Improving earlier non-invasive diagnosis of high-grade serous ovarian cancer

Elizabeth Kerr Moore

Darwin College

July 2018

This dissertation is submitted for the degree of Doctor of Philosophy

Improving earlier non-invasive diagnosis of high-grade serous ovarian cancer
Elizabeth Kerr Moore

The majority of women with ovarian cancer (OC) have advanced disease at diagnosis and 5-year survival rates of less than 25%. Women with stage I disease have significantly better 5-year survival rates of over 90%. Recent large studies using CA 125 and transvaginal ultrasound have failed to improve mortality in a screened population. There is therefore a pressing need for new diagnostic biomarkers in OC.

The primary aim of my project, as a first step in developing a diagnostic circulating tumour DNA (ctDNA) biomarker for high grade serous ovarian cancer (HGSOC), was to investigate low-cost high-throughput next generation sequencing assays in plasma samples collected from women with newly diagnosed OC. The secondary aim was to apply these methods to other non-invasive samples including cervical liquid based cytology samples that might contribute to earlier diagnosis or screening for women with OC.

ctDNA was detected in 30-49% of women with newly diagnosed OC from the UKOPS (n=54) and CTCR-OV04 (n=156) cohorts using targeted sequencing. Using the trimmed median absolute deviation (t-MAD) score, a quantitative measure of genome wide copy number aberration generated from shallow whole genome sequencing (sWGS) data, ctDNA was detected in 39–41% of the women with newly diagnosed disease.

To improve sensitivity of ctDNA detection I developed an optimised method for targeted sequencing that has the potential to lower the limit of detection of ctDNA in HGSOC by 100 fold. I have also shown that the size profile of HGSOC ctDNA fragments is different to that of wildtype DNA fragments and shown that selecting for DNA fragments between 90–150 bp can increase rates of ctDNA detection in HGSOC. ctDNA detection increased to 53–67% of women with newly diagnosed OC using the size selected t-MAD score.

I have evaluated the utility of cervical sampling for earlier diagnosis of OC by testing and optimising DNA extraction, library preparation and sequencing methods. I have detected tumour DNA in routine cervical cytology samples collected from women subsequently diagnosed with cervical and endometrial cancers.

In summary I have developed methods for ctDNA detection in women with newly diagnosed HGSOC that can be applied and refined in larger prospective studies of women undergoing follow-up for treated HGSOC, women with symptoms suggestive of OC and women at high risk of OC.

Acknowledgements

I would like to thank my supervisors James Brenton and Nitzan Rosenfeld for their support and guidance over the last few years. I would like to thank all current and past members of the Brenton and Rosenfeld labs. In particular Ania Piskorz, Teodora Goranova, Filipe Martins, Dineika Chandrananda, Florent Mouliere and Wendy Cooper. I would like to thank the core facilities at the CRUK CI in particular the Genomics core.

I would like to thank Charlotte Hodgkin, Karen Hoskings, Rachel Moore, Heather Biggs and Linda Jones from the research team.

I would like to thank the clinical team at Addenbrookes Hospital, in particular Robin Crawford. I would also like to thank Peter Baldwin, John Latimer, Krish Haldar, Helen Bolton, Helen Addley, Evis Sala and Ramona Woitek.

I would like to thank Target Ovarian Cancer, Medical Research Council and Cambridge Cancer Centre for funding my research.

Finally I would like to thank my family: Noah, Erin and Finlay — it's been an adventure. Jon — thank you for everything.

Contents

1 Abstract **21**

2 Introduction **23**

2.1 Ovarian cancer 23

2.2 CA 125 24

2.3 CA 125 based screening 25

2.4 Other biomarkers in ovarian cancer 26

2.5 Circulating tumour DNA 27

2.6 Circulating tumour DNA in ovarian cancer 28

2.7 Cervical sampling in ovarian cancer 29

2.8 Overview 30

2.9 Project aims 31

3 Methods **33**

3.1 Study populations 33

3.1.1 CTCR-OV04 33

3.1.2 UKOPS 33

3.1.3 CTCR-OV05 34

3.1.4 CS01 34

3.2 Sample collections 35

3.2.1 FFPE samples 35

3.2.2 Blood samples 35

3.2.3 Cervical samples 36

3.3 Cell culture . . . 36
3.4 Extraction of DNA . . . 37
3.4.1 Extraction of DNA from FFPE . . . 37
3.4.2 Extraction of DNA from cervical cytology . . . 38
3.4.3 Extraction of DNA from cervical mucous . . . 38
3.4.4 Extraction of DNA from cultured cells . . . 38
3.4.5 Extraction of DNA from plasma . . . 39
3.4.6 Quantification of DNA . . . 40
3.5 Library preparation . . . 40
3.5.1 TAm-Seq . . . 40
3.5.2 Whole genome library preparation . . . 43
3.5.3 Digital PCR . . . 44
3.5.4 Whole genome single stranded library preparation . . . 46
3.5.5 TAm-Seq V2 . . . 47
3.6 In vitro size selection . . . 49
3.7 p53 peptide microarray . . . 49
3.8 Bioinformatics analysis . . . 50
3.8.1 Analysis of TAm-Seq . . . 50
3.8.2 Analysis of shallow whole genome sequencing data . . . 50
3.8.3 In silico size selection . . . 51
3.8.4 Quantification of the 10 bp periodic oscillation . . . 51
3.8.5 Analysis of TAm-Seq V2 . . . 52

4 Detection of plasma circulating tumour DNA in women with newly diagnosed ovarian cancer 53

4.1 Introduction . . . 53
4.2 Results . . . 55
4.2.1 UKOPS cohort . . . 55
4.2.2 CTCR-OV04 cohort . . . 61
4.3 Discussion . . . 69

5 Improving sensitivity and specificity for circulating tumour DNA detection 73

5.1 Introduction . . . 73

5.2 Results . . . 75

5.2.1 TAm-Seq V2 . . . 75

5.2.2 Selective sequencing using DNA fragment sizes . . . 88

5.3 Discussion . . . 99

6 Improved sensitivity for detection of circulating tumour DNA in newly diagnosed ovarian cancer 103

6.1 Introduction . . . 103

6.2 Results . . . 104

6.2.1 UKOPS cohort . . . 104

6.2.2 CTCR-OV04 cohort . . . 113

6.3 Discussion . . . 127

7 Investigating cervical sampling to increase detection of ovarian cancer 131

7.1 Introduction . . . 131

7.2 Results . . . 132

7.2.1 Investigation of quality of cervical cytology DNA for sequencing . . . 132

7.2.2 Application of targeted sequencing to routinely collected cervical cytology samples . . . 139

7.2.3 Cervical samples in high-grade serous ovarian cancer . . . 139

7.3 Discussion . . . 144

8 Circulating tumour DNA as a prognostic and predictive biomarker in ovarian cancer 147

8.1 Introduction . . . 147

8.2 Results . . . 148

8.2.1 CTCR-OV04 cohort . . . 148

8.3 Discussion . . . 155

9 Further understanding of circulating tumour DNA and application for future clinical trials 157

9.1 Introduction . . . 157
9.2 Results . . . 158
9.2.1 Dynamics of circulating tumour DNA . . . 158
9.2.2 Collection tubes . . . 164
9.2.3 Serum . . . 166
9.3 Discussion . . . 171

10 Summary **175**
10.1 Future directions . . . 178

11 Bibliography **181**

12 Appendices **i**
12.1 CTCR-OV05 study documents . . . ii
12.1.1 Protocol . . . ii
12.1.2 Patient information sheet — referral . . . xvi
12.1.3 Consent form — referral . . . xxii
12.1.4 Patient information sheet — surgery . . . xxiv
12.1.5 Consent form — surgery . . . xxxi
12.2 CS01 study documents . . . xxxiii
12.3 TAm-Seq primer panel 1 . . . xlv
12.4 TAm-Seq primer panel 10 . . . xlvi
12.5 TAm-Seq UKOPS primer panel . . . xlvii
12.6 Mutation calls for UKOPS FFPE using TAm-Seq . . . xlviii
12.7 De novo mutation calls for UKOPS plasma using TAm-Seq . . . xlix
12.8 Specific variant mutation calls for UKOPS plasma using TAm-Seq . . . li
12.9 UKOPS copy number profiles: tumour, unselected plasma and size selected plasma . . . liii
12.10 De novo mutation calls for CTCR-OV04 plasma using TAm-Seq . . . lvii
12.11 Benign controls . . . lx
12.12 Specific variant calls for CTCR-OV04 plasma using TAm-Seq . . . lxi

12.13 Manuscript: Science Translational Medicine lxii
12.14 Copy number profiles for relapsed HGSOC cases with and without size selection xciv
12.15 TAm-Seq V2 optimisations cviii
12.15.1 Enzyme cviii
12.15.2 Annealing temperature and extension time cix
12.15.3 Primer concentrations cxii
12.15.4 Purification first round polymerase chain reaction (PCR) product . cxiii
12.15.5 Barcoding cxiii
12.15.6 Purification of final PCR product cxv
12.15.7 Primer pooling cxv
12.16 Mutation calls in cervical cytology samples cxvi
12.17 Specific variant calls for cervical samples using TAm-Seq cxviii
12.18 Mutation calls using TAm-Seq V2 in cervical samples cxix
12.19 p53 autoantibodies and ctDNA cxxv
12.20 Mutation calling using TAm-Seq for plasma samples around surgery cxxvii
12.21 Manuscript: Journal of Molecular Diagnostics cxxviii
12.22 Mutation calling using TAm-Seq for serum samples clix

List of Figures

3.1 Digital PCr protocol . . . 46

4.1 Schematic illustrating approach for detection of ctDNA in women with newly diagnosed OC . . . 54

4.2 UKOPS experimental overview . . . 58

4.3 Fragmented DNA in UKOPS HGSOC FFPE cases with no *TP53* mutation . . . 59

4.4 UKOPS sWGS analysis with t-MAD score . . . 60

4.5 CTCR-OV04 experimental overview . . . 62

4.6 CTCR-OV04 targeted sequencing analysis . . . 65

4.7 CTCR-OV04 correaltion of MAF with age at diagnosis and CA 125 . . . 66

4.8 CTCR-OV04 sWGS analysis using t-MAD score . . . 67

4.9 CTCR-OV04 tumour volume . . . 68

5.1 Approaches for increasing the sensitivity of ctDNA detection . . . 75

5.2 Experimental conditions investigated to optimise TAm-Seq V2 . . . 76

5.3 TAm-Seq V2 sequencing coverage comparison of different experimenatl conditions . . . 77

5.4 TAm-Seq V2 sequencing coverage comparison of different purification methods . . . 77

5.5 TAm-Seq V2 sequencing coverage comparison of different barcoding methods 78

5.6 TAm-Seq V2 sequencing coverage comparison of different pooling strategies 79

5.7 Comparison of expected and observed AF TAm-Seq V2 . . . 85

5.8 Mouse xenograft plasma . . . 88

5.9 Xenograft fragmentation profiles . . . 89

5.10 Fragmentation profiles HGSOC cases and healthy controls . . . 91

5.11 MAF before and after size selection . . . 92

5.12 Distribution of DNA fragments before and after size selection 93
5.13 Example of copy number profile before and after size selection 94
5.14 t-MAD score before and after size selection 95
5.15 ROC analysis in-vitro and in-silico size selected t-MAD score 96
5.16 SCNA alanlysis before and after size selection 97
5.17 Fragmentation feature analysis 98
6.1 UKOPS comparison size selected t-MAD and MAF 104
6.2 UKOPS comparison of unselected and size selected t-MAD score 105
6.3 UKOPS size selected t-MAD analysis 108
6.4 UKOPS ctDNA detection using size selected t-MAD score 109
6.5 Comparison of MAF and IchorCNA tumour fraction in UKOPS cohort 110
6.6 UKOPS ctDNA detection using size selected IchorCNA 111
6.7 UKOPS ROC analysis 112
6.8 CTCR-OV04 comparison of unselected and size selected t-MAD 114
6.9 CTCR-OV04 unselected and size selected t-MAD by stage 116
6.10 CTCR-OV04 ctDNA detection using size selected t-MAD 119
6.11 CTCR-OV04 tumour volume 120
6.12 CTCR-OV04 ROC analysis t-MAD 121
6.13 CTCR-OV04 fragmentation features 122
6.14 CTCR-OV04 ROC analysis fragmentation features 123
6.15 CTCR-OV04 IchorCNA analysis 124
6.16 CTCR-OV04 size selected IchorCNA 125
6.17 CTCR-OV04 ROC analysis IchorCNA 126
7.1 Overview of approach for investigating cervical sampling 132
7.2 Cervical cytology composition 133
7.3 Cervical cytology DNA sequencing coverage 134
7.4 Cultured cells in fixative 136
7.5 Sequencing coverage cultured cells in fixative 137
7.6 Fragmentation pattern of cultured cells in fixative 138

7.7 sWGS analysis cervical samples 142
7.8 TAm-Seq V2 analysis cervical samples 143
8.1 CTCR-OV04 comparison ctDNA and end of treatment response 149
8.2 CTCR-OV04 fragmentation features and end of treatment response 151
8.3 CTCR-OV04 change in MAF with neo-adjuvant chemotherapy 152
8.4 CTCR-OV04 change in t-MAD with neo-adjuvant chemotherapy 153
8.5 CTCR-OV04 change in fragmentation features with neo-adjuvant chemotherapy 154
9.1 Fragmentation features and ascites drainage 159
9.2 t-MAD score around surgery 162
9.3 Fragmentation features around surgery 163
9.4 Analysis of cell stabilisation tubes 165
9.5 Mouse xenograft serum 167
9.6 Serum experiment overview 168
9.7 Comparison of fragmentation profiles for plasma and serum and different extraction methods 170
12.1 TAm-Seq V2 primer pooling strategies cxv

List of Tables

3.1 Schedule of blood collection for surgery cohort of CTCR-OV05 36
3.2 Volume of reagments for DNA extraction from plasma 40
4.1 UKOPS summary of clinical data 56
4.2 CTCR-OV04 summary of clinical data 61
4.3 CTCR-OV04 cases with CA 125 <35 U/ml 63
5.1 Final V2 parameters 78
5.2 *TP53* digital PCR primers and probes 80
5.3 Digital PCR primer and probe optimisations 81
5.4 TAm-Seq V2 sensitivity using patient pool 81
5.5 Standard TAm-Seq of pool used for V2 optimisation 82
5.6 Expected MAF for pooled mutations in TAm-Seq V2 82
5.7 Number of expected positive wells TAm-Seq V2 optimisation 82
5.8 Number of positive wells TAm-Seq V2 83
5.9 Observed AF TAm-Seq V2 84
5.10 TAm-Seq V2 mutation calling 86
5.11 Digital PCR quantification of patient plasma samples 87
5.12 Xenograft samples and whole genome library preparation 90
6.1 UKOPS ctDNA detection using IchorCNA 106
6.2 UKOPS ctDNA detection using size selected IchorCNA 107
6.3 CTCR-OV04 percentage change in t-MAD with size selection 115
6.4 CTCR-OV04 ctDNA detection using IchorCNA tumour fraction 118
8.1 CTCR-OV04 clinical response 148

8.2 Percentage of responders and non-responders with >60% reduction in MAF after 1 (C1-C2) and 2 (C1-C3) cycles of chemotherapy. 150

9.1 MAF pre and post ascites drainage 158

9.2 MAF pre and post surgery 160

9.3 Residual disease 161

10.1 ctDNA summary 177

Acronyms

AF allele fraction.

AUC area under the curve.

CCCBPL Cambridge Cancer Centre Blood Processing Laboratory.

cfDNA cell free DNA.

CMA cervical mucous aspirate.

CMDL Cancer Molecular Diagnostics Laboratory.

ctDNA circulating tumour DNA.

EDTA ethylenediaminetetraacetic acid.

FFPE formalin fixed paraffin embedded.

H&E hematoxylin and eosin.

HGSOC high grade serous ovarian cancer.

HPV human papillomavirus.

IDS interval debulking surgery.

IGV Integrative Genomics Viewer.

IOTA International Ovarian Tumour Analysis.

LBC liquid based cytology.

LLETZ large loop excision of the transformation zone.

MAF mutant allele fraction.

NGS next-generation sequencing.

NIHR National Institute for Health Research.

NSCLC non small cell lung cancer.

OC ovarian cancer.

PBS phosphate buffered saline.

PCR polymerase chain reaction.

PD progressive disease.

PLCO the prostate, lung, colorectal and ovarian.

PR partial response.

RCT randomised control trial.

RMI Risk of Malignancy Index.

ROC receiver operating characteristic.

ROCA risk of ovarian cancer algorithm.

ROCkeTS Refining Ovarian Cancer Test accuracy Scores.

ROMA Risk of Ovarian Malignancy Algorithm.

SCNA somatic copy number alteration.

SD stable disease.

STIC serous tubal intraepithelial carcinoma.

sWGS shallow whole genome sequencing.

t-MAD trimmed median absolute deviation.

TTP time to progression.

TVUS transvaginal ultrasound scan.

UKCTOCS UK Collaborative Trial of Ovarian Cancer Screening.

UKOPS UK Ovarian Cancer Population Study.

VAF variant allele fraction.

WBC white blood cell.

Abstract

The majority of women with OC have advanced disease at diagnosis and 5-year survival rates of less than 25%. Women with stage I disease have significantly better 5-year survival rates of over 90%. Recent large studies using CA 125 and transvaginal ultrasound have failed to improve mortality in a screened population. There is therefore a pressing need for new diagnostic biomarkers in OC.

The primary aim of my project, as a first step in developing a diagnostic ctDNA biomarker for HGSOC, was to investigate low-cost high-throughput next generation sequencing assays in plasma samples collected from women with newly diagnosed OC. The secondary aim was to apply these methods to other non-invasive samples including cervical liquid based cytology samples that might contribute to earlier diagnosis or screening for women with OC.

ctDNA was detected in 30-49% of women with newly diagnosed OC from the UKOPS (n=54) and CTCR-OV04 (n=156) cohorts using targeted sequencing. Using the t-MAD score, a quantitative measure of genome wide copy number aberration generated from sWGS data, ctDNA was detected in 39–41% of the women with newly diagnosed disease.

To improve sensitivity of ctDNA detection I developed an optimised method for targeted sequencing that has the potential to lower the limit of detection of ctDNA in HGSOC by 100 fold. I have also shown that the size profile of HGSOC ctDNA fragments is different to that of wildtype DNA fragments and shown that selecting for DNA fragments between 90–150 bp can increase rates of ctDNA detection in HGSOC. ctDNA detection increased to 53–67% of women with newly diagnosed OC using the size selected t-MAD score.

I have evaluated the utility of cervical sampling for earlier diagnosis of OC by testing and optimising DNA extraction, library preparation and sequencing methods. I have detec-

ted tumour DNA in routine cervical cytology samples collected from women subsequently diagnosed with cervical and endometrial cancers.

In summary I have developed methods for ctDNA detection in women with newly diagnosed HGSOC that can be applied and refined in larger prospective studies of women undergoing follow-up for treated HGSOC, women with symptoms suggestive of OC and women at high risk of OC.

Introduction

2.1 Ovarian cancer

Ovarian cancer (OC) is the fifth most common cancer in women in the UK with approximately 7,000 new cases diagnosed each year (CRUK 2016). Approximately 60% of women already have stage III/IV disease at the time of diagnosis (CRUK 2016). Presentation with advanced disease is more common (>70%) in women with high grade serous ovarian cancer (HGSOC) and these cases account for the majority of mortality from OC. Although initially 70–80% of women respond well to chemotherapy ultimately most develop chemotherapy resistance leading to treatment failure. 5-year survival rates for women with advanced disease at the time of diagnosis are less than 25%, however women with early stage disease at the time of diagnosis have 5-year survival rates of over 90%, including significantly improved survival for women with HGSOC (CRUK 2016). Since the early 1970s mortality rates from OC have only slightly improved (CRUK 2016) with survival in the UK lagging behind the rest of Europe (Coleman et al. 2011). There is therefore a pressing need to develop methods that might enable earlier diagnosis of and potentially screening for OC.

HGSOC originates from the fallopian tube epithelium, specifically arising from intraepithelial tubal carcinoma, preferentially at the fimbrial end of the fallopian tube. High rates of co-existent serous ovarian carcinoma and serous tubal intraepithelial carcinoma (STIC) (Kindelberger et al. 2007; Przybycin et al. 2010) and primary serous peritoneal carcinoma and intrapepithelial tubal carcinoma (Carlson et al. 2008) have been identified. Identical *TP53* mutations have been identified in STIC and paired HGSOC samples (Kindelberger et al. 2007; Lee et al. 2007; Kuhn et al. 2012) suggesting a clonal relationship. Mouse models of HGSOC have shown that oophrectomy does not prevent the development of HGSOC but that bilateral salpingectomy is preventative (Perets et al. 2013; Kim et al. 2012).

The fallopian tube is in direct communication with the peritoneal cavity. As such symptoms of OC are generally non-specific and include abdominal bloating, loss of appetite, early satiety, abdominal pain, urinary frequency and urgency. One in two women aged between 50–70 present to their GP each year with these symptoms (Lim et al. 2014). The unique biology of HGSOC also results in early metastatic spread, particularity to the omentum. Diagnosis of OC is currently a challenge with only 1/3 patients with OC in the UK presenting through a primary care referral (Elliss-Brookes et al. 2012). Thus tests with improved sensitivity and specificity in women with non-specific symptoms are critical to improving mortality.

The gene *TP53* encodes the tumour suppressor protein p53, a transcription factor that regulates the expression of proteins involved in apoptosis and genomic integrity. *TP53* mutations are ubiquitous in HGSOC (Ahmed et al. 2010; Kobel et al. 2016) and have been identified as early drivers of the disease (Labidi-Galy et al. 2017). Other point mutations are uncommon in HGSOC which has been identified as a disease characterised by copy number alterations (Ciriello et al. 2013) and unlike most other cancer subtypes these are typically acquired during the first half of clonal evolution (Gerstung et al. 2017). The unique genomic features found in HGSOC present another significant challenge to the early detection of HGSOC.

2.2 CA 125

CA 125 is a glycoprotein encoded by the *MUC16* gene that was originally found to be elevated (>35 U/ml) in serum in 82% of OC cases but only 1% of healthy and 6% of benign controls (Bast et al. 1983). CA 125 is a well validated biomarker instrumental to the current clinical management of OC. Current guidelines recommend measuring serum CA 125 in primary care in all women with symptoms suggestive of OC followed by a transvaginal ultrasound scan (TVUS) in those with elevated levels (RCOG 2016; NICE 2011). The combination of CA 125, ultrasound findings (expressed as a score of 0, 1 or 3) and menopausal status (1 if premenopausal, 3 if postmenopausal), are used to calculate a Risk of Malignancy Index (RMI). RMI with a cut-off of 200 was found to have a sensitivity of 85% and specificity of 97% to discriminate OC cases from benign controls (Jacobs et al. 1990).

However CA 125 is limited in both sensitivity and specificity as a diagnostic biomarker. In both screening and pre-surgical studies it has been shown that >50% of women with

stage I disease do not have an elevated CA 125 (Jacobs et al. 1989; Jacobs et al. 1993) potentially resulting in delayed or missed diagnosis. Conversely CA 125 is elevated in a number of benign gynaecological and non-gynaecological conditions, particularly in premenopausal women, leading to further potentially unnecessary invasive investigations and associated anxiety.

