cycles/burst**
Condition	50	20.0	200

2.6.4 Library preparation using CAPP-Seq

ACCEPT libraries were prepared under supervision from Dr Anthony Cutts. ACCEPT germline and cell-free DNA libraries were generated using the KAPA library Preparation Kit (Kapa Biosystem) following the manufacturer's instructions, with minor modifications as described previously (Rossi et al. 2019). 200ng of gDNA was used for each germline library. For cell-free DNA libraries, a minimum input of 32ng was used, and adapter ligation carried out at 16°C for 16 hrs. Genomic DNA libraries underwent 7 PCR cycles and cell-free DNA libraries underwent 9 PCR cycles pre-capture, respectively.

Hybridisation capture was performed using SeqCap EZ Choice Library probes according to manufacturer's instructions, with some modifications. Libraries were multiplexed and underwent 2 rounds of capture at 47° C for 16 hours, where 5 PCR cycles were performed after the first capture and 14 PCR cycles were carried out after the second capture.

2.6.5 Library preparation using Thruplex-IDT capture-based hybridisation.

DNA libraries were prepared using Thruplex® Tag-seq (Takara Bio) kit for the LYM1 and the O-CHOP samples. The Thruplex® Tag-seq HV (Takara Bio) kit was used for STELLAR and OXPLORED samples. Both kits combine unique molecular identifiers, stem-loop adaptors and Illumina indexes.

Thruplex® Tag-seq method

Libraries were prepared from 20ng-75ng of DNA in 10µl according to the manufacturer's instructions with minor modifications. DNA was repaired by combining with 5µl Template Preparation Mastermix and incubating on a Thermocycler for 25 minutes at 22°C, followed by 20 minutes at 55°C. Each 15µl sample was combined with 2.5µl Library Synthesis Buffer and 2.5µl Library Synthesis Enzyme and incubated at 30°C for 40 minutes. During this step, stem-loop adaptors containing MBA were ligated to the 5' end of the DNA, leaving the 3' end open. The 20µl product of the Library Synthesis reaction was combined with 25µl of Library Amplification mix and 5µl Indexing reagent. In this reaction, the 3' ends of the DNA are extended and the Illumina-compatible indexes are incorporated. The library was amplified as described in Table 2.6.5.1 (overleaf).

Table 2.6.5.1: Thruplex® Library Amplification Reaction

	Stage	Temperature	Time	Number of Cycles
Extension & Cleavage	1	72	3 mins	1
	2	85	2 mins	1
Denaturation	3	98	2 mins	1
Addition of Indexes	4	98	20 sec	4
		67	20 sec	
		72	40 sec	
Library Amplification	5	98	20 sec	6-13
		72	50 sec	
Hold	6	4	Hold	1

The number of amplification cycles related to the quantity and source of DNA. I used two additional cycles more than recommended by the manufacturer for FFPE DNA and four additional cycles for cell-free DNA libraries. The unpurified libraries were quantified. If insufficient library was obtained, two further amplification cycles were performed. Library clean-up was performed with Agencourt AMPure XP beads (Beckman-Coulter) and 80% ethanol. The purified library was eluted in 22µl nuclease-free water (NF-H2O) and quantified using real-time PCR (Section 2.6.6).

Changes in Thruplex Tag-Seq HV method

In the Thruplex® Tag-Seq HV library preparation protocol, the library amplification was shortened, and final extension step prolonged (Table 2.6.5.2). Purified libraries were quantified by qPCR and brought forward to capture hybridisation (Section 2.6.7).

Table 2.6.5.2 Thruplex® Tag Seq HV Library Amplification Reaction

	Stage	Temperature	Time	Number of Cycles
Extension & Cleavage	1	72	3 mins	1
	2	85	2 mins	1
Denaturation	3	98	2 mins	1
Library Amplification	4	98 65	20 sec 75 sec	5-11*
Final extension	5	68	5 minutes	
Hold	6	4	Hold	1

**amplification cycles depended on input DNA*

2.6.6 Quantification of DNA libraries by qPCR

Libraries were quantified the real-time PCR Kapa Library Quantification kit (KK4854, KapaBiosystems, Woburn, MA, USA). Quantification was performed after Thruplex library preparation to ensure 500ng of library available for hybridisation and after capture hybridisation to provide accurate library quantification for sequencing.

Library quantification was performed by amplifying the six supplied DNA Standards and diluted library samples in triplicate, using the KAPA SYBR FAST qPCR Master mix and primers targeting the Illumina P5 and P7 oligo sequences. A No Template Control (NTC) was included. The 20 µl reaction mix for standards, test solutions and NTC was prepared as per manufacturer's instructions. Each of the six standards, each test sample (each library at two dilutions) and the NTC was analysed in triplicate.

The sealed and centrifuged qPCR plate was placed in the real-time thermocycler 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). The amplification conditions consisted of an initial denaturation step of 95°C for 5 minutes followed by 35 cycles of denaturation at 95°C for 30 seconds and primer annealing and extension at 60°C for 45 seconds. Data acquisition was performed during this last step of the cycle.

The KAPA Library Quantification Data Analysis Template was used to calculate the mean Cq score for each DNA standard (measured in triplicate) and to generate a standard curve by plotting against the $\log_{10}$ of the picomolar concentration. The standard curve was checked as follows; the mean Cq value between the DNA standards lies between 3.1-3.6; the calculated reaction efficiency was 90-110% and R2 ≥ 0.99. If necessary, one of a triplicate standard was removed from the standard curve calculation to achieve these values. This approach contrasts with the EuroMRD guidelines for IGH rearrangement, which recommends all three values are used except for obvious technical failure (van der Velden et al. 2021). However, qPCR here is used as an internal quality control procedure not to issue a clinical result.

The concentration of the library was calculated against a standard curve. The size-adjusted concentration (in pM) was calculated:

$$\frac{\text{Size of DNA Standard in bp (452bp)}}{\text{Average fragment length of the library in bp}}$$

2.6.7 Hybridisation capture of DNA libraries using xGen Lockdown Probes and Reagents

xGen Lockdown Probes (IDT) are synthetic 5'-biotinylated oligos for target capture applications. Panels are based on 1x tiling of the xGen Lockdown Probes. My custom xGen Lockdown Probes was resuspended in IDTE pH 8.0 to achieve a concentration of 0.75pmol/µL.

Target capture was performed on individually indexed O-CHOP cell-free DNA libraries. For gDNA libraries and STELLAR cell-free DNA libraries, six libraries were combined for capture.

Each 500ng adaptor ligated Thruplex DNA library was combined with 5µg cot-I DNA (Invitrogen) and 2µl of xGen Universal Blocking Oligo TS-Mix. XGen Universal Blockers TS-Mix binds to adaptor sequences to inhibit non-specific binding and improve the on-target reads. The combination was dried using a Speedvac on medium setting for at least 2 hours before being resuspended in a mixture of 1.8µl NF-H2O, 8.5µl NimbleGen 2x hybridisation buffer and 2.7µl Hybridisation Buffer Enhancer. The mixture was incubated at 95°C for 10 mins. 4µl of well-mixed custom xGen Lockdown Probes (RS-panel) was added. Each library pool was hybridised at 65˚C for 16 hours.

Post-hybridisation capture with Thruplex® Tag-Seq libraries

The Dynabeads M-270 Streptavidin beads acclimatised to room temperature for 30 minutes. Hybridised samples were bound to prepared streptavidin beads (Dynabeads M-270) at 65°C for 45 mins. Following removal of the supernatant using a magnet, 200µl 1× Stringent Wash Buffer was added and incubated at 65 °C for 5 min. The supernatant containing unbound DNA was removed and the wash repeated. Sequential washes of the bound DNA were performed using 200µl 1× Wash Buffer I; 200µl 1× Wash Buffer II; 200µl 1× Wash Buffer III. Between each wash, the supernatant was removed using magnetic separation.

The captured DNA on beads was resuspended in 20µl of NF-H2O and then combined with 25µl 2X KAPA HiFi HotStart ReadyMix, 10µM Illumina P5 primer and 10µl Illumina P7 primer for 12 cycles of post-capture amplification.

Table 2.6.7.1: Post capture amplification

	Number of cycles	**Temperature**	**Time**
Polymerase activation	1	98°C	45 sec
Amplification	12		
Denaturation		98°C	15 sec
Annealing		60°C	30 sec
Extension		72°C	30 sec
Final extension	1	72°C	1 minute
Hold	1	4°C	Hold

Post-capture PCR fragments were purified using magnetic bead separation (Agencourt AMPure XP beads, Beckman-Coulter) and resuspended in 22µl of Qiagen EB Buffer. The IDT capture was quantified by KAPA qPCR (Section 2.6.6) and diluted to the appropriate concentration for loading on the flow cell for sequencing.

2.6.8 Library loading and Illumina sequencing-by-synthesis.

Sequencing was performed on MiSeq, HiSeq and NextSeq 550 Illumina sequencers at OMDC. STELLAR, OXPLORED and LYM1 libraries were normalised and loaded for sequencing by myself, under supervision from OMDC staff. Other libraries were prepared for sequencing by OMDC staff.

Overview of Illumina sequencing-by-synthesis (SBS)

SBS consists of hybridising library constructs on a flow cell. Libraries are normalised, pooled and flowed across a solid surface, the 'flow cell.' The template fragments bind to the flow cell surface via adapter sequences added during library preparation. Following this a solid phase bridge amplification PCR process known as 'cluster generation' creates copies of each library template in tight clusters on the flow cell surface.

The sequencing uses universal primers binding to adaptor sequences and modified deoxyribonucleotide triphosphate (dNTPs) containing a terminator, blocking further polymerisation. In this way, only a single base can be added by a polymerase enzyme to each growing nucleotide strand. This facilitates the simultaneous sequencing of large numbers of template molecules spread across the solid surface of the flow cell (Figure 2.6.8.1).

The terminator contains a fluorescent label, detectable by a camera within the instrument. Following addition of dNTPs to the template, the images are recorded. After data acquisition, the terminators are removed, and a new cycle is initiated.

Paired-end sequencing is initially performed in one direction for a set number of cycles before the target is inverted and sequencing performed in the contralateral direction for an equal number of cycles. Sequencing data includes the index sequences, allowing demultiplexing and identification of individual patient results.

Figure 2.6.8.1 Schematic of sequencing-by-synthesis (SBS)

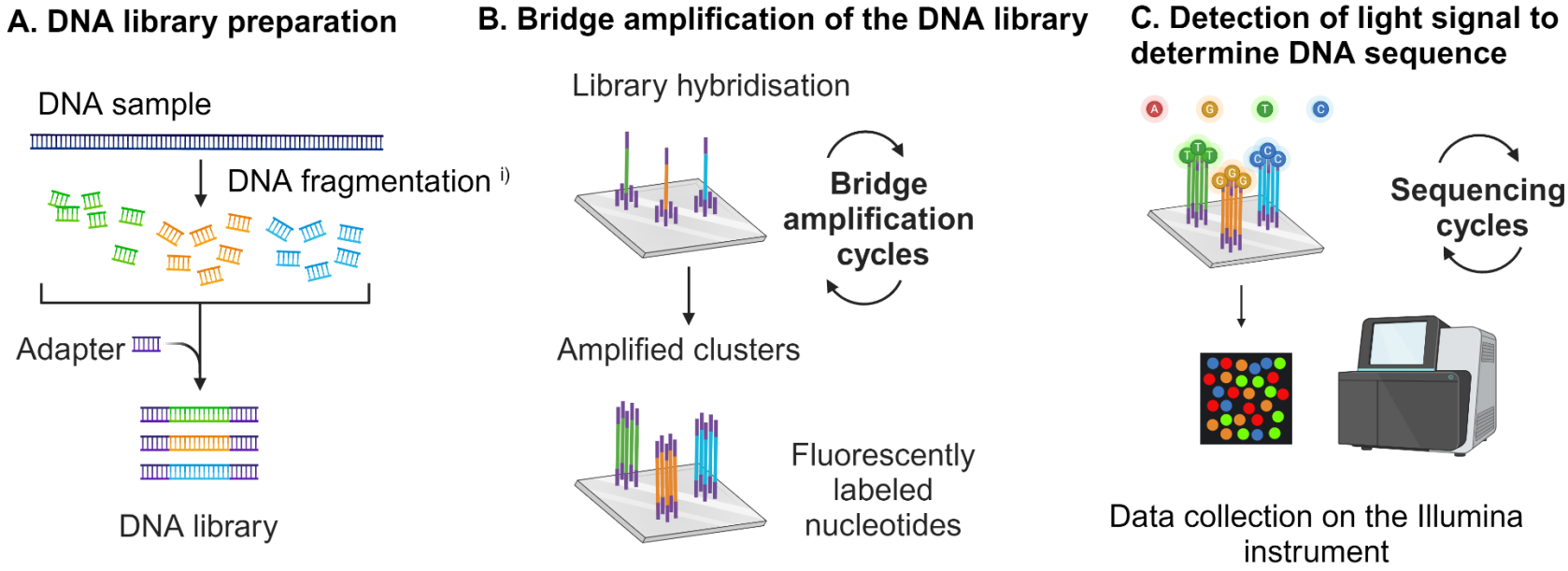


Figure 2.6.8.1: Visualisation of sequencing-by-synthesis (SBS). First, a DNA library is prepared by fragmenting DNA, then adding adapters to the fragments (A). When working with cell-free DNA the fragmentation step is not required (i). The DNA library with adapters sequences binds to the solid surface of a flow cell via complementary sequences in the adaptors (B). The free end bends to attach to a second oligo forming a bridge. Polymerase generates the complementary strand using ddNTPs with a fluorescent tag. Both strands release, straighten and forms a new bridge (bridge amplification). Hundreds of strands undergo bridge amplification simultaneously, generating a molecular cluster to allow the light signal from the fluorescent tag to be detected by the instrument (C). The colour emitted corresponds to the base incorporated. Afterwards, the fluorescent tag is cleaved, washed away and a new cycle begins. Figure created with BioRender.com

Principle of PhiX internal control

The PhiX library is a commercial control derived a from a bacteriophage genome. It has a mean size of 500bp and consists of 45:55 GC:AT bases. It provides a control for run quality monitoring and for low diversity libraries, PhiX provides additional fluorescent signals to improve overall run quality. The 20pM PhiX, added at 1-5% provides an internal control of the sequencing process. The proportion of PhiX sequences identified following sequencing, as well as the accuracy of sequencing, provides information on sequencing data quality. In low diversity libraries, the PhiX control helps counterbalance the limited of sequence diversity.
PhiX Control v3 (10nM) was prepared and diluted to 20pM PhiX as per manufacturer's instructions. For my cell-free DNA libraries, I used PhiX 5%.

Denaturing of library and PhiX prior to sequencing

5µl of pooled 4nM library is combined with 5µl of 0.2M NaOH and incubated at room temperature for five minutes. After incubation, 990µl of Illumina HT11 buffer is added to make a 20pM library. To prepare the 20pM library for loading, 315µl of 20pM library is combined with 12µl PhiX control (10.5pM) and 273µl HT1 buffer. This gives 10.5pM library in 600µl for loading into the reagents' cartridge.

Factors guiding choice of Illumina sequencing instrument

The choice of a specific instrument for SBS necessitates a balance between the size of the target panel, depth of sequencing required, number of samples per run and the cost of sequencing. The target depth of sequencing depends on the source of DNA (germline, tumour or cell-free DNA) and the presence or absence of UMIs. For example, the custom CLL/RT panel included UMIs requiring a greater depth of sequencing to achieve the target *deduplicated* depth.

The HiSeq 2500 at OMDC was decommissioned in 2019 and replaced by the NextSeq 550.

Table 2.6.8.1 Comparison of Illumina SBS instrument and flow cells used in the thesis

Instrument	MiSeq	HiSeq 2500	NextSeq 550	NextSeq 550
Flow cell	V2	High output v4	Mid Output v2	High Output v2
Maximum Output*	5.1 Gb	1 TB	19.5Gb	60Gb
Read length	2x150bp	2x125bp	2x150bp	2 x150bp
Libraries	LYM1 gDNA libraries O-CHOP tumour and germline	O-CHOP cell-free DNA libraries	STELLAR tumour and germline libraries ACCEPT tumour and germline libraries	STELLAR cell-free DNA libraries
*specifications provided by Illumina refer to performance using PhiX control library				

Sequencing-by-synthesis (SBS) using the Illumina MiSeq

Libraries from gDNA (O-CHOP and MBL/CLL LYM1 study) underwent sequencing using the MiSeq Reagent Kit v2 300-cycle kit (MS-102-2002, Illumina Inc., San Diego, CA, USA). Paired end sequencing was completed as described in the following: 150 cycles for DNA insert, 8 cycles for index i7, 8 cycles for index i5, and 150 cycles for DNA insert on the MiSeq instrument (SY-410-1003, Illumina Inc., San Diego, CA, USA).

SBS using the Illumina HiSeq

Sequencing of the Thruplex-IDT O-CHOP cell-free DNA libraries were performed on a HiSeq 2500 (Illumina Inc., San Diego, CA, USA) following a High Output v4 mode, 2x125 cycles (paired end, indexed). Cluster generation was performed on a cBot system (Illumina Inc, San Diego, CA, USA). Following the pre-run wash cycle, the reagent plate, flow cell and samples were loaded into the instrument. The reagents were loaded into the instrument. After the flow check was conducted, the sequencing flow cell was inserted.

SBS using the Illumina NextSeq

ACCEPT, STELLAR and OXPLORED libraries were sequenced on the NextSeq 550 (Illumina, San Diego, CA, USA) using the mid output kit v2 flow cell for gDNA and the high output kit v2 flow cell for cell-free DNA. DNA libraries were denatured using fresh 0.2N NaOH (Table 2.6.8.1) and incubated for five minutes at room temperature.

Table 2.6.8.2: Volumes for library preparation for NextSeq 550

Library concentration	**Library volume**	**0.2 N NaOH**	**200mM Tris-HCl, pH 7**	**Chilled HT1**
4nM	5µl	5µl	5µl	985µl
2nM	10µl	10µl	10µl	970µl
1nM	20µl	20µl	20µl	940µl

After incubation, the mixture was combined with the appropriate volume of 200mM Tris-HCl, pH 7 and vortexed. The appropriate volume of HT1 was added to the denatured libraries and mixed by vortexing to make a 20pM library.

For the NextSeq Mid Output kits, the 20pM library was diluted to using 97µl denatured library and 1203µl chilled HT1 to make 1.3ml at 1.5pM library. The PhiX control was similarly prepared to make 1.3ml at 1.5pM PhiX control. 1µl of denatured 20pM PhiX was combined with the 1299µl denatured 20pM gDNA libraries was combined for loading into the reagent cartridge.

The NextSeq 550 reagent cartridge was thawed overnight in a fridge. The NextSeq flow cell was inspected and cleaned with lens tissue as required, before loading onto the instrument. After tapping to remove air bubbles and inverting slowly, the reservoir well labelled ‘load library here’ was pierced with a 1ml pipette tip. The full 1.3ml of denatured library was loaded into the reagent cartridge and the cartridge was placed into position on the instrument. The sample sheet and indexes were loaded, and the run started. Sequencing was performed using NextSeq 550 v2.5, 2x151 cycles (paired end, indexed).

2.6.9 Post sequencing analysis of the capture hybridisation libraries

The Illumina BaseSpace app "FASTQ Generation" (Version: 1.0.0) generated demultiplexed and paired FASTQ files using bcl2fastq2.

Data analysis was performed using a custom bioinformatics pipelines developed at the OMDC for amazon cloud software (https://signin.aws.amazon.com/). Optimisation of the pipeline was done by the bioinformatics core team at OMDC. The coverage of each target region is assessed on a per sample and per run basis using Covermi software (https://github.com/eawilson/CoverMi). Read trimming was done using Udini v1.0.0 (https://github.com/eawilson/udini). Read alignment was performed using Burrows-Wheeler aligner BWA-mem v0.7.17 against the hg19 human reference genome (reference GRCh37) (Li and Durbin 2009). Deduplication and error correction were performed by elduderino v1.2.0 (https://github.com/eawilson/elduderino).

For the DLBCL panel used in Chapter 4, reads with the same start point were grouped together. If > 60% of the bases in reads with the same starting point matched, the reads were grouped together and deduplicated into a consensus read. Deduplicated consensus reads were grouped into read 'families'. For the 56-genes Thruplex-IDT panel used for CLL/RT (Chapter 3 and Chapter 5), the reads with identical UMI were collapsed into read 'families'. A minimum family size of two was used for deduplicated reads. In both cases, variants represented by two or more read families were retained. The choice of a family size of two represented a balance between detecting low frequency variants and minimising false positive calls.

Variant calling used Varscan2 v2.4.2 (Koboldt et al. 2012). Variant annotation was done using Ensemble Variant Effect Predictor (VEP v104) (McLaren et al. 2016) and Annovar (Wang, Li, and Hakonarson 2010).

Figure 2.6.9.1 Schematic of the OMDC bioinformatics pipeline

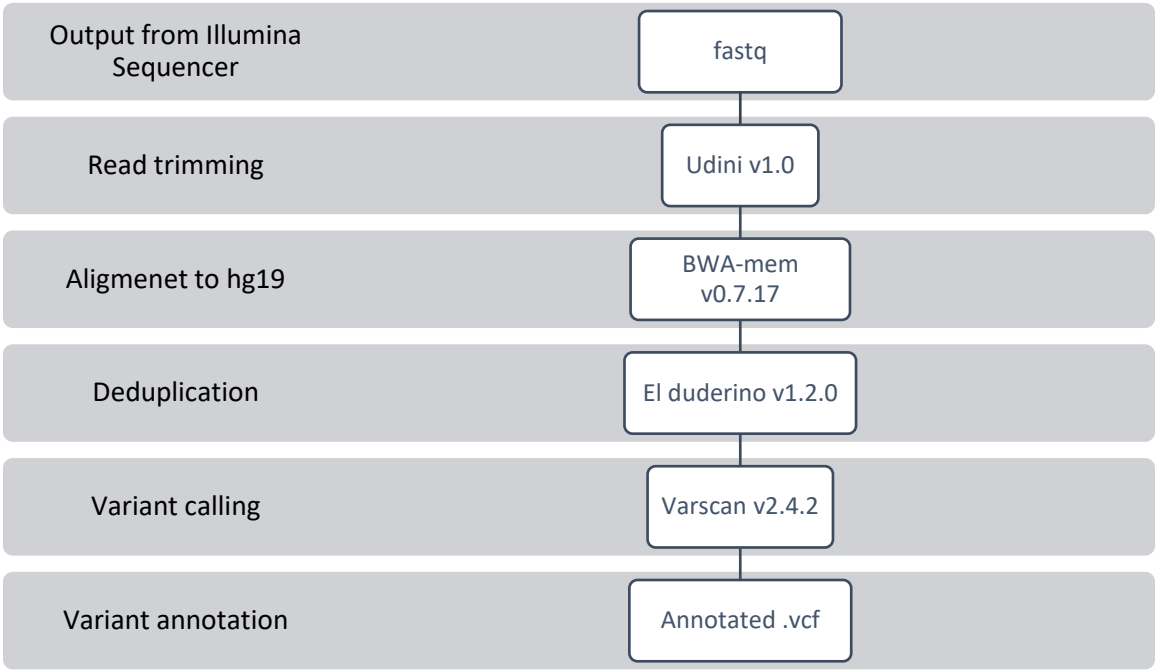


Figure 2.6.9.1 Visualisation of the steps taken in the post-sequencing analysis

2.6.10 Variant filtering

Using the annotated varscan files, I filtered variants aiming to detection low frequency variants while minimising false positives and sequencing artefacts (Section 2.6.9). Variants fulfilling the following criteria were retained; Quality Score > 30; non-synonymous variants, Fischer Strand bias score < 60 and Minor Allele Frequency < 0.01 Further filtering differed depending on DNA source (gDNA vs cell-free DNA) and the availability of a germline sample.

In Chapter 3, further variant filtering included a False Discovery test at a < 0.01 threshold, retaining variants with COSMIC ID previously reported in haematological cancer and visual inspection using Integrated Genome Viewer (IGV). For the RT tumour samples (Chapter 5), the FDR test was inadvertently removing pathogenic *TP53* mutations at higher AF so was omitted.

The CAPP-Seq library (Chapter 4), the false discovery test threshold was < 0.001. These libraries lacked UMIs. For each patient, I aimed to filter out germline variants and recurrent sequencing artefacts. Variants occurring in the control samples and recurrent variants at AF < 0.3 in the germline samples were considered artefactual and removed.

Variant filtering steps were chosen by me with guidance from the OMDC bioinformatics team (Dr Dimitris Vavoulis, Dr Kate Ridout and Dr Laura Lopez-Pascua).

2.6.11 Quantification of ctDNA burden (HGE/ml)

Cell-free DNA concentration was measured in haploid genomic equivalents per millilitre of plasma (HGE/ml) assuming one haploid genome weighs 3.3 picograms. Total cell-free DNA in haploid human genome equivalents per millilitre was calculated as follows:

$$\text{Total cell-free DNA} = \frac{\text{cell-free DNA (ng/µl) x cell-free DNA volume (µl)}}{\text{Plasma volume (µl)}}$$

$$\text{Total cell-free DNA in HGE/ml} = \text{Total cell-free DNA} \times 1000/3.3\text{pg}$$

Plasma ctDNA burden was calculated by multiplying the cell-free DNA concentration in plasma (pg/ml) by the median allele frequency of the patient-specific somatic variants identified, divided by one haploid genome and reported as a 10-base logarithm.

The median AF was chosen as it is less sensitive to outlying values. The RT cohorts were anticipated to have numerous aneuploidies leading to artefactual AF values e.g., an unexpectedly high *TP53* VAF in a patient with deletion of the contralateral *TP53* allele. The median ctDNA value was calculated from the median VAF of the detected variants and the HGE on a per sample basis. For visualisation, any samples with no detectable ctDNA variant were assigned a HGEctDNA value of 1, thus a $\log_{10}$(HGEctDNA) of zero.

The median pre-treatment ctDNA burden ($\log_{10}$HGE/ml) was calculated for each cohort. Patients were subdivided into 'high' or 'low' ctDNA subgroups based on whether the sample ctDNA burden was greater or lesser than the median value.

2.7 Nanopore sequencing of gDNA and cell-free DNA samples

2.7.1 IGHV and *TP53* primer design and PCR amplification

IGHV primer design

Primers targeting the full-length IGHV locus were designed with the IGH leader (forward) and consensus JH (reverse) regions and modified to include ONT barcoding tails (Table 2.7.2.1) (Campbell et al. 1992).

Table 2.7.1.1: Primer Sequences to amplify the IGHV locus

PRIMER ID	***Primer Sequence (5'-3')***
VH1/7	TTTCTGTTGGTGCTGATATTGCATGGACTGGACCTGGAGG
VH2	TTTCTGTTGGTGCTGATATTGCCACRCTCCTGCTGCTGACCA
VH3A	TTTCTGTTGGTGCTGATATTGCGCTGGGTTTTCCTTGTTGC
VH3B	TTTCTGTTGGTGCTGATATTGCATGGAGTTKGGRCTGAGCTG
VH4	TTTCTGTTGGTGCTGATATTGCGCTCCCAGATGGGGTCCTG
VH5	TTTCTGTTGGTGCTGATATTGCCTCCTCCTGGCTGTTCTCC
VH6	TTTCTGTTGGTGCTGATATTGCCTGTCTCCTTCCTCATCTTCC
JH	ACTTGCCTGTCGCTCTATCTTCCTTACCTGAGGAGACGGTGACC *Primer design by Dr Adam Burns, postdoctoral researcher*

PCR amplification of IGHV locus

All eight IGHV primers were pooled to a final concentration of 2µM. The final reaction contained 2.5µl 10X IGHV primer pool, 12.5µl 2x Qiagen Multiplex PCR Master Mix (Qiagen, Hilden, Germany), 2.5µl 5x Qiagen Q-solution (Qiagen, Hilden, Germany), 5.5µl molecular grade water and 2µl gDNA. Cycling conditions were: 95°C for 15 minutes followed by 30 cycles of 94°C for 30 seconds, 62°C for 90 seconds and 72°C for 90 seconds, with a final extension of 72°C for 10 minutes.

IGHV PCR products were confirmed to measure ~ 600bp using automated electrophoresis (Figure 2.7.1.1). following clean-up with AMPure XP beads, 100-200fmol (37-74ng) of IGHV PCR product was used as input for the ONT barcoding PCR protocol (Section 2.7.2).

Figure 2.7.1.1: Tapestation electrophenogram of BM2462 (66243) IGHV library, 598bp

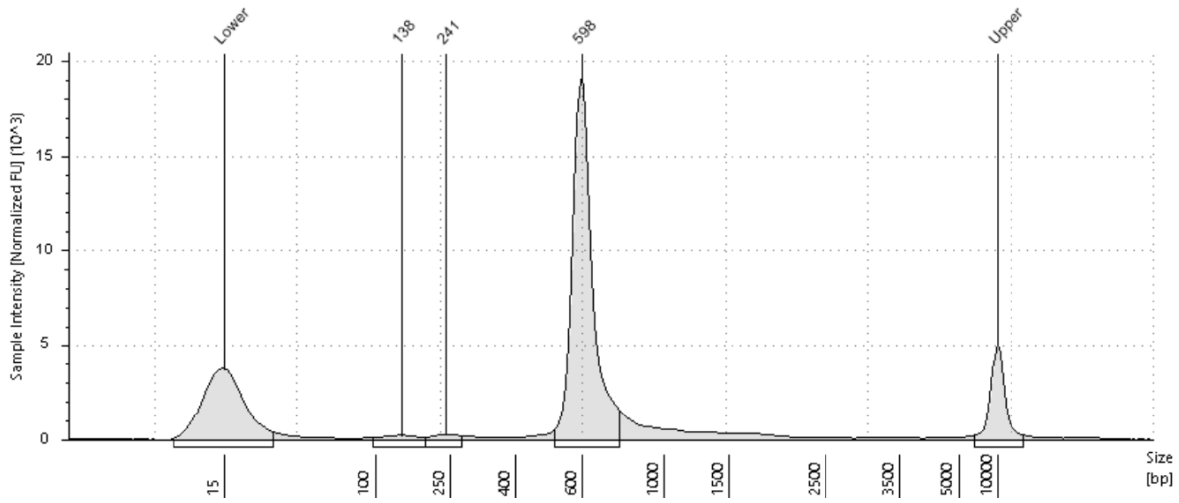


Figure 2.7.2.1: Automatic electrophoresis showing fluorescence intensity (y-axis) and size in bp (x-axis). The upper and lower markers correspond with the ladder. The peak at 598bp corresponds to the IGHV PCR products.

TP53 exonic primer design

Primers were designed for all exons of *TP53* in five amplicons, with each primer including the ONT barcoding tail sequence (Table 2.7.1.2).

Table 2.7.1.2: Primer Sequences to amplify *TP53*

PRIMER ID	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
TP53_EXON1	TTTCTGTTGGTGCTGATATTGCTGAATAGTGAGCTGCTCCGG	ACTTGCCTGTCGCTCTATCTTCGAATCCTGACTCTGCACCCT
TP53_EXON2-5	TTTCTGTTGGTGCTGATATTGCTCAGTGAGGAATCAGAGGCC	ACTTGCCTGTCGCTCTATCTTCAGTGTCTCATGCTGGATCCC
TP53_EXON5-6	TTTCTGTTGGTGCTGATATTGCGCCACTGACAACCACCCTTA	ACTTGCCTGTCGCTCTATCTTCATCCATCCCATCACACCCTC
TP53_EXON7-9	TTTCTGTTGGTGCTGATATTGCCCAGCTACTCAGTGGGCT	ACTTGCCTGTCGCTCTATCTTCTTGGGCCTGTGTTATCTCCT
TP53_EXON10_11	TTTCTGTTGGTGCTGATATTGCCTCCCTTGTTGTCATGCAGC	ACTTGCCTGTCGCTCTATCTTCGAGGTGGAGGTGGGTAGATC
	Primer design by Dr Adam Burns, postdoctoral researcher	

PCR amplification of TP53

The final *TP53* reaction mix contained 0.5µM forward primer, 0.5µM reverse primer, 12.5µl 2x Qiagen Multiplex Master Mix, 5µl 5x Q-solution, 3.5µl NF-H2O and 2µl gDNA. For the exon 7-9 and 10-11 reactions, the Q-solution was replaced with NF-H2O. Cycling conditions for all *TP53* amplicons were: 95°C for 15 minutes followed by 30 cycles of 94°C for 30 seconds, 62°C for 90 seconds and 72°C for 3 minutes, with a final extension of 72°C for 10 minutes. Following AMPure XP bead clean-up and Qubit quantification, each amplicon was normalised to 2nM and pooled on a per-sample basis. These pools were used as input for the barcoding PCR reaction.

2.7.2. Preparation of amplicon libraries

Libraries from the advanced CLL cohort were prepared using SQK-LSK108. MBL/early CLL libraries were prepared using SQK-LSK109 (SQK-LSK108 & SQK-LSK109, ONT, Oxford, UK).

Amplicon barcoding

Initially, 100-200fmol PCR product was combined with 2µl ONT barcode sequence, 50µl 2x LongAmp MasterMix (New England Biolabs, Ipswich, MA, USA) to a final volume of 100µl with NF-H2O. Cycling conditions were: 95°C for 3 minutes, 14 cycles of 95°C for 15 seconds, 62°C for 15 seconds and 65°C for 90 seconds, with a final extension time of 65°C for 10 minutes. Following AMPure XP bead clean-up, samples were quantified. Up to 12 barcoded IGHV amplicons were pooled for a final concentration of 1µg.

DNA repair and end-prep

Barcoded DNA amplicons were combined with 1µl DNA Control Strand (DNA CS, a 3.6kb standard amplicon), 3.5µl NEBNext FFPE DNA Repair Buffer, 2µl NEBNext FFPE DNA Repair Mix, 3.5µl Ultra II End-prep reaction buffer and 3µl Ultra II End-prep enzyme mix before incubating on a thermal cycler for 20°C for five minutes and 65°C for five minutes. Following AMPure XP bead clean-up, 60µl of repaired and end-prepped DNA amplicons was brought forward.

Adaptor ligation and clean-up

For adaptor ligation, the barcoded DNA amplicons were combined with 25µl Ligation Buffer, 10µl NEBNext Quick T4 DNA Ligase and 5µl AMX adaptor mix and incubated at room temperature for ten minutes. The ONT Short Fragment Buffer (SFB) was used with AMPure XP beads in the clean-up after adapter ligation. The final library was eluted in 15µl of ONT Elution Buffer and quantified using a Qubit fluorometer.

Priming the flow cell and loading the library for sequencing (all libraries)

The MinION Mk1B lid was opened and the flow cell positioned under the clip. After opening priming port cover and checking for air bubbles, the tip of a P1000 pipette set to 200µl was inserted into the priming port and turned slowly until a small volume of buffer entered the pipette tip or the dial shows 230µl, whichever occurred first. The flow cell was primed by combining 30µl Flush Tether with Flush Buffer, loading 800µl of the flush solution into the priming port and incubating at room temperature for five minutes. With the sample port open, the remaining 200µl was loading into the priming port of the flow cell.

Table 2.7.2.1: Preparation of the LSK-109 barcoded amplicon library for loading

Reagent	*Volume*
Sequencing Buffer (SQB)	37.5µl
Loading Beads (LB)	25.5µl
DNA library	12µl
Total	75µl

5-50fmol (approximately 2-20ng of 600bp IGHV library) in 12µl of the final library was mixed with 37.5µl Sequencing Buffer (SQB), 25.5µl Loading Beads (LLB) and loaded into the primed flow cell (Table 2.7.3.1). Where the calculated library concentration was > 50fmol, the library was diluted in NF-H2O to attain an appropriate concentration.

2.7.3 Whole Genome Library Preparation using the SQK-RAD-004 rapid sequencing kit (genomic DNA)

Libraries for shallow WGS were prepared from 400ng gDNA using the Rapid Sequencing protocol (SQK-RAD004, ONT, Oxford, UK).

Rapid whole genome library (RAD004) preparation

To tagment the DNA, 400ng gDNA in 7.5µl NF-H2O was combined with 2.5µl fragmentation mix (FRA) and incubated at 30°C for one minute and 80°C for one minute. Nanopore adaptors are added using 1µl rapid adaptor (RAP) before incubating at room temperature for five minutes.

Priming the flow cell and loading the RAD004 gDNA library

The flow cell was primed as described above. RAD004 library loading preparation is shown in Table 2.7.3.1.

Table 2.7.3.1 Preparation of the RAD004 library for loading

Reagent	*Volume*
Sequencing Buffer (SQB)	34µl
Loading Beads (LB)	25.5µl
NF-H2O	4.5µl
DNA library	11µl
Total	75µl

2.7.4 Preparing cell-free DNA libraries for shallow WGS using nanopore

Nanopore cell-free DNA library preparation using LSK-109

I prepared the fragmented gDNA and cell-free DNA libraries using the Native Barcoding Amplicons (EXP-NBA-104) and SQK-LSK109 protocols modified as follows:

- Greater than 200fmol cell-free DNA was used as input
- The bead/sample ratio was increased to 1.8x for all clean up steps optimising retention of short fragments.
- An additional 2.5x AMPure bead clean up step was included
- All available library was loaded into the MinION flow cell for sequencing

As the O-CHOP cell-free DNA samples were collected in EDTA, the total DNA quantity combines genuine cell-free DNA and longer DNA fragments leached from leukocytes. Electrophenograms reports 23-60% of the total DNA as cell-free DNA. The mean input DNA was 580.5 pmol, more than the recommended 200fmol input.

Four barcoded O-CHOP cell-free DNA libraries (14-43fmol) per flow cell were sequenced for 48 hours. For subsequent cell-free DNA libraries, one cell-free DNA library per flow cell was used, excess library was loaded and sequenced for 70 hours.

2.7.5 Sequencing on the MinION

Nanopore flow cell check

After loading, the flow cell was checked for viable nanopores. The flow cells had a minimum of 800 viable pores for WGS experiments. For sequencing of amplicon libraries, flow cells with less than 800 pores were used. A new mux scan was required prior to sequencing a library, to examine the quality of each channel and select the best pore for sequencing. A mux scan is repeated at intervals throughout the sequencing run. The default setting of bias voltage was –180mV was used for new flow cells and –190mV for previously used flow cells. The Q-score was set at 7 throughout.

Amplicon libraries for IGHV and TP53

For the advanced CLL samples, amplicon libraries were sequenced for 48 hours using a FLO-MIN107 flow cell on a MinION Mk1b instrument. For other samples, amplicon libraries were sequenced using a FLO-MIN109 R9.4 flow cell for 4-70 hours.

Combined IGHV, TP53 and sWGS assays in a single run

For patient CLL374, IGHV and *TP53* amplicon and sWGS libraries were sequenced synchronously. IGHV and *TP53* libraries were assigned individual barcodes and pooled following the barcoding PCR protocol. A WGS library was prepared using RAD004. The combination library was prepared for sequencing: 34µl sequencing buffer, 25.5µl loading beads, 2µl combined IGHV/*TP53* library, 2.5µl NF-H2O and 11µl WGS library. Libraries were sequenced for 48 hours on R9.4 flow cell.

Basecalling using MinKNOW software

Basecalling was performed live (during the sequencing experiment) on the local computer using the Guppy basecaller in 'fast basecalling' mode. For barcoded libraries, reads were split by barcode using MinKNOW software, sorting fastq and fast5 files into barcode-specific folders.

2.7.6 Post nanopore sequencing analysis

Data acquisition and basecalling using MinKNOW

Albacore v2.1.10 (Oxford Nanopore Technologies, Oxford, UK) was used for basecalling and de-multiplexing of raw sequencing files, with Porechop (v0.2.3, https://github.com/rrwick/Porechop) used for adaptor removal. Whole-genome datasets were aligned to hg19 with minimap2 (Li 2018)(https://github.com/lh3/minimap2/) and sorted and indexed with samtools (v1.19) (Danecek et al. 2021).

Developing a method to determining IGVH status

Two post-sequencing data analysis strategies were tested.

Method A*: Analysis using Canu to generate a consensus IGH sequence*

Each fastq file was down-sampled to 10,000 reads using Seqtk (https://github.com/lh3/seqtk) to account for variation in coverage. A preliminary consensus sequence was generated using

Canu-2.1.1 (Koren et al. 2017) which specialises in assembling a consensus sequence from data with relatively high per-base error rate. It performs correction, read trimming and assembly before generating a consensus contigs. Read correction generates consensus sequences from the overlapping sections of the raw nanopore reads. Read trimming retains regions of high-quality sequence. Sets of consistent overlaps are used to generate a consensus sequence. For the IGHV reads, canu was run with genome size 1000bp and minimum overlap of 500bp.

The canu consensus contig was used a reference for alignment of the fastq reads with Minimap2. The resulting SAM file was converted into BAM format, sorted and indexed with samtools (v1.19, https://github.com/samtools/). A final consensus sequence was generated with Nanopolish (v0.11.0, https://github.com/its/nanopolish). The international ImMunoGenTics (IMGT) database (https://www.imgt.org/IMGT_vquest/input) was used to determine the level of identity with respect to germline (Giudicelli, Brochet, and Lefranc 2011; Brochet, Lefranc, and Giudicelli 2008). The IMGT database contains curated and annotated Immunoglobin, T-cell receptor and Major Histocompatibility Complex (MHC) sequences from a range of vertebrate species. Each IGHV sequence was uploaded to IMGT and analysed using the advanced functionalities to determine CLL subsets.

Results and limitations of Method A are reported in Section 6.3.1. Principally, this method reported IGHV rearrangements from healthy controls.

Method B: *Determination of VH family from nanopore reads using IMGT/HighV-QUEST*

Secondly, I down-sampled each nanopore fastq file to 10,000 reads and made a fasta file using Seqtk. I uploaded the fasta file IMGT/HighV-QUEST to determine the VH family, presence of indels and germline homology (Li et al. 2013). The IMGT/HighV-QUEST is optimised for high-throughput sequencing data and output consists of 11 .csv text files per sample. The 'Summary' file contains the alignment score and identity percentage with the closest V and J genes, the CDR-IMGT lengths and a description of any insertions or deletions. I filtered the summary output file retaining mapped reads and VH families accounting for > 10% of mapped reads. VH families with > 1000 supporting reads were reported.

Results and limitations of Method B are reported in Section 6.3.1. Principally, this method reported all IGHV rearrangements as 80-90% germline homology.

Combined method to determine IGHV family and germline homology

Using IMGT/HighV-QUEST (Method B) to determine VH family followed by canu (Method A) to determine germline homology correctly called control and test samples, reported in Section 6.3.1 and was used for the clinical samples.

Post-sequencing analysis of TP53 amplicons

For SNV detection in *TP53*, reads were aligned to hg19 with Minimap2 (Li 2018). Samtools (v1.19, https://github.com/samtools/) compiled a pileup file using for each sample, which was input for VarScan v2.4.2 for variant calling (Koboldt et al. 2012). The following filtering criteria were used: minimum coverage 400x, minimum VAF 0.1, minimum average quality 20. All remaining variants were annotated with Annovar (Wang, Li, and Hakonarson 2010). To identify

indels, NGMLR (https://github.com/philres/ngmlr) with ONT-specific options was used to align our reads to hg19. Indels were called using Sniffles (Sedlazeck et al. 2018), retaining variants identified as 'precise'.

Copy number analysis using QDNASeq package

Coverage was calculated using the samtools (version 1.19) depth function. For copy-number analysis, I used the QDNAseq package (v1.24.0) (Scheinin et al. 2014). Briefly, aligned reads for each shallow whole-genome dataset are grouped into non-overlapping 100kb windows. The number of reads per window are counted, adjusted for GC content and mappability, and filtered to remove low quality regions, allowing regions of CNA to be called.

For gDNA experiments, the default cutoffs=log2(c(deletion = 0.5, loss = 1.5, gain = 2.5, amplification = 10)/2) for copy number calls were used. Modifications to the QDNAseq parameters for cell-free DNA are reported in Chapter 7.

2.8 Data analysis and statistical methods

2.8.1 General considerations

Median, inter-quartile range / range or means and standard deviations were reported for continuous variables. Frequencies and/or percentages were given for binary outcomes.

For most comparisons, non-parametric tests (Wilcoxon rank sum test, Kruskal-Wallis test) were employed. The use of parametric (Pearson, r) and non-parametric (Spearman, rho) tests for correlation are indicated in the figures and text. Adjustment for multiple testing was performed with the false discovery rate (FDR) method. For all statistical analysis, the null hypothesis was rejected when the p-value was less than 0.05.

A confusion matrix was used to in Chapter 4 visualise the relationship between eradication of ctDNA and clinical outcomes. A confusion matrix is a method to show the predicted versus the actual classification; in this case, the eradication of ctDNA at early post-treatment timepoints and the response at the end of treatment. Positive predictive value (PPV) was calculated from the true positive rate (i.e. the number of patients who eradicated ctDNA and responded) divided by the sum of the true positives and the false positives. The negative predictive value was calculated from the negative predictive rate (i.e., the number of patients with persistent ctDNA lacking response) divided by the sum of the true negatives plus the false negatives. Sensitivity was calculated as the correctly predicted outcomes based on a positive test result (true positives) divided all positive tests. Specificity was calculated as the correctly predicted outcomes based on a negative test result (true negatives) divided by all negative tests (Table 2.7.2.1).

Co-occurrence of CLL biomarkers and mutations were calculated using the R package ‘cooccur’ and visualised as a heatmap (Griffith 2016). The package calculated the observed and expected frequencies (based on each variable being independent) of co-occurrence between each biomarker. Probabilities of co-occurrence that are higher or lower than obtained by chance were reported.

Table 2.8.2.1 Sample confusion matrix

		Predicted outcome	
Actual outcome	Total population = P+ N	Positive (PP)	Negative (NP)
	Test positive (P)	True positive (TP)	False negative (TN)
	Test negative (N)	False positive (FP)	True negative (TN)
Sensitivity = TP/P; Specificity = TN/N; Positive Predictive Value = TP/ (TP+FP); Negative Predictive Value = NP/ (NP + FN)			

2.8.2 Data visualisation

Variant filtering and data visualisation were performed in R using packages tidyverse, ggpubr, gridExtra. Dot plots, box plots and bar plots were drawn with R package 'ggplot2'. Forest plot visualisations of the Cox regression analyses were drawn with R package 'forestplot.'

2.8.3 Survival analysis

Treatment Free Survival (TFS) was calculated from the time of sampling (LYM1 cohort) or date of enrolment (OXPLORED cohort) until the date of first treatment, respectively.

Progression Free Survival (PFS) for the trial cohorts was defined as the time from randomisation to the date of progression or death. Progression was defined using modified Cheson criteria (Cheson et al. 2014). Participants withdrawn from study, lost to follow-up or in going remission will be censored at the date of last follow-up. Overall Survival (OS) was calculated from the date of sampling (LYM1 cohort) or date of enrolment (all trial cohorts) to date of death from any cause.

TFS, PFS and OS distributions were plotted using Kaplan–Meier estimates and visualised using the R packages 'survminer' (v0.4.9) and 'survival' (v3.4.0). Univariate analysis of survival (PFS and OS) by subgroups was performed using proportional hazards models (Cox proportional-hazards model). Cox regression provided in the R package 'survminer' was used for univariate analyses.

Chapter 3: Sequencing based detection of molecular features that inform prognosis in Monoclonal B-cell Lymphocytosis and early Chronic Lymphocytic Leukaemia

3.1 Introduction

3.1.1 Prognostic models of MBL and early CLL

The majority of chronic lymphocytic leukaemia (CLL) patients in Europe are diagnosed with early asymptomatic disease (da Cunha-Bang et al. 2016; Howell et al. 2013). Guidelines recommend active surveillance, deferring treatment until the patient develops cytopenias, bulky lymphadenopathy, fevers, weight loss or drenching sweats (Hallek et al. 2018; Schuh et al. 2018).

Recent CLL prognostic scoring systems incorporate clinical and molecular features to refine risk stratification (Table 3.1.2.1). Systems combining biological and genomic data to predict progression in newly diagnosed patients are based on biomarker performance as independent covariates in multi-variable regression models. Many scores were developed from clinical trial populations, who fulfil criteria for intervention (Molica et al. 2017; Pflug et al. 2014; International 2016). These models have limited discrimination capacity when applied to early disease. Predictive power ranged 60%-68%, with tools incorporating IGHV and *TP53* status demonstrating greater utility. Only three scoring systems have undergone external validation, the CLL-International Prognostic Index (CLL-IPI), the Barcelona-Brno and the MDACC 2007 scores (Kreuzberger et al. 2020). On an individual patient level, the current prognostic scoring systems will misclassify up to one third of early CLL patients.

Monoclonal B-cell lymphocytosis (MBL) and low-risk Rai stage CLL have comparable progression free survival (PFS) at 2 years (97.2% vs 93.1%, p = 0.07) (Fung et al. 2007). Although few prognostic scores have been tested in MBL, the CLL-IPI is associated with shorter time to first treatment (TTFT) (Parikh et al. 2021; Kleinstern et al. 2020).

As well as disease progression, people with MBL and CLL experience increased serious infections (Agius et al. 2020; Andersen et al. 2019; Shanafelt et al. 2021) and other cancers (Besson et al. 2021; Brewer et al. 2012; Wang et al. 2021). Infections remain a major cause of death in treatment-naïve CLL, even accounting for age and comorbidities (Strati et al. 2017). Risk of death from infection parallels increasing CLL-IPI score (Wang et al. 2021). One third of treatment-naïve CLL patients experience severe infection, with a 30-day mortality of 10% (Andersen et al. 2019). The CLL-Treatment Infection Model (CLL-TIM) employed a machine learning based approach to identify patients at risk of either progression or severe infection (Agius et al. 2020). Patients at high-risk by this model, but not fulfilling IWCLL treatment criteria are eligible for PreVent-ACaLL a randomised trial of fixed duration acalabrutinib and venetoclax (Da Cunha-Bang et al. 2019).

Early identification of long-term stable disease (SD) is equally clinically valuable. The proposed IPS-E, developed specifically for early asymptomatic CLL is based on degree of lymphocytosis, IGHV status and palpable lymph nodes and reported 8.4% of low-risk CLL patients required treatment at 5 years. A Belgian study of 156 newly diagnosed patients added normal β2M to the IPS-E parameters to identify a low-risk subgroup, none of whom required treatment after 7.7 years (Galle 2018). The Danish group developed the CLL-WONT score comprising age, Binet

stage, LDH, lymphocyte count, β2M, IGHV status and del(11q) or del(17p) to identify CLL patients at low-risk of progression (10% requiring treatment at 5 years) suitable for monitoring in primary care (Brieghel et al. 2022).

3.1.2. Imprecise prognostic tools and toxicity of interventions limited the benefit of early intervention for CLL

Previous clinical trials failed to demonstrate a benefit for early chemotherapeutic intervention (Dighiero et al. 1998; Hoechstetter et al. 2017; Herling et al. 2020; Langerbeins et al. 2022).

The reasons early intervention failed to improve outcomes for CLL are twofold. As systems selecting patients for early intervention lack precision, studies inevitably enrol people with indolent CLL who would remain asymptomatic without treatment. One third of CLL1 participants assigned a high-risk score (based on diffuse marrow infiltration or lymphocyte doubling time < 12 months, with thymidine kinase > 7.0U/l or β2 microglobulin > 3.5mg/L) and randomised to placebo had not required treatment after 8.6 years (Hoechstetter et al. 2017). Of the CLL12 participants designated as high-risk using GSCLL score (Table 3.1.1.2) and randomised to placebo, 60% had not progressed at 3 years (Langerbeins et al. 2022). This rate of progression is lower than predicted, underlying the need for more robust prognostic tools.

Secondly, chemotherapeutic agents confer significant morbidity. In the chlorambucil for early CLL trial, an excess of deaths (relative risk 1.14, 95% CI 0.92-1.41) and second malignancies was observed in the treated group (Dighiero et al. 1998). More potent chemotherapy agents increased the toxicity. In CLL1 trial, 91.5% fludarabine treated patients reported Grade 3-4 AE (Hoechstetter et al. 2017). In the CLL7 trial for early CLL, 74.4% of FCR treated patients experienced Grade 3-5 AE and treated related deaths occurred in 2.4% (Herling et al. 2020). With ibrutinib for early CLL, 21.5% of ibrutinib treated patients had cardiac arrhythmias versus 7.7% of the placebo group (Langerbeins et al. 2022).

Table 3.1.1.2: CLL Prognostic Scoring Systems and Performance in Predicting Time to First Treatment (TTFT)

Features included in CLL Prognostic Scores

Scoring system	Age	Gender	Binet or Rai Staging	Lymphocyte count	β2M	LDH	IGHV status	*TP53*	Other features	Accuracy at Predicting TTFT % (95% CI)
CLL-IPI	Yes	--	Yes	--	Yes	--	Yes	Deletion & Mutation	--	72% (67-77)
IPS-E	--	--	#	Yes	--	--	Yes	--	--	
Belgian (Galle) CLL-WONT	--	--	Yes	Yes	Yes	--	Yes	--	--	NR
	Yes	--	Yes	Yes	Yes	Yes	Yes	deletion	Del(11q)	
MDACC 2007	Yes	Yes	Yes	--	Yes	--	--	--	--	64% (59-69)
MDACC 2011	--	--	#	--	--	Yes	Yes	Deletion	Del(11q)	70% (56-76)
Barcelona -Brno	--	--	--	--	--	--	Yes	Deletion	Del(11q)	74% (70-79)
O-CLL1			Yes	Yes	Yes		Yes	Deletion	Del(11q)	63% (58-69)
GCLLSG	Yes	Yes	--	--	Yes	--	Yes	Deletion	Del(11q) ECOG PS Serum Thymidine Kinase	72% (67-77)

 # IPS-E includes presence of lymphadenopathy, MDACC includes number & size of lymph nodes
NR not reported Table adapted from ***(Molica et al. 2017)***

3.1.3 Mutations in CLL-driver genes are detectable at early stages of disease and may refine prognosis

Established CLL biomarkers such as chromosomal deletions and IGHV mutational status are detectable in MBL. Like indolent CLL, the most frequently identified CNA in MBL are del(13q14) and trisomy 12 (Lanasa et al. 2011). Unmutated IGHV (u-IGHV) and the hierarchical classification of cytogenetic aberrations are risk factors for progression from MBL to CLL and for shorter TTFT in CLL (Rawstron et al. 2008; Shanafelt et al. 2010). The majority of MBL carries m-IGHV and the IGHV repertoire for MBL mirrors that of favourable prognosis CLL (Rawstron et al. 2008; Condoluci et al. 2020).

Mutations in CLL-driver genes conferring growth or survival advantages are detectable in MBL, albeit at lower frequency. *NOTCH1* mutations are detectable by Sanger Sequencing (SS) in 3.2% high-count MBL patients (Rasi et al. 2012). Using heat-melt curve analysis and SS, mutations in *NOTCH1* and *SF3B1* were found in 3.5% and 7% of MBL respectively (Nils Winkelmann 2013). Targeted NGS performed on paired samples at MBL diagnosis and follow-up identified mutations in *NOTCH1*, *BIRC3* and *SF3B1*, signifying a shortened TTFT. *TP53* and *XPO1* mutations were absent, implying these are secondary events or that they occur below the limit of detection. Non-progression was associated with absence of driver mutations and sub-clonal stability (Barrio et al. 2017). Another WES study of 20 ultra-stable CLL patients noted CLL-driver mutations were absent, although analysis was restricted to mutations with VAF > 0.20 (Agathangelidis, Ljungstrom, et al. 2018).

Detection of somatic variants in CLL-driver genes may identify patients who will require treatment. In early CLL, mutated *ATM* was significantly associated with a shorter TTFT (Hu et al. 2019). Other studies of early CLL reported limited impact of mutations in *ATM*, *MYD88, NOTCH1* and *SF3B1* on TTFT, which may reflect the limited sensitivity of the testing method used. Only cases with allele frequency > 0.10 were considered positive, missing sub-clonal mutations (Condoluci and Rossi 2019). No impact *of MYD88, XPO1* or *EGR2* mutations were observed in a Danish cohort of newly diagnosed CLL (Brieghel et al. 2020).

Clonal evolution and expansion of CLL subclones carrying CLL-driver mutations is recognised to drive relapsed CLL after chemotherapy (Landau et al. 2013) and BTKi therapy (Black et al. 2022). Progression from indolent disease to CLL requiring therapy is associated with acquisition of new mutations and was preceded by sub-clonal expansion in 11/47 (23.4%) MBL (Barrio et al. 2017). Identification of changes in the mutational profile anticipated progression by an average of 60 months (Ojha et al. 2014). In addition to specific somatic mutations, the number of CLL-drivers and the number of functionally altered signalling pathways informs prognosis in newly diagnosed CLL. Shorter TTFT was observed with increasing number of driver mutations (Kleinstern et al. 2020) and increasing number of altered signalling pathways (Brieghel et al. 2020). Both studies utilised a hybridisation capture library preparation and targeted deep sequencing.

Taken together, these results imply that sensitive techniques to detect low frequency mutations in early-stage disease may identify MBL and early CLL at risk of progression. Reliable identification of MBL and early CLL at risk of progression is a prerequisite for any early intervention trial.

3.1.4 Hypothesis and objectives for Chapter 3

My work represents a pilot study using targeted capture hybridisation sequencing to evaluate whether:

- CLL-driver variants are detectable at low variant allele frequency in MBL and Rai 0 CLL using sequencing techniques optimised for identifying low frequency variants.
- If driver variants can be detected, does progression to symptomatic CLL requiring treatment correlate with the expansion of B-cell clones carrying CLL-driver mutations.
- Methodology from this pilot study will be applied to the retrospective MBL and Rai 0 CLL samples from the LYM1 cohort to assess whether progressors can be distinguished from non-progressors by
    - Number of driver mutations detected
    - Detection of the expansion of the clone containing the driver mutations

My work informs the laboratory approach to the OXPLORED study objectives. As it is exploratory, it is not powered to support or refute the hypothesis, rather the goal is to generate data to inform the sequencing approach for the full OXPLORED cohort (800 MBL/CLL participants).

3.2 Methods

3.2.1 Study population and data sources

Sequential PBMC samples obtained at diagnosis, follow-up and pre-treatment were available from 20 LYM1 study participants, representing an opportunity to evaluate sequential MBL and Rai 0 CLL samples as a pilot for the OXPLORED study. Samples from participants who experienced indolent disease for at least three years were selected. Data on lymphocyte count, date of progression or last follow-up, chromosomal deletions by FISH and IGVH status by SS were available.

Prospective MBL/early CLL samples came from five OXPLORED participants for whom sequential samples and clinical outcome data available (Section 2.2). Salivary germline and sorted CD19+ B-cell (tumour) samples were collected at study enrolment and at 24-month follow-up or disease progression, whichever occurred sooner (Table 2.2.2.2). Baseline clinical was obtained via OpenClinica. IGHV status was determined using IgCaller on Whole Genome Sequencing (WGS) datasets (Section 2.1.2).

Ethical approvals are described in Section 2.1 and 2.2

3.2.2 Capture hybridisation library preparation and sequencing

For the LYM1 samples, tumour DNA was extracted from PBMCs, and no germline was available. For the OXPLORED cohort, salivary DNA was used for germline and DNA extracted from CD19+ sorted B-cells was used as a source of tumour DNA (Section 2.4).

Panel design, DNA extraction and preparation of molecular barcoded Thruplex-IDT libraries are described in Section 2.6. Following bead clean-up and quantification, libraries were hybridised with custom probes and blocking oligos overnight. Post-hybridisation capture involving serial bead purification and wash buffers was performed prior to post-capture library amplification. Libraries were quantified by real-time PCR, normalised and pooled for sequencing.

The first two runs were sequenced on the Illumina MiSeq. Subsequent runs were sequenced on the Illumina NextSeq 550 using the Mid-Output flow cell (Section 2.6.8). Fastq files were analysed using the OMDC bioinformatics pipeline to generate annotated variant files (Section 2.6.9).

3.2.3 Variant filtering for the LYM1 and OXPLORED samples

Using the annotated varscan files, I filtered variants aiming to detect low frequency variants while minimising false positives and sequencing artefacts (Section 2.6.9). Variants fulfilling the following criteria were retained; Quality Score > 30; non-synonymous variants, Fischer Strand bias score < 60, Minor Allele Frequency < 0.01 and False Discovery < 0.01. Variants with a COSMIC ID were retained, visually inspected using Integrated Genome Viewer (IGV) and manually curated, prioritising mutations previously reported in haematological cancers in line with the approach described (Mansouri et al. 2023).

For the LYM1 cohort, no germline was available. A paired salivary germline was sequenced for each OXPLORED participant. For the sequential samples, shared homozygous and heterozygous synonymous variants were used to exclude sample mix-up.

3.2.4 Calculation of International Prognostic Score for Early CLL (IPS-E)

The IPS-E was chosen as it was developed for early CLL and used available datapoints (Condoluci et al. 2020). The scores were applied to MBL patients. Lacking β2M results, the CLL-IPI or CLL-WONT scores could not be determined.

3.2.5 Statistical analysis and data visualisation

Variant filtering and data visualisation were performed in R using packages tidyverse, and ggpubr. Dot plots, box plots and bar plots were drawn with R package 'ggplot2'. For the purposes of visualisation, variants that were undetected at one timepoint were assigned a VAF of –0.1. Correlation was calculated using Pearson's correlation and visualised as a heatmap using ggplot2. Treatment Free Survival (TFS) was estimated using the Kaplan-Meier method (Section 2.8).

3.3 Results

In this chapter, I examine using capture hybridisation based targeted sequencing two MBL and early CLL cohorts; the LYM1 cohort in whom prolonged clinical follow-up was available and selected samples from the ongoing OXPLORED trial.

Here, I confirm unmutated IGHV (u-IGHV) status delineates a subgroup who progress to treatment. MBL/CLL with u-IGHV status was enriched for somatic mutations. The majority of patients with MBL and early CLL in this cohort do not have a somatic mutation detectable. All variants detected in the LYM1 cohort occurred at allele frequency (AF) > 5%. When sorted B-cells were used as a source of input DNA, variants at AF < 5% were seen. Sequential samples showed increases, decreases and stable AF.

3.3.1 The retrospective MBL and CLL cohort from the LYM1 study

Patient characteristics of the LYM1 cohort

Sample workflow is shown in Figure 3.3.1.1. Clinical characteristics of the 17 MBL/CLL patients sequenced are shown in Table 3.3.1.1. Eleven of 17 progressed to treatment after a median of 9.8 years follow-up.

Sequencing results for the LYM1 cohort

Libraries were sequenced on four runs, with a mean of 90.8% reads ≥ Q30 (range 87.7-92.9%). Depth of sequencing per gene is summarised in Figure 3.3.1.2. Post-sequencing filtered was performed as described (Section 3.2.3), prioritising variants with a COSMIC ID (Mansouri et al. 2023) and illustrated in Figure 3.3.1.3.

Figure 3.3.1.1: CONSORT diagram of participants/samples

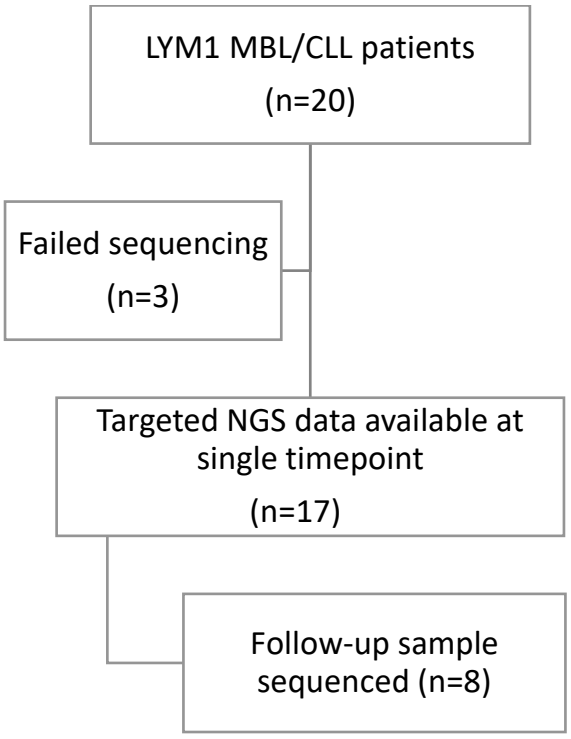


Figure 3.3.2.1: Visualisation of the LYM1 cohort samples and sequencing

Table 3.3.1.1: Clinical characteristics of the LYM1 Cohort sequenced at baseline (n=17)

Age	
Median (IQR)	70.5 years, (66.8-76.8 years)
Diagnosis	
Monoclonal B-cell lymphocytosis	5 patients
Chronic Lymphocytic Leukaemia	10 patients
Size of clonal B-cell population unknown	2 patients
Lymphocytosis	
Median (IQR)	14.9 x 10^9/L, (5.5-15.6 x10^9/L)
IGHV status	
Mutated-IGHV (<98%)	5 patients
Unmutated-IGHV (>98%)	10 patients
Unknown	2 patients
FISH result	
No abnormality	2 patients
Del(13q) as sole abnormality	4 patients
Del(11q)	3 patients
Trisomy 12	6 patients
Del(17p)	0 patients
No data	2 patients
International Prognostic Score for early CLL	
Low-risk	3 patients
Intermediate-risk	9 patients
High-risk	2 patients
Unable to determine due to missing data	3 patients
Duration of follow-up	
Median (IQR)	9.75 years (7.6-10.7 years)
Status at 10 years follow-up	
Treated	11 patients

IQR = interquartile range; FISH = fluorescence in-situ hybridisation; IGHV = Immunoglobulin heavy chain rearrangement

Figure 3.3.1.2: Boxplot of coverage of target genes

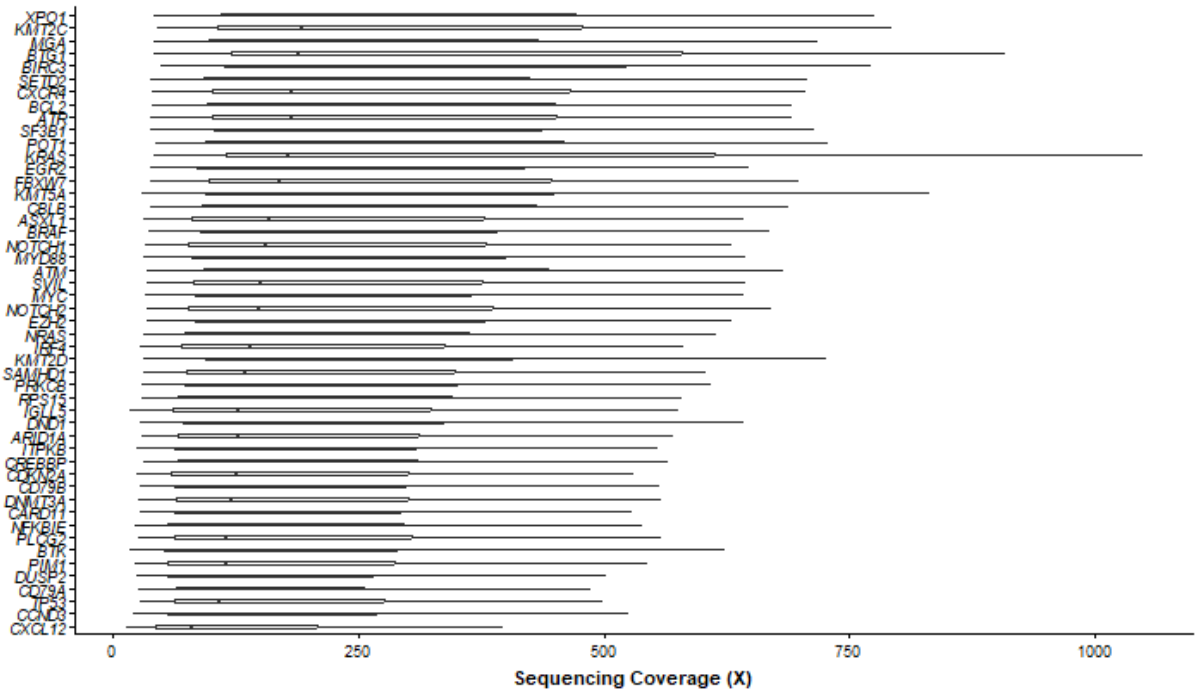


Figure 3.3.1.2: Sequencing coverage of target genes (y-axis). Boxplot: centre line, median; box limits upper and lower quartiles

Figure 3.3.1.3: Flow diagram of variant filtering for the LYM1 baseline samples

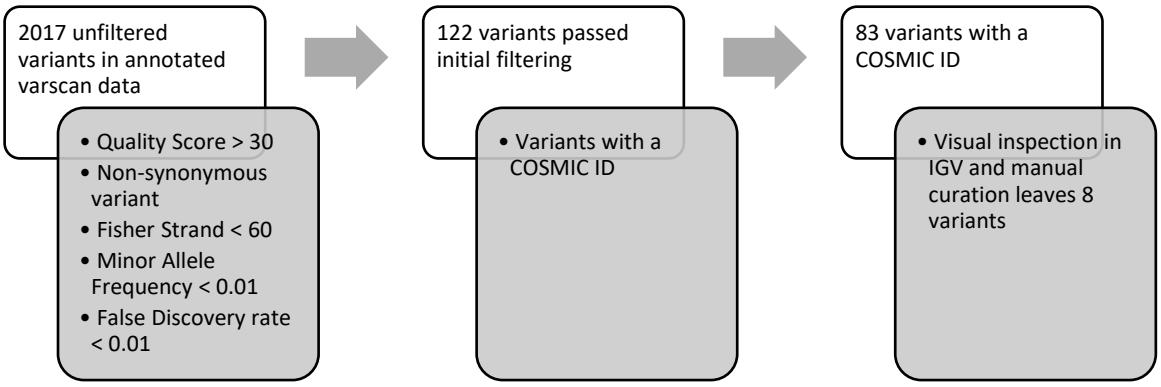


Figure 3.3.1.3: Visualisation of post-sequencing filtering for the LYM1 and OXPLORED samples

3.3.2 Mutations and patterns of cooccurrence in the LYM1 cohort

No variant was identified in 11/17 baseline samples. Eight variants were detected in six baseline samples (Figure 3.3.2.1). The 2bp frameshift mutation in *NOTCH1* was the most frequently identified variant, detected in three patients (Table 3.3.2.1). Variants occurred at median AF of 0.19 (IQR 0.09 – 0.37). No variants were found in the five patients with m-IGHV. In two patients, the IGHV status could not be determined, one of whom carried a *BIRC3* mutation. The remaining seven variants co-occurred with u-IGHV (Figure 3.3.2.1 and Figure 3.3.2.2). One u-IGHV patient (BM-0306) had 3 variants detected in *NOTCH1*, *FBXW7* and *RPS15.*

Figure 3.3.2.1: Number of variants by IGHV status (n=17)

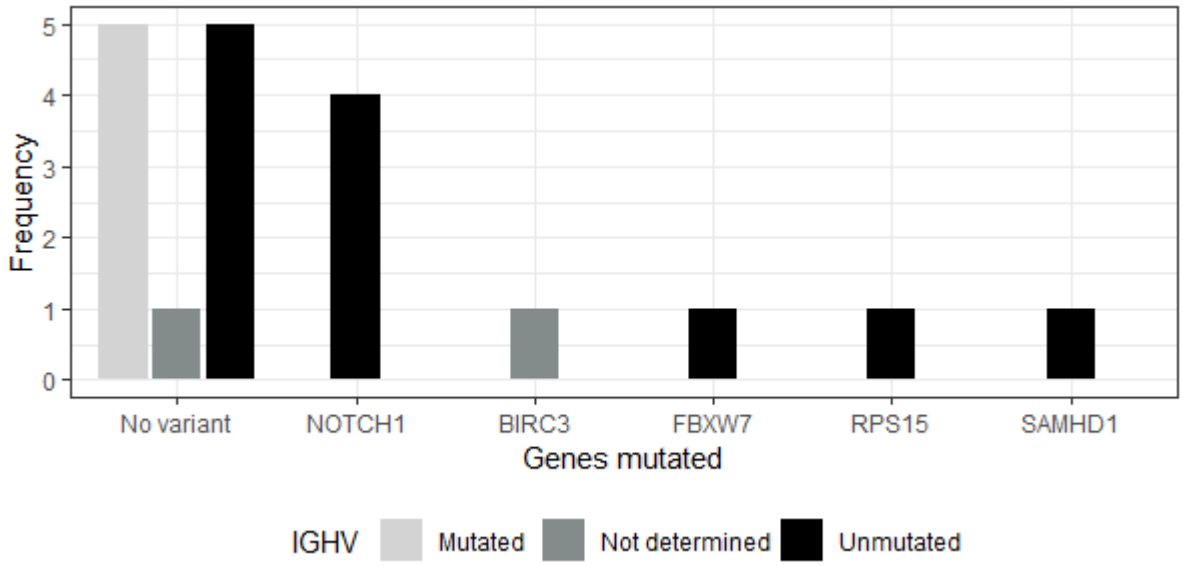


Figure 3.3.2.1: Barplot showing number of variants by m-IGHV (light grey) and u-IGHV (black) status. The IGHV status was undetermined for 2 patients (dark grey).

Table 3.3.2.1: Variants identified in the baseline LYM1 samples

LYM1 Patient ID	Gene	Chromosome: Position: Change	VAF	HGVSc	HGVSp	COSMIC
BM-0160	*SAMHD1*	chr20:35533825:C/T	0.1586	NM_015474.4:c.1352G>A	NP_056289.2:p.Arg451His	COSV53429981
BM-0306	*FBXW7*	chr4:153247288:C/A	0.0729	NM_018315.5:c.1274G>T	NP_060785.2:p.Arg425Leu	COSV55901126
	NOTCH1	chr9:139390152:T/C	0.0693	NM_017617.5:c.*371A>G	NA	COSV53038282
	RPS15	chr19:1440436:C/A	0.3734	NM_001018.5:c.413C>A	NP_001009.1:p.Ser138Tyr	COSV51832669
BM-0552	*NOTCH1*	chr9:139390649:AG/-	0.0966	NM_017617.5:c.7541_7542del	NP_060087.3:p.Pro2514ArgfsTer4	COSV53024776
BM-1596	*BIRC3*	chr11:102207657:C/-	0.2456	NM_001165.5:c.1639del	NP_001156.1:p.Gln547AsnfsTer21	COSV104374664
BM-2462	*NOTCH1*	chr9:139390649:AG/-	0.2143	NM_017617.5:c.7541_7542del	NP_060087.3:p.Pro2514ArgfsTer4	COSV53024776
BM-3399	*NOTCH1*	chr9:139390649:AG/-	0.2688	NM_017617.5:c.7541_7542del	NP_060087.3:p.Pro2514ArgfsTer4	COSV53024776

Figure 3.3.2.2: Co-occurrence heatmap from the baseline LYM1 samples

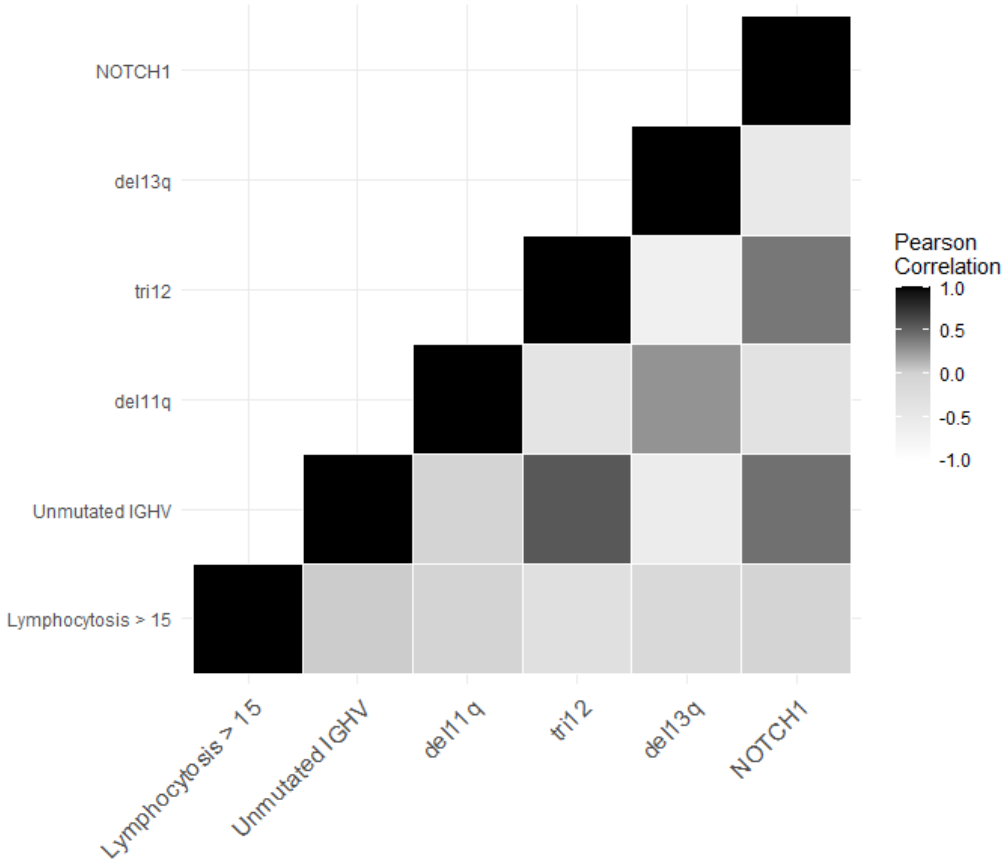


Figure 3.3.2.2: Co-occurrence heatmap of lymphocytosis, unmutated IGHV, mutated genes and chromosomal aberrations in the LYM1 cohort of MBL and early CLL.