2.3 CA 125 based screening

A mortality benefit to OC screening using CA 125 has yet to be elucidated. The prostate, lung, colorectal and ovarian (PLCO) cancer screening trial of over 60,000 women using annual TVUS and CA 125 found no increase in cancers detected at an earlier stage and no improvement in mortality (Buys et al. 2011). This study was limited in that a single threshold was used for CA 125, there was a lack of central protocol for management of a positive result and there was a long follow up with 40% of OC diagnoses occurring after the end of screening. An increase in detection of early stage cancer was found in the Japanese study of OC screening of over 80,000 women again using annual ultrasound and CA 125 (Kobayashi et al. 2008) however, the mortality rates are unknown. An improved survival in women undergoing annual screening with ultrasound was found in The Kentucky screening study of over 20,000 women (Nagell et al. 2007). This however was not a randomised control trial (RCT) and findings may be biased by a healthy volunteer effect.

The UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) is the largest screening study with over 200,000 postmenopausal women aged 50–74 years recruited from 13 centres in England, Wales and Northern Ireland. Women were randomised to annual multimodal screening using CA 125 interpreted using the risk of ovarian cancer algorithm (ROCA), annual ultrasound or no screening in a 1:1:2 ratio and followed up for a median of 11.1 years. This study initially reported encouraging data with an increase in sensitivity and specificity, 89.4% and 99.8% respectively, for the diagnosis of OC in women in the multimodal screening arm (Menon et al. 2009). Primary analysis did not reveal a significant reduction in mortality in the multimodal screening arm (Jacobs et al. 2016). However, an encouraging mortality reduction of 23% was seen in years 7–14 compared to 8% in years 0–7 suggesting that a mortality benefit might be seen after longer follow-up.

2.4 Other biomarkers in ovarian cancer

Significant work is currently being undertaken to identify new protein biomarkers that can potentially be used in combination with CA 125 and RMI to improve triage of OC. One example is HE4, a serum protein biomarker, that has been found to have increased sensitivity for the diagnosis of OC compared to RMI when used in combination with menopausal status and CA 125 to calculate a Risk of Ovarian Malignancy Algorithm (ROMA) (Moore et al. 2010).

However, despite decades of effort, CA 125 remains the single-best biomarker for OC with models derived from the most promising markers failing to show improvement over serum CA 125 alone (Cramer et al. 2011; Zhu et al. 2011).

Refining Ovarian Cancer Test accuracy Scores (ROCkeTS) is a National Institute for Health Research (NIHR) funded study that has been developed by the University of Birmingham with four phases:

1. Systematic review of the literature.
2. Interrogation of datasets/samples from UK Ovarian Cancer Population Study (UKOPS), UKCTOCS and International Ovarian Tumour Analysis (IOTA) to refine model using tests identified in phase 1.
3. Prospective study to collect serum and perform ultrasound with IOTA criteria to validate new models.
4. Analysis and pathway generation.

Recruitment to the prospective part of the ROCkeTS study commenced in June 2015 and as of March 2018 1,021 postmenopausal and 737 premenopausal women had been recruited. This will provide a valuable resource to validate CA 125 and other protein biomarkers for triage of symptomatic women in secondary care. Interim analysis of current recruitment to ROCkeTS showed a prevalence of OC of 8% among postmenopausal women and 1.5% in premenopausal women (personal communication Sudha Sundar, Univeristy of Birmingham).

Alternative biomarkers to serum proteins include the detection of tumour-associated antibodies in serum. Proteins produced by tumour cells can be mutated or over-expressed.

Release of these into the circulation can trigger an autologous immune response generating autoantibodies. It has been shown that p53 specific autoantibodies can be detected in 24–42% of OC cases (Angelopoulou et al. 1996; Anderson et al. 2010; Tsai-Turton et al. 2009) including 16% of UKCTOCS cases not detected using the multi-modal screening strategy (Yang et al. 2017).

2.5 Circulating tumour DNA

An alternative strategy for earlier detection and/or screening is the detection of circulating tumour DNA (ctDNA). It has been known since 1948 that blood contains free circulating DNA (Mandel et al. 1948). In 1977 it was shown that there are higher levels of free circulating DNA in the blood of patients with cancer (Leon et al. 1977) and studies have consistently found higher levels of free circulating DNA in the blood of patients with cancer (mean 219 ng/ml, range 10–1200 ng/ml) compared to healthy controls (mean 3.7 ng/ml, range 0–100 ng/ml) (Jahr et al. 2001). In 1989 it was first shown that tumour DNA can be identified in the blood of patients with cancer (Stroun et al. 1989) however, tumour DNA can make up a tiny fraction of the total cell free DNA (cfDNA). Using next-generation sequencing (NGS) techniques it is now possible to amplify small amounts of free circulating DNA in the blood to identify molecular alterations observed in tumour DNA and identify these from the large amount of non-tumour cfDNA in the circulation. ctDNA is a highly specific biomarker of cancer and could therefore provide a powerful non-invasive method for earlier OC diagnosis.

The detection of ctDNA has had an extensive impact in oncology with widespread use in research studies of advanced and relapsed disease (Wan et al. 2017). ctDNA has been shown to be useful for the prediction of clinical response in metastatic breast cancer (Dawson et al. 2013), prediction of disease relapse in colorectal cancer (Diehl et al. 2008) and breast cancer (Garcia-Murillas et al. 2015) and for the detection of resistance mechanisms in colorectal cancer (Diaz et al. 2012; Kuang et al. 2009).

There has also been considerable recent interest in using ctDNA for early detection of cancer (Wan et al. 2017). There are significant challenges in applying this approach to detection of early disease where levels of ctDNA are expected to be very low and in many cases below detection thresholds of current methodologies. Strategies applied to increase the detection of ctDNA in early stage cancers include the detection of patient specific muta-

tions (Bettegowda et al. 2014; Abbosh et al. 2017), deep sequencing of multiple genes (Phallen et al. 2017) and the combination of ctDNA and protein based biomarkers (Cohen et al. 2018). These powerful methods, able to detect down to an allele fraction (AF) of 0.1–0.01%, have been able to detect ctDNA in 40–50% of stage I cancers. Although these studies have shown feasibility for detection of ctDNA in early stage cancers, results in HGSOC are preliminary. There is not yet any evidence that cancers can be detected earlier, or that this will lead to improved clinical outcomes. These methods are also unlikely to be cost effective. It is possible that sensitivity also may not be sufficient for the detection of very low tumour volumes. Using linear modelling it has been predicted that a tumour with a volume of 10 cm^3 will have a variant allele fraction (VAF) of 0.1% and a tumour with a volume of 1 cm^3 will have a VAF of 0.008% (Abbosh et al. 2017).

2.6 Circulating tumour DNA in ovarian cancer

Original studies found ctDNA to be present in 17–30% of plasma and serum samples obtained from women with OC (Otsuka et al. 2004; Swisher et al. 2005). These studies examined the frequency of *TP53* mutations in samples but were limited to only a few nucleotide substitutions and were relatively insensitive.

A targeted sequencing method (TAm-Seq) developed by the labs uses a two-stage amplification process for amplification and deep sequencing of exons of driver genes from fragmented and/or low abundance input DNA such as plasma cfDNA. Prior knowledge of the patients tumour specific mutation is not required. Control samples are used to assess background sequencing error rates and each sample is analysed in duplicate to control for sequencing errors and increase specificity. The original method using a sequencing depth of approximately 700× was able to detect ctDNA in 20/38 (53%) of women with advanced HGSOC down to AF of 2% with a specificity of >95% (Forshew et al. 2012). TAm-Seq is a cost effective method and may be suitable for clinical application.

Digital PCR has a higher sensitivity for ctDNA detection. Using prior knowledge of the patients tumour specific mutation to design tumour-specific assays for *TP53* ctDNA can be detected in >80% of women with HGSOC including women with newly diagnosed disease (Parkinson et al. 2016). This method cannot be used in a diagnostic setting as prior knowledge of the patients tumour specific mutation is required. This study does however provide evidence that ctDNA is present in plasma samples collected before treatment for

HGSOC.

In relapsed HGSOC it has been shown that there is a relationship between tumour volume measured by 3D CT volumetric analysis and ctDNA levels. Using digital PCR for *TP53* 45/47 patients with a tumour volume of >32 cm^3 had detectable ctDNA. Only 2/10 cases with a tumour volume of <32 cm^3 had detectable ctDNA and one of these can potentially be explained by the presence of large volume ascites. The *TP53* mutant allele fraction (MAF) was moderately correlated with the tumour volume (correlation coefficient 0.59) and was higher in cases without ascites (correlation coefficient 0.82). CA 125 was less well correlated with tumour volume (correlation coefficient 0.52 and 0.51 respectively). Interestingly the median *TP53* MAF per tumour volume was much higher in the relapsed HGSOC cases (0.04% per cm^3) compared to newly diagnosed HGSOC cases (0.0008% per cm^3) (Parkinson et al. 2016).

These studies all focused on the detection of ctDNA in relapsed HGSOC. There is an opportunity for earlier detection of HGSOC if appropriate methods with high enough sensitivity can be developed. Persistent *TP53* mutations in fallopian tubal epithelial cells have been conclusively demonstrated to be the first pathological change in STIC. STIC is the precursor or preinvasive lesion to HGSOC (Lee et al. 2007). p53 staining of fallopian tube sections to identify abnormal staining patterns for p53 protein is established in clinical practice to identify STIC lesions. A recent publication using whole exome sequencing for detection of somatic mutations and copy number aberrations for matched STIC and fallopian tube cancers has confirmed the presence of *TP53* mutations in STIC lesions. Using evolutionary analysis, the authors estimate that the average time between development of STIC and invasive carcinoma to be 7 years (Labidi-Galy et al. 2017). There is therefore evidence to suggest that preinvasive lesions to HGSOC do harbour *TP53* mutations and that there is a significant window of opportunity for earlier detection of OC.

2.7 Cervical sampling in ovarian cancer

Routine cervical screening was introduced as a method of detecting pre-malignant changes in ectocervical cells that could progress to cervical cancer if not treated. In 2008 the conventional Pap smear was replaced by liquid based cytology (LBC) as a means of reducing the number of inadequate samples obtained. In addition the use of LBC has allowed the collection of DNA as well as cytological evaluation of cervical samples. This has been util-

ised for human papillomavirus (HPV) triage and test of cure.

Cells and DNA shed from the distal fallopian tube can pass through the uterus and cervix to the vagina where they can be sampled. Occasionally OC is diagnosed following an abnormal cervical cytology sample in asymptomatic women with otherwise normal clinical investigations. It is estimated that <0.1% of all routine cervical smear samples are thought to show glandular abnormalities (NHS 2016) which suggest possible endocervical, endometrial or ovarian abnormalities. Approximately 30–40% of these are subsequently diagnosed with pre-invasive or invasive disease (DeSimone et al. 2006; Boyraz et al. 2017), the minority of these being ovarian.

Another approach for increasing the sensitivity of detection of early stage disease is to collect samples proximal to the suspected location of tumour, in which tumour DNA is likely to be enriched. Maritschnegg et al. 2015 found a 60% sensitivity for the detection of OC using lavage of the uterine cavity and Wang et al. 2018 found a sensitivity of 45% for the detection of OC by performing direct sampling of the intrauterine cavity using a Tao brush. These procedures are not suitable for widespread use in symptomatic women as they are invasive procedures that can only be performed by trained healthcare professionals. These procedures are also uncomfortable and expensive.

Studies looking at less invasive methods of sampling at the cervical os have shown that tumour DNA and somatic mutations can be identified in cervical cytology samples (Kinde et al. 2013) and vaginal tampons (Erickson et al. 2014) from women with advanced OC. More recently it has been shown that 33% of OCs (n=245) were detected using a combination of somatic mutation detection in 18 genes and aneuploidy detection in cervical cytology samples (Wang et al. 2018). This included detection in 34% of early stage cancers. Combining cervical cytology and plasma increased the sensitivity to 63% in 83 women with OC. This suggests that cervical sampling could potentially be used as part of a diagnostic or triage tool for OC.

2.8 Overview

In summary existing approaches for the detection and screening for early stage OC have a low sensitivity and little impact on mortality. Tests with improved sensitivity for the detection of early stage disease as well as tests with a higher specificity are required in primary care to ensure that all women are triaged appropriately and in a timely manor

following presentation with symptoms. To achieve an improvement in both sensitivity and specificity it is likely that a combination of tests will be required including both protein markers and potentially ctDNA.

Earlier detection of HGSOC poses a number of challenges relating to the unique biology and genomic features of the disease. Methods for the detection of ctDNA in early stage cancers are now being developed. However, these are currently limited by complex library preparation methods, complex bioinformatics analysis, the requirement for deep sequencing, and ultimately are not cost-effective for routine use as a diagnostic test. I am going to investigate alternative opportunities to increase the specificity and sensitivity of detection of HGSOC using plasma ctDNA and cervical sampling. The long term aim is to investigate the combination of multiple biomarkers to improve earlier detection of OC.

2.9 Project aims

1. To determine if ctDNA can be detected in plasma samples collected from women with newly diagnosed OC.
2. To develop methods with increased sensitivity for the detection of ctDNA in women with newly diagnosed HGSOC.
3. To establish whether tumour DNA can be detected in cervical samples.

Methods

3.1 Study populations

3.1.1 CTCR-OV04

CTCR-OV04 (Molecular analysis of response to treatment in ovarian cancer) (REC 08/H0306/61) is an ongoing single centre observational study at Addenbrookes Hospital recruiting all patients with known or suspected ovarian, fallopian tube or primary peritoneal cancer. To date over 900 patients have been recruited and plasma, tissue, and ascites samples have been banked at multiple treatment time points.

3.1.2 UKOPS

UK Ovarian Cancer Population Study (UKOPS) is a biobanking study that recruited from 10 centres across England, Wales, and Northern Ireland between 2006 and 2008. Inclusion criteria were:

1. Women with an adnexal mass suspicious of OC about to undergo surgery.
2. Women with a confirmed diagnosis of primary invasive or borderline OC.
3. Women with a probable diagnosis of primary invasive OC not undergoing surgery.
4. Women with a possible benign or borderline adnexal mass about to undergo surgery.

5. Apparently healthy women recruited from the multimodal arm of UKCTOCS when they attended for annual screening.

In total this study banked samples from 1,200 epithelial OCs (300 of which were recruited pre-operatively), 500 benign or borderline neoplasms, and 2,800 healthy controls. Through a collaboration with Usha Menon at UCL we have had access to formalin fixed paraffin embedded (FFPE) tissue and matched plasma samples obtained prior to the diagnosis of epithelial OC.

3.1.3 CTCR-OV05

I am the principle investigator for CTCR-OV05 (Genetic biomarkers for gynaecological conditions) (REC 14/EE/1248). I designed and developed study documents for CTCR-OV05, a single centre observational study recruiting all women referred to Addenbrookes Hospital or undergoing surgery at Addenbrookes Hospital for suspected OC. Between June 2015 and September 2017 111 women were recruited to CTCR-OV05. Plasma, tissue, and cervical samples were collected from participants at multiple treatment time points. In total I collected 94 cervical cytology samples from women recruited to CTCR-OV05. See appendix 12.1 for CTCR-OV05 study documents.

3.1.4 CS01

I am the principle investigator for CS01 (Molecular analysis of cervical smear samples) (REC 15/LO/0635) an observational study designed to collect all routine cervical smear samples from West Anglia Pathology Service Cervical Cytology Laboratory showing non-cervical glandular abnormalities. The study has been developed to assess if molecular analysis of routine cervical smear samples is a sensitive enough method for use in the diagnosis of OC. See appendix 12.2 for CS01 study documents. In total 76 cervical cytology samples were received and analysed between December 2015 and March 2018.

3.2 Sample collections

3.2.1 FFPE samples

FFPE blocks were obtained from the pathology archives at Addenbrookes Hospital for samples collected through CTCR-OV04, and from the pathology archives at UCL for samples collected through UKOPS. FFPE blocks were reviewed by the study pathologist and the block most suitable for DNA extraction (largest area of highest cellularity tumour) was selected. Based on the tumour area and cellularity between 3–15 8 µm sections on plain glass slides were cut by Addenbrookes Tissue Bank or CRUK CI histopathology core. Tumour regions were marked by the study pathologist on an adjacent hematoxylin and eosin (H&E) section.

3.2.2 Blood samples

Banked plasma samples were retrieved from the Cambridge Cancer Centre Blood Processing Laboratory (CCCBPL) from patients recruited to CTCR-OV04. Patients recruited to the referral cohort of CTCR-OV05 had 9 ml blood collected in an ethylenediaminetetraacetic acid (EDTA) tube for plasma and buffy coat isolation, 2.5 ml blood collected in an EDTA tube for whole blood collection and 7 ml blood collected in a serum tube. Patients recruited to the surgery cohort of CTCR-OV05 had the schedule of blood collection outlined in table 3.1.

EDTA tubes for plasma isolation were kept at room temperature and processed within one hour of collection or kept at 4 °C and processed within six hours of collection. Samples collected in EDTA tubes were centrifuged at 820 g for 10 minutes, 1.5 ml aliquots of plasma were subsequently centrifuged at 14,000 rpm for 10 minutes. The supernatant was transferred to sterile 1.5 ml screw cap tubes and stored at −80 °C before DNA extraction. Following the initial centrifugation step the buffy coat layer was removed and transferred to sterile 1.5 ml screw cap tube and stored at −80 °C. Serum tubes are left to stand at room temperature to allow the blood to clot. Samples are subsequently centrifuged at 3,000 rpm for 10 minutes. 1.5 ml aliquots are then transferred to sterile 1.5 ml screw cap tubes and stored at -80°C. Whole blood is stored in 1 ml aliquots in sterile 1.5 ml screw cap tubes at −80 °C.

Time point	Samples Collected
Start of Operation	9.0 ml EDTA for plasma and buffy coat
	2.5 ml EDTA for whole blood
	7.0 ml serum
End of Operation	9.0 ml EDTA for plasma
2 hours post-operative	9.0 ml EDTA for plasma
Day 1 post-operative	9.0 ml EDTA for plasma
	7.0 ml serum
Day 2 post-operative and daily whilst inpatient	9.0 ml EDTA for plasma

Table 3.1: Schedule of blood collection for surgery cohort of CTCR-OV05

3.2.3 Cervical samples

Patients recruited to both the referral and surgery cohort of CTCR-OV05 had a cervical mucous aspirate (CMA) and cervical cytology sample collected at the initial sampling time point. CMAs were collected using a cervical mucous sampling device and transferred to a cryovial for storage at −80 °C. Cervical cytology samples were collected by routine liquid based cytology method using a brush and transferred to 20 ml PreservCyt fixative solution. Samples were stored at room temperature and processed for DNA extraction within 3 weeks of collection.

3.3 Cell culture

Frozen cell pellets for Hela, OVCAR3 and SKOV3 were thawed. 5 ml of the relevant media was added to each cell pellet and the sample was centrifuged at 1,000 rpm for 3 minutes at room temperature. The media was aspirated to leave the cell pellet which was resuspended in 7 ml media and incubated at 37 °C with 5% CO_2. Samples were cultured to 90% confluence and then split by washing with warm phosphate buffered saline (PBS), trypsinising with 0.05% trypsin, neutralising with an equal volume of media, centrifugation at 1,300 rpm for 3 minutes, aspirating the media and resuspending in the appropriate volume of media for the number of cells present and size of flask being using for replating. Cells were cultured by this processing until sufficient cells were present for subsequent experiments. Cells were counted using Vi-Cell according to manufacturers instructions.

3.4 Extraction of DNA

3.4.1 Extraction of DNA from FFPE

DNA was extracted from FFPE sections using scrape macro dissection followed by the QIAamp DNA Micro Kit (Qiagen). Firstly tumour areas marked by the pathologist were macro dissected by scraping with a scalpel into a 1.5 ml micro-centrifuge tube containing 1 ml xylene to removed the paraffin. Samples were mixed by vortexing for 30 seconds and centrifuged for 5 minutes at full speed. The supernatant was removed and 1 ml 100% ethanol added to remove the remaining xylene. The sample was mixed by vortexing for 30 seconds and centrifuged for 5 minutes at full speed. The supernatant was removed and the pellet dried for 10 minutes at 30 °C. The pellet was resuspended in 180 µl ATL buffer, mixed by vortexing and spun down briefly. The sample was incubated at 95 °C for 15 minutes to repair the cross-linking damage caused by the formalin. The sample was briefly centrifuged and 20 µl Proteinase K added when the sample reached room temperature. The sample was mixed by vortexing, centrifuged briefly and incubated at 56 °C for 12–24 hours mixing at 500 rpm. The sample was incubated at 90 °C for 1 hour to reverse the formaldehyde DNA modification. The sample was centrifuged briefly, 4 µl RNase A (100 mg/ml) added and incubated at room temperature for 2 minutes. 200 µl AL buffer and 200 µl 100% ethanol were added, mixed by vortexing for 30 seconds and briefly centrifuged. The sample was transferred to a QIAamp MinElute column and centrifuged for 1 minute at 6,000 g. The flow through was discarded and the column placed in a clean collection tube. 500 µl AW1 buffer was added and centrifuged for 1 minute at 6,000 g. The flow through was discarded and the column placed in a clean collection tube. 500 µl AW2 buffer was added and centrifuged for 1 minute at 6,000 g. The flow through was discarded and the column placed in a clean collection tube and centrifuged for 3 minutes at full speed. The QIAmp Min Elute column was placed in a 1.5 ml micro-centrifuge tube and 20 µl ATE buffer added to the membrane, incubated at room temperature for 5 minutes and centrifuged for 1 minute at full speed. A further 20 µl ATE buffer was added to the membrane, incubated at room temperature for 5 minutes and centrifuged for 1 minute at full speed. The sample was stored at −80 °C.

3.4.2 Extraction of DNA from cervical cytology

DNA was extracted from cervical cytology samples within 3 weeks of collection using the QIAamp DNA Micro Kit (Qiagen). Samples were transferred to a 50 ml falcon tube and centrifuged at full speed for 5 minutes. The fixative was aspirated and the cell pellet re-suspended in 1 ml 100% ethanol. The protocol for DNA extraction from FFPE samples was then followed (see section 3.4.1, page 37).

3.4.3 Extraction of DNA from cervical mucous

DNA was extracted from CMA samples using the QIAamp DNA Micro Kit (Qiagen). 180 µl ATL buffer and 20 µl Proteinase K were added to the sample and transferred to a 1.5 ml micro-centrifuge tube. The sample was incubated at 56 °C for 2 hours. 4 µl of RNase A was added at room temperature and incubated for 2 minutes. 200 µl AL buffer and 200 µl 100% ethanol were added to the sample, mixed by vortexing and briefly centrifuged. The sample was transferred to a QIAamp MinElute column and centrifuged for 1 minute at full speed. The flow through was discarded and the column placed in a clean collection tube. 500 µl AW1 buffer was added and centrifuged for 1 minute at full speed. The flow through was discarded and the column placed in a clean collection tube. 500 µl AW2 buffer was added and centrifuged for 1 minute at full speed. The flow through was discarded and the column placed in a clean collection tube and centrifuged for 3 minutes at full speed. The QIAamp MinElute column was placed in a 1.5 ml micro-centrifuge tube and 20 µl ATE buffer added, incubated at room temperature for 5 minutes and centrifuged for 1 minute at full speed. A further 20 µl ATE buffer was added, incubated at room temperature for 5 minutes and centrifuged for 1 minute at full speed. The sample was stored at −80 °C.

3.4.4 Extraction of DNA from cultured cells

DNA was extracted from cultured cells using the DNeasy Blood & Tissue kit. Cells were centrifuged for 5 minutes at 300 g. The pellet was resuspended in 200 µl PBS and 20 µl Proteinase K added. 200 µl AL buffer and 200 µl 100% ethanol were added, mixed by vortexing and incubated at 50 °C for 10 minutes. The sample was placed in a DNeasy Mini spin column placed in a 2 ml collection tube and centrifuged at 6,000 g for 1 minute. The flow through was discarded and the column placed in a new collection tube. 500 µl AW1 buffer

was added and centrifuged for 1 minute at 6,000 g. The flow through was discarded and the column placed in a new collection tube. 500 µl AW2 buffer was added and centrifuged for 3 minute at 20,000 g. The spin column was placed in a clean 1.5 ml micro-centrifuge tube and 200 µl AE buffer applied to the membrane. The column was incubated at room temperature for 1 minute and the centrifuged for 1 minute at 6,000 g. The sample was stored at −20 °C.

3.4.5 Extraction of DNA from plasma

Circulating nucleic acids were extracted from between 1.2–4 ml of plasma with either the QIAvac 24 plus vacuum manifold and the QIAamp Circulating Nucleic Acid kit (Qiagen) or the QIAsymphony (Qiagen).

For the QIAvac extraction samples was adjusted to final volume using PBS. Proteinase K and ACL buffer containing 5.6 µg of carrier RNA were added to the sample and mixed by vortexing for 30 seconds (Table 3.2). The sample was briefly centrifuged and ACB buffer added (Table 3.2). The sample was mixed by pulse vortexing for 30 seconds. The sample was incubated on ice for 5 minutes. The lysate was applied to the QIAamp Mini column. A vacuum was applied to draw lysate through the column. 600 µl ACW1 buffer was applied to the column and the pump used to draw the sample through the column. 750 µl ACW2 buffer was applied to the column and the pump used to draw the sample through the column. 750 µl 100% ethanol was applied to the column and the pump used to draw the sample through the column. The column was placed in a clean 2 ml collection tube and centrifuged at full speed for 3 minutes. The column was placed in a new 2 ml collection tube and incubated at 56 °C with the lid open for 10 minutes to dry the membrane. The column was placed in a clean 1.5 ml micro-centrifuge tube, 50 µl AVE buffer applied to the membrane and incubated at room temperature for 3 minutes. The column was centrifuged for 1 minute at 20,000 g. The eluate was reapplied to the membrane, incubated at room temperature for 3 minutes and centrifuged at 20,000 g for 1 minute. The sample was stored at −80 °C.

The QIAsymphony extractions were performed by the Cancer Molecular Diagnostics Laboratory (CMDL) according to fixed protocols.

Volume of sample (ml)	Proteinase K (µl)	ACL buffer (ml)	ACB buffer (ml)
1	100	0.8	1.8
2	200	1.6	3.6
3	300	2.4	5.4

Table 3.2: Volume of reagents added depending on initial starting volume of sample.

3.4.6 Quantification of DNA

DNA was quantified using the Qubit dsDNA BR assay kit or HS assay kit. The Qubit working solution was prepared by diluting the relevant dsDNA reagent 1:200 in the relevant dsDNA buffer. Appropriate standards were prepared using 190 µl of working solution and 10 µl standard. Samples were prepared using 199 µl working solution and 1 µl sample. Standards and samples were mixed by vortexing for 2–3 seconds. The standards and samples were read using the Qubit Fluorometer.

3.5 Library preparation

3.5.1 TAm-Seq

The TAm-Seq library preparation method has previously been published (Forshew et al. 2012).

For FFPE and buffy coat samples TAm-Seq was performed using 5 µl of DNA at a concentration of 10 ng/µl or from 5 µl of DNA for low concentration samples. Owing to the low concentrations of plasma and serum samples TAm-Seq was performed using 5 µl of DNA. All samples are run in duplicate.

Firstly a pre-amplification step was performed by combining 5 µl of sample with 5 µl of pre-amplification master mix (see below). The primer mix is specific to the primer panel for specific experiments (see appendices 12.3, 12.4, 12.5 for primer panel details).

Pre-amplification master mix	µl
10× FastStart High Fidelity Reaction Buffer with $MgCl_2$ (Roche)	1.00
25 mM $MgCl_2$ (Roche)	1.08
DMSO (Roche)	0.50
10 mM PCR Grade Nucleotide Mix (Roche)	0.20
5 U/µl FastStart High Fidelity Enzyme Blend (Roche)	0.10
Primer mix (made by combining 1 µl of each primer (100 µM stock concentration))	0.48
H_2O	1.64

The following pre-amplification protocol was then run.

95 °C	10 min	
95 °C	15 sec	15 cycles
60 °C	4 min	

5 µl of each pre-amplified sample was added to 2 µl of ExoSAP-IT and incubated at 37 °C for 15 minutes followed by 80 °C for 15 minutes.

To dilute the pre-amplification products 18 µl PCR certified water was added to each sample. 1 µl of the diluted pre-amplification product was combined with 4 µl of pre-sample master mix solution (see below).

Pre-Sample Master Mix	µl
10× FastStart High Fidelity Reaction Buffer without $MgCl_2$ (Roche)	0.50
25 mM $MgCl_2$ (Roche)	0.90
DMSO (Roche)	0.25
10 mM Grade Nucleotide Mix (Roche)	0.10
5 U/µl FastStart High Fidelity Enzyme Blend (Roche)	0.05
20× Access Array Loading Reagent (Fluidigm)	0.25
H_2O	1.95

A 48×48 Access Array IFC was primed, 4 µl of 20× primer solution added to each of the primer inlets and 4 µl of sample plus master mix added to each of the sample inlets. The Access Array was loaded using the IFC Controller AX and then placed in the Fluidigm

Thermal Cycler and the AA 48×48 Standard v1 protocol run. After the thermal cycling had finished 2 µl 1× Access Array Harvest Reagent was added to each of the sample inlets. Samples were harvested using the IFC Controller AX with approximately 10 µl of harvested PCR product retrieved per sample. 1 µl of PCR product was diluted in 149 µl of PCR certified water. 1 µl of diluted harvested PCR product was added to 7 µl of pre-sample master mix (see below) and 2 µl of the relevant Access Array Barcode Library for the Illumina Genome Analyser (2 µM).

Pre-Sample Master Mix	µl
10× FastStart High Fidelity Reaction Buffer without $MgCl_2$ (Roche)	1.00
25 mM $MgCl_2$ (Roche)	1.80
DMSO (Roche)	0.50
10 mM PCR Grade Nucleotide Mix (Roche)	0.20
5 U/µl FastStart High Fidelity Enzyme Blend (Roche)	0.10
H_2O	3.40

The following PCR protocol was run.

95 °C	10 min	
95 °C	15 sec	
60 °C	30 sec	15 cycles
72 °C	1 min	
72 °C	3 min	

1 µl of each PCR product was pooled to create the library (48 µl total). The library was purified by adding 86.4 µl (ratio of 1:1.8) AMPure XP magnetic beads. The bead mix solution was mixed by vortexing and placed on a magnet for 5–10 minutes to separate the beads from the solution. The supernatant was aspirated and discarded. The beads were washed twice using 200 µl of 70% ethanol. The beads were left to air dry for 20–30 minutes. 20 µl PCR certified water was added to the beads and mixed by vortexing. The bead mix solution was placed on the magnet for 5–10 minutes. The clear solution was collected and stored at −20 °C. The library was quantified either using the Bioanalyzer Agilent DNA 1000 kit or Tapestation Agilent D1000 ScreenTape System according to manufacturers protocol. Libraries were diluted to 10 nM , pooled and submitted to the genomics core for sequencing using Illumina MiSeq, HiSeq 2500 or HiSeq 4000.

3.5.2 Whole genome library preparation

Whole genome libraries were prepared using Rubicon ThruPlex DNA-Seq kit. A single replicate for each sample was run using this protocol.

For FFPE samples DNA was first sheared using the Covaris LE220. 75 ng of DNA in 15 µl was placed in a microTUBE-15 AFA bead strip V2 well and spun down. The settings used are listed below. The same protocol was used for shearing DNA from cervical cytology samples but the treatment time was extended to 180 seconds.

Target BP, peak (bp)	200–250
Peak Incident Power (W)	180
Duty Factor (%)	30
Cycles per Burst	50
Treatment time (sec)	120
Temperature (°C)	20
Water level	4
Y Dither (mm)	5
X-Y Dither Speed (mm/ sec)	20
X Dither (mm)	0
X-Y Dwell (sec)	0

10 µl of sheared DNA or 10µl of plasma or serum DNA was then combined with 2 µl of template preparation buffer and 1 µl of template preparation enzyme. The samples were incubated at 22 °C for 25 minutes followed by 55 °C for 20 minutes. 1 µl of library synthesis buffer and 1 µl of library synthesis enzyme were added to each sample and incubated at 22 °C for 40 minutes. 25 µl of library amplification buffer, 1 µl of library amplification enzyme and 4 µl of nuclease free water was added to each sample with 5 µl of the appropriate indexing reagent. The library amplification reaction was performed with the number of PCR cycles varying depending on the DNA input according to manufacturers protocol.

Library amplification reaction		
72 °C	3 min	
85 °C	2 min	
98 °C	2 min	
98 °C	20 sec	
67 °C	20 sec	4 cycles
72 °C	40 sec	
98 °C	20 sec	n cycles
72 °C	50 sec	
4 °C	hold	

Libraries were purified using AMPure XP magnetic beads (1:1 ratio). The bead mix solution was mixed by vortexing for 2 minutes and incubated at room temperature for 1 minutes. The sample was placed on a magnet for 5 minutes to separate the beads from the solution. The supernatant was aspirated and discarded. The beads were washed twice using 200 µl of 80% ethanol. The beads were left to air dry for 20 minutes. 20 µl of library dilution buffer was added to the beads, mixed by vortexing for 2 minutes and incubated at room temperature for 2 minutes. The bead mix solution was placed on the magnet for 5 minutes. The supernatant was collected and stored at −20 °C. The libraries were quantified using the KAPPA library quantification kit or the Tapestation Agilent D5000 ScreenTape System according to manufacturers protocol. Libraries were diluted to 10 nM, pooled and submitted to the genomics core for sequencing using Illumina HiSeq 4000 generating 150 bp paired-end reads.

3.5.3 Digital PCR

Firstly the assay mix was prepared.

20× assay mix	
F primer (100 µM)	4.50 µl
R primer (100 µM)	4.50 µl
Probe (100 µM)	1.25 µl
H_2O	14.75 µl

Depending on the number of samples to be run either the 37K chip or the 12.765 chip was used. Two replicates were performed for all samples following this protocol.

For the 37K chip firstly the chip was primed. Next the PCR master mix was prepared.

Master Mix for 37K chip (volume per inlet)	µl
20× PCR Master Mix	3.0
20× GE Sampling Loading Reagent	0.6
20× Gene specific assay 1	0.3
20× gene specific assay 2 or nuclease free water if using 1 probe	0.3

The master mix was mixed by vortexing and 4.2 µl aliquoted into PCR strips for the number of reactions required. 200 µl 1× GE sample loading reagent was prepared by dilution 10 µl 20× GE sample loading reagent in 190 µl nuclease free water. 2.3 µl of DNA was aliquoted for each reaction into PCR strips. The DNA was denatured at 95 °C for 1 minute and then transferred to ice for 1 minute. The samples were briefly centrifuged. 1.8 µl sample was transferred to the PCR strips containing the master mix. The samples were mixed by vortexing and briefly centrifuged. Once primed the chip was removed from the IFC Controller and 10 µl 1× GE sample loading reagent added to all hydration inlets. 4.6 µl sample master-mix was added to each sample inlet. Bubbles were removed from all inlets before loading. The chip was placed in the IFC Controller MX and the load (167×) script run. Once complete the chip was removed from the controller and sticky tape used to remove any dust/debris from the chip surface. The chip was placed in the Biomark and the PCR protocol run (Figure 3.1).

Master Mix for 12.765 chip (volume per inlet)	µl
20× PCR Master Mix	5.0
20× GE Sampling Loading Reagent	0.5
20× Gene specific assay 1	1.0
Nuclease free water	0.5

For the 12.765 chip the master mix was prepared first. The master mix was mixed by vortexing and 7 µl aliquoted into PCR strips for the number of reactions required. The chip was primed using the IFC Controller MX. 3.5 µl of DNA was then added to a PCR strip and heat denatured at 95 °C for 1 minute. The samples were then transferred to ice for at least

1 minute and then briefly centrifuged. 3 µl of the heat denatured sample was transferred to the PCR strips containing the master mix. The samples were mixed by vortexing and briefly centrifuged. Once primed the chip was removed from the IFC Controller and 9.5 µl water to the H wells. 9.5 µl sample master-mix was added to each sample inlet. Bubbles were removed from all inlets and the samples loaded using the IFC Controller MX. Once complete the chip was removed and sticky tape used to remove any dust/debris from the chip surface. The chip was placed in the Biomark and the PCR protocol run (Figure 3.1).

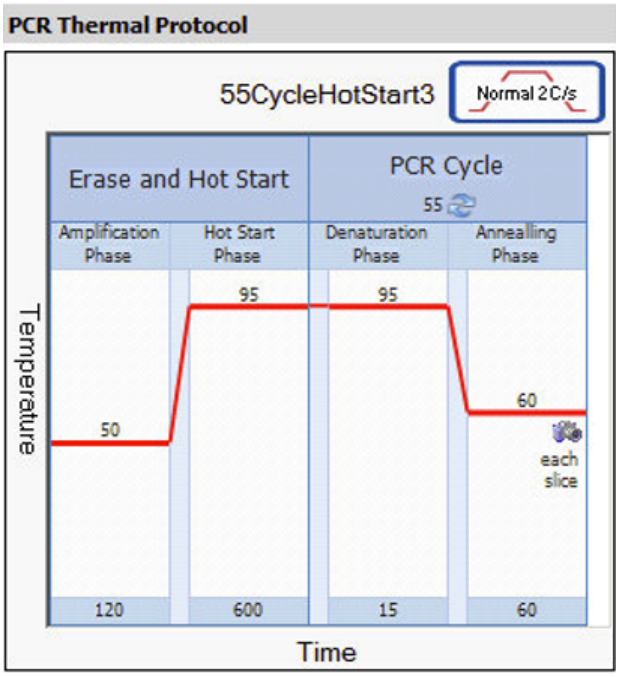


Figure 3.1: Digital PCR protocol.

3.5.4 Whole genome single stranded library preparation

Single stranded libraries were prepared using the DNA SMART ChIP-Seq Kit. A single replicate of each sample was run using this protocol. Up to 10 ng DNA in a volume of 20 µl was aliquoted into PCR tubes. The sample was incubated in a preheated hot lid thermal cycler at 94 °C for 2 minutes to convert dsDNA to ssDNA. The samples were removed and immediately placed on ice for at least 2 minutes. The tubes were briefly centrifuged. Next 3.25 µl DNA SMART buffer and 0.75 µl shrimp alkaline phosphatase were added to each sample. The samples were mixed by vortexing and briefly centrifuged. The samples were placed in a preheated thermal cycler at 37 °C for 10 minutes followed by 65 °C for 5 minutes to dephosphorylate the 3′ end of the ssDNA in preparation for T-tailing. Next 1 µl of DNA SMART T-Tailing mix and 1 µl Terminal Deoxyynucleotidyl Transferase was added to each sample.

The samples were mixed by vortexing and briefly centrifuged. The samples were placed in a preheated thermal cycler and heated to 37 °C for 20 minutes then 70 °C for 10 minutes to add poly(T) tail to the ssDNA to provide a site for the DNA SMART Pold(dA) primer. 2 µl of DNA SMART Poly(dA) Primer was added to each sample, mixed by vortexing and briefly centrifuged. The samples were incubated in a preheated, hot lid thermal cycler at 94 °C for 1 minute for the DNA SMART Poly(dA) primer to anneal to the ssDNA template. The sample was immediately placed on ice for at least 2 minutes and then centrifuged briefly. 6 µl DNA SMART buffer, 6 µl DNA SMART oligonucleotide mix and 4 µl SMARTScribe Reverse Transcriptase were added to each sample, mixed by vortexing and centrifuged briefly. The samples were placed in a preheated thermal cycler at 42 °C for 90 minutes then 70 °C for 15 minutes. 50 µl SeqAmp PCR buffer and 2 µl SeqAmp DNA polymerase was added to each sample with 2 µl of each of the relevant forward and reverse primers. The samples were mixed by vortexing and briefly centrifuged. The following PCR protocol was run with the number of cycles varying depending on the DNA input (according to manufacturers protocol).

94 °C	1 min	
98 °C	15 sec	
55 °C	15 sec	n cycles
68 °C	30 sec	
4 °C	hold	

The samples were then purified using AMPure XP magnetic beads at a ratio of 3:1 sample:beads. The library was quantified either using the Bioanalyzer Agilent DNA 1000 kit or Tapestation Agilent D5000 ScreenTape System according to the manufacturers protocol. Libraries were diluted to 10 nM, pooled and submitted to the genomics core for sequencing using Illumina MiSeq or HiSeq 4000.

3.5.5 TAm-Seq V2

DNA was diluted to 40 copies/µl based on digital PCR quantification of both ends of *TP53* using CXP039A and Amplicon023 primers and probes (see section 5.2.1, page 80). PCR master mix was prepared using the following reagents per well. The total number of replicate wells varied for each sample depending on the expected MAF and amount of material available. A minimum of 48 replicate wells were performed for each sample with more replicate

wells included if sufficient material was available and the MAF was expected to be very low (eg. <0.5%). Reactions were performed in a 384 well plate. Two primer pools were run in parallel.

TAm-Seq V2 master mix	µl
5× Q5 buffer	1.00
dNTPs (10 mM)	0.10
HS-Q5	0.05
Primers (1 µM)	2.40
H_2O	0.45
DNA	1.00

The following PCR cycling conditions were then run.

98 °C	30 sec	
98 °C	10 sec	
67 °C	15 sec	30 cycles
72 °C	5 sec	
72 °C	2 min	

Primers were tagged with a well-specific barcode to enable pooling of 1 µl from up to 48 wells (pool 1 and 2) per sample (96 µl total). Purification was performed using Pippin 2% Agarose 100–600 bp 20B cassette (HTC2010) selecting for 140–250 bp following manufacturers instructions. Master mix for barcoding was then prepared.

TAm-Seq V2 barcoding master mix	µl
5× Q5 buffer	2.00
dNTPs (10 mM)	0.20
HS-Q5	0.10
H_2O	4.20
Barcodes	2.50
DNA (direct from Pippin)	1.00

The following PCR protocol was run.

98 °C	30 sec	
98 °C	10 sec	
67 °C	30 sec	15 cycles
72 °C	8 sec	
72 °C	2 min	

9 µl of PCR product was added to 11 µl of water and 28 µl of AMPure XP magnetic beads. The bead mix solution was mixed by vortexing and placed on a magnet for 5–10 minutes to separate the beads from the solution. The supernatant was aspirated and discarded. The beads were washed twice with 200 µl of 70% ethanol and then left to air dry for up to 5 minutes. 20 µl PCR certified water was added to the beads and mixed by vortexing. The bead mix solution was placed on the magnet for 5–10 minutes. The supernatant was collected and stored at −20 °C. The libraries were quantified using the Tapestation Agilent D1000 ScreenTape System according to the manufacturers protocol. Libraries were diluted to 10 nM , pooled and submitted to the genomics core for sequencing using Illumina MiSeq or HiSeq 4000.

3.6 In vitro size selection

In-vitro size selection of DNA was performed using the PippinHT (Sage Bioscience). For each sample between 10–20 ng DNA in 20 µl H_2O was combined with 5 µl of the relevant marker solution. The protocol was selected depending on the loading cassette used and the size selection parameters required. The cassette was prepared as per manufacturers guidelines. Samples were eluted in 30 µl Tris-TAPS buffer and collected.

3.7 p53 peptide microarray

The p53 peptide microarray was developed in the Wandall laboratory (Pedersen et al. 2013). Briefly, a peptide array covering the whole of p53 with 15-mer peptides and 10 amino acid overlaps was designed. Serum samples were incubated on the microarray with conjugated goat anti-human IgG. Slides were scanned and the mean value of relative fluorescence intensity recorded. Values higher than 3 standard deviations over the mean were

considered positive.

3.8 Bioinformatics analysis

3.8.1 Analysis of TAm-Seq

The TAm-Seq analysis pipeline has previously been published (Forshew et al. 2012). Briefly, reads are demultiplexed and aligned to the human reference genome (GRCh37). Aligned reads are separated into the constituent amplicons and an alignment pileup generated. Frequencies of non-reference alleles are calculated using a base quality and a mapping quality cut off. The normal and Poisson distribution of non-reference alleles is modelled and the probability of obtaining the observed non-reference alleles determined. Mutations that pass a probability cut off of 0.9995 are kept for further analysis, ranked by observed frequency and corrected for median frequency across all samples for each locus. Known SNPs are discarded. A mutation is called if it ranks highly in non-reference frequency in both replicates. The Integrative Genomics Viewer (IGV) (Thorvaldsdottir et al. 2012) is used for manual curation of all mutation calls and inspection of samples that had no mutations called by the analysis pipeline.

When performing specific variant calling for a known mutation the frequency of the non-reference allele can be read out directly from the desired locus. The background noise at each locus is calculated by determining the frequency of non-reference reads at that locus in all other samples. A mutation is only called if the frequency is larger than the maximum background value at that specific locus.

3.8.2 Analysis of shallow whole genome sequencing data

Sequence data was analysed using an in-house pipeline that carried out the following; paired end sequence reads were aligned to the human reference genome (GRCh37) using BWA-mem following the removal of contaminating adapter sequences. PCR and optical duplicates were marked using MarkDuplicates (Picard Tools) feature and these were excluded from downstream analysis along with reads of low mapping quality and supplementary alignments. When necessary, reads were down-sampled for comparison purposes. Somatic copy number aberration analysis was performed in R using a software suite

for shallow whole genome sequencing (sWGS) analysis named CNAclinic (Chandrananda 2017) as well as the QDNAseq pipeline. Sequencing reads were randomly sampled to 10 million reads per dataset and allocated into equally sized (30 Kbp) non-overlapping bins throughout the length of the genome. Read counts in each bin were corrected to account for sequence GC content and mapability, and bins overlapping 'blacklisted' regions prone to alignment artefacts (derived from the ENCODE project + 1000 Genomes database) were excluded from downstream analysis. Read counts in test samples were normalized by the counts from an identically processed healthy individual and $\log_2$ transformed to obtained copy number ratio values per genomic bin. Read counts in healthy controls were normalized by their median genome-wide count. Next, bins were segmented using both Circular Binary Segmentation and Hidden-Markov Model based algorithms, and an averaged $\log_2$R value per bin was calculated. An in-house empirical blacklist of aberrant read count regions was constructed as follows: 46 sWGS datasets from healthy plasma were used to calculate median read counts per 30 Kbp genomic bin as a function of GC content and mappability. A 2D LOESS surface is applied and the difference between the actual count and the LOESS fitted values were calculated. The median of these residual values across the 46 controls were calculated per genomic bin. Regions with median residuals greater than 4 standard deviations were blacklisted. The averaged segmental $\log_2$R values in each test sample that overlap this cfDNA blacklist were trimmed and the median absolute value was calculated. This score is defined as t-MAD or the trimmed median absolute deviation from $\log_2 R = 0$.

3.8.3 In silico size selection

Paired-end reads were generated by sequencing DNA from both ends of the fragments present in the library. The original length of the DNA can was inferred using the mapping locations of the read ends in the genome. Once alignment was complete, Samtools software was used to select paired reads that correspond to fragment lengths in a specific range.