3.3.3 Relationship between IGHV, mutations, IPS-E and treatment-free survival

The impact of IGHV status on treatment-free survival (TFS) approached significance (p=0.05, Figure 3.3.3.1). No difference in TFS was observed in patients with or without any variant (p=0.11), and with or without *NOTCH1* mutations (p=0.11), however all mutations co-occurred with u-IGHV. Grouping patients by IGHV status and presence/absence of mutations did not improve the discriminative capacity (Figure 3.3.3.2, p=0.14). No patient with low-risk IPS-E progressed at 10 years follow-up. The difference in TFS by IPS-E score was not significant (Figure 3.3.3.3, p = 0.18).

The OXPLORED cohort

At this point, my data did not support the hypothesis that deep sequencing techniques detect mutations at low VAF at the earliest stages of disease. All variants detected occurred at VAF > 0.07. DNA libraries for the LYM1 cohort were made from unsorted PBMC and represented a mixture of neoplastic B-cells and normal leukocytes. In this cohort with relatively modest lymphocytosis (median lymphocytosis 14.9 x 10^9/L), using unsorted cells as a source of DNA potentially limited the capacity for detection of low VAF variants.

To assess whether this limitation could be overcome by preparing libraries from sorted B-cells, I therefore went on to use CD19+ B-cell samples from the ongoing OXPLORED study. B-cells sorted to >90% purity was anticipated to provide a better source of DNA for detecting low VAF variants in early CLL.

Figure 3.3.3.1: Treatment free survival in the LYM1 cohort by IGHV status(n=15)

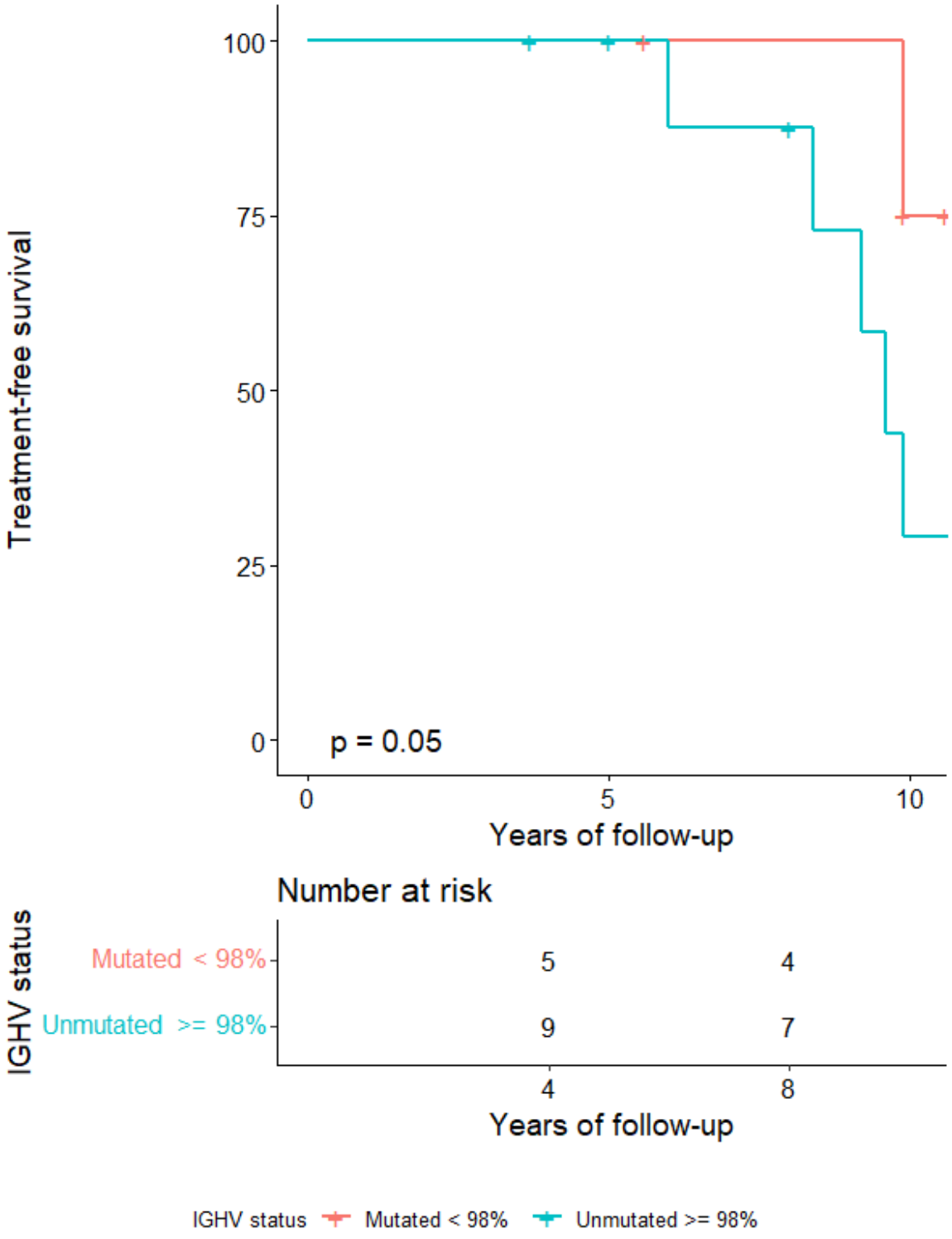


Figure 3.3.3.1: Kaplan-Meier curve of TFS stratified by IGHV mutational status. IGHV rearrangement status was defined as > or < 98% homology to the germline. Patients with mutated IGHV status are shown in red. Patients with unmutated in blue. The IGHV status is unknown for two patients, BM-1596 and BM-4073. BM-1087 completed 3.7 years of follow-up and was stable at last follow-up.

Figure 3.3.3.2: Treatment free survival by IGHV status and presence of mutations in the LYM1 cohort (n=15)

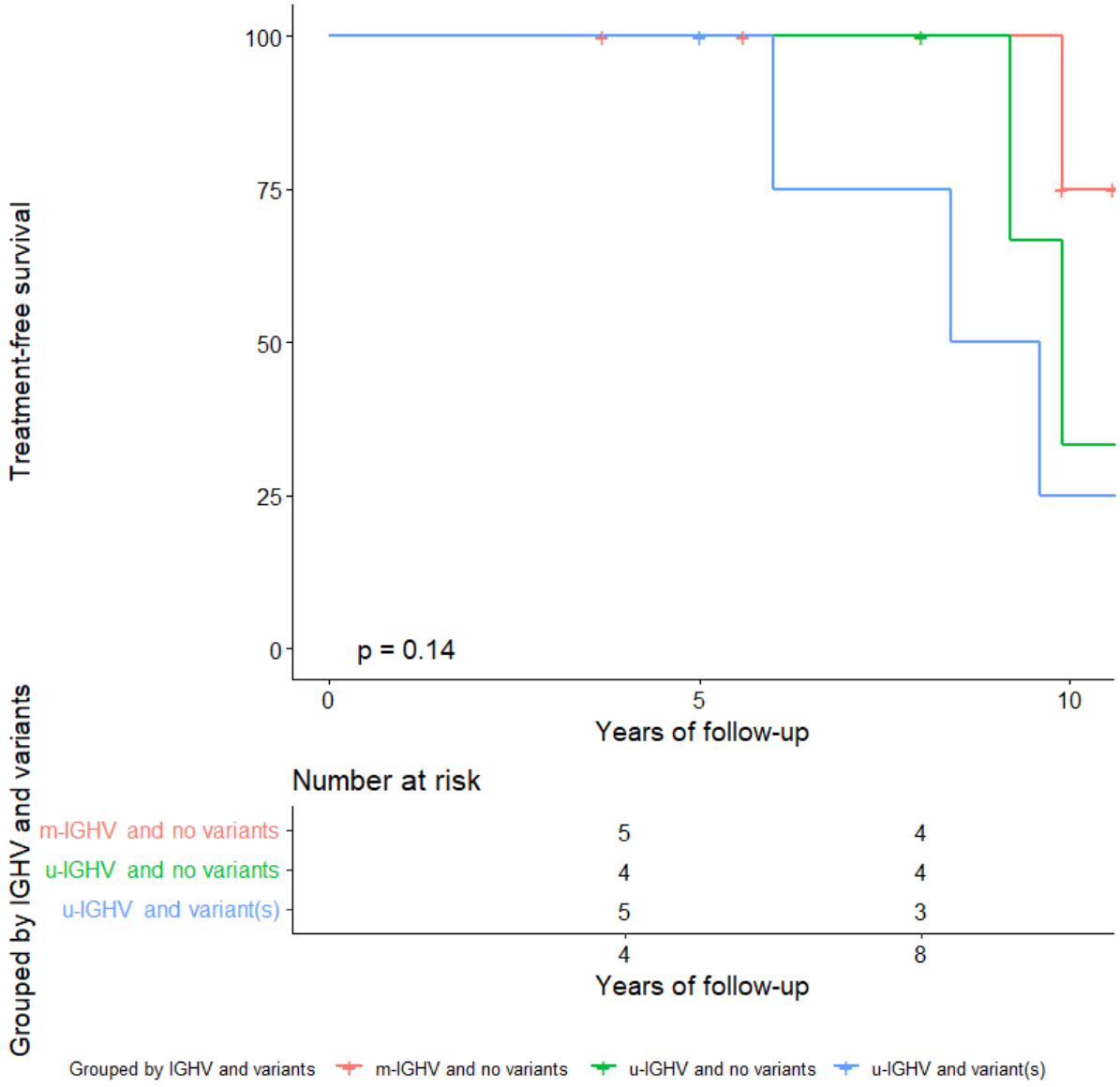


Figure 3.3.3.2: Kaplan-Meier curve of TFS stratified by IGHV mutational status and somatic mutations. Patients were grouped by IGHV rearrangement status and presence/absence of somatic variants into m-IGHV and no variants (red), u-IGHV and no variant (green) and u-IGHV and any variant (blue). IGHV rearrangement status was defined as > or < 98% homology to the germline. The IGHV status is unknown for two patients, BM-1596 and BM-4073. BM-1087 completed 3.7 years of follow-up and was stable at last follow-up.

Figure 3.3.3.3: Treatment free survival by IPS-E score in the LYM1 cohort (n=14)

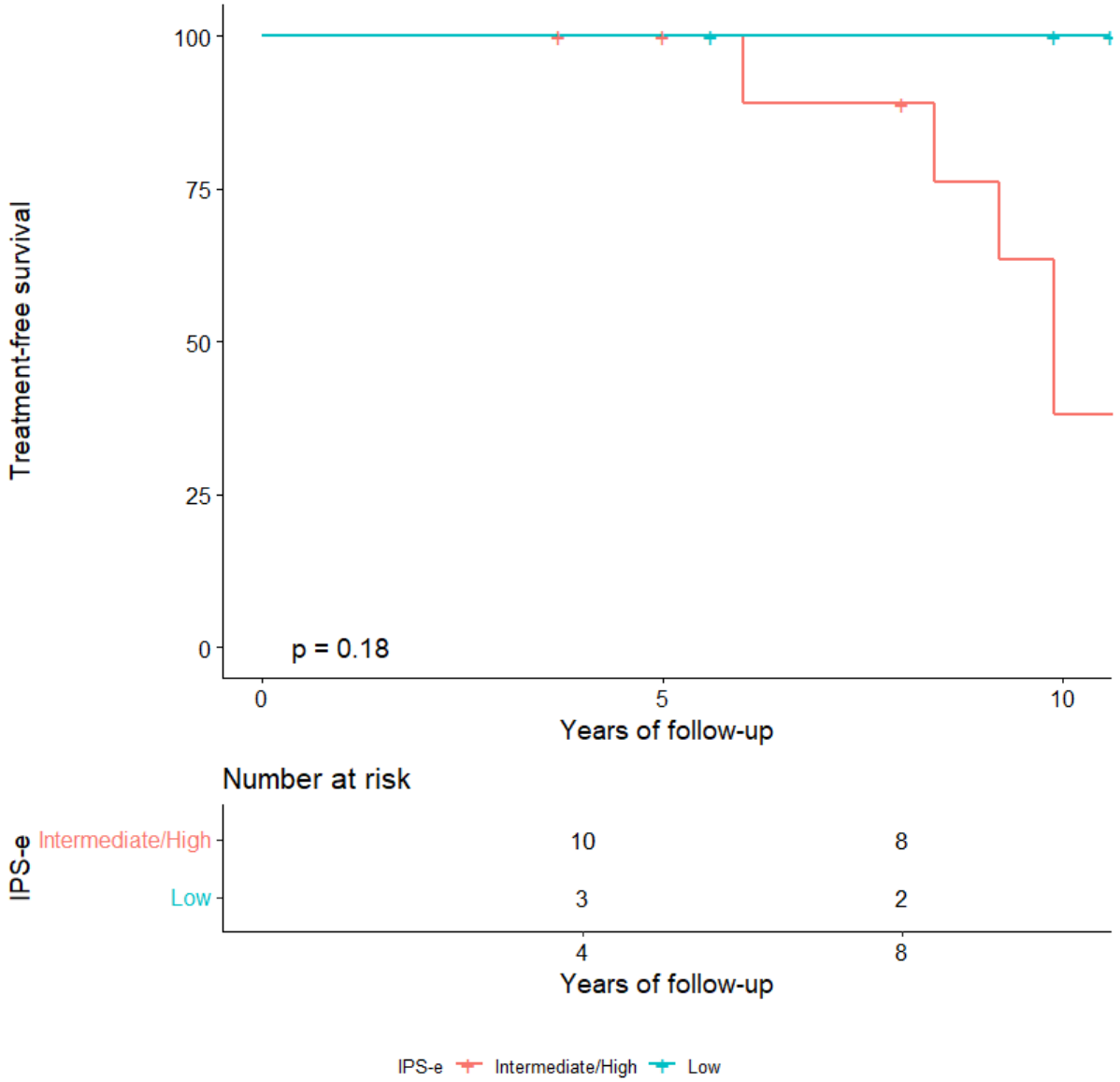


Figure 3.3.3.3: Kaplan-Meier curve of TFS stratified by IPS-E score. All LYM1 participants had no lymphadenopathy. Mutated IGHV and lymphocytes < $15x10^9$/L scored 0 or low-risk (blue). Unmutated IGHV and/or lymphocytes > $15x10^9$/L scored 1 point each. Patients with 1 (intermediate) or 2 (high) IPS-E scores are considered together for visualisation. IPS-E could not be determined for 3 participants due to unknown IGHV status (n=2) or missing lymphocyte count (n=1). BM-1087 completed 3.7 years of follow-up and was stable at last follow-up.

3.3.4 Clinical features of the OXPLORED cohort

Five asymptomatic CLL OXPLORED participants for whom sequential samples and clinical outcomes were available were selected. The participants experienced stable early CLL (n=2), progression to CLL requiring treatment (n=2) and high-grade transformation (n=1) (Figure 3.3.4.1.). Stable was defined as not requiring treatment prior to the 24-month follow-up visit. Salivary DNA provided the germline sample (Section 2.4.3). Clinical features are summarised in Table 3.3.4.1. below.

Table 3.3.4.1: Clinical features of the OXPLORED participants sequenced

	Nodes	IGHV status	Lymph $x10^9/L$	IPS-E	β2M g/L	Outcome at 24-months
OXP-10002	Yes	Unmutated (98.3%) IGHV2-5	122.69	High	2	Stable
OXP-10008	No	Mutated (94.4%) IGHV4-34	3.98	Low	1.7	Stable
OXP-10010	Yes	Unmutated (100%) IGHV1-69	11.64	Intermediate	2.8	Progressed
OXP-10029	Yes	Unmutated (98.9%) IGHV1-3	11.99	Intermediate	1	Richter's Transformation
OXP-10114	Yes	Unmutated (100%) IGHV1-69 #subset3	110.85	High	5.1	Progressed

Figure 3.3.4.1: CONSORT diagram of OXPLORED participants/samples

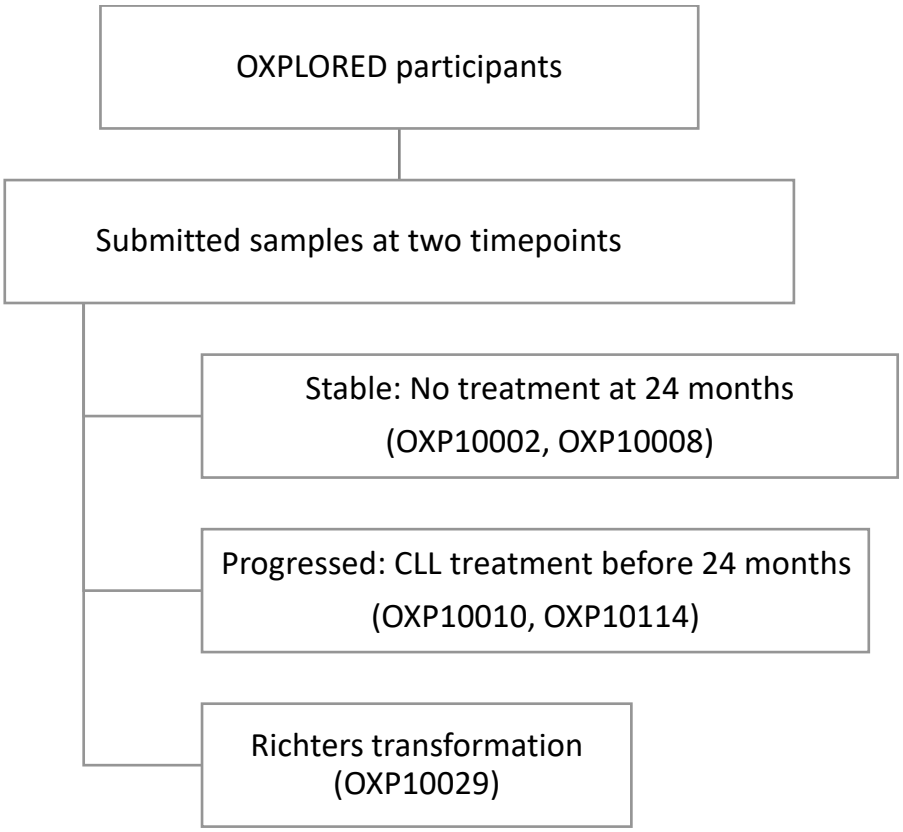


Figure 3.3.4.1: Visualisation of the samples submitted and outcomes for the five OXPLORED participants

3.3.5 Mutations and patterns of cooccurrence in the OXPLORED cohort

At baseline, 0-9 variants were detected in the CD19+ tumour cells (Figure 3.3.6.1 and Table 3.3.6.2). One m-IGHV patient (OXP-10008) had an *IGLL5* variant. All other variants were found in u-IGHV patients. In OXP-10029, nine variants were detectable at enrolment, 16.2 months prior to the diagnosis of RT. The pattern of mutation impacts the *NOTCH1* and *MYC* related signalling, both of which are implicated in the pathogenesis of high-grade transformation (Klintman et al. 2021; Rossi, Spina, and Gaidano 2018; Nadeu et al. 2022). This finding implies that patients at risk of high-grade transformation may be identifiable in advance.

Figure 3.3.5.1: Variants detected at baseline in OXPLORED samples

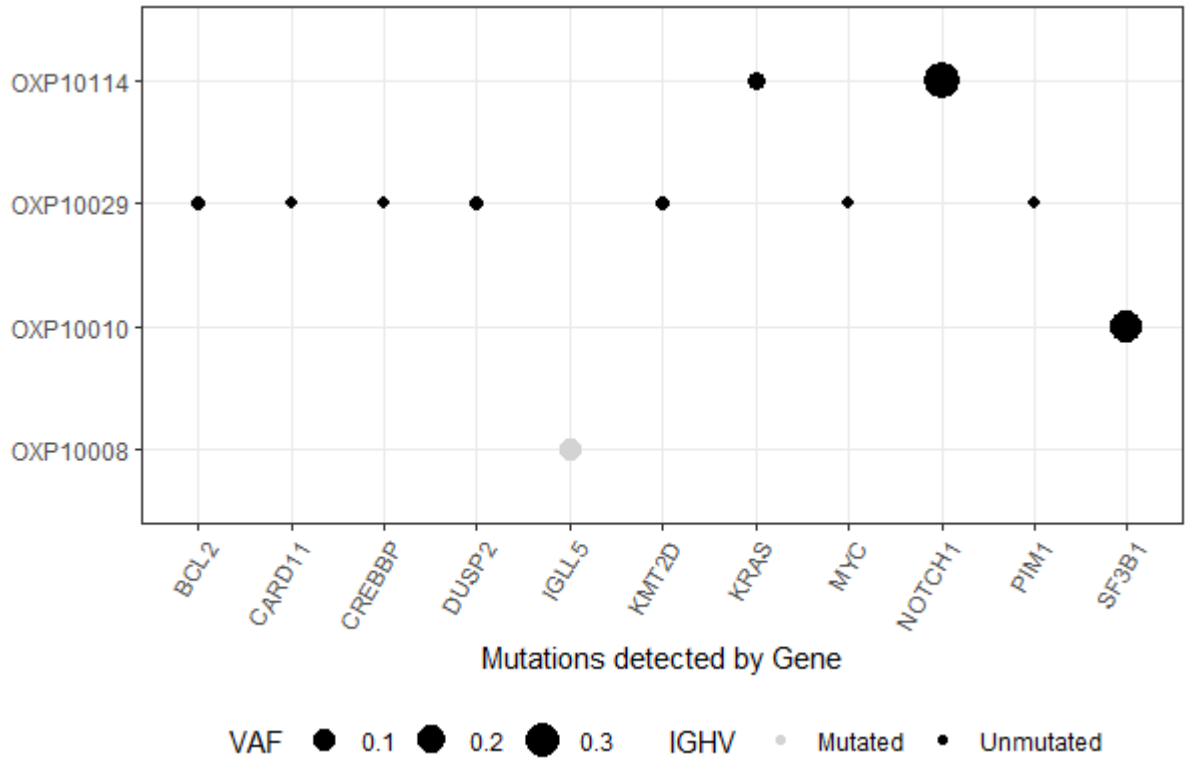


Figure 3.3.5.1: Bubble plot of the variants detected at baseline in the OXPLORED participants. Study participants on the y-axis. Genes with variants detected on the x-axis. Size of the circle reflects the Variant Allele Frequency (VAF). IGHV status is coloured black for u-IGVH and light grey for m-IGHV. No variants were detected in OXP-10002.

Table 3.3.5.1: Variants identified in OXPLORED CD19+ B-cells at baseline

Participant	Gene	Chr: Position	VAF	HGVSc	HGVSp	COSMIC
OXP10002	No variant detected					
OXP10008	IGLL5	chr22:23230405	0.09	NM_001178126.2:c.172A>G	NP_001171597.1:p.Ser58Gly	COSV73250237
OXP10010	SF3B1	chr2:198266834	0.28	NM_012433.4:c.2098A>G	NP_036565.2:p.Lys700Glu	COSV59205318
OXP10029	DUSP2	chr2:96809972	0.02	NM_004418.4:c.651C>G	NP_004409.1:p.Phe217Leu	COSV56619739
	PIM1	chr6:37139249	0.02	NM_002648.4:c.589G>C	NP_002639.1:p.Val197Leu	COSV65165639
	CARD11	chr7:2976758	0.02	NM_032415.7:c.1254G>C	NP_115791.3:p.Arg418Ser	COSV67796802
	MYC	chr8:128751013	0.02	NM_002467.6:c.550A>G	NP_002458.2:p.Ser184Gly	COSV99422123
	KMT2D	chr12:49435108	0.03	NM_003482.4:c.6445T>A	NP_003473.3:p.Ser2149Thr	COSV104396916
	KMT2D	chr12:49448385	0.02	NM_003482.4:c.326T>C	NP_003473.3:p.Val109Ala	COSV56407793
	CREBBP	chr16:3900611	0.01	NM_004380.3:c.485C>T	NP_004371.2:p.Ala162Val	COSV52137833
	BCL2	chr18:60985385	0.02	NM_000633.3:c.515A>G	NP_000624.2:p.Asn172Ser	COSV61372150
	BCL2	chr18:60985462	0.02	NM_000633.3:c.438G>C	NP_000624.2:p.Arg146Ser	COSV61381975
OXP10114	KRAS	chr12:25398284	0.06	NM_004985.5:c.35G>A	NP_004976.2:p.Gly12Asp	COSV55497369
	KRAS	chr12:25398285	0.01	NM_004985.5:c.34G>T	NP_004976.2:p.Gly12Cys	COSV55497461
	NOTCH1	chr9:139390649	0.37	NM_017617.5:c.7541_7542del	NP_060087.3:p.Pro2514ArgfsTer4	COSV53024776

3.3.6 Sequential samples (LYM1 and OXPLORED cohorts)

Paired samples at baseline and follow-up were available for 13 patients, 11 of whom of which were successfully sequenced at both timepoints. Seven LYM1 patients sequenced at follow-up (median 8 years) and four OXPLORED participants (median 15 months) were sequenced. No significant difference in median depth of sequencing (137x baseline vs 99x follow-up, wilcoxon, p= 0.37) or capture efficiency (proportion of off-target reads 0.44 baseline vs 0.47 follow-up, wilcoxon, p = 0.23) was seen. Identical homozygous variants were confirmed at both timepoints, excluding sample mix-up.

Three patients (BM-1087; BM-4082; BM-4478) had no variants detected in either sample. BM-1087 and BM-4082 remained stable at 3.7 and 5.5 years respectively. BM-4478 progressed after 9.9 years follow-up.

Eight patients had a mutation detectable in at least one sample (Table 3.3.6.1). In the two non-progressors, variants affecting *SAMHD1* (BM-0160) and *IGLL5* (OXP10008) were found at higher VAF in the follow-up sample. For the six progressors, no consistent pattern of change in VAF was observed. In OXP10002 and BM-0552, new variants affecting *NOTCH1* and *TP53* respectively were identified at progression. In BM-0306, the VAF of the *FBXW7* and *RPS15* was stable. For OXP10029, who developed RT at 16.2 months follow-up, only one variant (*CREBBP* COSV52137833) was shared between both timepoints.

Table 3.3.6.1 Changes in variants detected in the sequential samples (n=8)

Patient ID	Gene	HGVSc	COSMIC ID	Baseline VAF	Follow-up VAF	Clinical outcome	Follow-up (years)
BM-0160	SAMHD1	NM_015474.4:c.1352G>A	COSV53429981	0.16	0.25	Stable	5.0 years
BM-0306	FBXW7	NM_018315.5:c.1274G>T	COSV55901126	0	0.07	Stable	8.0 years
	NOTCH1	NM_017617.5:c.*371A>G	COSV53038282	0.07	0.04		
	RSP15	NM_001018.5:c.413C>A	COSV51832669	0.37	0.44		
BM-0552	CARD11	NM_032415.7:c.343T>A	COSV62716639	0.08	ND	Progressed	6.8 years
	NOTCH1	NM_017617.5:c.*378A>G	COSV53104793	0.21	0.22		
	NOTCH1	NM_017617.5:c.7541_7542del	COSV53024776	0.10	ND		
	MGA	NM_001164273.2:c.2738del	COSV54952539	0.01	0.14		
	TP53	NM_000546.6:c.395A>G	COSV52666637	ND	0.29		
BM-3399	NOTCH1	NM_017617.5:c.7541_7542del	COSV53024776	0.27	0.21	Stable	6.0 years
OXP10002	NOTCH1	NM_017617.5:c.7541_7542del	COSV53024776	ND	0.02	Stable	2 years
OXP10008	IGLL5	NM_001178126.2:c.172A>G	COSV73250237	0.09	0.45	Stable	2 years
OXP10029	DUSP2	NM_004418.4:c.651C>G	COSV56619739	0.02	ND	Richter Transformation	1.3 years
	PIM1	NM_002648.4:c.589G>C	COSV65165639	0.02	ND		
	CARD11	NM_032415.7:c.1254G>C	COSV67796802	0.02	ND		
	MYC	NM_002467.6:c.550A>G	COSV99422123	0.02	ND		
	KMT2D	NM_003482.4:c.6445T>A	COSV104396916	0.03	ND		
	KMT2D	NM_003482.4:c.326T>C	COSV56407793	0.02	ND		
	CREBBP	NM_004380.3:c.485C>T	COSV52137833	0.01	0.07		
	BCL2	NM_000633.3:c.515A>G	COSV61372150	0.02	ND		
	BCL2	NM_000633.3:c.438G>C	COSV61381975	0.02	ND		
OXP10114	KRAS	NM_004985.5:c.35G>A	COSV55497369	0.06	0.074	Progressed	0.42 years
	KRAS	NM_004985.5:c.34G>T	COSV55497461	0.01	0.01		
	NOTCH1	NM_017617.5:c.7541_7542del	COSV53024776	0.37	0.434		

ND not detected

3.4 Discussion

This work provides preliminary data on the potential impact of CLL-driver mutations on clinical outcome in a selected cohort of MBL and early CLL. The main objective of the work presented in this chapter was to i) detect variants at low VAF in early CLL using capture based NGS ii) describe the relationship between the presence of low VAF variants (number detected and change in the allele frequency over time) and progression of CLL.

Lack of a germline sample limited my ability to confidently discriminate between private germline polymorphisms and somatic variants in the LYM1 cohort, necessitating filtering based on previously reported variants. With prospectively collected MBL/CLL cohorts, choice of germline samples is critical. PB samples, even when B-cell depleted may be contaminated by tumour cells. The choice of salivary germline for OXPLORED balanced the competing demands of an uncontaminated sample with the pragmatic sampling of large numbers of people.

In this pilot study, targeted sequencing from unsorted PBMCs yielded few mutations and variants at AF < 0.05 were absent. Changing from PBMCs to CD19+ sorted cells providing enhanced and uniform B-cell purity facilitated low frequency variants detection for OXPLORED participants. Ten of fourteen baseline variants occurred at AF < 0.05. These results imply that sorted B-cells are necessary to detect low frequency variants even with a deep targeted approach. Sorting CD19+ B-cells has been incorporated in the OXPLORED study workflow.

Non-tumour cells, notably monocyte derived nurse-like cells and mesenchymal stromal cells play a role in survival of CLL cells (Burger and Gribben 2014). Phenotypic and functional changes in non-tumour cells, notably T-cells are described in CLL. T-cells in CLL exhibit an exhausted phenotype and impaired capacity for immune synapse formation (Vlachonikola et al. 2021; Roessner and Seiffert 2020). The decision to sort circulating B-cells in OXPLORED potentially limits the capacity to study crosstalk between the neoplastic B-cells and non-tumour cells. However the interplay between CLL and the tumour microenvironment will be more appropriately examined using bone marrow samples from the OXPLORED study, which will remain unsorted.

Despite the limitations of my small dataset, some findings merit consideration. Mirroring other published works, variants occurred more frequently in u-IGHV CLL (Mansouri et al. 2023; Brieghel et al. 2020; Robbe et al. 2022; Knisbacher et al. 2022). In the LYM1 cohort, no variant was detected in 11/17 (64.7%) patients contrasting with other targeted deep sequencing studies where 70-73% had a variant detected (Brieghel et al. 2020; Kleinstern et al. 2020). The LYM1 cohort were patients selected for indolent disease, potentially enriching for absence of mutations and may not represent the broader early CLL population. Although panel size was similar (56 genes versus 59 genes), the approach to variant calling differed. I reported variants described in haematological cancers on COSMIC, an approach modelled on the ERIC in HARMONY report which facilitated consistent reporting between laboratories (Mansouri et al. 2023). The OXPLORED cohort, all diagnosed within three years of enrolment had variants in four of five patients, paralleling other studies (Brieghel et al. 2020; Kleinstern et al. 2020).

Four of 17 LYM1 patients and one OXPLORED patient had *NOTCH1* variants at baseline, higher than other published series of early CLL (Hu et al. 2019; Rasi et al. 2012; Knisbacher et al. 2022), potentially reflecting the higher proportion of u-IGHV patients. All *NOTCH1* variants occurred in u-IGHV samples. Baseline *NOTCH1* variants were detected at VAF > 0.10 and no impact on TTFT was observed in univariate analysis (p=0.11). Consensus on the influence of *NOTCH1* mutations on TTFT is lacking, perhaps reflecting the co-occurrence of u-IGHV and other mutations. Some studies reported *NOTCH1* mutated CLL had a shorter TTFT (Nadeu, Delgado et al. 2016, (Kleinstern et al. 2020) but others did not find an association (Condoluci and Rossi 2019). Stratifying by IGHV status, *NOTCH1* mutations were associated with a shorter TTFT only in m-IGHV (Mansouri et al. 2023).

Mutations affecting *TP53* and *SF3B1* were absent from the baseline samples, like other cohorts of early CLL (Rossi, Bruscaggin, et al. 2011; Hu et al. 2019). One progression sample (BM1596) carried an inactivating *TP53* mutation, (p.Lys132Arg NM_000546.6:c.395A>G) affecting the DNA binding domain.

Other studies report that the number of CLL-drivers independently impacts on clinical outcomes, when adjusting for clinical stage, IGHV status and CLL-IPI (Landau et al. 2013; Puente et al. 2015; Kleinstern et al. 2020). Tumour mutational burden (defined as number of somatic mutations on a 59 gene panel) was associated with shorter TTFT in CLL and MBL, after adjusting for CLL-IPI (Kleinstern et al. 2020). Brieghel et al further refined this, reporting the number of signalling pathways affected by driver gene mutations negatively impacted on TTFT (Brieghel et al. 2020). Small numbers and few mutations limited the ability to replicate these findings.

No consistent change in VAF was observed in the paired samples sequenced. Possible changes in the proportion of clonal B-cells between the baseline and the follow-up samples could be a confounder for the LYM1 cohort, but as the paired OXPLORED samples were sorted CD19+ B-cells with > 90% purity, this does not fully explain the observation. Expansion of variants was seen without clinical progression and progression occurred with a stable variant profile. Similar heterogeneity was observed using WES on sequential samples from eight MBL patients; stable *BIRC3* and *FBXW7* mutant allele fractions occurred with clinical progression and expansion of mutant allele fractions without treatment requirement after five years follow-up was observed (Ojha et al. 2014). My data does not support the hypothesis that progression in MBL/CLL is mediated by expansion of a clone carrying a driver variant.

My findings suggest that detection of variants alone is insufficient to reliably identify people with MBL and early CLL who will experience early disease progression and that alternative approaches are needed. Compared to solid tumours, CLL has a relatively low mutational burden. Somatic mutations are not detected by TGS in half of MBL (Kleinstern et al. 2020) and one third of CLL (Brieghel et al. 2022). Over half of candidate driver mutations occur in <2% of CLL, arguing for whole genome approaches. WES reported 59/82 drivers occur in <2% patients (Knisbacher et al. 2022). Going forward with OXPLORED, these results suggest that deep targeted sequencing approaches are unlikely to yield sufficient data to discriminate early progressors.

Whole genome approaches facilitate the detection of wide range of somatic variants including non-coding variants. Whole genome sequencing (WGS) enables the simultaneous

detection of multiple prognostic parameters. CNAs, coding and non-coding variants, IGHV status, telomerase and mutational signatures can be obtained from a single test. Non-coding variants affecting CLL-driver genes are detectable in MBL and CLL, where they are implicated in TTFT and OS (Puente et al. 2015; Larrayoz et al. 2017). Our group used non-negative matrix factorisation to integrate multiple biomarkers from WGS datasets of CLL requiring treatment into prognostics subgroups (Robbe et al. 2022). Previous WGS studies of MBL and early CLL reported a 2400-3449 variants per sample, sufficient somatic mutations per patient to perform molecular subgrouping (Agathangelidis, Ljungstrom, et al. 2018; Robbe et al. 2022). WGS of the OXPLORED dataset will extend the value of WGS-derived biological subgroups into the MBL/early CLL space, identifying at diagnosis people with an adverse mutational profile. Based on down sampling and saturation analysis reported by Lawrence et al (Lawrence et al. 2014) and assuming 1 mutation/Mb for CLL samples, the OXPLORED sample size would enable detection of ≥ 90% tumour mutations occurring in ≥ 2% samples with 90% power.

Prospective validation of prognostic scores for early CLL is a key goal for the OXPLORED study. Assuming the proportion of newly diagnosed CLL with high risk IPS-E (20%) and their rate of progression (14% at one year) mirrors the original Condoluci et al. publication and assuming a sensitivity of 0.76, the full OXPLORED study population (n=800) will have an 82% power to determine if IPS-E correctly identifies progressors. The study collects clinical examination findings, β2M, LDH at enrolment and integrates it with molecular data to allow calculation of the IPS-E, CLL-IPI and CLL-WONT scores. Ongoing work using samples from the OXPLORED study aims to integrate clinical parameters with WGS-derived prognostic subgroups. A reliable prognostic tool to identify at diagnosis, CLL patients who progress quickly would provide a foundation for an early intervention study.

Chapter 4: Targeted sequencing of circulating tumour DNA as minimally invasive tool for the assessment of Diffuse Large B-cell lymphoma

4.1 Introduction

Detection of circulating tumour DNA (ctDNA) in lymphoma is a rapidly evolving area of interest. It has the advantage of being minimally invasive and, combined with sensitive NGS-based analysis can detect genetic aberrations without the need for tissue biopsy, track changes in response to treatment and anticipate clinical relapse.

Potential clinical applications of ctDNA analysis in lymphoma patients include identification of tumour mutations and CNAs, sequential ctDNA quantification to stratify treatment response, monitoring residual disease after therapy and tracking emerging treatment resistance mutations.

Cancer Personalised Profiling by Deep Sequencing (CAPP-seq) is a targeted capture deep NGS method (Newman et al. 2016b; Scherer et al. 2016). CAPP-Seq employs a set of targets chosen to cover recurrent mutations in the tumour. The targets are amplified and sequenced in the cell-free DNA facilitating the quantification of ctDNA burden from somatic variants and the identification of a personalised tumour profile. DLBCL ctDNA genotyping using CAPP-Seq has reported a recovery rate more than 90% of tumour mutations. Mutations absent from the ctDNA tend to occur a < 0.05 AF in the tumour (Rossi et al. 2019).

Circulating tumour DNA assessment of disease burden, mutations and prognostic scores

In DLBCL, ctDNA burden measured as human genome equivalents (HGE/mL) correlate with International Prognostic Index (IPI) and metabolic tumour volume (MTV) by PET-CT (Scherer et al. 2016; Camus et al. 2021; Meriranta et al. 2022; Kurtz et al. 2018; Alig et al. 2021). Quantification of ctDNA can predict survival independent of IPI and distinguish between histological variants and molecular subtypes of DLBCL (Meriranta et al. 2022).

Concordance between mutations in the diagnostic tissue biopsy and the ctDNA profile is observed in >80% of cases, despite the limitations of FFPE tissue biopsies and intra-tumour heterogeneity (Rossi et al. 2017; Scherer et al. 2016). Using capture based sequencing and a bespoke gene panel of 125 genes, the ctDNA profile distinguished activated B-cell (ABC) and germinal centre B-cell (GCB) subtypes of DLBCL but also identified MDC/CD5-like and EZB/C3-like DLBCL subgroups (Meriranta et al. 2022).

Circulating tumour DNA assessment of response and early detection of relapse

Early dynamic changes in ctDNA offer a minimally invasive tool to assess response (Rossi et al. 2017; Roschewski et al. 2015). A 2log reduction in ctDNA after two R-CHOP cycles identified patients with prolonged disease control (Kurtz et al. 2018). In individual ibrutinib treated DLBCL patients, BTK drug-binding site mutations were detectable in ctDNA prior to relapse illustrating the capability of ctDNA analysis to expand tracking of emerging disease to nodal malignancies (Scherer et al. 2016).

4.1.1 Targeted sequencing of ctDNA on the ACCEPT clinical trial

In this chapter, I describe capture-based target sequencing of plasma cell-free DNA samples from the ACCEPT trial. The ACCEPT study is a Phase 1b/2 trial of acalabrutinib in combination with R-CHOP for newly diagnosed DLBCL (Davies et al. 2020). Acalabrutinib is an oral covalent BTK inhibitor (BTKi) which causes lymphocytosis in the early weeks of treatment as neoplastic B-cells redistribute from nodal tissue into the circulation. In the Phoenix trial of R-CHOP with ibrutinib, another BTKi, patients with MCD and N1 subtypes preferentially benefitted from BTKi treatment (Wilson et al. 2021).

Using CAPP-seq, I quantify baseline ctDNA burden and explore the relationship between pre-treatment ctDNA quantification and clinical endpoints including Overall Response Rate (ORR) and Progression Free Survival (PFS). I examine, in a tumour agnostic fashion, sequential plasma samples from 31 patients to determine whether dynamic changes in ctDNA after two cycles of treatment informs clinical outcomes.

4.1.2 Hypothesis and objectives for Chapter 4

In this chapter, I explore the hypotheses that:

1. Quantification of the pre-treatment plasma ctDNA informs ORR and PFS
2. Changes in the plasma ctDNA profile during treatment inform ORR and PFS

4.2 Methods:

4.2.1 Sample selection

31 DLCBL patients enrolled in the ACCEPT trial provided germline and sequential plasma cell-free DNA samples at baseline, cycle 2 day 1 (C2D1), cycle 3 day 1 (C3D1) and end of treatment (EOT, week 27). PBMC provided the germline DNA.

Acalabrutinib solubility decreases with increasing gastric pH. Concomitant use of protein pump inhibitors (PPIs) reduces peak acalabrutinib concentration and area under the curve (AUC). The manufacturers and the ACCEPT protocol recommend avoiding PPIs however three patients were inadvertently prescribed PPI medication on study. The ACCEPT steering committee decided to exclude these patients from the exploratory pharmacokinetic studies and from the outcome analysis. As the PPI interaction occurred after C3D1 and thus after submission of the plasma samples of interest, I have included the three PPI exposed patients in the ctDNA assessments (Figure 4.2.1.2).

Non-cancer cell-free DNA control samples were obtained from the Suspected CANcer (SCAN) pathway for the prompt investigation of potential cancer symptoms (Nicholson et al. 2018). Controls were selected from participants found to be cancer-free after clinical assessment, screening bloods and CT TAP.

Figure 4.2.1.1 Schematic for ACCEPT trial

Adapted from the ACCEPT protocol v5 8-Nov-2018

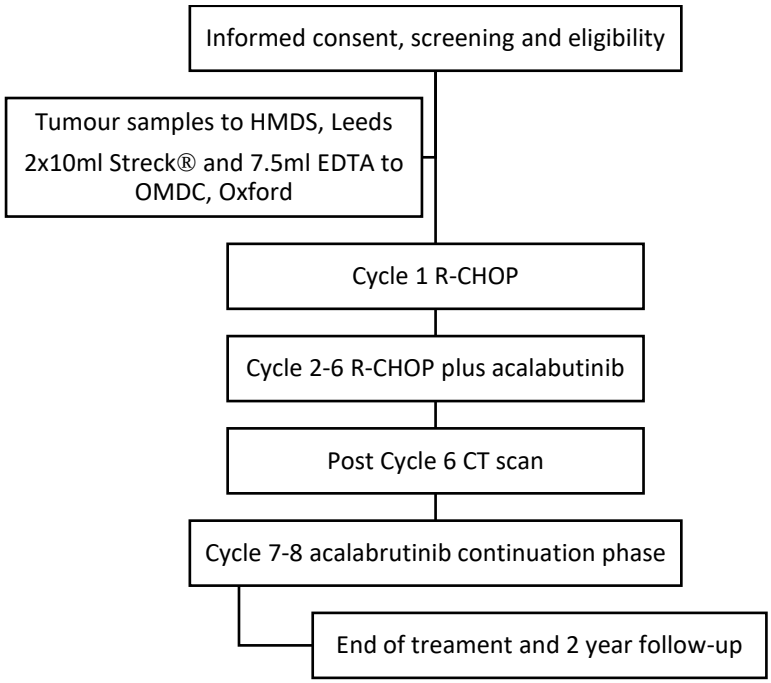


Figure 4.2.1.1: Schematic of the ACCEPT trial. HMDS, Haematological Malignancy Diagnostic Centre, OMDC Oxford Molecular Diagnostics Centre, R-CHOP rituximab, cyclophosphamide, doxorubicin, vincristine and prednisolone

Figure 4.2.1.2. ACCEPT patient and sample flow-chart

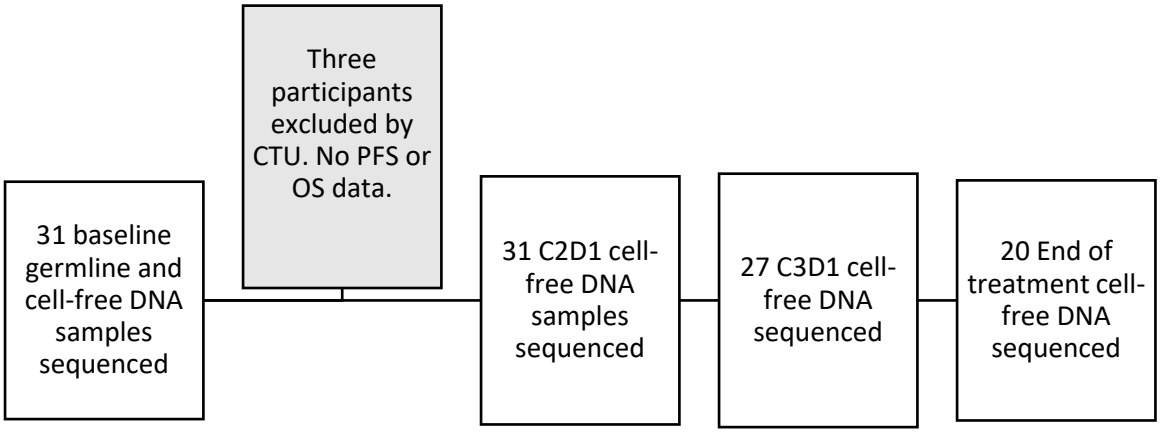


Figure 4.2.1.2: Schematic of samples sequenced at each timepoint. CTU = clinical trials unit; C2D1 = Cycle 2 Day 1 of treatment; C3D1 = Cycle 3 Day 1 of treatment; PFS = progression free survival; OS = overall survival

4.2.2 DNA extraction, library preparation and sequencing

DNA extraction, quality assessment, library preparation and sequencing are described in Section 2.4-2.6. Briefly, germline and cell-free DNA libraries were generated using the KAPA library Preparation Kit (Kapa Biosystem) following the manufacturer's instructions, with minor modifications. 200µg of genomic DNA was used for each germline library. A minimum of 32ng of cell-free DNA was used for each cell-free DNA library.

Target enriched libraries from ACCEPT patients and controls were sequenced on the NextSeq 550 with paired-end 2x151bp read length. Data analysis was performed using the bioinformatics pipelines developed at the OMDC (Section 2.6.9). As both vardict and mutect2 variant callers reported abundant variants in the control samples, only the output from varscan2 was retained for further analysis.

4.2.3 Post-sequencing analysis and variant filtering strategy

Using the annotated varscan files, I filtered variants aiming to detect low frequency variants while minimising false positives and sequencing artefacts, using criteria described in Section 2.6.10. Variants fulfilling the following criteria were retained; Quality Score > 30; non-synonymous variants, Fischer Strand bias score < 60 and Minor Allele Frequency < 0.05. For the ACCEPT cohort the False Discovery threshold was < 0.001. Next, variants were visually inspected using Integrated Genome Viewer (IGV), discarding variants in repetitive or

homopolymer regions. All *TP53* variants were confirmed as pathogenic or deleterious using Seshat (Soussi, Leroy, and Taschner 2014).

For the 25 participants with germline samples available, subtraction of variants found in the germline sample was performed. Further filtering methods were adapted depending on the availability of matched germline samples. For each patient, I aimed to filter out germline variants and recurrent sequencing artefacts. All variants occurring in the control cell-free DNA samples and recurrent variants at AF < 0.3 in the germline samples were considered artefactual and discarded (Figure 4.2.3.1).

Quantification of ctDNA and molecular response is described in Section 2.6.11. Briefly, the ctDNA burden was calculated from the median VAF of the detected variants and the quantify of cell-free DNA (in human genome equivalents) on a per sample basis.

Figure 4.2.3.1: Schematic illustrating the variant filtering strategy for the ACCEPT cell-free DNA libraries

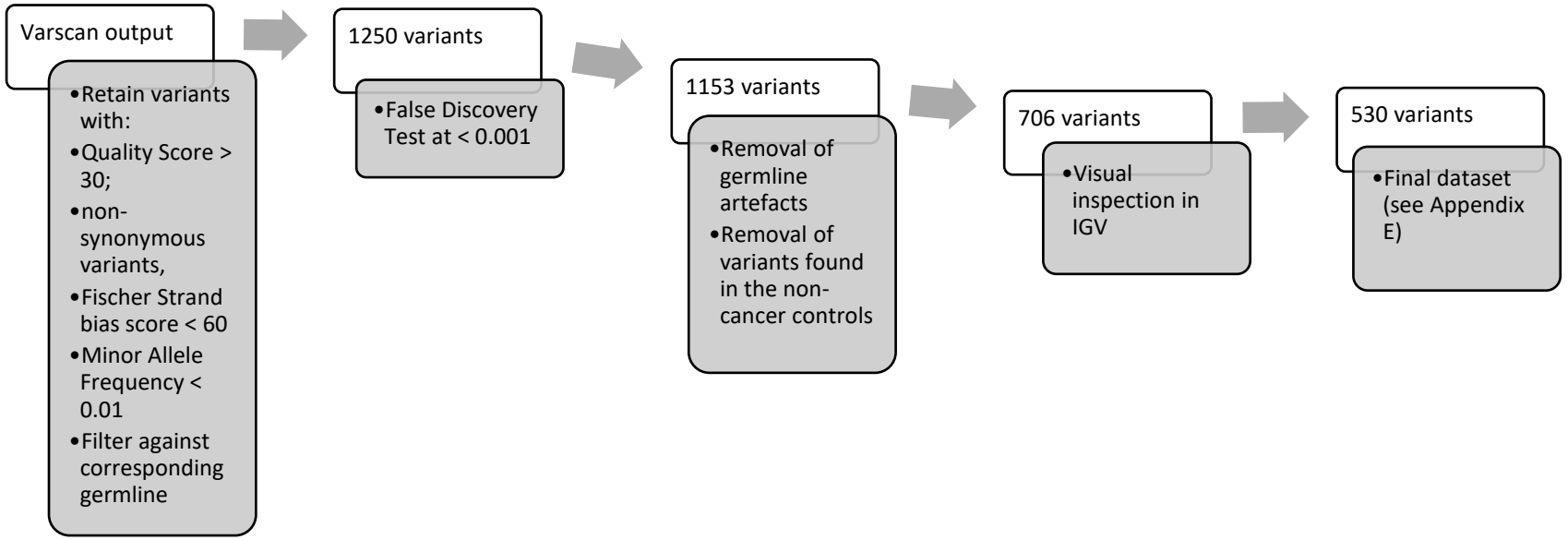


Figure 4.2.3.1 Flow chart illustrating the variant filtering steps applied to the ACCEPT cell-free DNA sequencing output. IGV = Integrated Genome Browser

4.2.4 Integration of clinical information with sequencing data

Baseline clinical data and outcomes for the ACCEPT cohort was received from the University of Southampton CTU. Outcome data was censored on 21st July 2022, when the last patient completed 24-months of follow-up.

Progression Free Survival (PFS) and Overall Survival (OS) was defined as the time from randomisation to the date of progression or death. Response was assessed using modified Cheson criteria (Cheson et al. 2014). Overall Response Rate included Complete Response (CR) and Partial Response (PR).

Relationship between ctDNA quantification and mutational profiles and baseline parameters were evaluated. Where outcome data was available, the link between dynamic changes in ctDNA values and clinical outcomes was assessed.

4.2.5 Statistical tests and data visualisation

Median, inter-quartile range or means and standard deviations were reported for continuous variables. Frequencies and/or percentages were given for binary outcomes. Variant filtering and data visualisation were performed in R using packages tidyverse, ggpubr, gridExtra and ggalluvial. Dot plots, box plots and bar plots were drawn with R package 'ggplot2'.

The Wilcoxon unpaired test was used to compare two samples. The Kruskal-Wallis test compared two or more groups. Kaplan-Meier survival curves were produced for PFS using the r-package 'survminer' (v0.4.9). The confusion matrix for the relationship between early clearance of ctDNA and clinical outcomes was applied as described in Section 2.8.

4.3 Results of the ACCEPT cohort

4.3.1. Clinical features of the ACCEPT cohort

31/38 (81.6%) patients enrolled provided pre-treatment plasma samples. The ORR was 27/31 (87.1%) for the cohort submitting plasma samples. Their principal features are summarised in Table 4.3.1.1. At a median follow up of 30 months (minimum of 24-months for survivors), 5/31 (16.1%) progressed and two died (Figure 4.3.1.1).

Table: 4.3.1.1 Clinical features of the ACCEPT cohort (n=31)

Demographics	
Gender	24 males 6 females No data supplied 1 patient
Age in years; median, range	64 years (24-78 years)
Performance status	ECOG 0 16 patients ECOG 1 11 patients ECOG 2 3 patients No data supplied 1 patient
DLBCL features	
Cell-of-origin	Non-GCB* 14 patients GCB* 14 patients Unclassified 1 patient Undetermined/test failed 2 patients
Clinical Stage	Stage I 1 patient Stage II 7 patients Stage III 8 patients Stage IV 14 patients No data supplied 1 patient
IPI-Score	Low (0-1) 6 patients Low-intermediate (2) 7 patients High-intermediate (3) 9 patients High (4-5) 8 patients No data supplied 1 patient
Extranodal disease	17 patients
Elevated LDH	20 patients
Bulky disease	6 patients
End of Treatment (EOT) response	
Complete response (CR)	23 patients
Partial response (PR)	4 patients
Stable disease (SD)	1 patient
Progressive disease (PD)	3 patients
*data on GCB and non-GCB subtypes was provided by the CRTU and based on immunohistochemical testing performed at each study site.	

Figure 4.3.1.1 Flow chart summarising outcomes for the ACCEPT cohort

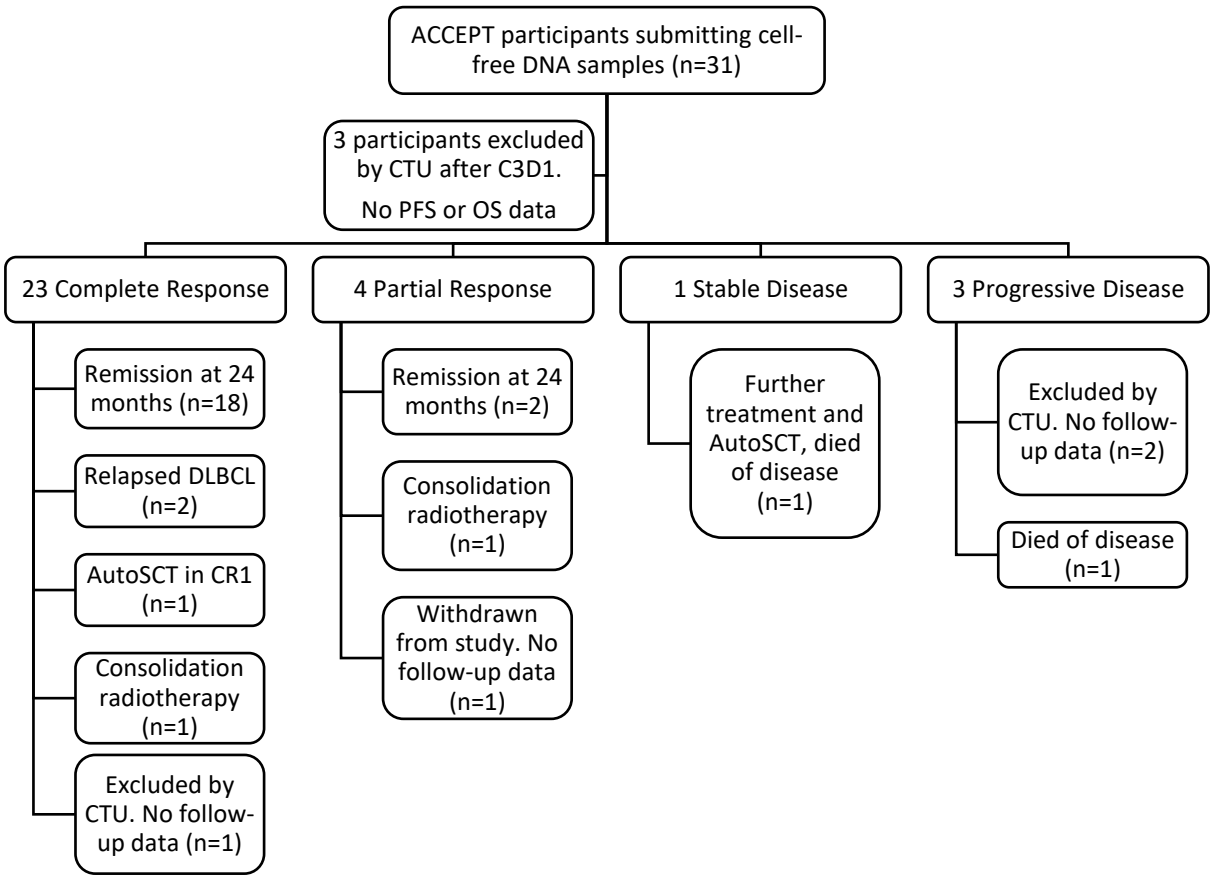


Figure 4.3.1.1. Flow diagram of the ACCEPT cohort, the responses at end of treatment and outcomes at 24 months follow-up. Responses are investigator assessed radiological responses using modified Cheson criteria after six cycles of treatment (see Section 2.4.1). Three participants were excluded by the Clinical Trials Unit (CTU) for co-prescription of proton pump inhibitor medication (n=2) and history of follicular lymphoma (n=1) respectively. No progression-free survival (PFS) or overall survival (OS) is available on these three participants and an additional fourth participant who was withdrawn from study.

4.3.2 Germline sequencing, non-cancer controls and artefacts

Twenty-five germline DNA samples were sequenced to a mean depth of 490x. 445 germline variants passed filtering. 338 variants were found at VAF < 0.30 in the germline sample and were considered artefacts. Recurrent variants at any VAF occurring 8 or more times in the germline samples were designated as artefacts.

The 338 germline variants at VAF < 0.30 were combined with recurrent variants to generate an artefact masterfile of 445 variants, used for cell-free DNA variant filtering.

Plasma cell-free samples from non-cancer controls (SCAN061 and SCAN067) were sequenced to median mean depths of 687X and 548X, respectively. Following initial filtering, 11 variants in SCAN061 and 9 variants in SCAN067 remained. As no germline sample was available for the two non-cancer controls, they were filtered against the masterfile of 445 germline variants using the same criteria as above. This left 2 variants in SCAN061 and 2 variants in SCAN067. The remaining four variants were inspected in IGV. In SCAN061, the *NOTCH2* chr1:120612046 G/T occurs in a GC rich region and was designated as artefact. The *MYC* chr8:128750607 G/C variant occurred at VAF 0.015 in a repetitive region and was designated as artefact. Both SCAN067 variants (rs41268121 and rs4645959) occurred at heterogeneous allele frequencies and were interpreted as SNPs.

All 338 germline variants at VAF < 0.30, recurrent germline variants and the two artefacts present in the SCAN controls were combined to a masterfile of 447 artefacts for downstream filtering of the cell-free DNA results.

4.3.3. Pre-treatment ctDNA burden and baseline clinical parameters (n=31)

Plasma cell-free DNA libraries were sequenced to a mean depth of 604X. Filtering was performed as described in Section 2.6.9 and Figure 4.2.3.1, retaining 530 ctDNA variants.

Baseline ctDNA was identified in 22/31 patients. The mean baseline ctDNA was 2.29 log_{10}HGE/ml (range: 1.7-4.37log_{10}HGE/ml). Nine patients had no detectable ctDNA at baseline. Since the nine patient samples with undetectable ctDNA were prepared and sequenced at the same time as patients with detectable ctDNA, lack of detectable ctDNA is assumed to be due to ctDNA levels below the limit of detection rather than technical issues.

Higher baseline ctDNA burdens, assessed as a continuous variable correlated with increasing IPI score (Figure 4.3.3.1A, kruskal-wallis, p =0.0017), elevated LDH (Wilcoxon, p = 0.00014) and presence of extranodal disease (wilcoxon, p= 0.046). The baseline ctDNA burden did not differ significantly based on clinical stage (kruskal-wallis, p = 0.13), increasing age (pearson, R = 0.15, p =0.42), gender (wilcoxon test, p =0.98), ECOG performance status (kruskal-wallis, p= 0.49), cell of origin (kruskal-wallis p = 0.67). Baseline ctDNA burden was lower in patients who went on to attain a Complete response at EOT (Figure 4.3.3.1B, kruskal-wallis, p = 0.0049).

No difference in PFS was seen between patients with undetectable ctDNA versus detectable ctDNA at baseline (p = 0.94). Treating baseline ctDNA burden as greater or less than the median value, no difference in PFS (Figure 4.3.3.2, p = 0.18) or OS (p = 0.34) was observed. The relationship between detectable ctDNA, ORR and outcome at 24 months is summarised in Figure 4.3.3.3.

Figure 4.3.3.1 Baseline ctDNA burden and IPI score (A) and response at EOT (B)

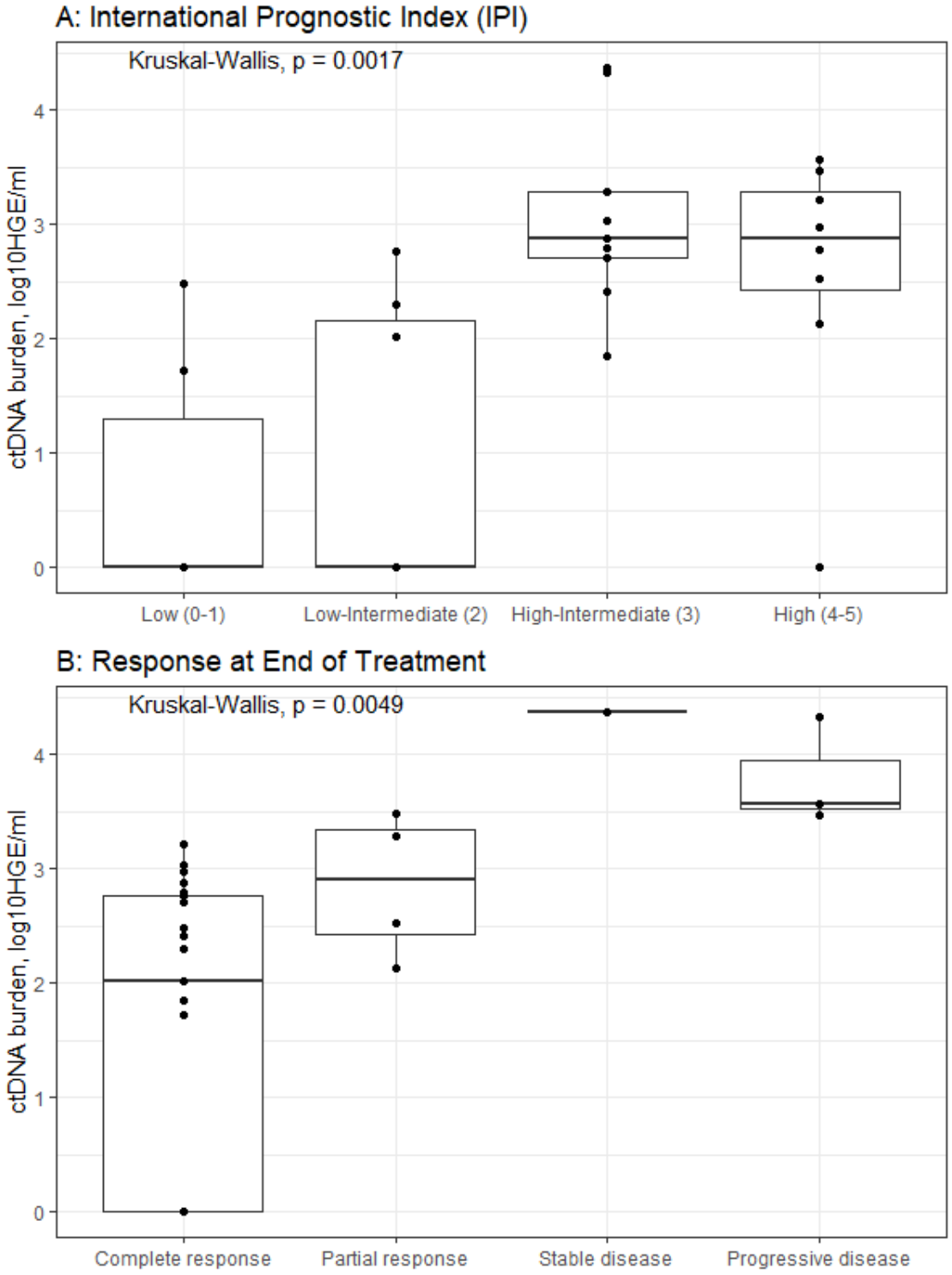


Figure 4.3.3.1: Boxplots of pre-treatment ctDNA burden (measured in HGE/ml, y-axis) across DLBCL International Prognostic Index categories (A) and response category (B).

Figure 4.3.3.2: Baseline ctDNA burden and Progression Free Survival

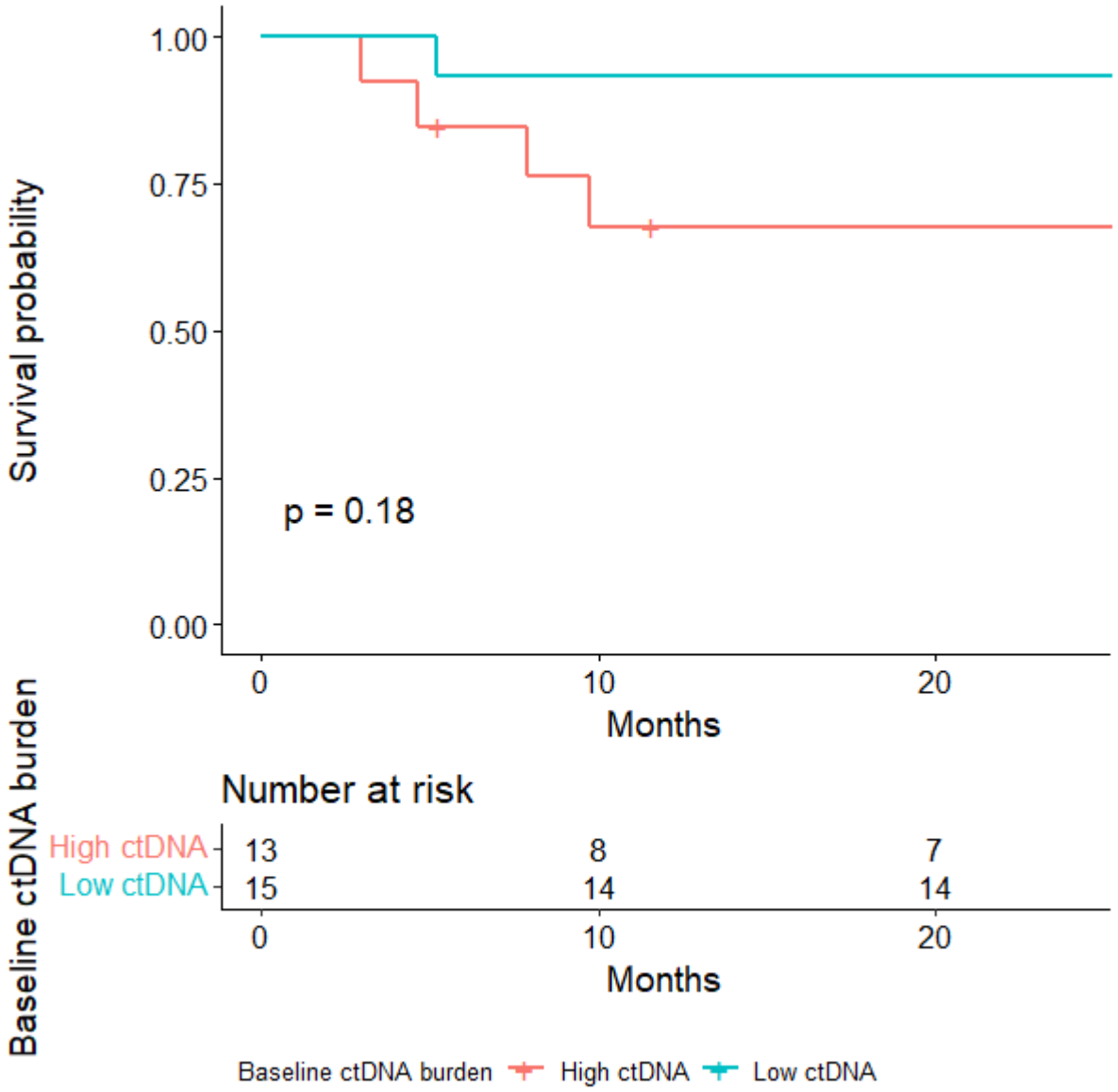


Figure 4.3.3.2: Kaplan-Meier estimates of PFS in patients above the median (High ctDNA, in red) compared with below the median (Low ctDNA, in blue). The 9 patients with undetectable ctDNA are included with the low ctDNA subgroup. No outcome data is available for 3 PPI exposed participants (n=28).

Figure 4.3.3.3. Flow diagram showing patients by ctDNA status at baseline, radiological responses after treatment and outcome at 24 months follow-up.

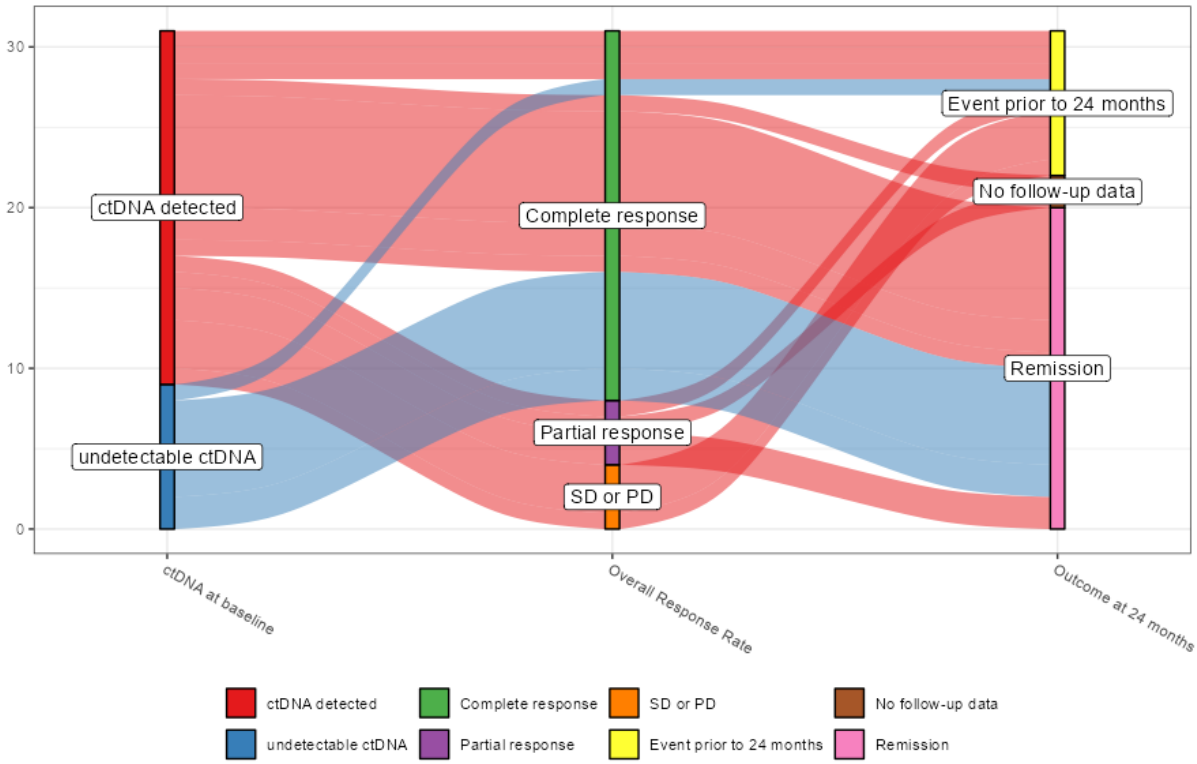


Figure 4.3.3.3. Alluvial plot illustrating the flow of participants from enrolment to 24 months follow-up. The x-axis shows key study points (enrolment, response assessment after six cycles and end of minimum follow-up). The y-axis shows numbers of patients. Blue flows represent patients with undetectable ctDNA at baseline and red flows represent patients in whom ctDNA was detected. Complete response is coloured in green, Partial response in purple and Stable disease (SD) or progressive disease (PD) in orange. Events prior to 24 months included disease progression and/or further lymphoma treatment. No follow-up data was outcome data was available for four participants.

4.3.4 Molecular profiling using plasma ctDNA

In all, 265 ctDNA variants were identified in 22 patients at baseline of which 244 were distinct. The most frequently mutated genes were *PIM1, BCL2* and *BCL6* (Figure 4.3.4.1). A median of four ctDNA variants were detected per sample (range 0-30) and the median VAF was 0.07 (range 0.008-0.44) (Figure 4.3.4.2). *TP53* mutations were detected in 8/31 (25.8%) patients (Table 4.3.4.1). No difference in PFS was seen between *TP53* mutated and wild-type patients (p = 0.87).

FFPE tumour biopsies collected as part of the ACCEPT trial were subject to targeted sequencing using a different mutational panel and gene expression profiling using whole transcriptome profiling at HMDS, Leeds. At the time of writing, no results were available for direct comparison with the ctDNA results.

Figure 4.3.4.1 Baseline ctDNA variants in each gene, displayed by frequency (n=22)

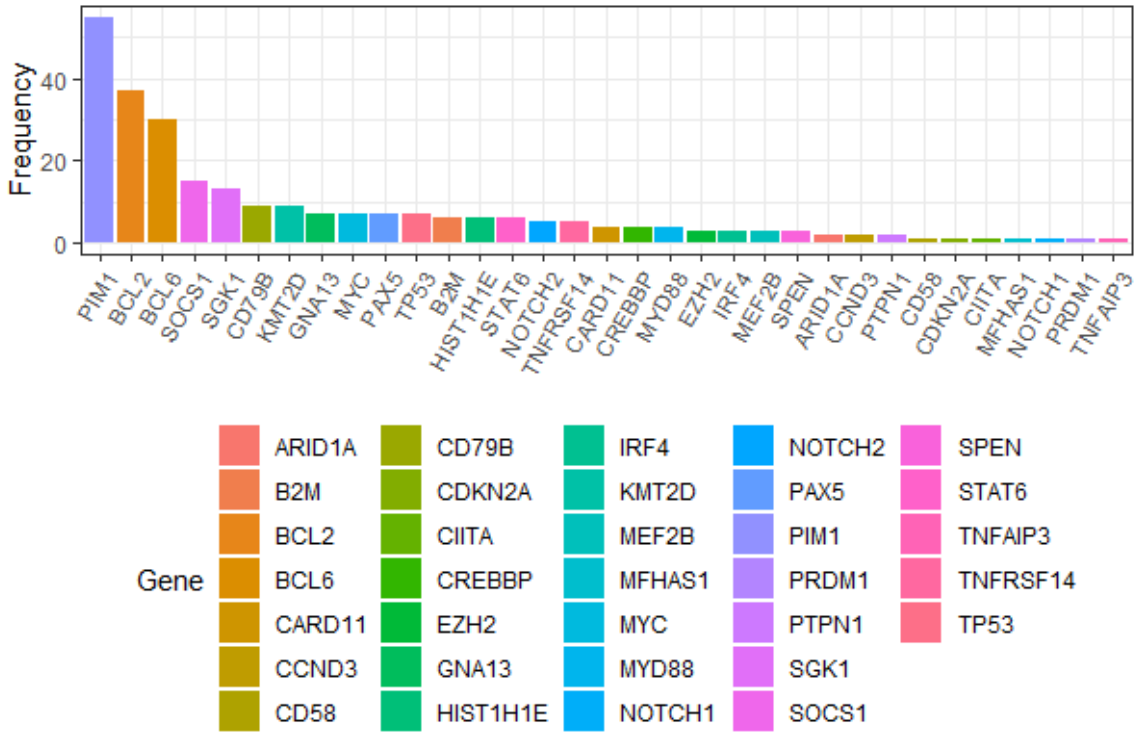


Figure 4.3.4.1: Plasma ctDNA variants detected at baseline (n=22) by gene (x-axis) and frequency (y-axis). No variants were detectable in 9 patients. More than one variant per gene per patient could be detected the six most frequently mutated genes.

Table 4.3.4.1 Plasma ctDNA *TP53* variants

Trial ID	HGVSc nomenclature	Protein	Type	COSMIC ID	ctDNA VAF
10010006	NM_000546.6:c.724T>G	p.C242G	missense	COSV52660956	0.35
10010014	NM_000546.6:c.97-4_131del	p.S33fs	frameshift	--	0.08
10090001	NM_000546.6:c.746G>T	p.R249M	missense	COSV52666526	0.23
10090025	NM_00546.6:c422G>A	p.C141Y	missense	--	#0.58
10090036	NM_000546.6:c.524G>A	p.R175H	missense	COSV52661038	0.16
10400008	NM_000546.6:c.784G>A	p.G262S	missense	COSV52841309	0.05
10460004	NM_000546.6:c.661_662insT	p.Q221VfsTer4	frameshift	COSV52977274	0.38
10460035	NM_000546.6:c.878_897del	p.G293AfsTer6	frameshift	--	0.11

#variant absent from germline sample

Figure 4.3.4.2 Baseline ctDNA variants by patient

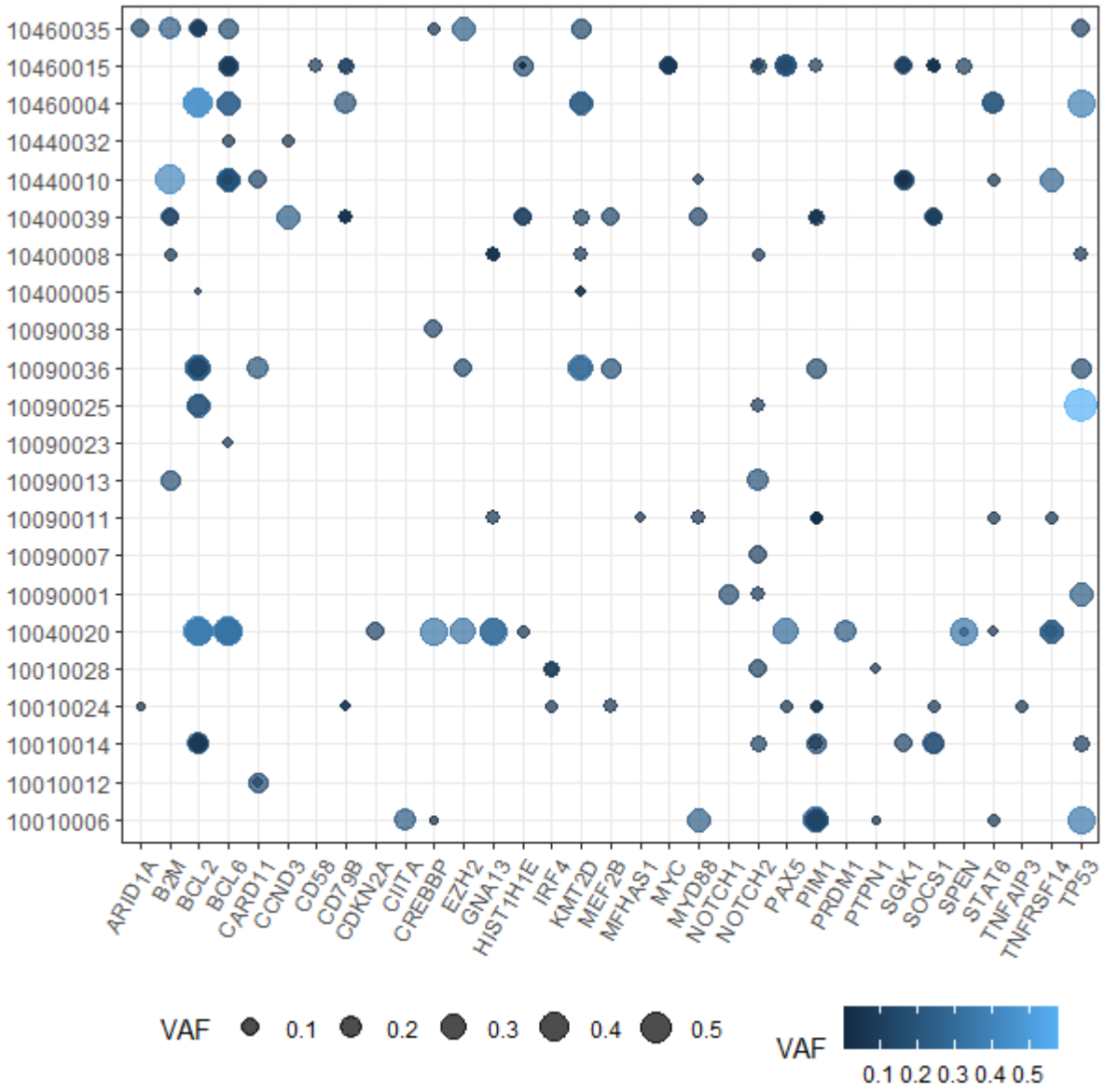


Figure 4.3.4.2: Plasma ctDNA variants by Gene (x-axis) and Trial ID (y-axis). VAF is displayed as point size and colour. The 9 patients with undetectable ctDNA at baseline are not displayed.

4.3.5 Early eradication of plasma ctDNA in sequential samples

To describe how changes in ctDNA reflect treatment responses, I examined sequential plasma samples collected after one cycle (C2D1), two cycles (C3D1) and at the EOT (week 27, Day 189). Early Molecular Response (EMR) is defined as $2\log_{10}$ reduction in ctDNA burden prior to C2D1 and Major molecular response (MMR) as attainment of a 2.5-$\log_{10}$ reduction in ctDNA burden by cycle 3 (Kurtz et al. 2018).

Of the nine patients with no baseline ctDNA detectable, all nine achieved a CR and remain in remission with no further treatment at last follow-up. One participant with no baseline ctDNA detectable, 10460022, was withdrawn from acalabrutinib treatment on study at 74 days due to safety concerns for participants over 65 years following publication of the PHOENIX trial (Wilson et al. 2021).

After one R-CHOP cycle, 16/22 patients had detectable ctDNA. Two of these sixteen patients (10090011 and 10460035) attained an EMR and both attained CR and experienced durable remissions. After two cycles of treatment (R-CHOP & R-CHOP+A), two additional patients (10090001 and 10090036) reached MMR. Patient 10090001 recorded a CR but received consolidation radiotherapy at Day 131. Patient 10090036 had undetectable ctDNA by C3D1 but ctDNA re-emerged by Day 189. This patient had SD at EOT and proceeded to salvage R-GDP and BEAM autologous stem cell transplant (AutoSCT) but died of disease 11 months after enrolment.

Taken together, all 9 patients with no baseline ctDNA, both patients in EMR and 2/3 patients in MMR (n=13) attained CR (Figure 4.3.6.1, Figure 4.3.6.2). Conversely, of the 18 patients not reaching EMR or MMR, 11 had CR and four had PR. By grouping the 9 patients with undetectable ctDNA at baseline with patients eradicating ctDNA at early timepoints (EMR and MMR), the positive predictive value of undetectable ctDNA in identifying responding patients is 92.3% (CI 67.6%-98.6%)(Table 4.3.5.1).

Table 4.3.5.1: Confusion matrix for dynamic ctDNA changes and EOT response

	Response at EOT (CR or PR)	**Non-response at EOT (SD or PD)**	**Patients**
Molecular response (EMR +/- MMR)	12*	1	13
No molecular response (detectable ctDNA)	15	3	18
Number of patients	27	4	31
Sensitivity 44.4% (range 25.5-64.7%) Specificity 75.0% (range 19.4-99.4%) **PPV 92.3% (67.6% - 98.6%)** NPV 16.7% (9.4%-27.9%)			
*9 patients with no detectable ctDNA at baseline were included with Molecular Response			

No significant different was observed with respect to PFS for EMR (p =0.94) or MMR (p=0.1). However, there are important confounding factors. No outcome data was provided for two patients found to be ineligible, both of whom experienced early disease progression. Patients 10040020 and 10090025 were excluded based on PPI co-prescription and a history

of follicular lymphoma respectively. By removing patients with early disease progression, the capacity of early ctDNA changes to detect progression is compromised. Secondly, two patients received planned consolidation radiotherapy and two patients autologous stem cell transplant after completing trial treatment.

4.3.6 Plasma ctDNA changes and overall response

At EOT (day 189), 23 patients were in CR, 4 had PR, 1 SD and 3 PD. The last submitted samples were at EOT (day 189) from 20 patients and at C3D1 from 10 patients. In these samples, ctDNA variants were detected in all three patients with PD, one of four with PR and 7/23 patients in CR (Figure 4.3.6.1, Figure 4.3.6.2). In the PD patients, both persistent ctDNA variants throughout treatment (10040020) and rapid recurrence of ctDNA variants (10090025 and 10460004) are seen. It is unclear from data available whether Patient 10090036 with SD commenced salvage R-GDP prior to Day 189.

Patient 10010028, with PR had persistent *NOTCH2* variant (NM_024408.4:c.73+17803A>G) after two cycles of treatment and withdrew due to toxicity at day 89 on study. Of the three PR patients with undetectable ctDNA, one patient (10400008) received consolidation radiotherapy at Day 228 to site of persistent disease on imaging. Patient 10440032 and Patient 10090013 remain in remission at 30.2 and 30.8 months respectively.

Of the four patients who submitted EOT plasma samples and received consolidation, one Patient 10010014 had detectable ctDNA, hinting at a potential role in ctDNA for selecting patients for post-induction therapy.

Of the six patients in CR with detectable ctDNA at the end of treatment (Day 189), five have received no further treatment at 24-months follow-up. Patient 10010014, in CR with detectable *NOTCH2* and *PTPN1* variants at Day 189 received a consolidation AutoSCT in first remission. One additional patient (10090038) had detectable ctDNA at C3D1 but did not submit an end of treatment sample.

Figure 4.3.6.1 Dynamic changes in ctDNA by variant during treatment (n=31)

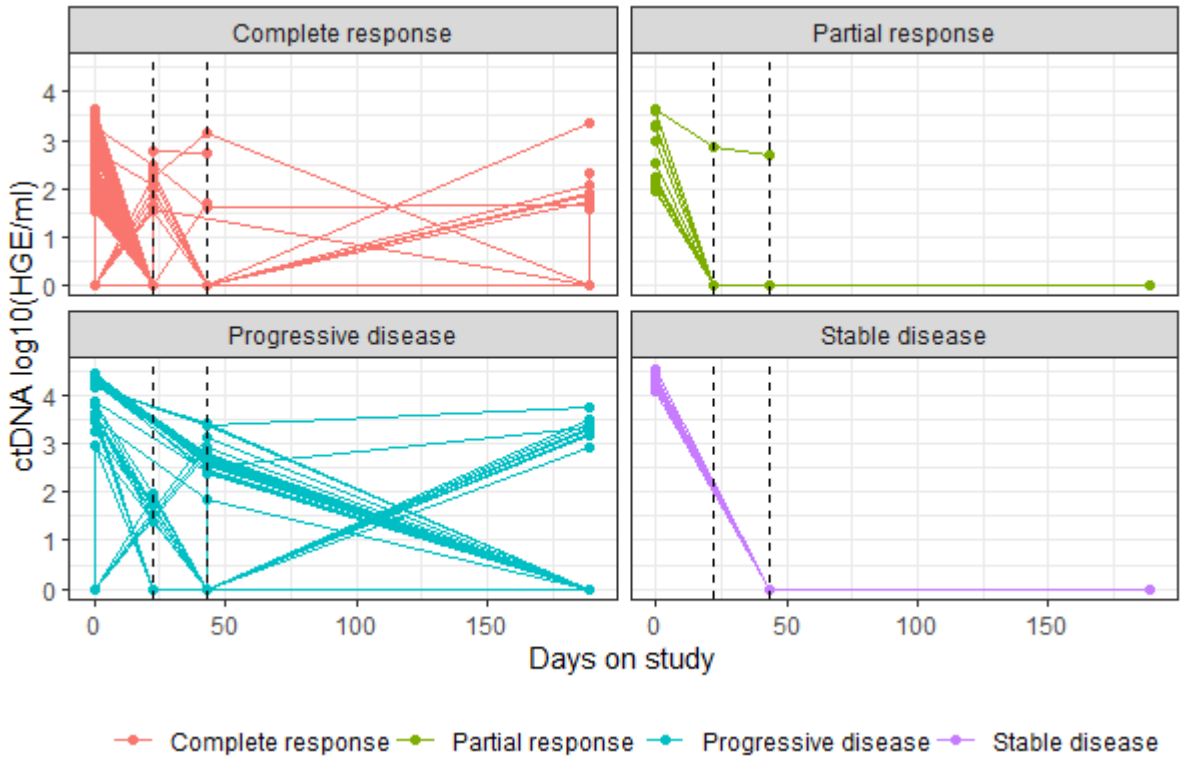


Figure 4.3.6.1: Changes in the ctDNA variants (quantified as $\log_{10}$(HGE/ml), y-axis) during treatment, displayed by Overall Response at end of treatment. C2D1 (Day 22) and C3D1 (Day 43) are illustrated using dashed vertical lines. 7/23 patients in CR had detectable ctDNA at end of treatment.

Figure 4.3.6.2 Dynamic changes in ctDNA by participant during treatment (n=31)

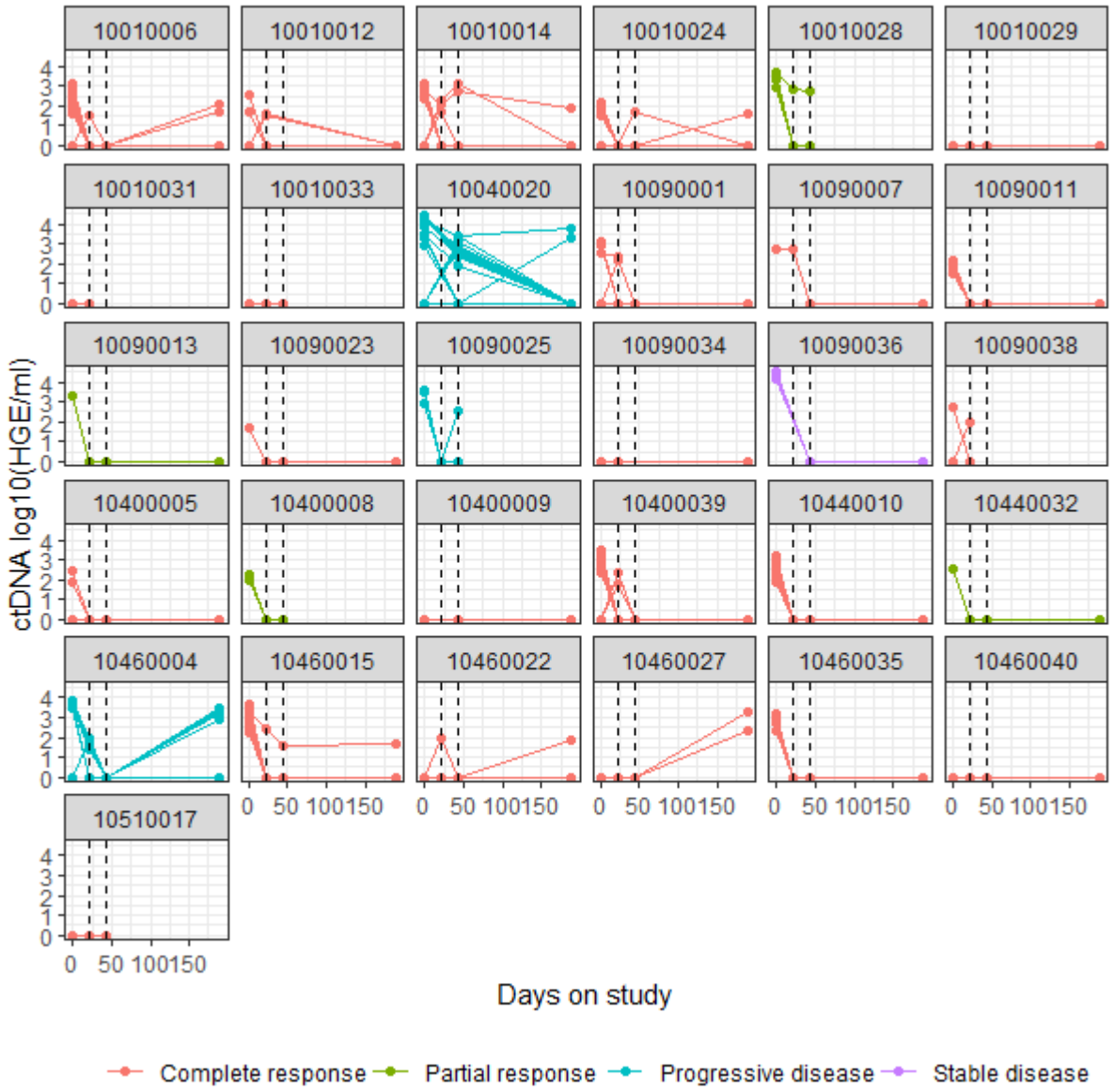


Figure 4.3.6.2: Changes in the ctDNA variants (quantified as $\log_{10}$(HGE/ml), y-axis) during treatment, displayed by Participant Trial ID. C2D1 (Day 22) and C3D1 (Day 43) are illustrated using dashed vertical lines. 6/23 patients in CR had detectable ctDNA at end of treatment (Day 189).

Table 4.3.6.1. Variants in detected in ctDNA at end of treatment, Day 189 (n=6)

Patient ID	Investigator Assessed Overall Response	Gene	VAF at enrolment	VAF at Day 189	HGVSc	HGVSp	COSMIC
10010006	Complete response	KMT2D	ND	0.0179	NM_003482.4:c.1799T>C	NP_003473.3:p.Phe600Ser	--
		CREBBP	0.0112	0.0082	NM_004380.3:c.6743_6745del	NP_004371.2:p.Gln2248del	--
10010014	Complete response	NOTCH2	0.0923	0.0778	NM_024408.4:c.73+17748C>T	--	--
10010024	Complete response	PTPN1	ND	0.0198	NM_002827.4:c.963T>A	NP_002818.1:p.Asn321Lys	--
10460015	Complete response	NOTCH2	0.0755	0.0703	NM_024408.4:c.73+17803A>G	--	COSV56684083
10460022	Complete response	HIST1H1	ND	0.0373	NM_005321.3:c.*17A>C	--	--
10460027	Complete response	STAT6	ND	0.0211	NM_001178078.2:c2453T>G	NP_001171549.1:p.Leu818Trp	--
		PIM1	ND	0.2182	NM_002648.4:c.-221A>C	--	--

ND not detected

4.4 Discussion

This work demonstrates ctDNA based assessment of B-cell malignancies is feasible in clinical trial populations. The work informs the future studies, including the ctDNA analysis plan for the NCRI REMoDL-A trial, randomising DLBCL to R-CHOP and R-CHOP plus acalabrutinib.

Minimally invasive molecular profiling

Conventional diagnostic approaches based on FFPE tissue have limited capacity to capture detailed molecular profiles and intra-tumour heterogeneity, constraining the delivery of molecular targeted treatments for lymphoma trials. In the REMoDL-B study, tissue biopsy was insufficient for molecular subclassification in 14.6%, meaning these patients could not participate in the randomisation (Davies et al. 2023). As clinical diagnostics shifts towards core biopsies, sensitive sequencing of ctDNA may circumvent some of the limitations of small tissue biopsies or allow mutation analysis when tissue is unavailable. In the work presented in this chapter, ctDNA profiling identified somatic variants in key driver genes in four patients when tissue biopsy was insufficient for molecular studies.

Capture based sequencing

With the advent of NGS technologies, sensitive DNA-based testing refines the diagnosis, informs prognosis and may select patients for specific targeted therapies. Capture-based enrichment panels using probes to hybridise the targets of interest have been developed and facilitate the sequencing of larger targets than amplicon sequencing, are less prone to bias in GC-rich and repetitive regions and in some cases, detect SV. Assays incorporating molecular barcoding or unique molecular identifiers enable detection of variants at allele frequencies < 5%. Capture-based sequencing with CAPP-SEQ has been successfully used in for minimally invasive profiling, assessment of pre-treatment ctDNA burden and changes in ctDNA on treatment for DLBCL (Kurtz et al. 2018; Scherer et al. 2016; Rossi et al. 2017). The work presented here uses molecular barcoding, as described in Rossi 2019 (Rossi et al. 2019). The depth of sequencing attained was lower than reported in other ctDNA publications, compromising the ability to detect very low frequency variants. The bioinformatics pipeline retains deduplicated reads with a minimum family size of 2 or greater. This decision was taken to minimise calling of artefactual or spurious variants. Consequently, a variable proportion of reads were discarded. Thus, the coverage achieved in the ACCEPT cohort is lower than published works using CAPP-SEQ for DLCBL (Kurtz et al. 2018; Scherer et al. 2016; Rossi et al. 2017). Alternative variant callers (vardict, mutect2) called abundant variants in the control cell-free DNA and germline DNA samples.

Pre-treatment ctDNA burden

Quantification of the baseline ctDNA burden adds additional prognostic parameters and informs response in other studies of DLBCL (Kurtz et al. 2019; Sereika et al. 2022). Higher baseline ctDNA level correlates with higher IPI scores and inversely related to response at EOT in univariate analysis in this cohort. Higher ctDNA levels are seen with elevated LDH and extranodal disease, both components of the IPI score. Given the small sample size, it was not possible to separate the effects of IPI score from ctDNA burden on clinical endpoints. In addition, patients received additional treatment with four patients receiving consolidation radiotherapy and one proceeding to AutoSCT.

Mutations in *TP53* are an emerging biomarker in DLCBL (Xu-Monette et al. 2022). Pathogenic *TP53* variants were detected in the ctDNA of 8/31 (25.8%) patients but no

difference in PFS was observed. Two previous studies of high-risk DLCBL and relapsed high-grade B-cell lymphoma respectively detected *TP53* mutations in ctDNA in 25% of patients (Bruscaggin et al. 2021; Meriranta et al. 2022), where it heralded poorer outcomes (discussed in more detail in Chapter 5). The impact of *TP53* mutations on BTKi treated DLBCL will be explored in the randomised REMoDL-A trial.

Dynamic changes in ctDNA

My work was unable to replicate previously published observations on early eradication of ctDNA and clinical outcomes (Kurtz et al. 2018; Meriranta et al. 2022). The majority of the ACCEPT patients had persistent ctDNA at C3D1 but experienced excellent outcomes. The small cohort size and the use of consolidation treatment confounds clinical endpoints.

Co-administration of the BTKi acalabrutinib, leading to early lymphocytosis and redistribution of the neoplastic lymphocytes and ongoing apoptosis of tumour cells in the lymphoid tissue may partly explain the difference in ctDNA dynamics seen. Few studies report sequential ctDNA measurements in BTKi treated lymphoid cancers and the majority examined later timepoints. Yeh et al reported that the changes in ctDNA mirrored radiological response, not lymphocytosis in ibrutinib treated CLL and RT, but did not measure ctDNA until Day 100 (Yeh, Dickinson, et al. 2017). In mantle cell lymphoma, ctDNA (quantified by median AF) falls within four weeks of ibrutinib treatment in responding patients (Agarwal et al. 2019). The LYM1002 trial studied ctDNA dynamics in ibrutinib and nivolumab treatment using a comparable capture hybridisation ctDNA sequencing method in relapsed DLBCL and other high-risk lymphomas (Bruscaggin et al. 2021). 28.6% of patients achieved a $2\log_{10}$ reduction in ctDNA burden by Day 28 and reduction in ctDNA burden informed radiological response and PFS. By contrast, over half of CLL patients treated with sequential obinutuzumab, acalabrutinib and venetoclax continued to have detectable ctDNA by ddPCR after four cycles, implying slower clearance of ctDNA with BTKi therapy although the sensitivity of the method differs (Furstenau et al. 2022). The C3D1 timepoint falls on Day 21 of acalabrutinib treatment (as acalabrutinib began at Day 22 of the protocol) which may be too early to identify clinical outcomes in BTKi treated patients.

The Meriranta study anticipates a role for post-treatment ctDNA in selecting patients for re-biopsy and/or salvage treatment. In this study, ten DLCBL patients underwent biopsy of FDG-avid masses at the EOT. Nine had undetectable ctDNA and just one patient had both detectable ctDNA and viable lymphoma on tissue biopsy. In my work, of the three patients who received consolidation radiotherapy, one attained MMR and had undetectable ctDNA at EOT, the second had detectable ctDNA and the third did not submit a plasma cell-free DNA sample. Using ctDNA results to spare lymphoma patients biopsies of post-treatment masses or select who can omit consolidation radiotherapy remains to be tested in a clinical trial.

Future directions

When tumour burden is anticipated to be low, the sensitivity of targeted NGS remains suboptimal. Duplex sequencing, phased variants and cell-free fragmentomics have attempted to reduce the limit of detection. Phased Variant Enrichment and Detection Sequencing (PhasED-seq) detects two or more variants on a single cell-free DNA strand, maximising DNA recovery while facilitate extremely low detection limits (Kurtz et al. 2021). As B-cell lymphomas accumulate mutations through the action of activation induced cytidine deaminase (AID) and the process of somatic hypermutation, the detection of

phased ctDNA variants might be particularly suitable for B-cell lymphomas. More recently, phased ctDNA variant detection was combined with ctDNA fragmentation patterns to enable low frequency ctDNA variant detection (Meriranta et al. 2022).

The ctDNA variant calling was tumour agnostic, meaning that ctDNA variants were not confirmed or refuted using the tissue biopsy. Nonetheless, identical variants were detected in sequential samples for many patients. Other studies of ctDNA in DLBCL report an imperfect correlation between ctDNA and tumour mutational profiles, due in part to intra-tumour heterogeneity. Variants in the ctDNA and absent from tumour may arise from other sources including clonal haematopoiesis of indeterminate potential (CHIP). CHIP is reported in up to 10% of older adults (Jaiswal et al. 2014; Genovese et al. 2014; Xie et al. 2014), 12% of newly diagnosed high-grade lymphoma patients (Amini et al. 2020) and 34%-43% of lymphoma patients undergoing salvage therapies (Teipel et al. 2022; Lackraj et al. 2022). The ACCEPT panel did not contain targets recurrently mutated in CHIP (*TET2, DNMT3A, ASXL1*) but this has been rectified for the STELLAR ctDNA work (see Chapter 5) and the ongoing REMoDL-A trial.

Despite the limitations of my work, exploring ctDNA profiles has promise. The storage of plasma cell-free DNA at ambient temperature facilitates the prompt transfer of samples to a centralised laboratory for analysis. In the DIRECT trial, most plasma samples arrived for central analysis within three days of phlebotomy (Dan Hodson, oral presentation 2022) which may facilitate the delivery of molecular results in a timely fashion. The lower sequencing costs associated with lower depth expands the availability of the technology Therefore, both baseline ctDNA and dynamic changes in ctDNA merit exploration in clinical trial cohorts and this work is planned for the REMoDL-A trial.

As the panel design for my thesis pre-dated the PhaseED-seq publication (Kurtz et al. 2021), there are limited AID targets included and I was unable to study phased ctDNA variants in the ACCEPT dataset. Similarly, the panel was not designed to enable detection of molecular subgroups of DLBCL (Chapuy et al. 2018; Schmitz et al. 2018). Phased variants and the detection of immunoglobulin loci breakpoints are incorporated into the targeted panel for the current randomised R-CHOP with acalabrutinib REMoDL-A trial.

Chapter 5: Targeted sequencing of circulating tumour DNA as minimally invasive tool for the assessment of Richter's Transformation

5.1 Introduction

Richter's Transformation (RT, formerly Richter's syndrome) is an aggressive, chemotherapy resistant high-grade B-cell lymphoma arising in chronic lymphocytic leukaemia (CLL) (Alaggio et al. 2022). RT occurs in 4-20% of CLL patients and frequently causes disease progression on B-cell receptor (BCR) pathway inhibitor therapy (Maddocks et al. 2015).

Ninety percent of RT share molecular lesions with the founding CLL-related clone, with mutations in *TP53* (60% to 80%), *CDKN2A* (30%), *MYC* (30%), *NOTCH1* (30%), or *MGA* (7%) most commonly identified (Rossi, Spina, et al. ; Fabbri et al. 2013; Chigrinova et al. 2013; Klintman et al. 2021). We previously performed whole genome sequencing (WGS) on paired circulating lymphocytes (PB-CLL) and tissue DNA from RT patients and observed variable relationships between the circulating and tissue-based neoplastic B-cells, including clonal expansion of subclones with mutations in DNA damage response genes and the MAPK pathway (Klintman et al. 2021). Integrating genomic and transcriptomic data from sequential CLL and RT samples, high mutational burden and frequent chromothripsis and chromoplexy were observed in RT (Nadeu et al. 2022). Mutational signature analysis revealed new mutational signatures associated with prior therapies (Nadeu et al. 2022; Parry et al. 2023b). Using single cell sequencing technologies, driver mutations and transcriptional patterns of RT were found in a minor population of CLL cells years prior to RT diagnosis (Nadeu et al. 2022).

Few studies have examined cell-free DNA in CLL and RT. Studying sequential ctDNA samples using a seven gene panel, Yeh et al reported that ctDNA, quantified as mutant allele fractions correlated with the extent of nodal disease, not lymphocytosis in CLL. An increased lymphocyte count is observed in the early weeks of Bruton tyrosine kinase inhibitor (BTKi) treatment, reflecting redistribution of neoplastic B-cells from the lymph nodes and bone marrow (BM) into the circulation. Despite increasing lymphocytosis, ctDNA declined in ibrutinib treated CLL, suggesting that dynamic changes in ctDNA might be suitable tool for monitoring on BTKi (Yeh, Hunter, et al. 2017).

Discordant results between the ctDNA and flow-cytometry based evaluations of MRD were seen in CLL. Yeh et al reported approximately 20% of timepoints had detectable ctDNA mutation without evidence of disease in peripheral blood (PB) by flow-cytometry. Furstenau et al observed that discordant results, with detectable ctDNA but PB MRD-negative by flow cytometry at 10^{-4} were more likely to be seen in patients with nodal disease, suggesting that ctDNA might offer a window into residual CLL surviving the nodes (Furstenau et al. 2022). By contrast, no additional benefit to plasma-based monitoring of immunoglobulin heavy chain sequences over PB and BM samples for residual disease assessment was observed using next generation sequencing (NGS), perhaps reflecting different sensitivities (Thompson et al. 2019). The potential role of baseline ctDNA assessment in CLL is understudied, although reports are anticipated for the CLL2-BAAG trial (Furstenau et al. 2022).

Published literature on ctDNA in RT is limited to individual patient reports and fifteen RT patients were treated with ibrutinib and nivolumab on the LYM1002 trial. In this study, sequential ctDNA samples were analysed using capture hybridisation sequencing. Mutations in ctDNA affecting *TP53* were detected in 47% of RT and persisted throughout treatment. RT patients with high ctDNA burden (defined as > 1900HGE/ml) had a shorter progression free

survival (PFS) but the difference did not reach significance and both groups did poorly (3.3 vs 8 months, p=0.07) (Bruscaggin et al. 2021).

In this exploratory study, pre-treatment plasma ctDNA from 31 RT patients enrolled in the O-CHOP and STELLAR clinical trials were interrogated using a customised targeted next-generation sequencing panel. Unlike the panel used for DLBCL ctDNA sequencing (Chapter 4), the panel used for the RT cohort included variants recurrently reported in CHIP and genes associated with resistance to BCR pathway inhibitor treatments. Sequential samples from the STELLAR trial allowed tracking of ctDNA changes following treatment. As STELLAR is recruiting, clinical outcome data was available for the O-CHOP cohort only. In contrast with the favourable outcomes described in the ACCEPT cohort (Chapter 4) but in keeping with other studies of RT, clinical outcomes were poor, underlining the need for early identification of candidates for salvage treatments in this cohort. After a median follow up of 6.5 months (interquartile range (IQR) 3.7-20.6 months), 17/21 died including 13 patients who died of RT (Eyre et al. 2016).

5.1.2 Hypothesis and objectives for ctDNA in RT

In this chapter, I explore the hypotheses that:

1. Quantification of the pre-treatment plasma ctDNA parallels clinical stage in RT and informs early radiological response (O-CHOP and STELLAR cohorts)
2. Quantification of the pre-treatment plasma ctDNA informs PFS and Overall Survival (OS) (O-CHOP cohort only)
3. Driver mutations are detectable in the ctDNA which are not detectable in the tissue (O-CHOP and STELLAR cohorts).
4. Unlike DLBCL, $2\log_{10}$ reductions in ctDNA burden on treatment will be infrequent in RT (STELLAR cohort only).

5.2 Methods:

5.2.1 Ethical approval and study populations

The O-CHOP study was a multi-centre Phase II trial of ofatumumab with CHOP for DLBCL-type Richter's transformation (RT) (Eyre et al. 2016)(Section 2.1.2).

The STELLAR trial is a prospective trial for DLBCL-type RT (Section 2.3 and Appendix B)(Appleby et al. 2019). Eligible RT patients are randomised to receive R-CHOP or R-CHOP plus acalabrutinib.

5.2.2 Plasma cell-free DNA and tumour samples

Available samples comprise 16 RT tumours and 31 pre-treatment plasma samples. Twenty-three post-treatment cell-free DNA samples were sequenced from the STELLAR cohort (See Figure 5.2.5.1). Salivary germline samples from the STELLAR cohort were sequenced. DNA extraction and QC is described in Section 2.5.

5.2.3 Sequencing target design

Panel targets were chosen to reflect variants contributing to the pathogenesis of B-cell malignancy (Landau et al. 2013; Nadeu et al. 2018; Ljungstrom et al. 2016; De Paoli et al. 2013; Fabbri et al. 2013), reported in RT (Klintman et al. 2021) or amenable to targeted therapy (Woyach et al. 2017; Jones et al. 2017; Liu et al. 2015; Mohanty et al. 2016). UTRs were included for targets where non-coding mutations impacted clinical outcomes and/or highlighted regions of kataegis in B-cell regulators (Puente et al. 2015). Four genes recurrently mutated in CHIP (*ASXL1, DNMT3A, JAK2* and *TET2*) were included to facilitate distinction between CHIP and tumour variants. (Methods 2.6.2 and Appendix D).

5.2.4 Library preparation, target enrichment and sequencing

Tumour, germline and plasma cell-free DNA libraries were prepared using the Thruplex-IDT Tag-Seq kit, with minor modifications. Library preparation, quantification and hybridisation capture are described in Section 2.6.5-2.6.7. O-CHOP cell-free DNA libraries were sequenced on the HiSeq 2500 in rapid run mode. STELLAR cell-free DNA libraries were sequenced on the NextSeq 550.

5.2.5 Post-sequencing data processing and variant filtering

Data analysis was performed using a custom bioinformatics pipelines developed at the OMDC for amazon cloud software (described in Section 2.6.9). Briefly, reads are aligned against the hg19 human reference genome using Burrows-Wheeler Aligner BWA-mem v0.7.17 (Li and Durbin 2009). Reads with identical UMI were collapsed into read 'families'. A minimum family size of two was used for deduplicated reads. Deduplication and error correction was performed using elduderino v1.2.0 https://github.com/eawilson/elduderino). Variant calling was with Varscan2 v2.4.2 (Koboldt et al. 2012). Variant annotation was done using Ensemble Variant Effect Predictor (VEP v104) (McLaren et al. 2016) and Annovar (Wang, Li, and Hakonarson 2010).

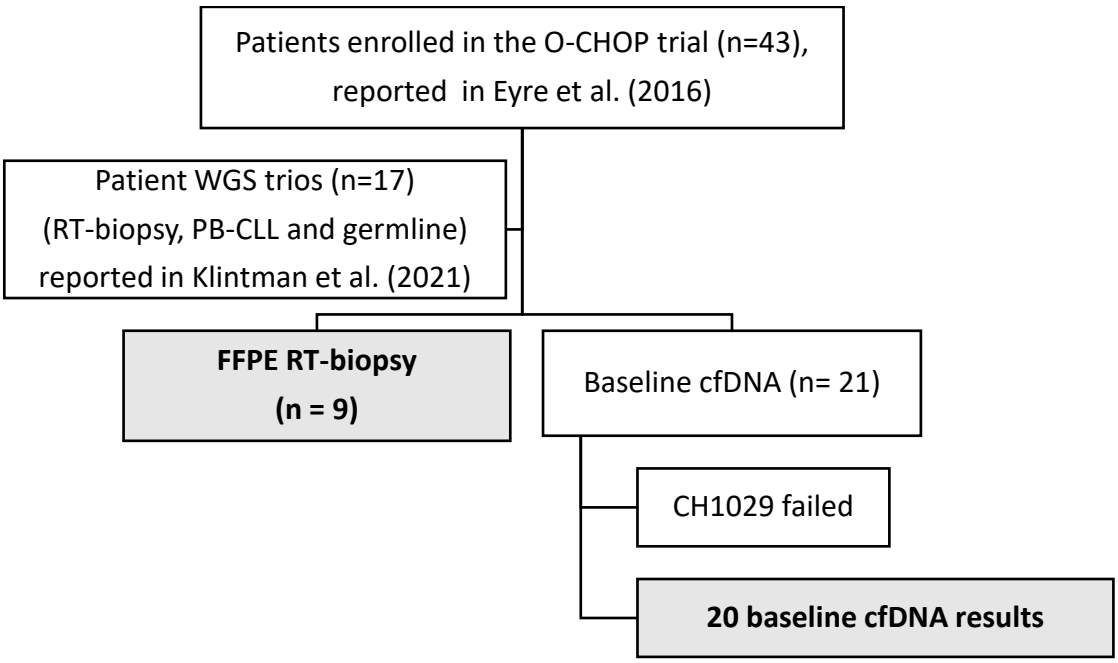
Variant filtering for the cell-free DNA libraries is summarised in Figure 5.2.5.2. As germline samples were available for the STELLAR cohort only, no tumour-germline subtraction was performed. Recurrent variants at VAF < 0.3 in the germline samples and the control normal tissue were designated as sequencing artefacts. Nine tumour samples were FFPE and seven derived from FF tissue biopsies. To account for artefactual changes with FFPE, only variants at AF > 0.05 were retained for the tumour samples.

Plasma ctDNA concentration was measured as described in Section 2.6.11. Plasma ctDNA burden was calculated by multiplying the total plasma cell-free DNA concentration (pg/ml) by the median allele fraction of the variants detected on a per patient basis at each timepoint. The ctDNA burden was reported in human genome equivalents (HGE) per ml on a log10 scale.

Pathogenicity of variants affecting *TP53* was determined using Seshat, an online tool which evaluates the pathogenicity of each variant according to either general prediction algorithms (Provean, Sift, Polyphen2, FATHMM, MutationAssessor) or algorithms developed exclusively for *TP53* (Soussi, Leroy, and Taschner 2014). Variants classed as pathogenic, deleterious and insufficient evidence for classification of this coding variant as pathogenic or benign (variants of uncertain significance) are retained for visual inspection using IGV. Visual inspection removed variants occurring in homopolymer regions and repetitive regions.

Figure 5.2.5.1: Source of Richter's Transformation tumour and plasma cell-free DNA samples

A Samples from the O-CHOP trial



B. Samples from the STELLAR trial

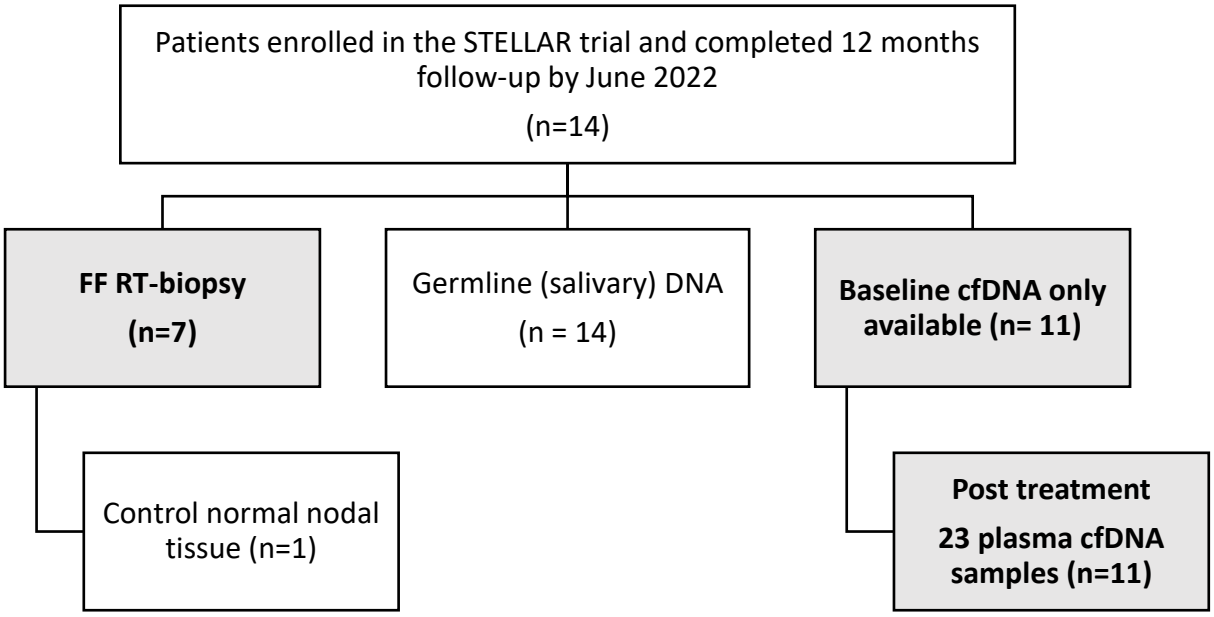


Figure 5.2.5.1: Schematic of trial populations and samples included in the Chapter 5. FF = fresh, frozen. FFPE = formalin fixed paraffin embedded. RT-biopsy = tissue biopsy establishing the diagnosis of Richter's transformation. cfDNA = cell-free DNA.

Figure 5.2.5.2: Schematic illustrating the variant filtering strategy for the RT cell-free DNA libraries

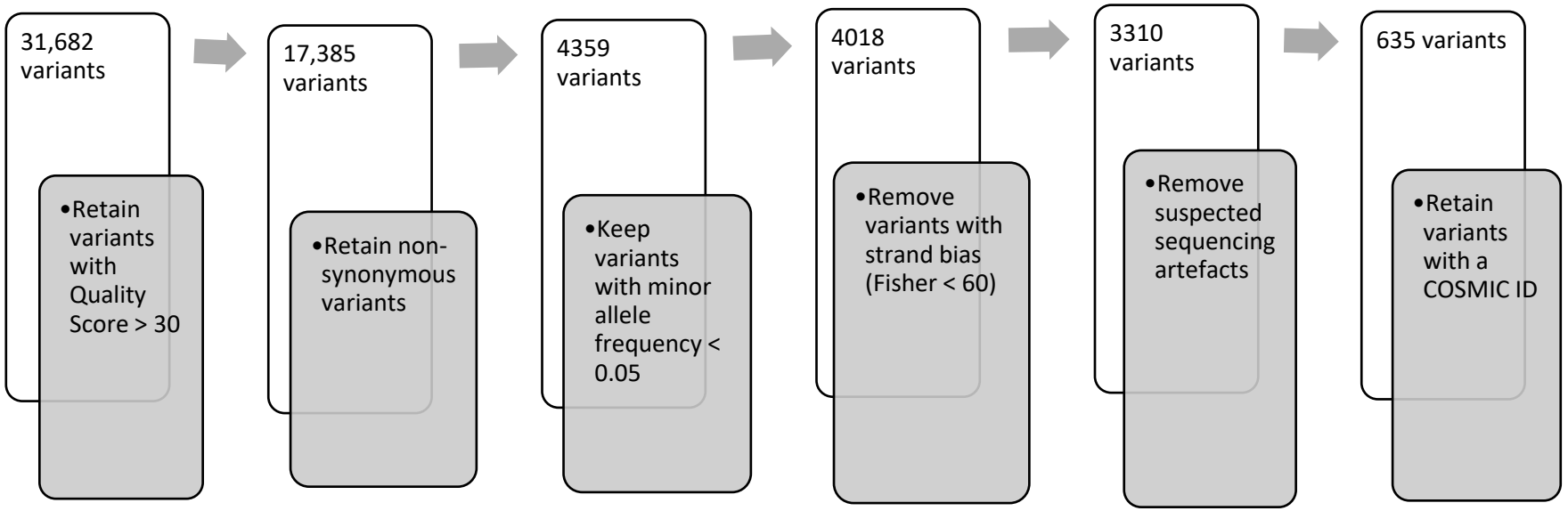


Figure 5.2.5.2 Flow chart illustrating the variant filtering steps applied to the RT cell-free DNA sequencing output.

5.2.6 Clinical parameters, response rates and patient outcomes

Baseline clinical parameters, response rates and patient specific outcomes were obtained from the Oxford Clinical Trials Office for the O-CHOP patient cohort.

As STELLAR is an ongoing trial, baseline parameters and response at interim assessment only was provided by the Clinical Trials Research Unit, University of Birmingham. Interim assessment was performed by CT-TAP after four cycles of treatment and reported according to Cheson criteria (Cheson et al. 2014). The Overall Response Rate (ORR) comprised patients attaining a Complete Response (CR) or Partial Response (PR) on imaging. Overall survival (OS) was defined as the interval between enrolment and data of death or last follow-up. Progression-free survival (PFS) was defined as the interval between enrolment and date of progression, date of death from any cause or date of last follow-up.

5.2.7 Statistical analysis and data visualisation

Median and inter-quartile range or means and standard deviations were reported for continuous variables. Frequencies and/or percentages were given for binary outcomes.

For most comparisons of clinical parameters and ctDNA burden, non-parametric tests (Wilcoxon rank sum test, Kruskal-Wallis test) were employed. The use of parametric (Pearson, r) and non-parametric (Spearman, rho) tests for correlation are indicated in the figures and text. The relationship between ctDNA and ORR is reported as an odds ratio with 95% confidence interval.

Dot plots, box plots and bar plots were drawn with R package ‘ggplot2’. The Venn diagram was drawn using the ‘ggvenn’ package.

TFS, PFS and OS distributions were plotted using Kaplan–Meier estimates and visualised using the R package ‘survminer’ (v0.4.9) and ‘survival’ (v3.4.0). Univariate analysis of survival (PFS and OS) by subgroups was performed using proportional hazards models (Cox proportional-hazards model). Forest plot visualisations of the Cox regression analyses were drawn with R package ‘forestplot’.

5.3 Results

Baseline features of the 31 RT patients providing pre-treatment plasma ctDNA are described in Table 5.3.1

Table 5.3.1 Clinical features of the RT patients submitting baseline plasma samples

	O-CHOP trial (n=20)	**STELLAR trial (n=11)**
Age in years (median, range)	64.5 years (51-90 years)	81 years (65-84 years)
Gender	14 males 6 females	8 males 3 females
Ann Arbor Stage		
Stage I	2 patients	0
Stage II	3 patients	3 patients
Stage III	7 patients	0 patients
Stage IV	7 patients	6 patients
Unknown	1 patient	2 patients
ECOG Performance Status		
ECOG 0	8 patients	2 patients
ECOG 1	7 patients	5 patients
ECOG 2	2 patients	4 patients
ECOG 3	3 patients	0 patients
Bloods at enrolment median, range		
Haemoglobin	109 g/dL (82-147g/dL)	124 g/dL (101-143g/dL)
Lymphocytes	3.1 x 10^9/L (0.4-88.5 x10^9/L)	2.1 x 10^9/L (0.5-66.4x10^9/L)
Platelets	86 x 10^9/L (18-439 x10^9/L)	210 x 10^9/L(75-357x10^9/L)
LDH	438iu/L (288-2312iu/L)	395iu (176-769iu/L)
Beta2 microglobulin	4.2mg/L (0.9-5.7mg/L)	4.1mg/L (2.3-10mg/L)
Previous CLL treatment	13 patients	5 patients
B-symptoms at diagnosis	11 patients	4 patients
Bulky disease > 5cm	9 patients	6 patients
Extranodal disease	8 patients	7 patients
Response after six cycles		
Complete Response	3 patients	ND
Partial Response	1 patient	
Stable Disease	0	
Progressive disease or treatment withdrawal prior to C6	16 patients	

**data on TP53 mutation status from the clinical data supplied to the trial management units*
ND: no data supplied

Tumour libraries were sequenced to mean deduplicated depth of 139X (IQR 78.5-165X) (Figure 5.3.1). For the O-CHOP cell-free DNA libraries, captures were performed individually for cell-free DNA libraries and libraries were sequenced on the Illumina HiSeq 2500 in rapid run mode to a median deduplicated depth of 449.2X (IQR 313-558.2X). The fragment size (mean 163bp, IQR 161-174bp) is consistent with plasma cell-free DNA, despite the O-CHOP samples being collected in EDTA. For the STELLAR cell-free DNA libraries, six post-Thruplex libraries were pooled for hybridisation and capture and libraries were sequenced on the Illumina NextSeq 550 to a mean deduplicated depth of 893X (IQR 331-1050X) (Figure 5.3.2).

The median cell-free DNA obtained from the cell-free DNA samples was 648.8 HGE/mL (IQR 221–4000). The ctDNA burden was calculated from the cell-free DNA quantity in HGE/mL and the median VAF of variants detected in the ctDNA on a per sample basis. median pre-treatment ctDNA burden was 24.86 HGE/mL (IQR range 23.6–443.2).

738 ctDNA variants were detected in 28/31 RT patients. No ctDNA was detected in three baseline samples, CH1005, CH1007 and CH1013 and was assumed to be below the limit of detection.

Figure 5.3.1 Boxplot of sequencing coverage of the Richter's tumours by gene

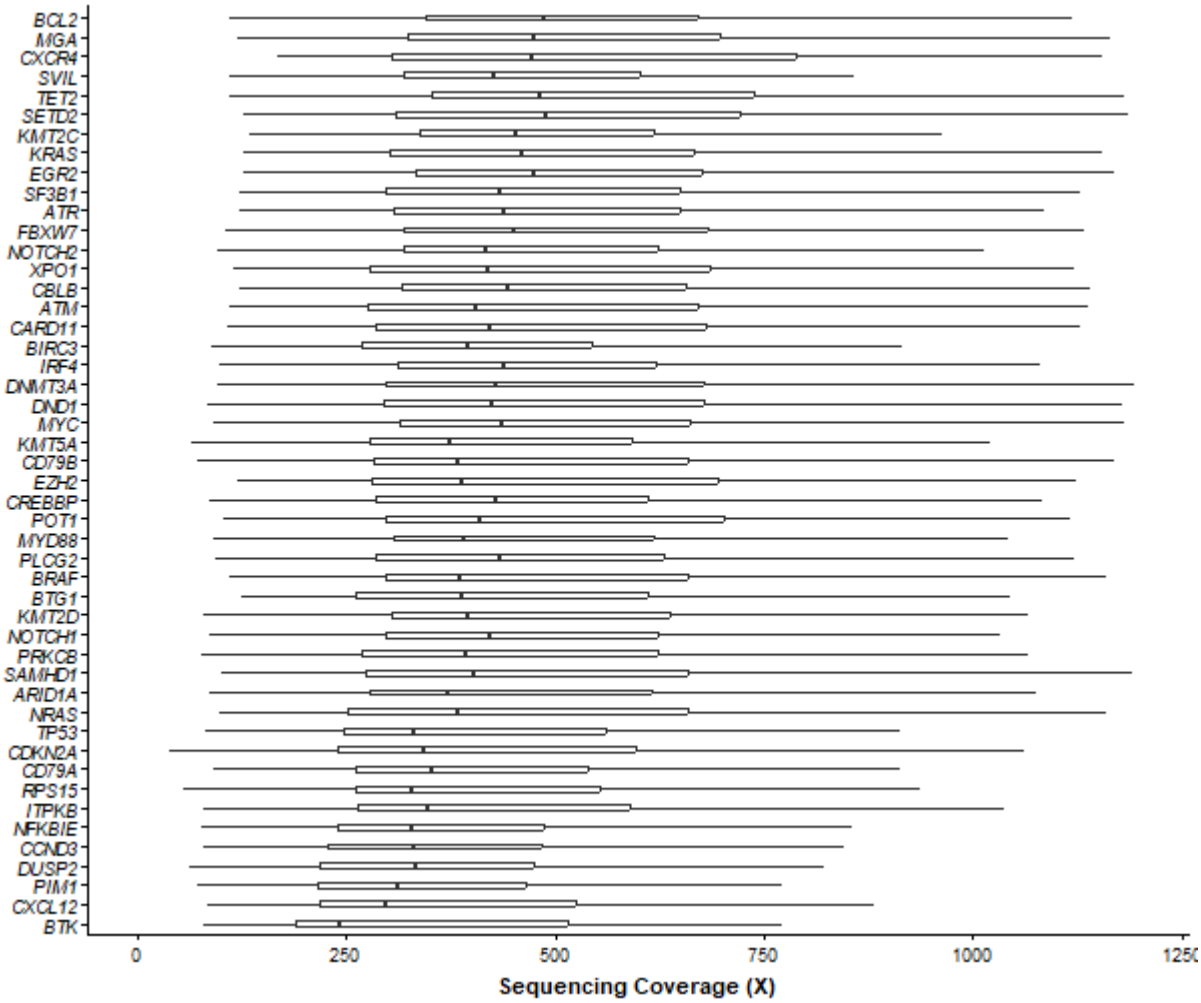


Figure 5.3.1: Sequencing coverage of target genes (y-axis). Reads were collapsed into read families sharing the same UMIs. Read families with two or more identical reads were retained. Boxplot: centre line, median; box limits upper and lower quartiles

Figure 5.3.2 Boxplot of sequencing coverage of Richter's cell-free DNA samples by gene

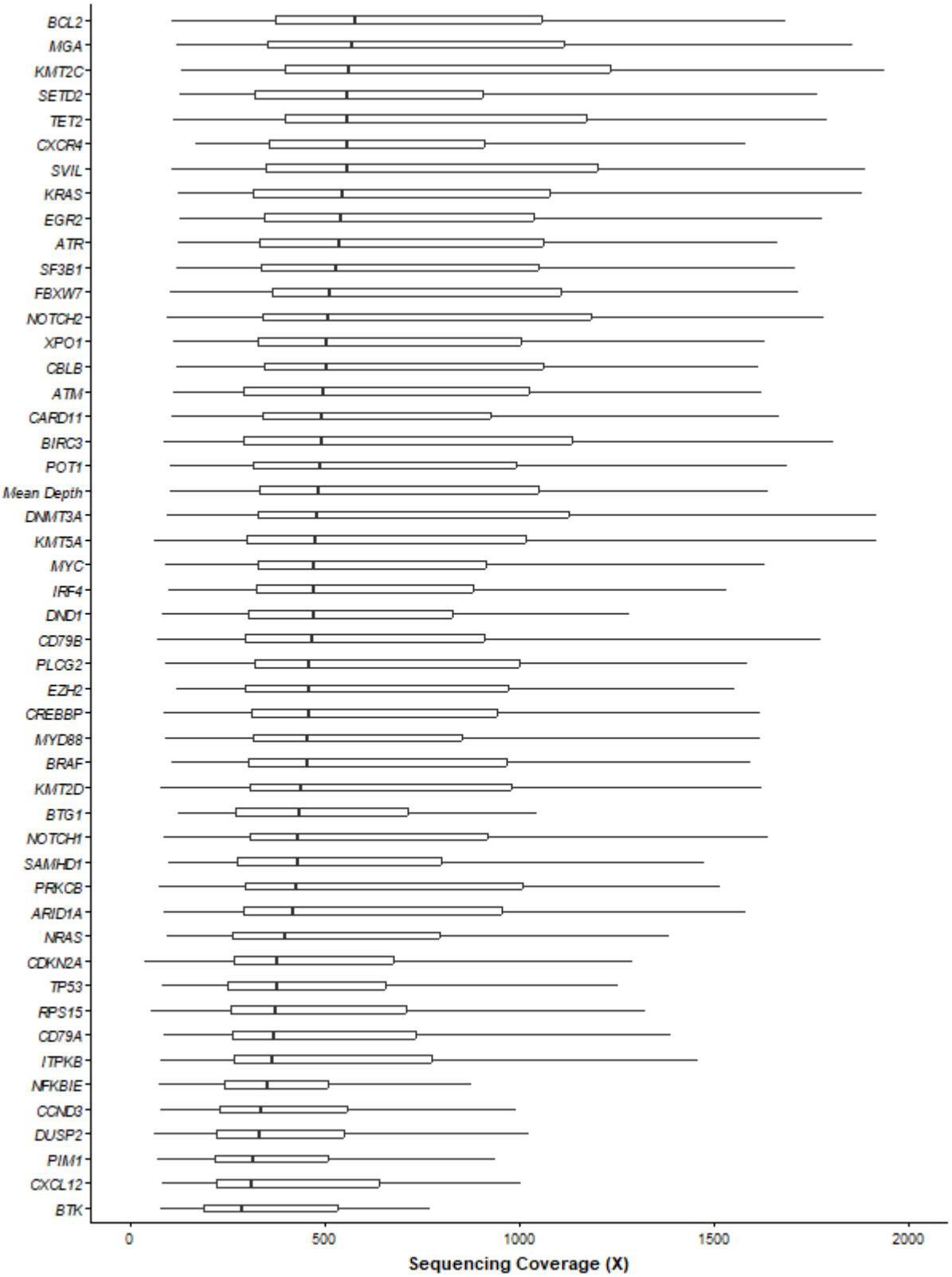


Figure 5.3.2: Sequencing coverage of target genes (y-axis). Reads were collapsed into read families sharing the same UMIs. Read families with two or more identical reads were retained. Boxplot: centre line, median; box limits upper and lower quartiles

5.3.1. Pre-treatment ctDNA burden and baseline clinical parameters (n=31)

No difference was observed in pre-treatment ctDNA burden based on gender (wilcoxon, p = 0.5), presence of B-symptoms (wilcoxon, p = 0.8), extranodal disease (wilcoxon, p = 0.19), ECOG performance status (kruskal-wallis, p = 0.2) and clinical stage (kruskal-wallis, p = 0.2) (Figure 5.3.1.1). Pre-treatment ctDNA burden ranged widely, even in patients with the same clinical stage. Within Stage III patients for example, pre-treatment ctDNA burden ranged from 0.30 to 3.27 log_{10}HGE/ml.

Increasing age was weakly inversely correlated with ctDNA burden (R = -0.39, pearson, p = 0.031). The relationship between lymphocytosis and increasing ctDNA burden is non-significant (spearman, R =0.24, p = 0.22). Pre-treatment LDH, as a continuous variable weakly correlated with ctDNA burden (spearman, R =0.46, p = 0.043).

Figure 5.3.1.1: Relationship between pre-treatment ctDNA and Clinical Stage

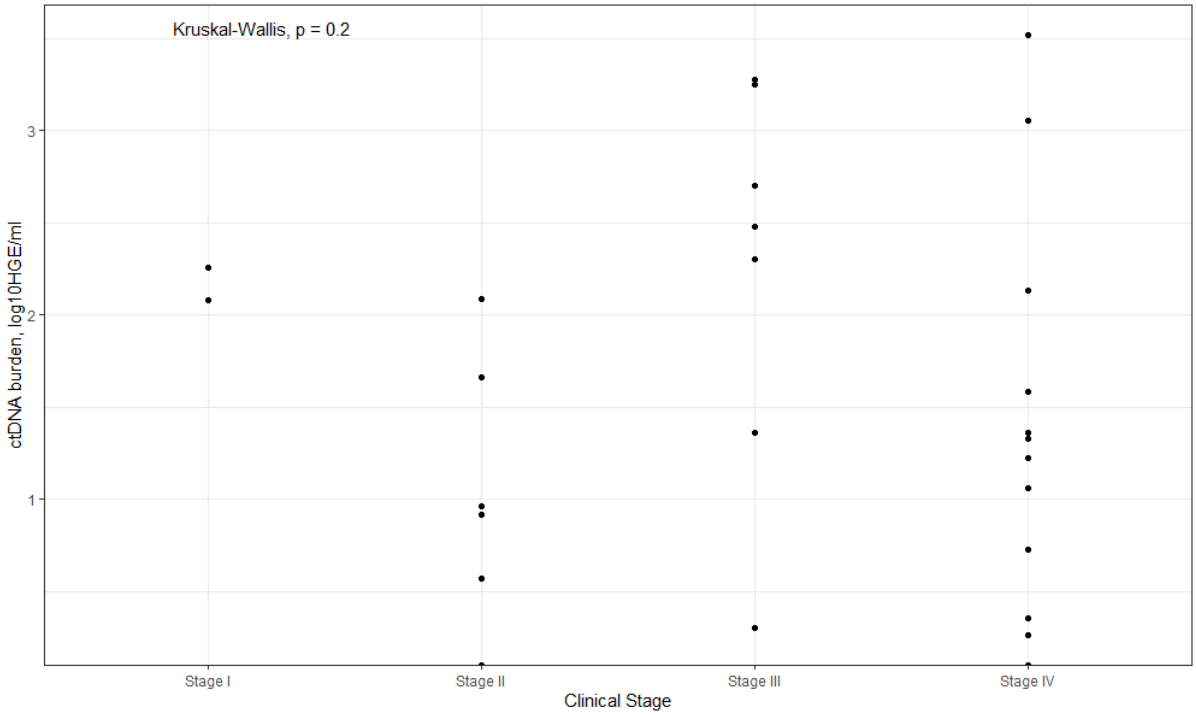


Figure 5.3.1.1: Dot plot of pre-treatment ctDNA in log_{10}HGE/ml and clinical stage. Log_{10}HGE/ml, log10-transformed haploid genome equivalents per millilitre of plasma.

5.3.2. Pre-treatment ctDNA burden and early radiological response (n=31)

Next, I examined the relationship between baseline ctDNA burden and response at interim assessment after four cycles and end of treatment (EOT). Responders included CR and PR. Non-responders had clinical progression prior to assessment or imaging showed stable (SD) or progressive disease (PD).

18/31 (58.1%) RT patients responded at interim assessment. Pre-treatment ctDNA burden, assessed as continuous variable, did not translate to response at interim assessment (Figure 5.3.2.1, OR 0.93 95% confidence interval 0.49-1.8). High pre-treatment ctDNA levels were defined as greater than or equal to than the median ctDNA value (24.86 HGE/mL or 1.4 log_{10}HGE/mL). The relationship between high pre-treatment ctDNA and response at interim assessment was non-significant (OR 1.34, 95% confidence interval 0.35-5.35, p = 0.67).

Figure 5.3.2.1: Cox proportional hazards of clinical and ctDNA features and their relationship with response at interim radiological assessment (n=31)

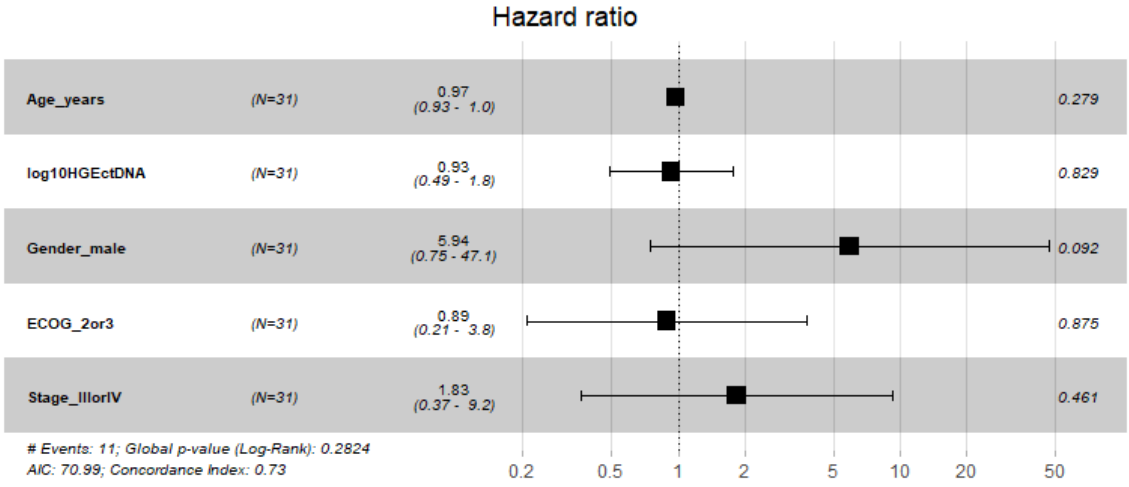


Figure 5.3.2.1: Forest plot showing the relationship between age, ctDNA burden, gender, ECOG Performance Status and advanced stage and Response at Interim Assessment.

5.3.3 Pre-treatment ctDNA burden, ORR, PFS and OS (O-CHOP, n=20)

Overall response rate (ORR), PFS and OS were available for the O-CHOP cohort only (n=20). The ORR was 8/20 (40%) after six cycles of O-CHOP. 4/12 patients with high ctDNA attained CR or PR after six cycles, compared with 4/8 of those with low pre-treatment ctDNA burden (OR 0.91, 95% confidence interval 0.19-4.2, p = 0.9).

High pre-treatment ctDNA did not correlate with PFS (Figure 5.3.3.1A, median PFS 4.1 vs 3.6 months, p = 0.53) or OS (Figure 5.3.3.1B; median OS 5.5 vs 7 months, p = 0.64). The aggressive nature of RT meant that the PFS and OS data was similar, and the Kaplan-Meier plots appear nearly identical over the timescale used. Considered as a continuous variable, pre-treatment ctDNA burden (log_{10}HGE/ml) did not impact PFS or OS (Figure 5.3.3.2A and B). Response at interim assessment predicted both PFS (p = 0.005) and OS (p = 0.003).

Figure 5.3.3.1: PFS and OS by ctDNA burden in the O-CHOP cohort (n=20)

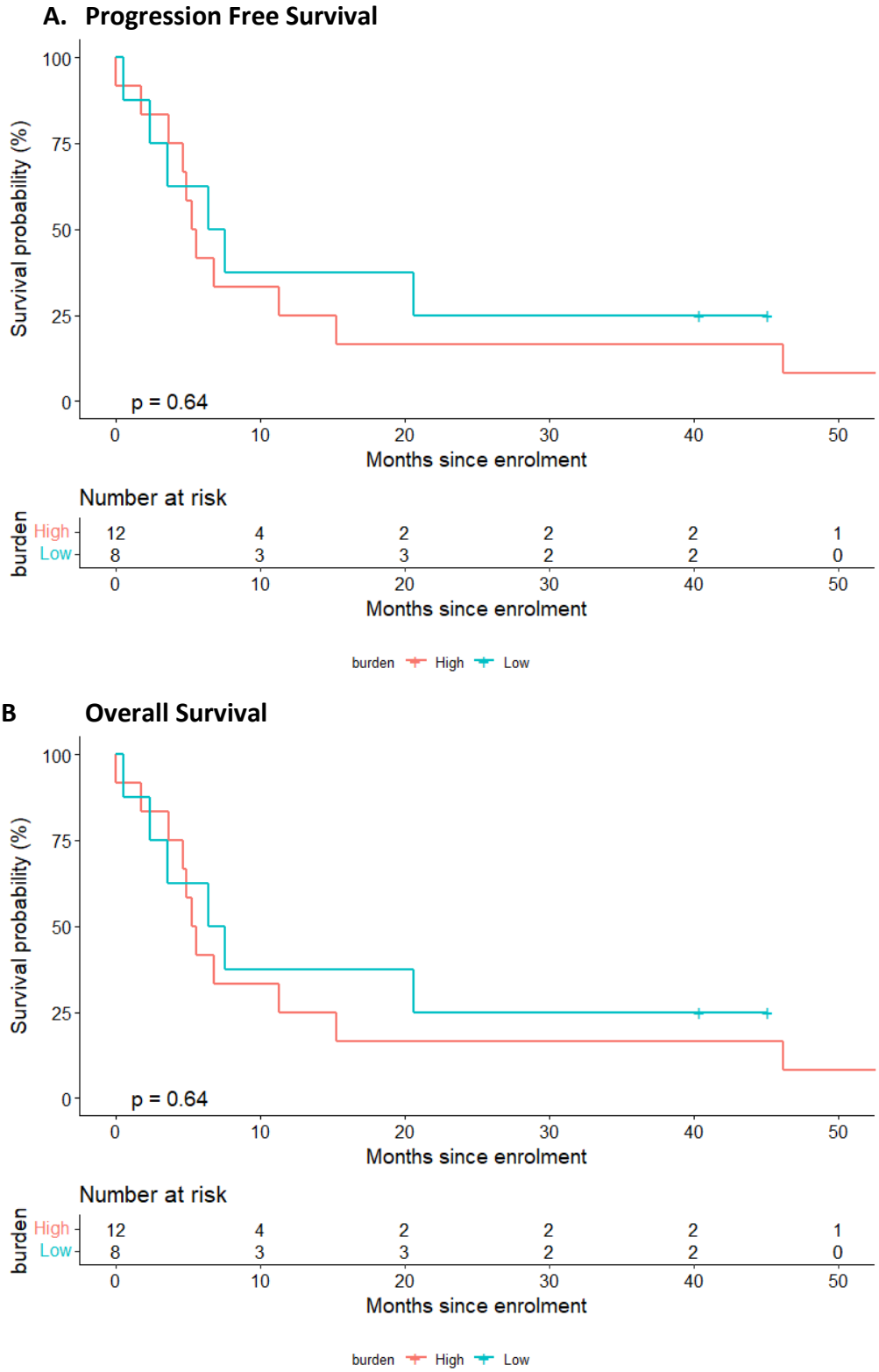
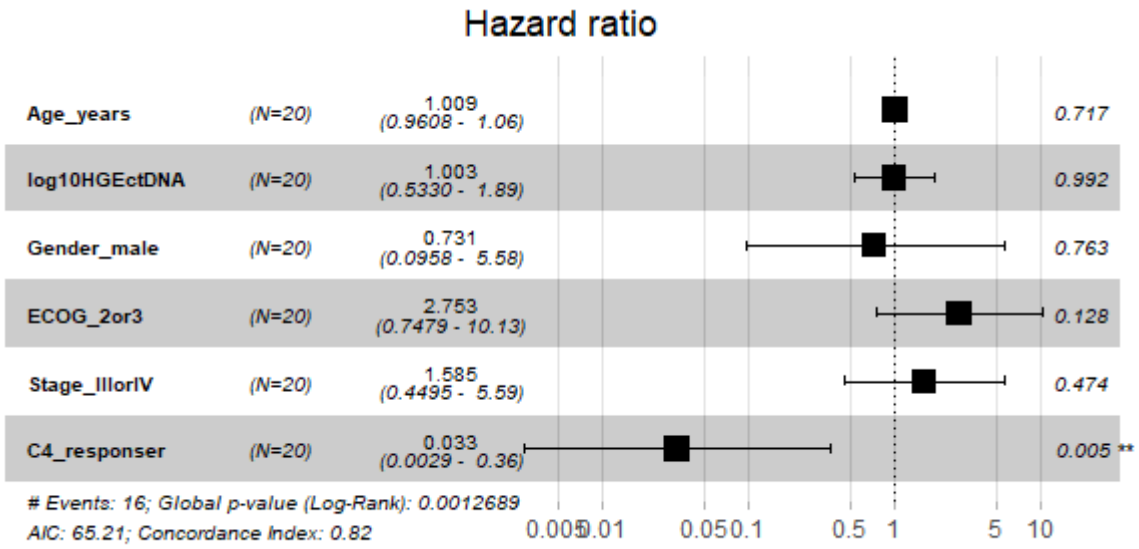


Figure 5.3.1.1: Kaplan-Meier estimates of PFS (A, upper) and OS (B, lower) in patients above the median (High ctDNA, in red) compared with below the median (Low ctDNA, in blue). The three patients with undetectable ctDNA at enrolment are included with the low ctDNA subgroup.

Figure 5.3.3.2: Cox proportional hazards of factors impacting clinical outcomes (n=20)

A Progression free survival



B Overall survival

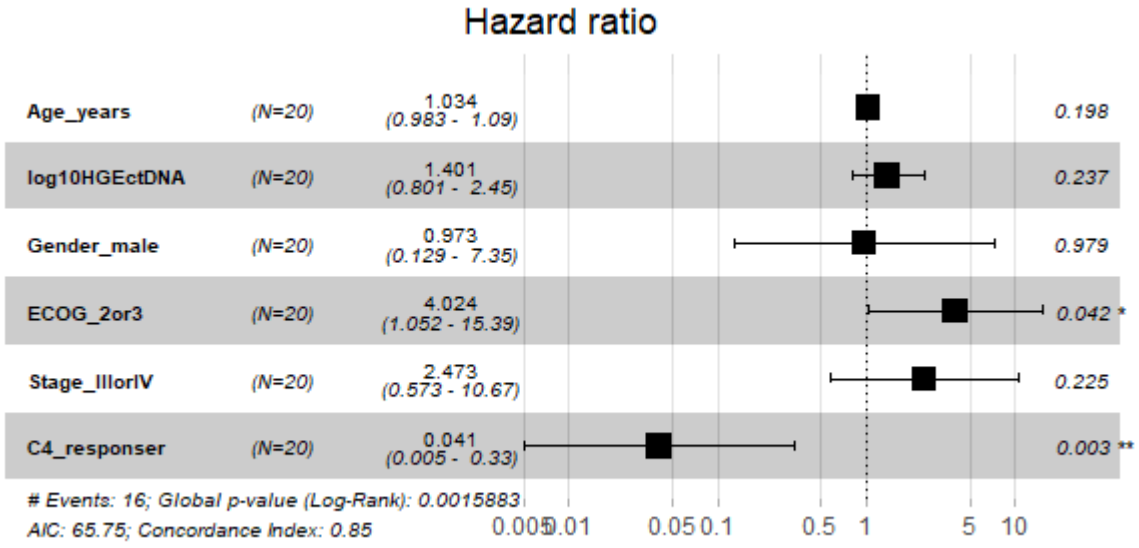


Figure 5.3.3.2: Forest plot showing the relationship between age, ctDNA burden, gender, ECOG Performance Status, advanced stage, and response at Cycle 4 and PFS (A, upper) and OS (B, lower).

5.3.4 Pre-treatment ctDNA detects additional mutations

Variants detected in the RT plasma cell-free DNA

After variant filtering, 635 ctDNA variants from 28/31 (90.3%) RT plasma samples remained (Table 5.3.4.1, Appendix F). The most frequently mutated genes were *KMT2D* (79 variants in 7 patients), *CDKN2A* (48 variants in 8 patients) and *TP53* (33 variants in 13 patients) (Figure 5.3.4.1). Following pathogenicity assessment and manual curation, 15 *TP53* mutations in the ctDNA were retained (Table 5.3.4.2).

Figure 5.3.4.1: Frequency barplot of variants detected in ctDNA by gene

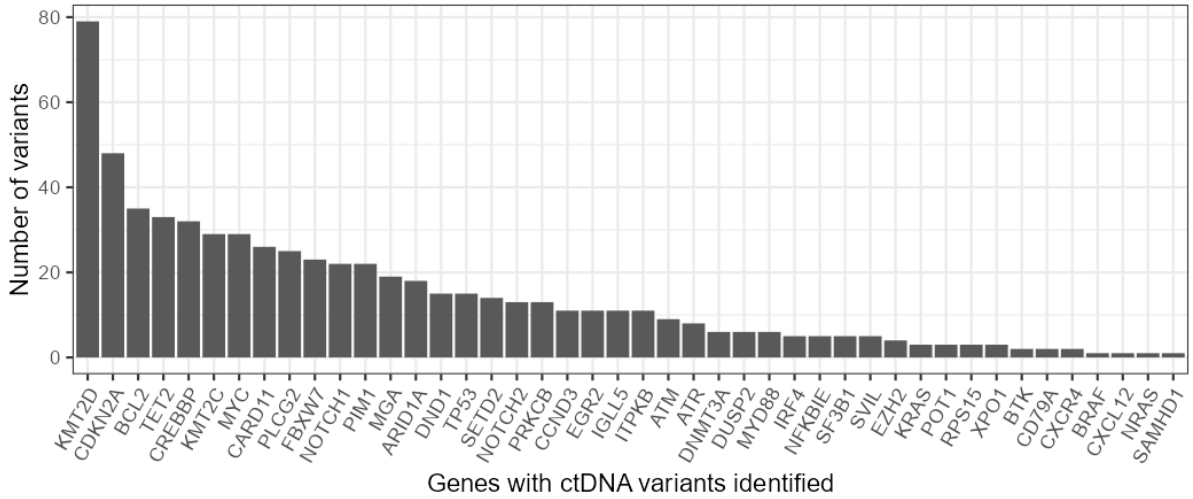


Figure 5.3.4.1: Plasma ctDNA variants detected at baseline (n=31) by gene (x-axis) and frequency (y-axis). No variants were detectable in 3 O-CHOP patients (CH1005, CH1007 and CH1013).