3.8.4 Quantification of the 10 bp periodic oscillation

The amplitude of the 10 bp periodic oscillation observed in the size distribution of cfDNA samples was determined from the sWGS data as follows: the local maxima and minima in

the range 75 bp to 150 bp were calculated. The average of their positions across the samples was calculated: (minima: 84, 96, 106, 116, 126, 137, 148, and maxima: 81, 92, 102, 112, 122, 134, 144). To compute the amplitude of the oscillations below 150 bp with 10 bp periodicity we calculated the sum of the height of the maxima and subtracted the sum of the minima. The height of the peak is defined as the number of fragments with that specific length divided by the total number of fragments. To define local maxima, we selected the positions y such that y was the largest value in the interval [y-2, y+2]. The same approach was used to pick minima.

3.8.5 Analysis of TAm-Seq V2

For analysis of TAm-Seq V2 firstly well demultiplexing was performed using an in house pipeline followed by alignment using BWA-MEM. The coverage per amplicon was calculated using bedtools requiring 80% coverage of the region. Mutation calling was then performed by calculating the background frequency of non-reference alleles in a cohort of control samples. Changes above the background rate were then determined in the samples using a beta distribution to determine a probability at each base. Scores were filtered based on parameters as outlined below.

1. Minimum depth $= 20$
2. Mutant reads > 2
3. Frequency > 0.01
4. p-value < 0.01
5. Same variant called on F and R reads

All mutations called in over two wells were reviewed manually.

Detection of plasma circulating tumour DNA in women with newly diagnosed ovarian cancer

4.1 Introduction

It has previously been shown that plasma ctDNA can be detected in >80% of women with HGSOC (Parkinson et al. 2016). Importantly, for the development of ctDNA as a diagnostic biomarker, this study detected ctDNA in cases with newly diagnosed disease. However, the majority of women in this study had relapsed HGSOC, and those few women with newly diagnosed HGSOC had stage III or IV disease. The median *TP53* MAF was an order of magnitude lower in the newly diagnosed cases compared to the relapsed cases, even after adjusting for volume of disease (Parkinson et al. 2016). Therefore sensitivity of ctDNA detection in newly diagnosed disease may be challenging.

Developing a method for detection of ctDNA in a diagnostic setting is also challenging. Parkinson et al. 2016 used digital PCR which requires prior knowledge of a patients tumour specific mutation, and is therefore not applicable in a diagnostic setting. Targeted sequencing for *TP53* using TAm-Seq is a low-cost, high-throughput technique that does not require prior knowledge of a patients tumour specific mutation so therefore could be used in a diagnostic setting. However, using this method the sensitivity for detection of ctDNA in relapsed HGSOC was lower at 53% (Forshew et al. 2012). Since the original publication the TAm-Seq method has been optimised by increasing the depth of sequencing and including primers for shorter overlapping amplicons. Recently the Rosenfeld and Brenton laboratories have also developed sWGS for the detection of ctDNA in plasma samples (Tsui

et al. 2018).

In this chapter I will discuss ctDNA detection in plasma samples collected from women with newly diagnosed OC from the UKOPS cohort (see section 3.1.2) and the CTCR-OV04 cohort (see section 3.1.1) using targeted sequencing and sWGS (Figure 4.1). This will provide information about the prevalence of ctDNA in women at the time of diagnosis of OC using high-throughput, low-cost sequencing methods that do not require prior knowledge of the patients tumour specific mutation. These are therefore methods that could be directly translated to routine clinical use as a diagnostic test in the future.

To determine the specificity of ctDNA detection the OC cohorts were compared to a group of 46 healthy controls for which plasma was obtained commercially and 32 benign controls recruited through CTCR-OV04 at the time of primary surgery for suspected OC who were subsequently diagnosed with benign disease.

For the UKOPS cohort sWGS data was down-samples to 3 million reads to allow direct comparison between patients. For the CTCR-OV04 sWGS data was down-sampled to 10 million reads.

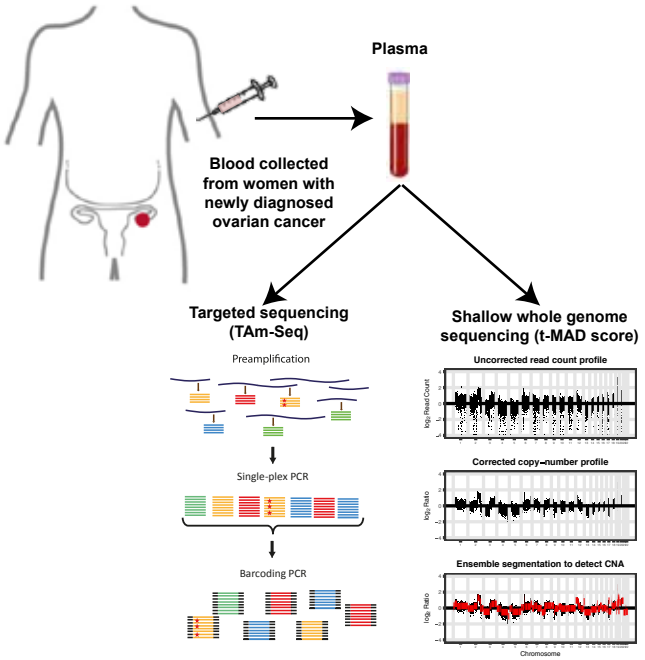


Figure 4.1: Schematic illustrating approach for detection of ctDNA in women with newly diagnosed OC.

For these cohorts the TAm-Seq analysis pipeline was run by James Morris. The t-MAD score is a quantitative score calculated from sWGS data that is used to measure the extent

of copy number aberration across the whole genome. The t-MAD score was developed by Dineika Chandrananda. See section 3.8.2, page 50 for description of t-MAD score calculation. For the cohorts discussed in this chapter sWGS analysis and calculation of the t-MAD score was performed by Dineika Chandrananda. DNA from some plasma samples from the CTCR-OV04 cohort were extracted by the QIASymphony at CMDL. Radiological assessment and tumour volume measurements were calculated by Ramona Woitek. The cases for inclusion from the UKOPS cohort were identified by Aleksandra Gentry-Maharaj and Chloe Karpinskyj from UCL. For the UKOPS cohort the histology described is the definitive histological diagnosis provided by UCL where available.

4.2 Results

4.2.1 UKOPS cohort

Fifty-four cases from the UKOPS cohort with epithelial OC were identified (see section 3.1.2, page 33). These cases all had an elevated CA 125, pre-treatment plasma sample available for ctDNA analysis and access to a FFPE tissue block. Table 4.1 summarises the demographic data for the selected cohort.

FFPE review and sequencing using TAm-Seq primer panel 1 (Appendix 12.3) was performed for the 54 OC cases (Figure 4.2, step 1). 22 cases were HGSOC and 73% of these had a *TP53* mutation identified in the FFPE tissue sample (Appendix 12.6). This detection rate is lower than expected (Kobel et al. 2016). The highly fragmented DNA found in these HGSOC with a non-detected *TP53* mutation may account for this (Figure 4.3).

Plasma DNA was extracted using the QIAvac 24 Plus vacuum manifold and the QIAamp Circulating Nucleic Acid kit (1 ml protocol) and targeted sequencing performed (Figure 4.2, step 2) using the UKOPS TAm-Seq primer panel (Appendix 12.5).

Using de novo mutation calling ctDNA was detected in 15/50 (30%) of plasma samples (Appendix 12.7) including 10/22 (45%) of the HGSOC cases. 82% matched the corresponding tumour specific FFPE mutation. Two (18%) plasma mutations did not match the corresponding tumour specific FFPE mutation and are therefore either false positive calls or subclonal mutations not identified in the FFPE sample.

As the sensitivity of ctDNA detection was low using targeted sequencing of plasma

Total number of women	54
Age at diagnosis, median (IQR) (years)	63 (57–72)
Stage at Diagnosis (number of women)	
I	22
II	5
III	17
IV	3
unknown	7
Histology (number of women)	
High grade serous	22
Endometrioid	5
Clear cell	6
Mucinous	7
Other	10
Unknown	4

Table 4.1: Summary of demographic data for UKOPS selected cohort

DNA I investigated whether whole genome amplification (Rubicon ThruPlex DNA Seq kit with 8 PCR cycles) followed by targeted sequencing could increase the sensitivity of detection (Figure 4.2, step 3). Using de novo mutation calling ctDNA was detected in 31/50 (62%) of samples (Appendix 12.7) with 1–4 mutations called in each sample. Thirty-two of these mutations did not match the corresponding tumour specific FFPE mutation. Again it is a possibility that these are low cellularity clones however, the additional PCR cycles from whole genome amplification would have increased the number of PCR errors increasing the likelihood that these are false positive calls.

When performing specific variant calling for the known tumour specific mutation the ctDNA detection rate increased to 76% for the targeted sequencing performed directly from the samples and to 70% for targeted sequencing performed on the whole genome library (Appendix 12.8).

Targeted sequencing of plasma DNA with de novo mutation calling had a low sensitivity for detection of ctDNA in this cohort and targeted sequencing of whole genome amplified plasma DNA had a high false positive detection rate. I therefore wanted to investigate whether quantitative analysis of copy number changes from sWGS data using the t-MAD score could be used to increase the sensitivity of ctDNA detection (Figure 4.2, step 4).

Using sWGS and the t-MAD score with a cut off of the highest t-MAD score calculated for 46 healthy controls the rate of ctDNA detection for the whole cohort was 41% and 56% for the HGSOC cases (Figure 4.4a).

Importantly the copy number changes seen in the plasma samples did appear to represent ctDNA in a sample as plasma profiles were similar to profiles seen for matched FFPE samples (Appendix 12.7). Secondly the unselected t-MAD score was highly correlated with the MAF determined by targeted sequencing in the HGSOC cases but not correlated in the other histological subtypes (Figure 4.4b). This may reflect the fact that HGSOC is predominately driven by copy number changes but other OC subtypes are not.

To summarise ctDNA was detected in 30% of cases using targeted sequencing and in 41% of cases using sWGS. Two cases with a t-MAD score below the maximum healthy control cut off had a mutation detected by targeted sequencing. The combination of the two assays therefore increased the detection rate to 46% in this cohort.

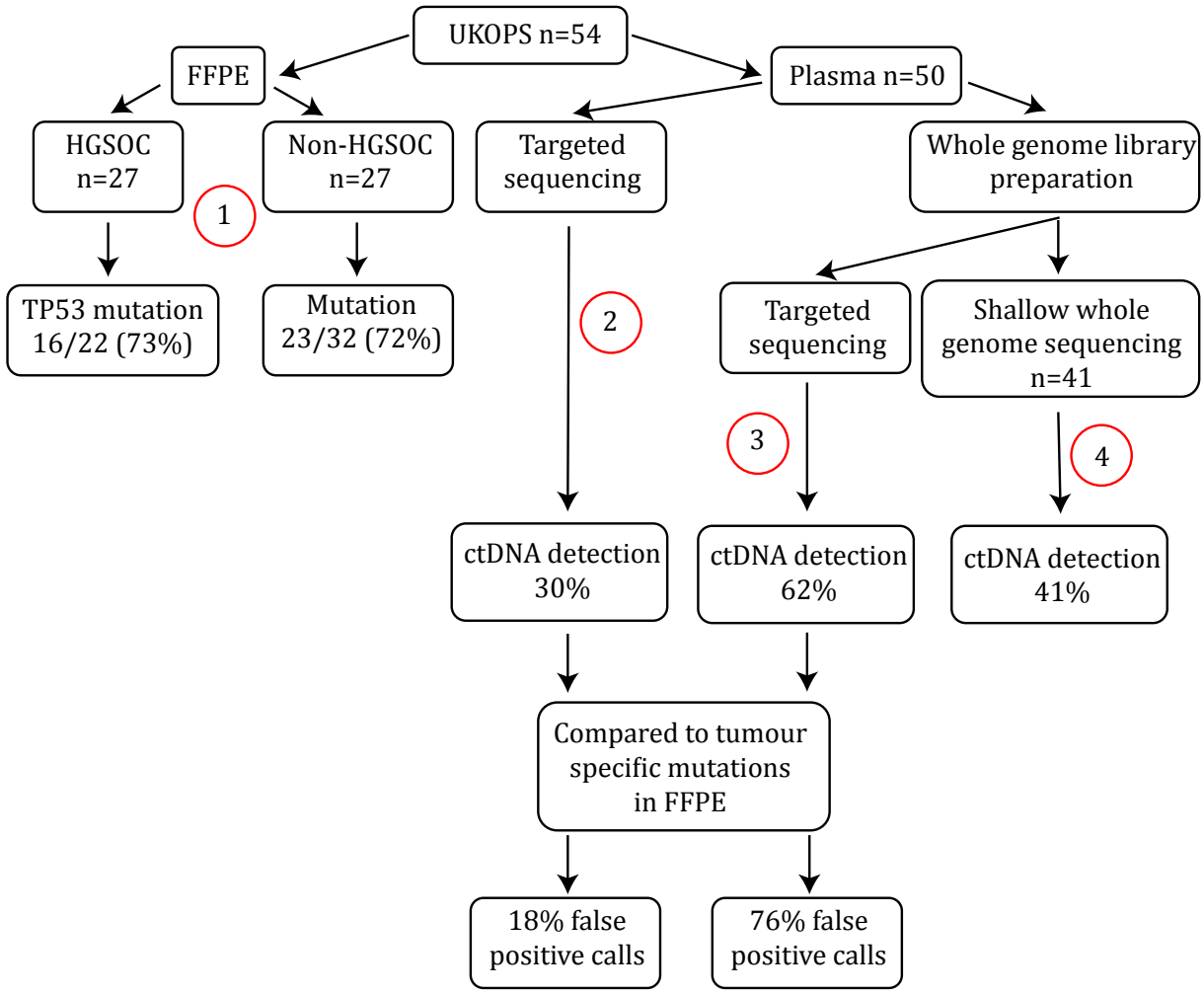


Figure 4.2: Experimental approach to UKOPS cohort, n=54. **1.** DNA was extracted from a representative FFPE block for each cases and sequenced using TAm-Seq primer panel 1. Mutations were detected in 72% of cases. **2.** DNA was extracted from <1 ml plasma for 50 of the cases. Targeted sequencing was performed using the UKOPS TAm-Seq primer panel (Appendix 12.5). **3.** Targeted sequencing of whole genome amplified samples was performed using the UKOPS TAm-Seq primer panel. Mutations called in steps 2 and 3 were compared to the patients tumour specific FFPE mutations. **4.** sWGS was performed. Reads were down-sampled to 3 million to allow comparison. Nine samples were excluded due to low read count.

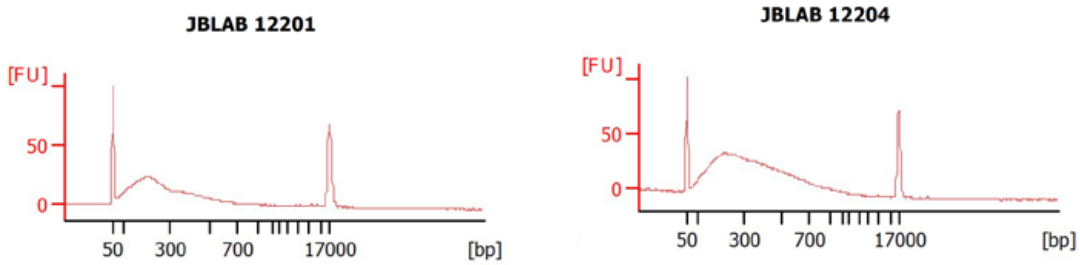


Figure 4.3: Bioanalyzer traces for two HGSOC FFPE samples in which a *TP53* mutation was not identified showing a pattern of highly fragmented DNA.

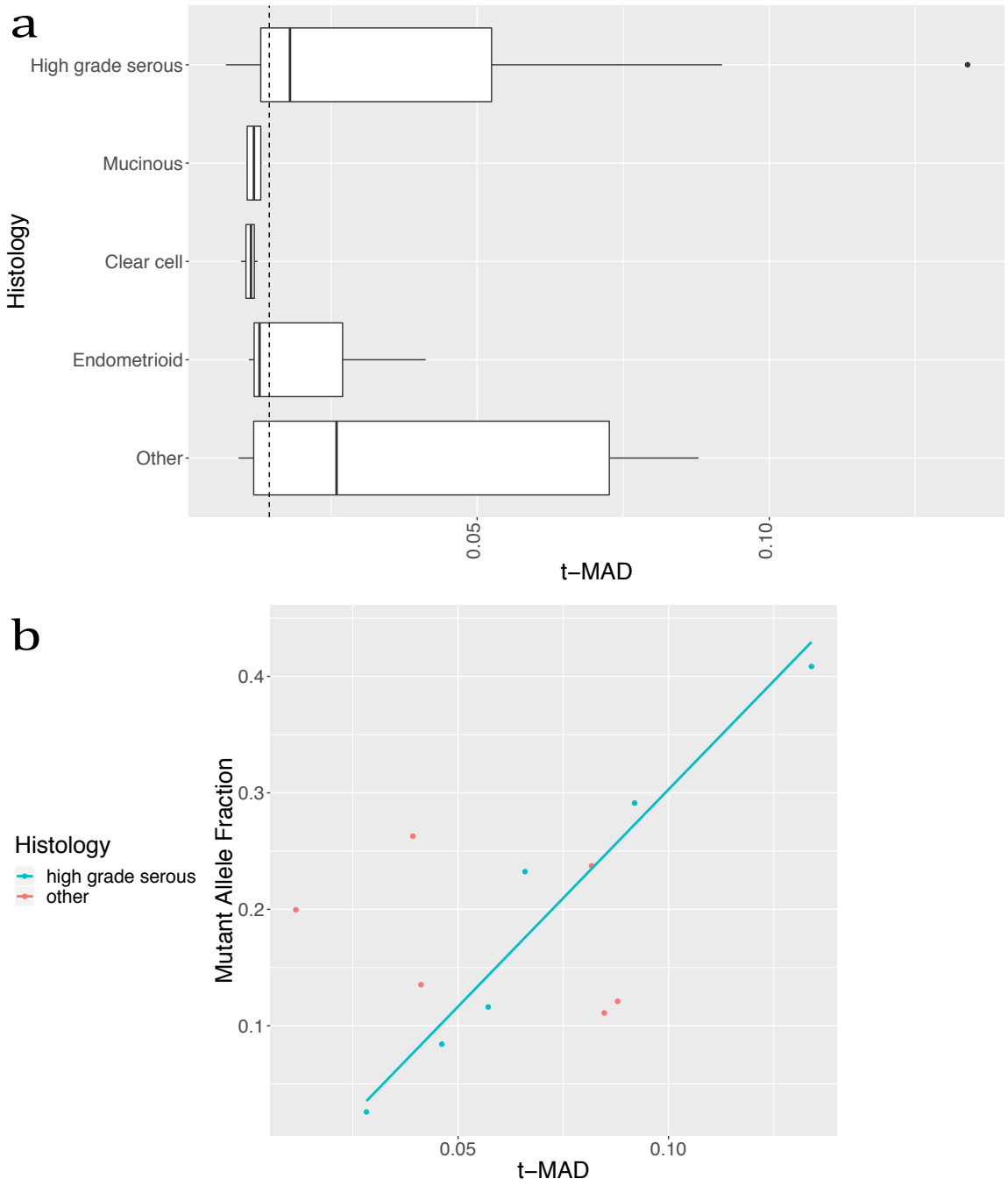


Figure 4.4: **a)** t-MAD score for all UKOPS cases calculated after down sampling to 3 million reads by histological subtype. Dashed line indicates the highest t-MAD score calculated in a cohort of 46 healthy controls. **b)** Comparison of unselected t-MAD score and MAF where ctDNA detected by targeted sequencing for HGSOC cases (correlation coefficient 0.97, $p<0.005$) and cases with other histological subtypes (correlation coefficient −0.39, $p=0.444$).

4.2.2 CTCR-OV04 cohort

The UKOPS cohort consists of samples collected from women with a range of histological subtypes of OC. HGSOC is the most common subtype of OC and accounts for the majority of mortality. It is characterised by ubiquitous *TP53* mutations (Kobel et al. 2016) and complex genomic rearrangements (Ciriello et al. 2013). I therefore decided to focus on the detection of ctDNA in women with newly diagnosed HGSOC. Cases from the CTCR-OV04 cohort were identified (see section 3.1.1). Inclusion criteria were:

1. Women undergoing primary surgery for HGSOC with a pre-treatment plasma sample available and CT scan within 28 days of the baseline plasma sample.
2. Women being treated with neo-adjuvant chemotherapy for HGSOC with plasma samples available before cycle 1, 2 and 3 of chemotherapy and a CT scan within 28 days of the baseline plasma sample.

Table 4.2 summarises the demographic data for the selected cohort.

Total number of women	156
Age at diagnosis, median (IQR) (years)	67 (58–73)
Stage at Diagnosis (number of women)	
I	9
II	3
III	93
IV	50
Treatment (number of women)	
Neo-adjuvant chemotherapy	127
Primary surgery	28

Table 4.2: Summary of demographic data for CTCR-OV04 selected cohort

I performed targeted sequencing for *TP53* and sWGS for baseline plasma samples collected from the whole cohort (Figure 4.5).

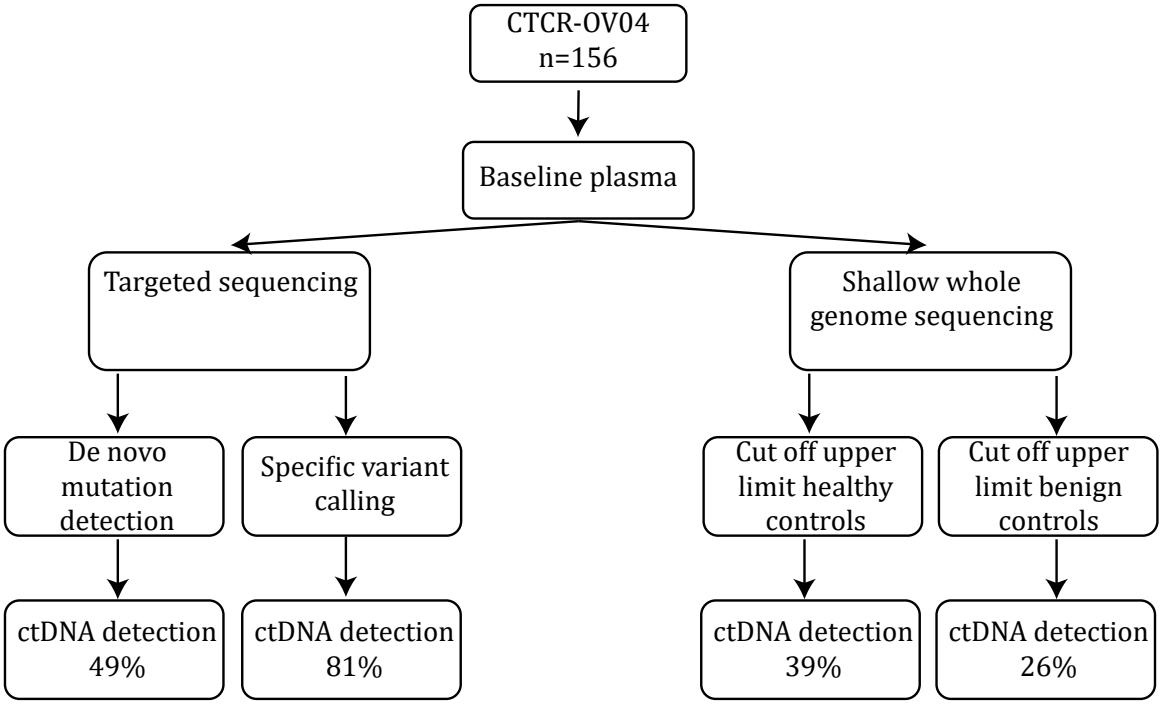


Figure 4.5: Experimental overview for CTCR-OV04 newly diagnosed cohort, n=156.

ctDNA was detected in 49% of the cohort using de novo mutation calling of targeted sequencing with TAm-Seq primer panel 10 (Figure 4.6a) with a MAF ranging from 0.64 to 0.006 (Figure 4.6c, e). See appendix 12.10 for all de novo mutation calls. Importantly ctDNA could be detected in cases with early stage disease (Figure 4.6a). Sensitivity for ctDNA detection was slightly higher in this cohort compared to the HGSOC cases in the UKOPS cohort (detection rate 45%), however remained too low to be useful as a diagnostic biomarker for HGSOC. However, specificity of ctDNA detection using this method is high with no *TP53* mutations detected in 32 plasma samples collected from women before undergoing primary surgery for suspected OC who were subsequently diagnosed with benign disease (Figure 4.6a). See appendix 12.11 for details about histological diagnosis for the benign controls.

There was no difference in the rate of ctDNA detection by menopausal status (Figure 4.7a) and no correlation between age at diagnosis and *TP53* MAF (Figure 4.7b). There was no difference in the range of pre treatment CA 125 values between women with and without a *TP53* mutation detected (Figure 4.7c). There was no correlation between CA 125 and *TP53* MAF in those women with a detectable mutation in plasma (Figure 4.7d). 12 wo-

OV04	Stage	CA 125 (U/ml)	TP53 mutation status	Mutant allele fraction
95	I	30	Not Detected	NA
133	I	18	Not Detected	NA
254	III	5	Not Detected	NA
430	III	26	Not Detected	NA
469	III	33	Not Detected	NA
547	III	28	Not Detected	NA
559	III	9	Detected	0.025
629	IV	25	Detected	0.040
709	III	11	Not Detected	NA
776	IV	27	Not Detected	NA
805	I	10	Detected	0.023
870	III	8	Detected	0.015

Table 4.3: CTCR-OV04 cases with with CA 125 <35 U/ml

men in the cohort had a CA 125 <35 U/ml (Table 4.3). Four of these had ctDNA detected by targeted sequencing.

5/53 (9%) of the mutations detected in plasma did not match the mutations detected in the corresponding FFPE sample. One of these mutations had a MAF of 0.008 which is close to the limit of detection using this method and could therefore potentially be a false positive call. Four of these mutations had a MAF of >0.01 which is above the limit of detection suggesting they could be low cellularity clonal mutations not identified in the FFPE sample as a result of heterogeneity within the tissue.

The rate of plasma ctDNA detection increased to 81% (64/79) (Figure 4.6b) when variant calling for the patients tumour specific *TP53* mutation was performed. The lowest MAF detected was 0.001 (Figure 4.6d, f). See appendix 12.12 for all specific variant calls.

In this cohort targeted sequencing had a high specificity and positive predictive value meaning that it would give very few false positive results if used in a diagnostic setting. However, due to the low sensitivity >50% of cases would be missed if this method was used in isolation as a diagnostic biomarker. sWGS was therefore performed for all basleine plasma samples. The t-MAD score was highly correlated with the *TP53* MAF (correlation coefficient of 0.92) (Figure 4.8a) providing confidence in the t-MAD score as a means of quantifying the levels of ctDNA in a sample.

Using the highest t-MAD score in a cohort of 46 healthy controls as a cut off ctDNA was detected in 61/156 (39%) of the whole cohort (Figure 4.8b). The rate of ctDNA detection decreased to 26% when using the highest t-MAD score in a cohort of 32 benign controls as a cut off, excluding one outlier (Figure 4.8b). As with targeted sequencing ctDNA could be detected in HGSOC cases with early stage disease (Figure 4.8b).

Although the sensitivity of ctDNA detection using sWGS was lower than when using targeted sequencing it is likely that the combination of the two assays will increase the sensitivity of detection. 10/79 (13%) of the cases not detected by targeted sequencing (including 1 case with stage I disease) had detectable ctDNA by sWGS using the upper limit of the healthy controls as a cut off (Figure 4.8c). Even if the upper limit of the benign controls was used as a cut off, excluding one outlier, the addition of sWGS to targeted sequencing would have detected an additional three cases (Figure 4.8c).

In relapsed OC the *TP53* MAF measured by digital PCR is highly correlated with volume of disease measured by 3-D reconstruction of CT images after exclusion of cases with ascites (Parkinson et al. 2016). Using digital PCR the limit of detection was an MAF of 0.003 and a tumour volume of 30 cm^3. When developing a diagnostic biomarker it is important to be able to detect low volume/early stage disease. I therefore investigated the limit of detection using TAm-Seq and sWGS by comparing the *TP53* MAF and t-MAD score to tumour volume.

Tumour volume was measured for 60 of the newly diagnosed CTCR-OV04 cohort. The median tumour volume was 338.5 cm^3 (IQR 152.2–670.6 cm^3) (Figure 4.9a). ctDNA was detected in a cases with a tumour volume of 25.4 cm^3 using targeted sequencing and in a case with a tumour volume of 67.9 cm^3 using sWGS (Figures 4.9b–c, 4.8d). However, there was no correlation between *TP53* MAF and tumour volume (correlation coefficient 0.12) or t-MAD score and tumour volume (correlation coefficient 0.12) even accounting for the presence of ascites or pleural effusions at baseline (Figures 4.9b–c, 4.8d). For example a case with a tumour volume of 2,093.3 cm^3 did not have a *TP53* mutation detected in plasma. Before ctDNA can be used as a diagnostic biomarker it is therefore important to better understand the relationship between newly diagnosed disease and ctDNA.

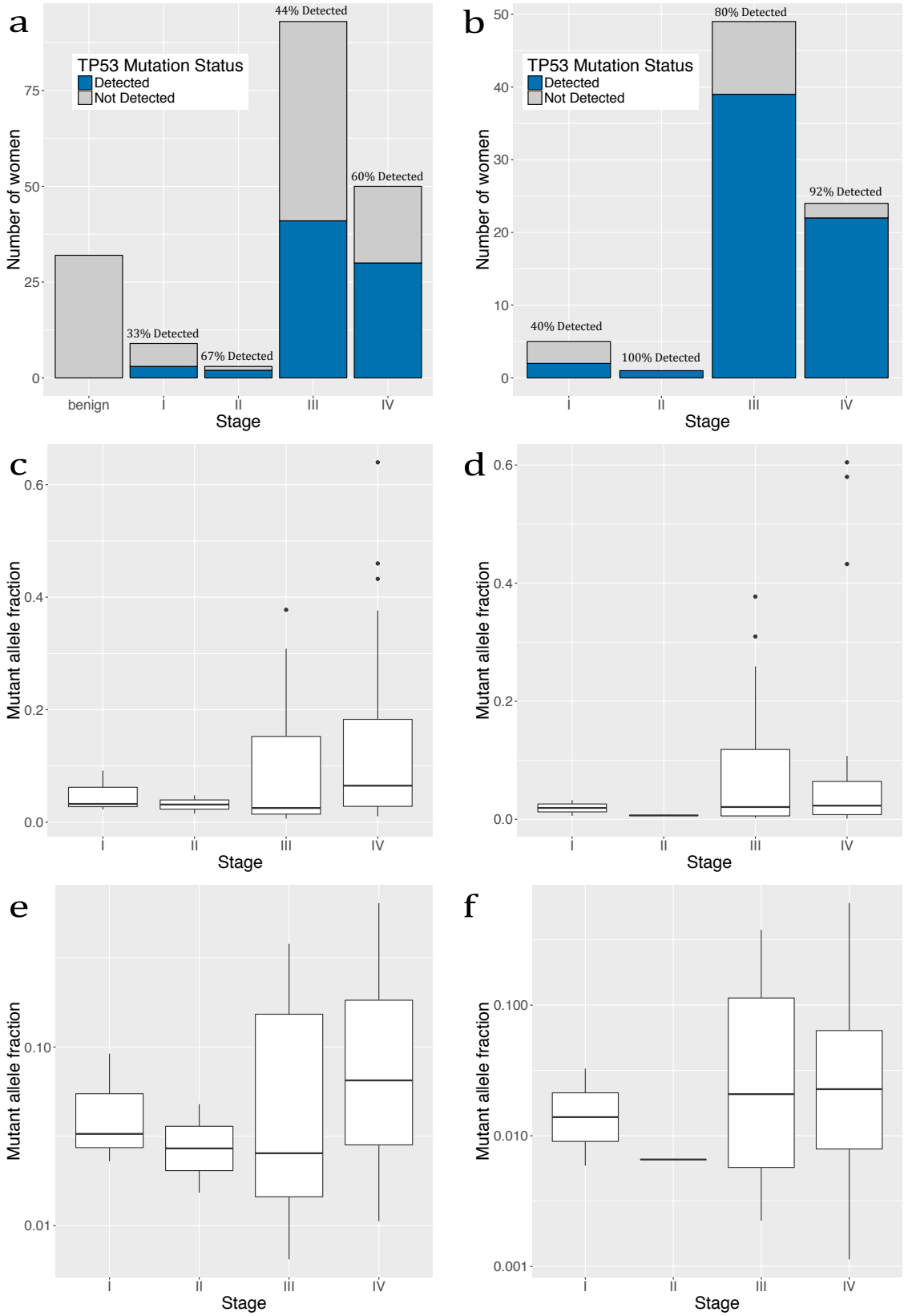


Figure 4.6: Targeted sequencing for *TP53*. **a)** Number of cases with *TP53* mutation detected by de novo mutation calling by stage of disease. **b)** Number of cases with *TP53* mutation detected by identification of patients tumour specific mutation by stage of disease. **c)** Boxplot of *TP53* MAF by de novo mutation calling in cases with ctDNA detected by stage of disease. **d)** Boxplot of *TP53* MAF by patient specific mutation calling in cases with ctDNA detected by stage of disease. **e)** Plot c with logarithmic scale. **f)** Plot d with logarithmic scale.

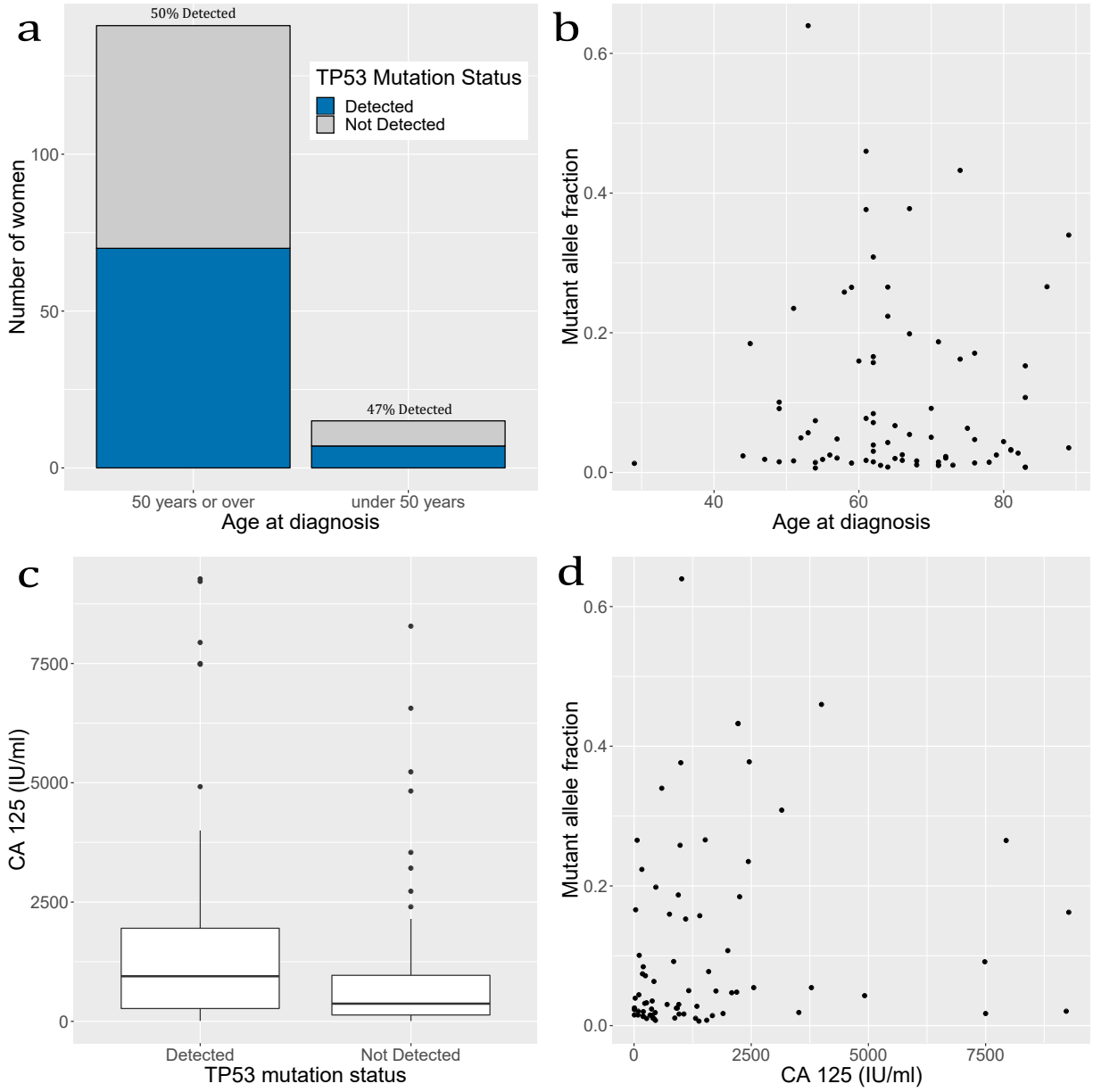


Figure 4.7: **a)** Number of cases with *TP53* mutation detected by menopausal status defined as age under 50 years (premenopausal) or age 50 years or over (postmenopausal). **b)** Correlation between age at diagnosis and *TP53* MAF from targeted sequencing (correlation coefficient -0.01, p=0.92). **c)** CA 125 pre diagnosis by *TP53* mutation status. **d)** Correlation between pre diagnosis CA 125 and *TP53* MAF from targeted sequencing (correlation coefficient 0.15,p=0.21).

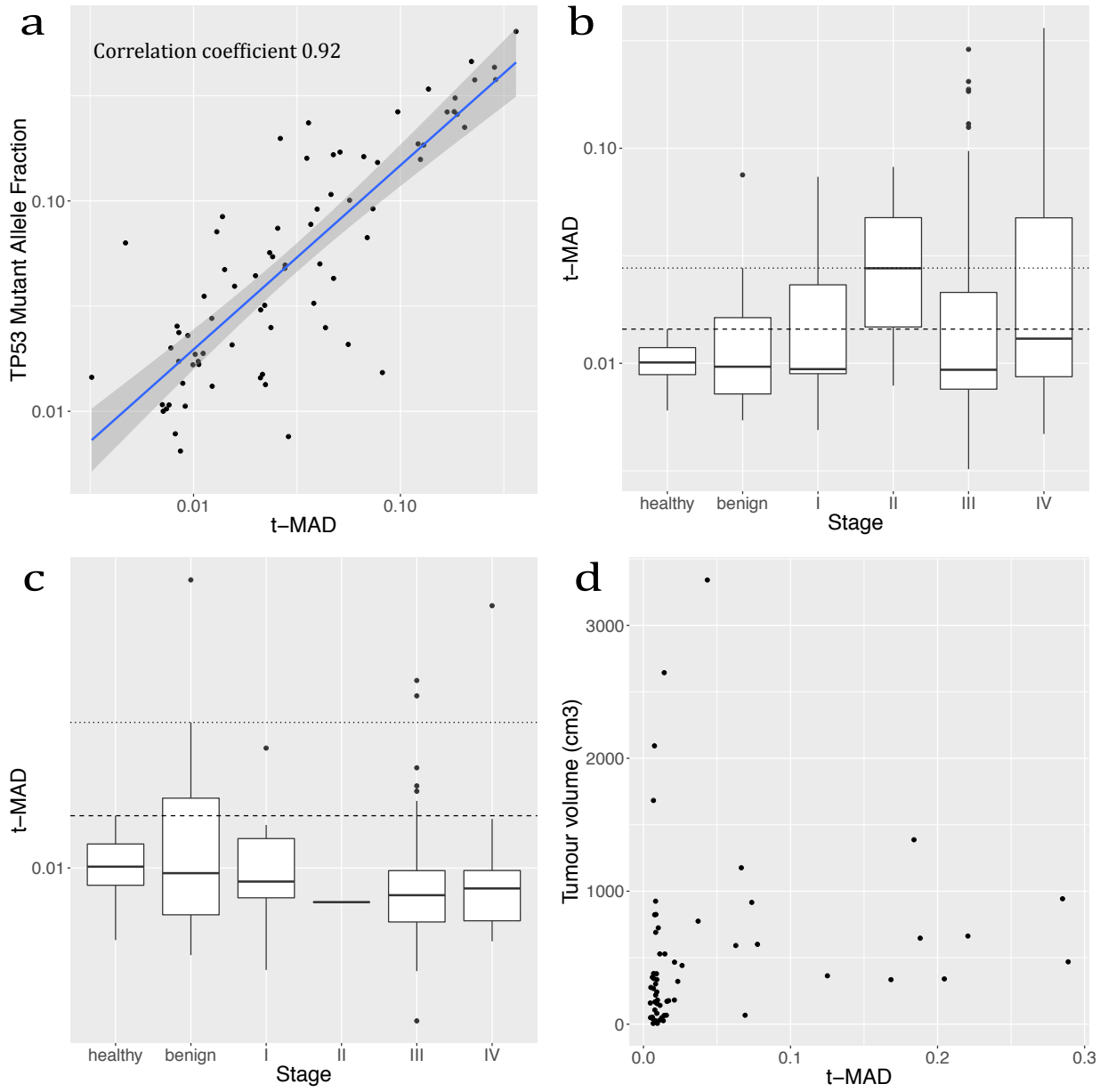


Figure 4.8: **a)** Correlation between unselected t-MAD score from sWGS data and *TP53* MAF from targeted sequencing. **b)** t-MAD score calculated from sWGS data for 156 HGSOC cases by stage, 29 benign controls and 46 healthy controls. Dashed line indicates the highest t-MAD score calculated in the cohort of healthy controls. Dotted line indicates the highest t-MAD score calculated in the cohort of benign controls excluding one outlier. **c)** t-MAD score calculated from sWGS data for 79 HGSOC cases with undetected *TP53* mutation by targeted sequencing by stage, 29 benign controls and 46 healthy controls. Dashed line indicates the highest t-MAD score calculated in the cohort of healthy controls. Dotted line indicates the highest t-MAD score calculated in the cohort of benign controls excluding one outlier. **d)** Comparison of t-MAD score with total tumour volume measured by 3-D CT reconstruction (n=60). Correlation coefficient=0.12.

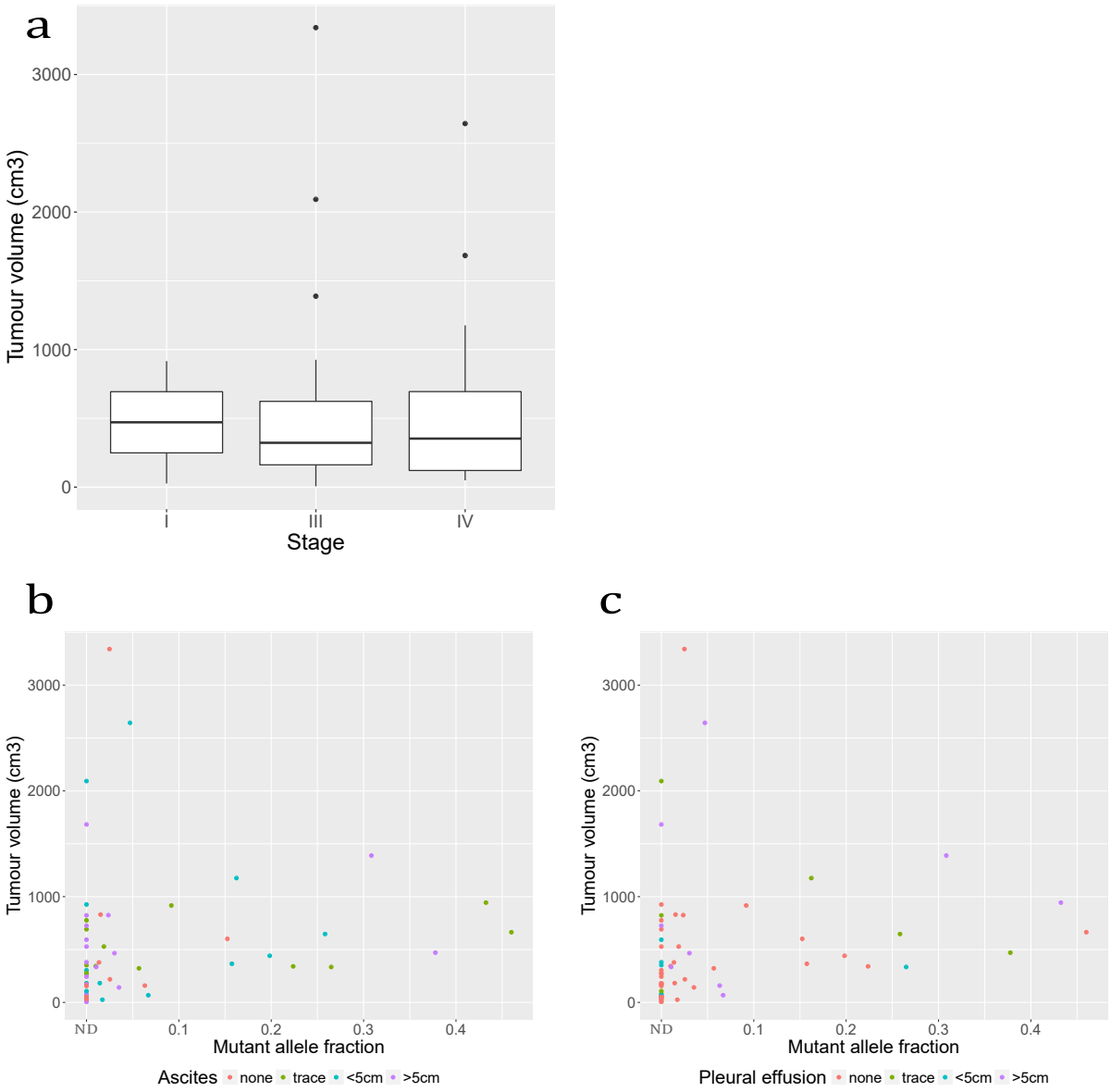


Figure 4.9: **a)** Tumour volume measured by 3-D CT reconstruction by stage of disease. **b)** Correlation of *TP53* MAF from targeted sequencing with total tumour volume including semi-quantitative assessment of ascites at baseline. ND = ctDNA not detected. **b)** Correlation of *TP53* MAF from targeted sequencing with total tumour volume including semi-quantitative assessment of pleural effusion at baseline.

4.3 Discussion

It has previously been shown that ctDNA can be detected in over 80% of women with HGSOC (Parkinson et al. 2016). The majority of women in this study had relapsed disease I therefore wanted to investigate whether ctDNA could be detected in women with newly diagnosed HGSOC as a first step towards developing a diagnostic biomarker. Parkinson et al. 2016 also used digital PCR to detect patient specific *TP53* mutations which cannot be applied in a diagnostic setting. I wanted to investigate low-cost, high throughput assays that can be directly translated to clinical use in a diagnostic setting.

Targeted sequencing for *TP53* does not require prior knowledge of the patients tumour specific mutation. It has previously been shown that ctDNA can be detected by targeted sequencing in 53% of women with relapsed HGSOC down to a MAF of 2% (Forshew et al. 2012). The compromise with this method is reduced sensitivity compared to digital PCR, Parkinson et al. 2016 were able to detect a minimum MAF of 0.9%.