Table 5.3.4.1 Variants detected in baseline plasma samples from RT patients by frequency (top 20 variants displayed)

Gene	HGVSc	HGVSp	COSMIC	Number
CDKN2A	NM_058195.4:c.223G>C	NP_478102.2:p.Gly75Arg	COSV58700022	6
CDKN2A	NM_058195.4:c.262C>G	NP_478102.2:p.Arg88Gly	COSV58718388	6
CDKN2A	NM_058195.4:c.349G>A	NP_478102.2:p.Ala117Thr	COSV58724994	6
KMT2D	NM_003482.4:c.5285G>A	NP_003473.3:p.Ser1762Asn	COSV56442519	6
MYC	NM_002467.6:c.415T>C	NP_002458.2:p.Phe139Leu	COSV52380970	6
ARID1A	NM_006015.6:c.3575A>C	NP_006006.3:p.Asn1192Thr	COSV61373793	5
ARID1A	NM_006015.6:c.4033A>G	NP_006006.3:p.Thr1345Ala	COSV100250806	5
BCL2	NM_000633.3:c.515A>G	NP_000624.2:p.Asn172Ser	COSV61372150	5
CARD11	NM_032415.7:c.1254G>C	NP_115791.3:p.Arg418Ser	COSV67796802	5
CARD11	NM_032415.7:c.2947G>A	NP_115791.3:p.Val983Met	COSV67803819	5
CCND3	NM_001760.5:c.785G>C	NP_001751.1:p.Ser262Thr	COSV57828228	5
CDKN2A	NM_058195.4:c.286C>A	NP_478102.2:p.Arg96Ser	COSV58688708	5
CREBBP	NM_004380.3:c.4807G>A	NP_004371.2:p.Ala1603Thr	COSV52117791	5
CREBBP	NM_004380.3:c.485C>T	NP_004371.2:p.Ala162Val	COSV52137833	5
CREBBP	NM_004380.3:c.7005C>G	NP_004371.2:p.Ile2335Met	COSV52139232	5
DND1	NM_194249.3:c.365C>T	NP_919225.1:p.Pro122Leu	COSV56907925	5
KMT2C	NM_170606.3:c.1017G>C	NP_733751.2:p.Lys339Asn	COSV51276408	5
KMT2D	NM_003482.4:c.14395G>A	NP_003473.3:p.Val4799Met	COSV56419469	5
MGA	NM_001164273.2:c.9007G>A	NP_001157745.1:p.Ala3003Thr	COSV54961488	5
MYC	NM_002467.6:c.144G>C	NP_002458.2:p.Gln48His	COSV52369984	5

Variants detected in the RT tumours

Twelve of 15 (80%) RT tumours had detectable mutations. 37 variants (median VAF 0.34, IQR 0.16-0.47) affecting 12 tumours were found (Figure 5.3.4.2, Table 5.3.4.1). *TP53* mutations occurred in seven tumours, followed by *EGR2* and *NOTCH1* affecting four tumours each.

Figure 5.3.4.2 Frequency barplot of variants detected in the tumour by genes

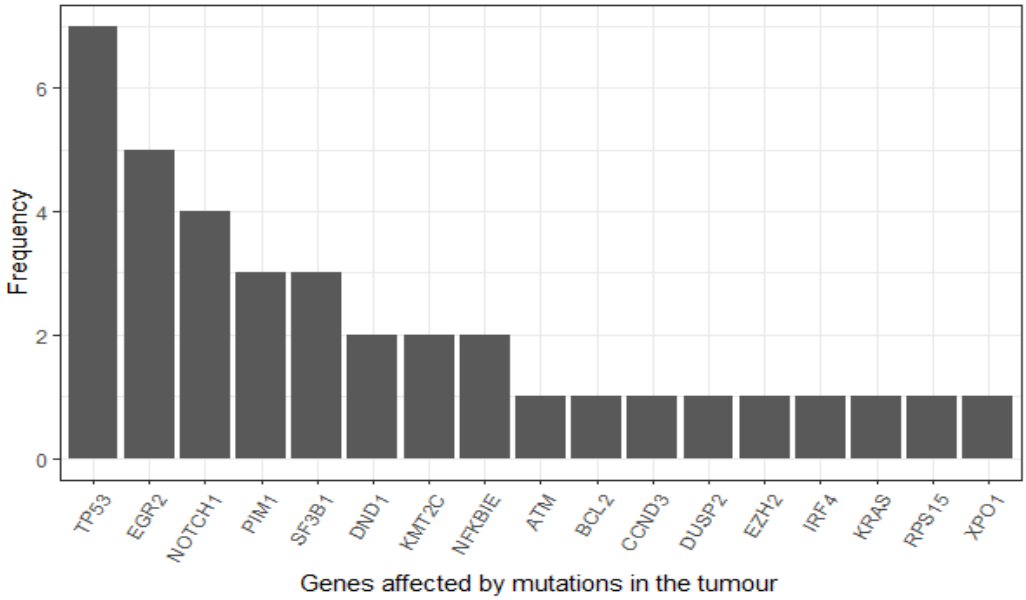


Figure 5.3.4.2: Variants detected in Richter's tumour tissue (n=15) by gene (x-axis) and frequency (y-axis). No variants were detectable in STELLAR004, STELLAR014 and CH1032.

Table 5.3.4.2. Thirty-seven variants detected in the Richters tumours

Trial ID	Gene	Consequences	HGVSc	HGVSp	COSMIC	VAF
CH1001	PIM1	missense	NM_002648.4:c.237G>C	NP_002639.1:p.Glu79Asp	COSV100998754	0.15
	KMT2C	missense	NM_170606.3:c.2968A>G	NP_733751.2:p.Ser990Gly	COSV51283026	0.09
	NOTCH1	frameshift	NM_017617.5:c.7541_7542del	NP_060087.3:p.Pro2514ArgfsTer4	COSV53024776	0.35
	PIM1	missense	NM_002648.4:c.403G>A	NP_002639.1:p.Glu135Lys	COSV65164498	0.1
	PIM1	missense	NM_002648.4:c.437G>A	NP_002639.1:p.Ser146Asn	COSV65164555	0.13
	IRF4	missense	NM_002460.4:c.208C>G	NP_002451.2:p.Leu70Val	COSV66704349	0.14
CH1007	TP53	missense	NM_000546.6:c.817C>T	NP_000537.3:p.Arg273Cys	COSV52662066	0.67
CH1009	TP53	missense	NM_000546.6:c.731G>C	NP_000537.3:p.Gly244Ala	COSV52661058	0.74
	KRAS	missense	NM_004985.5:c.35G>A	NP_004976.2:p.Gly12Asp	COSV55497369	0.18
	DND1	stop	NM_194249.3:c.854G>A	NP_919225.1:p.Trp285Ter	COSV56906398	0.17
CH1017	NOTCH1	frameshift	NM_017617.5:c.7541_7542del	NP_060087.3:p.Pro2514ArgfsTer4	COSV53024776	0.16
	EGR2	missense	NM_000399.5:c.1066G>A	NP_000390.2:p.Glu356Lys	COSV54346083	0.18
	DUSP2	missense	NM_004418.4:c.419C>T	NP_004409.1:p.Pro140Leu	COSV56619408	0.18
	BCL2	intron variant	NM_000633.3:c.-286-85G>A	NA	COSV61386812	0.28
CH1018	RPS15	missense	NM_001018.5:c.413C>T	NP_001009.1:p.Ser138Phe	COSV51832669	0.26
CH1019	NOTCH1	frameshift	NM_017617.5:c.7541_7542del	NP_060087.3:p.Pro2514ArgfsTer4	COSV53024776	0.75
	EGR2	missense	NM_000399.5:c.1150C>A	NP_000390.2:p.His384Asn	COSV54346140	0.64
	XPO1	missense	NM_003400.4:c.1711G>A	NP_003391.1:p.Glu571Lys	COSV68944533	0.36
CH1021	KMT2C	missense	NM_170606.3:c.2968A>G	NP_733751.2:p.Ser990Gly	COSV51283026	0.1
	TP53	missense	NM_000546.6:c.527G>A	NP_000537.3:p.Cys176Tyr	COSV52660760	0.09
	SF3B1	missense	NM_012433.4:c.2680G>A	NP_036565.2:p.Asp894Asn	COSV59214349	0.4
CH1022	NOTCH1	frameshift	NM_017617.5:c.7541_7542del	NP_060087.3:p.Pro2514ArgfsTer4	COSV53024776	0.35
	CCND3	stop gained	NM_001760.5:c.778C>T	NP_001751.1:p.Gln260Ter	COSV57827644	0.13
STELLAR002	TP53	missense	NM_000546.6:c.743G>A	NP_000537.3:p.Arg248Gln	COSV52661091	0.56
	ATM	missense	NM_000051.4:c.5558A>T	NP_000042.3:p.Asp1853Val	COSV53725804	0.23
	EGR2	missense	NM_000399.5:c.1066G>A	NP_000390.2:p.Glu356Lys	COSV54346083	0.35
STELLAR005[#]	TP53	missense	NM_000546.6:c.817C>T	NP_000537.3:p.Arg273Cys	COSV52662066	0.45

	EGR2	missense	NM_000399.5:c.1150C>A	NP_000390.2:p.His384Asn	COSV54346140	0.48
	DND1	stop	NM_194249.3:c.854G>A	NP_919225.1:p.Trp285Ter	COSV56906398	0.09
	SF3B1	missense	NM_012433.4:c.2318T>G	NP_036565.2:p.Leu773Arg	COSV59223584	0.26
	NFKBIE	upstream variant	NA	NA	COSV99240993	0.38
STELLAR015[#]	TP53	Missense	NM_000546.6:c.817C>T	NP_000537.3:p.Arg273Cys	COSV52662066	0.83
	EGR2	Missense	NM_000399.5:c.1150C>A	NP_000390.2:p.His384Asn	COSV54346140	0.57
	SF3B1	missense	NM_012433.4:c.2318T>G	NP_036565.2:p.Leu773Arg	COSV59223584	0.38
	NFKBIE	upstream variant	NA	NA	COSV99240993	0.47
STELLAR016	TP53	missense	NM_000546.6:c.724T>C	NP_000537.3:p.Cys242Arg	COSV52660956	1
	EZH2	missense	NM_004456.5:c.1937A>T	NP_004447.2:p.Tyr646Phe	COSV57445929	0.36

STELLAR005 was randomised to the standard arm of the STELLAR trial. At progression, they joined the acalabrutinib monotherapy (Cohort 1) and were assigned a new study identification number, STELLAR015 (See also Figure 2.3.1.1).
HGVS, Human Genome Variation Society. HGVSc coding sequence. HGVSp protein

Variants affecting *TET2* or *DNM3TA* and previously reported in myeloid stem cell disorders were detected in the ctDNA of 6/31 (19.4%) RT patients.

Paired tumour and cell-free DNA samples were available for 8/31 (25.8%) RT patients. While 20/33 (60.6%) of the tumour variants could be detected in the paired ctDNA sample, 85/105 (81%) of ctDNA variants were not identified in the tumour sample (Figure 5.3.4.3). Three of these 85 variants occurred in *TET2*, were previously reported in MDS/AML and were interpreted as reflecting CHIP.

Figure 5.3.4.3: Venn diagram of mutations for paired RT tumour and ctDNA samples (n=8 patients)

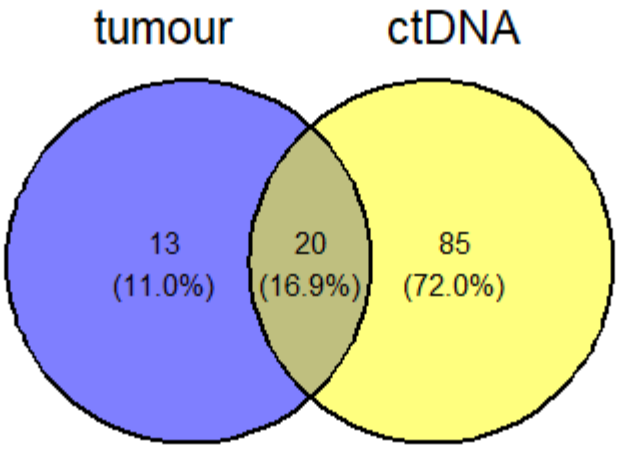


Figure 5.3.4.3: Venn diagram showing mutations detected in the tumour only (blue, n= 13), shared mutations (overlap, n=20) and mutations found in the ctDNA alone (yellow, n= 85) for eight RT patients with paired samples sequenced.

Additional TP53 variants in the ctDNA

Notably, *TP53* variants were detected in the ctDNA of 13/31 (41.9%) RT patients. Twelve variants in ten RT patients could be confirmed as pathogenic or deleterious using Seshat (Table 5.3.4.3). The remaining three *TP53* variants were variants of uncertain significance (VUS).

Paired tumour samples were sequenced for four patients (STELLAR004, STELLAR005, STELLAR015, STELLAR016) and *TP53* variants were detected in three of the four tumours. STELLAR004 had two *TP53* mutations detected in the ctDNA at VAF 0.25 and 0.27, respectively. Neither variant was detected in the tumour.

For the RT patients lacking a paired tumour sample, results were compared with the clinical dataset provided by the trials unit. Three additional O-CHOP RT patients (CH1008, CH1023 and CH1024) had *TP53* ctDNA mutations compared to assessment of biopsy or PB-CLL material alone previously reported (Eyre et al. 2016). In the STELLAR trial dataset, STELLAR007 is recorded as *TP53* mutated. STELLAR005, STELLAR010 and STELLAR015 are recorded as *TP53* wild-type and the remaining samples as unknown.

This finding suggests that the frequency of *TP53* mutations in RT patients may be higher than is found from sequencing tumour DNA alone, further implicating impaired DNA damage response pathways and offering a mechanism explaining the resistance to or early progression after chemotherapy. No relationship between *TP53* disruption and survival was observed in the O-CHOP cohort, although outcomes were poor for both groups (median OS 6.4 vs 5.1 months, p 0.91).

Figure 5.3.4.4: Frequency of variants detected in ctDNA by patient (n=28)

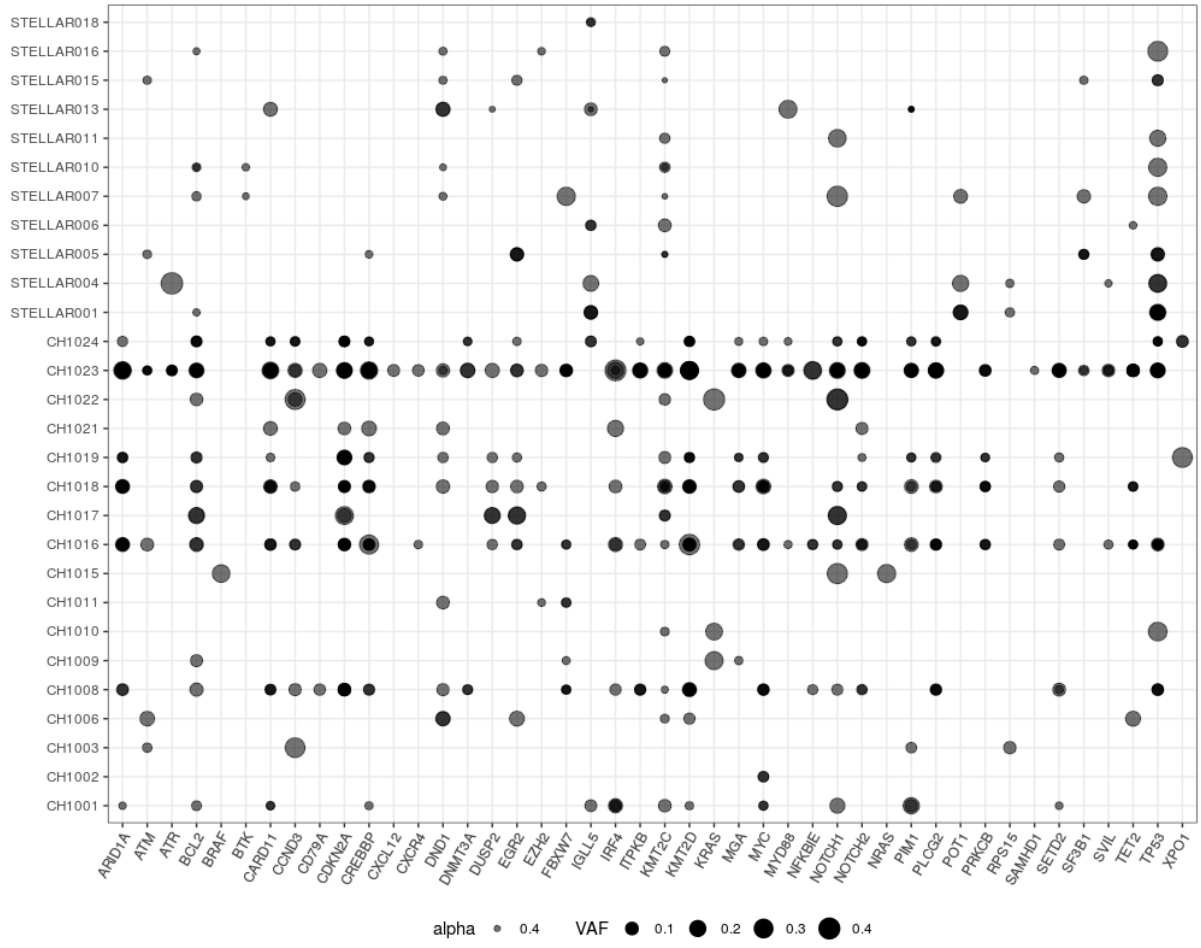


Figure 5.3.4.4: Bubbleplot of genes mutated in pre-treatment circulating tumor DNA. Each row represents a participants pre-treatment sample, and each column represents a gene. The size of the circle represents the VAF detected.

Table 5.3.4.3: Fifteen *TP53* variants detected in the ctDNA of Richter's Transformation

Trial_ID	HGVS DNA	HGVS Protein	COSMIC ID	VAF	Pathogenicity assessment by Seshat
CH1008	NM_000546.6:c.497C>T	NP_000537.3:p.Ser166Leu	COSV52661795	0.07	VUS
CH1010	NM_000546.6:c.731G>C	NP_000537.3:p.Gly244Ala	COSV52661058	0.30	Deleterious
CH1016	NM_000546.6:c.707A>T	NP_000537.3:p.Tyr236Phe	COSV52662150	0.11	VUS
CH1023	NM_000546.6:c.551_552insCTA	NP_000537.3:p.Asp184_Ser185insTyr	COSV53138655	0.08	Pathogenic
CH1024	NM_000546.6:c.707A>T	NP_000537.3:p.Tyr236Phe	COSV52662150	0.04	VUS
STELLAR001	NM_000546.6:c.752T>G	NP_000537.3:p.Ile251Ser	COSV52728902,	0.12	Deleterious
	NM_000546.6:c.701A>G	NP_000537.3:p.Tyr234Cys	COSV52661201	0.16	Pathogenic
STELLAR004	NM_000546.6:c.752T>G	NP_000537.3:p.Ile251Ser	COSV52728902	0.27	Deleterious
	NM_000546.6:c.701A>G	NP_000537.3:p.Tyr234Cys	COSV52661201	0.25	Pathogenic
STELLAR005	NM_000546.6:c.817C>T	NP_000537.3:p.Arg273Cys	COSV52662066	0.11	Pathogenic
STELLAR007	NM_000546.6:c.517G>A	NP_000537.3:p.Val173Met	COSV52676535	0.29	Pathogenic
STELLAR010	NM_000546.6:c.730G>A	NP_000537.3:p.Gly244Ser	COSV52668392	0.28	Pathogenic
STELLAR011	NM_000546.6:c.809T>G	NP_000537.3:p.Phe270Cys	COSV52661272	0.20	Deleterious
STELLAR015	NM_000546.6:c.817C>T	NP_000537.3:p.Arg273Cys	COSV52662066	0.07	Pathogenic
STELLAR016	NM_000546.6:c.724T>C	NP_000537.3:p.Cys242Arg	COSV52660956	0.35	Deleterious

Table 5.3.4.1: *TP53* variants detected in ctDNA and their classification. Deleterious = Multiple lines of computational evidence are suggestive of a deleterious effect on *TP53.* Pathogenic = Published research as well as database analysis provide sufficient evidence for classification of this variant as pathogenic. VUS = variant of uncertain significance.

5.3.5 Changes in ctDNA burden on treatment (STELLAR cohort only, n=11/14)

At interim assessment, three patients attained CR, seven had PR and two patients had Stable Disease (SD). Two patients, STELLAR004 and STELLAR018 progressed or withdrew from the study prior to the interim radiological assessment at the fourth cycle of treatment. Neither submitted a post-treatment plasma sample. As I had not observed a relationship between baseline ctDNA burden and early responses, next, I examined sequential ctDNA samples in the STELLAR cohort to learn if dynamic changes in the ctDNA after treatment linked with outcomes.

Twenty-three post treatment plasma cell-free DNA samples from 11 patients were sequenced. One of 11 RT patients, STELLAR006 eradicated ctDNA after two cycles of treatment and attained a CR at interim assessment. Two other patients (STELLAR010 and STELLAR014) attained a radiological CR with detectable ctDNA.

All seven patients in PR and both patients with SD at interim assessment had persistent ctDNA after treatment. Changes in the ctDNA burden were heterogenous. One patient, STELLAR011 had a 1log reduction in ctDNA after one cycle of treatment and recorded a PR at interim assessment. STELLAR011 had persistent detectable ctDNA at week 24. Two patients in PR had rising ctDNA burdens (STELLAR001 and STELLAR012) with the remainder showing stable ctDNA levels (less than one log change from pre-treatment levels) (Figure 3.5.5.1, by ctDNA burden and Figure 3.5.5.2, by variant).

Figure 5.3.5.1: Sequential ctDNA burden at each timepoint in the STELLAR samples (n=11)

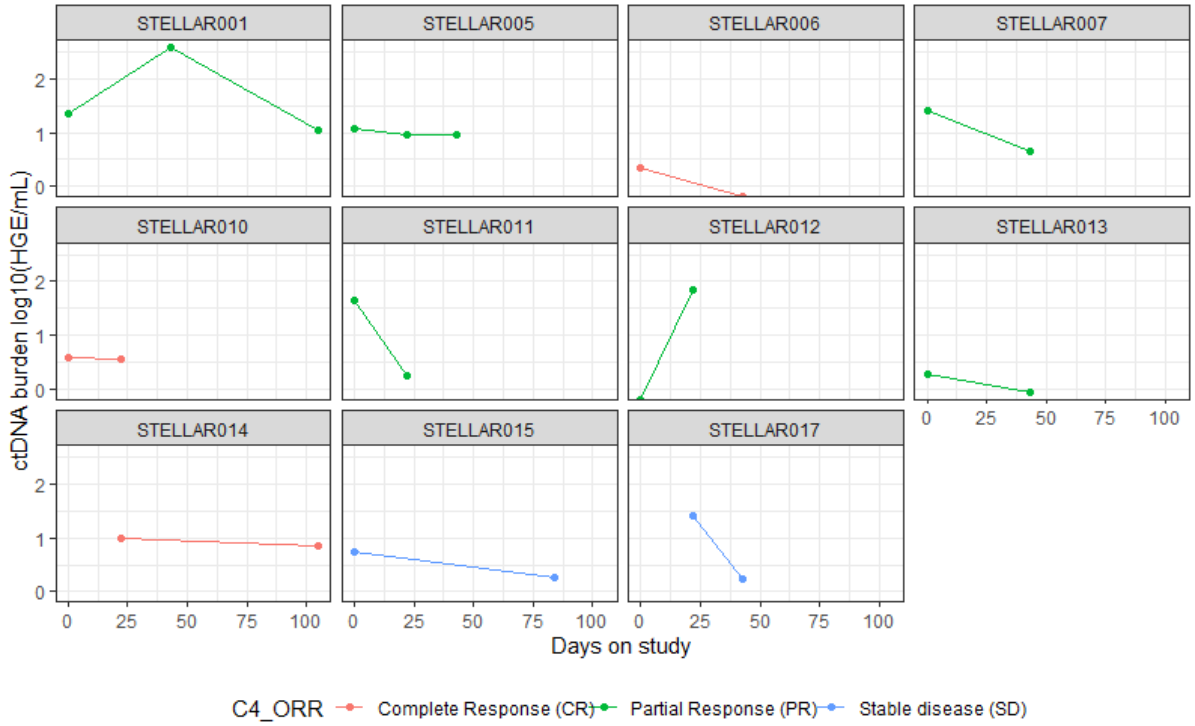


Figure 5.3.5.1: Changes in the ctDNA burden (measured the product of the median ctDNA VAF by the cell-free DNA in ng/µl and displayed on as log_{10}(HGE/ml), y-axis) during treatment, coloured by Response at Interim Assessment. Patients in CR are shown in red, PR in green and SD in blue. The first sample for STELLAR014 and STELLAR017 is prior to Cycle 2.

Figure 5.3.5.2: Dynamic changes in ctDNA by variant during treatment in the STELLAR samples (n=11)

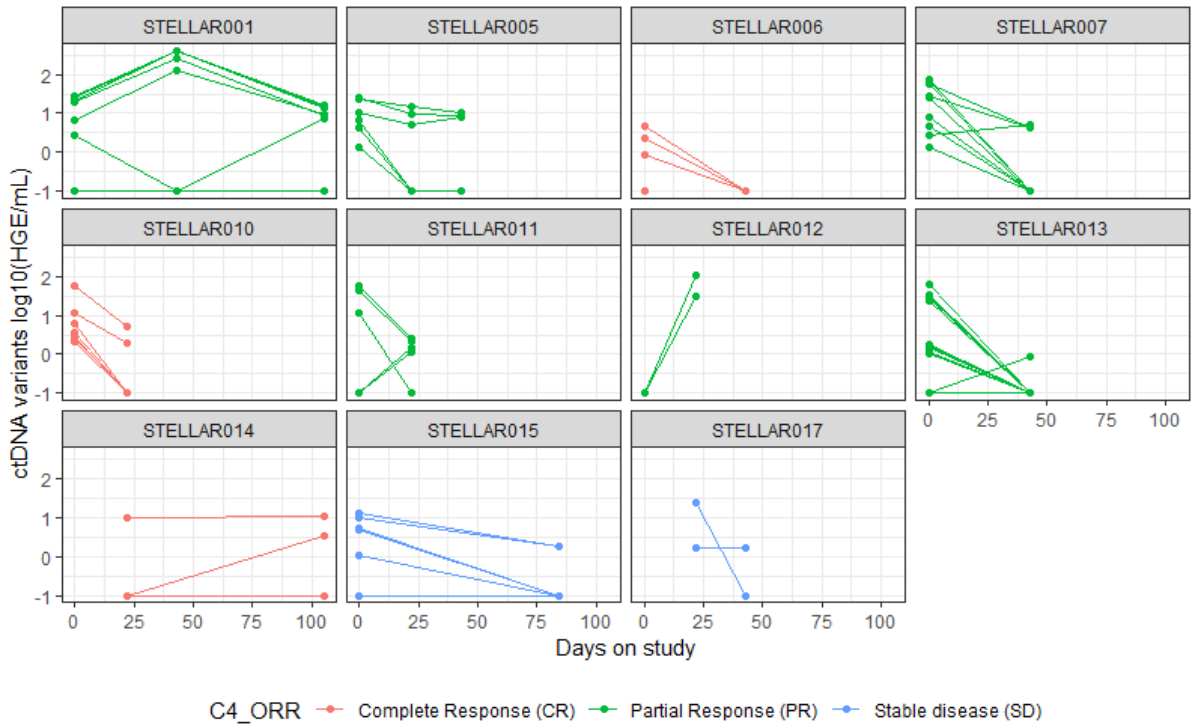


Figure 5.3.5.2: Changes in the ctDNA variants (quantified as log_{10}(HGE/ml), y-axis) during treatment, coloured by Response at Interim assessment. Patients in CR are shown in red, PR in green and SD in blue. The first sample for STELLAR014 and STELLAR017 is prior to Cycle 2. For visualisation, variants not detected at one timepoint are depicted at VAF = -1.

5.4 Discussion

Quantification of ctDNA burden and dynamic changes in ctDNA is an emerging biomarker in lymphoma (Lauer, Mutter, and Scherer 2022) but few studies have examined ctDNA in RT. Published data is limited to 15 RT patients enrolled the LYM1002 trial for aggressive B-cell lymphoma (Bruscaggin et al. 2021) and individual cases (Yeh, Hunter, et al. 2017; Parry et al. 2023b; Nadeu et al. 2022). In this exploratory work, I sequenced plasma cell-free DNA samples from RT patients enrolled in two clinical trials. My findings differ from existing literature on ctDNA in DLBCL and CLL, arguing for a unique role for ctDNA in RT.

5.4.1 Pre-treatment plasma ctDNA burden, clinical stage and interim response

In this RT cohort, ctDNA did not parallel increasing radiological stage or other biomarkers and my data does not support the hypothesis that ctDNA burden parallels radiological stage in RT. This finding contrasts with literature on ctDNA and other lymphomas. Increasing baseline ctDNA burden, measured in haploid human genome equivalents per millilitre (HGE/ml) parallels increasing clinical stage, total metabolic tumour volume (TMTV) by PET-CT in DLBCL (Scherer et al. 2016; Meriranta et al. 2022) and International Prognostic Indexes other nodal lymphomas (Fernandez-Miranda et al. 2023; Lakhotia et al. 2022; Spina et al. 2018). Other studies quantified ctDNA using droplet digital PCR and reported ctDNA correlated with TMTV (Delfau-Larue et al. 2018). The LYM1002 trial does not report ctDNA quantification by stage for the 15 RT patients (Bruscaggin et al. 2021).

RT differs from nodal lymphomas in that the disease burden is distributed across nodal, marrow and circulating PB compartments. Radiological staging is based on the identification of nodal disease and potentially underestimates the disease burden of marrow and circulating neoplastic cells. In this context, the ctDNA which captures neoplastic mutations from all disease compartments might provide a more comprehensive assessment of disease burden. My sample size is small and the relationship between ctDNA quantification, clinical stage and other prognostic features will be addressed using further samples from the ongoing STELLAR study.

5.4.2 Impact of pre-treatment plasma ctDNA burden on PFS and OS

Numerous studies in DLCBL and other lymphomas report pre-treatment ctDNA quantity predicts clinical response to frontline chemoimmunotherapy (Scherer et al. 2016; Kurtz et al. 2018; Meriranta et al. 2022; Fernandez-Miranda et al. 2023). Conversely, RT patients with both high and low baseline ctDNA did poorly in the LYM1002 trial (median PFS 3.1 vs 8.8 months, $p = 0.235$) (Bruscaggin et al. 2021). Similarly, RT patients in the O-CHOP trial did poorly irrespective of pre-treatment ctDNA burden (median PFS 4.1 vs 3.6 months, $p = 0.53$) underlining the aggressive nature of the disease. Overall, my data does not support the hypothesis that plasma ctDNA burden informs PFS and OS in RT.

Radiological response at early interim assessment, not pre-treatment ctDNA burden predicted PFS and OS in the O-CHOP cohort. However, 9/31 (29%) patients experienced disease progression, died or withdrew from treatment prior to four cycles of chemo-immunotherapy.

As STELLAR is an ongoing study, the CTU did not provide any data on ORR at end of treatment, PFS or OS. Data on radiological responses and clinical outcomes will be available when the trial has completed, and this work is planned.

5.4.3 Additional driver mutations are detectable in the ctDNA

Previous studies in DLCBL showed a high degree of concordance between the mutational profile from the tumour and ctDNA sequencing (Rossi et al. 2017; Scherer et al. 2016). In contrast, I observed limited concordance (20/33, 60.6% of tumour variants detected in ctDNA) between the tumour and ctDNA profiles in the two RT cohorts. All the O-CHOP cohort carried related IGHV rearrangements in the paired PB-CLL (Klintman et al. 2021), suggesting that the differences observed cannot be attributed to the RT arising from a clone with a distinct VDJ rearrangement. No data on clonal relatedness, determined by IGHV rearrangements, was available for the STELLAR cohort. Potential factors include the use of FFPE tumour DNA in the O-CHOP cohort, different sequencing depths and biological heterogeneity.

Variants found in the ctDNA but absent from the nodal tissue might reflect variants from the circulating neoplastic B-cells. I did not have access to PB-CLL cells taken at the time of RT diagnosis for targeted sequencing and some discordant results may be resolved by sequencing DNA from the circulating PB-CLL. Six RT patients had variants previously reported in CHIP. Only a minority of the ctDNA only variants might be explained by clonal haematopoiesis.

Spatial heterogeneity in RT may be greater than observed in DLCBL. Using WGS, 507 clonal tissue-RT variants were detected in 10/14 (71.4%) of RT patients. 40/507 (7.9%) variants were observed at subclonal frequency (and thus unlikely to be detectable in cell-free DNA) and 27 were absent from the paired PB-CLL. Clonal stability, defined as change in cancer cell fraction of –0.1 to 0.1 between the PB-CLL and nodal RT was observed in a minority of sequenced tumours (Klintman et al. 2021).

It is possible my variant filtering strategy did not completely exclude artefactual calls in the ctDNA results. Consensus guidelines on ctDNA sequencing for lymphoma recommend panels are tested on 30 control samples (Rossi et al. 2020). This was not possible for this project for reasons of cost and limited availability of suitable controls. As outlined earlier (Methods and Chapter 4), three variant callers (vardict, mutect2 and varscan2) were tested on non-cancer control plasma samples. Varscan2 was selected as the other callers generated abundant variant calls in the control and germline samples.

My variant filtering strategy sought to minimise false positive calls in the ctDNA. by retaining only variants with COSMIC ID previously reported in haematological malignancy. To minimise the influence of FFPE artefact, I retained variants at VAF > 0.05 from the tumour samples. A more permissive filtering strategy, including tumour variants at VAF > 0.01 did not reveal additional shared variants.

My findings agree with two other publications reporting additional ctDNA variants compared with sequencing of genomic DNA in RT. Using a seven-gene targeted panel, Yeh et al showed ctDNA detects additional mutations compared to circulating leukocytes in

relapsed/refractory CLL and detection of new mutations in ctDNA anticipates high-grade transformation. RT copy number gains and losses are detectable in the ctDNA by shallow WGS (Yeh, Hunter, et al. 2017; Parry et al. 2023b).

In this cohort, 13/31 (41.9%) RT patients had a ctDNA *TP53* variant and ctDNA revealed additional mutations in *TP53* not detected by standard diagnostic tests (FISH and targeted sequencing) at study enrolment. The LYM1002 trial detected *TP53* mutations using CAPP-SEQ in the ctDNA of 7/15 (46.7%) of RT patients which persisted during ibrutinib and nivolumab treatment. In the same study, follicular lymphoma and DLBCL with *TP53* mutations in ctDNA experienced shorter PFS, underlining that mutations detectable in ctDNA only may be clinically important. The difference was not significant in RT, but both *TP53*-mutated and *TP53* wild-type did poorly.

In addition, ctDNA detected the BTK^{C481S} mutation associated with resistance to covalent BTKi drugs in STELLAR007 illustrating a potential role for plasma cell-free DNA in detecting of treatment resistant mutations. The *PGCL2* variants detected are not previously reported in the context of BTK inhibitor resistance and their clinical significance is uncertain. Small studies using PCR-based approaches to detect *BTK* and *PGCL2* resistance mutations have yielded conflicting results. Plasma cell-free DNA detected additional mutations compared to PB-CLL in one study (Albitar et al. 2017) but failed to detected *BTK* mutations in another (Trummer et al. 2022).

5.4.4 Early reductions in plasma ctDNA burden on treatment are uncommon in RT

Dynamic changes in the ctDNA burden during treatment delineates subgroups with favourable and unfavourable outcomes in DLCBL (Kurtz et al. 2018). In CLL clinical trials, flow cytometry to 10^{-4} complements clinical assessment of response, informs time to next treatment and can direct treatment duration. Dynamic changes in the number of residual CLL cells detectable by flow-cytometry has been used to predict duration of remission after treatment (Gregory 2020; Rawstron 2020).

In the STELLAR participants studied, only one patient (STELLAR006, *TP53* wild-type) eradicated ctDNA by cycle 3 and attained a CR at interim assessment. Early 2-$\log_{10}$ reductions in ctDNA were infrequent in RT and in *TP53*-mutated DLBCL patients in the LYM002 trial (Bruscaggin et al. 2021). In this cohort, almost one third patients progressed prior to the interim assessment scan (cycle 4 or day 84 of planned treatment). Response rates to chemoimmunotherapy are poorer in RT than DLCBL, making early identification of non-progressors critical to allow suitable candidates to receive salvage treatments including cellular therapies. Further sequential plasma samples collected will be used to report the relationship between post-treatment ctDNA and clinical outcomes in the full STELLAR cohort (Appleby et al. 2019).

Three publications contrast ctDNA and flow-cytometry assessments of MRD in CLL patients treated with targeted therapies. Yeh and colleagues examined 32 CLL patients treated with either ibrutinib or venetoclax and observed that ctDNA based assessed showed greater sensitivity than four colour flow-cytometry in patients with residual lymphadenopathy (Yeh, Hunter, et al. 2017). In the GCLLSG trial combining obinutuzumab, acalabrutinib and

venetoclax, ctDNA using droplet digital PCR detected MRD in 29% of patients classed as MRD-negative by flow-cytometry (Furstenau et al. 2022). By contrast, plasma cell-free DNA offered no advantage over VDJ-based MRD sequencing of PB and BM in 32 FCR treated CLL patients (Thompson et al. 2019). The STELLAR trial will address the further address relationship between ctDNA, flow-cytometric MRD and outcomes in RT.

5.4.5 Limitations of this work

Molecular studies on RT are limited by the rarity of the disease. The largest WES study to date included 98 patients with DLBCL-type RT from six large academic institutions (Parry et al. 2023b). My work reports ctDNA sequencing from 31 RT patients, which has limited the opportunity for some results to reach statistical significance.

Plasma cell-free DNA analysis was not an exploratory objective of the O-CHOP study. Plasma was obtained from samples collected in EDTA tubes, risking contamination of the plasma with longer DNA fragments from circulating leukocytes, including circulating CLL cells. Evidence of this can been seen in the higher cell-free DNA quantities obtained from the EDTA plasma and the variable degree of long fragments on O-CHOP cell-free DNA electrophenograms. However, the library preparation includes multiple bead clean-up steps optimised to retain short fragments and the final library size is consistent with genuine cell-free DNA and adapters. Plasma cell-free DNA was obtained from Streck tubes for the STELLAR trial.

STELLAR recruited more slowly than anticipated due in part to the impact of the COVID-19 pandemic. Only 18 patient samples were available for my thesis. The target recruitment for the STELLAR randomised cohort is sixty participants, enabling detection of an improvement in PFS from 20% in the standard of care (R-CHOP) arm to 40% in the intervention arm (R-CHOP plus acalabrutinib) at 80% power (Appleby et al. 2019). Sequencing of cell-free DNA is an exploratory outcome for the study and samples from the fully recruited STELLAR trial offer the chance to reexamine the relationship between ctDNA and clinical outcomes in a larger cohort.

Changes in availability of consumable kits and progress at the OMDC meant that the library preparation and sequencing differ between the O-CHOP and the STELLAR cohorts. Capture hybridisation was performed individually for the O-CHOP cell-free DNA libraries whereas the STELLAR cell-free DNA libraries were pooled. The Thruplex-HV library preparation used for the STELLAR libraries required fewer PCR cycles than the Thruplex kit used for the O-CHOP libraries. Furthermore, the Thruplex-HV libraries require temperature control for the initial post capture washes. Finally, while O-CHOP cell-free DNA libraries were sequenced on the Illumina HiSeq, the OMDC had transitioned to the Illumina NextSeq for high-throughput sequencing by the time the STELLAR libraries were available. The numerous differences in the sample source, library preparation and sequencing platform limit direct comparisons between the two RT cohorts.

Depth of ctDNA sequencing attained is lower than reported in other lymphoma ctDNA studies using capture hybridisation. High proportions of off-target reads can be seen with capture hybridisation libraries, meaning that many reads sequenced fall outside the

selector, or target of interest. The median proportion of off-target reads for my ctDNA sequencing was 32.6% (IQR 29.4-60.1%). It means that although the number of reads suggested deep sequencing, many reads were off-target and thus uninformative. Although the depth of sequencing could be recovered by altering parameters of the bioinformatics pipeline (reducing the minimum family size to 1 read), I retained the minimum family size of two deduplicated reads to minimise the potential for a false positive call. In addition, post-sequencing filtering kept variants with COSMIC ID previously reported in haematological cancers, to further reduce of the risk.

Although confirmatory work is required in larger prospective cohorts, our hypothesis-generating study suggests that examination of ctDNA in RT has the potential to reveal additional, clinically significant CLL-driver mutations. Longer follow-up in the STELLAR trial dataset will reveal whether all RT patients with rising or stable ctDNA burdens after treatment will progress.

Chapter 6: Detection of clinically relevant biomarkers in chronic lymphocytic leukaemia by nanopore sequencing

6.1 Introduction

6.1.1 Detection of CLL biomarkers requires multiple diagnostic tests

The principal established biomarkers for CLL are specific Copy Number Alterations (CNAs) (del13q, tri12 and del11q), the status of the rearranged immunoglobulin heavy chain gene (IGHV) and the presence of deletions and/or mutations in *TP53*. Complex karyotype (determined by karyotyping of stimulated lymphocytes) and genomic complexity (derived from WGS or WES datasets) are emerging biomarkers of high-risk disease.

In clinical diagnostics, multiple tests interrogate a single patient sample before results are integrated into a unified diagnosis. In this chapter, I will evaluate whether nanopore sequencing (NS) can provide an alternative method to determine these markers accurately.

Immunoglobulin heavy chain status in chronic lymphocytic leukaemia

Immunoglobulin heavy chain genes are produced by the rearrangement of the variable, diversity and joining regions (VH, D, and JH, respectively) during early B-cell differentiation. Immunoglobulin diversity is determined by the number of germline VH, D and J segments, their combinations and somatic hypermutation of these rearranged genes (Section 1.1.3). CLL is broadly classified into IGHV mutated (m-IGHV) and unmutated (u-IGHV), based on the sequence homology to the germline, reflecting the degree of somatic hypermutation of the immunoglobulin locus due to activation-induced cytidine deaminase (AID) activity in the germinal centres. The mutational status of the rearranged immunoglobulin heavy chain variable (IGHV) gene has strong prognostic value, influencing time to first treatment (TTFT), progression free survival (PFS) and overall survival (OS) (Condoluci et al. 2020; Rotbain et al. 2020). IGHV status informs patient counselling, treatment choices and may be an entry criterion for clinical trials (Huber et al. 2022).

Recent ERIC guidelines for immunoglobulin gene sequence analysis acknowledges some limitations of this approach. While dividing CLL into two distinct groups is convenient, the binary approach is overly simplistic. So called ‘borderline cases’, with 97-99% germline homology should be highlighted in a clinical report (Agathangelidis et al. 2022). Assessing germline homology as a continuous variable refines the prognosis (Jain, Nogueras Gonzalez, et al. 2018). Cases with germline homology < 98% but stereotyped IGHV rearrangements such as subset #2 or subset #169 also carry an adverse clinical prognosis (Langerak et al. 2011; Agathangelidis et al. 2021; Jaramillo et al. 2020).

TP53 aberrations and chronic lymphocytic leukaemia

Both deletions of the *TP53* locus (17p13.1) and/or mutations in *TP53* confer inferior prognosis (Nadeu et al. 2016; Dohner et al. 2000; Chiaretti et al. 2014). Loss of *TP53* function facilitates genomic instability and enables cells to tolerate the metabolic consequences of aneuploidy (Kastenhuber and Lowe 2017). Clonal and subclonal pathogenic SNVs/indels affecting *TP53* DNA-binding domains are clinically relevant, necessitating sensitive techniques for detection.

Detection of Copy Number Alterations and Genomic Complexity in Chronic Lymphocytic Leukaemia

The prognostic role of specific copy number gains and losses, determined by fluorescence *in-situ* hybridisation (FISH) is established (Dohner et al. 2000) and testing of specific loci is recommended before treatment (Hallek et al. 2018). An increasing body of evidence suggests genomic complexity (GC), derived from WGS or WES provides important prognostic information in CLL (Baliakas et al. 2019; Robbe et al. 2022), and cannot be revealed by existing targeted approaches.

6.1.2 Limitations of current laboratory techniques

In clinical diagnostics, detection of all recommended and desirable molecular CLL biomarkers (Hallek et al. 2018) requires combination of FISH, sequencing and conventional karyotyping (with specific stimulation of lymphocytes). These conventional diagnostic tests have significant drawbacks with respect to overall cost, speed, sensitivity and quantity and quality of input material required. FISH probes only target the regions of interest and reporting of a CNA is heavily dependent on the skill of the analyst. Only deletions affecting > 20% of the lymphocytes can be reliably detected by FISH. Lack of sensitivity of SS does not facilitate the detection of clinically important subclonal *TP53* variants (Landau et al. 2013; Nadeu et al. 2018).

The established method for determining somatic hypermutation of the clonotypic IGHV gene comprises two steps. First, PCR amplification of the clonal rearrangement is performed, followed by SS of the clonotypic IGHV gene. The sequencing result is compared to the reference germline IGH gene, using 2% difference from the reference sequence (also known as 98% germline homology) as the threshold for discriminating m-IGHV from u-IGHV (Hamblin et al. 1999; Rosenquist et al. 2017; Damle et al. 1999). The 98% cut-off was chosen based on Sanger sequencing to exclude potential polymorphic variant sequences and avoid the need to analyse the corresponding germline gene in each patient (Agathangelidis, Sutton, et al. 2018; Rosenquist et al. 2017). SS cannot determine the IGHV status in a 3-4% of CLL patients (Langerak et al. 2011). In a Spanish nationwide study, SS was unable to determine the IGHV status due to either insufficient sample quality (3.1%), polyclonal arrangements (1.96%), non-productive (1.56%) or biclonal mutations (4.04%) (Bosch et al. 2021).

Widespread IGHV analysis in clinical diagnostic laboratories is limited by the technical complexity, limited scalability and labour intensiveness of the method. Outside trials, IGHV status was tested in just 12-21% of pre-treatment CLL patients in the informCLL registry (Mato et al. 2020; Tang 2017). *TP53* deletions were tested by FISH in 28%-60% and mutation testing performed in 11% (Mato et al. 2020; Tejaswi et al. 2020; Tang 2017).

Technological developments facilitate the adoption of next-generation-sequencing (NGS), usually using sequencing-by-synthesis (SBS) into existing workflows within diagnostic laboratories. Advantages of NGS platforms for IGHV include: the ability to batch samples; direct clone determination and the ability to detect subclonal rearrangements. Laboratory capacity is a bottleneck for widespread access to IGHV testing for CLL patients, even in well-resourced healthcare systems (Bosch et al. 2021).

The introduction of short-read SBS methods facilitates greater understanding of disease pathobiology but some aspects of the clinical interpretation of immunogenetics determined by SBS remain uncertain. Deep targeted sequenced facilitates the identification of secondary immunoglobulin rearrangements in 24.4% of CLL patients (Stamatopoulos et al. 2017). Discordant cases (co-existing m-IGHV and one u-IGHV productive rearrangements) exist and are challenging to report (Agathangelidis et al. 2022).

The short-read length of SBS presents significant technical challenges to the analysis of privately rearranged loci such as the IGHV, which may poorly align with a reference genome. The costs of establishing and validating an NGS platform for IGH testing are prohibitive outside large academic centres. NGS-based testing for IGHV status becomes feasible when the laboratory receives sufficient samples to exploit the batch capacity of the platform and sequence large numbers of samples in a clinically relevant TAT. Furthermore, post sequencing analysis is complex.

6.1.3 Rationale for exploring Nanopore Sequencing

The MinION is a benchtop nanopore sequencer, capable of producing up to 40Gb data, at a rate of 450bp per second, depending on the flow cell used (Minervini et al. 2020). The initial investment costs are lower than the SBS platforms. With barcoding, multiple samples can be simultaneously sequenced. Nanopore sequencing (NS) is not limited by read length, enabling the sequencing of long linear read lengths (1-100kbp). The ability to sequence longer continuous reads may enable more accurate determination of the IGHV mutational status and may facilitate accurate identification of structural variants (SV), GC and chromothripsis (Cretu Stancu et al. 2017; Sakamoto, Sereewattanawoot, and Suzuki 2020). Low coverage sequencing screens for copy number alterations (CNA) and SV across solid tumours and acute leukaemia (Euskirchen et al. 2017; Norris et al. 2016; Au et al. 2019). Cryptic translocation can be revealed by NS using rapid shallow whole genome sequencing (sWGS) (Au et al. 2019). Detection of SV is not restricted by the comparatively higher error rate of NS (De Coster et al. 2019).

The principal limitation of NS remains the comparatively higher error rate although improvements in sequencing chemistry, the development of improved base-calling algorithms and post-sequencing error correction bioinformatics has increased the sequencing accuracy rate to 99.6%. Rapid sequencing times and low initial cost of the equipment makes MinION an attractive candidate for diagnostic use, particularly laboratories constrained by limited financial or staff resources.

6.1.4 Hypothesis and objective for Chapter 6

In this chapter, I explore the hypothesis that

- NS enables identification of *TP53* mutations in CLL
- NS facilitates reporting of IGHV status (VH family and germline homology) in CLL
- Shallow WGS using NS allows detection of Copy Number Alterations

6.2 Methods

6.2.1 Sample selection and sources of data for comparison

Burkitt's lymphoma cell line (TC413) gDNA, chosen for high tumour purity and known karyotype and CNA was used to evaluate sequencing parameters (Zviran et al. 2020).

MBL and early CLL: LYM1 and OXPLORED cohort

DNA was extracted from PBMC from MBL and Rai 0 CLL participants in the LYM1 study (Section 2.4). Data on IGHV family and germline homology determined by SS was provided by the study team. For OXPLORED samples, IGHV family and germline homology was determined by IgCaller from WGS data.

Advanced CLL

Advanced CLL samples were previously characterised by SBS and/or SNP arrays (Stamatopoulos et al. 2017; Burns et al. 2018; Stamatopoulos et al. 2018). Samples with known CNA, *TP53* aberrations and IGHV rearrangements were chosen.

Eleven advanced CLL samples were sequenced using nanopore by Dr Adam Burns (Burns et al. 2019). I reanalysed the data from these 11 samples and expanded the cohort using LYM1 and OXPLORED samples.

Table 6.2.1.1 Sources of samples and data for comparison

Cohort	***TP53* mutation**	**IGHV status**	**Copy Number Alterations**
LYM1 Study (n=25)	Targeted NGS (Chapter 3)	Sanger sequencing[1]	FISH[1]
OXPLORED (n=7)	--	IgCaller on WGS using SBS[3]	--
Advanced CLL (n=11)	SBS[2]	SBS[2]	SNP array/WGS[2]

Samples and data sources are further described in Section 2.1.1.

1. *Data provided from LYM1 study cohort*
2. *Data provided by Dr Adam Burns and previously published (Stamatopoulos et al. 2017; Burns et al. 2018; Stamatopoulos et al. 2018)*
3. *WGS performed with Genomics England and the National Genome Reference Library. IgCaller output fasta files generated by Dr Kate Ridout. I analysed the fasta files using IMGT/V-QUEST and interpreted results as per ERIC guidelines*

6.2.2 Primers, library preparation and sequencing

Based on previous publication on leader primers, six primers homologous to the conserved areas at the five prime of the leader sequences for each of the six human IGHV families were designed (Campbell et al. 1992) (Section 2.7.2). Nanopore libraries were prepared for each DNA sample using PCR amplification and LSK-109 library kit and sequenced on the MinION (Section 2.7.3 -2.7.5).

6.2.3 Identification of *TP53* mutations

Nanopore fastq files were aligned to hg19 using the minimap2 and NGMLR. Sniffles was used to call *TP53* indels and VarScan called SNVs (Section 2.7.8).

6.2.4 Determining IGHV status

Two post-sequencing data analysis strategies were developed and assessed. To account for variation in coverage, each nanopore fastq file was downsized to 10,000 reads.

(Method A) Analysis using Canu to generate a consensus IGH sequence

This method was developed with Dr Adam Burns. Briefly, a preliminary consensus sequence from the fastq using canu v2.1.1 was generated. Alignment of fastq, using the canu consensus sequence as a reference, was performed with Minimap2. The SAM file was converted into BAM format, sorted and indexed with samtools. A second and final consensus sequence was generated with Nanopolish. The IMGT/V-QUEST database determined the level of germline homology, per ERIC guidelines (Rosenquist et al. 2017).

(Method B) Determination of VH family from nanopore reads using IMGT High V-QUEST

I developed this method independently. Using Seqtk, I converted the fastq file to fasta format for uploading to IMGT/HighV-QUEST, analysed including 'insertions and deletions' parameters. The resulting summary file was filtered to identity the most frequently occurring V-GENE family.

Finally, I used IMGT/HighV-QUEST (Method B) to identify the VH family and generation a polished consensus sequence (Method A) for the germline homology (Figure 6.2.3.1). Further details can be found in Section 2.7.8.

Figure 6.2.4.1 Summary of the post-sequencing analysis for IGHV status

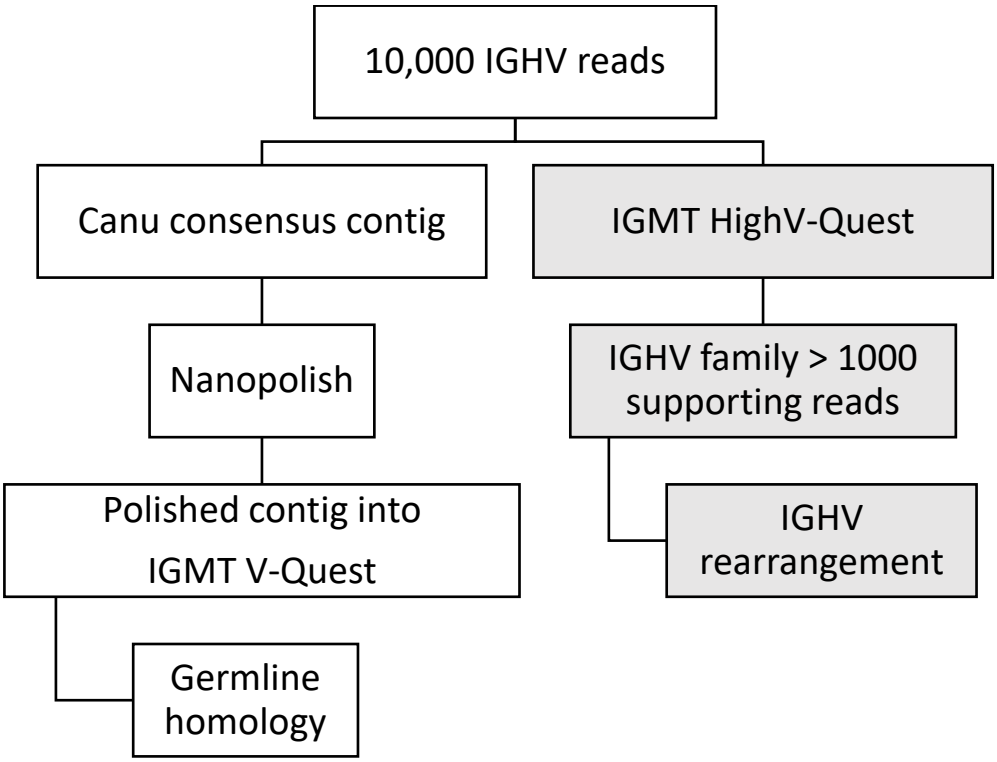


Figure 6.2.4.1: Flow chart summarising the methods to determine germline homology (left, white boxes) and immunoglobulin rearrangement (right, grey boxes) from nanopore sequencing of IGHV amplicon libraries.

6.2.5 Copy Number calling using NS and QDNASeq

This method was developed with assistance from Dr Adam Burns and initial results were presented at American Society for Hematology meeting 2018 (Burns 2018).

Shallow WGS libraries were produced from long-fragment gDNA using the RAD-004 protocol (Section 2.7.4). Coverage was calculated using the samtools (version 1.19) depth function. CNA were determined using QDNASeq (version1.24.0) and visualised using the integrated Rplots (R version 4.0.3) and in Integrative Genomics Viewer (IGV). For the gDNA experiments, the default cutoffs=log2(c(deletion = 0.5, loss = 1.5, gain = 2.5, amplification = 10)/2) for copy number calls were used.

For CLL374, *TP53* and IGHV amplicons were combined with a RAD004 WGS library and sequenced on a single run (Section 2.7.7).

6.2.6 Statistical analysis

Filtering of IGHV reads and data visualisation were performed in R using packages tidyverse and ggpubr. Relationships between the germline homology by NS versus SBS and NS versus SS were calculated using Pearson's correlation coefficients.

6.3 Results

Key biomarkers used for prognostication and treatment selection in CLL include IGHV status, *TP53* mutations and copy number aberrations at del17p (*TP53* locus), del11q, tri12 and del13q. Detection of 5 ≥ CNAs by karyotyping identifies high-risk CLL (Baliakas et al. 2019).

In this chapter, I report that NS of *TP53* amplicons correctly identified all *TP53* mutations in selected CLL samples. NS also identified immunoglobulin heavy chain rearrangements and germline homology in CLL, provided 10% of the input DNA carries the clonal rearrangement. The method failed to correctly identify VH4 family rearrangements, suggesting further optimisation of the primers or amplification conditions is needed.

Finally, as proof of principle, I demonstrate the feasibility of determining IGHV, *TP53* and CNAs in a single sequencing run.

6.3.1 Development of method for post-sequencing identification of IGHV status using control samples and synthetic fastq

To understand how the proportion of clonal reads impacted on the post-sequencing analysis and to optimise the analysis pipeline, a series of fastq files to represent varying proportions of normal polyclonal sequences and clonal B-cell sequences were created (Table 6.3.1.1). Sample OXP10073 was chosen as the source of clonal VH sequences as 9156/10,000 (91.56%) sequences mapped to IGH-gene. Sample OMDC-control1 was used as the polyclonal control in which no IGHV rearrangement should be detected. The three Test fastq files were analysed by Method A: Canu & nanopolish pipeline and Method B: IMGT/HighV-QUEST alone.

Table 6.3.1.1. Fastq files created to estimate limit of detection

Name	Number of clonal OXP10073 reads	Number of control (polyclonal) reads	Total
Test50	5000	5000	10,000
Test25	2500	7500	10,000
Test10	1000	9000	10,000
OMDC control 1	0	10000	10,000

Performance of (Method A) Analysis using Canu to generate a consensus IGH sequence

Using the canu and nanopolish tools, the correct IGHV family and germline homology can be detected in Test50 and Test25 but is not detected in Test10. However, an additional IGHV3-33 clone is reported in the Test10 and Test25. The polyclonal sample incorrectly records a IGHV3-NL1*01 F (Table 6.3.1.2). Refining the canu parameters to increase the overlap required did not alter the result. Canu creates a consensus contig, without regard to the number of supporting reads.

Table 6.3.1.2 Method A: IGHV calls using canu and nanopolish in the test files

Sample	*IGHV-GENE and allele*	*Germline homology % (with ins/del events)*	*Functionality*
OXP10073 ('Test100')	IGHV3-72*01	93.86%	P
Test50	IGHV3-72*01	93.86%	P
Test25	IGHV3-72*01	93.86%	P
	IGHV3-33*08	98.25%	UP
Test10	IGHV3-33*08	98.25%	UP
OMDC control 1	IGHV3-NL1*01 F	96.83%	UP

P productive, UP unproductive

Performance of (Method B) VH family from nanopore reads using IMGT/High V-QUEST

Next, to correct for the errors in Method A, I developed and tested an alternative method (Method B) to identify the IGHV family using IGMT/HighV-QUEST was developed (Section 2.7.8). Taking the output of IMGT/HighV-QUEST, I filtered the sequences by IGHV family, retaining VH families with > 100 supporting reads. This approach correctly called no rearrangement in the polyclonal sample and increasing numbers of reads supporting IGHV3-72*01 in the test samples (Table 6.3.1.3). However, germline homology was low, and all rearrangements were called as unproductive.

Table 6.3.1.3 Method B: IGHV calls using unpolished nanopore fastq and IGMT/HighV-QUEST

Name	*Reads*	*IGHV-Gene and allele*	*Germline homology % (with ins/del events)*	*Functionality*
OXP10073 ('Test100')	9156	IGHV3-72*01	89.2%	UP
Test50	4378	IGHV3-72*01	88.36%	UP
Test25	2271	IGHV3-72*01	88.19%	UP
Test10	894	IGHV3-72*01	88.03%	UP
OMDC Control 1	--	No rearrangement	--	--

P productive, UP unproductive

Combination of the two approaches is needed to determine VH family and germline homology

Then, I tested a final method which combined the two approaches, using IMGT/HighV-QUEST output to identify a dominant VH family and the canu and nanopolish packages to report germline homology (Figure 6.3.2.1 and Figure 6.3.1.1). This method successfully identified the correct IGHV rearrangement and germline homology for all test samples except Test 10. In the Test 10 sample, the number of reads supporting the VH family was less than 1000 reads (or less than 10% of the total number of reads) and 1000 reads was set as the threshold for reliable detection of a dominant VH family.

Figure 6.3.1.1: Number of reads supporting IGHV3-72*01 (red) in OXP10073, the three test files and the normal control by IMGT/HighV-QUEST

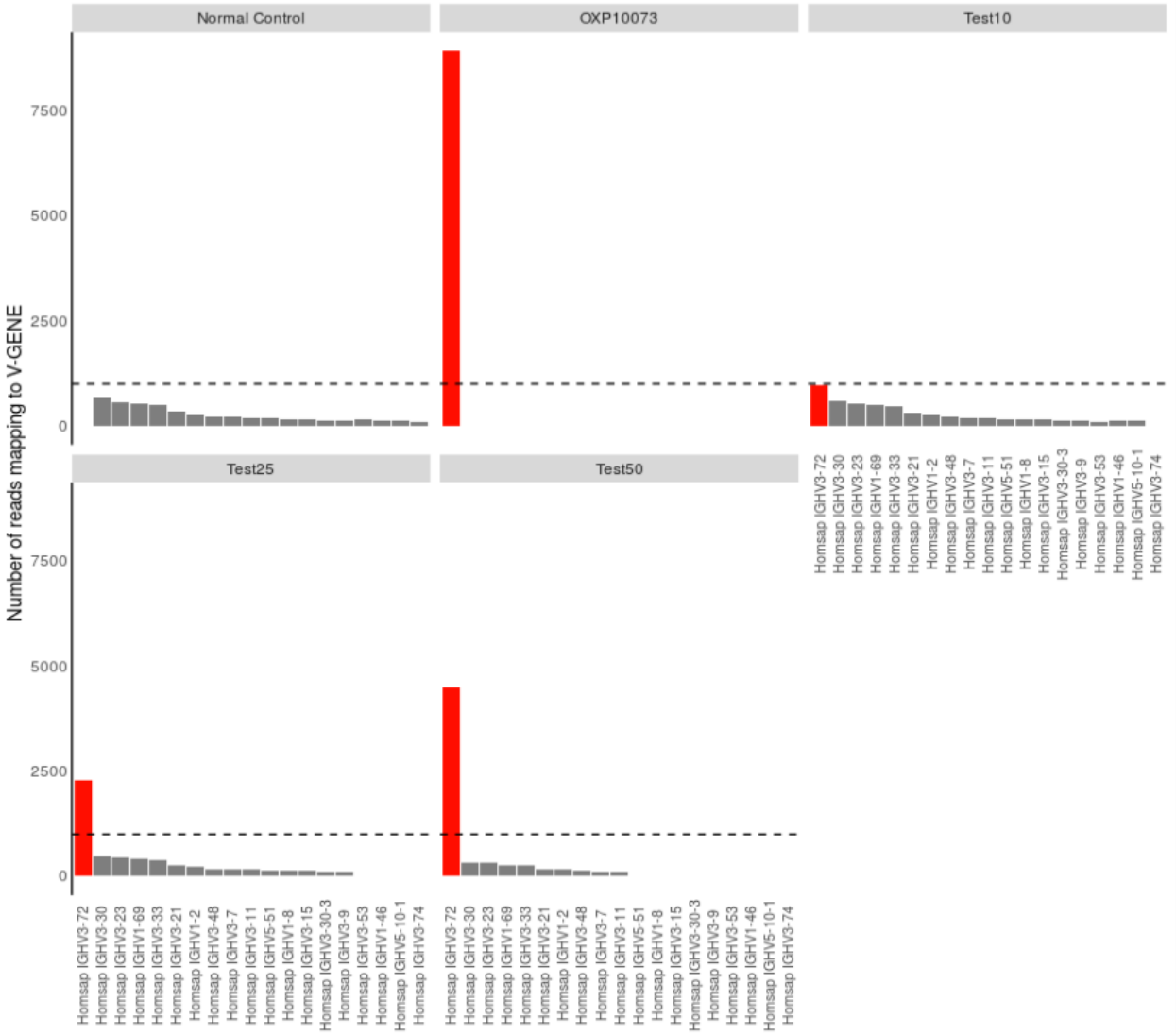


Figure 6.3.1.1: Frequency bar chart of rearranged VH-gene families called by IGMT HighV-Quest from the unpolished nanopore reads. The normal control is shown in the top left panel. The correct VH-gene family is highlighted in red and other reads shown in grey. The dashed line represents 1000 supporting reads, the threshold set to call a VH-gene family result.

6.3.2 Nanopore sequencing determines IGHV status in clinical CLL samples: Sequential application of IMGT/HighV-Quest, custom filtering and canu and nanopolish

Next, I applied the combined method to 11 advanced CLL samples and 33 early CLL samples in whom the IGHV status had been previously determined (Figure 6.3.2.1). In Table 6.3.2.1, samples in which the dominant VH family and germline homology were concordant with the SS or SBS result are listed. Samples with discordant results are displayed in Table 6.3.2.3.

NS reports VH family and stereotypy

For advanced CLL cohort, the combined approach was concordant with the SBS result in all 11 patient samples (Table: 6.3.2.1). In addition, both VH rearrangements were detected in CLL351, CLL371 and ARC477 using SBS were also found using NS.

In the early CLL samples, NS was concordant in 23/32 (75%) patients. In two cases (BM-3999, BM-1087), NS identified an additional clonal VH rearrangement undetected by SS (Table 6.3.2.2). Results from all seven OXPLORED samples were all concordant with SBS, including V3-21 #subset2 in OXP10067 (Figure 6.3.2.2, Table 6.3.2.1). No immunoglobulin rearrangement was detected by NS or SBS in OXP10074 and OXP10076.

Results were discordant with SS (n=4) or no dominant sequence was identified by NS (n=5) for nine LYM1 patients (Table 6.3.2.3). In seven of nine patients, SS identified a VH4 family, implying a problem with the VH4 primer or PCR cycling conditions. Three of the four discordant results were MBL samples (BM-0590, BM-4193 and BM-4990), with total lymphocyte count of 4.2-4.9 $x10^9$/L, suggesting the clonal B-cell population may be below the 10% limit of detection identified in Section 6.3.1. BM-0409 had sufficient lymphocytes (lymphocytes 10.3$x10^9$/L) to provide a representative sample and PCR amplification of the nanopore library was successful but less than 1000 reads mapped to the correct VH family (V3-9). A later sample from the same patient, taken when the lymphocyte count was 28$x10^9$/L and sequenced using NS was able to identify VH3-9 as the dominant clone.

NS reports germline homology

In samples when the V-gene family identified by NS agreed with the SS or SBS result, the germline homology obtained from the polished nanopore reads was highly concordant with SBS ($R=1$, $p = 1.1x10^{-15}$, Figure 6.3.2.3) and with SS ($R = 0.99$, $p = 6.5x10^{-13}$, Figure 6.3.2.4).

Results are consistent across sequencing runs

Next, I prepared an amplicon library using a new DNA aliquot and sequenced the library on a different MinION run. In all four cases, the VH family obtained was identical and germline homology concordant (Table 6.3.2.5).

Figure 6.3.2.1: Flowchart of clinical samples and sources of IGHV rearrangement data

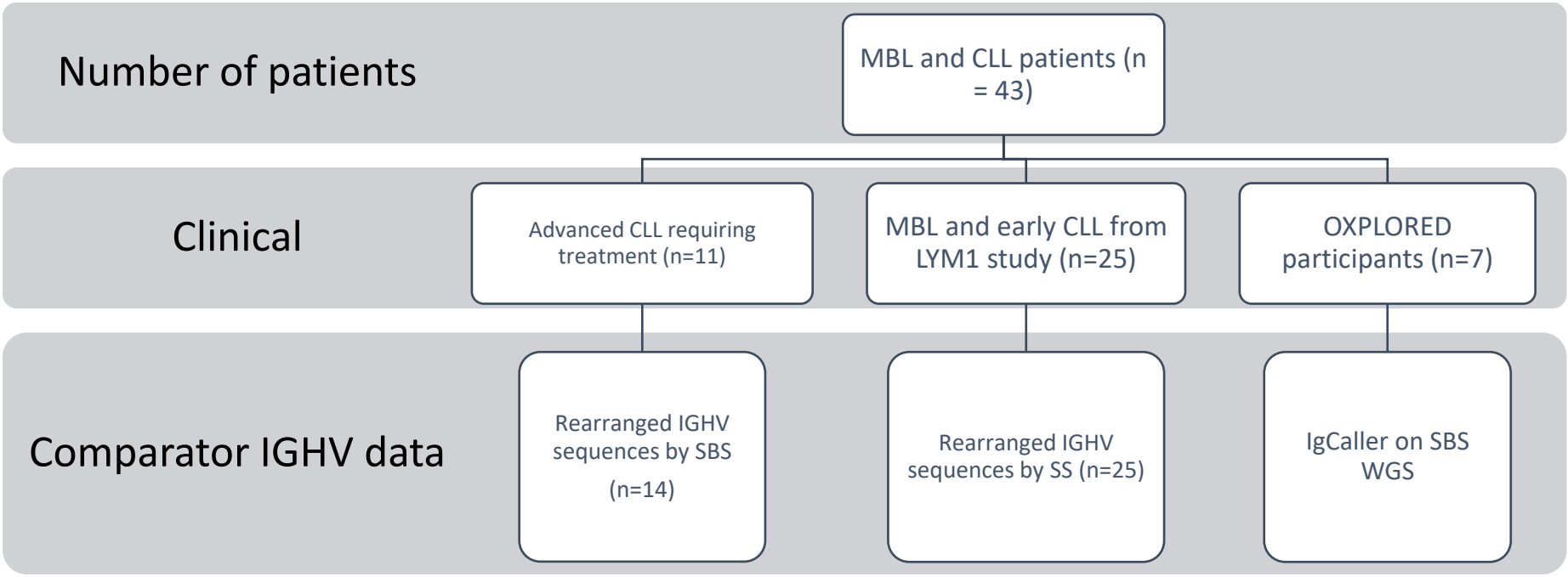


Figure 6.3.2.1: Flowchart summarising samples sequenced and sources of comparative data. SS Sanger sequencing; SBS sequencing-by-synthesis

Table 6.3.2.1. Comparison of nanopore sequencing IGHV results with other methods

	Nanopore sequencing (Combined approach)				Comparator		
Patient ID	Family	Supporting reads	Germline homology	Productivity	Family	Germline homology	Method
CLL145a	IGHV3-23	3330	96.17%	P	IGHV3-32	96.17%	SBS
CLL331[a]	IGHV3-72	2118	95.92%	P	IGHV3-72	95.83%	SBS
CLL345[a]	IGHV3-49	4133	100.00%	P	IGHV3-49	100.00%	SBS
CLL347[a]	IGHV3-9	3665	100.00%	P	IGHV3-9	100.00%	SBS
CLL374[a]	IGHV3-11	4732	98.96%	P	IGHV3-11	99.00%	SBS
ADM293[a]	IGHV1-69	3679	100.00%	UP	IGHV1-69	100.00%	SBS
ADM455[a]	IGHV3-21 #subset2	5552	96.88%	UP	IGHV3-21 #subset2	96.70%	SBS
ARC603[a]	IGHV3-7	5510	100.00%	UP	IGHV3-7	100.00%	SBS
CLL351[a]	IGHV3-72	4297	92.18%	P	IGHV3-72	92.38%	SBS
	IGHV3-13	1998	91.7%	P	IGHV3-13	91.81%	SBS
CLL371[a]	IGHV3-23	1983	92.36%	P	IGHV3-23	92.36%	SBS
	IGHV3-30	1953	90.94%	UP	IGHV3-30	90.88%	SBS
ADM477[a]	IGHV3-30	1209	98.96%	UP	IGHV3-30	98.56%	SBS
	IGHV1-3	3161	99.65%	P	IGHV1-3	99.65%	SBS
BM-0306	IGHV1-69	5025	99.65%	P	IGHV1-69	99.65%	SS
BM-0363	IGHV3-9	6730	100.00%	P	IVHV3-9	99.65%	SS
BM-0409	IGHV3-9	5087	96.17%	UP	IGHV3-9	96.53%	SS
BM-0434	IGHV2-5	3925	91.00%	P	IGHV2-5	91.07%	SS
BM-0981	IGHV3-23	4612	91.61%	UP	IGHV3-23	ND	SS
BM-1087	IGHV3-11	2116	100.00%	P	IGHV3-11	100.00%	SS
	IGHV1-8	3767	100.00%	P	--	--	--
BM-1993	IGHV3-30	2112	92.98%	UP	IGHV3-30	93.06%	SS
BM-2101	IGHV5-51	6672	95.14%	P	IGHV5-51	94.76%	SS
BM-2462	IGHV3-9	4314	100.00%	UP	IGHV3-9	100.00%	SS
BM-2483	IGHV1-69	6242	100.00%	P	IGHV1-69	100.00%	SS
BM-3999	IGHV3-30	613	100.00%	P	IGHV3-30	100.00%	SS

	IGHV1-3	7300	100.00%	P	--	--	--
BM-3214	IGHV3-74	5800	94.10%	P	IGHV3-74	94.10%	SS
BM-3943	IGHV1-69	3922	100.00%	P	IGHV1-69	100.00%	SS
BM-3980	IGHV3-30	6328	100.00%	P	IGHV3-30	100.00%	SS
BM-4208	IGHV3-33	8046	100.00%	P	IGHV3-33	100.00%	SS
BM-4478	IGHV1-69	4758	91.67%	P	IGHV1-69	93.75%	SS
OXP10067	IGHV3-21*01 #subset2	5554	97.90%	P	IGHV3-21*01 #subset2		WGS and IgCaller
OXP10070	IGHV3-30*03	8240	95.47%	P	IGHV3-30*3	95.83%	WGS and IgCaller
OXP10071	IGHV3-53*01	8127	92.61%	P	IGHV 3-53*01	92.98%	WGS and IgCaller
OXP10072	IGHV3-72*01	8444	93.84%	P	IGHV3-72*01	94.98%	WGS and IgCaller
OXP10073	IGHV3-72*01	8743	93.86%	P	IGHV3-72*01	94.22%	WGS and IgCaller
OXP10074	No dominant rearrangement		--	--	No dominant rearrangement		WGS and IgCaller
OXP10075	IGHV3-9*01	6633	96.50%	P	IGHV3-9*01	96.53%	WGS and IgCaller
OXP10076	No dominant rearrangement				No dominant rearrangement		

a. *Denotes advanced CLL samples. For these samples, the library preparation was done by Dr Adam Burns and I reanalysed the fastq data.*

Table 6.3.2.2 Secondary IGHV rearrangements detected by NS in five samples

	IGHV by nanopore sequencing				Comparator method		
Sample	**VH family**	**Supporting reads**	**Germline homology**	**Functionality**	**VH family**	**Germline homology**	**Method**
CLL351	IGHV3-72	4297	92.18%	P	IGHV3-72	92.38%	SBS
	IGHV3-13	1998	91.7%	P	IGHV3-13	91.81%	SBS
CLL371	IGHV3-23	1983	92.36%	P	IGHV3-23	92.36%	SBS
	IGHV3-30	1953	90.94%	UP	IGHV3-30	90.88%	SBS
ADM477	IGHV3-30	1209	98.96%	UP	IGHV3-30	98.56%	SBS
	IGHV1-3	3161	99.65%	P	IGHV1-3	99.65%	SBS
BM-3999	IGHV3-30	613	100.00%	P	IGHV3-30	100.00%	SS
	IGHV1-3	7300	100.00%	P	ND	--	--
BM-1087	IGHV3-11	2116	100.00%	P	IGHV3-11	100.00%	SS
	IGHV1-8	3767	100.00%	P	ND	--	--
ND = not detected; SBS = sequencing-by-synthesis; SS = Sanger sequencing; P = productive rearrangement; UP = unproductive rearrangement							

Figure 6.3.2.2 Dominant VH clones by NS in the OXPLORED cohort

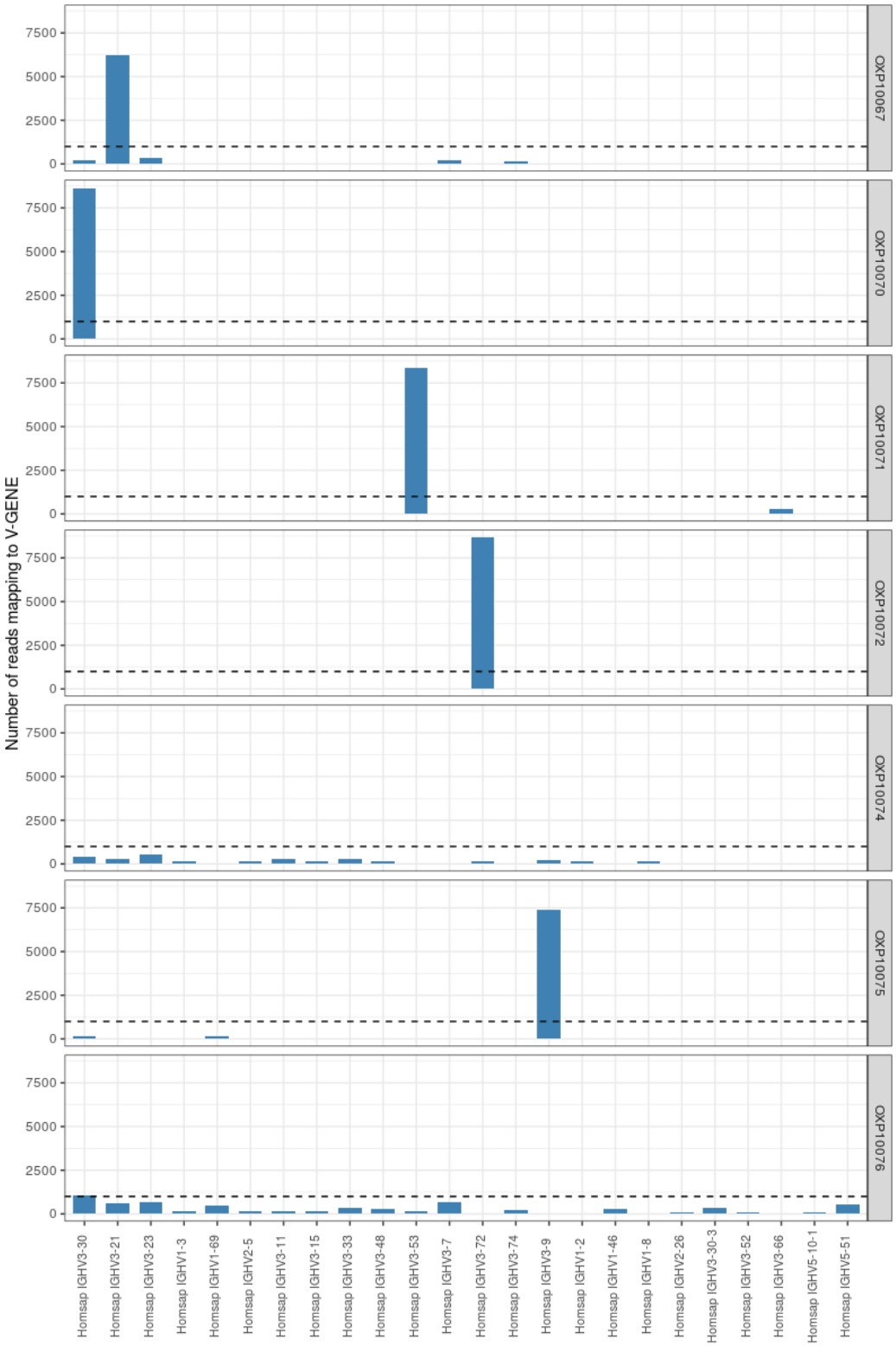


Figure 6.3.2.2: Frequency bar chart of rearranged VH-gene families called by IGMT HighV-Quest from the unpolished nanopore reads. The dashed line represents 1000 supporting reads, the threshold set to call a VH-gene family result.