Since this original publication improvements have been made to the TAm-Seq method including optimisation of the primers and increasing the depth of sequencing. In this chapter I have shown that using this optimised TAm-Seq method I can detect ctDNA with a minimum MAF of 0.6%. Despite this improvement compared to Forshew et al. 2012 I detected ctDNA in only 45% of newly diagnosed HGSOC cases from the UKOPS cohort and 49% of newly diagnosed cases from the CTCR-OV04 cohort. Importantly ctDNA was detected by targeted sequencing in 4/12 cases with a CA 125 <35 U/ml. The majority of women in the CTCR-OV04 cohort had an elevated CA 125 before diagnosis reflecting the cohort selection of women with a confirmed HGSOC diagnosis. In future studies it will be important to compare the performance of ctDNA to CA 125 and investigate the combination of the two markers for detection of OC in women who do not already have a diagnosis. It is also important to note that *TP53* mutations are not specific to HGSOC and can be found in other cancer types. However, it could be argued that detection of a *TP53* mutation in an individual within the general population would be important irrespective of the cancer of origin and referral would then be tailored depending on symptoms and results of other triage tests.

In relapsed HGSOC a strong correlation between *TP53* MAF and tumour volume was observed (Parkinson et al. 2016). It is therefore possible that the lower rate of ctDNA detection in newly diagnosed HGSOC is due to lower disease volume. However, I found no

correlation between *TP53* MAF and tumour volume in the CTCR-OV04 newly diagnosed HGSOC cohort. Parkinson et al. 2016 found a lower correlation between ctDNA levels and tumour volume in women with ascites. I did not find the presence of ascites or pleural effusions to correlate with *TP53* MAF.

It is possible that the nature and extent of release of ctDNA from newly diagnosed tumours is different to relapsed disease. This is supported by Parkinson et al. 2016 who found that the median *TP53* MAF per volume was lower (0.0008%) in women with newly diagnosed disease compared to women with relapsed disease (0.04%). In newly diagnosed non small cell lung cancer (NSCLC) only a moderate correlation between tumour volume and mean clonal VAF was found (Abbosh et al. 2017). However, in lung cancer proliferation was highly correlated with levels of ctDNA with both a high ki67 proliferation index and high FDG-PET avidity independent predictors of ctDNA detection. Although proliferation is generally high in HGSOC it is possible that cases with undetected ctDNA are cases with low tumour proliferation. There is no reliable measure of tumour proliferation across undiagnosed patients. Ongoing work is looking at cellular proliferation at multiple tumour sites in FFPE tissue collected at the time of primary surgery for HGSOC.

sWGS is another low cost, high-throughput sequencing assay. HGSOC is a cancer characterised by copy number changes which can be detected by sWGS. The t-MAD score is a new quantitative measure of copy number changes across the whole genome (unpublished). I found a high correlation between *TP53* MAF and the t-MAD score suggesting that the t-MAD score is a reliable measure of ctDNA levels. The addition of the t-MAD score to the *TP53* MAF increased the rate of ctDNA detection to 60% in the UKOPS HGSOC cohort and 56% in the CTCR-OV04 cohort. sWGS data was down-sampled to 3 million reads in the UKOPS cohort which may have effected the sensitivity of ctDNA detection.

Other studies looking at women with newly diagnosed OC have identified ctDNA in 71–100% of cases (Bettegowda et al. 2014; Phallen et al. 2017). The improved sensitivity in these studies is a result of parallel analysis of multiple genes and very deep sequencing. This results in complex library preparation methods, complex bioinformatics analysis and high cost rendering these methods incompatible with clinical use in their current iteration.

The lower rates of detection in the UKOPS and CTCR-OV04 cohorts are likely to be for a number a reasons. Firstly limitations of the cohorts themselves. The UKOPS cohort included in this thesis is a small number of samples with a range of histological subtypes. Only 22 of the cases had HGSOC. The CTCR-OV04 cohort although larger in number and restricted to HGSOC cases remains a small cohort in comparison to other published cohorts.

In particular the number of cases with early stage disease was low, however, this reflects the current real world clinical situation with very few women having early stage HGSOC at the time of diagnosis. Limitations also relate to the samples used. Plasma samples from the UKOPS cohort were <1ml in volume so sensitivity of ctDNA detection is limited by the number of available DNA molecules in the sample. Plasma samples from the CTCR-OV04 cohort ranges from 2–4ml in volume which is also lower than that used in recent publications. Limitations of the methods are also apparent. The panel of primers used for targeted sequencing was developed with a focus on HGSOC. It is therefore not optimised for detection of ctDNA is plasma samples collected from women with other histological subtypes of OC. Similarly the t-MAD score is only relevant for detecting ctDNA in cancers that exhibit significant copy number rearrangements. This is not the cases for non-HGSOCs.

In an attempt to overcome some of the limitations of ctDNA detection relating to the methods used I wanted to investigate strategies that could be applied to targeted sequencing and sWGS methods that might improve the sensitivity for detection of ctDNA in women with newly diagnosed disease that could be easily translated to a diagnostic setting.

Improving sensitivity and specificity for circulating tumour DNA detection

5.1 Introduction

Interest in the use of ctDNA for the detection of early stage cancers has recently increased (Wan et al. 2017). Approaches to increase the sensitivity of ctDNA detection in early stage disease include: the use of patient specific assays, the interrogation of multiple genes in one assay, deep sequencing and the combination of protein biomarkers with ctDNA assays (Bettegowda et al. 2014; Phallen et al. 2017; Abbosh et al. 2017; Cohen et al. 2018). These methods have allowed the detection of ctDNA in 40–50% of stage I cancers. These methods involve complex library preparation techniques and bioinformatics analysis. There is little data about how these assays perform in a real world situation and they are offered by commercial providers with no option to improve the technique to focus on a specific cancer type. They are also costly, so may not be suitable for deployment in the NHS.

One alternative strategy to increase sensitivity for ctDNA detection is to overcome background sequencing error rates. One way to do this is by limiting the number of input DNA molecules in a reaction to ensure that the minimum AF is sufficiently above the level of background noise which currently defines the limit of detection. Multiple reactions can be performed in parallel in order to screen sufficient molecules to observe low AF mutations (Figure 5.1a). This optimised targeting sequencing method had previously been conceptualised by the lab and the technology filed and granted as a patent by CRT (Rosenfeld et al. 2016). In this chapter I will discuss the optimisation of this method (referred to as TAm-Seq V2). Optimisation of this method was performed in conjunction with Wendy Cooper and the bioinformatics analysis performed by James Morris.

A second strategy for increasing the sensitivity of ctDNA detection is to enrich for cell free tumour DNA in a sample using selective sequencing. cfDNA fragments are commonly around 167 bp in length suggesting release from cells undergoing apoptosis (Jahr et al. 2001; Lo et al. 2010; Chandrananda et al. 2015; Jiang et al. 2016). However, circulating fetal DNA has been shown to be shorter than maternal plasma cfDNA and this difference has been utilised to improve the sensitivity of non-invasive prenatal testing (Lo et al. 2010; Yu et al. 2014; Lun et al. 2008; Minarik et al. 2015). In cancer it has also been shown that circulating tumour DNA fragments can be a different size to non-tumour cfDNA fragments but the results are conflicting (Mouliere et al. 2011; Umetani et al. 2006; Giacona et al. 1998; Underhill et al. 2016; Jiang et al. 2015; Mouliere et al. 2014).

In this chapter I will discuss the investigation of the features of ctDNA fragments compared to non-tumour cfDNA fragments in plasma (Figure 5.1b). I will discuss the use of selective sequencing based on these observations as a means of improving sensitivity of detection of ctDNA in women with newly diagnosed HGSOC. This work was part of a larger investigation of the fragmentation features of cfDNA seen in a range of cancer types led by Florent Mouliere, Dineika Chandrananda and Anna Piskorz. Dineika Chandrananda performed the in-silico size selection, calculation of t-MAD and determined the fragmentation features presented in this chapter. This manuscript has been accepted for publication in Science Translation Medicine (Appendix 12.13).

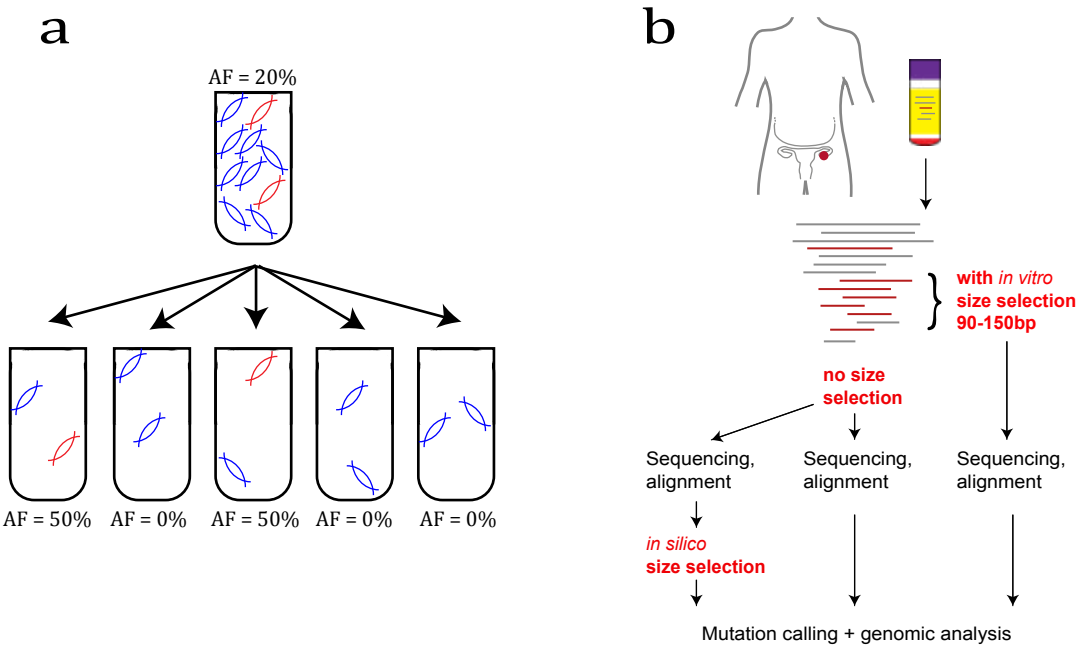


Figure 5.1: Approaches for increasing the sensitivity of ctDNA detection. **a)** TAm-Seq V2. If original sample with MAF of 20% contains 10 copies of DNA, two are mutant and 8 wildtype. If the sample is split into 5 replicates each containing two DNA molecules this would give two positive wells each with a MAF of 50% and three negative wells with an MAF of 0%. In practice this means that a sample with an MAF of 0.05% which is currently below the limit of detection could be partitioned into 250 wells each with 40 molecules (if sufficient molecules were available) and we would expect to detect 5 positive wells each with an MAF of 2.5% which is above the limit of detection. **b)** Selective sequencing of ctDNA fragments using size selection.

5.2 Results

5.2.1 TAm-Seq V2

Optimisation of method

Extensive optimisations were performed to determine the experimental conditions that provided the most even coverage across all amplicons for the TAm-Seq V2 method. The broad approach is summarised in Figure 5.2. Details of the experimental conditions are discussed in appendix 12.15. Optimisations were performed using healthy control serum

samples obtained commercially.

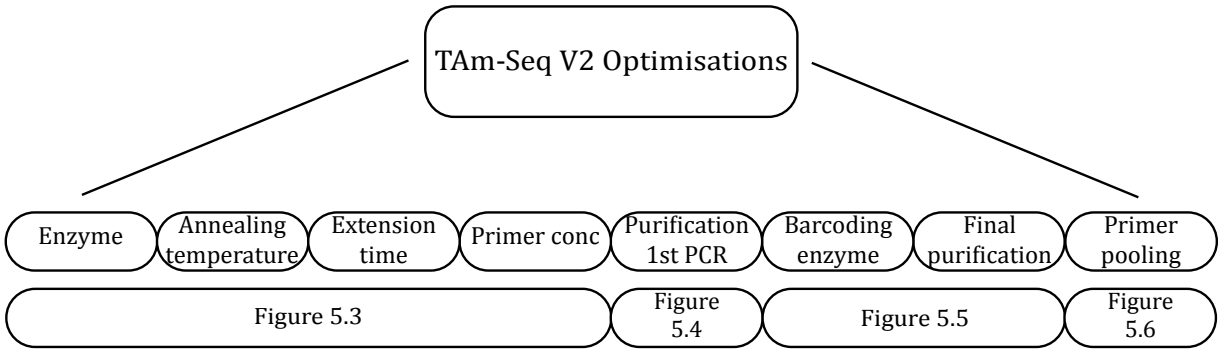


Figure 5.2: Experimental conditions investigated to optimise TAm-Seq V2 method. Further detail in appendix 12.15.

Investigation of different enzymes, annealing temperatures, extension times and primer concentrations found the Q5 HS enzyme with an annealing temperature of 67 °C and extension time of 10 seconds and the Q5 HS enzyme with an annealing temperature of 63 °C and extension time of 5 seconds to have the most even coverage across amplicons (Figure 5.3). Coverage decreased with increasing amplicon length for all experimental conditions.

Initial optimisations selected for a range of 140–190 bp using the PippinHT for purification of the first round PCR product as the PCR product was expected to range from 145–185 bp (based on amplicon size and barcode (6 bp each end) and CS1/CS2 tag (22 bp each end)). As coverage was poorer for the longer amplicons I hypothesised that these were being lost during during purification of the first round PCR product. I therefore investigated whether selecting for different size ranges would improve the coverage of the longer amplicons. Coverage for the longer amplicons was higher when selecting for fragments between 140–250 bp (Figure 5.4). There is no disadvantage in selecting for longer fragments as the main aim of this purification step is to remove the short primer dimer fragments. Interestingly two amplicons have significantly lower coverage compared to the other amplicons for all purification ranges.

Investigation of different barcoding enzymes and different methods for purification of the final PCR product revealed that the AMPure bead purification provided the most even coverage across amplicons (Figure 5.5). Although the Roche enzyme performed slightly better than the Q5 HS in terms of evenness of coverage the HS Q5 was selected as it is predicted to have a lower rate of PCR errors.

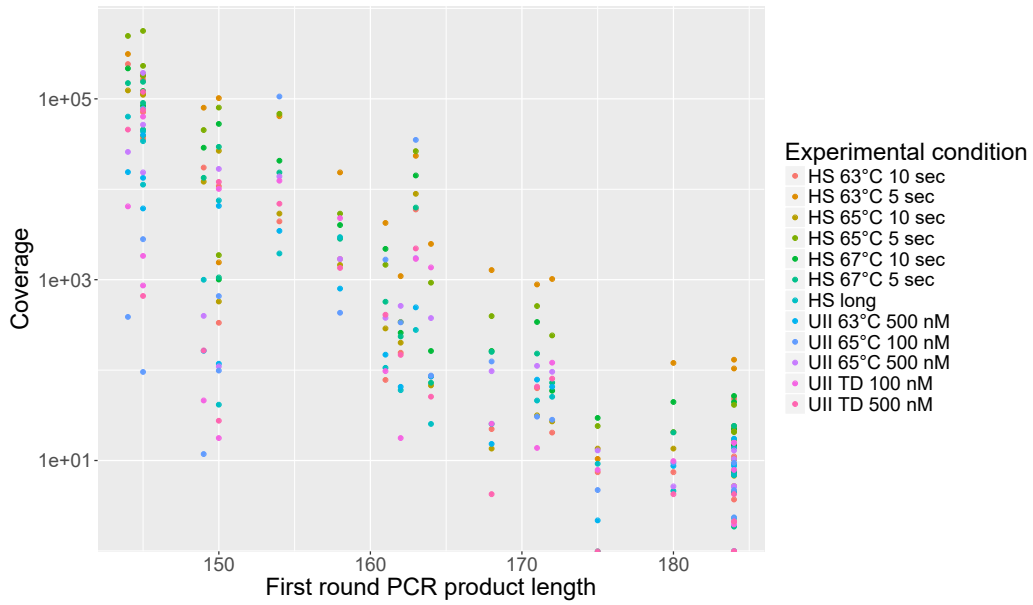


Figure 5.3: Sequencing coverage by first round PCR product length for different enzymes (HS = Q5 HS, UII = Ultra II), annealing temperatures, extension times and primer concentrations. See appendix 12.15, sections 12.15.1, 12.15.2, 12.15.3, page cviii for description of methodology.

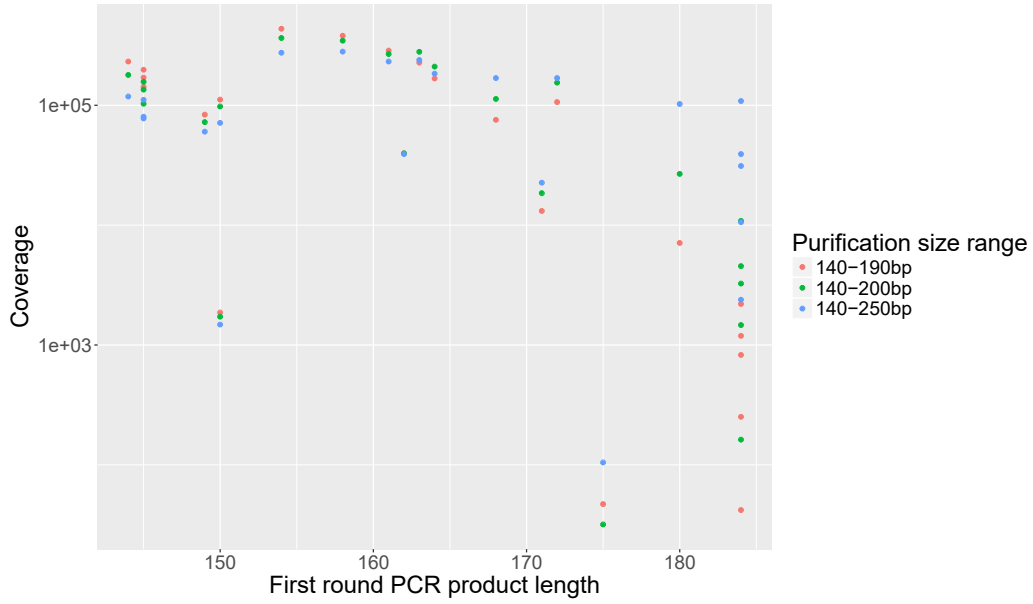


Figure 5.4: Sequencing coverage by first round PCR product length for different size ranges used for purification of first round PCR product. See appendix 12.15, section 12.15.4, page cxiii for description of methodology.

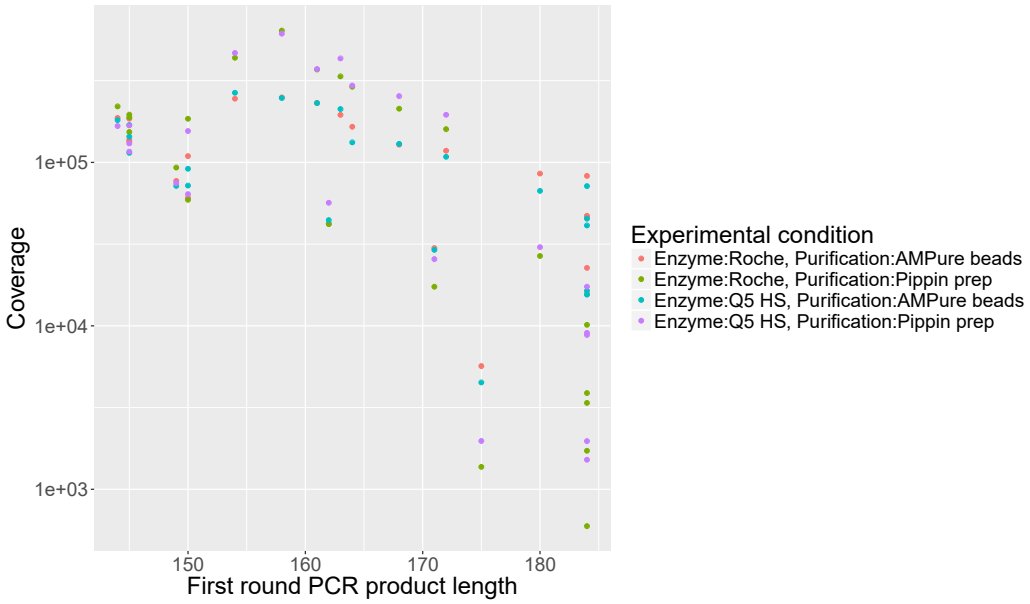


Figure 5.5: Sequencing coverage by first round PCR product length for different barcoding enzymes and different final purification methods. See appendix 12.15, sections 12.15.5 and 12.15.6, page cxiii for description of methodology.

Investigation of different primer pooling strategies found strategy 4 (Appendix 12.15) to provide the most even coverage across all amplicons (Figure 5.6a). An annealing temperature of 67 °C and an extension time of 5 seconds provided the most even coverage within this pooling strategy (Figure 5.6b).

Table 5.1 shows the final parameters chosen to take forward for the TAM-Seq V2 method following all optimisations.

Enzyme	Q5 HS
Annealing temperature	67 °C
Extension time	5 seconds
Primer concentration	100nM
Purification 1st round PCR product	PippinHT 140–250 bp
Barcoding enzyme	Q5 HS
Final purification	AMPure bead
Primer pooling	Strategy 4

Table 5.1: Final parameters for TAm-Seq V2 method.

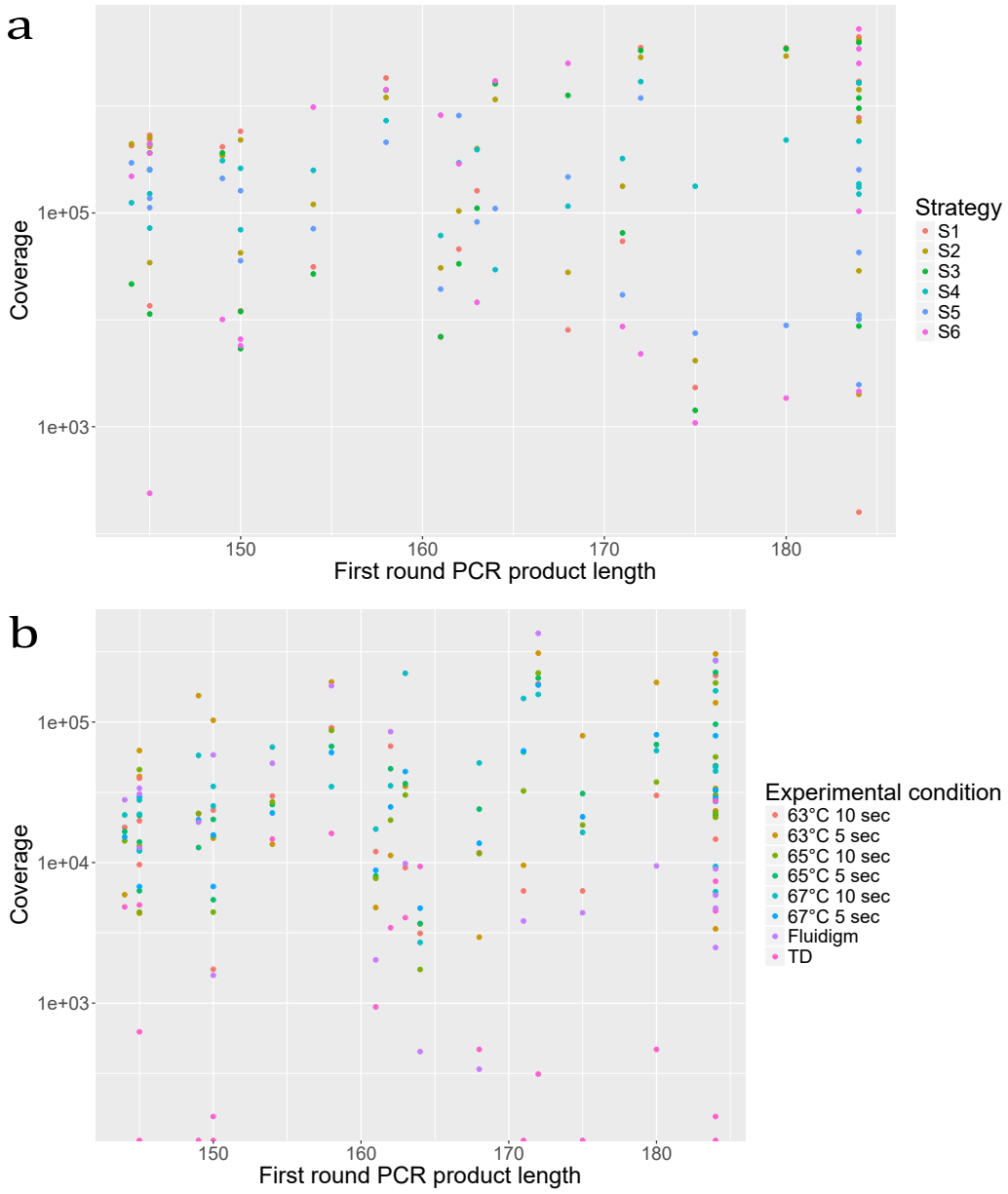


Figure 5.6: **a)** Sequencing coverage per product length for different primer pooling strategies. See appendix 12.15, section 12.15.7, page cxv for description of methodology. **b)** Sequencing coverage per product length for different experimental conditions for primer pooling strategy 4. TD=touchdown PCR protocol. See appendix 12.15, section 12.15.2, page cix for description of methodology.