Table 6.3.2.3 Discordant VH family results between Nanopore and Sanger Sequencing (SS) are predominantly seen in the V4 family rearrangements

	IGHV by NS			IGHV by SS		
Sample	**VH family**	**Supporting reads**	**Germline homology**	**VH family**	**Germline homology**	**Comment**
BM-0590	IGHV3-9	2016	99.70%	IGHV4-34	92.28%	Discordant
BM-3644	IGHV6-1	1456	98.60%	IGHV4-39	90.04%	Discordant
BM-4193	IGHV3-21	1019	90.30%	IGHV1-18	93.06%	Discordant
BM-4990	IGHV3-13	2964	100.00%	IGHV4-39	100.00%	Discordant
BM-0160	No dominant rearrangement			IGHV4-38	99.63%	
BM-0409	No dominant rearrangement			IGHV3-9	95.92%	
BM-0680	No dominant rearrangement			IGHV4-34	96.53%	
BM-3856	No dominant rearrangement			IGHV4-34	91.58%	
BM-3823	No dominant rearrangement			IGHV4-34	95.92%	

Figure 6.3.2.3: Comparison of the percentage germline homology for IGHV rearrangements detected by SBS and by Nanopore Sequencing

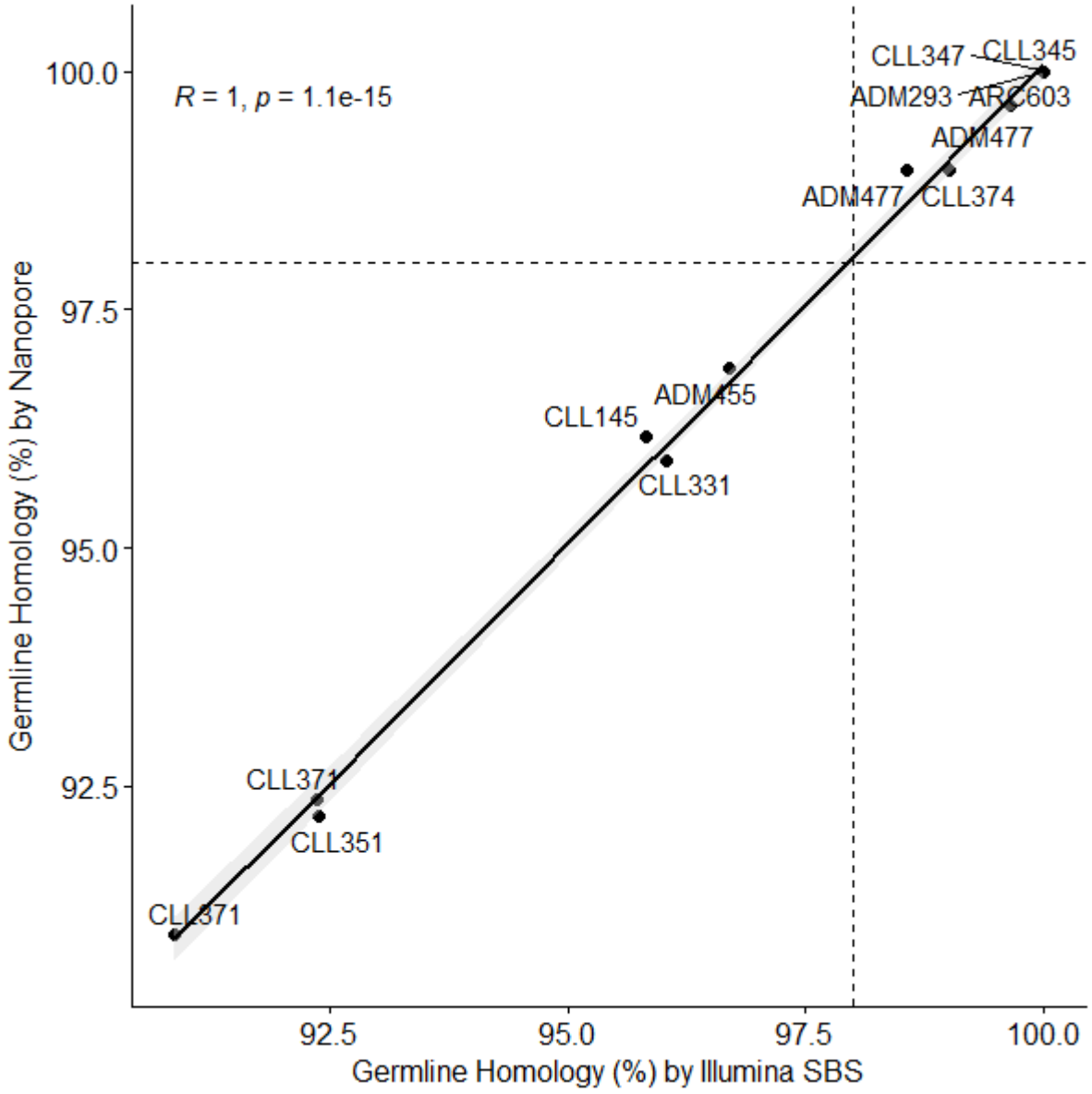


Figure 6.3.2.3: Germline homology of the 13 IGHV sequences detected by both nanopore sequencing (y-axis) and Illumina Sequencing-by-synthesis (x-axis) are depicted. The second rearrangement in CLL351 is not depicted as it was detectable by one method only. The dashed line represents the 98% threshold. R = pearsons correlation coefficient.

Figure 6.3.2.4: Comparison of the percentage germline homology for IGHV rearrangements detected by SS and by Nanopore Sequencing

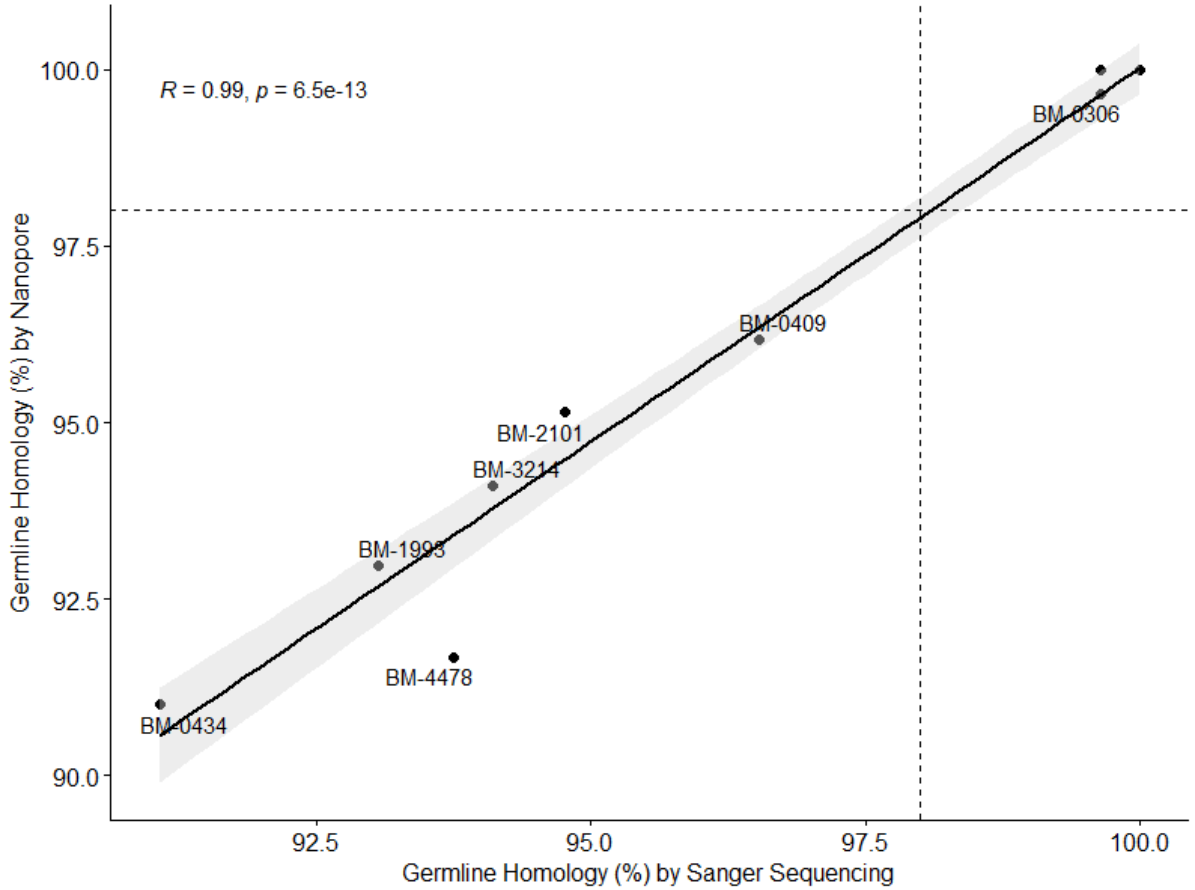


Figure 6.3.2.4: Germline homology of the 16 IGHV sequences detected by both nanopore sequencing (y-axis) and Sanger sequencing (x-axis) are depicted. Patient labels for the 8 patients with 100% germline homology overlap and are not displayed in the figure. The dashed line represents the 98% threshold. R = pearsons correlation coefficient.

Table 6.3.2.5 Inter-run variability for NS to detect IGHV status

PATIENT ID	RUN ID	VH FAMILY	SUPPORTING READS	GERMLINE HOMOLOGY
BM-0981	24	IGHV3-23	2071	91.61%
	17	IGHV3-23	4612	91.67%
BM-4208	23	IGHV3-33	3922	100%
	25	IGHV3-33	8046	100%
BM-1087	16	IGHV1-8 IGHV3-11	3767 2116	100% 92.4%
	24	IGHV1-8 IGHV3-11	2588 1804	100% 92.23%
BM-3644	16	IGHV6-1	1456	98.2%
	24	IGHV6-1	939	98.6%

6.3.3 Nanopore sequencing detects *TP53* SNVs and indels

Seven pathogenic variants in *TP53* were detected by SBS in the advanced CLL samples and these seven samples were tested by NS. All seven *TP53* mutations were found by NS at similar allele frequencies (R= 0.8, p= 0.03)(Table 6.3.3.1, Figure 6.3.3.1).

Sixteen early CLL samples previously characterised as *TP53* wild-type using capture hybridisation SBS (Chapter 3) were subject to *TP53* testing by NS. NS correctly identified the common *TP53* polymorphism (chr17:g.7578210A>G) found at heterozygous allele frequency using SBS in BM3214. However, NS called a pathogenic *TP53* variant chr17:g.7577094C>T in one Rai 0 CLL (BM-4478) sample tested, which was absent from capture hybridisation sequencing (Chapter 3) and was interpreted as a false positive.

Table 6.3.3.1 *TP53* variants detected by nanopore and Illumina SBS

Patient	COSMIC	HGVSc	Protein	VAF by Nanopore Sequencing	VAF by Illumina SBS
CLL374[a]	COSM10812	NM_000546.6:c.722C>T	p.Ser241Phe	0.76	0.79
CLL345[a]	COSM11224	NM_000546.6:c.394A>C	p.Lys132Gln	0.60	0.54
CLL331[a]	COSM45622	--	Glu154Serfs*6	0.97	0.93
CLL145[a]	COSM10660	NM_000546.6:c.818G>A	p.Arg273His	0.47	0.48
ARC603[a]	COSM10654	NM_000546.6:c.637C>T	p.Arg213Ter	0.54	0.82
ADM477[a]	COSM10722	NM_000546.6:c.853G>A	p.Glu285Lys	0.53	0.64
ADM293[a]	COSM46030	NM_000546.6:c.974G>T	p.Gly325Val	0.45	0.50
BM-4478	COSM10704	NM_00546.6:c.844C>T	p.Arg282Trp	0.13	ND

ND= not detected

a. Denotes advanced CLL samples. The library preparation was done by Dr Adam Burns.

Taken together, my results show clinically relevant *TP53* variants can be detected using NS but NS is limited by the error rate of the platform. Guidelines for *TP53* variant testing recommend a sequencing assay which reliably calls variants detectable by SS and avoids false-positive calls above the SS detection limit (i.e. 10%) (Malcikova et al. 2018). NS fulfils the former requirement but reported a variant at 0.13, absent from the SBS (Table 6.3.3.1).

Figure 6.3.3.1 Comparison of *TP53* variant allele frequencies by Illumina SBS and by nanopore sequencing

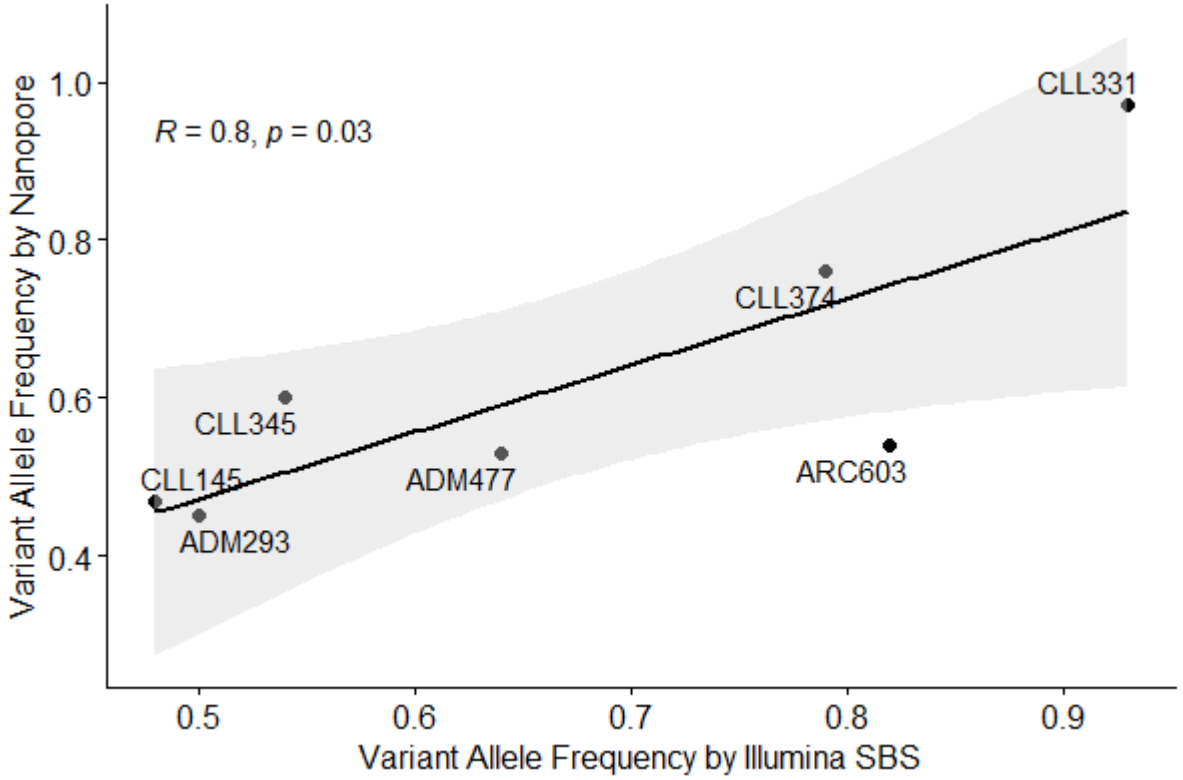


Figure 6.3.3.1: Concordance between the Allele Frequencies of the *TP53* variants detected by nanopore sequencing (y-axis) and Illumina SBS (x-axis) BM-4478 is not depicted. R = Pearson's correlation coefficient.

6.3.4. Nanopore sequencing detects of Copy Number Alterations in gDNA

A cell line DNA with known CNAs (CA46; TC413, Burkitt lymphoma, Figure 7.3.1.1) was used initially to assess the ability of NS to detect CNAs and to establish the number of reads needed for studying patient samples. I prepared a RAD004 library from CA46 (TC413) gDNA and sequenced it for 70 hours on the MinION Mk1b.

Figure 6.3.4.1: Published Consensus Karyotype for Burkitt lymphoma cell line CA46 (TC413)

46,X,–Y, dup(1)(q21q32),dup(7)(q?21q?22),t(8;14)(q24;q32), i(13)(q10),+16/47,sl,+der(7;13)(q10;q10).ish dup(1)(q31.3q11) (RP11-198A7++,D1Z1++),dup(7)(q11q22.1)(RP11-229D13++,RP5-1099C19++),der(13)dup(13)(q31.3q21)del(13)(q31.3q34)(RP11-29P20++,RP11-307D17++,RP11-195A5–,RP11-650D9–)/der(13)dup(13)del(13)×2/ider(13)(q10)dup(13)del(13)

From (Zimonjic, Keck-Waggoner, and Popescu 2001)

289,804 reads of median length 8.71kb, mapped to the human genome and were used to generate Copy Number calls using QDNASeq. The mean depth was 1.35x. QDNASeq showed the gains at Chr1, Chr7, Chr13 and Chr16 and the loss at Chr13 (Figure 6.3.4.1). The losses at Chr4 and Chr6 correspond to artefacts (further described in Chapter 7).

Next, I estimated the number of reads required to call CNA using the high-purity, long reads gDNA and QDNASeq. I down sampled the TC413 long-read gDNA fastq file to create synthetic fastq files representing 200,000 to 50,000 reads respectively (Table 6.3.4.1) As read numbers fell, the precision in detected CNA declined. At 100,000 reads, corresponding to depth of 1.11x, the gain affecting 11q appears as a gain of the whole chromosome. The amplification at Chr7 is visible in the QDNASeq output (Figure 6.3.4.1, Table 6). Potential false positive CNA affecting chr6, chr15 and chr20 are reported as depth reduces (Table 6.3.4.1) but are absent from the karyotyping report and the x1.35 depth sWGS of TC413 gDNA. By 50,000 reads (depth 1.05x) only a gain of chr7 was detected and the complex changes affecting chromosome 7 and chromosome 13 are absent (Figure 6.3.4.3, Table 6.3.4.1). Again, false positive CN calls were observed. Based on this modelling, I targeted 250,000 reads for subsequent gDNA experiments.

Figure 6.3.4.2: NS and QDNASeq analysis of Burkitt cell line gDNA TC413 (1.35x depth)

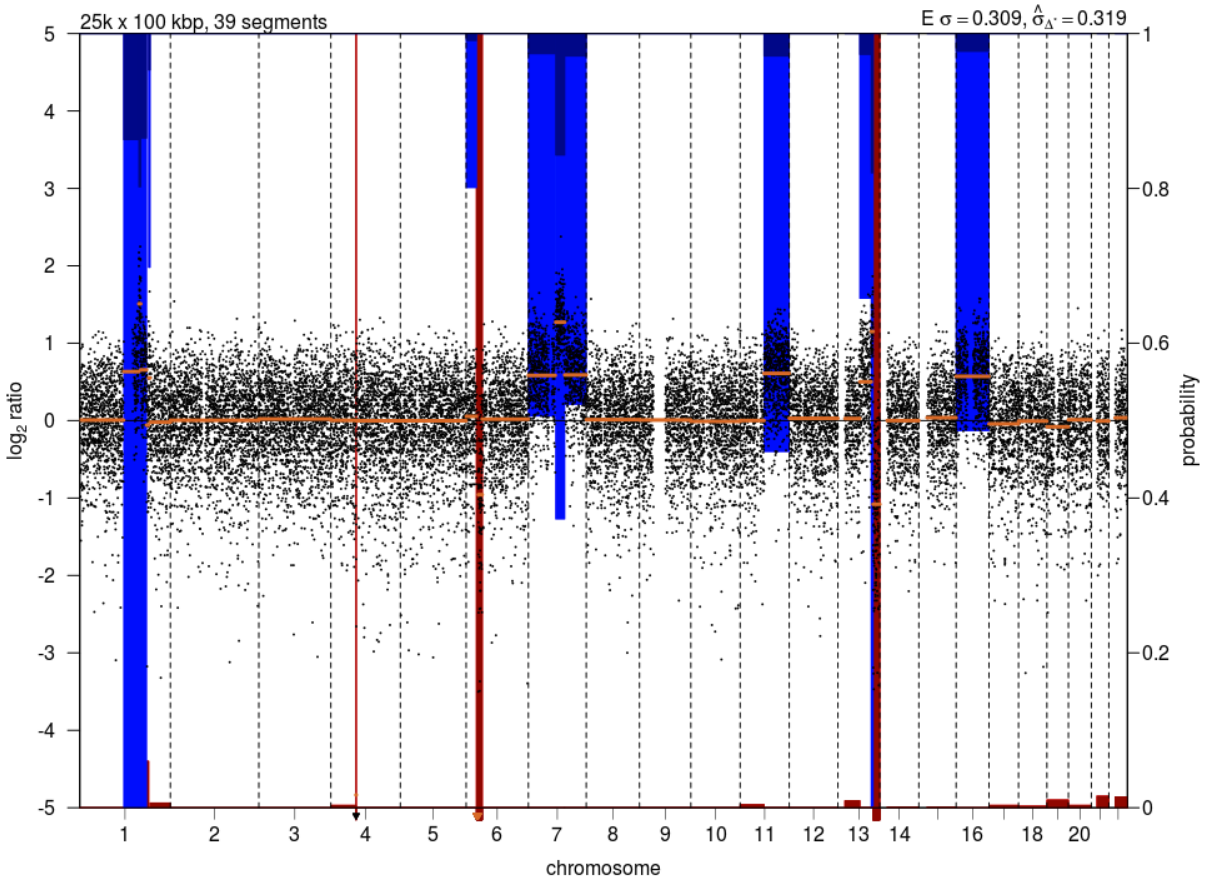


Figure 6.3.4.2: Relative CN values at each autosomal chromosome (1.35x depth). The log2 equals zero line indicates a likely diploid genome value. Blue represents regions of copy number gain and wine-red shows copy number losses.

Figure 6.3.4.3: NS and QDNASeq analysis of Burkitt cell line gDNA (TC413) (1.05x depth)

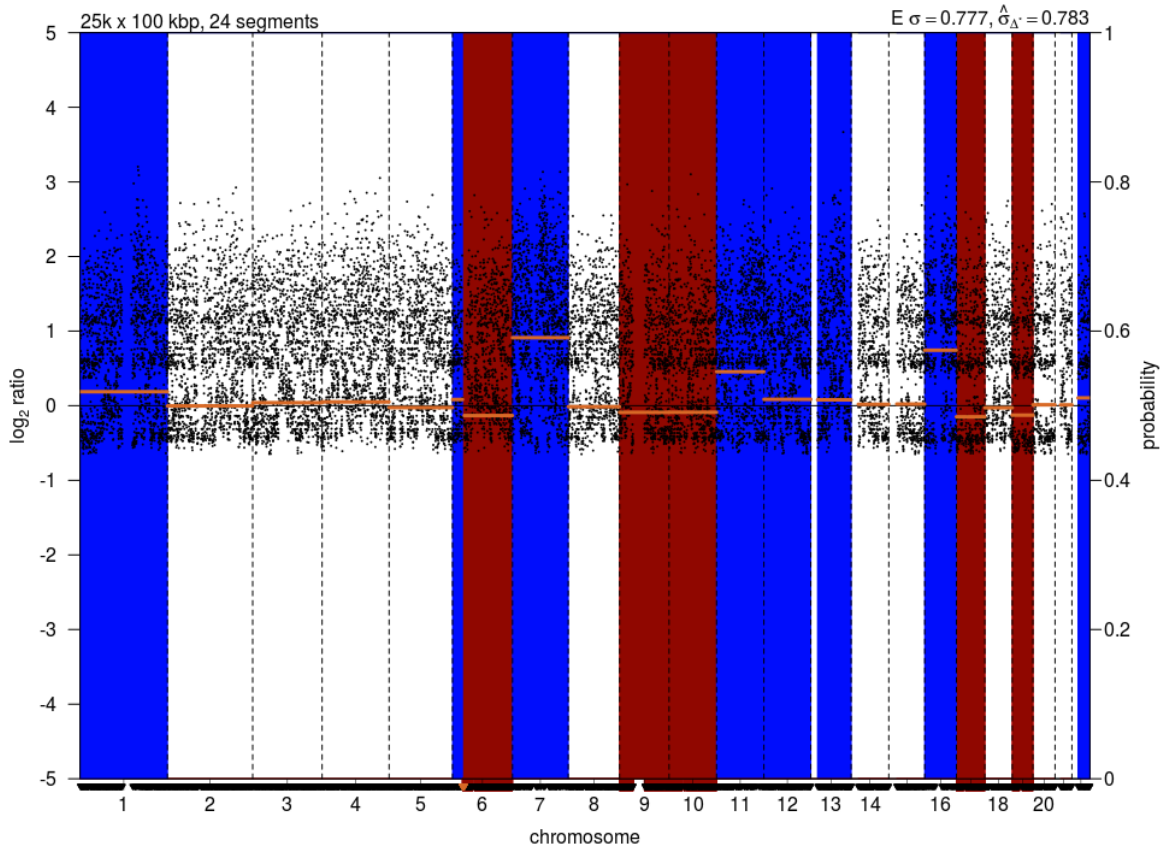


Figure 6.3.4.3: Relative CN values at each autosomal chromosome (1.05x depth). The log2 equals zero line indicates a likely diploid genome value. Blue represents regions of copy number gain and wine-red shows copy number losses.

Table 6.3.4.1: Summary QDNASeq copy number calls from TC413 gDNA (long-reads) with declining coverage

Chrom	Start position	SV Type	SV Length	Original 1.35x	100,000 reads 1.11x	50,000 reads 1.05x
1	**121,200,001***	**CN Gain**	**41,400,000**	**Detected**	**Detected[i)]**	**Detected[ii)]**
1	**162,600,001***	**CN Gain**	**5,200,000**	**Detected**	**Detected**	
1	**167,800,001***	**CN Gain**	**25,700,000**	**Detected**	**Detected[i)]**	
6	400,001	CN Loss	28,100,000	Detected	Detected	Detected
7	**1***	**CN Gain**	**74,000,000**	**Detected**	**Detected**	
7	**75,200,001***	**CN Gain**	**24,600,000**	**Detected**	**Detected**	**Detected[ii)]**
7	**100,000,001***	**CN Gain**	**59,100,000**	**Detected**	**Detected**	
11	65,800,001	CN Gain	69,000,000	Detected	ND	Detected
13	**60,300,001***	**CN Gain**	**72,300,000**	**Detected**	**Detected[i)]**	
13	**91,700,001***	**CN Gain**	**5,000,000**	**Detected**	**Detected**	**Detected[ii)]**
13	**96,700,001***	**CN Loss**	**18,469,878**	**Detected**	**Detected**	
16	**100,001***	**CN Gain**	**90,000,000**	**Detected**	**Detected**	**Detected**
19	300,001	CN Loss	62,900,000	Detected	Detected	Detected

Only CNA larger than 5Mb are displayed. CNAs marked * are detected in the published karyotype
At 100,000 reads, the CNA marked i) have different start positions or lengths.
At 50,000 reads, the multiple CNA merged to form a single CN Gain, marked ii) spanning the whole chromosome.
ND not detected

Having established conditions for detection of CNAs using NS, I applied the method to selected advanced CLL samples. This work was done in collaboration with colleagues at OMDC and has been reported in a preliminary capacity (Burns 2018). Advanced CLL samples were selected for known CNAs, previously determined using array or Illumina SBS WGS (Burns et al. 2018).

NS achieved a mean depth of coverage of 1.6x (range 0.1-3.4x) with a mean read length of 6.3kbp. Del(17p13.1) events, of 20.3Mb and 21.3Mb respectively, were identified in two samples (CLL331 and ARC603) (Figure 6.3.4.4). Del(13q) was identified in four samples (CLL331, CLL345, CLL374 and ARC603) with an 800kbp minimally deleted region (MDR) encompassing *MIR16*, *MIR15A* and *DLEU1*. NS and QDNASeq analysis (on default callBins thresholds) failed to call one instance of del(17p13.1) affecting CLL347. However, nanopore WGS depth for CLL347 was 0.7x. By adjusting the thresholds on QDNAseq for CLL347, the loss affecting the *TP53* locus could be seen despite the shallow depth of sequencing.

Sequencing of early CLL samples from the LYM1 cohort (n=2), selected for CNA identified by FISH was performed independently. NS and QDNASeq correctly identified the tri12 in BM2483 (Figure 6.3.5.2) but failed to call del13q in BM4082, even with adjusting QDNAseq thresholds. The lymphocyte counts were similar between the two samples (lymphocytes $16.8x10^9$/L in BM4082, lymphocyte $15x10^9$/L in BM2483) but the proportion of cells observed carrying the CNA was lower at 36% in BM4082. Although my numbers are small and results should be interpreted with caution, this implies a threshold for detection using the NS and QDNAseq method (further explored in Chapter 7).

Figure 6.3.4.4: Visualisation of tri12 in early CLL case (BM-2483)

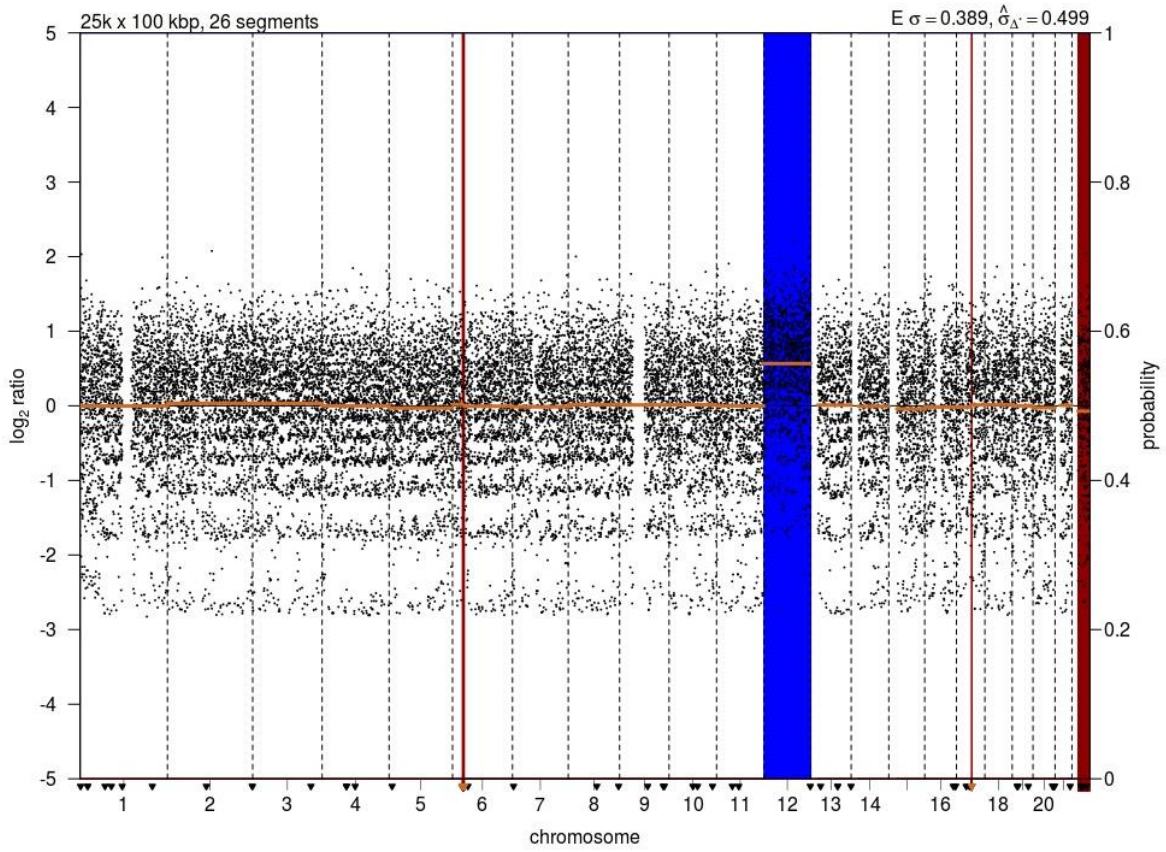


Figure 6.3.4.4: Graphical representation of tri(12): Log2 ratio plots from QDNAseq visualising the copy number profile for BM-2483, with blue indicating copy number gain. The tri12 was detected in 53% of cells using FISH. X-axis = genomic location by chromosome, primary Y-axis = log2 transformed read depth, secondary Y-axis = CNA call confidence value.

6.3.5 Proof of principle: combining *TP53*, *IGHV* and shallow WGS to detect clinically relevant molecular features in a single run

For CLL374, *TP53* and IGHV amplicons were combined with a RAD004 WGS library and sequenced on a single run (Section 2.7.7). Concordant VH family, germline homology, TP53 variant and del17p results were obtained by NS and SBS (Table 6.3.5.1).

Table 6.3.5.1 *TP53,* IGHV and CNA detected by NS in a single run in CLL374

Test	Nanopore sequencing	Sequencing-by-synthesis
***IGHV* status**	IGHV3-11 98.96% Unmutated	IGHV3-11 99.00% Unmutated
***TP53* mutation**	COSM10812 NM_000546.6:c.722C>T p.Ser241Phe VAF 0.76	COSM10812 NM_000546.6:c.722C>T p.Ser241Phe VAF 0.79
Copy number alterations	Copy Number loss Chr17:1 -chr17:22,000,000	Copy number loss Chr17:1-24,000,000

6.4 Discussion

6.4.1 Overview

This work explores the feasibility of NS using MinION to identify key CLL biomarkers from a single sequencing run, as an alternative to SS and SBS.

Equitable access to modern diagnostics requires a sequencing technology accessible to patients and healthcare professionals, deliverable in resource constrained settings and reporting results within a standardised framework. The high cost of equipment purchase, consumables and maintenances is a barrier to the widespread availability of molecular diagnostic testing. The ONT MinION platform is competitively priced and comparatively cheap to maintain. Minervini et al estimated the cost of IGHV sequencing using the MinION at €20 per sample (Minervini et al. 2021).

6.4.2 IGHV status

This chapter indicates NS can identify VH family (using unprocessed nanopore reads and IGMT/HighV-QUEST) and determine germline homology (using canu and nanopolish informatics tools sequentially). A lower limit of 1000 reads (10% of 10,000 reads per sample) is the minimum required to call a VH family and distinguish clonal rearrangements from background non-specific reads. The intrinsic error rate of NS can be circumvented using bioinformatics tools to build a consensus contig. The IGHV method presented in this chapter had concordant germline homology results with short-read sequencing in the 14 VH rearrangements found in the 11 pre-treatment CLL cohort. In the MBL and early CLL cohort, NS determined germline homology in one sample (BM-0981) unclassified by SS.

For the OXPLORED samples, the IGHV status was determined using IgCaller, a python tool to characterise immunoglobulin gene rearrangements from short-read WGS datasets (Nadeu et al. 2020). My NS results provided the same IGHV status in all cases. NS was able to detect IGHV status in OXP-10075, which was undetermined by IgCaller.

The ERIC consortium produces guidelines standardising the performance and interpretation of IGHV mutational analysis (Agathangelidis, Sutton, et al. 2018; Langerak et al. 2011; Rosenquist et al. 2017; Agathangelidis et al. 2022). The nanopore-based method differs from the ERIC guidance in two aspects: nucleic acid starting material and use of leader primers. While both gDNA and complementary DNA (cDNA) are suitable for IGHV analysis, cDNA is recommended as cDNA preferentially identifies functional rearrangements and can define the isotype of the functional rearrangement (Rosenquist et al. 2017). Analysis of gDNA avoids the reverse transcription step needed to prepare cDNA and is also suitable for use on stored materials. Using gDNA facilitated the inclusion of archived samples from the retrospective cohort. The non-productive or non-transcribed rearrangement on the second allele is detectable using gDNA. Using gDNA and NS, additional non-productive IGHV rearrangements were found in seven CLL samples. The nanopore method to determine IGHV status was not formally tested with cDNA.

My work uses leader primers targeting the IGHV region. Leader primers, recommended in the recent update of the ERIC guidelines (Agathangelidis et al. 2022; Langerak 2022), permit sequencing of the IGHV region, providing greater precision in determining the homology to the closest germline gene. Their principal disadvantage is the reduction in PCR efficacy. Heut

et al conducted a comparative study of both FR1 and leader primers on 34 CLL samples, focusing on samples with germline homology 97-99% range. Longer amplicons improved the accuracy of classification for CLL patients in the 97-99% germline homology range. However, they were unable to determine the mutational status in 8/34 (24%) of cases when the longer leader primers are used (Huet et al. 2020). Conversely, IGHV status was undetermined for five samples examined using nanopore. My nanopore primers have not undergone the intensive optimisation of the BIOMED-2 consortium (van Dongen et al. 2003). Optimisation of primers and PCR conditions to improve efficacy (and improve amplification of the VH4-34 group) may reduce the number of failures.

In contrast, framework-region 1 (FR-1) primers capture only a proportion of the IGHV region meaning that germline homology will be approximated and risking underestimating somatic hypermutation (Rosenquist et al. 2017). Their higher PCR efficacy means they are frequently used in diagnostic laboratories. The ERIC guidelines recommend the use of FR1 primers are restricted to the samples where multiplex PCR with leader primers has failed (Agathangelidis et al. 2022).

In summary, my work demonstrates high levels of concordance of the NS approach, with standard SS and SBS methods. However, samples with VH4-34 by SS produced discordant results by NS, which is most probably due to ineffective PCR amplification of the longer primer. Further work is therefore needed to optimise these primers.

6.4.3 Detection of *TP53* mutations

In CLL*, TP53* aberrations resulting from either deletion of the 17p13 locus and/or mutation-s within the *TP53* gene, are associated with poor responses to chemoimmunotherapy (Fabbri and Dalla-Favera 2016; Brieghel et al. 2019; Nadeu et al. 2016). International guidelines mandate *TP53* testing prior to first and subsequent lines of therapy (Hallek et al. 2018; Malcikova et al. 2018). In line with ERIC 2018 recommendations for *TP53* analysis, gDNA is used and exons 1-11, with splicing sites are covered (Malcikova et al. 2018).

I show reproducibility in detecting at high VAF, *TP53* mutations between nanopore and Illumina SBS. My results agree with one previous study comparing SS and nanopore in 12 *TP53* disrupted CLL samples (Minervini et al. 2016), in which the lowest *TP53* variant detectable by NS after error correction was at 31% (Minervini et al. 2016). The error correction strategies employed in this paper differed from mine in that they were developed for sequencing microbial genomes.

ERIC *TP53* sequencing guidelines recommend NGS assays should at minimum, reliably identify variants detectable by SS and avoid false positives calls with VAFs about the Sanger detection limit (i.e. > VAF 10%). The intrinsic error rate of the nanopore platform creates vulnerabilities to reporting false positive *TP53* mutations, especially close to the error threshold. In the BM-4478 sample, the variant at VAF 0.13 could not be identified in the corresponding SBS file. This error was not due to insufficient material, since the patient had lymphocyte count 33.3 $x10^9$/L (above the recommended threshold of $10x10^9$/L) (Malcikova et al. 2018) and 99% of the lymphocyte gate co-expressed CD5/CD19, implying that the clonal B-cell population was well-represented in the input DNA. The NS method therefore

needs further improvement to reliably detect variants at the recommended allele frequency.

6.4.4 Detection of CNAs

Genomic structural changes inform the diagnosis and prognosis of many B-lymphoid malignancies. In clinical diagnostics, fluorescence *in situ* hybridisation (FISH) for translocations and copy number alterations (CNA) is the most widely available tool for identification of genomic structural changes. Chromosomal banding analysis (CBA) is reserved for lymphoid malignancies that reliably yield sufficient metaphases in culture. Nanopore based shallow WGS offers an accessible platform to determine multiple copy number changes simultaneously and derive genomic complexity in CLL. The method is applicable to other cancers where multiple large chromosomal gains or losses are prognostic (Ben-David and Amon 2020; Manier et al. 2017; Roberts et al. 2008) and influence treatment decisions (Sonneveld et al. 2016; Dohner et al. 2022). The laboratory equipment and workflow are more straightforward than conventional karyotyping or WGS using SBS potentially expanding access to genome-wide data for resource constrained settings.

Translocations and fusion genes detected by FISH are critical to define and manage haematological cancers (Alaggio et al. 2022; Khoury et al. 2022). Although not explored in this work, nanopore based WGS has been used to detect translocations in acute leukaemia (Au et al. 2019) and pre-implantation genetic testing (Xia et al. 2023).

6.4.5 Summary and future directions

Immunogenetic analysis plays a central role in understanding and managing CLL. My results show that long-read NS can detect VH family and germline homology. Further work is required to optimise the PCR efficiency of the VH4 primer and facilitate detection of the full range of VH families.

My results demonstrate that using current MinION platform and post-sequencing informatics tools, NS can detect *TP53* variants, but this is limited to detecting *TP53* variants at high AF. NS is unable to reliably detect sub-clonal variants, which retain clinical importance and remains vulnerable to false positives. Detection of CNA using NS has already been demonstrated in our laboratory (Burns 2018). As an extension of this, in a single proof-of-principle experiment, *TP53* mutations, IGHV status and CNA by shallow WGS (see also Chapter 7) for CLL374 could be determined in a single sequencing run. NS can provide all clinically relevant CLL biomarkers in one sequencing run on one flow cell.

My work has several additional limitations. First, as a proof-of-principle study, sample number is small and precludes accurate quantification of sensitivity or specificity to detect SV. The threshold for detection is estimated using artificial datafiles but not confirmed in a dilution experiment. A prospective and multi-centric evaluation of the approach presented here is needed to demonstrate robustness across laboratories.

Chapter 7: Development and application of a nanopore sequencing method for shallow Whole Genome Sequencing using cell-free DNA

7.1 Introduction

Shallow whole genome sequencing (sWGS) of tumour gDNA provides genome-wide data and circumvents the need for cell culture (Maclachlan et al. 2021; Elnenaei et al. 2019). However, availability of high-quality tumour DNA and the turnaround time (TAT) limit the application of sequencing-based techniques into clinical practice. Most biopsy samples are formalin, fixed and paraffin embedded (FFPE), limiting the DNA quality available. DNA extracted from FFPE specimens is subject to nucleic acid fragmentation, DNA cross-links, localised denaturation and strand breaks and C>T deamination artefacts (Do and Dobrovic 2015). In a study of paired FF and FFPE tissue biopsies, WGS of the FFPE DNA identified less than half of the CNAs found in the corresponding FF due to non-uniform coverage and the disrupted baseline log2ratio signal (Robbe et al. 2022). In addition, collection of fresh tissue from nodal biopsies is hampered by the heterogenous routes lymphoma patients take to diagnosis (Howell et al. 2013). Another factor hindering sequencing efforts is that to deliver cost-effective SBS, libraries are sequenced when Illumina flow cells are at or near capacity. Translating this into clinical practice requires centralisation at a laboratory serving a large population size. As sufficient sample numbers need to be collected and prepared before a sequencing run is economical, this leads to delays in the delivery of clinical results.

In order to overcome the problems with DNA quality and heterogeneity, analysis of ctDNA by sWGS might offer an alternative, minimally invasive platform for extensive genomic characterisation and is being explored using sequencing-by-synthesis (SBS) in solid tumours (Raman et al. 2020; Van Roy et al. 2017). The potential delay arising from the economic need to sequence multiple samples in parallel can be circumvented by using NS which sequences fewer samples per run on less expensive flow-cells.

7.1.1 Detection of copy number alterations from plasma cell-free DNA

The long-read ability of nanopore sequencing (NS) may facilitate accurate identification of SV, GC and chromothripsis (Cretu Stancu et al. 2017; Sakamoto, Sereewattanawoot, and Suzuki 2020). Low coverage sequencing can screen for CNA and SV across a range of solid tumours (Euskirchen et al. 2017; Norris et al. 2016).

Given that many of the advantages of NS derive from the ability to capture longer DNA fragments, it seems counter-intuitive to explore the role of NS in cell-free DNA which circulates as short fragments. However, the potential advantages of NS over SBS merit investigation. Preliminary studies report that large genomic events such as aneuploidy and CNAs can be detected. In a feasibility study, NS of maternal plasma cell-free DNA identified 0.11% of short reads mapping to the Y chromosome and correctly identified mothers carrying a male foetus (Cheng et al. 2015).

7.1.2 Genomic features differ between nodal tissue and circulating lymphocytes in Richter transformation (RT)

Reporting of CNAs and SVs is crucial in RT. Indeed, recent studies using WGS of RT tumours revealed a median of 18 CNAs and 37 SVs, far higher than the 4 CNAs and 5 SV seen in the preceding CLL sample (Nadeu et al. 2022). WGS of synchronous circulating neoplastic lymphocytes cells (PB-CLL) and nodal tissue (RT-tissue) showed RT is characterised by increased mutational burden and CNAs (Klintman et al. 2021).Application of NS to cell-free DNA offers potential for minimally invasive detection of tumour CNA. In the context of nodal lymphomas like RT, it may provide a minimally invasive assessment of aneuploidy and GC in the nodal tissue.

7.1.3 Hypothesis and objectives

In this chapter, I aim to evaluate shallow WGS of plasma cell-free DNA using NS, using samples from the O-CHOP clinical trial. I hypothesised that tumour derived CNAs would be detectable in ctDNA and that ctDNA would reveal additional CNAs compared with WGS performed on the tumour (RT-tissue) and PB-CLL.

There are two principal issues to be overcome in applying NS to cell-free DNA: the problem of low abundance and the problem of short fragments. I performed a series of experiments using fragmented gDNA to elucidate the impact of short fragments and variable tumour content on sequencing and copy number detection.

My main objectives were

- Develop an NS method for detection of aneuploidy and CNA in fragmented cell line gDNA.
- Apply this method to the detection of large CNAs (and thereby derive Complex Karyotype) in plasma cell-free DNA from RT clinical trial patients, previously characterised by Illumina SBS WGS.

7.2 Methods

7.2.1 Sample selection

I began with fragmented gDNA extracted from the Burkitt's lymphoma cell line (TC413) described in Chapter 6. Cell-line DNA was chosen for high tumour purity, known karyotype and likely clonal CNAs. Next, I used pre-treatment cell-free DNA samples from the O-CHOP RT clinical trial (described in Chapters 2 and 5) to identify CNA in cell-free DNA.

Figure 7.2.1.1 Summary of the samples and workflow for Chapter 7

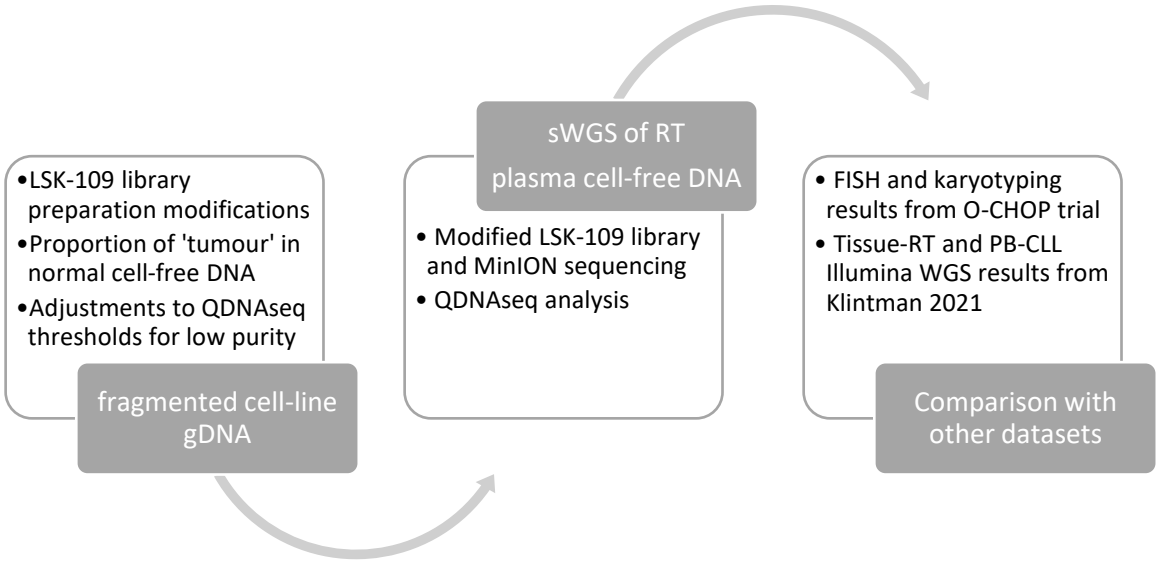


Figure 7.2.1.1: Schematic showing the workflow for Chapter 7. gDNA = genomic DNA, sWGS = shallow WGS, O-CHOP = ofatumumab and CHOP trial for RT described in Chapter 2, RT = Richters transformation, PB-CLL = DNA extracted from circulating neoplastic lymphocytes, Tissue-RT = DNA extracted from the tissue biopsy showing Richter's transformation

7.2.2 Preparation of DNA library for nanopore sequencing

Extraction of gDNA and cell-free DNA, quantification and quality assurance procedures are described in Section 2.4. The library preparation for cell-free DNA and loading on the MinION for sequencing is described in Section 2.7.

Briefly, I modified the ONT LSK-109 Native Barcoding Amplicons protocol for fragmented gDNA and cell-free DNA. Modifications include using all available cell-free DNA as input, changing the ratio of AMPure bead ratio to 1.8x and addition of 2.5x AMPure bead clean-up step after DNA end-repair in order to retain shorter fragments and 'overloading' the MinION flow-cell for sequencing (Martignano et al. 2021).

7.2.3 In-silico experiment evaluates the number of reads required to call CNA

To simulate varying proportion of ctDNA in a plasma cell-free DNA, I combined the down sampled fastq files in varying ratios (Burkitt fragmented gDNA: control) to simulate variable degrees of sWGS coverage and tumour fraction.

Table 7.2.3.2: Variable proportions of the fragmented gDNA (Burkitt cell line TC413) and normal non-cancer control cell-free DNA

File name	Fragmented gDNA (Burkitt cell line TC413)	Pooled normal cell-free DNA	Ratio
'0.33 ctDNA'	250,000 reads	500,000 reads	1:2
'0.25 ctDNA'	125,000 reads	500,000 reads	1:4
'0.2 ctDNA'	100,000 reads	500,000 reads	1:5
'0.1 ctDNA'	50,000 reads	500,000 reads	1:10

7.2.4 Copy number calls using QDNASeq

CNA were determined using QDNASeq (version1.24.0) using autosomal chromosomes and bin size of 100kbp (Section 2.7.5). Data was visualised using the integrated Rplots (R v4.0.3) and in Integrative Genomics Viewer (IGV).

Interpretation of the log2CNT threshold

In the fragmented gDNA and cell-free DNA experiments, the log2CNT threshold for calling CNA was adjusted to detect aberrations with the following cutoffs:
homozygous deletion < -0.15 < loss < -0.1 < normal < 0.1 gain < 1 < amplification.

In mixture of diploid tumour and normal DNA sample, I would expect to detect two copies of each chromosome from the non-tumour DNA and two copies from the tumour, 4 detected/4 expected or log2(1) = 0.

If the tumour had a chromosomal gain and the sample comprised 50% tumour, I would anticipate 5 copies of the chromosome, not four copies which gives 5 detected/4 expected or log2(5/4) = 0.32.

With a cell-free DNA sample, values of 50% tumour are unlikely. The ranges were chosen to reflect an anticipated tumour fraction of 25-30%. This would mean 4.3 copies detected/4

expected or log2(4.3/4) = 0.1. This is referred to as the 'optimised threshold' throughout Section 7.3 and 7.4.

7.2.5 Comparison of copy number alterations from nanopore sequencing with copy number alterations from the O-CHOP trial dataset

FISH results from O-CHOP trial samples were obtained from the Oxford Clinical Trials Office. Karyotyping by CBA results were available for CH1005 and CH1007.

Illumina SBS WGS datasets prepared from FFPE tumour DNA (mean depth 88x) and PB-CLL (mean depth 87x) have been previously published (Klintman et al. 2021). Library preparation, sequencing and copy number calling are outlined in Section 2.1.3. These WGS results were considered the 'ground truth.'

For comparison with the plasma cell-free DNA, the work focused on CN gains and losses larger than 10Mb, clinically tested CNA (del13q, del11q, trisomy 12 and del17p) and presence or absence of Complex Karyotype (CK). Samples with ≥ 5 CNA of at least 10Mb in length were considered to have CK. Copy neutral loss of heterozygosity events were excluded. The CNAs greater than 10Mb were compared between ctDNA and tissue-RT DNA, and ctDNA and PB-CLL. CNAs were concordant if the same CNA type (gain or loss) reported and ≥ 50% overlapping region, as described (Smolander et al. 2021).

To resolve discordance calls from the ctDNA sWGS, Illumina SBS WGS bam files from four O-CHOP tissue-RT and PB-CLL samples (CH1001, CH1005, CH1007 and CH1011) were re-analysed using QDNASeq with same 100kbp bin size.

7.3 Results

Cell-free DNA is characterised by both short fragment sizes and low-abundance. Both attributes represent challenges to NS. In addition, the majority of cell-free DNA comes from non-tumour cells and tumour-derived ctDNA fragments represents a minority of the total.

Therefore, in Section 7.3.1 I tested the impact of short reads on shallow WGS. I fragmented the TC413 gDNA and adapted the amplicon library preparation for shorter reads. Next, I sought to establish the proportion of ctDNA required to call CNAs in Section 7.3.2. I created "dilutions" with variable proportions of reads from fastq files of fragmented gDNA and normal cell-free DNA. I performed a corresponding wet lab experiment. My optimisation established that I needed to generate a minimum of 500,000 reads and use samples with ctDNA concentration of 10% or more for a pilot clinical evaluation presented in 7.3.3.

7.3.1. Detection of copy number alterations in fragmented gDNA from the Burkitt cell line (TC413) using nanopore shallow Whole Genome Sequencing (sWGS)

I fragmented the Burkitt cell line TC413 gDNA (as described in Section 2.6.3) and prepared libraries of fragmented gDNA using LSK-109 Native Barcoding Amplicon protocol, modified as outlined in Section 2.7. 25fmol of library was loaded onto the MinION and sequenced for 70 hours. The median read length was 280bp, consistent with the fragmented gDNA and ONT barcodes and adaptors. Over 3.9 $x10^6$ reads aligned to hg19, corresponding to a mean depth of 1.29x.

Eighteen chromosomal regions affected by CNAs were detected by NS using either unfragmented or fragmented gDNA from TC413. Of these, 16 were seen in the unfragmented gDNA (1.35x depth) and 18 in the fragmented gDNA (1.29x depth) (Table 7.3.1.1). By adjusting the log2R threshold, the gains at Chr1 and Chr7 in the fragmented gDNA were revealed as multiple closely related gains of comparable total length (Figure 7.3.1.1). The fragmented gDNA identified additional gains at Chr2, Chr10 and Chr17 and an additional deletion at Chr3, absent from the unfragmented gDNA. Eleven of the 18 CNAs paralleled changes described in the published karyotype (Zimonjic, Keck-Waggoner, and Popescu 2001).

Minor size differences were observed between the original and fragmented gDNA results. For example, in the gDNA, the gain at Chr13 runs from Chr13q31.2 to Chr13q32.1 (5Mb in length) but in the corresponding fragmented gDNA the corresponding gain measures 4.9Mb in length. At this point, it was possible to identify CNAs using short reads (~280bp in length) in highly pure samples. Next, I investigated the number of reads required to call a copy number change in the fragmented gDNA. I down sampled the fragmented gDNA TC413 fastq file to 1,000,000 reads, 500,000 reads and 250,000 reads. I ran each subsampled file through the QDNASeq copy number caller. At the 1,000,000 reads (1.11x depth), CNAs of 0.5Mb or fewer are no longer detectable. The duplication at Chr2:7,200,001 (Chr2p25.2) and the deletion at Chr3:176,900,001 (Chr3q26.32) are absent (Table 7.3.1.1). Differences in CNA length are seen, with the duplication at Chr1 measuring 2.3Mb shorter.

By adjusting the log2CNT threshold to call copy number changes, the complex changes affecting Chr1, Chr7 and Chr13 could be correctly identified with 250,000 reads of

fragmented gDNA. Deletions less than 1Mb (for example, Chr3q26.32 of 0.5Mb) were not detected at 250,000 reads even with the optimised log2CNT threshold. Based on these results, a minimum of 250,000 reads (1.02x coverage) is required to call CNA in the highly pure fragmented gDNA sample. To account for potential run-to-run variation, I targeted a minimum of 500,000 reads for cell-free DNA sequencing on a single MinION flow cell.

Figure 7.3.1.1 Visualisation of Copy Number Alternations detected using NS and QDNAseq analysis of TC413 fragmented gDNA

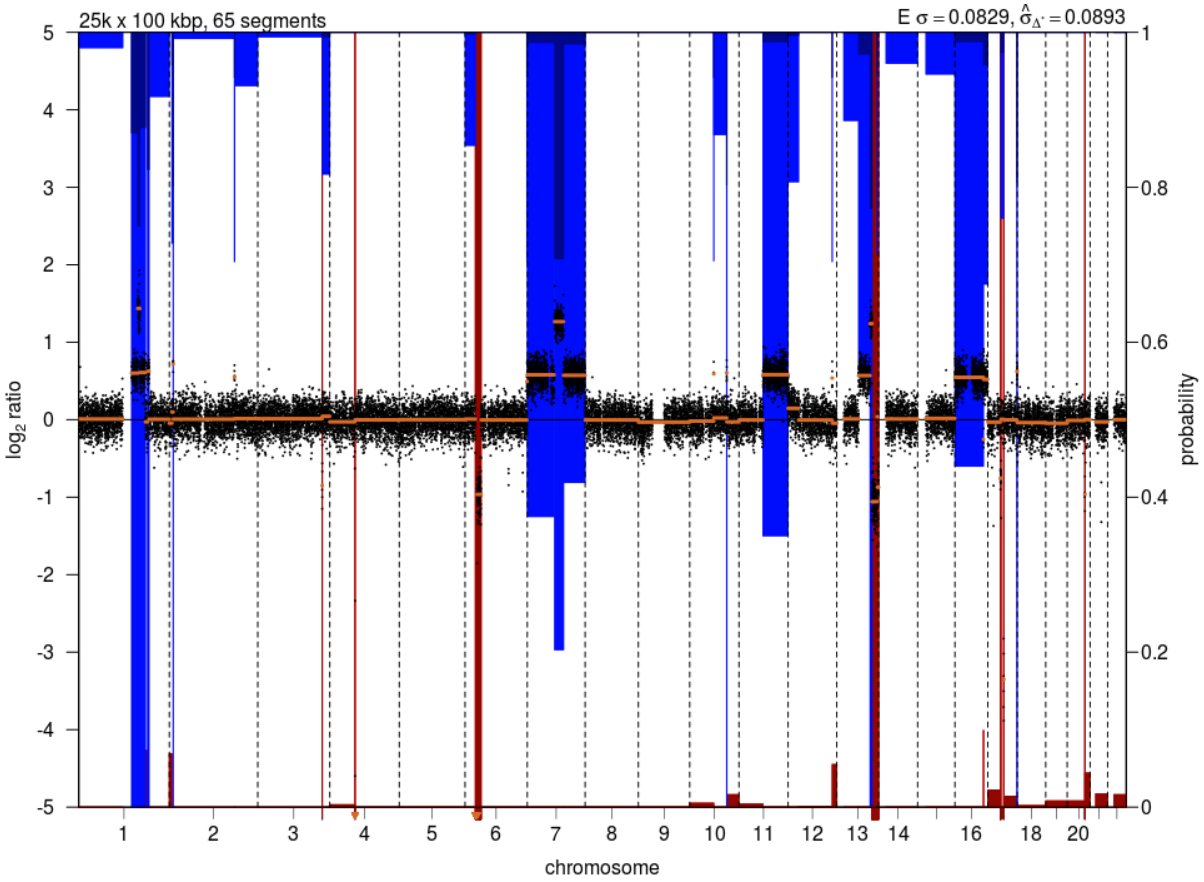


Figure 7.3.1.1. Visualisation of Copy Number Alterations from TC413 fragmented gDNA. Relative CN values at each autosomal chromosome (1.29x depth and QDNAseq default thresholds). X-axis = genomic location, primary Y-axis = log2 transformed read depth with the log2 equals zero line indicating a likely diploid genome value, secondary Y-axis = CNA call confidence value. Blue represents regions of copy number gain and wine-red shows copy number losses.

Table 7.3.1.1. Copy number alterations detected by QDNASeq in TC413 gDNA compared to fragmented gDNA and as coverage declines

				Original unfragmented TC413 gDNA	Fragmented TC413 gDNA 3.9x 10^6 reads	Fragmented TC413 gDNA 1x10^6 reads	Fragmented TC413 gDNA 0.5x10^6 reads	Fragmented TC413 gDNA 0.25x10^6 reads
Depth of sequencing				1.35x	1.29x	1.11x	1.05x	1.02x
Chr	**Start Position**	**SV Type**	**SV Length**					
1	**121,200,001***	**CN Gain**	**41.4 Mb**	**D**	**D**	**D**	**D**	**D**
1	**162,600,001***	**CN Gain**	**5.2 Mb**	**D**	**D**	**D**	**D**	**D**
1	**167,800,001***	**CN Gain**	**25.5 Mb**	**D**	**D**	**D**	**D**	**D**
2	7,200,001	CN Gain	172.5 Mb	ND	D	ND	ND	ND
3	176,900,001	CN Loss	5 Mb	ND	D	ND	ND	ND
6	400,001	CN Loss	28.1 Mb	D	D	ND	ND	ND
7	**1***	**CN Loss**	**74 Mb**	**D**	**D**	**D**	**D**	**D**
7	**75,200,001***	**CN Gain**	**24.6 Mb**	**D**	**D**	**D**	**D**	**D**
7	**100,000,001***	**CN Loss**	**59.1 Mb**	**D**	**D**	**D**	**D**	**D**
10	**101,600,001***	**CN Gain**	**0.6 Mb**	**D**	**D**	**ND**	**ND**	**ND**
11	200,001	CN Loss	134.6 Mb	D	D	D	D	ND
13	**60,200,001***	**CN Gain**	**72.3 Mb**	**D**	**D**	**ND**	**ND**	**D#**
13	**91,800,001***	**CN Gain**	**5 Mb**	**D**	**D**	**D**	**D**	
13	**96,700,001***	**CN Loss**	**18.5 Mb**	**D**	**D**	**D**	**D**	**D**
16	**100,001***	**CN Gain**	**78.3 Mb**	**D**	**D**	**D**	**D**	**D**
17	79,900,001	CN Gain	12 Mb	D	D	D	D	ND
19	300,001	CN Loss	62.9 Mb	D	D	D	D	D

D = Detected with at least 50% overlapping region, ND = not detected, Mb = mega base pair
** Changes also detected (>50% overlap) in the published consensus karyotype*
#One CN gain of 77Mb called in fragmented gDNA at 1.02x

7.3.2 Estimating the tumour fraction required to detect CNAs using synthetic fastq files and serial dilutions

Unlike the cell-line DNA, DNA samples from cancer patients contain variable proportions of tumour-derived DNA admixed with DNA from non-clonal cells. In addition, clinically significant CNAs may be present only in a minor population of tumour cells.

In-silico estimation of tumour fraction required

First, I investigated the threshold of sequencing noise, or false positive calls in non-cancer cell-free DNA. I created a 'non-cancer control karyotype' fastq file by combining fastq files from five EBV positive infectious mononucleosis cell-free DNA samples. Three deletions were called from non-cancer control cell-free DNA sWGS (Table 7.3.2.1). The deletion at Chr6:28,500,001 approximates the human HLA loci (chr6:28,477,797-33,448,354) and likely reflects the limitations of mapping to a highly diverse region of the genome. The deletion at Chr17: 43,400,001 maps to a highly repetitive region. The deletion at Chr4: 69,200,001 maps to *YTHDC1*, YTH domain containing 1, a gene involved in mRNA export from nucleus; mRNA splice site selection; and regulation of gene expression. All three regions were designated as artefacts for analysis.

Table 7.3.2.1 Artefactual copy number calls of normal control cell-free DNA

Chromosome	Start Position	SV Type	SV Length	LOG2CNT
4	69,200,001	CN Loss	700,000	-5.1
6	28,500,001	CN Loss	4,900,000	-1021.98
17	43,400,001	CN Loss	900,000	-5.11

Next, I created a series of synthetic fastq files containing reads from the normal control fastq and Burkitt cell-line fragmented gDNA reads to represent tumour fractions of 0.33, 0.25, 0.20, 0.15 and 0.10. By altering the log2 thresholds to call copy number changes in QDNASeq, gains at Chr1, Chr7, Chr13 and Chr16 could be visualised at tumour fraction of 33% and depth of 1.05x (Figure 7.3.2.1; Table 7.3.2.2). Importantly, no false positive calls were made.

At tumour fraction of 0.25 (depth 1.04x) and using the optimised log2 thresholds, the gain in Chr7 is seen and a Chr19 deletion emerges which was not detected in the CBA or the gDNA sWGS. The synthetic 'Tumour Fraction 0.10' and normal controls produce identical outputs on optimised thresholds. These results imply that high proportion of tumour-derived reads (0.33) are required to call CNAs and below that level, false positive calls emerge.

Figure 7.3.2.1. Visualisation of Copy Number Alterations detected using NS and QDNASeq analysis of synthetic 'Tumour fraction 0.33 ctDNA' fastq file

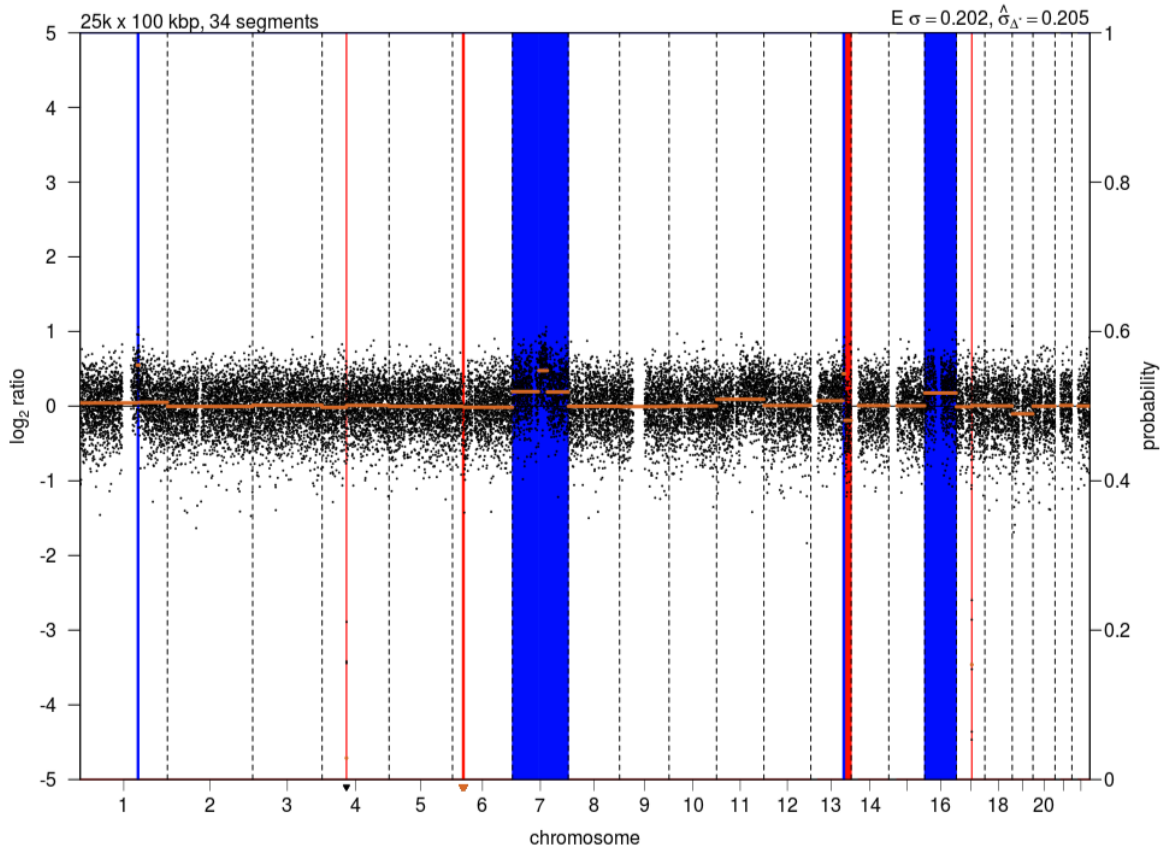


Figure 7.3.2.1: Visualisation of Copy Number Alterations detected using NS and QDNASeq from synthetic 'Tumour fraction 0.33 ctDNA' fastq file (TC413 fragmented gDNA reads admixed with control cell-free DNA reads at 1:2 ratio). Relative CN values at each autosomal chromosome (1.05x depth and QDNAseq optimised thresholds), with wine-red indicating CN loss and blue indicating CN gain. X-axis = genomic location, primary Y-axis = log2 transformed read depth with the log2 equals zero line indicating a likely diploid genome value, secondary Y-axis = CNA call confidence value.

Table: 7.3.2.2 NS and QDNASeq analysis for synthetic Tumour Fraction 0.33 ctDNA

Chromosome	Start Position	SV Type	SV Length	Log2CNT
1	163,100,001	CN Gain	4.6 Mb	0.55
7	1	CN Gain	159.1 Mb	0.43
13	91,800,001	CN Gain	4.9 Mb	0.43
13	96,700,001	CN Loss	18.5 Mb	-0.19
16	100,001	CN Gain	90.0 Mb	0.18
Artefactual Copy number changes are not included				

Wet-lab experiment to estimate tumour fraction required

Next, I compared these theoretical results to those obtained in the lab by making serial dilutions of fragmented gDNA T413 in non-cancer control cell-free DNA to represent 0.1 to 0.25 tumour fragments. Applying the optimised log2 cutoff thresholds to the 'Diluted tumour fraction 0.25 ctDNA' dilution bam file, the gains affecting Chr7. Ch11 and Chr16 could be seen (Figure 7.3.2.4). At 'Diluted tumour fraction 0.15 ctDNA', calling using the optimised thresholds detected two copy number gains. The depth was identical in at 1.02x. At 'Diluted tumour fraction 0.10 ctDNA' the QDNASeq output was indistinguishable from the non-cancer control cell-free DNA using default and optimised thresholds. By further lowering the log2 thresholds, (homozygous deletion < -0.06 < loss < -0.06 < normal < 0.06 gain < 1 < amplification) the gains at Chr7 and Chr11 could be called in the 'Diluted tumour fraction 0.10 ctDNA' dataset (Table 7.3.2.3).

The implications of both experiments (summarised in Table 7.3.2.3) are that detection of CNA by nanopore sWGS is limited primarily by the proportion of representative ctDNA in the sample. At 'tumour' fraction 10%, only the CN gains greater than 100Mb in length can be detected in the dilution experiment.

Figure 7.3.2.2: Visualisation of Copy Number Alterations detected using NS and QDNAseq analysis for the 'Diluted 0.25 tumour fraction' ctDNA

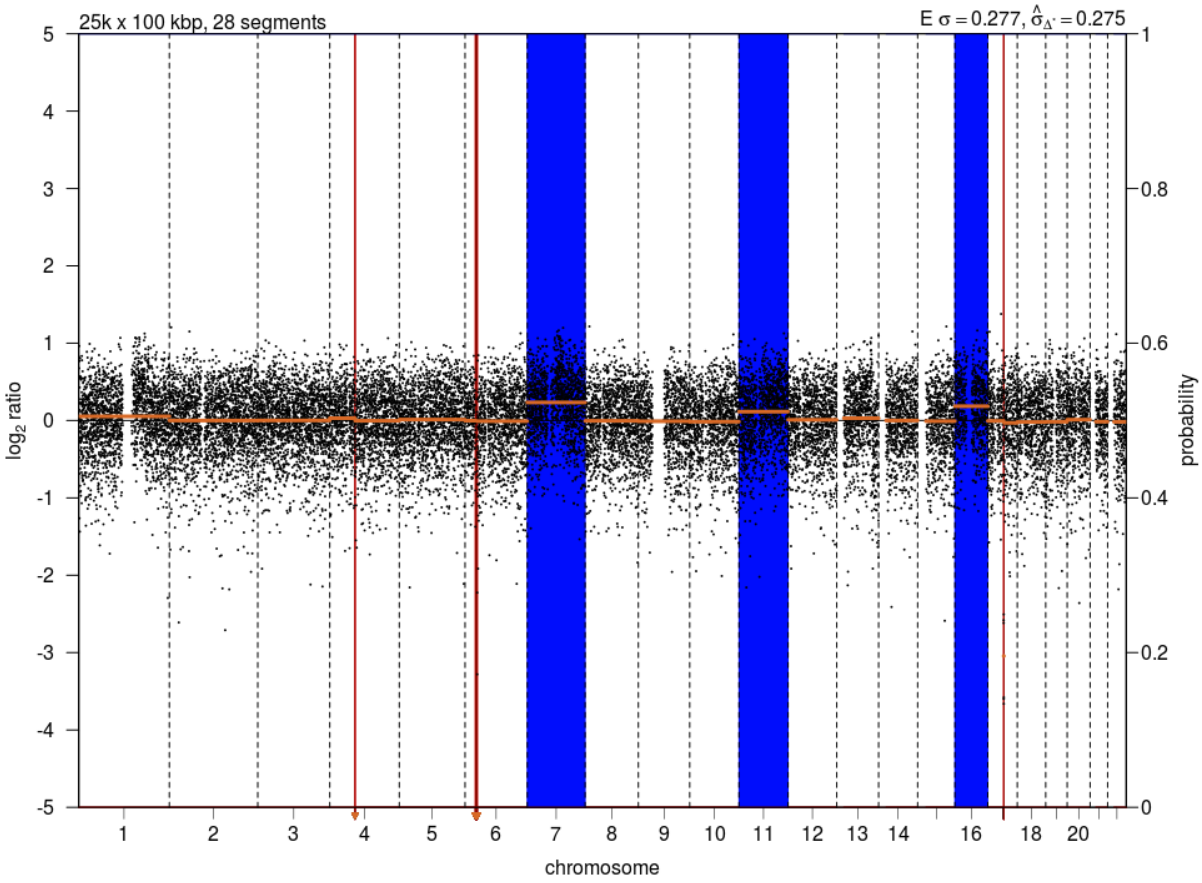


Figure 7.3.2.2: Visualisation of Copy Number Alterations detected using NS and QDNASeq from wet-lab 'Diluted tumour fraction 0.25 ctDNA' fastq file (TC413 fragmented gDNA admixed with non-cancer control cell-free DNA at 1:3 ratio). Relative CN values at each autosomal chromosome (1.02x depth and QDNAseq optimised thresholds) with wine-red indicated CN loss and blue indicating CN gain. X-axis = genomic location, primary Y-axis = log2 transformed read depth with the log2 equals zero line indicating a likely diploid genome value, secondary Y-axis = CNA call confidence value.

Table 7.3.2.3 Summary of Copy Number gains and losses detected as proportion of tumour fragments decreases

Synthetic dilution experiment (reads from fragmented gDNA fastq mixed with reads from normal cell-free DNA fastq)					
Copy Number Alteration			Synthetic 25% Optimised threshold	Synthetic 15% Optimised threshold	Synthetic 10% log2(homozygous deletion < -0.06 < loss < -0.06 < normal < 0.06 gain < 1 < amplification
Chr: Pos	SV Type	SV length (Mb)			
7:1	CN Gain	159.1;	Detected	Detected	Detected
16:100,001	CN Gain	90.0	ND	Detected	ND
19: 300,001	CN Loss	58.8	Detected	ND	Detected
Wet-lab dilution experiment (fragmented cell-line cell-free DNA in pooled normal cell-free DNA)					
			Diluted to 25% 1.02x, optimised threshold	Diluted to 15% 1.02x, optimised thresholds	Diluted to 10% 1.02x calling CNA with log2(homozygous deletion < -0.06 < loss < -0.06 < normal < 0.06 gain < 1 < amplification)
Chr: Pos	SV Type	SV length (Mb)			
7:1	CN Gain	159.1	Detected	Detected	Detected
11:200,001	CN Gain	134.8	Detected	ND	Detected
16: 100,001	CN Gain	90.0	Detected	Detected	ND

Chr:Pos, Chromosome:Start Position, SV = Structural Variant, CN = Copy number, ND = not detected

Table 7.3.2.4: Recommendations for shallow WGS of cell-free DNA using NS based on fragmented gDNA experiments

	Cell-free DNA
Input DNA	All available cell-free DNA
Library preparation method	LSK-109 gDNA, modified.
Sequencing run	Minimum 40 hours
Target number of reads	1,000,000 reads
Target depth	1.0x
Proportion of tumour DNA in the sample	Only samples with greater than 0.25 ctDNA are likely to call CNA

7.3.3 Copy number changes in the cell-free DNA samples from the RT clinical trials

Following on from the optimisation above, I proceeded to test clinical samples. Using pre-treatment plasma cell-free DNA samples previously examined by SBS (Chapter 5), I attempted to detect CNA and CK using NS of ctDNA. Plasma cell-free DNA remained for seven O-CHOP patients (CH1001, CH1005, CH1007, CH1011, CH1016, CH1017 and CH1018). Data on CNA was available from Illumina WGS on the tumour (tissue-RT) and PB-CLL DNA for all seven O-CHOP patients (Klintman et al. 2021).

The mean input cell-free DNA was 580.5 pmol for the O-CHOP cohort. Libraries were prepared using the LSK-109 NBA protocol, modified for cell-free DNA (Section 2.7.5). 14-43fmol of final library was sequenced for 48hours. The median depth for RT cell-free DNA sWGS was 1.04x (range 0.7-1.08x).

CNA were called from the sWGS data using QDNASeq with the optimised log2 thresholds. CNA with the following cut-offs: loss < -0.15 < normal < 0.14 gain < 0.58 < duplication < 1 < amplification were called. In the seven cell-free DNA samples sequenced, 49 CNA >10Mb were detected.

7.3.4 Comparison of O-CHOP cell-free DNA results with the FISH, karyotyping and previously published SBS WGS results performed as part of the clinical study

I compared CNA gains and losses > 10Mb in length, as 10Mb is the limit of detection of karyotyping by CBA. The Illumina WGS data recorded 64 CNA > 10Mb in length affecting the seven paired PB-CLL and tumour O-CHOP samples. Thirty-three CNA were detected in both PB-CLL and tumour, 24 were found in the tumour only and eleven were identified in the PB-CLL only. The median length was 20.42Mb (IQR 13.01Mb -42.49Mb). Three patients (CH1007, CH1011 and CH1017) had ≥ 5 CNA or CK. Twenty-three (23/49, 46.9%) CNAs detected in the RT cell-free DNA samples were confirmed in the tissue-RT WGS (Figure 7.3.4.1).

Figure 7.3.4.1: Relationship between CNAs detected in the three DNA sources in RT

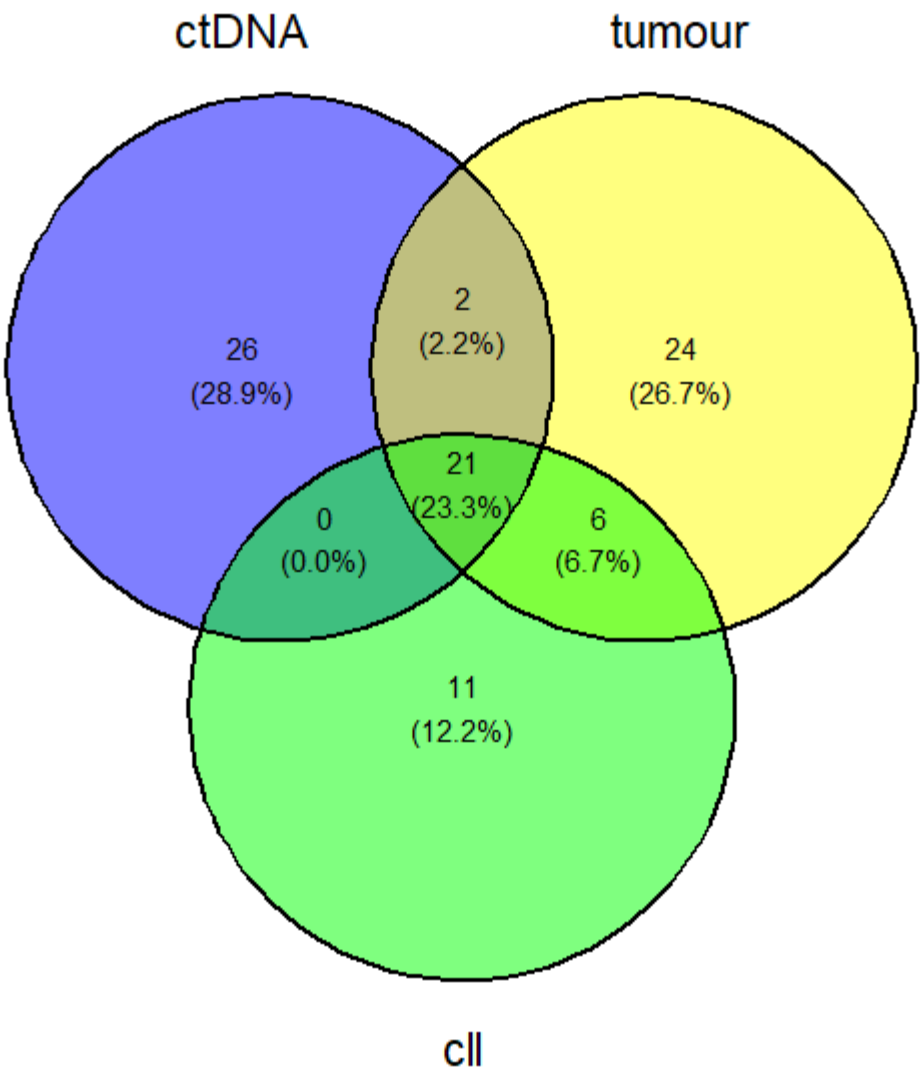


Figure 7.3.4.1: Venn diagram showing the relationship between CN gains and losses in the plasma ctDNA (blue), tumour tissue biopsy (yellow) and circulating CLL cells (green). CNAs were considered concordant if the same type (gain/loss) was detected and the region had at least 50% overlap. For visualisation, if two adjacent CNAs in the tumour tissue corresponded to one larger CNA in the ctDNA, this was counted as one shared CNA.

Additional findings in the ctDNA

Copy number changes tested by FISH (del13q, del11q, tri12 and del17p) are considered clinically relevant. Next, I ascertained whether ctDNA could detect additional clinically relevant CNAs (compared to FISH) or identify additional patients with ≥ 5 CNAs larger than 10Mb.

Shallow WGS of cell-free DNA detected four additional clinically relevant CNAs in 3/7 patients; a del13q and tri12 in CH1007 and copy number losses affecting the *TP53* locus in CH1005 and CH1018. Concordance of clinically relevant CNAs detectable by FISH and additional findings are displayed in Table 7.3.4.1.

Based on the cell-free DNA, one additional patient (CH1005) had ≥ 5 CNAs compared to the Illumina WGS results. For two patients (CH1005 and CH1007), the bone marrow karyotyping result supported the copy number profile of the cell-free DNA, showing additional changes absent from the Illumina WGS of nodal tissue and PB-CLL.

For CH1005, two CN gains at chr1:p36.33-p31.1 and chr1:q21.2-q44 were called from the Illumina WGS of the tissue biopsy only. However, sWGS of the ctDNA reported 23 CNAs > 10Mb in length including large CN gain at chr12, deletion at chr17p (*TP53* locus) and loss of the short arm of chr9 which includes the *CDKN2A* locus (Figure 7.3.4.2). Karyotyping results from CH1005 bone marrow records two clones, one 46XY and a second complex karyotype clone with duplication at chr1 and variant translocation t(1;9;17)(p36;q11;p13) affecting the *TP53* locus.

For CH1007, 13/24 ctDNA CNAs were found in the corresponding gDNA sequencing. Additional findings in the ctDNA sWGS included gains of large segments of chr12 and losses on chr13 encompassing the *DLEU2* locus. Again, karyotyping of CH1007 report a second complex clone including tri12. These results imply that sWGS of the ctDNA captured CNAs arising from the second, marrow-based clone and absent from the tissue biopsy or circulating CLL cells.

Figure 7.3.4.2: Visualisation of Copy Number Alterations detected using NS and QDNASeq analysis for CH1005 cell-free DNA

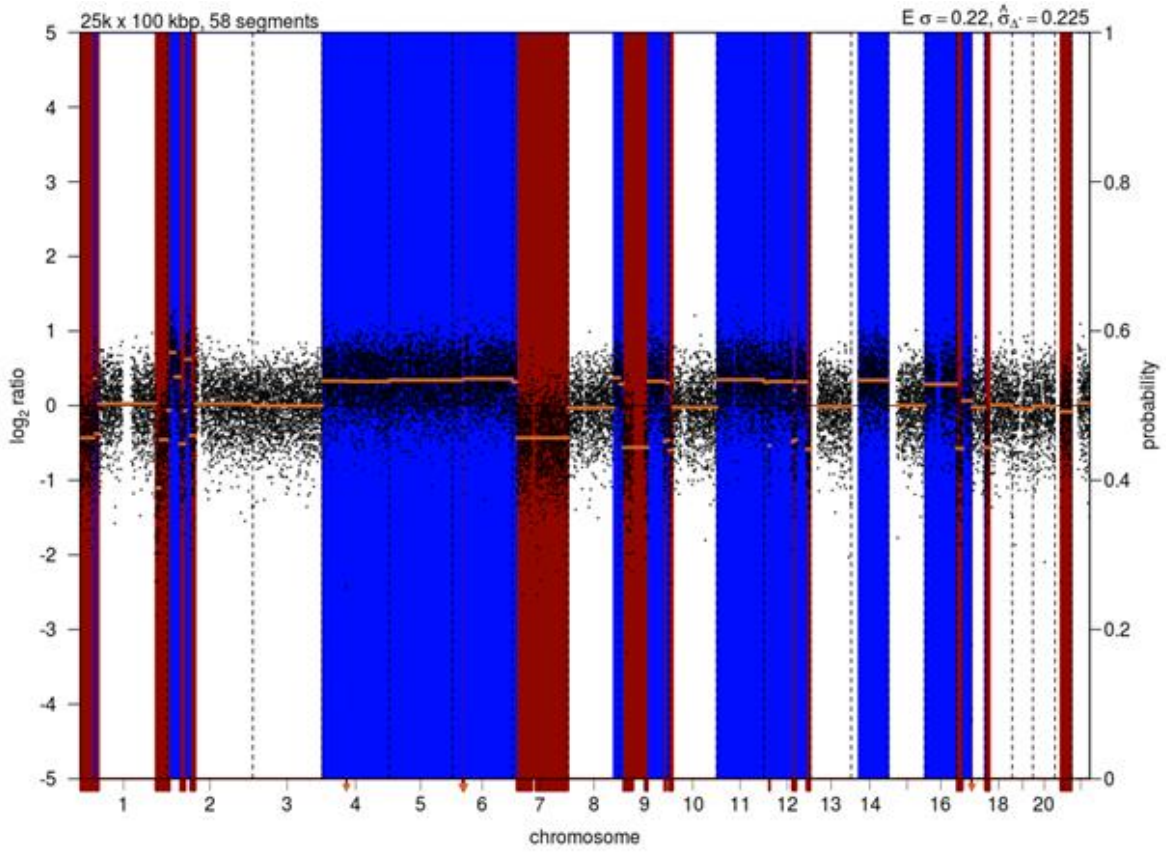


Figure 7.3.4.2: Visualisation of Copy Number Alterations detected using NS and QDNASeq from plasma cell-free DNA from CH1005. Relative CN values at each autosomal chromosome (1.03x depth and QDNAseq optimised thresholds). X-axis = genomic location, primary Y-axis = log2 transformed read depth with the log2 equals zero line indicating a likely diploid genome value, secondary Y-axis = CNA call confidence value. Blue and wine-red show the ranges of relative gains and losses respectively.

Table 7.3.4.1: Comparison of clinically reported CNA ctDNA sWGS (nanopore), PB-CLL WGS and tDNA WGS (Illumina)

	Source	del11q	tri12	del13q	del17p	Comparison between ctDNA and tumour DNA	Comment
CH1001	CtDNA sWGS	-	-	-	-	Concordant result	-
	PB-CLL WGS	-	-	-	-		
	Tumour DNA WGS	-	-	-	-		
CH1005	CtDNA sWGS	-	-	-	+ a	Discordant result	Additional del17p detected in ctDNA
	PB-CLL WGS	-	-	-	-		
	Tumour DNA WGS	-	-	-	-		
CH1007	CtDNA sWGS	+	+ b	+	+	ctDNA detected both tumour CNA	Additional tri12 and del13q detected in ctDNA
	PB-CLL WGS	-	-	-	+		
	Tumour DNA WGS	+	-	-	+		
CH1011	CtDNA sWGS	-	-	-	-	Discordant	ctDNA sequencing did not detect the del13q and del17p reported in the tumour
	PB-CLL WGS	-	-	-	+		
	Tumour DNA WGS	-	-	+	+		
CH1016	CtDNA sWGS	-	-	-	-	Discordant	ctDNA sequencing did not detect the tri12 reported in the tumour
	PB-CLL WGS	-	-	-	-		
	Tumour DNA WGS	-	+	-	-		
CH1017	CtDNA sWGS	+	-	-	-	Concordant results	-
	PB-CLL WGS	+	-	-	-		
	Tumour DNA WGS	+	-	-	-		
CH1018	CtDNA sWGS	-	+	-	+	ctDNA detected tumour CNA	Additional del17p detected in ctDNA
	PB-CLL WGS	-	-	-	-		
	Tumour DNA WGS	-	+	-	-		

sWGS = shallow Whole Genome Sequencing; PB-CLL = circulating lymphocytes at the time of RT and Illumina WGS; Tumour DNA = tumour DNA from FFPE tissue biopsy and Illumina WGS;+ = detected; - = not detected

a: Karyotyping of bone marrow from CH1005 shows complex karyotype including a translocation affecting chr17p13

b: Karyotyping of bone marrow from CH1007 shows complex karyotype including tri12

Clinically significant CNAs undetected by sWGS of ctDNA

To understand why sWGS failed to detect CNAs in two samples (del13q and del17p in CH1011 and the trisomy 12 in CH1016), I looked at the mean, median and max VAF from the targeted NGS (Chapter 5) to estimate the proportion of tumour derived fragments in the plasma cell-free DNA. No threshold for reliably detecting CNA was observed.

Two samples (CH1005 and CH1007) in which no ctDNA variants passed filtering on TGS, sWGS identified trisomy 12, and del11q and del17p respectively, illustrating an independent and complementary role for sWGS to deep targeted sequencing.

Attempting to resolve discordant findings

Additional findings in the ctDNA may represent false positive calls in the ctDNA or represent genuine intra-tumour heterogeneity. To resolve this question, Illumina SBS WGS bam files from tumour DNA were reanalysed using QDNAseq to understand if differences in the post sequencing analysis were responsible. Using this approach, a loss affecting the *TP53* locus is detected in the tumour but not PB-CLL of CH1005, suggesting the CN loss detected in the ctDNA is genuine and tumour-derived. Conversely, the tri12 and del13q are absent from the tissue-RT and PB-CLL of CH1007 using this approach.

7.4 Discussion

In this chapter, I explored the potential for the use of nanopore sWGS to detect aneuploidy, CNAs and GC from plasma cell-free DNA samples and show that sWGS can detect clinically relevant CNAs in RT. Shallow WGS of plasma cell-free DNA offers a minimally invasive platform to determining genomic features in lymphomas and solid tumours.

Sequencing plasma cell-free DNA samples from cancer patients is challenging as; cell-free DNA exists in short-fragments, cell-free DNA exists in low abundance and tumour-fragments are heavily diluted by cell-free DNA from non-tumour cells. The challenges are compounded with NS, as the technique is optimal for long DNA fragments. To assess the feasibility of using NS, therefore, I used combination of fragmented gDNA from a cell line with a known karyotype, in-silico down sampling experiments, synthetic fastq files and a dilution experiment to study the impact of each of these parameters on sequencing and the copy number calls. Based on my results, I concluded that, nanopore sWGS can successfully be performed on short-fragment gDNA and cell-free DNA generating ~ 1X depth of sequencing on one single flow cell on MinION.

In these preliminary studies, I demonstrated that the principal limitation of copy number calling is the proportion of ctDNA in the cell-free DNA sample, a problem which is common to all methods of ctDNA analysis. I therefore chose to examine pre-treatment samples from RT, the majority of whom had Stage IV disease and were anticipated to have high proportions of tumour-derived fragments in their ctDNA. In this cohort, sWGS of cell-free DNA detected additional tri12 and del(13q) in CH1007 and additional loss of *TP53* in CH1005 and CH1018. In addition, CH1005 cell-free DNA revealed > 5 large CNAs, greater than observed in PB-CLL or tissue WGS using SBS but in keeping with the bone marrow karyotyping report. In these samples, the use of cfDNA and NS to determine CNA is feasible.

Limitations of this work

My preliminary findings require further confirmation in larger cohorts with improved sample collection and storage protocols.

O-CHOP plasma samples were collected in EDTA, which is not optimal for cell-free DNA sequencing as the plasma becomes contaminated by leukocyte DNA as the interval between phlebotomy and processing increases. Despite this limitation, tumour-associated CNAs absent from PB-CLL were detectable in 3/7 cell-free DNA samples (CH1005, CH1007 and CH1018). Plasma samples from the STELLAR trial cohort (Chapter 2 and Chapter 5), collected in tubes with preservatives to minimise leakage from leukocytes have a lower cell-free DNA quantity (median 198.9HGE/mL) and provide a suitable cohort to further explore the optimum method for sWGS.

While I have modelled the minimum number of reads required, I have not established the minimum cell-free DNA input required to detect CNA. The cell-free DNA obtained from the O-CHOP cell-free DNA samples had a median 648.8 HGE/mL (IQR 221-4000HGE/mL) or 10.5ng/µl, which is higher quantities than reported in DLCBL (Bohers et al. 2018; Kurtz et al. 2018; Rossi et al. 2019), reflecting both disease burden and the effects of EDTA. Similarly

high ctDNA quantities may not be obtainable in other cancers, limiting the role of this method.

Copy number changes detected in the ctDNA alone may reflect tumour heterogeneity or represent false positives. My work was unable to fully resolve the potential false positives observed in the plasma cell-free DNA samples. As a control while the method is undergoing validation, sWGS by SBS of plasma cell-free DNA should be performed in parallel to facilitate direct comparison between the methods, including comparison of time-to-result and cost per sample. This, in combination with the available gDNA WGS datasets will allow a consensus on the validity of NS sWGS to be reached.

My work was limited by the depth of sequencing achievable using a R9.4 flow cell on a MinION sequencer. Since this work was completed, ONT launched the R10.3 flow cell and MinION $1D^2$ protocols, resulting in improved per-base Phred quality scores (corresponding to 98.6% accuracy) and improved ability to capture single nucleotide polymorphisms. The updated ONT flow cell chemistry and base-calling have been used to validate NS for HLA typing (Matern et al. 2020). For cell-free DNA sequencing, the increased per base accuracy might facilitate alignment of more reads to the reference genome and thus greater capacity to identify CNA from sWGS.

A further improvement could result from sequencing cell-free DNA on ONT high-output sequencers (GridION or PromethION) would yield additional depth of coverage, potentially permitting detection of further genomic features and detection of aneuploidy at lower tumour contents and has been explored in a range of solid tumours (Valle-Inclan et al. 2021; Katsman et al. 2022; Martignano et al. 2021).

There is potential for further optimisation of post sequencing analysis. My work used minimap2 for both gDNA and cell-free DNA alignment. While recommended for alignment of long nanopore reads, it may be suboptimal for short cell-free DNA reads. Martignano reported higher mean percentage of unique mapped cell-free DNA reads using BWA compared to minimap2 (98.5% versus 85.6%) (Martignano et al. 2021).

Alternative copy number callers for NS are available. NanoGLADIATOR is a package of python scripts developed specifically for the detection of CNA from NS (Magi et al. 2019). In contrast to QDNASeq which uses complete fastq files, NanoGLADIATOR can be applied to data generated while the nanopore sequencer is running, enabling rapid results. In a test dataset of six haematological cancers previously characterised by Agilent CGH, sWGS of gDNA and CNA using NanoGLADIATOR recorded high correlation (R =0.7). Increasing sequencing time and number of reads increased the ability to detect CNA (Magi et al. 2019). In this study, the limit of detection was predicted to be AF 0.1, but it was not formally tested. More recently, NanoGLADIATOR was tested on sWGS obtained from lung cancer cell-free DNA (Martignano et al. 2021).

ICHOR is an R-based bioinformatics package optimised detection of large CNA from ultra-low WGS of cell-free DNA (0.1x). The ichorCNA algorithm divides the genome into non-overlapping bins of 1Mb and uses probabilistic modelling for calling CNA. Providing samples contained > 10% tumour-derived fragments, concordance of > 80% for CNA detected in the

tumour gDNA could be obtained from WES datasets at genome-wide coverage of 0.1x (Adalsteinsson et al. 2017). Like NanoGLADIATOR, iCHOR can be applied without a paired normal control. My attempts to apply these bioinformatics tools to my cell-free DNA datasets were unsuccessful.

Plasma samples from the STELLAR trial cohort (Chapter 2 and Chapter 5), collected in tubes with preservatives to minimise leakage from leukocytes have a lower cell-free DNA quantity (median 198.9HGE/mL) and provide a suitable cohort to further explore the optimum method for sWGS.

Shallow WGS of plasma cell-free DNA in using nanopore

Cell-free DNA sequencing using the PromethION or GridION, ONT high-output sequencer has been examined in lung, ovarian, prostate and bladder cancers (Martignano et al. 2021; Katsman et al. 2022; van der Pol 2022; Valle-Inclan et al. 2021). In keeping with my findings, recurrent CNA were seen in normal cell-free DNA suggesting that they arise from the technique itself (Martignano et al. 2021). Further sWGS of control cell-free DNA is required to understand artefactual CNAs.

Only Martignano et al reported estimations of the limit of detection of their method. Simulated ctDNA was made by manipulating fastq reads to introduce artificial copy number changes. The simulated ctDNA reads were mixed with reads from a single healthy control at 1-50% concentrations. The artificial copy number changes were detectable when the simulated ctDNA represented > 5% of the reads. However, at the 5% threshold, three and four copies of a genomic region were indistinguishable, leading the authors to conclude that ctDNA fraction of 10% is the minimum required (Martignano et al. 2021).

Shallow WGS using Illumina SBS of cell-free DNA in RT

Few studies using Illumina SBS evaluate the role of sWGS of plasma cell-free DNA in RT. Yeh at al detected nodal (RT-tissue) associated CNAs using sWGS of plasma cell-free DNA, absent from the PB-CLL and bone marrow in two of three RT patients sequenced (Yeh, Hunter, et al. 2017).Tumour associated CNAs, absent from the PB-CLL are detectable in 6/8 (75%) plasma cell-free DNA at RT diagnosis (Parry et al. 2023b). Using Illumina SBS at 0.1x and the iCHOR bioinformatics package, whole genome duplication, chromothripsis and CN changes arising from the Richter's tumour were identified in plasma (Parry et al. 2023b).

It is notable despite the frequency of CNAs in RT tumours, not every RT plasma sequenced revealed CNAs, albeit only small numbers are available (Yeh, Hunter, et al. 2017; Parry et al. 2023b). This is consistent with my results, where no CNAs were detected in two of seven samples.

Diagnosis of RT relies on a high index of clinical suspicion and choice of the correct nodal site for biopsy. No strategies exist to facilitate early diagnosis or sequential monitoring of genomic features heralding transformation. In two patients, sWGS of plasma samples taken during treatment for CLL revealed RT CNAs 4-6 months in advance of the diagnostic tissue biopsy (Parry et al. 2023b). In combination with Yeh et al, where somatic variants of RT were

detectable in plasma prior to diagnosis, these findings imply a potential role for ctDNA in early detection of RT (Yeh, Hunter, et al. 2017).

Potential clinical applications of sWGS of plasma cell-free DNA in cancer

Aneuploidy is prognostic for disease-free and OS in breast, endometrial, renal cell clear and colorectal cancers (Ben-David and Amon 2020). Nanopore sWGS of cell-free DNA provides a rapid screening tool for a large structural gains and losses in high-grade lymphoma, myeloma or other tumours with aneuploidy. My work, based on a shallower depth of sequencing detected CNA when the ctDNA fraction > 0.25. Further optimisation to enable nanopore based detection of CNA at lower ctDNA fractions to facilitate lymphoma diagnosis is going as part of the Aggressive Infection-Related East Africa Lymphoma (AI-REAL) programme.

My preliminary work demonstrates the feasibility of using sWGS on the MinION platform to detect large chromosomal gains and deletions and merits further exploration in a larger cohort.

Chapter 8: Discussion

8.1 Overview

While targeted treatment options have expanded the therapeutic landscape in chronic lymphocytic leukaemia (CLL), similar precision approaches have yielded limited success in diffuse large B-cell lymphoma (DLBCL). In Richter's Transformation (RT), or high-grade transformation of CLL, the clinical outcomes remain dismal.

Biological heterogeneity has hampered historical efforts to select the appropriate patients with early CLL for clinical trials of early intervention. In Chapter 3, I used a capture hybridisation based targeted panel to identify variants in CLL-driver genes in two small Monoclonal B-cell Lymphocytosis (MBL) and early CLL cohorts. Variants at <0.05 allele frequency were seen only when sorted CD19+ B-cells provided the tumour DNA. The eight sequential samples showed variable changes in VAF between enrolment and follow-up samples. The number of mutations detected had no significant impact on treatment free survival (TFS) in univariate analysis. While my data does not support the hypothesis that progressors would be identifiable from the number of mutations detected by targeted sequencing, presence of mutations is tightly linked with unmutated immunoglobulin heavy chain gene rearrangement (u-IGHV).

Delivery of precision therapy for B-cell lymphoma is challenged by the heterogeneity of the disease which is poorly captured by conventional diagnostic tools. Limited ability to capture the biological heterogeneity of mature B-cell malignancies constrains the delivery of therapeutics based on an individualised molecular profile. Sequencing circulating tumour DNA (ctDNA) circumvents the problems of FFPE induced DNA damage and reflects the profile of all tumour sites. As ctDNA is minimally invasive, it can be repeated throughout treatment, expanding the monitoring of measurable residual disease (MRD) to lymphoma. I use disease-specific targeted panels to study the role of ctDNA in DLBCL (Chapter 4) and RT (Chapter 5) respectively.

Immunoglobulin heavy chain gene rearrangement (IGHV) status and *TP53* disruption are key CLL biomarkers, however barriers to universal testing for CLL patients persist. Sanger sequencing (SS) is limited by sensitivity. Outside of specialist centres, laboratories may lack the sample throughput to make next generation sequencing (NGS) cost effective. Resource constrained healthcare systems may be unable to invest in sequencing-by-synthesis (SBS) technology. In Chapter 6, I explored the use of Oxford Nanopore Technology's (ONT) MinION platform to determine IGHV status, detect *TP53* mutations and large copy number gains and losses in CLL. Although limited by the proportion of clonal B-cells in the sample, the technology offers an affordable and rapid assessment of IGHV status in CLL and detects clonal *TP53* mutations. Significantly, it can provide a result within five working days and IGHV and *TP53* can be sequenced simultaneously. Finally, I adapted the nanopore amplicon sequencing method for shallow whole genome sequencing (sWGS) using cell-free DNA. Using plasma samples from RT, copy number changes arising from nodal and marrow sites of disease were detectable in individual patients.

The key findings described in the preceding chapters lay the foundation for the genomics questions of the three NCRI multi-centre studies of DLBCL (REMoDL-A), MBL and CLL (OXPLORED) and RT (STELLAR). NS for haematological cancers will be further explored as

part of our groups NIHR RIGHT Global Health programme grant (Schuh 2019; Legason et al. 2022).

8.2 Unmutated IGVH rearrangements co-exist with variants in key genes in Monoclonal B-cell lymphocytosis and early CLL

Early identification of people with MBL and CLL who will experience disease progression is challenging. Non-progression events including serious infections lead to significant morbidity in MBL and CLL (Andersen et al. 2019) . Randomised trials of early intervention reported fewer progressors in the untreated arm than predicted by scoring systems (Hoechstetter et al. 2017; Langerbeins et al. 2022). Robust tools to identify the minority of CLL patients who will experience rapid progression are needed to enable future trials of targeted treatment for this population.

My thesis attempted two strategies to address the problem. First, I established a large observational study for early lymphoproliferative disorders, OXPLORED. This prospective trial will generate a rich clinical and genomic dataset. Data on non-progression outcomes are collected in tandem.

Secondly and to inform the sequencing strategy for the OXPLORED samples, I evaluated the potential for deep targeted sequencing for early identification of progressors in the retrospective LYM1 cohort. Variants at allele frequencies < 0.05 were absent from DNA obtained from unsorted cells, outlining the necessity of sorting CD19+ B-cells for the prospective OXPLORED study. Targeted sequencing with a 59-gene panel was uninformative in 11/17 (64.7%) of the LYM1 cohort and one of five OXPLORED samples sequenced. Fewer mutations were found in the LYM1 cohort than other studies of newly diagnosed CLL (Brieghel et al. 2020; Kleinstern et al. 2020), possibly reflecting samples chosen for indolent clinical behaviour, low proportion of clonal B-cells in the unsorted samples and strict filtering for previously reported variants. Mutations frequently co-existed with u-IGHV in my study population, in common with other publications (Brieghel et al. 2020; Kleinstern et al. 2020; Robbe et al. 2022; Knisbacher et al. 2022).

At least 58 recurrently mutated, putative driver genes have been identified in CLL (Robbe et al. 2022; Landau et al. 2015). Half of candidate driver mutations are found in <2% of CLL cases (Knisbacher et al. 2022), implying the even large panels will miss rare variants and arguing for a whole-genome approach to detect clinically significant mutations and to identify MBL/early CLL at risk of progression.

Whole genome sequencing (WGS) enables the simultaneous detection of IGHV status, mutations and copy number alterations (CNAs) as well as the ability to characterise mutational signatures and genomic complexity. Building on the Genomics England 100,000 Genome Programme and the laboratory and analytical packages established to perform WGS on almost 500 CLL clinical trial patients, the OXPLORED study has secured funding for WGS of paired germline and tumour samples.

8.3 Plasma ctDNA provides a minimally invasive tool for with potential for disease assessment in high-grade B-cell lymphoma

My thesis explored ctDNA quantification and dynamic changes during treatment in high-grade B-cell lymphoma. The ACCEPT (DLBCL) and the STELLAR (RT) cohorts received similar treatments (R-CHOP with or without acalabrutinib) and provided sequential plasma samples.

In both DLBCL and RT groups, capture hybridisation-based sequencing targeted specific genomic loci using a disease-specific panel. Baseline ctDNA burden correlated with DLBCL International Prognostic Index (IPI) in ACCEPT, in keeping with other studies of ctDNA in DLBCL (Kurtz et al. 2018; Meriranta et al. 2022). By contrast, quantification of ctDNA in RT did not correlate with other adverse biomarkers.

Assessment of ctDNA revealed additional mutations in both groups. In the four ACCEPT participants with insufficient tumour biopsy for molecular analysis at HMDS, ctDNA detected mutations affecting *BCL2, NOTCH1*, *CARD11* and *TP53. TP53* mutations were detected in ctDNA in 8/31 (25.5%) DLBCL and 13/31 (41.9%) RT patients. Five of 13 RT patients had the mutation identified in the ctDNA alone. In published work, mutations affecting *TP53* were detected in the ctDNA of 25% of high-risk DLBCL enrolled in NLG-LBC-05 Phase II trial and a similar proportion (25.4%) in the LYM1002 trial for relapsed high-grade B-cell lymphoma (Meriranta et al. 2022; Bruscaggin et al. 2021). Mutated *TP53* was reported to correlate with high baseline ctDNA burden and metabolic tumour volume by PET-CT in high-risk DLBCL (Meriranta et al. 2022). Despite the differing treatment strategies (intense chemoimmunotherapy in NLG-LBC-05 versus ibrutinib and nivolumab in LYM1002), patients with *TP53* mutations in baseline ctDNA experienced inferior clinical outcomes in both studies. No difference in PFS was observed between *TP53* mutated and wild-type DLCBL patients in the ACCEPT study. Nor do my results demonstrate inferior outcomes for high baseline ctDNA or mutated *TP53* in RT, although outcomes in the O-CHOP trial were poor in both groups.

Additional BTKi mutations were seen in the ctDNA of two STELLAR participants (STELLAR007 and STELLAR010). Other studies testing for BTK resistance mutations using plasma cell-free DNA reported contrasting results. Using qPCR on 20 paired PB-CLL and plasma cell-free DNA, a greater detection rate was seen using circulating lymphocytes (Trummer et al. 2022) while another study found plasma cell-free DNA more reliable (Albitar et al. 2017). The fully recruited STELLAR cohort (n=60) offers the opportunity to examine the impact of mutated *TP53* and the optimal strategy for detection of BTK resistance mutations.

8.4 Long-read nanopore sequencing can provide comprehensive molecular profiling

In Chapter 6, I refined methods to determine IGHV status and detect *TP53* mutations using NS. As proof of principle, the combination of IGHV and *TP53* sequencing with shallow WGS (detecting CNA) facilitated the detection of clinically significant parameters in a single sequencing run.

I focused on IGHV and *TP53* as they have the greatest impact on clinical decision making. Although raw nanopore reads have a higher error rate than short reads obtained from Illumina sequencers, the error rate is random. This enables bioinformatics tools such as canu and nanopolish to generate a consensus sequence, or contig for the region of interest.

The principal limitations of IGHV and *TP53* variant detection are the proportion of clonal DNA in the sample. My work suggests a threshold of > 10%, making it comparable to SS but inferior to SBS for detection of low allele frequency *TP53* variants. Both tests require optimisation (of the VH4 primers for *IGHV* and detection of variants at 0.05 to 0.2 allele frequency for *TP53*) before NS can be considered for development into clinical diagnostics.

8.5 Large copy number alterations may be detected in the cell-free DNA using nanopore sequencing

Shallow WGS detects CNA and GC, both of which have prognostic relevance for CLL. I adapted Oxford Nanopore Technology's amplicon sequencing protocol to perform shallow WGS of the cell-free DNA from RT patients. For some patients, the sWGS ctDNA results mirrored the PB-CLL and for the others, it reflected the LN or marrow profile. Our groups previous work shows that some cases of RT have concordant profiles across the PB-CLL and LN sites of disease, whereas the profiles are divergent in others (Klintman et al. 2021). Convincing evidence of tissue-based CNA in the cell-free DNA was seen in three out of ten RT plasma samples (CH1005, CH1018 and STELLAR004).

8.6 Limitations of this work

All work presented in the thesis is exploratory and not intended to deliver definite answers. The impact of COVID-19 on UK health services and clinical research resulted in slower than predicted recruitment for the OXPLORED and STELLAR cohorts and fewer samples for analysis. Additionally, closure of the University laboratories in 2020, followed by restricted reopening mandated a reduction in planned sequencing experiments and made experiments with shorter workflow such as nanopore more feasible.

The work presented in my thesis laid the foundations for the observational OXPLORED study (target recruitment 800 MBL/CLL). OXPLORED is not intended to impact on the participant's routine clinical care and the decision to initiate treatment remains with the participants primary haematologist. Inter-physician heterogeneity has been observed in the decision to initiate treatment. TTFT is prolonged for CLL patients cared for by CLL-specialist compared with patients whose primary physician is a general haematologist (median TTFT 9.2 vs 6.1 years, $p < 0.001$) (Shanafelt et al. 2012). While this mirrors real-world clinical care, variability between physicians may impact on the study's ability to identify progressors.

For both high-grade lymphoma cohorts, a small number of non-cancer controls and germline DNA samples were used to identify recurrent sequencing artefacts. This work preceded the consensus guideline advising testing of ctDNA panels on a minimum of 30 normal controls (Rossi et al. 2020), a recommendation with significant cost implications for smaller ctDNA projects. The ideal control would be age-matched and proven to be cancer-free (not just asymptomatic). Identification of suitable control samples is challenging, as both CHIP and small clonal B-cell populations occurs in up to 5% of adults > 60 years (Rawstron and Hillmen 2010). A small number of plasma cell-free DNA from SCAN participants (who had normal CT TAP and routine bloods) were available. For the REMoDL-A DLBCL trial ctDNA work, commercial age-matched control plasma samples have been sourced.

For the DLBCL ctDNA work, the panel design preceded the publication of Phase-Seq (Kurtz et al. 2021). The panel used in Chapter 4 lacks sufficient AID targets to enable detection of phased ctDNA variants. Nor can the panel used for ACCEPT detect *MYC* and immunoglobulin loci breakpoints. This has been updated for the REMoDL-A ctDNA work.

RT is a rare disease. Our group previously reported 37 patients, the largest uniformly treated clinical trial cohort (Eyre et al. 2016). The largest genomic study on RT to date reported on 98 patients from six large academic centres (Parry et al. 2023b). My work established the STELLAR randomised study, with a target recruitment of 60 RT patients and concurrent platform studies to investigate promising novel agents. As STELLAR is an ongoing study, I was unable to access data on the treatment arm and available outcome data was limited to the interim imaging assessment at cycle 4. More mature outcome data will be available in late 2024.

Unlike PET-CT to guide response adapted treatment, no standardised method for reporting ctDNA exists. The first international consortium attempting to harmonise ctDNA work provided recommendations on specimen collection and processing and choice of assay for MRD detection, but deferred guidance on post-sequencing analysis (Rossi et al. 2020). This work reports ctDNA burden in HGE/ml, paralleling published works using capture hybridisation in high-grade lymphoma.

No comparator method for tracking ctDNA variants was used. Sequential monitoring on the clonotypic IGH rearrangement has been used to monitor response to treatment in DLBCL (Roschewski et al. 2015). Capture hybridisation sequencing is reported to show greater sensitivity than IGH monitoring in DLBCL (Kurtz et al. 2015) but the literature in CLL is conflicting. Comparison with flow-cytometry based assessment yielded discordant results (Trummer et al. 2022; Furstenau et al. 2022). Sequential monitoring using sensitive flow-cytometry is ongoing for the STELLAR trial, generating a dataset for future comparison.

8.7 Implications of this work

This work facilitated the establishment of the OXPLORED longitudinal study for early lymphoproliferative disorders. In the LYM1 cohort examined in Chapter 3, the targeted sequencing panel revealed the limitations of a targeted approach for a cancer with a relatively low mutation burden as 64.7% of MBL/early CLL had no mutations detected. Variants at low allele frequency were not observed in the LYM1 cohort, where DNA was extracted from unsorted PBMCs. This finding, coupled with the desire to identify low frequency mutations informed the decision to sort B-cells for OXPLORED MBL and early CLL samples.

The CLL project within the 100,000 Genomes programme used WGS and non-negative matrix factorisation to derive a comprehensive multi-modal prognostic score (Robbe et al. 2022). The patient cohort comprised 485 CLL patients enrolled in UK-wide clinical trials. Just twelve patients enrolled in the CLEAR study were on active surveillance or ‘watch and wait.’ Ongoing analysis of WGS data from the OXPLORED cohort explores the role of this WGS-derived prognostic score for identification of early progressors. IGHV status and CLL-driver

mutations, in conjunction with parameters derived from WGS including mutational signature, genomic complexity and telomere length are being examined. The biological samples combined with the high-quality clinical outcome data form a robust prospective dataset to assess the WGS-derived multi-modal prognostic score.

The OXPLORED dataset offers the opportunity for prospective validation of other early CLL prognostic models such as the IPS-E (Condoluci et al. 2020) and using the prospectively collected hospitalization data, the chance to examine clinically important non-progression outcomes such as serious infection and second malignancies.

My work on ctDNA in DLBCL describes the relationship between baseline ctDNA burden and clinical features and outlines the potential role of ctDNA eradication as an early indication of treatment response in a BTKi treated population. It also demonstrates, at least for regimens containing BTKi, that persistence of ctDNA during treatment is not associated with a high-risk of relapse, a phenomenon also observed with persistent MRD by flow-cytometry or NGS in BTKi treated CLL.

The work informs the exploratory endpoints and ctDNA analysis plan of the randomised Phase 2 NCRI REMoDL-A trial for DLCBL (Davies 2020). All 453 participants will receive one cycle of R-CHOP. Participants whose tumours are molecularly typed as GCB, ABC or Unclassified will be randomised to either R-CHOP or R-CHOP plus acalabrutinib. For REMoDL-A, the cohort of cell-free DNA controls has been expanded and the ctDNA panel refined to facilitate detection of phased variants and translocations. To remain cost-effective, samples will be batched for sequencing. Delivery of real-time ctDNA results that can be incorporated into trial randomisations or used to direct treatment decisions remains a barrier to the clinical application of ctDNA analysis and is not part of REMoDL-A.

In the context of RT, pre-treatment ctDNA burden did not inform ORR and outcomes in the O-CHOP study. The detection of recognised variants in CLL-driver genes in plasma cell-free DNA in the O-CHOP cohort, including variants absent from the circulating lymphocytes hinted that ctDNA might offer an additional tool to understand RT and suggested ctDNA sequencing merited further exploration in an RT clinical trial. Sequential plasma cell-free DNA sampling was incorporated into the STELLAR clinical trial design to examine whether ctDNA sequencing would detect additional clinically relevant genomic features and whether dynamic changes in the ctDNA profile influenced clinical outcomes. High pre-treatment ctDNA burden correlates with recognised adverse prognostic features in the 11 STELLAR participants examined in this thesis. STELLAR remains open to recruitment and it is too early to answer the question of dynamic changes in ctDNA and outcomes. However, the post-sequencing analysis strategy developed for my thesis will enable the question to be addressed when the full trial cohort has been sequenced and outcome data is available. As clinical trial populations were used, clinical and outcome data are collected to a high standard.

NS for IGHV status determines germline homology and VH family in up to twelve patients simultaneously on a four-hour sequencing run. The technology facilitates detection of clinically relevant CLL biomarkers in resource constrained settings. The IWCLL has committed to expanding access to therapeutic advances in CLL globally and building local diagnostic and research capacity (Schuh 2022). Our research group has an NIHR RIGHT

programme grant to develop molecular diagnostics capacity in Tanzania and Uganda (Legason et al. 2022). My nanopore work forms the foundation for other projects on NS in haematological disorders.

In addition, shallow WGS using nanopore detects chromosomal aneuploidy in RT cell-free DNA samples. By virtue of the rapid TAT, shallow WGS could be used to identify GC non-invasively and select patients with GC for specific treatment modalities or determine duration of treatment in a clinical trial context.

Clinical outcomes remain dismal, and RT represents an unmet need for effective therapies. Molecular profiles obtained from ctDNA shows a high proportion of *TP53* mutated disease, arguing for therapies acting independently of the *TP53* pathway. Combinations of drugs targeting BCR signalling, BCL2 inhibitors and others are being tested in ongoing Phase 2 trials (NCT05388006; NCT04271956).

8.8 Conclusions

Overall, this thesis and my work in establishing the OXPLORED study has laid the foundations for comprehensive characterisation of MBL and early CLL, creating a national platform for studying early lymphoproliferative disorders. My work on ctDNA describes the potential complementary role of ctDNA burden in high-grade B-cell lymphoma and highlights limitations in the application of ctDNA analysis to inform outcomes. In both DLCBL and RT, plasma ctDNA revealed additional *TP53* mutations. In contrast to DLBCL, early eradication of ctDNA is rare in RT. My work informs the ctDNA analysis of ongoing NCRI clinical trials on DLCBL (REMoDL-A) and RT (STELLAR). My exploration of NS opens the door for rapid 'diagnostic' biomarkers and WGS in CLL and minimally invasive WGS in RT. After additional validation, the technology could be adopted to risk-stratify patients with lymphoma, solid tumours and in the context of Global Health.

References:

Abbosh, C., N. J. Birkbak, G. A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, J. Le Quesne, D. A. Moore, S. Veeriah, R. Rosenthal, T. Marafioti, E. Kirkizlar, T. B. K. Watkins, N. McGranahan, S. Ward, L. Martinson, J. Riley, F. Fraioli, M. Al Bakir, E. Gronroos, F. Zambrana, R. Endozo, W. L. Bi, F. M. Fennessy, N. Sponer, D. Johnson, J. Laycock, S. Shafi, J. Czyzewska-Khan, A. Rowan, T. Chambers, N. Matthews, S. Turajlic, C. Hiley, S. M. Lee, M. D. Forster, T. Ahmad, M. Falzon, E. Borg, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, D. Hafez, A. Naik, A. Ganguly, S. Kareht, R. Shah, L. Joseph, A. Marie Quinn, P. A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, D. Oukrif, A. U. Akarca, J. A. Hartley, H. L. Lowe, S. Lock, N. Iles, H. Bell, Y. Ngai, G. Elgar, Z. Szallasi, R. F. Schwarz, J. Herrero, A. Stewart, S. A. Quezada, K. S. Peggs, P. Van Loo, C. Dive, C. J. Lin, M. Rabinowitz, Hjwl Aerts, A. Hackshaw, J. A. Shaw, B. G. Zimmermann, T. RACERx consortium, Peace consortium, and C. Swanton. 2017. 'Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution', *Nature*, 545: 446-51.

Adalsteinsson, V. A., G. Ha, S. S. Freeman, A. D. Choudhury, D. G. Stover, H. A. Parsons, G. Gydush, S. C. Reed, D. Rotem, J. Rhoades, D. Loginov, D. Livitz, D. Rosebrock, I. Leshchiner, J. Kim, C. Stewart, M. Rosenberg, J. M. Francis, C. Z. Zhang, O. Cohen, C. Oh, H. Ding, P. Polak, M. Lloyd, S. Mahmud, K. Helvie, M. S. Merrill, R. A. Santiago, E. P. O'Connor, S. H. Jeong, R. Leeson, R. M. Barry, J. F. Kramkowski, Z. Zhang, L. Polacek, J. G. Lohr, M. Schleicher, E. Lipscomb, A. Saltzman, N. M. Oliver, L. Marini, A. G. Waks, L. C. Harshman, S. M. Tolaney, E. M. Van Allen, E. P. Winer, N. U. Lin, M. Nakabayashi, M. E. Taplin, C. M. Johannessen, L. A. Garraway, T. R. Golub, J. S. Boehm, N. Wagle, G. Getz, J. C. Love, and M. Meyerson. 2017. 'Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors', *Nat Commun*, 8: 1324.

Agarwal, R., Y. C. Chan, C. S. Tam, T. Hunter, D. Vassiliadis, C. E. Teh, R. Thijssen, P. Yeh, S. Q. Wong, S. Ftouni, E. Y. N. Lam, M. A. Anderson, C. Pott, O. Gilan, C. C. Bell, K. Knezevic, P. Blombery, K. Rayeroux, A. Zordan, J. Li, D. C. S. Huang, M. Wall, J. F. Seymour, D. H. D. Gray, A. W. Roberts, M. A. Dawson, and S. J. Dawson. 2019. 'Dynamic molecular monitoring reveals that SWI-SNF mutations mediate resistance to ibrutinib plus venetoclax in mantle cell lymphoma', *Nat Med*, 25: 119-29.

Agathangelidis, A., A. Chatzidimitriou, T. Chatzikonstantinou, C. Tresoldi, Z. Davis, V. Giudicelli, S. Kossida, C. Belessi, R. Rosenquist, P. Ghia, A. W. Langerak, F. Davi, K. Stamatopoulos, and the European Research Initiative on C. L. L. Eric. 2022. 'Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: the 2022 update of the recommendations by ERIC, the European Research Initiative on CLL', *Leukemia*, 36: 1961-68.

Agathangelidis, A., A. Chatzidimitriou, K. Gemenetzi, V. Giudicelli, M. Karypidou, K. Plevova, Z. Davis, X. J. Yan, S. Jeromin, C. Schneider, L. B. Pedersen, R. C. Tschumper, L. A. Sutton, P. Baliakas, L. Scarfo, E. J. van Gastel, M. Armand, E. Tausch, B. Biderman, C. Baer, D. Bagnara, A. Navarro, A. Langlois de Septenville, V. Guido, G. Mitterbauer-Hohendanner, A. Dimovski, C. Brieghel, S. Lawless, M. Meggendorfer, K. Brazdilova, M. Ritgen, M. Facco, C. Tresoldi, A. Visentin, A.

Patriarca, M. Catherwood, L. Bonello, A. Sudarikov, K. Vanura, M. Roumelioti, H. Skuhrova Francova, T. Moysiadis, S. Veronese, K. Giannopoulos, L. Mansouri, T. Karan-Djurasevic, R. Sandaltzopoulos, C. Bodor, F. Fais, A. P. Kater, I. Panovska, D. Rossi, S. Alshemmari, P. Panagiotidis, P. Costeas, B. Espinet, D. Antic, L. Foroni, M. Montillo, L. Trentin, N. Stavroyianni, G. Gaidano, P. Francia di Celle, C. Niemann, E. Campo, A. Anagnostopoulos, C. Pott, K. Fischer, M. Hallek, D. Oscier, S. Stilgenbauer, C. Haferlach, D. Jelinek, N. Chiorazzi, S. Pospisilova, M. P. Lefranc, S. Kossida, A. W. Langerak, C. Belessi, F. Davi, R. Rosenquist, P. Ghia, and K. Stamatopoulos. 2021. 'Higher-order connections between stereotyped subsets: implications for improved patient classification in CLL', *Blood*, 137: 1365-76.

Agathangelidis, A., N. Darzentas, A. Hadzidimitriou, X. Brochet, F. Murray, X. J. Yan, Z. Davis, E. J. van Gastel-Mol, C. Tresoldi, C. C. Chu, N. Cahill, V. Giudicelli, B. Tichy, L. B. Pedersen, L. Foroni, L. Bonello, A. Janus, K. Smedby, A. Anagnostopoulos, H. Merle-Beral, N. Laoutaris, G. Juliusson, P. F. di Celle, S. Pospisilova, J. Jurlander, C. Geisler, A. Tsaftaris, M. P. Lefranc, A. W. Langerak, D. G. Oscier, N. Chiorazzi, C. Belessi, F. Davi, R. Rosenquist, P. Ghia, and K. Stamatopoulos. 2012. 'Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies', *Blood*, 119: 4467-75.

Agathangelidis, A., V. Ljungstrom, L. Scarfo, C. Fazi, M. Gounari, T. Pandzic, L. A. Sutton, K. Stamatopoulos, G. Tonon, R. Rosenquist, and P. Ghia. 2018. 'Highly similar genomic landscapes in monoclonal B-cell lymphocytosis and ultra-stable chronic lymphocytic leukemia with low frequency of driver mutations', *Haematologica*, 103: 865-73.

Agathangelidis, A., L. A. Sutton, A. Hadzidimitriou, C. Tresoldi, A. W. Langerak, C. Belessi, F. Davi, R. Rosenquist, K. Stamatopoulos, and P. Ghia. 2018. 'Immunoglobulin Gene Sequence Analysis In Chronic Lymphocytic Leukemia: From Patient Material To Sequence Interpretation', *J Vis Exp*.

Aggarwal, C., C. D. Rolfo, G. R. Oxnard, J. E. Gray, L. M. Sholl, and D. R. Gandara. 2021. 'Strategies for the successful implementation of plasma-based NSCLC genotyping in clinical practice', *Nat Rev Clin Oncol*, 18: 56-62.

Agius, R., C. Brieghel, M. A. Andersen, A. T. Pearson, B. Ledergerber, A. Cozzi-Lepri, Y. Louzoun, C. L. Andersen, J. Bergstedt, J. H. von Stemann, M. Jorgensen, M. E. Tang, M. Fontes, J. Bahlo, C. D. Herling, M. Hallek, J. Lundgren, C. R. MacPherson, J. Larsen, and C. U. Niemann. 2020. 'Machine learning can identify newly diagnosed patients with CLL at high risk of infection', *Nat Commun*, 11: 363.

Ahn, I. E., M. Z. H. Farooqui, X. Tian, J. Valdez, C. Sun, S. Soto, J. Lotter, S. Housel, M. Stetler-Stevenson, C. M. Yuan, I. Maric, K. R. Calvo, P. Nierman, T. E. Hughes, N. S. Saba, G. E. Marti, S. Pittaluga, S. E. M. Herman, C. U. Niemann, L. B. Pedersen, C. H. Geisler, R. Childs, G. Aue, and A. Wiestner. 2018. 'Depth and durability of response to ibrutinib in CLL: 5-year follow-up of a phase 2 study', *Blood*, 131: 2357-66.

Alaggio, R., C. Amador, I. Anagnostopoulos, A. D. Attygalle, I. B. O. Araujo, E. Berti, G. Bhagat, A. M. Borges, D. Boyer, M. Calaminici, A. Chadburn, J. K. C. Chan, W. Cheuk, W. J. Chng, J. K. Choi, S. S. Chuang, S. E. Coupland, M. Czader, S. S. Dave, D. de Jong, M. Q. Du, K. S. Elenitoba-Johnson, J. Ferry, J. Geyer, D. Gratzinger, J. Guitart, S. Gujral, M. Harris, C. J. Harrison, S. Hartmann, A. Hochhaus, P. M. Jansen, K. Karube, W. Kempf, J. Khoury, H. Kimura, W. Klapper, A. E. Kovach, S. Kumar, A. J. Lazar, S. Lazzi, L. Leoncini, N. Leung, V. Leventaki, X. Q. Li, M. S. Lim, W. P. Liu, A. Louissaint, Jr., A. Marcogliese, L. J. Medeiros, M. Michal, R. N. Miranda, C.

Mitteldorf, S. Montes-Moreno, W. Morice, V. Nardi, K. N. Naresh, Y. Natkunam, S. B. Ng, I. Oschlies, G. Ott, M. Parrens, M. Pulitzer, S. V. Rajkumar, A. C. Rawstron, K. Rech, A. Rosenwald, J. Said, C. Sarkozy, S. Sayed, C. Saygin, A. Schuh, W. Sewell, R. Siebert, A. R. Sohani, R. Tooze, A. Traverse-Glehen, F. Vega, B. Vergier, A. D. Wechalekar, B. Wood, L. Xerri, and W. Xiao. 2022. 'The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms', *Leukemia*, 36: 1720-48.

Albitar, A., W. Ma, I. DeDios, J. Estella, I. Ahn, M. Farooqui, A. Wiestner, and M. Albitar. 2017. 'Using high-sensitivity sequencing for the detection of mutations in BTK and PLCgamma2 genes in cellular and cell-free DNA and correlation with progression in patients treated with BTK inhibitors', *Oncotarget*, 8: 17936-44.

Alborelli, I., D. Generali, P. Jermann, M. R. Cappelletti, G. Ferrero, B. Scaggiante, M. Bortul, F. Zanconati, S. Nicolet, J. Haegele, L. Bubendorf, N. Aceto, M. Scaltriti, G. Mucci, L. Quagliata, and G. Novelli. 2019. 'Cell-free DNA analysis in healthy individuals by next-generation sequencing: a proof of concept and technical validation study', *Cell Death Dis*, 10: 534.

Alig, S., C. W. Macaulay, D. M. Kurtz, U. Duhrsen, A. Huttmann, C. Schmitz, M. C. Jin, B. J. Sworder, A. Garofalo, M. Shahrokh Esfahani, B. Y. Nabet, J. Soo, F. Scherer, A. F. M. Craig, O. Casasnovas, J. R. Westin, G. Gaidano, D. Rossi, M. Roschewski, W. H. Wilson, M. Meignan, M. Diehn, and A. A. Alizadeh. 2021. 'Short Diagnosis-to-Treatment Interval Is Associated With Higher Circulating Tumor DNA Levels in Diffuse Large B-Cell Lymphoma', *J Clin Oncol*, 39: 2605-16.

Alizadeh, A. A., M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, Jr., L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt. 2000. 'Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling', *Nature*, 403: 503-11.

Amini, R. M., V. Ljungstrom, M. Abdulla, L. Cavelier, T. Pandzic, P. Hollander, G. Enblad, and P. Baliakas. 2020. 'Clonal hematopoiesis in patients with high-grade B-cell lymphoma is associated with inferior outcome', *Am J Hematol*, 95: E287-E89.

Andersen, M. A., C. E. Moser, J. Lundgren, and C. U. Niemann. 2019. 'Epidemiology of bloodstream infections in patients with chronic lymphocytic leukemia: a longitudinal nation-wide cohort study', *Leukemia*, 33: 662-70.

Appleby, N., T. A. Eyre, M. Cabes, A. Jackson, R. Boucher, F. Yates, S. Fox, A. Rawstron, P. Hillmen, and A. Schuh. 2019. 'The STELLAR trial protocol: a prospective multicentre trial for Richter's syndrome consisting of a randomised trial investigation CHOP-R with or without acalabrutinib for newly diagnosed RS and a single-arm platform study for evaluation of novel agents in relapsed disease', *BMC Cancer*, 19: 471.

Araf, S., J. Wang, K. Korfi, C. Pangault, E. Kotsiou, A. Rio-Machin, T. Rahim, J. Heward, A. Clear, S. Iqbal, J. K. Davies, P. Johnson, M. Calaminici, S. Montoto, R. Auer, C. Chelala, J. G. Gribben, T. A. Graham, T. Fest, J. Fitzgibbon, and J. Okosun. 2018. 'Genomic profiling reveals spatial intra-tumor heterogeneity in follicular lymphoma', *Leukemia*, 32: 1258-63.

Armand, P., Y. Oki, D. S. Neuberg, M. Faham, C. Cummings, M. Klinger, L. Weng, S. Bhattar, A. S. Lacasce, E. D. Jacobsen, M. S. Davids, C. Jacobson, D. C. Fisher, J. R. Brown, N. H. Fowler, M. Alma Rodriguez, M. J. Wallace, S. S. Neelapu, S. Rodig, A. Younes, and A. S. Freedman. 2013. 'Detection of circulating tumour DNA in patients with aggressive B-cell non-Hodgkin lymphoma', *Br J Haematol*, 163: 123-6.

Au, C. H., D. N. Ho, B. B. K. Ip, T. S. K. Wan, M. H. L. Ng, E. K. W. Chiu, T. L. Chan, and E. S. K. Ma. 2019. 'Rapid detection of chromosomal translocation and precise breakpoint characterization in acute myeloid leukemia by nanopore long-read sequencing', *Cancer Genet*, 239: 22-25.

Baliakas, P., B. Espinet, C. Mellink, M. Jarosova, A. Athanasiadou, P. Ghia, A. P. Kater, D. Oscier, C. Haferlach, and K. Stamatopoulos. 2022. 'Cytogenetics in Chronic Lymphocytic Leukemia: ERIC Perspectives and Recommendations', *Hemasphere*, 6: e707.

Baliakas, P., S. Jeromin, M. Iskas, A. Puiggros, K. Plevova, F. Nguyen-Khac, Z. Davis, G. M. Rigolin, A. Visentin, A. Xochelli, J. Delgado, F. Baran-Marszak, E. Stalika, P. Abrisqueta, K. Durechova, G. Papaioannou, V. Eclache, M. Dimou, T. Iliakis, R. Collado, M. Doubek, M. J. Calasanz, N. Ruiz-Xiville, C. Moreno, M. Jarosova, A. C. Leeksma, P. Panayiotidis, H. Podgornik, F. Cymbalista, A. Anagnostopoulos, L. Trentin, N. Stavroyianni, F. Davi, P. Ghia, A. P. Kater, A. Cuneo, S. Pospisilova, B. Espinet, A. Athanasiadou, D. Oscier, C. Haferlach, K. Stamatopoulos, and the European Research Initiative on C. L. L. Eric. 2019. 'Cytogenetic complexity in chronic lymphocytic leukemia: definitions, associations, and clinical impact', *Blood*, 133: 1205-16.

Barrans, S., S. Crouch, A. Smith, K. Turner, R. Owen, R. Patmore, E. Roman, and A. Jack. 2010. 'Rearrangement of MYC is associated with poor prognosis in patients with diffuse large B-cell lymphoma treated in the era of rituximab', *J Clin Oncol*, 28: 3360-5.

Barrio, S., T. D. Shanafelt, J. Ojha, K. G. Chaffee, C. Secreto, K. M. Kortum, S. Pathangey, D. L. Van-Dyke, S. L. Slager, R. Fonseca, N. E. Kay, and E. Braggio. 2017. 'Genomic characterization of high-count MBL cases indicates that early detection of driver mutations and subclonal expansion are predictors of adverse clinical outcome', *Leukemia*, 31: 170-76.

Beekman, R., V. Chapaprieta, N. Russinol, R. Vilarrasa-Blasi, N. Verdaguer-Dot, J. H. A. Martens, M. Duran-Ferrer, M. Kulis, F. Serra, B. M. Javierre, S. W. Wingett, G. Clot, A. C. Queiros, G. Castellano, J. Blanc, M. Gut, A. Merkel, S. Heath, A. Vlasova, S. Ullrich, E. Palumbo, A. Enjuanes, D. Martin-Garcia, S. Bea, M. Pinyol, M. Aymerich, R. Royo, M. Puiggros, D. Torrents, A. Datta, E. Lowy, M. Kostadima, M. Roller, L. Clarke, P. Flicek, X. Agirre, F. Prosper, T. Baumann, J. Delgado, A. Lopez-Guillermo, P. Fraser, M. L. Yaspo, R. Guigo, R. Siebert, M. A. Marti-Renom, X. S. Puente, C. Lopez-Otin, I. Gut, H. G. Stunnenberg, E. Campo, and J. I. Martin-Subero. 2018. 'The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia', *Nat Med*, 24: 868-80.

Ben-Dali, Y., M. H. Hleuhel, C. da Cunha-Bang, C. Brieghel, C. B. Poulsen, E. Clasen-Linde, H. H. N. Bentzen, H. Frederiksen, I. Christiansen, L. H. Nielsen, L. Enggaard, M. Helleberg, M. Clausen, M. Frederiksen, R. S. Pedersen, C. U. Niemann, and M. A. Andersen. 2020. 'Richter's transformation in patients with chronic lymphocytic leukaemia: a Nationwide Epidemiological Study', *Leuk Lymphoma*, 61: 1435-44.

Ben-David, U., and A. Amon. 2020. 'Context is everything: aneuploidy in cancer', *Nat Rev Genet*, 21: 44-62.

Besson, C., A. Moore, W. Wu, C. M. Vajdic, S. de Sanjose, N. J. Camp, K. E. Smedby, T. D. Shanafelt, L. M. Morton, J. D. Brewer, L. Zablotska, E. A. Engels, J. R. Cerhan, S. L. Slager, J. Han, S. I. Berndt, and manuscript InterLymph Consortium. Full authors list is given at the end of the. 2021. 'Common genetic polymorphisms contribute to the association between chronic lymphocytic leukaemia and non-melanoma skin cancer', *Int J Epidemiol*, 50: 1325-34.

Binet, J. L., A. Auquier, G. Dighiero, C. Chastang, H. Piguet, J. Goasguen, G. Vaugier, G. Potron, P. Colona, F. Oberling, M. Thomas, G. Tchernia, C. Jacquillat, P. Boivin, C. Lesty, M. T. Duault, M. Monconduit, S. Belabbes, and F. Gremy. 1981. 'A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis', *Cancer*, 48: 198-206.

Black, G. S., X. Huang, Y. Qiao, S. Tarapcsak, K. A. Rogers, S. Misra, J. C. Byrd, G. T. Marth, D. M. Stephens, and J. A. Woyach. 2022. 'Subclonal evolution of CLL driver mutations is associated with relapse in ibrutinib- and acalabrutinib-treated patients', *Blood*, 140: 401-05.

Bodor, C., L. Kotmayer, T. Laszlo, F. Takacs, G. Barna, R. Kiss, E. Sebestyen, T. Nagy, L. L. Hegyi, G. Mikala, S. Fekete, P. Farkas, A. Balogh, T. Masszi, J. Demeter, J. Weisinger, H. Alizadeh, B. Kajtar, Z. Kohl, R. Szasz, L. Gergely, T. Gurbity Palfi, A. Sulak, B. Kollar, M. Egyed, M. Plander, L. Rejto, L. Szerafin, P. Ilonczai, P. Tamaska, P. Pettendi, D. Levai, T. Schneider, A. Sebestyen, P. Csermely, A. Matolcsy, Z. Matrai, and D. Alpar. 2021. 'Screening and monitoring of the BTK(C481S) mutation in a real-world cohort of patients with relapsed/refractory chronic lymphocytic leukaemia during ibrutinib therapy', *Br J Haematol*, 194: 355-64.

Bohers, E., P. J. Viailly, S. Becker, V. Marchand, P. Ruminy, C. Maingonnat, P. Bertrand, P. Etancelin, J. M. Picquenot, V. Camus, A. L. Menard, E. Lemasle, N. Contentin, S. Lepretre, P. Lenain, A. Stamatoullas, H. Lanic, J. Libraire, S. Vaudaux, L. F. Pepin, P. Vera, H. Tilly, and F. Jardin. 2018. 'Non-invasive monitoring of diffuse large B-cell lymphoma by cell-free DNA high-throughput targeted sequencing: analysis of a prospective cohort', *Blood Cancer J*, 8: 74.

Bosch, F., B. Navarro, M. Crespo, M. Alcoceba, J. B. Sanchez, B. Tazon, A. Serrano, M. G. Alvarez, L. G. Serrano, P. Alonso-Torres, M. Villanueva, C. Loriente, P. Abrisqueta, M. Peiro, J. A. Garcia-Marco, M. Gonzalez, and M. J. Terol. 2021. 'Building a network of TP53 and IGHV testing reference centers across Spain: the Red53 initiative', *Ann Hematol*, 100: 825-30.

Bouzidi, A., K. Labreche, M. Baron, M. Veyri, J. A. Denis, M. Touat, M. Sanson, F. Davi, E. Guillerm, S. Jouannet, F. Charlotte, F. Bielle, S. Choquet, P. Y. Boelle, J. Cadranel, V. Leblond, B. Autran, J. M. Lacorte, J. P. Spano, F. Coulet, and Ideation study group. 2021. 'Low-Coverage Whole Genome Sequencing of Cell-Free DNA From Immunosuppressed Cancer Patients Enables Tumor Fraction Determination and Reveals Relevant Copy Number Alterations', *Front Cell Dev Biol*, 9: 661272.

Brewer, J. D., T. D. Shanafelt, C. C. Otley, R. K. Roenigk, J. R. Cerhan, N. E. Kay, A. L. Weaver, and T. G. Call. 2012. 'Chronic lymphocytic leukemia is associated with decreased survival of patients with malignant melanoma and Merkel cell carcinoma in a SEER population-based study', *J Clin Oncol*, 30: 843-9.

Brieghel, C., C. da Cunha-Bang, C. W. Yde, A. Y. Schmidt, S. Kinalis, F. Nadeu, M. A. Andersen, L. O. Jacobsen, M. K. Andersen, L. B. Pedersen, J. Delgado, T. Baumann, M. Mattsson, L. Mansouri, R. Rosenquist, E. Campo, F. C. Nielsen, and C. U. Niemann. 2020. 'The Number of Signaling Pathways Altered by Driver Mutations in Chronic Lymphocytic Leukemia Impacts Disease Outcome', *Clin Cancer Res*, 26: 1507-15.

Brieghel, C., V. Galle, R. Agius, C. da Cunha-Bang, M. A. Andersen, P. Vlummens, M. Mattsson, R. Rosenquist, K. E. Smedby, C. D. Herling, J. Bahlo, M. Hallek, J. D. Lundgren, F. Offner, and C. U. Niemann. 2022. 'Identifying patients with chronic lymphocytic leukemia without need of treatment: End of endless watch and wait?', *Eur J Haematol*, 108: 369-78.

Brieghel, C., S. Kinalis, C. W. Yde, A. Y. Schmidt, L. Jonson, M. A. Andersen, C. da Cunha-Bang, L. B. Pedersen, C. H. Geisler, F. C. Nielsen, and C. U. Niemann. 2019. 'Deep targeted sequencing of TP53 in chronic lymphocytic leukemia: clinical impact at diagnosis and at time of treatment', *Haematologica*, 104: 789-96.

Brochet, X., M. P. Lefranc, and V. Giudicelli. 2008. 'IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis', *Nucleic Acids Res*, 36: W503-8.

Bruscaggin, A., L. T. di Bergamo, V. Spina, B. Hodkinson, G. Forestieri, F. Bonfiglio, A. Condoluci, W. Wu, M. C. Pirosa, M. R. Faderl, R. Koch, M. Schaffer, J. D. Alvarez, N. Fourneau, B. Gerber, G. Stussi, E. Zucca, S. Balasubramanian, and D. Rossi. 2021. 'Circulating tumor DNA for comprehensive noninvasive monitoring of lymphoma treated with ibrutinib plus nivolumab', *Blood Adv*, 5: 4674-85.

Burger, J. A., and J. G. Gribben. 2014. 'The microenvironment in chronic lymphocytic leukemia (CLL) and other B cell malignancies: insight into disease biology and new targeted therapies', *Semin Cancer Biol*, 24: 71-81.

Burns, A., R. Alsolami, J. Becq, B. Stamatopoulos, A. Timbs, D. Bruce, P. Robbe, D. Vavoulis, R. Clifford, M. Cabes, H. Dreau, J. Taylor, S. J. L. Knight, R. Mansson, D. Bentley, R. Beekman, J. I. Martin-Subero, E. Campo, R. S. Houlston, K. E. Ridout, and A. Schuh. 2018. 'Whole-genome sequencing of chronic lymphocytic leukaemia reveals distinct differences in the mutational landscape between IgHV(mut) and IgHV(unmut) subgroups', *Leukemia*, 32: 332-42.

Burns, A., Bruce D., Robbe, P., Timbs, A., Stamatopoulos, B., Clifford R., Lopopolo, M., Parkes., D., Ridout, K., Schuh, A. 2018. 'Detection of Clinically Relevant Molecular Alterations in Chronic Lymphocytic Leukemia (CLL) By Nanopore Sequencing', *Blood*, 132.

Burns, Adam, Daniella DiSalvo-Williams, David Bruce, Pauline Robbe, Adele Timbs, Basile Stamatopoulos, Ruth Clifford, Maria Lopopolo, Duncan Parkes, Kate Ridout, and Anna Schuh. 2019. 'The diagnostic chronic lymphocytic leukaemia genome by nanopore sequencing', *bioRxiv*: 750059.

Byrd, J. C., B. Harrington, S. O'Brien, J. A. Jones, A. Schuh, S. Devereux, J. Chaves, W. G. Wierda, F. T. Awan, J. R. Brown, P. Hillmen, D. M. Stephens, P. Ghia, J. C. Barrientos, J. M. Pagel, J. Woyach, D. Johnson, J. Huang, X. Wang, A. Kaptein, B. J. Lannutti, T. Covey, M. Fardis, J. McGreivy, A. Hamdy, W. Rothbaum, R. Izumi, T. G. Diacovo, A. J. Johnson, and R. R. Furman. 2016. 'Acalabrutinib (ACP-196) in Relapsed Chronic Lymphocytic Leukemia', *N Engl J Med*, 374: 323-32.

Byrd, J. C., J. A. Woyach, R. R. Furman, P. Martin, S. O'Brien, J. R. Brown, D. M. Stephens, J. C. Barrientos, S. Devereux, P. Hillmen, J. M. Pagel, A. Hamdy, R. Izumi, P. Patel, M. H. Wang, N. Jain, and W. G. Wierda. 2021. 'Acalabrutinib in treatment-naive chronic lymphocytic leukemia', *Blood*, 137: 3327-38.

Bystry, V., A. Agathangelidis, V. Bikos, L. A. Sutton, P. Baliakas, A. Hadzidimitriou, K. Stamatopoulos, N. Darzentas, and C. L. L. European Research Initiative on. 2015. 'ARResT/AssignSubsets: a novel application for robust subclassification of chronic lymphocytic leukemia based on B cell receptor IG stereotypy', *Bioinformatics*, 31: 3844-6.

Campbell, M. J., A. D. Zelenetz, S. Levy, and R. Levy. 1992. 'Use of family specific leader region primers for PCR amplification of the human heavy chain variable region gene repertoire', *Mol Immunol*, 29: 193-203.

Camus, V., M. Viennot, J. Lequesne, P. J. Viailly, E. Bohers, L. Bessi, B. Marcq, P. Etancelin, S. Dubois, J. M. Picquenot, E. L. Veresezan, M. Cornic, L. Burel, J. Loret, S. Becker, P. Decazes, P. Lenain, S. Lepretre, E. Lemasle, H. Lanic, A. L. Menard, N.

Contentin, H. Tilly, A. Stamatoullas, and F. Jardin. 2021. 'Targeted genotyping of circulating tumor DNA for classical Hodgkin lymphoma monitoring: a prospective study', *Haematologica*, 106: 154-62.

Chabon, J. J., E. G. Hamilton, D. M. Kurtz, M. S. Esfahani, E. J. Moding, H. Stehr, J. Schroers-Martin, B. Y. Nabet, B. Chen, A. A. Chaudhuri, C. L. Liu, A. B. Hui, M. C. Jin, T. D. Azad, D. Almanza, Y. J. Jeon, M. C. Nesselbush, L. Co Ting Keh, R. F. Bonilla, C. H. Yoo, R. B. Ko, E. L. Chen, D. J. Merriott, P. P. Massion, A. S. Mansfield, J. Jen, H. Z. Ren, S. H. Lin, C. L. Costantino, R. Burr, R. Tibshirani, S. S. Gambhir, G. J. Berry, K. C. Jensen, R. B. West, J. W. Neal, H. A. Wakelee, B. W. Loo, Jr., C. A. Kunder, A. N. Leung, N. S. Lui, M. F. Berry, J. B. Shrager, V. S. Nair, D. A. Haber, L. V. Sequist, A. A. Alizadeh, and M. Diehn. 2020. 'Integrating genomic features for non-invasive early lung cancer detection', *Nature*, 580: 245-51.