Amplicon	Forward Primer	Reverse Primer	Probe
Amplicon001	GTGGGGAACAAGAAGTGGAGA	CTCCCTGCTTCTGTCTCCTAC	ATGTCAGTCTGAGTCAGGCC
Amplicon001.2	TCAGTCTGAGTCAGGCCCTTCT	AAGTCCAAAAAGGGTCAGTCTACCT	TTGAACATGAGTTTTTTATGGC
CXP039A	CAGCTCTCGGAACATCTCGAA	CAACAGAGGAGGGGGAGAAGT	CTCACGCCCACGGAT
Amplicon023	CTAGGATCTGACTGCGGCTC	TGGAAGTGTCTCATGCTGGATC	GCTGCTGCAAGAGGAAAAGT

Table 5.2: Digital PCR primers and probes for quantification of both ends of *TP53*.

Quantification of DNA for input into PCR

Due to the principle of TAm-Seq V2 accurate quantification of the number of copies of DNA input into the reaction is required. *RPP30* is a stable region in most genomes and primers and probes against *RPP30* are routinely used in the laboratory for quantification of total copies of DNA. In HGSOC *RPP30* is not a stable region. I therefore designed digital PCR primers and probes for both ends of *TP53* in order to quantify the number of copies of *TP53* input into the reaction (Table 5.2). Primers and probes were tested in singleplex and multiplex using cell line DNA and DNA from healthy control serum samples obtained commercially (Table 5.3). Amplicon001 and Amplicon023 performed well in singleplex but were not as efficient in multiplex. The same pattern was seen with Amplicon001.2 and Amplicon023. CXP039A and Amplicon023 however, multiplexed well and was therefore the assay used for quantification of DNA for subsequent experiments.

Investigation of sensitivity

To investigate the sensitivity of ctDNA detection using TAm-Seq V2 DNA from three patient samples was pooled (Table 5.4) and diluted using healthy control DNA. Targeted sequencing of the pooled dilutions was performed using the standard TAm-Seq method to provide a comparison for MAF and sensitivity (Table 5.5). The lowest MAF detected was 0.87%. Mutations were not detected in the 1:25 dilution where the MAF was expected to be 0.4% or lower. This is consistent with previous limits of detection observed using this method (Section 4.2).

The expected MAF for the mutations at different dilutions was extrapolated from the TAm-Seq results for the neat sample (Table 5.6). From this the number of expected positive wells was determined for each of the three mutations (Table 5.7).

Sample	RPP30 AC/µl	Amplicon001 singleplex AC/ul	Amplicon001 multiplex AC/µl	Amplicon023 multiplex AC/µl	Amplicon023 singleplex AC/µl
HeLa(10 ng/µl)	9044	8239	5023	8818	8874
SKOV3(10 ng/µl)	4100	5918	3423	6167	6144
Female control	200	221	180	195	200
Male control	64	69	74	92	64
Sample	RPP30 AC/µl	Amplicon001.2 singleplex AC/µl	Amplicon001.2 multiplex AC/µl	Amplicon023 multiplex AC/µl	Amplicon023 singleplex AC/µl
CIOV1 (10 ng/µl)		4195	1223	3759	4244
CIOV1 (1 ng/µl)		221	154	167	218
CIOV1 (0.1 ng/µl)		31	18	23	26
COV362 (10 ng/µl)		10	1210	3541	3
COV362 (1 ng/µl)		392	259	269	44
COV362 (0.1 ng/µl)		26	15	21	18
H_2O		0	0	0	0
Sample	RPP30 AC/µl	CXP039A singleplex AC/µl	CXP039A multiplex AC/µl	Amplicon023 multiplex AC/µl	Amplicon023 singleplex AC/µl
CIOV1 (10 ng/µl)		976	1024	788	962
CIOV1 (10 ng/µl)		1028	1024	755	1036
CIOV1 (1 ng/µl)		72	100	46	63
CIOV1 (1 ng/µl)		68	76	52	50
CIOV1 (0.1 ng/µl)		6	5	10	7
CIOV1 (0.1 ng/µl)		5	1	3	8
H_2O		NA	0	0	NA
H_2O		NA	0	0	NA

Table 5.3: Digital PCR primer and probe optimisations. AC = amplifiable copies.

Sample	Mutation	RPP30 AC/µl	Amplicon001 AC/µl	Amplicon023 AC/µl	Expected MAF in pool (%)
OV04 68	7578534C>G, c.G396C, p.K132N	1182	513	831	5.00
OV04 114	7578406C>T, c.G524A, p.R175H	1126	218	564	7.76
OV04 70	7577538C>T, c.G743A, p.R248Q	2846	1682	2149	11.58

Table 5.4: Three patient serum samples quantified by digital PCR and pooled to use in dilution experiment to asses sensitivity of TAm-Seq V2.

Mutation	Neat	1:5	1:25	1:100	1:400
7578534C>G	5.32	0.87	ND	ND	NA
7578406C>T	11.07	2.55	ND	ND	NA
7577538C>T	10.31	2.41	ND	ND	NA

Table 5.5: MAF (%) detected in neat pool and diluted pool samples using standard TAm-Seq. ND=not detected. NA=not run as this was expected to be well below the limit of detection using standard TAm-Seq.

Mutation	Neat	1:5	1:25	1:100	1:400
7578534C>G	5.32	1.06	0.21	0.05	0.01
7578406C>T	11.07	2.21	0.44	0.11	0.03
7577538C>T	10.31	2.06	0.41	0.10	0.03

Table 5.6: Expected MAF (%) for the three mutations in the pool at different dilutions based on the standard TAm-Seq neat sample MAF and the dilution factor.

Mutation	Neat	1:5	1:25	1:100	1:400
7578534C>G	42	17	4	4	2
7578406C>T	47	28	8	8	4
7577538C>T	47	27	7	8	4
Total wells analysed	48	48	48	192	384

Table 5.7: Number of positive wells expected for each mutation at different dilutions based on the total number of wells and the expected MAF.

Mutation	Neat	1:5	1:25	1:100	1:400
7578534C>G	34	6	0	0	0
7578406C>T	44	23	3	7	2
7577538C>T	48	20	2	5	2
Total wells analysed	48	48	48	192	384

Table 5.8: Number of positive wells for each mutations at different dilutions.

Mutation calling was performed using the TAm-Seq V2 method for the pooled dilutions. All mutations called in at least two wells were reviewed. The MAF was calculated by dividing the total number of non-reference alleles across all wells by the total number of wells multiplied by the number of copies of DNA in each well (Table 5.9). The observed AF correlated well with the expected AF (Figure 5.7). With approximately 30,000 input molecules (384 replicates of 2 pools of 40 molecules) the limit of detection was a MAF of approximately 0.01–0.02%. No unexpected mutations were observed in the 1:5, 1:25, 1:100 and 1:400 dilutions in >2 wells. However, in the undiluted (neat) sample 3 unexpected mutations were observed each in 2 wells out of the 48 analysed (7578190T>C (c.A659G, p.Y220C), 7577114C>T (c.G824A, p.C275Y) and 7577548C>T (c.G733A, p.G245S)). These could represent false positive calls or could be true positive mutations that were not identified in the FFPE sample due to tumour heterogeneity. If a threshold of >2 wells is used to call a true positive mutation these potential false positive calls are eliminated along with three true positive calls (Table 5.8). The coverage at position 7578534 was low in the control samples which may explain why this mutation was not observed in the patient pool at the lower dilutions.

Mutation	Neat	1:5	1:25	1:100	1:400
7578534C>G	5.26	0.69	ND	ND	ND
7578406C>T	8.84	1.84	0.35	0.16	0.02
7577538C>T	7.32	2.06	0.26	0.14	0.01

Table 5.9: Calculated AF (%) for each mutation at different dilutions based on the number of positive wells, total number of wells screened, number of mutant reads and total number of reads.

To investigate if these unexpected mutations were true or false positive calls a further 48 replicates of the undiluted pool were performed. (This experiment was done with the optimised primer pooling strategy that had not been used for the first experiments of the pooled samples). The following mutations were detected:

Position	Mutation	Number of positive wells
7578534	C>G (c.G396C, p.K132N)	43
7578406	C>T (c.G524A, p.R175H)	48
7577538	C>T(c.G743A, p.R248Q)	48
7577548	C>T (c.G733A, p.G245S)	4
7579325	G>T (c.C362A, p.S121Y)	2

7578406C>T and 7577538C>T were expected mutations detected in approximately the same numbers of wells as with the initial pooling strategy and experimental conditions. 7578534C>G was previously detected in 34 wells in the undiluted sample and not at all in the 1:25, 1:100 and 1:400 dilutions due to low coverage at this locus in the controls. Using the new pooling strategy 7578534C>G was detected in 43 wells which suggests that mutation calling across *TP53* may now be more even. 7577548C>T was previously detected in 2 wells in the undiluted sample which suggests this may be a true positive mutation. 7579325G>T was not previously called which suggests that this is a false positive call and suggests that the threshold for defining a positive test should be 3 or more positive wells.

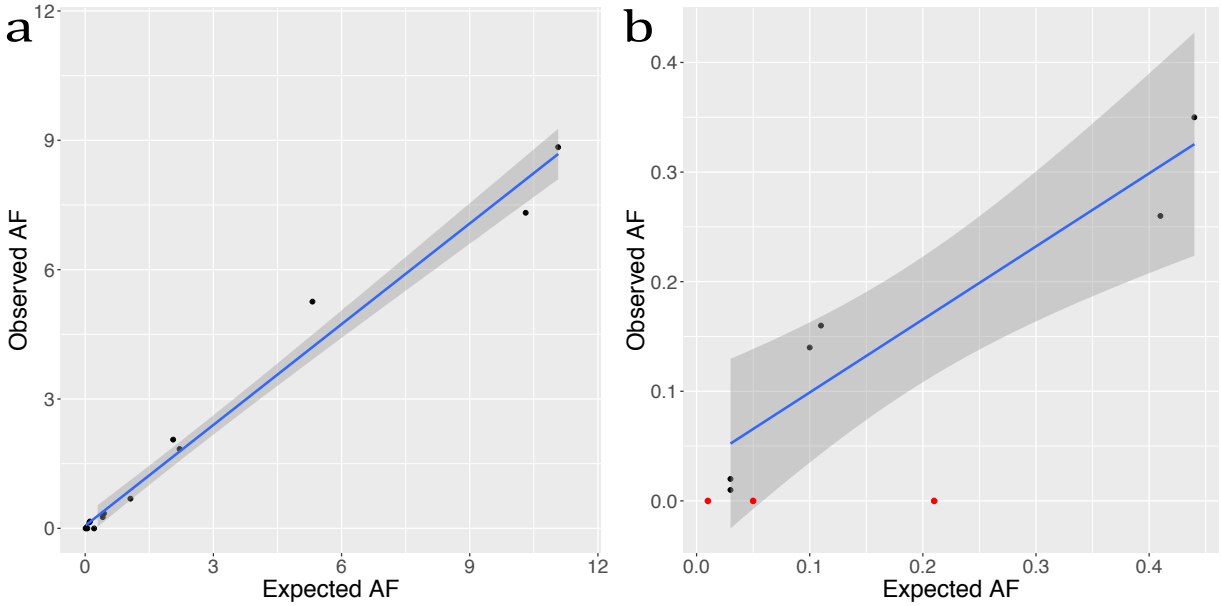


Figure 5.7: **a)** Expected AF extrapolated from standard TAm-Seq mutation calling and observed AF for all calls. **b)** Expected AF extrapolated from standard TAm-Seq mutation calling and observed AF for calls with an AF of <1%. Those marked in red were not detected by TAm-Seq V2.

Application to clinical samples

The TAm-Seq V2 method was then applied to plasma samples collected from three women with HGSOC where a *TP53* mutation had been detected at close to the limit of detection using standard TAm-Seq. A mutation was identified in >3 of the replicate wells for each sample and in each case an identical mutation was identified at a similar AF using standard TAm-Seq (Table 5.10). When using a threshold of 2 wells to define a true positive mutation three other alterations were observed in JBLAB-16679 (7573927C>A, 7578517G>A and 7578393A>C). It was encouraging that the same *TP53* mutations were called at a similar MAF using both the standard TAm-Seq and TAm-Seq V2 methods. However, the *TP53* MAF in these samples was within the limit of detection of standard TAm-Seq so not testing the increased sensitivity of TAm-Seq V2.

I therefore wanted to apply the TAm-Seq V2 method to plasma samples in which a *TP53* mutation had not been detected using TAm-Seq V1.

Sample	Mutation	Positive wells/ wells analysed	Observed AF (%)	AF (%) by standard Tam-Seq
JBLAB-16676	7577539G>A	10/48	0.83	0.78
JBLAB-16679	7577124C>T	7/95	0.31	0.54
JBLAB-16710	7578413C>A	13/43	0.73	0.64

Table 5.10: TAm-Seq V2 mutation calling results for three patient plasma samples.

In order to detect mutations at an AF of 0.01–0.02% 384 replicate wells needed to be analysed. Each well contains 40 copies of DNA and is run in duplicate for the two different primer pools. 30,720 DNA molecules therefore need to be screened to reach this level of sensitivity. Digital PCR for both end of *TP53* was performed to quantify DNA from plasma samples from 23 women with HGSOC (Table 5.11). Samples had been extracted from 0.7–3 ml of plasma and contained between 1154–41385 total DNA molecules. Only one sample contained sufficient molecules to perform 384 replicate TAm-Seq V2 reactions.

The median number of DNA copies per ml of plasma was 3590 (range 1648–13795) (Table 5.11). A median of 9 ml of plasma is therefore required for extraction in order to obtain sufficient DNA molecules to screen for the detection of low AF mutations.

Sample	Volume of plasma (ml)	Total TP53 copies	Copies/ml plasma
JBLAB-16679	3.0	41385	13795
JBLAB-14752	3.0	29538	9846
JBLAB-16676	3.0	27308	9103
JBLAB-16710	3.0	22846	7615
JBLAB-15562	3.0	17769	5923
JBLAB-15108	3.0	15000	5000
JBLAB-16616	1.5	14769	9846
JBLAB-15067	3.0	10462	3487
JBLAB-15441	3.0	10077	3359
JBLAB-16628	0.9	9385	10427
JBLAB-15711	3.0	9000	3000
JBLAB-16687	3.1	9000	2903
JBLAB-16632	2.1	6154	2930
JBLAB-16642	1.7	6000	3529
JBLAB-16627	1.5	5385	3590
JBLAB-16614	1.5	5231	3487
JBLAB-16635	0.7	4308	6154
JBLAB-15179	3.0	3769	1256
JBLAB-16648	0.8	3538	4423
JBLAB-16640	1.2	3231	2809
JBLAB-16622	0.9	2923	3248
JBLAB-16629	0.7	2769	3956
JBLAB-16625	0.7	1154	1648

Table 5.11: Plasma samples from 23 women with HGSOC. Total number of *TP53* mutations based on digital PCR for both ends of *TP53*. Copies of *TP53* per ml plasma.

5.2.2 Selective sequencing using DNA fragment sizes

I used xenograft models to test the hypothesis that selective sequencing of ctDNA could be performed by leveraging differences in DNA fragmentation patterns between tumour derived and non-tumour derived cfDNA fragments (Figure 5.8). In this model tumour derived cfDNA fragments align to the human genome and non-tumour derived cfDNA fragments align to the host genome (in this case mouse). Whole genome libraries were prepared from plasma DNA using the Rubicon ThruPlex DNA-kit (Table 5.12) and submitted for paired-end 150 sequencing on the Illumina MiSeq.

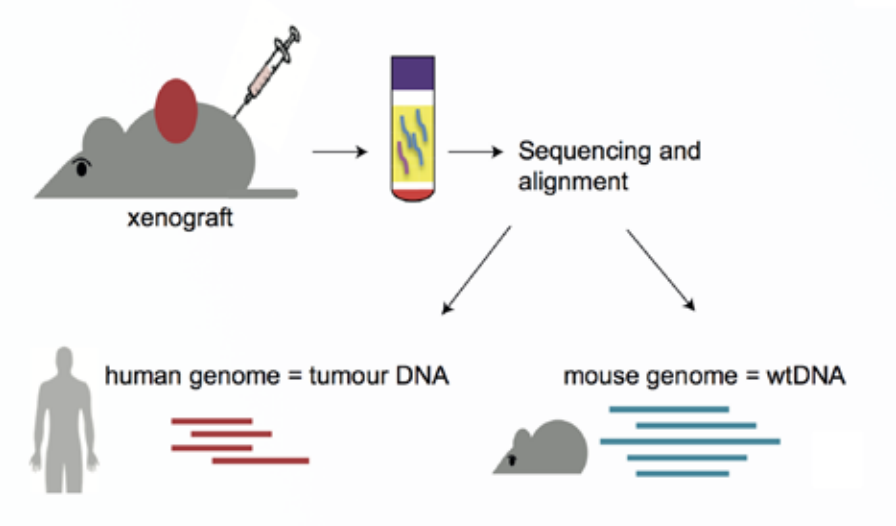
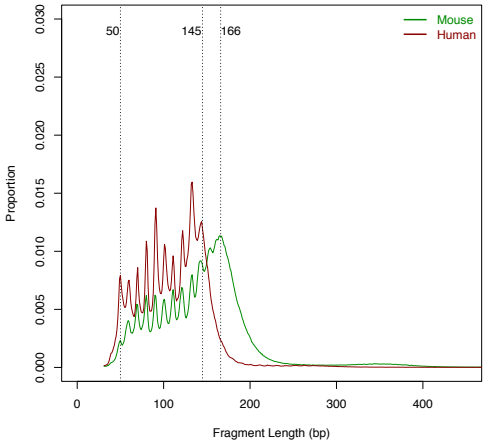


Figure 5.8: Mouse model with xenografted tumour cells enabled the discrimination of tumour derived cfDNA fragments from non-tumour cfDNA fragments.

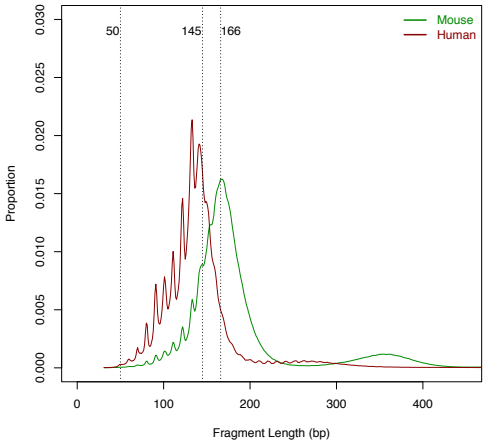
Figure 5.9 shows the proportion of reads aligning to the human genome (tumour) and the proportion of reads aligning to the mouse genome (non-tumour) against fragment length. The median difference between the proportion of tumour derived and non-tumour derived cfDNA fragments between 90–150 bp was 30.6%. In this model system, tumour derived DNA fragments were shorter than non-tumour cfDNA fragments.

a



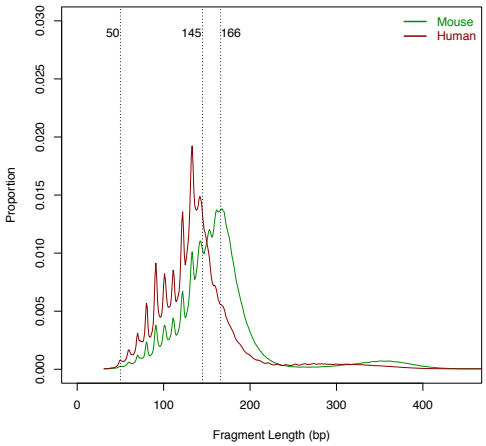
Proportion of fragments between 90-150bp:
Tumour: 57.3%
Non-Tumour: 34.9%

b



Proportion of fragments between 90-150bp:
Tumour: 65.1%
Non-Tumour: 20.7%

c



Proportion of fragments between 90-150bp:
Tumour: 58.2%
Non-Tumour: 33.3%

Figure 5.9: Xenograft fragmentation profiles **a)** Plasma DNA from mouse implanted with OVCAR3 cultured cells **b)** Plasma DNA from mouse implanted with ascites collected from a women before treatment for HGSOC **c)** Plasma DNA from mouse implanted with tissue collected at the time of primary surgery for HGSOC.

Mouse	Xenograft model	Sample	Sample volume (µl)	DNA concentration (ng/µl)	DNA input whole genome library prep (ng)	Number of PCR cycles
a	OVCAR3	plasma	55	0.015	0.3	16
a	OVCAR3	serum	62	<0.005 [a]	unknown	16
b	ascites	plasma	270	0.338	1.0	11
b	ascites	serum	160	0.618	1.0	11
c	solid tumour	plasma	180	0.490	1.0	11
c	solid tumour	serum	90	0.678	1.0	11

Table 5.12: Xenograft samples and whole genome library preparation. OVCAR3 is a HGSOC cell line implanted into a mouse. Ascites was collected from a women before treatment for HGSOC and implanted into a mouse. Solid tumour was obtained at the time of primary surgery for HGSOC and implanted into a mouse.

[a] too low to quantify

The next step was to investigate if there is difference in fragment length between tumour-derived and non-tumour derived cfDNA in plasma samples from patients with cancer. It is not easy to distinguish tumour derived cfDNA fragments from non-tumour derived cfDNA fragments within a single sample without the identification of multiple mutations and ultra-deep sequencing. I therefore performed sWGS to compare the fragment sizes of DNA identified in plasma samples from 45 women with relapsed HGSOC to 46 healthy controls (Figure 5.10a). The proportion of DNA fragments <150 bp was higher for the HGSOC cases compared to the healthy controls (Figure 5.10b). Although this approach cannot confirm that these shorter fragments are tumour derived cfDNA fragments the fact that the same pattern is not seen in cfDNA from healthy controls supports the hypothesis that ctDNA fragments may be shorter than non-tumour derived cfDNA fragments.