Chaganti, S., T. Illidge, S. Barrington, P. McKay, K. Linton, K. Cwynarski, A. McMillan, A. Davies, S. Stern, K. Peggs, and Haematology British Committee for Standards in. 2016. 'Guidelines for the management of diffuse large B-cell lymphoma', *Br J Haematol*, 174: 43-56.

Chapuy, B., C. Stewart, A. J. Dunford, J. Kim, A. Kamburov, R. A. Redd, M. S. Lawrence, M. G. M. Roemer, A. J. Li, M. Ziepert, A. M. Staiger, J. A. Wala, M. D. Ducar, I. Leshchiner, E. Rheinbay, A. Taylor-Weiner, C. A. Coughlin, J. M. Hess, C. S. Pedamallu, D. Livitz, D. Rosebrock, M. Rosenberg, A. A. Tracy, H. Horn, P. van Hummelen, A. L. Feldman, B. K. Link, A. J. Novak, J. R. Cerhan, T. M. Habermann, R. Siebert, A. Rosenwald, A. R. Thorner, M. L. Meyerson, T. R. Golub, R. Beroukhim, G. G. Wulf, G. Ott, S. J. Rodig, S. Monti, D. S. Neuberg, M. Loeffler, M. Pfreundschuh, L. Trumper, G. Getz, and M. A. Shipp. 2018. 'Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes', *Nat Med*, 24: 679-90.

Chen, S., M. Liu, and Y. Zhou. 2018. 'Bioinformatics Analysis for Cell-Free Tumor DNA Sequencing Data', *Methods Mol Biol*, 1754: 67-95.

Cheng, S. H., P. Jiang, K. Sun, Y. K. Cheng, K. C. Chan, T. Y. Leung, R. W. Chiu, and Y. M. Lo. 2015. 'Noninvasive prenatal testing by nanopore sequencing of maternal plasma DNA: feasibility assessment', *Clin Chem*, 61: 1305-6.

Cheson, B. D., R. I. Fisher, S. F. Barrington, F. Cavalli, L. H. Schwartz, E. Zucca, T. A. Lister, Australasian Leukaemia Alliance, Group Lymphoma, Group Eastern Cooperative Oncology, Consortium European Mantle Cell Lymphoma, Foundation Italian Lymphoma, Research European Organisation for, Group Treatment of Cancer/Dutch Hemato-Oncology, Osea Grupo Espanol de Medula, Group German High-Grade Lymphoma Study, Group German Hodgkin's Study, Group Japanese Lymphorra Study, Association Lymphoma Study, Ncic Clinical Trials Group, Group Nordic Lymphoma Study, Group Southwest Oncology, and Institute United Kingdom National Cancer Research. 2014. 'Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification', *J Clin Oncol*, 32: 3059-68.

Chiaretti, S., M. Marinelli, I. Del Giudice, S. Bonina, A. Piciocchi, M. Messina, M. Vignetti, D. Rossi, V. Di Maio, F. R. Mauro, A. Guarini, G. Gaidano, and R. Foa. 2014. 'NOTCH1, SF3B1, BIRC3 and TP53 mutations in patients with chronic lymphocytic leukemia undergoing first-line treatment: correlation with biological parameters and response to treatment', *Leuk Lymphoma*, 55: 2785-92.

Chigrinova, E., A. Rinaldi, I. Kwee, D. Rossi, P. M. Rancoita, J. C. Strefford, D. Oscier, K. Stamatopoulos, T. Papadaki, F. Berger, K. H. Young, F. Murray, R. Rosenquist, T. C. Greiner, W. C. Chan, E. M. Orlandi, M. Lucioni, R. Marasca, G. Inghirami, M.

Ladetto, F. Forconi, S. Cogliatti, H. Votavova, S. H. Swerdlow, S. Stilgenbauer, M. A. Piris, A. Matolcsy, D. Spagnolo, E. Nikitin, A. Zamo, V. Gattei, G. Bhagat, G. Ott, E. Zucca, G. Gaidano, and F. Bertoni. 2013. 'Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome', *Blood*, 122: 2673-82.

Choi, W. W., D. D. Weisenburger, T. C. Greiner, M. A. Piris, A. H. Banham, J. Delabie, R. M. Braziel, H. Geng, J. Iqbal, G. Lenz, J. M. Vose, C. P. Hans, K. Fu, L. M. Smith, M. Li, Z. Liu, R. D. Gascoyne, A. Rosenwald, G. Ott, L. M. Rimsza, E. Campo, E. S. Jaffe, D. L. Jaye, L. M. Staudt, and W. C. Chan. 2009. 'A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy', *Clin Cancer Res*, 15: 5494-502.

Clifford, R., T. Louis, P. Robbe, S. Ackroyd, A. Burns, A. T. Timbs, G. Wright Colopy, H. Dreau, F. Sigaux, J. G. Judde, M. Rotger, A. Telenti, Y. L. Lin, P. Pasero, J. Maelfait, M. Titsias, D. R. Cohen, S. J. Henderson, M. T. Ross, D. Bentley, P. Hillmen, A. Pettitt, J. Rehwinkel, S. J. Knight, J. C. Taylor, Y. J. Crow, M. Benkirane, and A. Schuh. 2014. 'SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage', *Blood*, 123: 1021-31.

Cohen, J. D., L. Li, Y. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A. A. Javed, F. Wong, A. Mattox, R. H. Hruban, C. L. Wolfgang, M. G. Goggins, M. Dal Molin, T. L. Wang, R. Roden, A. P. Klein, J. Ptak, L. Dobbyn, J. Schaefer, N. Silliman, M. Popoli, J. T. Vogelstein, J. D. Browne, R. E. Schoen, R. E. Brand, J. Tie, P. Gibbs, H. L. Wong, A. S. Mansfield, J. Jen, S. M. Hanash, M. Falconi, P. J. Allen, S. Zhou, C. Bettegowda, L. A. Diaz, Jr., C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. M. Lennon, and N. Papadopoulos. 2018. 'Detection and localization of surgically resectable cancers with a multi-analyte blood test', *Science*, 359: 926-30.

Condoluci, A., and D. Rossi. 2019. 'Genetic mutations in chronic lymphocytic leukemia: impact on clinical treatment', *Expert Rev Hematol*, 12: 89-98.

Condoluci, A., L. Terzi di Bergamo, P. Langerbeins, M. A. Hoechstetter, C. D. Herling, L. De Paoli, J. Delgado, K. G. Rabe, M. Gentile, M. Doubek, F. R. Mauro, G. Chiodin, M. Mattsson, J. Bahlo, G. Cutrona, J. Kotaskova, C. Deambrogi, K. E. Smedby, V. Spina, A. Bruscaggin, W. Wu, R. Moia, E. Bianchi, B. Gerber, E. Zucca, S. Gillessen, M. Ghielmini, F. Cavalli, G. Stussi, M. A. Hess, T. S. Baumann, A. Neri, M. Ferrarini, R. Rosenquist, F. Forconi, R. Foa, S. Pospisilova, F. Morabito, S. Stilgenbauer, H. Dohner, S. A. Parikh, W. G. Wierda, E. Montserrat, G. Gaidano, M. Hallek, and D. Rossi. 2020. 'International prognostic score for asymptomatic early-stage chronic lymphocytic leukemia', *Blood*, 135: 1859-69.

Corcoran, R. B., and B. A. Chabner. 2018. 'Application of Cell-free DNA Analysis to Cancer Treatment', *N Engl J Med*, 379: 1754-65.

Costa, A. F., A. Altemani, and M. Hermsen. 2011. 'Current concepts on dedifferentiation/high-grade transformation in salivary gland tumors', *Patholog Res Int*, 2011: 325965.

Cretu Stancu, M., M. J. van Roosmalen, I. Renkens, M. M. Nieboer, S. Middelkamp, J. de Ligt, G. Pregno, D. Giachino, G. Mandrile, J. Espejo Valle-Inclan, J. Korzelius, E. de Bruijn, E. Cuppen, M. E. Talkowski, T. Marschall, J. de Ridder, and W. P. Kloosterman. 2017. 'Mapping and phasing of structural variation in patient genomes using nanopore sequencing', *Nat Commun*, 8: 1326.

Cumbo, C., L. Impera, C. F. Minervini, P. Orsini, L. Anelli, A. Zagaria, N. Coccaro, G. Tota, A. Minervini, P. Casieri, C. Brunetti, A. R. Rossi, E. Parciante, G. Specchia, and F. Albano. 2018. 'Genomic BCR-ABL1 breakpoint characterization by a multi-strategy

approach for "personalized monitoring" of residual disease in chronic myeloid leukemia patients', *Oncotarget*, 9: 10978-86.

da Cunha-Bang, C., C. H. Geisler, L. Enggaard, C. B. Poulsen, P. de Nully Brown, H. Frederiksen, O. J. Bergmann, E. J. Pulczynski, R. S. Pedersen, L. H. Nielsen, I. Christiansen, and C. U. Niemann. 2016. 'The Danish National Chronic Lymphocytic Leukemia Registry', *Clin Epidemiol*, 8: 561-65.

Da Cunha-Bang, Caspar, Rudy Agius, Arnon P. Kater, Mark-David Levin, Anders Österborg, Mattias Mattsson, and Carsten Utoft Niemann. 2019. 'PreVent-ACaLL Short-term combined acalabrutinib and venetoclax treatment of newly diagnosed patients with CLL at high risk of infection and/or early treatment, who do not fulfil IWCLL treatment criteria for treatment. A randomized study with extensive immune phenotyping', *Blood*, 134: 4304-04.

Damle, R. N., T. Wasil, F. Fais, F. Ghiotto, A. Valetto, S. L. Allen, A. Buchbinder, D. Budman, K. Dittmar, J. Kolitz, S. M. Lichtman, P. Schulman, V. P. Vinciguerra, K. R. Rai, M. Ferrarini, and N. Chiorazzi. 1999. 'Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia', *Blood*, 94: 1840-7.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. 2021. 'Twelve years of SAMtools and BCFtools', *Gigascience*, 10.

Davies, A. 2020. 'Acalabrutinib in Combination With R-CHOP for Previously Untreated Diffuse Large B-cell Lymphoma (DLBCL) (REMoDL-A)', NIH, Accessed 17/09/2022.

Davies, A., S. Barrans, C. Burton, K. Mercer, J. Caddy, F. Chinnery, L. Day, D. Fernando, K. Ardeshna, G. Collins, J. Radford, S. Rule, A. McMillan, P. Johnson, and G. Griffiths. 2020. 'ACCEPT - combining acalabrutinib with rituximab, cyclophosphamide, doxorubicin, vincristine and prednisolone (R-CHOP) for Diffuse Large B-cell Lymphoma (DLBCL): study protocol for a Phase Ib/II open-label non-randomised clinical trial', *F1000Res*, 9: 941.

Davies, A., T. E. Cummin, S. Barrans, T. Maishman, C. Mamot, U. Novak, J. Caddy, L. Stanton, S. Kazmi-Stokes, A. McMillan, P. Fields, C. Pocock, G. P. Collins, R. Stephens, F. Cucco, A. Clipson, C. Sha, R. Tooze, M. A. Care, G. Griffiths, M. Q. Du, D. R. Westhead, C. Burton, and P. W. M. Johnson. 2019. 'Gene-expression profiling of bortezomib added to standard chemoimmunotherapy for diffuse large B-cell lymphoma (REMoDL-B): an open-label, randomised, phase 3 trial', *Lancet Oncol*, 20: 649-62.

Davies, A. J., S. Barrans, L. Stanton, J. Caddy, S. Wilding, G. Saunders, C. Mamot, U. Novak, A. McMillan, P. Fields, G. P. Collins, R. Stephens, F. Cucco, C. Sha, M. van Hoppe, R. Tooze, J. R. Davies, G. Griffiths, A. Schuh, C. Burton, D. R. Westhead, M. Q. Du, and P. W. M. Johnson. 2023. 'Differential Efficacy From the Addition of Bortezomib to R-CHOP in Diffuse Large B-Cell Lymphoma According to the Molecular Subgroup in the REMoDL-B Study With a 5-Year Follow-Up', *J Clin Oncol*: JCO2300033.

De Coster, W., P. De Rijk, A. De Roeck, T. De Pooter, S. D'Hert, M. Strazisar, K. Sleegers, and C. Van Broeckhoven. 2019. 'Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome', *Genome Res*, 29: 1178-87.

De Paoli, L., M. Cerri, S. Monti, S. Rasi, V. Spina, A. Bruscaggin, M. Greco, C. Ciardullo, R. Fama, S. Cresta, R. Maffei, M. Ladetto, M. Martini, L. Laurenti, F. Forconi, R. Marasca, L. M. Larocca, F. Bertoni, G. Gaidano, and D. Rossi. 2013. 'MGA, a

suppressor of MYC, is recurrently inactivated in high risk chronic lymphocytic leukemia', *Leuk Lymphoma*, 54: 1087-90.

Del Giudice, I., M. Marinelli, J. Wang, S. Bonina, M. Messina, S. Chiaretti, C. Ilari, L. Cafforio, S. Raponi, F. R. Mauro, V. Di Maio, M. S. De Propris, M. Nanni, C. Ciardullo, D. Rossi, G. Gaidano, A. Guarini, R. Rabadan, and R. Foa. 2016. 'Inter- and intra-patient clonal and subclonal heterogeneity of chronic lymphocytic leukaemia: evidences from circulating and lymph nodal compartments', *Br J Haematol*, 172: 371-83.

Delfau-Larue, M. H., A. van der Gucht, J. Dupuis, J. P. Jais, I. Nel, A. Beldi-Ferchiou, S. Hamdane, I. Benmaad, G. Laboure, B. Verret, C. Haioun, C. Copie-Bergman, A. Berriolo-Riedinger, P. Robert, R. O. Casasnovas, and E. Itti. 2018. 'Total metabolic tumor volume, circulating tumor cells, cell-free DNA: distinct prognostic value in follicular lymphoma', *Blood Adv*, 2: 807-16.

Dicker, F., H. Herholz, S. Schnittger, A. Nakao, N. Patten, L. Wu, W. Kern, T. Haferlach, and C. Haferlach. 2009. 'The detection of TP53 mutations in chronic lymphocytic leukemia independently predicts rapid disease progression and is highly correlated with a complex aberrant karyotype', *Leukemia*, 23: 117-24.

Dighiero, G., K. Maloum, B. Desablens, B. Cazin, M. Navarro, R. Leblay, M. Leporrier, J. Jaubert, G. Lepeu, B. Dreyfus, J. L. Binet, and P. Travade. 1998. 'Chlorambucil in indolent chronic lymphocytic leukemia. French Cooperative Group on Chronic Lymphocytic Leukemia', *N Engl J Med*, 338: 1506-14.

Do, H., and A. Dobrovic. 2015. 'Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization', *Clin Chem*, 61: 64-71.

Dohner, H., S. Stilgenbauer, A. Benner, E. Leupolt, A. Krober, L. Bullinger, K. Dohner, M. Bentz, and P. Lichter. 2000. 'Genomic aberrations and survival in chronic lymphocytic leukemia', *N Engl J Med*, 343: 1910-6.

Dohner, H., A. H. Wei, F. R. Appelbaum, C. Craddock, C. D. DiNardo, H. Dombret, B. L. Ebert, P. Fenaux, L. A. Godley, R. P. Hasserjian, R. A. Larson, R. L. Levine, Y. Miyazaki, D. Niederwieser, G. Ossenkoppele, C. Rollig, J. Sierra, E. M. Stein, M. S. Tallman, H. F. Tien, J. Wang, A. Wierzbowska, and B. Lowenberg. 2022. 'Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN', *Blood*, 140: 1345-77.

Duhrsen, U., S. Muller, B. Hertenstein, H. Thomssen, J. Kotzerke, R. Mesters, W. E. Berdel, C. Franzius, F. Kroschinsky, M. Weckesser, D. Kofahl-Krause, F. M. Bengel, J. Durig, J. Matschke, C. Schmitz, T. Poppel, C. Ose, M. Brinkmann, P. La Rosee, M. Freesmeyer, A. Hertel, H. G. Hoffkes, D. Behringer, G. Prange-Krex, S. Wilop, T. Krohn, J. Holzinger, M. Griesshammer, A. Giagounidis, A. Raghavachar, G. Maschmeyer, I. Brink, H. Bernhard, U. Haberkorn, T. Gaska, L. Kurch, D. M. E. van Assema, W. Klapper, D. Hoelzer, L. Geworski, K. H. Jockel, A. Scherag, A. Bockisch, J. Rekowski, A. Huttmann, and Petal Trial Investigators. 2018. 'Positron Emission Tomography-Guided Therapy of Aggressive Non-Hodgkin Lymphomas (PETAL): A Multicenter, Randomized Phase III Trial', *J Clin Oncol*, 36: 2024-34.

Eertink, J. J., C. N. Burggraaff, M. W. Heymans, U. Duhrsen, A. Huttmann, C. Schmitz, S. Muller, P. J. Lugtenburg, S. F. Barrington, N. G. Mikhaeel, R. Carr, S. Czibor, T. Gyorke, L. Ceriani, E. Zucca, M. Hutchings, L. Kostakoglu, A. Loft, S. Fanti, S. E. Wiegers, S. Pieplenbosch, R. Boellaard, O. S. Hoekstra, J. M. Zijlstra, and H. C. W. de Vet. 2021. 'Optimal timing and criteria of interim PET in DLBCL: a comparative study of 1692 patients', *Blood Adv*, 5: 2375-84.

el Rouby, S., A. Thomas, D. Costin, C. R. Rosenberg, M. Potmesil, R. Silber, and E. W. Newcomb. 1993. 'p53 gene mutation in B-cell chronic lymphocytic leukemia is

associated with drug resistance and is independent of MDR1/MDR3 gene expression', *Blood*, 82: 3452-9.
Elnenaei, M. O., P. Knopf, S. D. Cutler, K. Sinclair, M. Abou El Hassan, W. Greer, M. Goudie, J. Wagner, D. White, S. Couban, N. Forward, D. Gaston, and C. J. V. Campbell. 2019. 'Low-depth sequencing for copy number abnormalities in multiple myeloma supersedes fluorescent in situ hybridization in scope and resolution', *Clin Genet*, 96: 163-68.
Euskirchen, P., F. Bielle, K. Labreche, W. P. Kloosterman, S. Rosenberg, M. Daniau, C. Schmitt, J. Masliah-Planchon, F. Bourdeaut, C. Dehais, Y. Marie, J. Y. Delattre, and A. Idbaih. 2017. 'Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing', *Acta Neuropathol*, 134: 691-703.
Eyre, T. A., R. Clifford, A. Bloor, L. Boyle, C. Roberts, M. Cabes, G. P. Collins, S. Devereux, G. Follows, C. P. Fox, J. Gribben, P. Hillmen, C. S. Hatton, T. J. Littlewood, H. McCarthy, J. Murray, A. R. Pettitt, E. Soilleux, B. Stamatopoulos, S. B. Love, A. Wotherspoon, and A. Schuh. 2016. 'NCRI phase II study of CHOP in combination with ofatumumab in induction and maintenance in newly diagnosed Richter syndrome', *Br J Haematol*, 175: 43-54.
Fabbri, G., and R. Dalla-Favera. 2016. 'The molecular pathogenesis of chronic lymphocytic leukaemia', *Nat Rev Cancer*, 16: 145-62.
Fabbri, G., H. Khiabanian, A. B. Holmes, J. Wang, M. Messina, C. G. Mullighan, L. Pasqualucci, R. Rabadan, and R. Dalla-Favera. 2013. 'Genetic lesions associated with chronic lymphocytic leukemia transformation to Richter syndrome', *J Exp Med*, 210: 2273-88.
Fenaux, P., C. Preudhomme, J. L. Lai, I. Quiquandon, P. Jonveaux, M. Vanrumbeke, C. Sartiaux, P. Morel, M. H. Loucheux-Lefebvre, F. Bauters, and et al. 1992. 'Mutations of the p53 gene in B-cell chronic lymphocytic leukemia: a report on 39 cases with cytogenetic analysis', *Leukemia*, 6: 246-50.
Fernandez-Miranda, I., L. Pedrosa, M. Llanos, F. F. Franco, S. Gomez, P. Martin-Acosta, F. R. Garcia-Arroyo, J. Guma, B. Horcajo, A. K. Ballesteros, L. Galvez, N. Martinez, M. Marin, S. Sequero, M. Navarro, N. Yanguas-Casas, V. Calvo, A. Rueda-Dominguez, M. Provencio, and M. Sanchez-Beato. 2023. 'Monitoring of Circulating Tumor DNA Predicts Response to Treatment and Early Progression in Follicular Lymphoma: Results of a Prospective Pilot Study', *Clin Cancer Res*, 29: 209-20.
Fleischhacker, M., and B. Schmidt. 2007. 'Circulating nucleic acids (CNAs) and cancer--a survey', *Biochim Biophys Acta*, 1775: 181-232.
Fontanilles, M., F. Marguet, E. Bohers, P. J. Viailly, S. Dubois, P. Bertrand, V. Camus, S. Mareschal, P. Ruminy, C. Maingonnat, S. Lepretre, E. L. Veresezan, S. Derrey, H. Tilly, J. M. Picquenot, A. Laquerriere, and F. Jardin. 2017. 'Non-invasive detection of somatic mutations using next-generation sequencing in primary central nervous system lymphoma', *Oncotarget*, 8: 48157-68.
Freeman, C. L., K. J. Savage, D. R. Villa, D. W. Scott, L. Srour, A. S. Gerrie, M. J. Brown, G. W. Slack, P. Farinha, B. Skinnider, J. Morris, F. Benard, C. Aquino-Parsons, A. Lo, T. Pickles, D. C. Wilson, P. Tonseth, J. M. Connors, and L. H. Sehn. 2021. 'Long-term results of PET-guided radiation in patients with advanced-stage diffuse large B-cell lymphoma treated with R-CHOP', *Blood*, 137: 929-38.
Friedmann-Morvinski, D., and I. M. Verma. 2014. 'Dedifferentiation and reprogramming: origins of cancer stem cells', *EMBO Rep*, 15: 244-53.
Fuentes, Azahara, Alicia Serrano, Blanca Ferrer, Blanca Navarro, Veronica Lendinez, Laura Olivares, Carmen Ivorra, Ana Isabel Teruel, Carlos Solano, Maria Jose Terol, and Javier F. Chaves. 2017. 'Development and Validation of a New Tool for the

Characterization of Immunoglobulin Gene Rearrangements in Patients with Chronic Lymphocytic Leukemia By Deep Next Generation Sequencing', *Blood*, 130: 1714.

Fung, S. S., K. L. Hillier, C. S. Leger, I. Sandhu, L. M. Vickars, P. F. Galbraith, C. H. Li, and H. A. Leitch. 2007. 'Clinical progression and outcome of patients with monoclonal B-cell lymphocytosis', *Leuk Lymphoma*, 48: 1087-91.

Furstenau, M., J. Weiss, A. Giza, F. Franzen, S. Robrecht, A. M. Fink, K. Fischer, C. Schneider, E. Tausch, S. Stilgenbauer, M. Ritgen, A. Schilhabel, M. Bruggemann, B. Eichhorst, M. Hallek, and P. Cramer. 2022. 'Circulating tumor DNA-based MRD assessment in patients with CLL treated with obinutuzumab, acalabrutinib and venetoclax', *Clin Cancer Res*.

Gaidano, G., P. Ballerini, J. Z. Gong, G. Inghirami, A. Neri, E. W. Newcomb, I. T. Magrath, D. M. Knowles, and R. Dalla-Favera. 1991. 'p53 mutations in human lymphoid malignancies: association with Burkitt lymphoma and chronic lymphocytic leukemia', *Proc Natl Acad Sci U S A*, 88: 5413-7.

Galle, V., Vlummens, P., Offner, O. 2018. 'Is it possible to predict who may never need treatment for B-CLL?', *Belgian Journal of Haematology*, 9: 124-9.

Genovese, G., A. K. Kahler, R. E. Handsaker, J. Lindberg, S. A. Rose, S. F. Bakhoum, K. Chambert, E. Mick, B. M. Neale, M. Fromer, S. M. Purcell, O. Svantesson, M. Landen, M. Hoglund, S. Lehmann, S. B. Gabriel, J. L. Moran, E. S. Lander, P. F. Sullivan, P. Sklar, H. Gronberg, C. M. Hultman, and S. A. McCarroll. 2014. 'Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence', *N Engl J Med*, 371: 2477-87.

Georgiadis, P., I. Liampa, D. G. Hebels, J. Krauskopf, A. Chatziioannou, I. Valavanis, Tmcm de Kok, J. C. S. Kleinjans, I. A. Bergdahl, B. Melin, F. Spaeth, D. Palli, R. C. H. Vermeulen, J. Vlaanderen, M. Chadeau-Hyam, P. Vineis, S. A. Kyrtopoulos, and consortium EnviroGenomarkers. 2017. 'Evolving DNA methylation and gene expression markers of B-cell chronic lymphocytic leukemia are present in pre-diagnostic blood samples more than 10 years prior to diagnosis', *BMC Genomics*, 18: 728.

Giacopelli, B., Q. Zhao, A. S. Ruppert, A. Agyeman, C. Weigel, Y. Z. Wu, M. M. Gerber, K. G. Rabe, M. C. Larson, J. Lu, J. S. Blachly, K. A. Rogers, W. G. Wierda, J. R. Brown, K. R. Rai, M. Keating, L. Z. Rassenti, T. J. Kipps, T. Zenz, T. D. Shanafelt, N. E. Kay, L. V. Abruzzo, K. R. Coombes, J. A. Woyach, J. C. Byrd, and C. C. Oakes. 2019. 'Developmental subtypes assessed by DNA methylation-iPLEX forecast the natural history of chronic lymphocytic leukemia', *Blood*, 134: 688-98.

Giudicelli, V., X. Brochet, and M. P. Lefranc. 2011. 'IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences', *Cold Spring Harb Protoc*, 2011: 695-715.

Gleeson, M., E. A. Hawkes, D. Cunningham, A. Jack, and D. Linch. 2015. 'Caution in the Use of Immunohistochemistry for Determination of Cell of Origin in Diffuse Large B-Cell Lymphoma', *J Clin Oncol*, 33: 3215-6.

Gopal, S., W. A. Wood, S. J. Lee, T. C. Shea, K. N. Naresh, P. N. Kazembe, C. Casper, P. B. Hesseling, and R. T. Mitsuyasu. 2012. 'Meeting the challenge of hematologic malignancies in sub-Saharan Africa', *Blood*, 119: 5078-87.

Gregory, W., Alexiou, D., Rawstron, A., Pivonka, D., Sail, K. . 2020. "Extrapolating Progression Free Survival Curves in CLL using Peripheral Blood MRD Measurements from Venetoclax Trials." In *European Haematology Association*.

Griffith, D. M., Veech, J. A., & Marsh, C. J. 2016. 'cooccur: Probabilistic Species Co-Occurrence Analysis in R.', *Journal of Statistical Software*, 69: 1-17.

Gutierrez-Garcia, G., T. Cardesa-Salzmann, F. Climent, E. Gonzalez-Barca, S. Mercadal, J. L. Mate, J. M. Sancho, L. Arenillas, S. Serrano, L. Escoda, S. Martinez, A. Valera, A. Martinez, P. Jares, M. Pinyol, A. Garcia-Herrera, A. Martinez-Trillos, E. Gine, N. Villamor, E. Campo, L. Colomo, A. Lopez-Guillermo, and I. Balears Grup per l'Estudi dels Limfomes de Catalunya. 2011. 'Gene-expression profiling and not immunophenotypic algorithms predicts prognosis in patients with diffuse large B-cell lymphoma treated with immunochemotherapy', *Blood*, 117: 4836-43.

Hallek, M., B. D. Cheson, D. Catovsky, F. Caligaris-Cappio, G. Dighiero, H. Dohner, P. Hillmen, M. Keating, E. Montserrat, N. Chiorazzi, S. Stilgenbauer, K. R. Rai, J. C. Byrd, B. Eichhorst, S. O'Brien, T. Robak, J. F. Seymour, and T. J. Kipps. 2018. 'iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL', *Blood*, 131: 2745-60.

Hamblin, T. J., Z. Davis, A. Gardiner, D. G. Oscier, and F. K. Stevenson. 1999. 'Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia', *Blood*, 94: 1848-54.

Hans, C. P., D. D. Weisenburger, T. C. Greiner, R. D. Gascoyne, J. Delabie, G. Ott, H. K. Muller-Hermelink, E. Campo, R. M. Braziel, E. S. Jaffe, Z. Pan, P. Farinha, L. M. Smith, B. Falini, A. H. Banham, A. Rosenwald, L. M. Staudt, J. M. Connors, J. O. Armitage, and W. C. Chan. 2004. 'Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray', *Blood*, 103: 275-82.

Herling, C. D., F. Cymbalista, C. Gross-Ophoff-Muller, J. Bahlo, S. Robrecht, P. Langerbeins, A. M. Fink, O. Al-Sawaf, R. Busch, R. Porcher, B. Cazin, B. Dreyfus, S. Ibach, S. Lepretre, K. Fischer, F. Kaiser, B. Eichhorst, C. M. Wentner, M. A. Hoechstetter, H. Dohner, V. Leblond, M. Kneba, R. Letestu, S. Bottcher, S. Stilgenbauer, M. Hallek, and V. Levy. 2020. 'Early treatment with FCR versus watch and wait in patients with stage Binet A high-risk chronic lymphocytic leukemia (CLL): a randomized phase 3 trial', *Leukemia*, 34: 2038-50.

Hillmen, P., A. Schuh, T. A. Eyre, J. M. Pagel, J. R. Brown, P. Ghia, J. N. Allan, W. Wierda, A. Hamdy, J. Huang, R. Izumi, P. Patel, M. H. Wang, and J. C. Byrd. 2016. 'Acalabrutinib Monotherapy in Patients with Richter Transformation from the Phase 1/2 ACE-CL-001 Clinical Study', *Blood*, 128.

Hoechstetter, M. A., R. Busch, B. Eichhorst, A. Buhler, D. Winkler, M. J. Eckart, U. Vehling-Kaiser, H. Schimke, U. Jager, H. J. Hurtz, G. Hopfinger, F. Hartmann, H. Fuss, W. Abenhardt, I. Blau, W. Freier, L. Muller, M. Goebeler, C. M. Wendtner, J. Bahlo, K. Fischer, M. Bentz, B. Emmerich, H. Dohner, M. Hallek, and S. Stilgenbauer. 2017. 'Early, risk-adapted treatment with fludarabine in Binet stage A chronic lymphocytic leukemia patients: results of the CLL1 trial of the German CLL study group', *Leukemia*, 31: 2833-37.

Howell, D. A., A. G. Smith, A. Jack, R. Patmore, U. Macleod, E. Mironska, and E. Roman. 2013. 'Time-to-diagnosis and symptoms of myeloma, lymphomas and leukaemias: a report from the Haematological Malignancy Research Network', *BMC Hematol*, 13: 9.

Howlett, C., S. J. Snedecor, D. J. Landsburg, J. Svoboda, E. A. Chong, S. J. Schuster, S. D. Nasta, T. Feldman, A. Rago, K. M. Walsh, S. Weber, A. Goy, and A. Mato. 2015. 'Front-line, dose-escalated immunochemotherapy is associated with a significant progression-free survival advantage in patients with double-hit lymphomas: a systematic review and meta-analysis', *Br J Haematol*, 170: 504-14.

Hu, B., K. P. Patel, H. C. Chen, X. Wang, R. Luthra, M. J. Routbort, R. Kanagal-Shamanna, L. J. Medeiros, C. C. Yin, Z. Zuo, C. Y. Ok, S. Loghavi, G. Tang, F. P. Tambaro, P. Thompson, J. Burger, N. Jain, A. Ferrajoli, P. Bose, Z. Estrov, M. Keating, and W. G.

Wierda. 2019. 'Association of gene mutations with time-to-first treatment in 384 treatment-naive chronic lymphocytic leukaemia patients', *Br J Haematol*, 187: 307-18.

Huber, H., S. Edenhofer, J. von Tresckow, S. Robrecht, C. Zhang, E. Tausch, C. Schneider, J. Bloehdorn, M. Furstenau, P. Dreger, M. Ritgen, T. Illmer, A. L. Illert, J. Durig, S. Bottcher, C. U. Niemann, M. Kneba, A. M. Fink, K. Fischer, H. Dohner, M. Hallek, B. Eichhorst, and S. Stilgenbauer. 2022. 'Obinutuzumab (GA-101), ibrutinib, and venetoclax (GIVe) frontline treatment for high-risk chronic lymphocytic leukemia', *Blood*, 139: 1318-29.

Huet, S., A. Bouvard, E. Ferrant, I. Mosnier, K. Chabane, G. Salles, A. S. Michallet, S. Hayette, and P. Sujobert. 2020. 'Impact of using leader primers for IGHV mutational status assessment in chronic lymphocytic leukemia', *Leukemia*, 34: 2257-59.

International, C. L. L. I. P. I. working group. 2016. 'An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data', *Lancet Oncol*, 17: 779-90.

Jain, Nitin, Alessandra Ferrajoli, Sreyashi Basu, Philip A. Thompson, Jan A. Burger, Tapan M. Kadia, Zeev E. Estrov, Naveen Pemmaraju, Wanda Lopez, Beenu Thakral, Joseph David Khoury, Carlos E. Bueso-Ramos, Jorge Blando, Susan M. O'Brien, Hagop M. Kantarjian, James P Allison, Michael J Keating, Padmanee Sharma, and William G. Wierda. 2018. 'A Phase II Trial of Nivolumab Combined with Ibrutinib for Patients with Richter Transformation', *Blood*, 132: 296-96.

Jain, P., M. J. Keating, W. G. Wierda, M. Sivina, P. A. Thompson, A. Ferrajoli, Z. Estrov, H. Kantarjian, S. O'Brien, and J. A. Burger. 2017. 'Long-term Follow-up of Treatment with Ibrutinib and Rituximab in Patients with High-Risk Chronic Lymphocytic Leukemia', *Clin Cancer Res*, 23: 2154-58.

Jain, P., G. M. Nogueras Gonzalez, R. Kanagal-Shamanna, U. Rozovski, N. Sarwari, C. Tam, W. G. Wierda, P. A. Thompson, N. Jain, R. Luthra, A. Quesada, G. Sanchez-Petitto, A. Ferrajoli, J. Burger, H. Kantarjian, J. Cortes, S. O'Brien, M. J. Keating, and Z. Estrov. 2018. 'The absolute percent deviation of IGHV mutation rather than a 98% cut-off predicts survival of chronic lymphocytic leukaemia patients treated with fludarabine, cyclophosphamide and rituximab', *Br J Haematol*, 180: 33-40.

Jaiswal, S., P. Fontanillas, J. Flannick, A. Manning, P. V. Grauman, B. G. Mar, R. C. Lindsley, C. H. Mermel, N. Burtt, A. Chavez, J. M. Higgins, V. Moltchanov, F. C. Kuo, M. J. Kluk, B. Henderson, L. Kinnunen, H. A. Koistinen, C. Ladenvall, G. Getz, A. Correa, B. F. Banahan, S. Gabriel, S. Kathiresan, H. M. Stringham, M. I. McCarthy, M. Boehnke, J. Tuomilehto, C. Haiman, L. Groop, G. Atzmon, J. G. Wilson, D. Neuberg, D. Altshuler, and B. L. Ebert. 2014. 'Age-related clonal hematopoiesis associated with adverse outcomes', *N Engl J Med*, 371: 2488-98.

Jamshidi, A., M. C. Liu, E. A. Klein, O. Venn, E. Hubbell, J. F. Beausang, S. Gross, C. Melton, A. P. Fields, Q. Liu, N. Zhang, E. T. Fung, K. N. Kurtzman, H. Amini, C. Betts, D. Civello, P. Freese, R. Calef, K. Davydov, S. Fayzullina, C. Hou, R. Jiang, B. Jung, S. Tang, V. Demas, J. Newman, O. Sakarya, E. Scott, A. Shenoy, S. Shojaee, K. K. Steffen, V. Nicula, T. C. Chien, S. Bagaria, N. Hunkapiller, M. Desai, Z. Dong, D. A. Richards, T. J. Yeatman, A. L. Cohn, D. D. Thiel, D. A. Berry, M. K. Tummala, K. McIntyre, M. A. Sekeres, A. Bryce, A. M. Aravanis, M. V. Seiden, and C. Swanton. 2022. 'Evaluation of cell-free DNA approaches for multi-cancer early detection', *Cancer Cell*, 40: 1537-49 e12.

Jaramillo, S., A. Agathangelidis, C. Schneider, J. Bahlo, S. Robrecht, E. Tausch, J. Bloehdorn, M. Hoechstetter, K. Fischer, B. Eichhorst, V. Goede, M. Hallek, H. Dohner, R. Rosenquist, P. Ghia, K. Stamatopoulos, and S. Stilgenbauer. 2020.

'Prognostic impact of prevalent chronic lymphocytic leukemia stereotyped subsets: analysis within prospective clinical trials of the German CLL Study Group (GCLLSG)', *Haematologica*, 105: 2598-607.

Jeck, W. R., J. Lee, H. Robinson, L. P. Le, A. J. Iafrate, and V. Nardi. 2019. 'A Nanopore Sequencing-Based Assay for Rapid Detection of Gene Fusions', *J Mol Diagn*, 21: 58-69.

Jiang, F., W. Liu, L. Zhang, Y. Guo, M. Chen, X. Zeng, Y. Wang, Y. Li, J. Xian, B. Du, Y. Xie, S. Ouyang, S. Li, Y. Yang, C. Zhang, F. Luo, and X. Sun. 2021. 'Noninvasive prenatal testing for beta-thalassemia by targeted nanopore sequencing combined with relative haplotype dosage (RHDO): a feasibility study', *Sci Rep*, 11: 5714.

Jones, D., J. A. Woyach, W. Zhao, S. Caruthers, H. Tu, J. Coleman, J. C. Byrd, A. J. Johnson, and G. Lozanski. 2017. 'PLCG2 C2 domain mutations co-occur with BTK and PLCG2 resistance mutations in chronic lymphocytic leukemia undergoing ibrutinib treatment', *Leukemia*, 31: 1645-47.

Juarez, M., C. Egoavil, M. Rodriguez-Soler, E. Hernandez-Illan, C. Guarinos, A. Garcia-Martinez, C. Alenda, M. Giner-Calabuig, O. Murcia, C. Mangas, A. Paya, J. R. Aparicio, F. A. Ruiz, J. Martinez, J. A. Casellas, J. L. Soto, P. Zapater, and R. Jover. 2017. 'KRAS and BRAF somatic mutations in colonic polyps and the risk of metachronous neoplasia', *PLoS One*, 12: e0184937.

Kastenhuber, E. R., and S. W. Lowe. 2017. 'Putting p53 in Context', *Cell*, 170: 1062-78.

Katsman, E., S. Orlanski, F. Martignano, I. Fox-Fisher, R. Shemer, Y. Dor, A. Zick, A. Eden, I. Petrini, S. G. Conticello, and B. P. Berman. 2022. 'Detecting cell-of-origin and cancer-specific methylation features of cell-free DNA from Nanopore sequencing', *Genome Biol*, 23: 158.

Khoury, J. D., E. Solary, O. Abla, Y. Akkari, R. Alaggio, J. F. Apperley, R. Bejar, E. Berti, L. Busque, J. K. C. Chan, W. Chen, X. Chen, W. J. Chng, J. K. Choi, I. Colmenero, S. E. Coupland, N. C. P. Cross, D. De Jong, M. T. Elghetany, E. Takahashi, J. F. Emile, J. Ferry, L. Fogelstrand, M. Fontenay, U. Germing, S. Gujral, T. Haferlach, C. Harrison, J. C. Hodge, S. Hu, J. H. Jansen, R. Kanagal-Shamanna, H. M. Kantarjian, C. P. Kratz, X. Q. Li, M. S. Lim, K. Loeb, S. Loghavi, A. Marcogliese, S. Meshinchi, P. Michaels, K. N. Naresh, Y. Natkunam, R. Nejati, G. Ott, E. Padron, K. P. Patel, N. Patkar, J. Picarsic, U. Platzbecker, I. Roberts, A. Schuh, W. Sewell, R. Siebert, P. Tembhare, J. Tyner, S. Verstovsek, W. Wang, B. Wood, W. Xiao, C. Yeung, and A. Hochhaus. 2022. 'The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms', *Leukemia*, 36: 1703-19.

Kiechle, M., M. Hinrichs, A. Jacobsen, J. Luttges, J. Pfisterer, F. Kommoss, and N. Arnold. 2000. 'Genetic imbalances in precursor lesions of endometrial cancer detected by comparative genomic hybridization', *Am J Pathol*, 156: 1827-33.

Killcoyne, S., E. Gregson, D. C. Wedge, D. J. Woodcock, M. D. Eldridge, R. de la Rue, A. Miremadi, S. Abbas, A. Blasko, C. Kosmidou, W. Januszewicz, A. V. Jenkins, M. Gerstung, and R. C. Fitzgerald. 2020. 'Genomic copy number predicts esophageal cancer years before transformation', *Nat Med*, 26: 1726-32.

Kleinstern, G., D. R. O'Brien, X. Li, S. Tian, B. F. Kabat, K. G. Rabe, A. D. Norman, H. Yan, C. M. Vachon, N. J. Boddicker, T. G. Call, S. A. Parikh, L. Bruins, C. Bonolo de Campos, J. F. Leis, T. D. Shanafelt, W. Ding, J. R. Cerhan, N. E. Kay, S. L. Slager, and E. Braggio. 2020. 'Tumor mutational load predicts time to first treatment in chronic lymphocytic leukemia (CLL) and monoclonal B-cell lymphocytosis beyond the CLL international prognostic index', *Am J Hematol*, 95: 906-17.

Klintman, J., N. Appleby, B. Stamatopoulos, K. Ridout, T. A. Eyre, P. Robbe, L. L. Pascua, S. J. L. Knight, H. Dreau, M. Cabes, N. Popitsch, M. Ehinger, J. I. Martin-Subero, E. Campo, R. Mansson, D. Rossi, J. C. Taylor, D. V. Vavoulis, and A. Schuh. 2021. 'Genomic and transcriptomic correlates of Richter transformation in chronic lymphocytic leukemia', *Blood*, 137: 2800-16.

Knight, S. J., C. Yau, R. Clifford, A. T. Timbs, E. Sadighi Akha, H. M. Dreau, A. Burns, C. Ciria, D. G. Oscier, A. R. Pettitt, S. Dutton, C. C. Holmes, J. Taylor, J. B. Cazier, and A. Schuh. 2012. 'Quantification of subclonal distributions of recurrent genomic aberrations in paired pre-treatment and relapse samples from patients with B-cell chronic lymphocytic leukemia', *Leukemia*, 26: 1564-75.

Knisbacher, B. A., Z. Lin, C. K. Hahn, F. Nadeu, M. Duran-Ferrer, K. E. Stevenson, E. Tausch, J. Delgado, A. Barbera-Mourelle, A. Taylor-Weiner, P. Bousquets-Munoz, A. Diaz-Navarro, A. Dunford, S. Anand, H. Kretzmer, J. Gutierrez-Abril, S. Lopez-Tamargo, S. M. Fernandes, C. Sun, M. Sivina, L. Z. Rassenti, C. Schneider, S. Li, L. Parida, A. Meissner, F. Aguet, J. A. Burger, A. Wiestner, T. J. Kipps, J. R. Brown, M. Hallek, C. Stewart, D. S. Neuberg, J. I. Martin-Subero, X. S. Puente, S. Stilgenbauer, C. J. Wu, E. Campo, and G. Getz. 2022. 'Molecular map of chronic lymphocytic leukemia and its impact on outcome', *Nat Genet*, 54: 1664-74.

Koboldt, D. C., Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan, L. Lin, C. A. Miller, E. R. Mardis, L. Ding, and R. K. Wilson. 2012. 'VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing', *Genome Res*, 22: 568-76.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, and A. M. Phillippy. 2017. 'Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation', *Genome Res*, 27: 722-36.

Kreuzberger, N., J. A. Damen, M. Trivella, L. J. Estcourt, A. Aldin, L. Umlauff, M. D. Vazquez-Montes, R. Wolff, K. G. Moons, I. Monsef, F. Foroutan, K. A. Kreuzer, and N. Skoetz. 2020. 'Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis', *Cochrane Database Syst Rev*, 7: CD012022.

Krishnakumar, R., A. Sinha, S. W. Bird, H. Jayamohan, H. S. Edwards, J. S. Schoeniger, K. D. Patel, S. S. Branda, and M. S. Bartsch. 2018. 'Systematic and stochastic influences on the performance of the MinION nanopore sequencer across a range of nucleotide bias', *Sci Rep*, 8: 3159.

Kulis, M., S. Heath, M. Bibikova, A. C. Queiros, A. Navarro, G. Clot, A. Martinez-Trillos, G. Castellano, I. Brun-Heath, M. Pinyol, S. Barberan-Soler, P. Papasaikas, P. Jares, S. Bea, D. Rico, S. Ecker, M. Rubio, R. Royo, V. Ho, B. Klotzle, L. Hernandez, L. Conde, M. Lopez-Guerra, D. Colomer, N. Villamor, M. Aymerich, M. Rozman, M. Bayes, M. Gut, J. L. Gelpi, M. Orozco, J. B. Fan, V. Quesada, X. S. Puente, D. G. Pisano, A. Valencia, A. Lopez-Guillermo, I. Gut, C. Lopez-Otin, E. Campo, and J. I. Martin-Subero. 2012. 'Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia', *Nat Genet*, 44: 1236-42.

Kurtz, D. M., M. S. Esfahani, F. Scherer, J. Soo, M. C. Jin, C. L. Liu, A. M. Newman, U. Duhrsen, A. Huttmann, O. Casasnovas, J. R. Westin, M. Ritgen, S. Bottcher, A. W. Langerak, M. Roschewski, W. H. Wilson, G. Gaidano, D. Rossi, J. Bahlo, M. Hallek, R. Tibshirani, M. Diehn, and A. A. Alizadeh. 2019. 'Dynamic Risk Profiling Using Serial Tumor Biomarkers for Personalized Outcome Prediction', *Cell*, 178: 699-713 e19.

Kurtz, D. M., M. R. Green, S. V. Bratman, F. Scherer, C. L. Liu, C. A. Kunder, K. Takahashi, C. Glover, C. Keane, S. Kihira, B. Visser, J. Callahan, K. A. Kong, M. Faham, K. S.

Corbelli, D. Miklos, R. H. Advani, R. Levy, R. J. Hicks, M. Hertzberg, R. S. Ohgami, M. K. Gandhi, M. Diehn, and A. A. Alizadeh. 2015. 'Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing', *Blood*, 125: 3679-87.

Kurtz, D. M., F. Scherer, M. C. Jin, J. Soo, A. F. M. Craig, M. S. Esfahani, J. J. Chabon, H. Stehr, C. L. Liu, R. Tibshirani, L. S. Maeda, N. K. Gupta, M. S. Khodadoust, R. H. Advani, R. Levy, A. M. Newman, U. Duhrsen, A. Huttmann, M. Meignan, R. O. Casasnovas, J. R. Westin, M. Roschewski, W. H. Wilson, G. Gaidano, D. Rossi, M. Diehn, and A. A. Alizadeh. 2018. 'Circulating Tumor DNA Measurements As Early Outcome Predictors in Diffuse Large B-Cell Lymphoma', *J Clin Oncol*, 36: 2845-53.

Kurtz, D. M., J. Soo, L. Co Ting Keh, S. Alig, J. J. Chabon, B. J. Sworder, A. Schultz, M. C. Jin, F. Scherer, A. Garofalo, C. W. Macaulay, E. G. Hamilton, B. Chen, M. Olsen, J. G. Schroers-Martin, A. F. M. Craig, E. J. Moding, M. S. Esfahani, C. L. Liu, U. Duhrsen, A. Huttmann, R. O. Casasnovas, J. R. Westin, M. Roschewski, W. H. Wilson, G. Gaidano, D. Rossi, M. Diehn, and A. A. Alizadeh. 2021. 'Enhanced detection of minimal residual disease by targeted sequencing of phased variants in circulating tumor DNA', *Nat Biotechnol*, 39: 1537-47.

Lackraj, T., S. Ben Barouch, J. J. F. Medeiros, S. Pedersen, A. Danesh, M. Bakhtiari, M. Hong, K. Tong, J. Joynt, A. Arruda, M. D. Minden, J. Kuruvilla, S. Bhella, V. Kukreti, M. Crump, A. Prica, C. Chen, Y. Deng, W. Xu, T. J. Pugh, A. Keating, J. E. Dick, S. Abelson, and R. Kridel. 2022. 'Clinical significance of clonal hematopoiesis in the setting of autologous stem cell transplantation for lymphoma', *Am J Hematol*, 97: 1538-47.

Lakhotia, R., C. Melani, K. Dunleavy, S. Pittaluga, N. Saba, L. Lindenberg, E. Mena, E. Bergvall, A. N. Lucas, A. Jacob, E. Yusko, S. M. Steinberg, E. S. Jaffe, A. Wiestner, W. H. Wilson, and M. Roschewski. 2022. 'Circulating tumor DNA predicts therapeutic outcome in mantle cell lymphoma', *Blood Adv*, 6: 2667-80.

Lanasa, M. C., S. D. Allgood, S. L. Slager, S. S. Dave, C. Love, G. E. Marti, N. E. Kay, C. A. Hanson, K. G. Rabe, S. J. Achenbach, L. R. Goldin, N. J. Camp, B. K. Goodman, C. M. Vachon, L. G. Spector, L. Z. Rassenti, J. F. Leis, J. P. Gockerman, S. S. Strom, T. G. Call, M. Glenn, J. R. Cerhan, M. C. Levesque, J. B. Weinberg, and N. E. Caporaso. 2011. 'Immunophenotypic and gene expression analysis of monoclonal B-cell lymphocytosis shows biologic characteristics associated with good prognosis CLL', *Leukemia*, 25: 1459-66.

Landau, D. A., S. L. Carter, P. Stojanov, A. McKenna, K. Stevenson, M. S. Lawrence, C. Sougnez, C. Stewart, A. Sivachenko, L. Wang, Y. Wan, W. Zhang, S. A. Shukla, A. Vartanov, S. M. Fernandes, G. Saksena, K. Cibulskis, B. Tesar, S. Gabriel, N. Hacohen, M. Meyerson, E. S. Lander, D. Neuberg, J. R. Brown, G. Getz, and C. J. Wu. 2013. 'Evolution and impact of subclonal mutations in chronic lymphocytic leukemia', *Cell*, 152: 714-26.

Landau, D. A., E. Tausch, A. N. Taylor-Weiner, C. Stewart, J. G. Reiter, J. Bahlo, S. Kluth, I. Bozic, M. Lawrence, S. Bottcher, S. L. Carter, K. Cibulskis, D. Mertens, C. L. Sougnez, M. Rosenberg, J. M. Hess, J. Edelmann, S. Kless, M. Kneba, M. Ritgen, A. Fink, K. Fischer, S. Gabriel, E. S. Lander, M. A. Nowak, H. Dohner, M. Hallek, D. Neuberg, G. Getz, S. Stilgenbauer, and C. J. Wu. 2015. 'Mutations driving CLL and their evolution in progression and relapse', *Nature*, 526: 525-30.

Langerak, A. W. 2022. *Immuno-genetics Methods and Protocols* (Humana Press: https://doi.org/10.1007/978-1-0716-2115-8).

Langerak, A. W., F. Davi, P. Ghia, A. Hadzidimitriou, F. Murray, K. N. Potter, R. Rosenquist, K. Stamatopoulos, C. Belessi, and C. L. L. European Research Initiative

on. 2011. 'Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases', *Leukemia*, 25: 979-84.

Langerbeins, P., R. Busch, N. Anheier, J. Durig, M. Bergmann, M. E. Goebeler, H. J. Hurtz, M. B. Stauch, S. Stilgenbauer, H. Dohner, A. M. Fink, P. Cramer, K. Fischer, C. M. Wendtner, M. Hallek, and B. Eichhorst. 2014. 'Poor efficacy and tolerability of R-CHOP in relapsed/refractory chronic lymphocytic leukemia and Richter transformation', *Am J Hematol*, 89: E239-43.

Langerbeins, P., C. Zhang, S. Robrecht, P. Cramer, M. Furstenau, O. Al-Sawaf, J. von Tresckow, A. M. Fink, K. A. Kreuzer, U. Vehling-Kaiser, E. Tausch, L. Muller, M. J. Eckart, R. Schlag, W. Freier, T. Gaska, C. Balser, M. Reiser, M. Stauch, C. M. Wendtner, K. Fischer, S. Stilgenbauer, B. Eichhorst, and M. Hallek. 2022. 'The CLL12 trial: ibrutinib vs placebo in treatment-naive, early-stage chronic lymphocytic leukemia', *Blood*, 139: 177-87.

Larrayoz, M., M. J. Rose-Zerilli, L. Kadalayil, H. Parker, S. Blakemore, J. Forster, Z. Davis, A. J. Steele, A. Collins, M. Else, D. Catovsky, D. G. Oscier, and J. C. Strefford. 2017. 'Non-coding NOTCH1 mutations in chronic lymphocytic leukemia; their clinical impact in the UK CLL4 trial', *Leukemia*, 31: 510-14.

Lauer, E. M., J. Mutter, and F. Scherer. 2022. 'Circulating tumor DNA in B-cell lymphoma: technical advances, clinical applications, and perspectives for translational research', *Leukemia*, 36: 2151-64.

Lawrence, M. S., P. Stojanov, C. H. Mermel, J. T. Robinson, L. A. Garraway, T. R. Golub, M. Meyerson, S. B. Gabriel, E. S. Lander, and G. Getz. 2014. 'Discovery and saturation analysis of cancer genes across 21 tumour types', *Nature*, 505: 495-501.

Leeksma, A. C., P. Baliakas, T. Moysiadis, A. Puiggros, K. Plevova, A. M. Van der Kevie-Kersemaekers, H. Posthuma, A. E. Rodriguez-Vicente, A. N. Tran, G. Barbany, L. Mansouri, R. Gunnarsson, H. Parker, E. Van den Berg, M. Bellido, Z. Davis, M. Wall, I. Scarpelli, A. Osterborg, L. Hansson, M. Jarosova, P. Ghia, P. Poddighe, B. Espinet, S. Pospisilova, C. Tam, L. Ysebaert, F. Nguyen-Khac, D. Oscier, C. Haferlach, J. Schoumans, M. Stevens-Kroef, E. Eldering, K. Stamatopoulos, R. Rosenquist, J. C. Strefford, C. Mellink, and A. P. Kater. 2021. 'Genomic arrays identify high-risk chronic lymphocytic leukemia with genomic complexity: a multi-center study', *Haematologica*, 106: 87-97.

Legason, I. D., M. D. Ogwang, C. Chamba, E. Mkwizu, C. El Mouden, H. Mwinula, L. Chirande, A. Schuh, and F. Chiwanga. 2022. 'A protocol to clinically evaluate liquid biopsies as a tool to speed up diagnosis of children and young adults with aggressive infection-related lymphoma in East Africa "(AI-REAL)"', *BMC Cancer*, 22: 484.

Lenartova, A., U. Randen, T. B. Johannesen, and G. E. Tjonnfjord. 2019. 'Richter syndrome epidemiology in a large population based chronic lymphocytic leukemia cohort from Norway', *Cancer Epidemiol*, 60: 128-33.

Li, C., K. R. Chng, E. J. Boey, A. H. Ng, A. Wilm, and N. Nagarajan. 2016. 'INC-Seq: accurate single molecule reads using nanopore sequencing', *Gigascience*, 5: 34.

Li, H. 2018. 'Minimap2: pairwise alignment for nucleotide sequences', *Bioinformatics*, 34: 3094-100.

Li, H., and R. Durbin. 2009. 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, 25: 1754-60.

Li, J., A. Smith, S. Crouch, S. Oliver, and E. Roman. 2016. 'Estimating the prevalence of hematological malignancies and precursor conditions using data from Haematological Malignancy Research Network (HMRN)', *Cancer Causes Control*, 27: 1019-26.

Li, S., M. P. Lefranc, J. J. Miles, E. Alamyar, V. Giudicelli, P. Duroux, J. D. Freeman, V. D. Corbin, J. P. Scheerlinck, M. A. Frohman, P. U. Cameron, M. Plebanski, B. Loveland, S. R. Burrows, A. T. Papenfuss, and E. J. Gowans. 2013. 'IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling', *Nat Commun*, 4: 2333.

Liu, T. M., J. A. Woyach, Y. Zhong, A. Lozanski, G. Lozanski, S. Dong, E. Strattan, A. Lehman, X. Zhang, J. A. Jones, J. Flynn, L. A. Andritsos, K. Maddocks, S. M. Jaglowski, K. A. Blum, J. C. Byrd, J. A. Dubovsky, and A. J. Johnson. 2015. 'Hypermorphic mutation of phospholipase C, gamma2 acquired in ibrutinib-resistant CLL confers BTK independency upon B-cell receptor activation', *Blood*, 126: 61-8.

Ljungstrom, V., D. Cortese, E. Young, T. Pandzic, L. Mansouri, K. Plevova, S. Ntoufa, P. Baliakas, R. Clifford, L. A. Sutton, S. J. Blakemore, N. Stavroyianni, A. Agathangelidis, D. Rossi, M. Hoglund, J. Kotaskova, G. Juliusson, C. Belessi, N. Chiorazzi, P. Panagiotidis, A. W. Langerak, K. E. Smedby, D. Oscier, G. Gaidano, A. Schuh, F. Davi, C. Pott, J. C. Strefford, L. Trentin, S. Pospisilova, P. Ghia, K. Stamatopoulos, T. Sjoblom, and R. Rosenquist. 2016. 'Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations', *Blood*, 127: 1007-16.

Lo, Y. M. D., D. S. C. Han, P. Jiang, and R. W. K. Chiu. 2021. 'Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies', *Science*, 372.

Lo, Y. M., F. M. Lun, K. C. Chan, N. B. Tsui, K. C. Chong, T. K. Lau, T. Y. Leung, B. C. Zee, C. R. Cantor, and R. W. Chiu. 2007. 'Digital PCR for the molecular detection of fetal chromosomal aneuploidy', *Proc Natl Acad Sci U S A*, 104: 13116-21.

Lu, H., F. Giordano, and Z. Ning. 2016. 'Oxford Nanopore MinION Sequencing and Genome Assembly', *Genomics Proteomics Bioinformatics*, 14: 265-79.

Maclachlan, K. H., E. H. Rustad, A. Derkach, B. Zheng-Lin, V. Yellapantula, B. Diamond, M. Hultcrantz, B. Ziccheddu, E. M. Boyle, P. Blaney, N. Bolli, Y. Zhang, A. Dogan, A. M. Lesokhin, G. J. Morgan, O. Landgren, and F. Maura. 2021. 'Copy number signatures predict chromothripsis and clinical outcomes in newly diagnosed multiple myeloma', *Nat Commun*, 12: 5172.

Maddocks, K. J., A. S. Ruppert, G. Lozanski, N. A. Heerema, W. Zhao, L. Abruzzo, A. Lozanski, M. Davis, A. Gordon, L. L. Smith, R. Mantel, J. A. Jones, J. M. Flynn, S. M. Jaglowski, L. A. Andritsos, F. Awan, K. A. Blum, M. R. Grever, A. J. Johnson, J. C. Byrd, and J. A. Woyach. 2015. 'Etiology of Ibrutinib Therapy Discontinuation and Outcomes in Patients With Chronic Lymphocytic Leukemia', *JAMA Oncol*, 1: 80-7.

Magi, A., D. Bolognini, N. Bartalucci, A. Mingrino, R. Semeraro, L. Giovannini, S. Bonifacio, D. Parrini, E. Pelo, F. Mannelli, P. Guglielmelli, and A. Maria Vannucchi. 2019. 'Nano-GLADIATOR: real-time detection of copy number alterations from nanopore sequencing data', *Bioinformatics*, 35: 4213-21.

Malcikova, J., E. Tausch, D. Rossi, L. A. Sutton, T. Soussi, T. Zenz, A. P. Kater, C. U. Niemann, D. Gonzalez, F. Davi, M. Gonzalez Diaz, C. Moreno, G. Gaidano, K. Stamatopoulos, R. Rosenquist, S. Stilgenbauer, P. Ghia, S. Pospisilova, and T. P. network European Research Initiative on Chronic Lymphocytic Leukemia. 2018. 'ERIC recommendations for TP53 mutation analysis in chronic lymphocytic leukemia-update on methodological approaches and results interpretation', *Leukemia*, 32: 1070-80.

Manier, S., K. Z. Salem, J. Park, D. A. Landau, G. Getz, and I. M. Ghobrial. 2017. 'Genomic complexity of multiple myeloma and its clinical implications', *Nat Rev Clin Oncol*, 14: 100-13.

Mansouri, L., B. Thorvaldsdottir, L. A. Sutton, G. Karakatsoulis, M. Meggendorfer, H. Parker, F. Nadeu, C. Brieghel, S. Laidou, R. Moia, D. Rossi, M. Catherwood, J. Kotaskova, J. Delgado, A. E. Rodriguez-Vicente, R. Benito, G. M. Rigolin, S. Bonfiglio, L. Scarfo, M. Mattsson, Z. Davis, A. Gogia, L. Rani, P. Baliakas, H. Foroughi-Asl, C. Jylha, A. Skaftason, I. Rapado, F. Miras, J. Martinez-Lopez, J. de la Serna, J. M. H. Rivas, P. Thornton, M. J. Larrayoz, M. J. Calasanz, V. Fesus, Z. Matrai, C. Bodor, K. E. Smedby, B. Espinet, A. Puiggros, R. Gupta, L. Bullinger, F. Bosch, B. Tazon-Vega, F. Baran-Marszak, D. Oscier, F. Nguyen-Khac, T. Zenz, M. J. Terol, A. Cuneo, M. Hernandez-Sanchez, S. Pospisilova, K. Mills, G. Gaidano, C. U. Niemann, E. Campo, J. C. Strefford, P. Ghia, K. Stamatopoulos, and R. Rosenquist. 2023. 'Different prognostic impact of recurrent gene mutations in chronic lymphocytic leukemia depending on IGHV gene somatic hypermutation status: a study by ERIC in HARMONY', *Leukemia*, 37: 339-47.

Martignano, F., U. Munagala, S. Crucitta, A. Mingrino, R. Semeraro, M. Del Re, I. Petrini, A. Magi, and S. G. Conticello. 2021. 'Nanopore sequencing from liquid biopsy: analysis of copy number variations from cell-free DNA of lung cancer patients', *Mol Cancer*, 20: 32.

Matern, B. M., T. I. Olieslagers, M. Groeneweg, B. Duygu, L. Wieten, M. G. J. Tilanus, and C. E. M. Voorter. 2020. 'Long-Read Nanopore Sequencing Validated for Human Leukocyte Antigen Class I Typing in Routine Diagnostics', *J Mol Diagn*, 22: 912-19.

Mato, A. R., J. C. Barrientos, N. Ghosh, J. M. Pagel, D. M. Brander, M. Gutierrez, K. Kadish, B. Tomlinson, R. Iyengar, D. Ipe, S. Upasani, C. I. Amaya-Chanaga, M. Sundaram, J. Han, N. Giafis, and J. P. Sharman. 2020. 'Prognostic Testing and Treatment Patterns in Chronic Lymphocytic Leukemia in the Era of Novel Targeted Therapies: Results From the informCLL Registry', *Clin Lymphoma Myeloma Leuk*, 20: 174-83 e3.

McKerrell, T., N. Park, T. Moreno, C. S. Grove, H. Ponstingl, J. Stephens, Group Understanding Society Scientific, C. Crawley, J. Craig, M. A. Scott, C. Hodkinson, J. Baxter, R. Rad, D. R. Forsyth, M. A. Quail, E. Zeggini, W. Ouwehand, I. Varela, and G. S. Vassiliou. 2015. 'Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis', *Cell Rep*, 10: 1239-45.

McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. 2016. 'The Ensembl Variant Effect Predictor', *Genome Biol*, 17: 122.

Meriranta, L., A. Alkodsi, A. Pasanen, M. Lepisto, P. Mapar, Y. N. Blaker, J. Jorgensen, M. L. Karjalainen-Lindsberg, I. Fiskvik, L. T. G. Mikalsen, M. Autio, M. Bjorkholm, M. Jerkeman, O. Fluge, P. Brown, S. Jyrkkio, H. Holte, E. Pitkanen, P. Ellonen, and S. Leppa. 2022. 'Molecular features encoded in the ctDNA reveal heterogeneity and predict outcome in high-risk aggressive B-cell lymphoma', *Blood*, 139: 1863-77.

Minervini, C. F., C. Cumbo, P. Orsini, L. Anelli, A. Zagaria, L. Impera, N. Coccaro, C. Brunetti, A. Minervini, P. Casieri, G. Tota, A. Russo Rossi, G. Specchia, and F. Albano. 2017. 'Mutational analysis in BCR-ABL1 positive leukemia by deep sequencing based on nanopore MinION technology', *Exp Mol Pathol*, 103: 33-37.

Minervini, C. F., C. Cumbo, P. Orsini, L. Anelli, A. Zagaria, G. Specchia, and F. Albano. 2020. 'Nanopore Sequencing in Blood Diseases: A Wide Range of Opportunities', *Front Genet*, 11: 76.

Minervini, C. F., C. Cumbo, P. Orsini, C. Brunetti, L. Anelli, A. Zagaria, A. Minervini, P. Casieri, N. Coccaro, G. Tota, L. Impera, A. Giordano, G. Specchia, and F. Albano. 2016. 'TP53 gene mutation analysis in chronic lymphocytic leukemia by nanopore MinION sequencing', *Diagn Pathol*, 11: 96.

Minervini, C. F., C. Cumbo, I. Redavid, M. R. Conserva, P. Orsini, A. Zagaria, L. Anelli, N. Coccaro, G. Tota, L. Impera, E. Parciante, F. Tarantini, A. Giordano, G. Specchia, P.

Musto, and F. Albano. 2021. 'Nanopore sequencing approach for immunoglobulin gene analysis in chronic lymphocytic leukemia', *Sci Rep*, 11: 17668.
Mohanty, A., N. Sandoval, M. Das, R. Pillai, L. Chen, R. W. Chen, H. M. Amin, M. Wang, G. Marcucci, D. D. Weisenburger, S. T. Rosen, L. V. Pham, and V. N. Ngo. 2016. 'CCND1 mutations increase protein stability and promote ibrutinib resistance in mantle cell lymphoma', *Oncotarget*, 7: 73558-72.
Molica, S., D. Giannarelli, L. Levato, R. Mirabelli, M. Gentile, and F. Morabito. 2017. 'Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance analysis of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI)', *Leuk Lymphoma*, 58: 1736-39.
Mollstedt, J., L. Mansouri, and R. Rosenquist. 2023. 'Precision diagnostics in chronic lymphocytic leukemia: Past, present and future', *Front Oncol*, 13: 1146486.
Mosquera Orgueira, A., B. Antelo Rodriguez, J. A. Diaz Arias, M. S. Gonzalez Perez, and J. L. Bello Lopez. 2019. 'New Recurrent Structural Aberrations in the Genome of Chronic Lymphocytic Leukemia Based on Exome-Sequencing Data', *Front Genet*, 10: 854.
Mouliere, F., B. Robert, E. Arnau Peyrotte, M. Del Rio, M. Ychou, F. Molina, C. Gongora, and A. R. Thierry. 2011. 'High fragmentation characterizes tumour-derived circulating DNA', *PLoS One*, 6: e23418.
Muchtar, E., N. E. Kay, and S. A. Parikh. 2021. 'Early intervention in asymptomatic chronic lymphocytic leukemia', *Clin Adv Hematol Oncol*, 19: 92-103.
Nadeu, F., G. Clot, J. Delgado, D. Martin-Garcia, T. Baumann, I. Salaverria, S. Bea, M. Pinyol, P. Jares, A. Navarro, H. Suarez-Cisneros, M. Aymerich, M. Rozman, N. Villamor, D. Colomer, M. Gonzalez, M. Alcoceba, M. J. Terol, B. Navarro, E. Colado, A. R. Payer, X. S. Puente, C. Lopez-Otin, A. Lopez-Guillermo, A. Enjuanes, and E. Campo. 2018. 'Clinical impact of the subclonal architecture and mutational complexity in chronic lymphocytic leukemia', *Leukemia*, 32: 645-53.
Nadeu, F., J. Delgado, C. Royo, T. Baumann, T. Stankovic, M. Pinyol, P. Jares, A. Navarro, D. Martin-Garcia, S. Bea, I. Salaverria, C. Oldreive, M. Aymerich, H. Suarez-Cisneros, M. Rozman, N. Villamor, D. Colomer, A. Lopez-Guillermo, M. Gonzalez, M. Alcoceba, M. J. Terol, E. Colado, X. S. Puente, C. Lopez-Otin, A. Enjuanes, and E. Campo. 2016. 'Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia', *Blood*, 127: 2122-30.
Nadeu, F., R. Mas-de-Les-Valls, A. Navarro, R. Royo, S. Martin, N. Villamor, H. Suarez-Cisneros, R. Mares, J. Lu, A. Enjuanes, A. Rivas-Delgado, M. Aymerich, T. Baumann, D. Colomer, J. Delgado, R. D. Morin, T. Zenz, X. S. Puente, P. J. Campbell, S. Bea, F. Maura, and E. Campo. 2020. 'IgCaller for reconstructing immunoglobulin gene rearrangements and oncogenic translocations from whole-genome sequencing in lymphoid neoplasms', *Nat Commun*, 11: 3390.
Nadeu, F., R. Royo, R. Massoni-Badosa, H. Playa-Albinyana, B. Garcia-Torre, M. Duran-Ferrer, K. J. Dawson, M. Kulis, A. Diaz-Navarro, N. Villamor, J. L. Melero, V. Chapaprieta, A. Dueso-Barroso, J. Delgado, R. Moia, S. Ruiz-Gil, D. Marchese, A. Giro, N. Verdaguer-Dot, M. Romo, G. Clot, M. Rozman, G. Frigola, A. Rivas-Delgado, T. Baumann, M. Alcoceba, M. Gonzalez, F. Climent, P. Abrisqueta, J. Castellvi, F. Bosch, M. Aymerich, A. Enjuanes, S. Ruiz-Gaspa, A. Lopez-Guillermo, P. Jares, S. Bea, S. Capella-Gutierrez, J. L. Gelpi, N. Lopez-Bigas, D. Torrents, P. J. Campbell, I. Gut, D. Rossi, G. Gaidano, X. S. Puente, P. M. Garcia-Roves, D. Colomer, H. Heyn, F. Maura, J. I. Martin-Subero, and E. Campo. 2022. 'Detection of

early seeding of Richter transformation in chronic lymphocytic leukemia', *Nat Med*, 28: 1662-71.

Navrkalova, V., K. Plevova, J. Hynst, K. Pal, A. Mareckova, T. Reigl, H. Jelinkova, Z. Vrzalova, K. Stranska, S. Pavlova, A. Panovska, A. Janikova, M. Doubek, J. Kotaskova, and S. Pospisilova. 2021. 'LYmphoid NeXt-Generation Sequencing (LYNX) Panel: A Comprehensive Capture-Based Sequencing Tool for the Analysis of Prognostic and Predictive Markers in Lymphoid Malignancies', *J Mol Diagn*, 23: 959-74.

Newman, A. M., A. F. Lovejoy, D. M. Klass, D. M. Kurtz, J. J. Chabon, F. Scherer, H. Stehr, C. L. Liu, S. V. Bratman, C. Say, L. Zhou, J. N. Carter, R. B. West, G. W. Sledge, Jr., J. B. Shrager, B. W. Loo, Jr., J. W. Neal, H. A. Wakelee, M. Diehn, and A. A. Alizadeh. 2016a. 'Integrated digital error suppression for improved detection of circulating tumor DNA', *Nat Biotechnol*, 34: 547-55.

Newman, A. M., A. F. Lovejoy, D. M. Klass, D. M. Kurtz, J. J. Chabon, F. Scherer, H. Stehr, C. L. Liu, S. V. Bratman, C. Say, L. Zhou, J. N. Carter, R. B. West, G. W. Sledge, J. B. Shrager, B. W. Loo, Jr., J. W. Neal, H. A. Wakelee, M. Diehn, and A. A. Alizadeh. 2016b. 'Integrated digital error suppression for improved detection of circulating tumor DNA', *Nat Biotechnol*, 34: 547-55.

Nicholson, B. D., J. Oke, C. Friedemann Smith, J. A. Phillips, J. Lee, L. Abel, S. Kelly, I. Gould, T. Mackay, Z. Kaveney, S. Anthony, S. Hayles, D. Lasserson, and F. Gleeson. 2018. 'The Suspected CANcer (SCAN) pathway: protocol for evaluating a new standard of care for patients with non-specific symptoms of cancer', *BMJ Open*, 8: e018168.

Nils Winkelmann, Mathew J Rose-Zerilli, Jade Forster, Mathew Parry, Anton Parker, Anne Gardiner, Zadie Davies, BSC, Andrew J Steele, Helen Parker, Andrew Collins, Nick C.P. Cross, David Oscier, Jonathan C Strefford. 2013. 'Mutations In XPO1, POT1, BIRC3 and FBXW7 collectively Predict Poor Outcome At Diagnosis In CLL and MBL Independent From The SF3B1 and NOTCH1 Status', *Blood*, 122: 4138.

Norris, A. L., R. E. Workman, Y. Fan, J. R. Eshleman, and W. Timp. 2016. 'Nanopore sequencing detects structural variants in cancer', *Cancer Biol Ther*, 17: 246-53.

Nowakowski, G. S., A. Chiappella, R. D. Gascoyne, D. W. Scott, Q. Zhang, W. Jurczak, M. Ozcan, X. Hong, J. Zhu, J. Jin, D. Belada, J. M. Bergua, F. Piazza, H. Mocikova, A. L. Molinari, D. H. Yoon, F. Cavallo, M. Tani, K. Yamamoto, K. Izutsu, K. Kato, M. Czuczman, S. Hersey, A. Kilcoyne, J. Russo, K. Hudak, J. Zhang, S. Wade, T. E. Witzig, and U. Vitolo. 2021. 'ROBUST: A Phase III Study of Lenalidomide Plus R-CHOP Versus Placebo Plus R-CHOP in Previously Untreated Patients With ABC-Type Diffuse Large B-Cell Lymphoma', *J Clin Oncol*, 39: 1317-28.

Nowakowski, G. S., F. Hong, D. W. Scott, W. R. Macon, R. L. King, T. M. Habermann, N. Wagner-Johnston, C. Casulo, J. L. Wade, G. G. Nagargoje, C. M. Reynolds, J. B. Cohen, N. Khan, J. E. Amengual, K. L. Richards, R. F. Little, J. P. Leonard, J. W. Friedberg, L. Kostakoglu, B. S. Kahl, and T. E. Witzig. 2021. 'Addition of Lenalidomide to R-CHOP Improves Outcomes in Newly Diagnosed Diffuse Large B-Cell Lymphoma in a Randomized Phase II US Intergroup Study ECOG-ACRIN E1412', *J Clin Oncol*, 39: 1329-38.

Ojha, J., C. Secreto, K. Rabe, J. Ayres-Silva, R. Tschumper, D. V. Dyke, S. Slager, R. Fonseca, T. Shanafelt, N. Kay, and E. Braggio. 2014. 'Monoclonal B-cell lymphocytosis is characterized by mutations in CLL putative driver genes and clonal heterogeneity many years before disease progression', *Leukemia*, 28: 2395-8.

Orsini, P., C. F. Minervini, C. Cumbo, L. Anelli, A. Zagaria, A. Minervini, N. Coccaro, G. Tota, P. Casieri, L. Impera, E. Parciante, C. Brunetti, A. Giordano, G. Specchia, and

F. Albano. 2018. 'Design and MinION testing of a nanopore targeted gene sequencing panel for chronic lymphocytic leukemia', *Sci Rep*, 8: 11798.

Ott, G., A. Rosenwald, and E. Campo. 2013. 'Understanding MYC-driven aggressive B-cell lymphomas: pathogenesis and classification', *Hematology Am Soc Hematol Educ Program*, 2013: 575-83.

Ott, G., M. Ziepert, W. Klapper, H. Horn, M. Szczepanowski, H. W. Bernd, C. Thorns, A. C. Feller, D. Lenze, M. Hummel, H. Stein, H. K. Muller-Hermelink, M. Frank, M. L. Hansmann, T. F. Barth, P. Moller, S. Cogliatti, M. Pfreundschuh, N. Schmitz, L. Trumper, M. Loeffler, and A. Rosenwald. 2010. 'Immunoblastic morphology but not the immunohistochemical GCB/nonGCB classifier predicts outcome in diffuse large B-cell lymphoma in the RICOVER-60 trial of the DSHNHL', *Blood*, 116: 4916-25.

Parikh, S. A., K. G. Rabe, N. E. Kay, T. G. Call, W. Ding, J. F. Leis, S. S. Kenderian, E. Muchtar, Y. Wang, A. B. Koehler, S. M. Schwager, C. E. Lesnick, G. Kleinstern, D. Van Dyke, C. A. Hanson, E. Braggio, S. L. Slager, and T. D. Shanafelt. 2021. 'The CLL International Prognostic Index predicts outcomes in monoclonal B-cell lymphocytosis and Rai 0 CLL', *Blood*, 138: 149-59.

Parry, E. M., I. Leshchiner, R. Guieze, C. Johnson, E. Tausch, S. A. Parikh, C. Lemvigh, J. Broseus, S. Hergalant, C. Messer, F. Utro, C. Levovitz, K. Rhrissorrakrai, L. Li, D. Rosebrock, S. Yin, S. Deng, K. Slowik, R. Jacobs, T. Huang, S. Li, G. Fell, R. Redd, Z. Lin, B. A. Knisbacher, D. Livitz, C. Schneider, N. Ruthen, L. Elagina, A. Taylor-Weiner, B. Persaud, A. Martinez, S. M. Fernandes, N. Purroy, A. J. Anandappa, J. Ma, J. Hess, L. Z. Rassenti, T. J. Kipps, N. Jain, W. Wierda, F. Cymbalista, P. Feugier, N. E. Kay, K. J. Livak, B. P. Danysh, C. Stewart, D. Neuberg, M. S. Davids, J. R. Brown, L. Parida, S. Stilgenbauer, G. Getz, and C. J. Wu. 2023a. 'Evolutionary history of transformation from chronic lymphocytic leukemia to Richter syndrome', *Nat Med*, 29: 158-69.

———. 2023b. 'Evolutionary history of transformation from chronic lymphocytic leukemia to Richter syndrome', *Nat Med*.

Pflug, N., J. Bahlo, T. D. Shanafelt, B. F. Eichhorst, M. A. Bergmann, T. Elter, K. Bauer, G. Malchau, K. G. Rabe, S. Stilgenbauer, H. Dohner, U. Jager, M. J. Eckart, G. Hopfinger, R. Busch, A. M. Fink, C. M. Wendtner, K. Fischer, N. E. Kay, and M. Hallek. 2014. 'Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia', *Blood*, 124: 49-62.

Phallen, J., A. Leal, B. D. Woodward, P. M. Forde, J. Naidoo, K. A. Marrone, J. R. Brahmer, J. Fiksel, J. E. Medina, S. Cristiano, D. N. Palsgrove, C. D. Gocke, D. C. Bruhm, P. Keshavarzian, V. Adleff, E. Weihe, V. Anagnostou, R. B. Scharpf, V. E. Velculescu, and H. Husain. 2019. 'Early Noninvasive Detection of Response to Targeted Therapy in Non-Small Cell Lung Cancer', *Cancer Res*, 79: 1204-13.

Puente, X. S., S. Bea, R. Valdes-Mas, N. Villamor, J. Gutierrez-Abril, J. I. Martin-Subero, M. Munar, C. Rubio-Perez, P. Jares, M. Aymerich, T. Baumann, R. Beekman, L. Belver, A. Carrio, G. Castellano, G. Clot, E. Colado, D. Colomer, D. Costa, J. Delgado, A. Enjuanes, X. Estivill, A. A. Ferrando, J. L. Gelpi, B. Gonzalez, S. Gonzalez, M. Gonzalez, M. Gut, J. M. Hernandez-Rivas, M. Lopez-Guerra, D. Martin-Garcia, A. Navarro, P. Nicolas, M. Orozco, A. R. Payer, M. Pinyol, D. G. Pisano, D. A. Puente, A. C. Queiros, V. Quesada, C. M. Romeo-Casabona, C. Royo, R. Royo, M. Rozman, N. Russinol, I. Salaverria, K. Stamatopoulos, H. G. Stunnenberg, D. Tamborero, M. J. Terol, A. Valencia, N. Lopez-Bigas, D. Torrents, I. Gut, A. Lopez-Guillermo, C. Lopez-Otin, and E. Campo. 2015. 'Non-coding recurrent mutations in chronic lymphocytic leukaemia', *Nature*, 526: 519-24.

Puente, X. S., M. Pinyol, V. Quesada, L. Conde, G. R. Ordonez, N. Villamor, G. Escaramis, P. Jares, S. Bea, M. Gonzalez-Diaz, L. Bassaganyas, T. Baumann, M. Juan, M. Lopez-Guerra, D. Colomer, J. M. Tubio, C. Lopez, A. Navarro, C. Tornador, M. Aymerich, M. Rozman, J. M. Hernandez, D. A. Puente, J. M. Freije, G. Velasco, A. Gutierrez-Fernandez, D. Costa, A. Carrio, S. Guijarro, A. Enjuanes, L. Hernandez, J. Yague, P. Nicolas, C. M. Romeo-Casabona, H. Himmelbauer, E. Castillo, J. C. Dohm, S. de Sanjose, M. A. Piris, E. de Alava, J. San Miguel, R. Royo, J. L. Gelpi, D. Torrents, M. Orozco, D. G. Pisano, A. Valencia, R. Guigo, M. Bayes, S. Heath, M. Gut, P. Klatt, J. Marshall, K. Raine, L. A. Stebbings, P. A. Futreal, M. R. Stratton, P. J. Campbell, I. Gut, A. Lopez-Guillermo, X. Estivill, E. Montserrat, C. Lopez-Otin, and E. Campo. 2011. 'Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia', *Nature*, 475: 101-5.

Queiros, A. C., N. Villamor, G. Clot, A. Martinez-Trillos, M. Kulis, A. Navarro, E. M. Penas, S. Jayne, A. Majid, J. Richter, A. K. Bergmann, J. Kolarova, C. Royo, N. Russinol, G. Castellano, M. Pinyol, S. Bea, I. Salaverria, M. Lopez-Guerra, D. Colomer, M. Aymerich, M. Rozman, J. Delgado, E. Gine, M. Gonzalez-Diaz, X. S. Puente, R. Siebert, M. J. Dyer, C. Lopez-Otin, C. Rozman, E. Campo, A. Lopez-Guillermo, and J. I. Martin-Subero. 2015. 'A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact', *Leukemia*, 29: 598-605.

Quesada, V., L. Conde, N. Villamor, G. R. Ordonez, P. Jares, L. Bassaganyas, A. J. Ramsay, S. Bea, M. Pinyol, A. Martinez-Trillos, M. Lopez-Guerra, D. Colomer, A. Navarro, T. Baumann, M. Aymerich, M. Rozman, J. Delgado, E. Gine, J. M. Hernandez, M. Gonzalez-Diaz, D. A. Puente, G. Velasco, J. M. Freije, J. M. Tubio, R. Royo, J. L. Gelpi, M. Orozco, D. G. Pisano, J. Zamora, M. Vazquez, A. Valencia, H. Himmelbauer, M. Bayes, S. Heath, M. Gut, I. Gut, X. Estivill, A. Lopez-Guillermo, X. S. Puente, E. Campo, and C. Lopez-Otin. 2011. 'Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia', *Nat Genet*, 44: 47-52.

Quesada, V., A. J. Ramsay, and C. Lopez-Otin. 2012. 'Chronic lymphocytic leukemia with SF3B1 mutation', *N Engl J Med*, 366: 2530.

Rai, K. R., A. Sawitsky, E. P. Cronkite, A. D. Chanana, R. N. Levy, and B. S. Pasternack. 1975. 'Clinical staging of chronic lymphocytic leukemia', *Blood*, 46: 219-34.

Raman, L., M. Van der Linden, K. Van der Eecken, K. Vermaelen, I. Demedts, V. Surmont, U. Himpe, F. Dedeurwaerdere, L. Ferdinande, Y. Lievens, K. Claes, B. Menten, and J. Van Dorpe. 2020. 'Shallow whole-genome sequencing of plasma cell-free DNA accurately differentiates small from non-small cell lung carcinoma', *Genome Med*, 12: 35.

Rasi, S., S. Monti, V. Spina, R. Foa, G. Gaidano, and D. Rossi. 2012. 'Analysis of NOTCH1 mutations in monoclonal B-cell lymphocytosis', *Haematologica*, 97: 153-4.

Rawstron, A. C., F. L. Bennett, S. J. O'Connor, M. Kwok, J. A. Fenton, M. Plummer, R. de Tute, R. G. Owen, S. J. Richards, A. S. Jack, and P. Hillmen. 2008. 'Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia', *N Engl J Med*, 359: 575-83.

Rawstron, A. C., and P. Hillmen. 2010. 'Clinical and diagnostic implications of monoclonal B-cell lymphocytosis', *Best Pract Res Clin Haematol*, 23: 61-9.

Rawstron, A. C., K. A. Kreuzer, A. Soosapilla, M. Spacek, O. Stehlikova, P. Gambell, N. McIver-Brown, N. Villamor, K. Psarra, M. Arroz, R. Milani, J. de la Serna, M. T. Cedena, O. Jaksic, J. Nomdedeu, C. Moreno, G. M. Rigolin, A. Cuneo, P. Johansen, H. E. Johnsen, R. Rosenquist, C. U. Niemann, W. Kern, D. Westerman, M. Trneny, S. Mulligan, M. Doubek, S. Pospisilova, P. Hillmen, D. Oscier, M. Hallek, P. Ghia, and

E. Montserrat. 2018. 'Reproducible diagnosis of chronic lymphocytic leukemia by flow cytometry: An European Research Initiative on CLL (ERIC) & European Society for Clinical Cell Analysis (ESCCA) Harmonisation project', *Cytometry B Clin Cytom*, 94: 121-28.

Rawstron, A., Webster, N., Dalal, S., Brock, K., Yates, F., . 2020. "Kinetics of Response in the Peripheral Blood Predicts Long Term Responses to Ibrutinib and Venetoclax Treatment for relapsed/refractory CLL in the Bloodwise TAP CLARITY Trial." In *European Haematology Association*.

Robbe, P., N. Popitsch, S. J. L. Knight, P. Antoniou, J. Becq, M. He, A. Kanapin, A. Samsonova, D. V. Vavoulis, M. T. Ross, Z. Kingsbury, M. Cabes, S. D. C. Ramos, S. Page, H. Dreau, K. Ridout, L. J. Jones, A. Tuff-Lacey, S. Henderson, J. Mason, F. M. Buffa, C. Verrill, D. Maldonado-Perez, I. Roxanis, E. Collantes, L. Browning, S. Dhar, S. Damato, S. Davies, M. Caulfield, D. R. Bentley, J. C. Taylor, C. Turnbull, A. Schuh, and Genomes Project. 2018. 'Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project', *Genet Med*, 20: 1196-205.

Robbe, P., K. E. Ridout, D. V. Vavoulis, H. Dreau, B. Kinnersley, N. Denny, D. Chubb, N. Appleby, A. Cutts, A. J. Cornish, L. Lopez-Pascua, R. Clifford, A. Burns, B. Stamatopoulos, M. Cabes, R. Alsolami, P. Antoniou, M. Oates, D. Cavalieri, Consortium Genomics England Research, C. L. L. pilot consortium, J. Gibson, A. V. Prabhu, R. Schwessinger, D. Jennings, T. James, U. Maheswari, M. Duran-Ferrer, P. Carninci, S. J. L. Knight, R. Mansson, J. Hughes, J. Davies, M. Ross, D. Bentley, J. C. Strefford, S. Devereux, A. R. Pettitt, P. Hillmen, M. J. Caulfield, R. S. Houlston, J. I. Martin-Subero, and A. Schuh. 2022. 'Whole-genome sequencing of chronic lymphocytic leukemia identifies subgroups with distinct biological and clinical features', *Nat Genet*, 54: 1675-89.

Roberts, P., S. A. Burchill, S. Brownhill, C. J. Cullinane, C. Johnston, M. J. Griffiths, D. J. McMullan, N. P. Bown, S. P. Morris, and I. J. Lewis. 2008. 'Ploidy and karyotype complexity are powerful prognostic indicators in the Ewing's sarcoma family of tumors: a study by the United Kingdom Cancer Cytogenetics and the Children's Cancer and Leukaemia Group', *Genes Chromosomes Cancer*, 47: 207-20.

Roessner, P. M., and M. Seiffert. 2020. 'T-cells in chronic lymphocytic leukemia: Guardians or drivers of disease?', *Leukemia*, 34: 2012-24.

Rogers, K. A., Y. Huang, A. S. Ruppert, G. Salem, D. M. Stephens, N. A. Heerema, L. A. Andritsos, F. T. Awan, J. C. Byrd, J. M. Flynn, K. J. Maddocks, and J. A. Jones. 2018. 'A single-institution retrospective cohort study of first-line R-EPOCH chemoimmunotherapy for Richter syndrome demonstrating complex chronic lymphocytic leukaemia karyotype as an adverse prognostic factor', *Br J Haematol*, 180: 259-66.

Roschewski, M., K. Dunleavy, S. Pittaluga, M. Moorhead, F. Pepin, K. Kong, M. Shovlin, E. S. Jaffe, L. M. Staudt, C. Lai, S. M. Steinberg, C. C. Chen, J. Zheng, T. D. Willis, M. Faham, and W. H. Wilson. 2015. 'Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study', *Lancet Oncol*, 16: 541-9.

Rosenquist, R., P. Ghia, A. Hadzidimitriou, L. A. Sutton, A. Agathangelidis, P. Baliakas, N. Darzentas, V. Giudicelli, M. P. Lefranc, A. W. Langerak, C. Belessi, F. Davi, and K. Stamatopoulos. 2017. 'Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations', *Leukemia*, 31: 1477-81.

Rossi, D., A. Bruscaggin, V. Spina, S. Rasi, H. Khiabanian, M. Messina, M. Fangazio, T. Vaisitti, S. Monti, S. Chiaretti, A. Guarini, I. Del Giudice, M. Cerri, S. Cresta, C.

Deambrogi, E. Gargiulo, V. Gattei, F. Forconi, F. Bertoni, S. Deaglio, R. Rabadan, L. Pasqualucci, R. Foa, R. Dalla-Favera, and G. Gaidano. 2011. 'Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness', *Blood*, 118: 6904-8.

Rossi, D., A. Condoluci, V. Spina, and G. Gaidano. 2019. 'Methods for Measuring ctDNA in Lymphomas', *Methods Mol Biol*, 1881: 253-65.

Rossi, D., F. Diop, E. Spaccarotella, S. Monti, M. Zanni, S. Rasi, C. Deambrogi, V. Spina, A. Bruscaggin, C. Favini, R. Serra, A. Ramponi, R. Boldorini, R. Foa, and G. Gaidano. 2017. 'Diffuse large B-cell lymphoma genotyping on the liquid biopsy', *Blood*, 129: 1947-57.

Rossi, D., D. M. Kurtz, M. Roschewski, F. Cavalli, E. Zucca, and W. H. Wilson. 2020. 'The development of liquid biopsy for research and clinical practice in lymphomas: Report of the 15-ICML workshop on ctDNA', *Hematol Oncol*, 38: 34-37.

Rossi, D., S. Rasi, V. Spina, A. Bruscaggin, S. Monti, S. Cresta, R. Fama, C. Deambrogi, M. Greco, M. Fangazio, C. Ciardullo, D. Piranda, G. M. Casaluci, M. Messina, I. D. Giudice, S. Chiaretti, M. Marinelli, A. Guarini, R. Foa, and G. Gaidano. 2012. 'The genome of chemorefractory chronic lymphocytic leukemia reveals frequent mutations of NOTCH1 and SF3B1', *Leuk Suppl*, 1: S26-8.

Rossi, D., E. Sozzi, A. Puma, L. De Paoli, S. Rasi, V. Spina, A. Gozzetti, M. Tassi, E. Cencini, D. Raspadori, V. Pinto, F. Bertoni, V. Gattei, F. Lauria, G. Gaidano, and F. Forconi. 2009. 'The prognosis of clinical monoclonal B cell lymphocytosis differs from prognosis of Rai 0 chronic lymphocytic leukaemia and is recapitulated by biological risk factors', *Br J Haematol*, 146: 64-75.

Rossi, D., V. Spina, C. Deambrogi, S. Rasi, L. Laurenti, K. Stamatopoulos, L. Arcaini, M. Lucioni, G. B. Rocque, Z. Y. Xu-Monette, C. Visco, J. Chang, E. Chigrinova, F. Forconi, R. Marasca, C. Besson, T. Papadaki, M. Paulli, L. M. Larocca, S. A. Pileri, V. Gattei, F. Bertoni, R. Foa, K. H. Young, and G. Gaidano. 2011. 'The genetics of Richter syndrome reveals disease heterogeneity and predicts survival after transformation', *Blood*, 117: 3391-401.

Rossi, D., V. Spina, and G. Gaidano. 2018. 'Biology and treatment of Richter syndrome', *Blood*, 131: 2761-72.

Rotbain, E. C., H. Frederiksen, H. Hjalgrim, K. Rostgaard, G. J. Egholm, B. Zahedi, C. B. Poulsen, L. Enggard, C. da Cunha-Bang, and C. U. Niemann. 2020. 'IGHV mutational status and outcome for patients with chronic lymphocytic leukemia upon treatment: a Danish nationwide population-based study', *Haematologica*, 105: 1621-29.

Roth, D. B. 2014. 'V(D)J Recombination: Mechanism, Errors, and Fidelity', *Microbiol Spectr*, 2.

Saito, H., A. Sekizawa, T. Morimoto, M. Suzuki, and T. Yanaihara. 2000. 'Prenatal DNA diagnosis of a single-gene disorder from maternal plasma', *Lancet*, 356: 1170.

Sakamoto, Y., S. Sereewattanawoot, and A. Suzuki. 2020. 'A new era of long-read sequencing for cancer genomics', *J Hum Genet*, 65: 3-10.

Samorodnitsky, E., B. M. Jewell, R. Hagopian, J. Miya, M. R. Wing, E. Lyon, S. Damodaran, D. Bhatt, J. W. Reeser, J. Datta, and S. Roychowdhury. 2015. 'Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing', *Hum Mutat*, 36: 903-14.

Scarfo, L., and P. Ghia. 2016. 'What does it mean I have a monoclonal B-cell lymphocytosis?: Recent insights and new challenges', *Semin Oncol*, 43: 201-8.

Scheinin, I., D. Sie, H. Bengtsson, M. A. van de Wiel, A. B. Olshen, H. F. van Thuijl, H. F. van Essen, P. P. Eijk, F. Rustenburg, G. A. Meijer, J. C. Reijneveld, P. Wesseling, D. Pinkel, D. G. Albertson, and B. Ylstra. 2014. 'DNA copy number analysis of fresh and

formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly', *Genome Res*, 24: 2022-32.

Scherer, F., D. M. Kurtz, M. Diehn, and A. A. Alizadeh. 2017. 'High-throughput sequencing for noninvasive disease detection in hematologic malignancies', *Blood*, 130: 440-52.

Scherer, F., D. M. Kurtz, A. M. Newman, H. Stehr, A. F. Craig, M. S. Esfahani, A. F. Lovejoy, J. J. Chabon, D. M. Klass, C. L. Liu, L. Zhou, C. Glover, B. C. Visser, G. A. Poultsides, R. H. Advani, L. S. Maeda, N. K. Gupta, R. Levy, R. S. Ohgami, C. A. Kunder, M. Diehn, and A. A. Alizadeh. 2016. 'Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA', *Sci Transl Med*, 8: 364ra155.

Schmitz, R., G. W. Wright, D. W. Huang, C. A. Johnson, J. D. Phelan, J. Q. Wang, S. Roulland, M. Kasbekar, R. M. Young, A. L. Shaffer, D. J. Hodson, W. Xiao, X. Yu, Y. Yang, H. Zhao, W. Xu, X. Liu, B. Zhou, W. Du, W. C. Chan, E. S. Jaffe, R. D. Gascoyne, J. M. Connors, E. Campo, A. Lopez-Guillermo, A. Rosenwald, G. Ott, J. Delabie, L. M. Rimsza, K. Tay Kuang Wei, A. D. Zelenetz, J. P. Leonard, N. L. Bartlett, B. Tran, J. Shetty, Y. Zhao, D. R. Soppet, S. Pittaluga, W. H. Wilson, and L. M. Staudt. 2018. 'Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma', *N Engl J Med*, 378: 1396-407.

Schuh, A. 2019. 'Evaluation and Transfer of mobile whole slide tissue scanners and liquid biopsies to deliver fast and precise diagnosis for improved outcomes of children and young adults with EBV-driven lymphoma: a joint Tanzanian and Ugandan Paediatric Oncology Network Initiative', NIHR, Accessed 17/09/2022.

———. 2022. 'International Workshop on CLL Global Partnerships Subcomittee', Accessed 17/09/2022. https://www.iwcll.org/about/subcommittee/global-partnerships-subcommittee/.

Schuh, A., J. Becq, S. Humphray, A. Alexa, A. Burns, R. Clifford, S. M. Feller, R. Grocock, S. Henderson, I. Khrebtukova, Z. Kingsbury, S. Luo, D. McBride, L. Murray, T. Menju, A. Timbs, M. Ross, J. Taylor, and D. Bentley. 2012. 'Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns', *Blood*, 120: 4191-6.

Schuh, A. H., N. Parry-Jones, N. Appleby, A. Bloor, C. E. Dearden, C. Fegan, G. Follows, C. P. Fox, S. Iyengar, B. Kennedy, H. McCarthy, H. M. Parry, P. Patten, A. R. Pettitt, I. Ringshausen, R. Walewska, and P. Hillmen. 2018. 'Guideline for the treatment of chronic lymphocytic leukaemia: A British Society for Haematology Guideline', *Br J Haematol*, 182: 344-59.

Schwarzenbach, H., D. S. Hoon, and K. Pantel. 2011. 'Cell-free nucleic acids as biomarkers in cancer patients', *Nat Rev Cancer*, 11: 426-37.

Sedlazeck, F. J., P. Rescheneder, M. Smolka, H. Fang, M. Nattestad, A. von Haeseler, and M. C. Schatz. 2018. 'Accurate detection of complex structural variations using single-molecule sequencing', *Nat Methods*, 15: 461-68.

Sereika, M., R. H. Kirkegaard, S. M. Karst, T. Y. Michaelsen, E. A. Sorensen, R. D. Wollenberg, and M. Albertsen. 2022. 'Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing', *Nat Methods*, 19: 823-26.

Sha, C., S. Barrans, F. Cucco, M. A. Bentley, M. A. Care, T. Cummin, H. Kennedy, J. S. Thompson, R. Uddin, L. Worrillow, R. Chalkley, M. van Hoppe, S. Ahmed, T. Maishman, J. Caddy, A. Schuh, C. Mamot, C. Burton, R. Tooze, A. Davies, M. Q. Du, P. W. M. Johnson, and D. R. Westhead. 2019. 'Molecular High-Grade B-Cell

Lymphoma: Defining a Poor-Risk Group That Requires Different Approaches to Therapy', *J Clin Oncol*, 37: 202-12.
Shanafelt, T. D., P. Ghia, M. C. Lanasa, O. Landgren, and A. C. Rawstron. 2010. 'Monoclonal B-cell lymphocytosis (MBL): biology, natural history and clinical management', *Leukemia*, 24: 512-20.
Shanafelt, T. D., N. E. Kay, S. A. Parikh, S. J. Achenbach, C. E. Lesnick, C. A. Hanson, G. Kleinstern, J. E. Olson, A. D. Norman, K. G. Rabe, S. M. Schwager, T. G. Call, and S. L. Slager. 2021. 'Risk of serious infection among individuals with and without low count monoclonal B-cell lymphocytosis (MBL)', *Leukemia*, 35: 239-44.
Shanafelt, T. D., N. E. Kay, K. G. Rabe, D. J. Inwards, C. S. Zent, J. F. Leis, S. M. Schwager, C. A. Thompson, D. A. Bowen, T. E. Witzig, S. L. Slager, and T. G. Call. 2012. 'Hematologist/oncologist disease-specific expertise and survival: lessons from chronic lymphocytic leukemia (CLL)/small lymphocytic lymphoma (SLL)', *Cancer*, 118: 1827-37.
Smolander, J., S. Khan, K. Singaravelu, L. Kauko, R. J. Lund, A. Laiho, and L. L. Elo. 2021. 'Evaluation of tools for identifying large copy number variations from ultra-low-coverage whole-genome sequencing data', *BMC Genomics*, 22: 357.
Snyder, M. W., M. Kircher, A. J. Hill, R. M. Daza, and J. Shendure. 2016. 'Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin', *Cell*, 164: 57-68.
Soilleux, E. J., A. Wotherspoon, T. A. Eyre, R. Clifford, M. Cabes, and A. H. Schuh. 2016. 'Diagnostic dilemmas of high-grade transformation (Richter's syndrome) of chronic lymphocytic leukaemia: results of the phase II National Cancer Research Institute CHOP-OR clinical trial specialist haemato-pathology central review', *Histopathology*, 69: 1066-76.
Sonneveld, P., H. Avet-Loiseau, S. Lonial, S. Usmani, D. Siegel, K. C. Anderson, W. J. Chng, P. Moreau, M. Attal, R. A. Kyle, J. Caers, J. Hillengass, J. San Miguel, N. W. van de Donk, H. Einsele, J. Blade, B. G. Durie, H. Goldschmidt, M. V. Mateos, A. Palumbo, and R. Orlowski. 2016. 'Treatment of multiple myeloma with high-risk cytogenetics: a consensus of the International Myeloma Working Group', *Blood*, 127: 2955-62.
Soussi, T., B. Leroy, and P. E. Taschner. 2014. 'Recommendations for analyzing and reporting TP53 gene variants in the high-throughput sequencing era', *Hum Mutat*, 35: 766-78.
Soverini, S., M. Baccarani, and G. Martinelli. 2016. 'Nilotinib against high dose imatinib for salvage therapy of chronic myeloid leukaemia', *Lancet Haematol*, 3: e554-e55.
Spence, T., S. Perera, J. Weiss, S. Grenier, L. Ranich, F. Shepherd, and T. L. Stockley. 2021. 'Clinical implementation of circulating tumour DNA testing for EGFR T790M for detection of treatment resistance in non-small cell lung cancer', *J Clin Pathol*, 74: 91-97.
Spina, V., A. Bruscaggin, A. Cuccaro, M. Martini, M. Di Trani, G. Forestieri, M. Manzoni, A. Condoluci, A. Arribas, L. Terzi-Di-Bergamo, S. L. Locatelli, E. Cupelli, L. Ceriani, A. A. Moccia, A. Stathis, L. Nassi, C. Deambrogi, F. Diop, F. Guidetti, A. Cocomazzi, S. Annunziata, V. Rufini, A. Giordano, A. Neri, R. Boldorini, B. Gerber, F. Bertoni, M. Ghielmini, G. Stussi, A. Santoro, F. Cavalli, E. Zucca, L. M. Larocca, G. Gaidano, S. Hohaus, C. Carlo-Stella, and D. Rossi. 2018. 'Circulating tumor DNA reveals genetics, clonal evolution, and residual disease in classical Hodgkin lymphoma', *Blood*, 131: 2413-25.
Stamatopoulos, B., T. Smith, E. Crompot, K. Pieters, R. Clifford, M. Mraz, P. Robbe, A. Burns, A. Timbs, D. Bruce, P. Hillmen, N. Meuleman, P. Mineur, R. Firescu, M.

Maerevoet, V. De Wilde, A. Efira, J. Philippe, B. Verhasselt, F. Offner, D. Sims, A. Heger, H. Dreau, and A. Schuh. 2018. 'The Light Chain IgLV3-21 Defines a New Poor Prognostic Subgroup in Chronic Lymphocytic Leukemia: Results of a Multicenter Study', *Clin Cancer Res*, 24: 5048-57.

Stamatopoulos, B., A. Timbs, D. Bruce, T. Smith, R. Clifford, P. Robbe, A. Burns, D. V. Vavoulis, L. Lopez, P. Antoniou, J. Mason, H. Dreau, and A. Schuh. 2017. 'Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia', *Leukemia*, 31: 837-45.

Stephens, D. M., Moseley, A., Hill, B. T., Pagel, J. M., Shadman, M., Fisch, M., Danilov, A.V., Ng, D., Mato, A.R., Danielle M. Brander, D. M., Othus, M., Coutre, S., O'Brien, S. M., Erba H. P.;. 2021. 'Randomized, Phase III Study of Early Intervention with Venetoclax and Obinutuzumab Versus Delayed Therapy with Venetoclax and Obinutuzumab in Newly Diagnosed Asymptomatic High-Risk Patients with Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma (CLL/SLL): Evolve CLL/SLL Study (SWOG S1925; NCT#04269902).', *Blood*, 138: 2630.

Stewart, J. P., J. Gazdova, N. Darzentas, D. Wren, P. Proszek, G. Fazio, S. Songia, M. Alcoceba, M. E. Sarasquete, P. Villarese, M. Y. van der Klift, K. C. Heezen, N. McCafferty, K. Pal, M. Catherwood, C. S. Kim, S. Srivastava, L. I. Kroeze, E. Hodges, K. Stamatopoulos, W. Klapper, E. Genuardi, S. Ferrero, M. van den Brand, G. Cazzaniga, F. Davi, L. A. Sutton, R. Garcia-Sanz, Pjta Groenen, E. A. Macintyre, M. Bruggemann, C. Pott, A. W. Langerak, D. Gonzalez, and N. G. S. Working Group EuroClonality. 2021. 'Validation of the EuroClonality-NGS DNA capture panel as an integrated genomic tool for lymphoproliferative disorders', *Blood Adv*, 5: 3188-98.

Strati, P., S. A. Parikh, K. G. Chaffee, N. E. Kay, T. G. Call, S. J. Achenbach, J. R. Cerhan, S. L. Slager, and T. D. Shanafelt. 2017. 'Relationship between co-morbidities at diagnosis, survival and ultimate cause of death in patients with chronic lymphocytic leukaemia (CLL): a prospective cohort study', *Br J Haematol*, 178: 394-402.

Sudha, P., A. Ahsan, C. Ashby, T. Kausar, A. Khera, M. H. Kazeroun, C. C. Hsu, L. Wang, E. Fitzsimons, O. Salminen, P. Blaney, M. Czader, J. Williams, M. I. Abu Zaid, N. Ansari-Pour, K. L. Yong, F. van Rhee, W. E. Pierceall, G. J. Morgan, E. Flynt, S. Gooding, R. Abonour, K. Ramasamy, A. Thakurta, and B. A. Walker. 2022. 'Myeloma Genome Project Panel is a Comprehensive Targeted Genomics Panel for Molecular Profiling of Patients with Multiple Myeloma', *Clin Cancer Res*, 28: 2854-64.

Sutton, L. A., V. Ljungstrom, A. Enjuanes, D. Cortese, A. Skaftason, E. Tausch, K. Stano Kozubik, F. Nadeu, M. Armand, J. Malcikova, T. Pandzic, J. Forster, Z. Davis, D. Oscier, D. Rossi, P. Ghia, J. C. Strefford, S. Pospisilova, S. Stilgenbauer, F. Davi, E. Campo, K. Stamatopoulos, R. Rosenquist, and Eric On Behalf Of The European Research Initiative On Cll. 2021. 'Comparative analysis of targeted next-generation sequencing panels for the detection of gene mutations in chronic lymphocytic leukemia: an ERIC multi-center study', *Haematologica*, 106: 682-91.

Swerdlow, S. H., E. Campo, S. A. Pileri, N. L. Harris, H. Stein, R. Siebert, R. Advani, M. Ghielmini, G. A. Salles, A. D. Zelenetz, and E. S. Jaffe. 2016. 'The 2016 revision of the World Health Organization classification of lymphoid neoplasms', *Blood*, 127: 2375-90.

Tang, B. 2017. 'Clinical and demographic characteristics of CLL patients receiving first-line bendamustine plus rituximab using the Flatiron Health database', *Journal of Clinical Oncology*, 35.

Teipel, R., F. Kroschinsky, M. Kramer, T. Kretschmann, K. Egger-Heidrich, T. Kruger, L. Ruhnke, S. Herold, S. Stasik, K. Sockel, J. M. Middeke, K. Trautmann-Grill, M.

Bornhauser, C. Thiede, and M. von Bonin. 2022. 'Prevalence and variation of CHIP in patients with aggressive lymphomas undergoing CD19-directed CAR T-cell treatment', *Blood Adv*, 6: 1941-46.

Tejaswi, V., D. P. Lad, N. Jindal, G. Prakash, P. Malhotra, A. Khadwal, A. Jain, S. Sreedharanunni, M. S. Sachdeva, S. Naseem, N. Varma, and S. Varma. 2020. 'Chronic Lymphocytic Leukemia: Real-World Data From India', *JCO Glob Oncol*, 6: 866-72.

Thompson, P. A., J. Srivastava, C. Peterson, P. Strati, J. L. Jorgensen, T. Hether, M. J. Keating, S. M. O'Brien, A. Ferrajoli, J. A. Burger, Z. Estrov, N. Jain, and W. G. Wierda. 2019. 'Minimal residual disease undetectable by next-generation sequencing predicts improved outcome in CLL after chemoimmunotherapy', *Blood*, 134: 1951-59.

Tilly, H., M. Gomes da Silva, U. Vitolo, A. Jack, M. Meignan, A. Lopez-Guillermo, J. Walewski, M. Andre, P. W. Johnson, M. Pfreundschuh, M. Ladetto, and Esmo Guidelines Committee. 2015. 'Diffuse large B-cell lymphoma (DLBCL): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up', *Ann Oncol*, 26 Suppl 5: v116-25.

Tilly, H., F. Morschhauser, L. H. Sehn, J. W. Friedberg, M. Trneny, J. P. Sharman, C. Herbaux, J. M. Burke, M. Matasar, S. Rai, K. Izutsu, N. Mehta-Shah, L. Oberic, A. Chauchet, W. Jurczak, Y. Song, R. Greil, L. Mykhalska, J. M. Bergua-Burgues, M. C. Cheung, A. Pinto, H. J. Shin, G. Hapgood, E. Munhoz, P. Abrisqueta, J. P. Gau, J. Hirata, Y. Jiang, M. Yan, C. Lee, C. R. Flowers, and G. Salles. 2022. 'Polatuzumab Vedotin in Previously Untreated Diffuse Large B-Cell Lymphoma', *N Engl J Med*, 386: 351-63.

Trotman, J., and S. F. Barrington. 2021. 'The role of PET in first-line treatment of Hodgkin lymphoma', *Lancet Haematol*, 8: e67-e79.

Trummer, A., W. Schier, J. Krauter, H. Hannig, and J. Christmann. 2022. 'Circulating versus cellular tumor DNA for the detection of BTK resistant CLL clones', *Leuk Res Rep*, 18: 100359.

Tsang, M., T. D. Shanafelt, T. G. Call, W. Ding, A. Chanan-Khan, J. F. Leis, G. S. Nowakowski, D. Bowen, M. Conte, S. M. Schwager, S. L. Slager, N. E. Kay, C. A. Hanson, and S. A. Parikh. 2015. 'The efficacy of ibrutinib in the treatment of Richter syndrome', *Blood*, 125: 1676-8.

Tsimberidou, A. M., and M. J. Keating. 2005. 'Richter syndrome: biology, incidence, and therapeutic strategies', *Cancer*, 103: 216-28.

Tsimberidou, A. M., W. G. Wierda, W. Plunkett, R. Kurzrock, S. O'Brien, S. Wen, A. Ferrajoli, F. Ravandi-Kashani, G. Garcia-Manero, Z. Estrov, T. J. Kipps, J. R. Brown, A. Fiorentino, S. Lerner, H. M. Kantarjian, and M. J. Keating. 2008. 'Phase I-II study of oxaliplatin, fludarabine, cytarabine, and rituximab combination therapy in patients with Richter's syndrome or fludarabine-refractory chronic lymphocytic leukemia', *J Clin Oncol*, 26: 196-203.

Tsimberidou, A. M., W. G. Wierda, S. Wen, W. Plunkett, S. O'Brien, T. J. Kipps, J. A. Jones, X. Badoux, H. Kantarjian, M. J. Keating, and Consortium Chronic Lymphocytic Leukemia Research. 2013. 'Phase I-II clinical trial of oxaliplatin, fludarabine, cytarabine, and rituximab therapy in aggressive relapsed/refractory chronic lymphocytic leukemia or Richter syndrome', *Clin Lymphoma Myeloma Leuk*, 13: 568-74.

Valle-Inclan, J. E., C. Stangl, A. C. de Jong, L. F. van Dessel, M. J. van Roosmalen, J. C. A. Helmijr, I. Renkens, R. Janssen, S. de Blank, C. J. de Witte, J. W. M. Martens, Mphm Jansen, M. P. Lolkema, and W. P. Kloosterman. 2021. 'Optimizing Nanopore

sequencing-based detection of structural variants enables individualized circulating tumor DNA-based disease monitoring in cancer patients', *Genome Med*, 13: 86.

van der Pol, Y., Tantyon, N.A.,. 2022. 'Real-time analysis of the cancer genome and fragmentome from plasma and urine short and long cell-free DNA using Nanopore sequencing', *medRxiv*.

van der Velden, V. H. J., M. Bruggemann, G. Cazzaniga, B. Scheijen, B. Tops, J. Trka, K. Pal, S. Hanzelmann, G. Fazio, S. Songia, A. W. Langerak, N. Darzentas, EuroMrd, and N. G. S. Working Group EuroClonality. 2021. 'Potential and pitfalls of whole transcriptome-based immunogenetic marker identification in acute lymphoblastic leukemia; a EuroMRD and EuroClonality-NGS Working Group study', *Leukemia*, 35: 924-28.

van Dongen, J. J., A. W. Langerak, M. Bruggemann, P. A. Evans, M. Hummel, F. L. Lavender, E. Delabesse, F. Davi, E. Schuuring, R. Garcia-Sanz, J. H. van Krieken, J. Droese, D. Gonzalez, C. Bastard, H. E. White, M. Spaargaren, M. Gonzalez, A. Parreira, J. L. Smith, G. J. Morgan, M. Kneba, and E. A. Macintyre. 2003. 'Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936', *Leukemia*, 17: 2257-317.

Van Roy, N., M. Van Der Linden, B. Menten, A. Dheedene, C. Vandeputte, J. Van Dorpe, G. Laureys, M. Renard, T. Sante, T. Lammens, B. De Wilde, F. Speleman, and K. De Preter. 2017. 'Shallow Whole Genome Sequencing on Circulating Cell-Free DNA Allows Reliable Noninvasive Copy-Number Profiling in Neuroblastoma Patients', *Clin Cancer Res*, 23: 6305-14.

Vardi, A., A. Agathangelidis, L. A. Sutton, P. Ghia, R. Rosenquist, and K. Stamatopoulos. 2014. 'Immunogenetic studies of chronic lymphocytic leukemia: revelations and speculations about ontogeny and clinical evolution', *Cancer Res*, 74: 4211-6.

Vermaat, J. S., S. T. Pals, A. Younes, M. Dreyling, M. Federico, I. Aurer, J. Radford, M. J. Kersten, and a Scientific Working Group of the European Hematology Association Eha Lymphoma Group. 2015. 'Precision medicine in diffuse large B-cell lymphoma: hitting the target', *Haematologica*, 100: 989-93.

Visentin, A., L. Bonaldi, G. M. Rigolin, F. R. Mauro, A. Martines, F. Frezzato, S. Pravato, L. R. Gargarella, M. A. Bardi, M. Cavallari, E. Volta, F. Cavazzini, M. Nanni, M. Facco, F. Piazza, A. Guarini, R. Foa, G. Semenzato, A. Cuneo, and L. Trentin. 2022. 'The complex karyotype landscape in chronic lymphocytic leukemia allows the refinement of the risk of Richter syndrome transformation', *Haematologica*, 107: 868-76.

Vlachonikola, E., E. Sofou, A. Chatzidimitriou, K. Stamatopoulos, and A. Agathangelidis. 2021. 'The Significance of B-cell Receptor Stereotypy in Chronic Lymphocytic Leukemia: Biological and Clinical Implications', *Hematol Oncol Clin North Am*, 35: 687-702.

Volden, R., T. Palmer, A. Byrne, C. Cole, R. J. Schmitz, R. E. Green, and C. Vollmers. 2018. 'Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA', *Proc Natl Acad Sci U S A*, 115: 9726-31.

Wan, J. C. M., C. Massie, J. Garcia-Corbacho, F. Mouliere, J. D. Brenton, C. Caldas, S. Pacey, R. Baird, and N. Rosenfeld. 2017. 'Liquid biopsies come of age: towards implementation of circulating tumour DNA', *Nat Rev Cancer*, 17: 223-38.

Wan, J. C. M., D. Stephens, L. Luo, J. R. White, C. M. Stewart, B. Rousseau, D. W. Y. Tsui, and L. A. Diaz, Jr. 2022. 'Genome-wide mutational signatures in low-coverage whole genome sequencing of cell-free DNA', *Nat Commun*, 13: 4953.

Wang, K., M. Li, and H. Hakonarson. 2010. 'ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data', *Nucleic Acids Res*, 38: e164.
Wang, L., M. S. Lawrence, Y. Wan, P. Stojanov, C. Sougnez, K. Stevenson, L. Werner, A. Sivachenko, D. S. DeLuca, L. Zhang, W. Zhang, A. R. Vartanov, S. M. Fernandes, N. R. Goldstein, E. G. Folco, K. Cibulskis, B. Tesar, Q. L. Sievers, E. Shefler, S. Gabriel, N. Hacohen, R. Reed, M. Meyerson, T. R. Golub, E. S. Lander, D. Neuberg, J. R. Brown, G. Getz, and C. J. Wu. 2011. 'SF3B1 and other novel cancer genes in chronic lymphocytic leukemia', *N Engl J Med*, 365: 2497-506.
Wang, Y., S. J. Achenbach, K. G. Rabe, T. D. Shanafelt, T. G. Call, W. Ding, S. S. Kenderian, E. Muchtar, J. F. Leis, A. B. Koehler, S. M. Schwager, J. R. Cerhan, S. L. Slager, N. E. Kay, and S. A. Parikh. 2021. 'Cause of death in patients with newly diagnosed chronic lymphocytic leukemia (CLL) stratified by the CLL-International Prognostic Index', *Blood Cancer J*, 11: 140.
Watanabe, J., M. Natsumeda, M. Okada, D. Kobayashi, Y. Kanemaru, Y. Tsukamoto, M. Oishi, A. Kakita, and Y. Fujii. 2019. 'High Detection Rate of MYD88 Mutations in Cerebrospinal Fluid From Patients With CNS Lymphomas', *JCO Precis Oncol*, 3: 1-13.
Wight, J. C., G. Chong, A. P. Grigg, and E. A. Hawkes. 2018. 'Prognostication of diffuse large B-cell lymphoma in the molecular era: moving beyond the IPI', *Blood Rev*, 32: 400-15.
Wilson, M. L., K. A. Fleming, M. A. Kuti, L. M. Looi, N. Lago, and K. Ru. 2018. 'Access to pathology and laboratory medicine services: a crucial gap', *Lancet*, 391: 1927-38.
Wilson, W. H., G. W. Wright, D. W. Huang, B. Hodkinson, S. Balasubramanian, Y. Fan, J. Vermeulen, M. Shreeve, and L. M. Staudt. 2021. 'Effect of ibrutinib with R-CHOP chemotherapy in genetic subtypes of DLBCL', *Cancer Cell*, 39: 1643-53 e3.
Woyach, J. A., A. S. Ruppert, D. Guinn, A. Lehman, J. S. Blachly, A. Lozanski, N. A. Heerema, W. Zhao, J. Coleman, D. Jones, L. Abruzzo, A. Gordon, R. Mantel, L. L. Smith, S. McWhorter, M. Davis, T. J. Doong, F. Ny, M. Lucas, W. Chase, J. A. Jones, J. M. Flynn, K. Maddocks, K. Rogers, S. Jaglowski, L. A. Andritsos, F. T. Awan, K. A. Blum, M. R. Grever, G. Lozanski, A. J. Johnson, and J. C. Byrd. 2017. 'BTK(C481S)-Mediated Resistance to Ibrutinib in Chronic Lymphocytic Leukemia', *J Clin Oncol*, 35: 1437-43.
Xia, Q., S. Li, T. Ding, Z. Liu, J. Liu, Y. Li, H. Zhu, and Z. Yao. 2023. 'Nanopore sequencing for detecting reciprocal translocation carrier status in preimplantation genetic testing', *BMC Genomics*, 24: 1.
Xie, M., C. Lu, J. Wang, M. D. McLellan, K. J. Johnson, M. C. Wendl, J. F. McMichael, H. K. Schmidt, V. Yellapantula, C. A. Miller, B. A. Ozenberger, J. S. Welch, D. C. Link, M. J. Walter, E. R. Mardis, J. F. Dipersio, F. Chen, R. K. Wilson, T. J. Ley, and L. Ding. 2014. 'Age-related mutations associated with clonal hematopoietic expansion and malignancies', *Nat Med*, 20: 1472-8.
Xu-Monette, Z. Y., L. Wei, X. Fang, Q. Au, H. Nunns, M. Nagy, A. Tzankov, F. Zhu, C. Visco, G. Bhagat, K. Dybkaer, A. Chiu, W. Tam, Y. Zu, E. D. Hsi, F. B. Hagemeister, X. Sun, X. Han, H. Go, M. Ponzoni, A. J. M. Ferreri, M. B. Moller, B. M. Parsons, J. H. van Krieken, M. A. Piris, J. N. Winter, Y. Li, B. Xu, M. Albitar, H. You, and K. H. Young. 2022. 'Genetic Subtyping and Phenotypic Characterization of the Immune Microenvironment and MYC/BCL2 Double Expression Reveal Heterogeneity in Diffuse Large B-cell Lymphoma', *Clin Cancer Res*, 28: 972-83.
Xu-Monette, Z. Y., L. Wu, C. Visco, Y. C. Tai, A. Tzankov, W. M. Liu, S. Montes-Moreno, K. Dybkaer, A. Chiu, A. Orazi, Y. Zu, G. Bhagat, K. L. Richards, E. D. Hsi, X. F. Zhao, W. W. Choi, X. Zhao, J. H. van Krieken, Q. Huang, J. Huh, W. Ai, M. Ponzoni,

A. J. Ferreri, F. Zhou, B. S. Kahl, J. N. Winter, W. Xu, J. Li, R. S. Go, Y. Li, M. A. Piris, M. B. Moller, R. N. Miranda, L. V. Abruzzo, L. J. Medeiros, and K. H. Young. 2012. 'Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: report from an International DLBCL Rituximab-CHOP Consortium Program Study', *Blood*, 120: 3986-96.

Xu, L., N. Tsakmaklis, G. Yang, J. G. Chen, X. Liu, M. Demos, A. Kofides, C. J. Patterson, K. Meid, J. Gustine, T. Dubeau, M. L. Palomba, R. Advani, J. J. Castillo, R. R. Furman, Z. R. Hunter, and S. P. Treon. 2017. 'Acquired mutations associated with ibrutinib resistance in Waldenstrom macroglobulinemia', *Blood*, 129: 2519-25.

Yeh, P., M. Dickinson, S. Ftouni, T. Hunter, D. Sinha, S. Q. Wong, R. Agarwal, R. Vedururu, K. Doig, C. Y. Fong, P. Blombery, D. Westerman, M. A. Dawson, and S. J. Dawson. 2017. 'Molecular disease monitoring using circulating tumor DNA in myelodysplastic syndromes', *Blood*, 129: 1685-90.

Yeh, P., T. Hunter, D. Sinha, S. Ftouni, E. Wallach, D. Jiang, Y. C. Chan, S. Q. Wong, M. J. Silva, R. Vedururu, K. Doig, E. Lam, G. M. Arnau, T. Semple, M. Wall, A. Zivanovic, R. Agarwal, P. Petrone, K. Jones, D. Westerman, P. Blombery, J. F. Seymour, A. T. Papenfuss, M. A. Dawson, C. S. Tam, and S. J. Dawson. 2017. 'Circulating tumour DNA reflects treatment response and clonal evolution in chronic lymphocytic leukaemia', *Nat Commun*, 8: 14756.

Younes, A., J. Brody, C. Carpio, A. Lopez-Guillermo, D. Ben-Yehuda, B. Ferhanoglu, A. Nagler, M. Ozcan, I. Avivi, F. Bosch, M. D. Caballero Barrigon, A. Hellmann, B. Kuss, D. D. F. Ma, F. Demirkan, M. Yagci, N. A. Horowitz, P. Marlton, R. Cordoba, T. Wrobel, D. Buglio, M. Streit, B. P. Hodkinson, M. Schaffer, J. Alvarez, R. Ceulemans, S. Balasubramanian, J. de Jong, S. S. Wang, N. Fourneau, and W. Jurczak. 2019. 'Safety and activity of ibrutinib in combination with nivolumab in patients with relapsed non-Hodgkin lymphoma or chronic lymphocytic leukaemia: a phase 1/2a study', *Lancet Haematol*, 6: e67-e78.

Younes, A., L. H. Sehn, P. Johnson, P. L. Zinzani, X. Hong, J. Zhu, C. Patti, D. Belada, O. Samoilova, C. Suh, S. Leppa, S. Rai, M. Turgut, W. Jurczak, M. C. Cheung, R. Gurion, S. P. Yeh, A. Lopez-Hernandez, U. Duhrsen, C. Thieblemont, C. S. Chiattone, S. Balasubramanian, J. Carey, G. Liu, S. M. Shreeve, S. Sun, S. H. Zhuang, J. Vermeulen, L. M. Staudt, W. Wilson, and Phoenix investigators. 2019. 'Randomized Phase III Trial of Ibrutinib and Rituximab Plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone in Non-Germinal Center B-Cell Diffuse Large B-Cell Lymphoma', *J Clin Oncol*, 37: 1285-95.

Zenz, T., B. Eichhorst, R. Busch, T. Denzel, S. Habe, D. Winkler, A. Buhler, J. Edelmann, M. Bergmann, G. Hopfinger, M. Hensel, M. Hallek, H. Dohner, and S. Stilgenbauer. 2010. 'TP53 mutation and survival in chronic lymphocytic leukemia', *J Clin Oncol*, 28: 4473-9.

Zenz, T., A. Krober, K. Scherer, S. Habe, A. Buhler, A. Benner, T. Denzel, D. Winkler, J. Edelmann, C. Schwanen, H. Dohner, and S. Stilgenbauer. 2008. 'Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia: results from a detailed genetic characterization with long-term follow-up', *Blood*, 112: 3322-9.

Zimonjic, D. B., C. Keck-Waggoner, and N. C. Popescu. 2001. 'Novel genomic imbalances and chromosome translocations involving c-myc gene in Burkitt's lymphoma', *Leukemia*, 15: 1582-8.

Zviran, A., R. C. Schulman, M. Shah, S. T. K. Hill, S. Deochand, C. C. Khamnei, D. Maloney, K. Patel, W. Liao, A. J. Widman, P. Wong, M. K. Callahan, G. Ha, S. Reed, D. Rotem, D. Frederick, T. Sharova, B. Miao, T. Kim, G. Gydush, J. Rhoades, K. Y.

Huang, N. D. Omans, P. O. Bolan, A. H. Lipsky, C. Ang, M. Malbari, C. F. Spinelli, S. Kazancioglu, A. M. Runnels, S. Fennessey, C. Stolte, F. Gaiti, G. G. Inghirami, V. Adalsteinsson, B. Houck-Loomis, J. Ishii, J. D. Wolchok, G. Boland, N. Robine, N. K. Altorki, and D. A. Landau. 2020. 'Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring', *Nat Med*, 26: 1114-24.

Appendix A: OXPLORED synopsis and statistical considerations



Full title: Oxford Pre-cancerous Lymphoproliferative Disorders: Analysis and Interception study

Short title: Oxford Pre-cancerous Lymphoproliferative Disorders study

Protocol Version & date:	V5.2_09Sep2022
Sponsor Protocol Number:	OCTO_091
REC Reference:	19/SC/0065
Clintrials.gov. No.	NCT04023747
IRAS Project ID:	242634

OXFORD Molecular Diagnostics Centre

Transforming patient healthcare, pioneering tomorrow's technology

octo Oncology Clinical Trials Office

Sponsored by the University of Oxford

Full Title of study:	Oxford Pre-cancerous Lymphoproliferative Disorders: Analysis and Interception study
Short Title:	Oxford Pre-cancerous Lymphoproliferative Disorders study
Study Acronym:	OxPLoreD
Objectives:	The primary objective of the study is to identify the clinical, genomic and immunological predictive markers of progression to malignant disease Secondary objectives of the study include: • To assess Quality of Life of patients with pre-cancerous lymphoid conditions • To study other clinical complications associated with pre-malignancy (infection, bone fractures, thrombosis and deaths not due to progression to lymphoid cancers) • To establish evidence-based guidance for monitoring and follow-up of patients with pre-cancerous conditions Exploratory objectives of OxPLoreD study include: • To identify novel surrogate markers of cure • To gather feasibility data in this subgroup at high-risk of progression to inform a future early intervention trial
Scientific rationale:	Identification and characterisation of the biological and genomic events that determine evolution to malignancy are essential to understanding the mechanisms of disease progression.
Clinical rationale:	Mature B-cell malignancies are preceded by preclinical clonal B-cell disorders, can have asymptomatic early-stage disease that is clinically managed with a 'watch-and-wait' approach, and can progress to clinically aggressive, incurable cancer. There is no consensus regarding optimal monitoring strategy for pre-cancerous lymphoproliferative disorders and practice varies widely. There is a paucity of data on quality of life in pre-cancerous lymphoproliferative disorders. Molecular analysis of patients with early lymphoproliferative disorder could be used to identify at-risk patients and suggest novel therapeutic strategies for these patients.
Study Design:	OxPLoreD is an observational cohort study
Participant Numbers:	Up to 1650 participants
Target Population:	The study targets 3 cohorts of early lymphoproliferative disorders: 1. Monoclonal B-cell lymphocytosis and asymptomatic Chronic Lymphocytic Leukaemia 2. IgM Monoclonal gammopathy and asymptomatic Waldenström's macroglobulinaemia 3. IgA or IgG Monoclonal gammopathy and smouldering myeloma

Inclusion and exclusion criteria	**Inclusion Criteria:** 1. Patients diagnosed within the previous three years with one of the following: a. High count monoclonal B-cell lymphocytosis (MBL) i.e. clonal B-cell population 0.5-4.9 10^9/L, b. Rai Stage 0-2/ Binet Stage A or B Chronic Lymphocytic Leukaemia not meeting the IWCLL criteria for treatment c. IgG or IgA Monoclonal Gammopathy of Uncertain Significance (see detailed eligibility criteria for these participants) d. IgM Monoclonal Gammopathy of Uncertain Significance (see detailed eligibility criteria for these participants) e. Asymptomatic smouldering Waldenström's Macroglobulinaemia not meeting the criteria for treatment f. Smouldering myeloma not meeting the criteria for treatment 2. ECOG performance status of 0, 1 or 2 3. Age 16 years and over 4. Signed written informed consent 5. The patient is willing and able to comply with the protocol for the duration of the study, and scheduled follow-up visits and examinations. **Exclusion Criteria:** 1. Pregnant or breast-feeding women. Pregnant or breast-feeding women may be re-screened following delivery and/or cessation of breastfeeding, as appropriate. 2. Previous chemotherapy or immunotherapy for any haematological cancers 3. Treatment with any other investigational agent, or participation in an interventional clinical trial within 28 days prior to enrolment. 4. Patients in cohort 2 or 3 on anticoagulation for a diagnosis of pulmonary embolus or deep vein thrombosis within the last 3 months or with a mechanical heart valve or any other condition causing a significant risk of thromboembolism. Participants who are anticoagulated for atrial fibrillation are eligible but will be asked to interrupt anticoagulation 3 days prior to bone marrow examination. 5. Other psychological, social or medical condition, physical examination finding or a laboratory abnormality that the Investigator considers would make the patient a poor study candidate or could interfere with protocol compliance or the interpretation of study results. 6. Any other malignancy that requires active surgical or chemotherapeutic treatment. Patients on long term hormone therapies (e.g. tamoxifen) are permitted to enrol at the discretion of the investigator, after considering the overall clinical context. 7. Any significant concurrent medical condition resulting in a life-expectancy of less than 5 years (including but not limited to renal, hepatic, haematological, gastrointestinal, endocrine, pulmonary, neurological, cerebral or psychiatric disease) 8. For cohort 3: Any contraindication for MRI- presence of any metallic foreign body, eGFR <30 and allergy to gadolinium contrast.
Duration on study:	Each participant will be on study for up to 60 months (5 years). Duration of the whole study will be 84 months (expected to be 5 years recruiting followed by 2 years follow-up of the last recruited participants).

Study Procedures and frequency:	Participants will undergo additional research radiological investigations (CT or MRI) if they consent and research blood and germline DNA samples will also be collected and quality of life (QoL) questionnaires completed. Additional bone marrow aspirate samples will also be collected at enrolment, if participants consent. Research bone marrow sampling will be offered to all participant cohorts at the Churchill Hospital site but only to cohort 1 participants at all other sites. Participants will have up to 2 scheduled follow-up visits (at 24 and 60 months) at which further blood samples and QoL questionnaires will be collected. If at any time during the study a participant experiences progression of their disease a disease progression visit will be arranged, to include repeat imaging and repeat bone marrow aspirate sample (if they consent), blood samples and QoL questionnaires. Imaging and bone marrow aspirate are standard of care procedures for some participants at enrolment and/or progression, but for other participants these will be optional research procedures. Where participants have imaging as standard of care the study team will, where possible, avoid an additional procedure for the participant by obtaining a copy of the standard of care imaging report.
Participant care post-study:	Participants with non-progressive early lymphoproliferative pre-malignancy will return to standard NHS care for patients at the conclusion of the study. This may be either hospital, telephone clinic or GP follow-up in line with local NHS practice. Participants in whom disease has progressed will return to standard NHS follow-up for patients with CLL, WM or MM.
No. of Study Site(s)	Multicentre. Recruitment will be encouraged through advertising the study via patient support groups and the NCRN clinical study groups.
End of study	When the last participant has either progressed or completed their 24-month follow-up visit. Participants will be on study for a maximum of 60 months.
Publication policy	The results of this study will be published in a peer-reviewed scientific journal and other academic research purposes as agreed by the Investigators.

Statistical considerations for the OXPLORED study

Sample size and power

All statistical analyses will be exploratory, and results will require validation in an independent cohort.

We will examine a number of known candidate predictors of early progression (see below). It is likely that none of these will have a predictive value that is sufficiently high on its own. We will therefore combine different predictors into a risk score for each group (MGUS/SM and MBL/early CLL).

For the power calculations presented below, we assume 1650 participants, 50% in each group. We assume that approximately 10% of patients in each group will progress to myeloma and CLL, respectively, and we wish to find associations between specific predictors and the progressor/non-progressor phenotype. We have also accounted for a 20% loss to follow-up.

The simplest approach will be to apply a χ^2 test on a 2×K contingency matrix, where 2 stands for the progressor/non-progressor phenotype, and K is the number of levels in each categorical predictor, as follows (in all cases, a significance level of 5% and two-sided tests are assumed):

Examples of individual candidate predictors are:

MGUS/SM:

1) **MYC aberration (present or absent):** We apply a χ^2 test with 1 degree of freedom on a 2×2 contingency table.
2) **MGUS prognostic score (0, 1, 2 or 3) based on clinical features:** We apply a χ^2 test with 3 degrees of freedom on a 2×4 contingency table.
3) **Number of genes on the target panel affected by a mutation (grouped as 0, 1 or ≥2):** We apply a χ^2 test with 2 degrees of freedom on a 2×3 contingency table.

Similarly**,** for the MBL/early CLL cases, we have:

1) **Absolute clonal B-cell count (continuous variable):** We can apply a two-sided t-test
2) **Expression of CD38 by immunophenotyping (present or absent):** Again, we can apply a χ^2 test with 1 degree of freedom on a 2×2 contingency table
3) **The CLL-IPI score (patients can score 1-8) based on age, clinical stage, IGHV mutation status, B2MG & TP53 status:** A χ^2 test with 7 degrees of freedom on a 2×8 contingency table.
4) **Number of genes on the targeted panel affected by a mutation (grouped as 0, 1, ≥2):** As above, we can apply a χ^2 test with 2 degrees of freedom on a 2×3 contingency table.
5) **Number of IGHV clones detectable by NGS (grouped in 5 categories):** Again, a χ^2 test with 4 degrees of freedom on a 2x5 contingency table.
6) On a subset of the patients, CT and MRI scan reports will also be available and could be integrated into this risk model.

Using these individual candidate predictors, the power to detect small effect sizes is above 30% and close to 100% for medium to large effect sizes for all candidate predictors. Importantly, by combining the different candidate predictors into a common risk score, we expect to significantly increase the power to predict small effect sizes from each individual predictor alone.

Appendix B: STELLAR trial protocol synopsis

STELLAR

A phase II, randomi**S**ed study of CHOP-R in combination with acalabru**T**inib compar**E**d to CHOP-R in patients with new**L**y diagnosed Richter's Syndrome and a p**LA**tfo**R**m for initial investigations into activity of novel treatments in relapsed/refractory and newly diagnosed Richter's Syndrome.

Version 5.0 16-Mar-2021

Chief Investigator:	Professor Anna Schuh
Sponsor:	University of Birmingham
EudraCT Number:	2017-004401-40
Sponsor Number:	RG_17-194
CAS Number:	HM2063
ISCRTN	ISRCTN52839057

This application is supported by the facilities funded through Birmingham Science City: Translational Medicine Clinical Research Infrastructure and Trials Platform, an Advantage West Midlands (AWM) *funded project which forms part of the Science City University of Warwick and University of Birmingham Research Alliance*

4. TRIAL SYNOPSIS

1. Title

A phase II, randomiSed study of CHOP-R in combination with acalabruTinib comparEd to CHOP-R in patients with newLy diagnosed Richter's Syndrome (RS) and a pLAtfoRm for initial investigations into activity of novel treatments in relapsed/refractory and newly diagnosed RS.

2. Trial Design

This is a prospective multicentre trial in patients in the UK. It has two elements: a randomised trial component and single-arm platform studies for initial, single-arm activity testing of novel agents (or novel combinations).

Randomised trial component

This is a prospective, phase II, randomised multicentre trial in patients in the UK with CLL and newly diagnosed diffuse large B-cell lymphoma (DLBCL)-type RS who are anthracycline-naïve.
Prior exposure to Bruton tyrosine kinase (BTK), phosphoinositide-3-kinase (PI3K), or BCL-2 inhibitors is permitted, with the caveat that patients who developed RS whilst on ibrutinib treatment (defined as diagnosis of RS within four weeks of last dose of ibrutinib) are **not eligible** for the **randomised component** of the STELLAR trial.

Ibrutinib-exposed CLL patients who (for example) discontinued ibrutinib due to adverse effects or progressive CLL, who then receive another CLL-directed therapy and later (more than four weeks) develop RS are not excluded from the randomised trial (see Section 4).

Patients with newly diagnosed RS who developed RS within four weeks of last dose of ibrutinib treatment may be eligible for recruitment into Cohort 2 of the trial (see below).

A total of 60 eligible participants will be 1:1 randomised between the current standard of care for RS, CHOP-R (cyclophosphamide, doxorubicin, vincristine, prednisolone and rituximab), and CHOP-R administered in combination with acalabrutinib. Participants will be treated with up to 6 cycles of CHOP-R, with assessment after 4 and 6 cycles of treatment, and followed-up for a minimum of 2 years. The primary outcome for assessment of activity for the randomised component of the STELLAR trial is progression-free survival (PFS).

Single-arm platform studies

In addition to the randomised component, this trial will act as a platform for initial, single-arm activity testing of novel agents (or novel combinations) in two cohorts of patients with RS who have progressed on, or are not eligible for, the main randomised trial component:

- Cohort 1 (up to 30 patients): Patients with RS in whom the disease has progressed following combination chemo-immunotherapy with anthracycline-based chemotherapy and anti-CD20 monoclonal antibody.

- Cohort 2 (up to 15 patients): Patients with newly diagnosed RS, which has developed whilst on ibrutinib treatment (defined as diagnosis of RS within four weeks of last dose of ibrutinib).

The purpose of these cohort investigations is to collect initial activity information to support development of the novel treatment(s) tested in the cohorts. For this reason, sample sizes for each cohort and novel treatment will depend on the planned development of the agent. The primary outcome for activity will be overall response rate (ORR) with Bayesian methods used for the main analyses of ORR.

In this protocol, reference to Cohort 1 and Cohort 2 refers to these cohorts treated with acalabrutinib monotherapy and CHOP-R in combination with acalabrutinib, respectively. Whilst the randomised trial is open to recruitment, if promising new treatments are identified for treatment of RS at a given tolerated dose level, these can be considered for further single-arm trials in Cohort 1 and/or Cohort 2 populations using the platform provided by this trial. The protocol and corresponding documents will be updated accordingly.

Cohort 1 (progressive RS following chemo-immunotherapy)

The first novel agent which will be tested in Cohort 1 (progressive RS following chemo-immunotherapy) is acalabrutinib monotherapy. Up to 30 participants are expected to be recruited into this cohort for treatment with acalabrutinib.

All participants in the STELLAR standard-of-care arm of the randomised study in whom RS has progressed or relapsed on CHOP-R, and who meet the eligibility criteria for Cohort 1, will be offered treatment with acalabrutinib in Cohort 1.

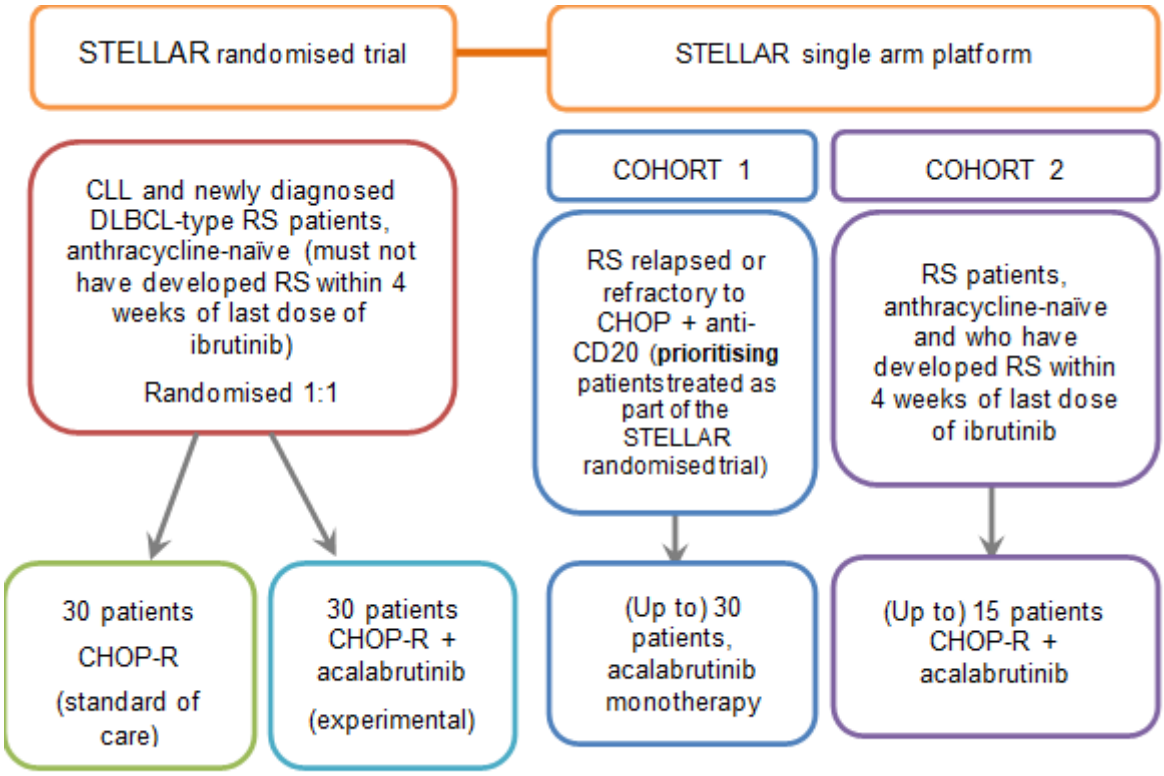
It is anticipated that of the 30 randomised trial participants allocated to the standard of care (CHOP-R) treatment, up to 21 patients will relapse or progress during the study.

If fewer than 21 of the 30 participants continue from the STELLAR randomised trial into Cohort 1 for treatment with acalabrutinib, then recruitment may be opened to suitable patients in whom RS has progressed or relapsed following combination chemo-immunotherapy with anthracycline-based chemotherapy and anti-CD20 monoclonal antibody outside of the randomised component of this trial.

Cohort 2 (anthracycline-naïve RS patients, diagnosed while on ibrutinib)

The first novel treatment to be tested in Cohort 2 (anthracycline-naïve RS patients, diagnosed while on ibrutinib) is CHOP-R in combination with acalabrutinib. Up to fifteen eligible participants will be recruited and registered into Cohort 2 for treatment with CHOP-R in combination with acalabrutinib. Recruitment to this cohort will close after 15 patients, or when recruitment to the randomised trial component closes, whichever occurs first.

Trial design schematic



Objectives

Primary objectives

Randomised component:
To compare progression free survival (PFS) in patients with RS receiving CHOP-R in combination with acalabrutinib with those receiving CHOP-R therapy alone.

Single-arm Cohort 1 (progressive RS following chemo-immunotherapy):
To assess the overall response rate (ORR) of acalabrutinib monotherapy.

Single-arm Cohort 2 (anthracycline-naïve RS patients, diagnosed while on ibrutinib):
To assess the ORR of CHOP-R in combination with acalabrutinib.

Secondary objectives

To evaluate overall survival (OS) in all cohorts, the ORR after cycle 6 for the randomised component only, and PFS for the single-arm cohorts. In all cohorts: Quality of Life (QoL), toxicity, and proportion of patients proceeding to allogeneic or autologous stem cell transplantation.

Exploratory objectives
Exploratory studies in all cohorts include patient response to treatment with regard to their underlying chronic lymphocytic leukaemia (CLL) and minimal residual disease (MRD), mutation analyses of RS tissue that may be predictive of outcome, and changes in and mutations of circulating lymphoma tumour DNA (ctDNA).

3. Outcome Measures

Primary outcome measures

Randomised component:
Progression free survival (PFS) defined as the time from randomisation to the date of progression or death from any cause. Progression will be defined by the modified Cheson criteria (Appendix 1). Patients who are alive and progression free at the time of analysis will be censored at date last seen. Any participants who withdraw or are lost to follow-up will also be censored at date last seen.

Single-arm Cohort 1 (progressive RS following chemo-immunotherapy):
Overall response after 24 weeks of treatment based on the modified Cheson criteria (Appendix 1). Complete response (CR) and partial response (PR) are considered to be positive responses for assessment of this outcome. Participants who withdraw prior to the 6-month assessment or do not have the 6-month assessment will be considered non-responders.

Single-arm Cohort 2 (anthracycline naïve RS patients, diagnosed whilst on ibrutinib):
Overall response after 6 cycles of treatment, defined by the modified Cheson criteria (Appendix 1). CR and PR are considered to be positive responses for assessment of this outcome. Participants who withdraw prior to the post-cycle-6 assessment or do not have the post-cycle-6 assessment will be considered non-responders.

Secondary outcome measures

Overall survival (OS) defined as time from date of randomisation (for randomised trial) or registration (to the relevant cohort for single-arm cohorts) to date of death from any cause. Participants who are alive at the time of analysis will be censored at date last seen.

- Note that participants randomised to receive CHOP-R, who progress on the standard of care arm and are eligible for Cohort 1, may enter single-arm Cohort 1 to receive acalabrutinib monotherapy. Their survival time for analysis of the randomised component will be from date of randomisation, and that used for analysis of Cohort 1 will be from date of registration to Cohort 1.

Overall response (randomised component only) after cycle 6, defined by the modified Cheson criteria (Appendix 1). CR and PR are considered to be positive responses for assessment of this outcome. Participants who withdraw prior to the post-cycle-6 assessment or do not have the post-cycle-6 assessment will be considered non-responders.

Overall response (Cohort 1 only) after 12 weeks, defined by the modified Cheson criteria (Appendix 1). CR and PR are considered to be positive responses for assessment of this outcome. Participants who withdraw prior to the 12-week assessment or do not have the 12-week assessment will be considered non-responders.

PFS (single-arm cohorts only) defined as the time from date of registration to date of progression or death from any cause. Progression will be defined by the modified Cheson criteria (Appendix 1). Participants who are alive and progression free at the time of analysis will be censored at date last seen. Any patients who withdraw or are lost to follow-up will also be censored at date last seen.

Quality of Life, (assessed using ECOG performance status [Appendix 6]), and the CLL17 and NHL-HG29 questionnaires (Appendix 2)) at the end of cycles 4 and 6 for participants receiving CHOP-R as part of their treatment (randomised cohorts and Cohort 2), and at 12 and 24 weeks for participants receiving acalabrutinib monotherapy (Cohort 1).

Toxicity defined as the number of participants who experience one or more adverse event grade 3 or higher or serious adverse event of any grade.

Proportion of participants proceeding to allogeneic or autologous stem cell transplantation

Exploratory outcome measures

- Proportion of participants with CLL MRD negativity (depletion of CLL cells below 10^{-4}) in the blood and bone marrow following treatment monitored using flow cytometry (randomised cohort and cohort 2).
- Identification of genes and pathways affected by mutations in RS tissue biopsies that might predict outcome.
- Changes in the level of detectable ctDNA during the course of the study.
- Presence of a specific mutation spectrum in ctDNA at diagnosis.
- Changes in the mutation spectrum of ctDNA during the course of the study.

4. Patient Population

Randomised trial component:
Anthracycline-naïve patients with CLL and newly diagnosed diffuse large B-cell lymphoma (DLBCL)-type RS are candidates for the randomised trial component. Prior exposure to Bruton tyrosine kinase (BTK), phosphoinositide-3-kinase (PI3K), or BCL-2 inhibitors is permitted, with the caveat that patients developing RS whilst on ibrutinib treatment are **not eligible** for the **randomised component** of the STELLAR trial.

CLL patients who discontinue treatment with ibrutinib due to toxicity are permitted to enrol in the randomised trial. Similarly, CLL patients who discontinue treatment with ibrutinib due to progressive CLL, receive another CLL-directed therapy and later (more than four weeks) develop RS (diagnosis of RS more than four weeks from the last dose of ibrutinib) may enrol in the randomised trial.
CLL patients developing RS on ibrutinib (diagnosis of RS less four weeks from the last dose of ibrutinib) are not eligible for the randomised trial and may be candidates for treatment on Cohort 2. (See Section 4 for guidance on eligibility for these cohorts).

Cohort 1 (Progressive RS following chemo-immunotherapy):
Patients with DLBCL-type RS who progressed following standard of care chemo-immunotherapy (anthracycline-based chemotherapy in combination with anti-CD20 monoclonal antibody) are candidates for Cohort 1, acalabrutinib monotherapy.

RS patients who participated in the randomised component of this trial and received the standard of care (CHOP-R) treatment arm may enrol in Cohort 1 and receive acalabrutinib monotherapy.
If fewer than the anticipated minimum number of participants continue from the STELLAR randomised trial component into Cohort 1 for treatment with acalabrutinib, recruitment may be opened to participants who received chemo-immunotherapy (anthracycline-based chemotherapy in combination with anti-CD20 monoclonal antibody) outside of the randomised trial.

Cohort 2 (anthracycline-naïve RS patients, diagnosed while on ibrutinib):
CLL patients, who are anthracycline-naïve and develop DLBCL-type RS while on ibrutinib treatment (RS diagnosis within four weeks of the last dose of ibrutinib) can be recruited into Cohort 2.

5. Sample Size

Randomised component:
A total of 60 eligible patients will be randomised on a 1:1 ratio between the standard of care (CHOP-R) and the experimental treatment arm (CHOP-R in combination with acalabrutinib).

Cohort 1 (progressive RS following chemo-immunotherapy):
Between 21 and 30 patients will be recruited into Cohort 1 to receive acalabrutinib monotherapy.

Cohort 2 (anthracycline-naïve RS patients, diagnosed while on ibrutinib):
Up to fifteen patients will be recruited into Cohort 2 to receive CHOP-R in combination with acalabrutinib.

Appendix C: Panel targets for the DLBCL panel

See supplementary file ‘AppendixC_Panel_targets_for_DLBCL_Chapter4.bed’

Appendix D: Panel targets for the CLL and RT panel

See supplementary file ‘AppendixD_Panel_targets_for CLL_Chapter3_Richters_Chapter5.bed’

Appendix E: Circulating tumour DNA variants detected in the ACCEPT (DLBCL) cohort

See supplementary file 'Chapter4_ctDNA_variants_detected_ACCEPT.tsv'

Appendix F: Circulating tumour DNA variants detected in the Richters Transformation cohorts

See supplementary file ‘AppendixF_Chapter5_ctDNA_variants_detected_Richters.tsv’

Appendix G: Examiners comments and actions taken for the revised thesis