Parallel experiments performed by other members of the lab in other cancer types support the generalisability of these findings and led to the hypothesis that selecting for DNA fragments between 90–150 bp could enrich for tumour derived cfDNA.

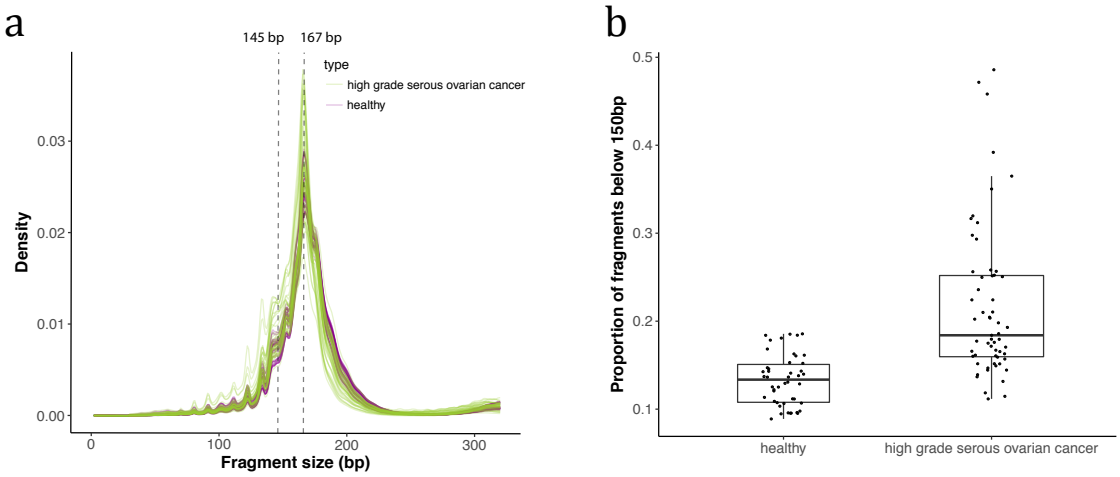


Figure 5.10: **a)** Fragmentation profiles for 45 HGSOC cases and 46 healthy controls. **b)** Proportion of cfDNA fragments below 150 bp for healthy controls and HGSOC cases.

To investigate this hypothesis I performed targeted sequencing for *TP53* using TAm-Seq primer panel 10 (Appendix 12.4) for plasma samples before in-vitro size selection and after in-vitro size selection (PippinHT 3% agarose gel selecting for bases 90–150 bp) collected from 13 women with relapsed HGSOC. An increase in MAF was observed for the majority of samples (at baseline and after one cycle of chemotherapy) following in-vitro size selection (Figure 5.11). Samples with a MAF <5% can generally not be used for more in depth analysis for example whole exome or whole genome sequencing. The dotted area in figure 5.11a illustrates samples with a MAF <5% before size selection which increased to >5% following in vitro size selection making them accessible for further wide-scale analysis.

sWGS was also performed for the samples before and after in-vitro size selection. Importantly we see that following in-vitro size selection the majority of cfDNA fragments lie within the range selected for (Figure 5.12).

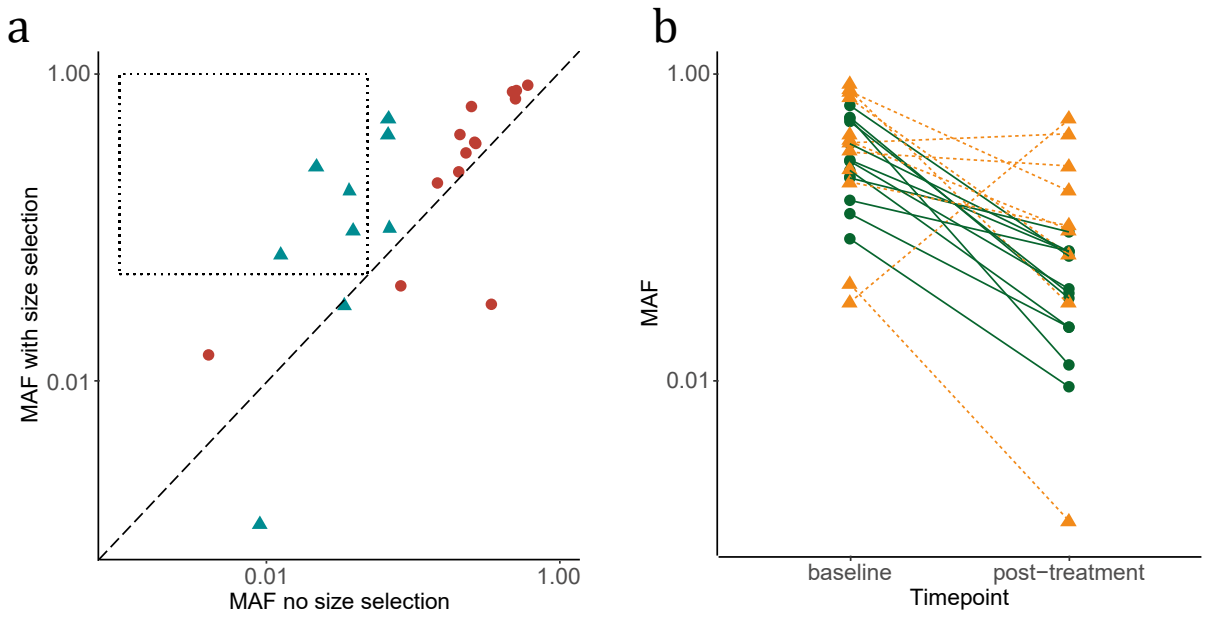


Figure 5.11: **a)** MAF before and after in-vitro size selection for plasma samples collected before (red circles) and after (blue triangles) starting chemotherapy for relapsed disease in 13 women with HGSOC. The dotted area indicates samples with a MAF <5% pre size selection and >5% post size selection. **b)** Comparison of MAF before and after treatment initiation with (yellow triangles) and without (green circles) in-vitro size selection.

Figure 5.13 shows an example of the copy number profiles generated from sWGS data for one patient. Figure 5.13a shows the baseline plasma sample with no size selection where the ctDNA levels are high with blue illustrating amplifications, orange deletions and black copy number neutral regions. In the post treatment plasma sample (Figure 5.13b) the level of ctDNA is reduced with fewer amplifications and deletions observed. Following in-vitro size selection of the post-treatment sample (Figure 5.13c) amplifications and deletions that match those observed in the pre-treatment plasma sample become visible suggesting an enrichment in ctDNA post size selection. Copy number plots for the remaining samples are shown in appendix 12.13.

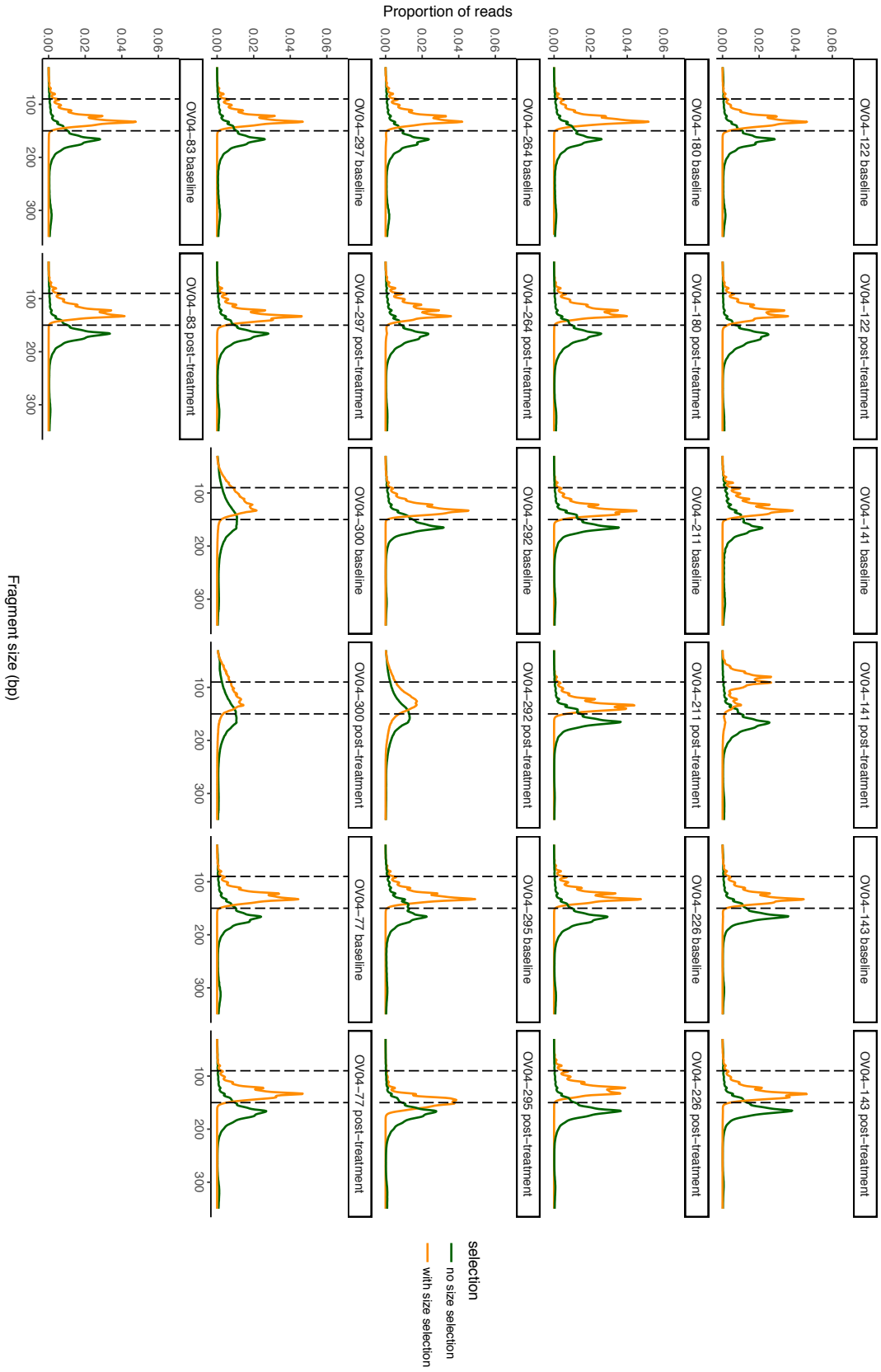


Figure 5.12: Distribution of DNA fragments determined by sWGS for plasma samples collected before and after treatment from 13 women with relapsed HGSOC. The distribution of cfDNA without size selection is shown in green and the distribution of the same sample after in-vitro size selection is shown in orange. The vertical lines represent the range of fragments selected with the PippinHT cassettes, between 90 and 150 bp.

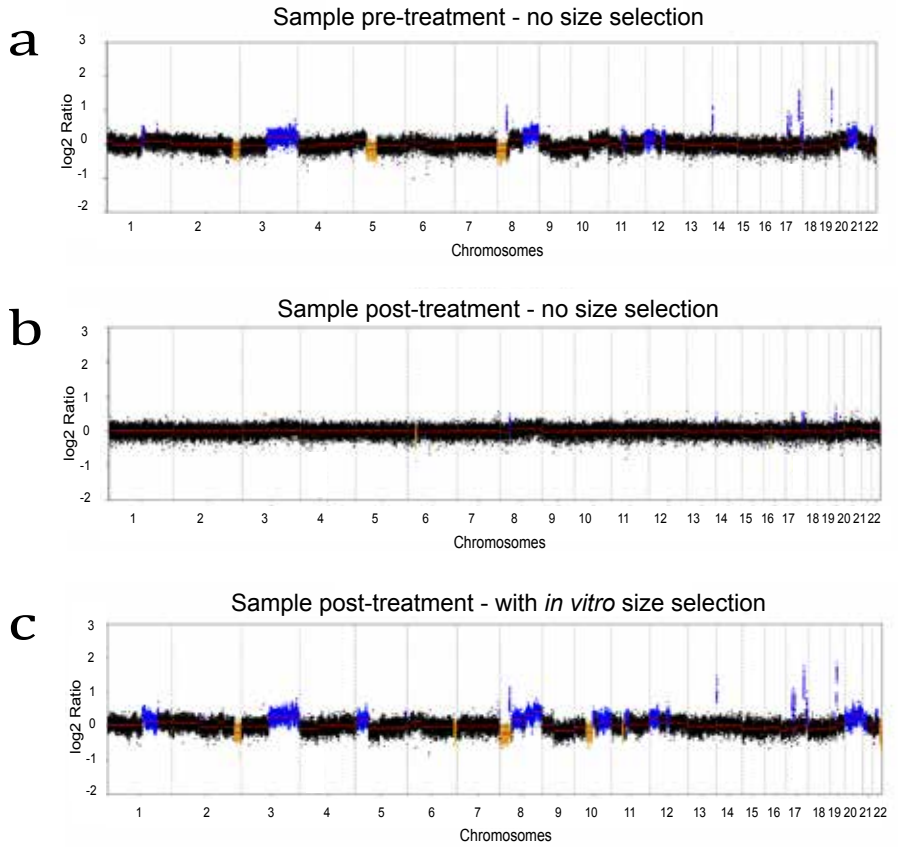


Figure 5.13: **a)** SCNA analysis of sWGS data from a plasma sample collected from a women with relapsed HGSOC before initiation of chemotherapy. Inferred amplifications are shown in blue and deletions in orange. **b)** SCNA analysis of a plasma sample from the same women after initiation of chemotherapy. **c)** SCNA analysis of the sample plasma sample as shown in b, after in-vitro size selection for DNA fragments between 90–150 bp.

sWGS was performed on a further 22 plasma samples collected from women with relapsed HGSOC before and after in-vitro size selection for fragments between 90–150 bp. The t-MAD score was calculated for all plasma samples to quantitatively asses the enrichment in ctDNA after size selection on a genome-wide scale. Following in-vitro size selection a significant increase in the t-MAD score was observed for the cohort (Figure 5.14b). The t-MAD score increased in 47/48 (98%) of the plasma samples (Figure 5.14a) with a median increase of 2.1 fold (Figure 5.14c). The degree of enrichment varied per sample but was higher for samples with a lower initial t-MAD score and therefore a lower level of original ctDNA (Figure 5.14c). This provides further evidence that selective sequencing of shorter fragments enriches for ctDNA in a sample.

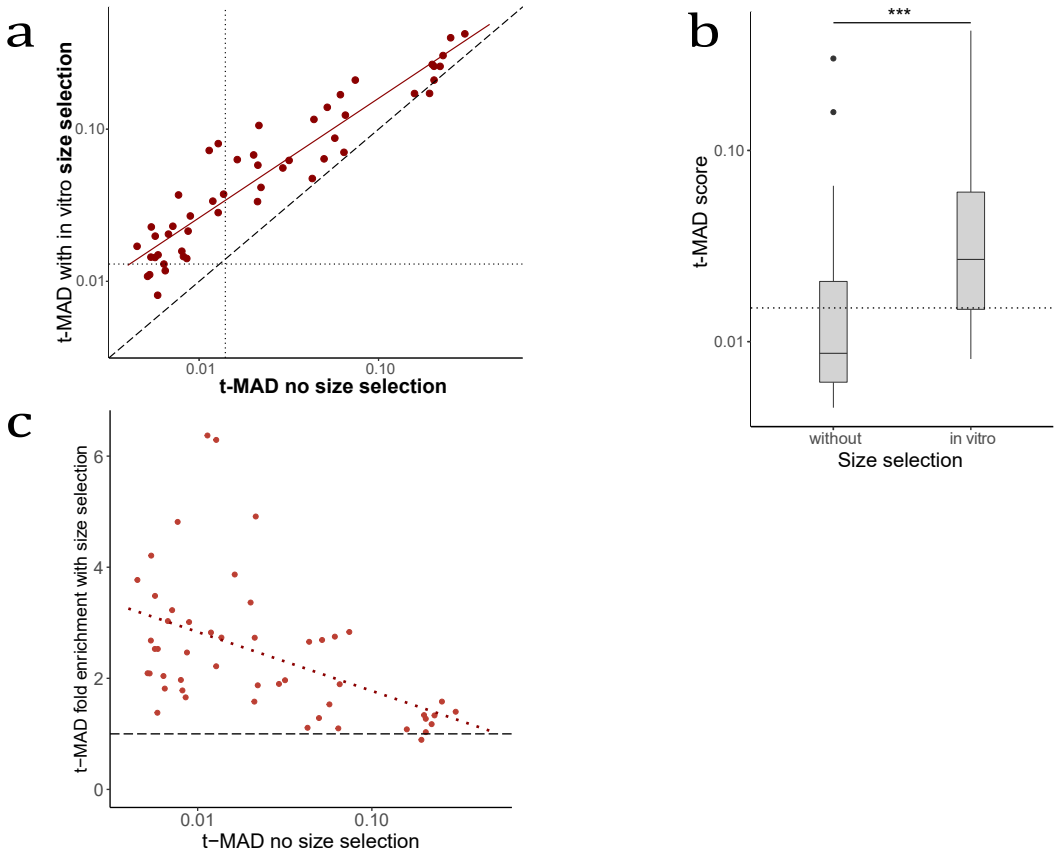


Figure 5.14: **a)** t-MAD score pre and post in-vitro size selection for 48 plasma samples collected from women with relapsed HGSOC. **b)** Comparison of t-MAD score with and without in-vitro size selection for the whole cohort. **c)** t-MAD score without size selection and t-MAD fold enrichment with in-vitro size selection.

To investigate if sWGS and size selection to enrich for ctDNA in a sample allowed for increased detection of cancer cases receiver operating characteristic (ROC) analysis comparing 48 women with relapsed HGSOC to 46 healthy controls was performed (Figure 5.15). A significant increase in area under the curve (AUC) was seen following in-silico size selection and in-vitro size selection compared to the unselected t-MAD score. A maximum sensitivity and specificity of 90% and 98% respectively was seen using the t-MAD score calculated for all samples following in vitro size selection. Again this illustrates the enrichment of tumour derived cfDNA in a sample following size selection (both in-vitro and in-silico) and suggests that size selection may improve ctDNA detection rates in women with newly diagnosed HGSOC.

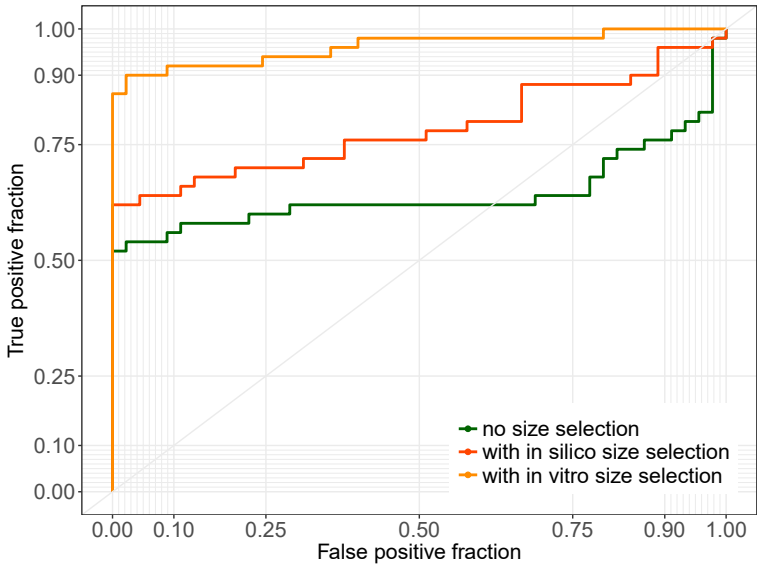


Figure 5.15: ROC analysis comparing 48 HGSOC cases and 46 healthy controls using t-MAD (AUC 0.64), t-MAD following in-silico size selection (AUC 0.78) and t-MAD following in-vitro size selection (AUC 0.97).

To assess if SCNAs with potential clinical value could be detected by sWGS following in vitro size selection the relative copy number values of 15 genes frequently altered in HGSOC were compared before and after size selection. A large number of SCNAs were revealed following size selection that were not detected in the same sample without size selection. This included amplifications in key genes including *NF1*, *TERT* and *MYC*. This could be useful clinically for patient stratification to clinical trials and treatment with targeted therapy.

In addition to the size of cfDNA fragments different fragmentation features (Figure 5.17a) were defined in the 48 plasma samples collected from women with relapsed HGSOC and the 46 healthy controls. The proportion (P) of DNA fragments observed by sWGS in multiple size ranges, the ratios of proportions of fragments in these size ranges and the amplitude of oscillations in fragment size density with 10 bp periodicity observed below 150 bp were calculated. The proportion (P) of fragments between 20–150 bp and the ratio of the proportion (P) of fragments between 20–150 bp and 160–180 bp (P(20–150)/P(160–180)) were best able to discriminate the HGSOC cases from the healthy controls (Figure 5.17b).

Selective sequencing of DNA fragments based on size and fragmentation features is therefore an option that may be useful in increasing the rate of detection of ctDNA in newly diagnosed HGSOC.

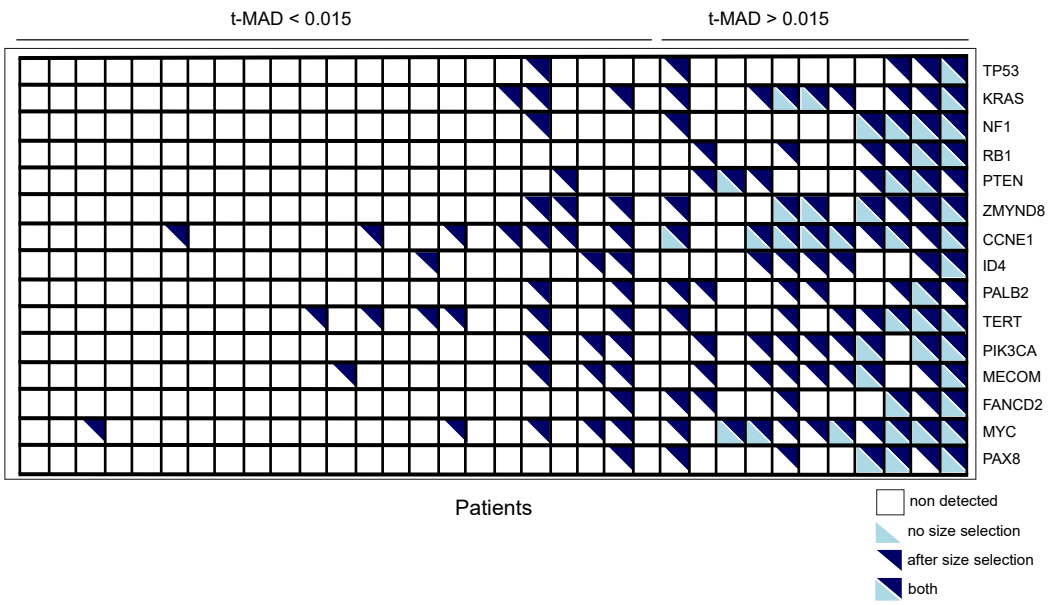


Figure 5.16: Detection of SCNA in 35 women with relapsed HGSOC across 15 genes frequently mutated in recurrent HGSOC. Patients ranked from left to right by increasing tumour fraction quantified by t-MAD. SCNA are labelled as detected for a gene if the relative copy number in that region was greater than 0.05. Empty squares represent copy number neutral regions, bottom left triangles in light blue indicate SCNAs that were detected without size selection and top right triangles in dark blue SCNAs that were detected after in-vitro size selection.

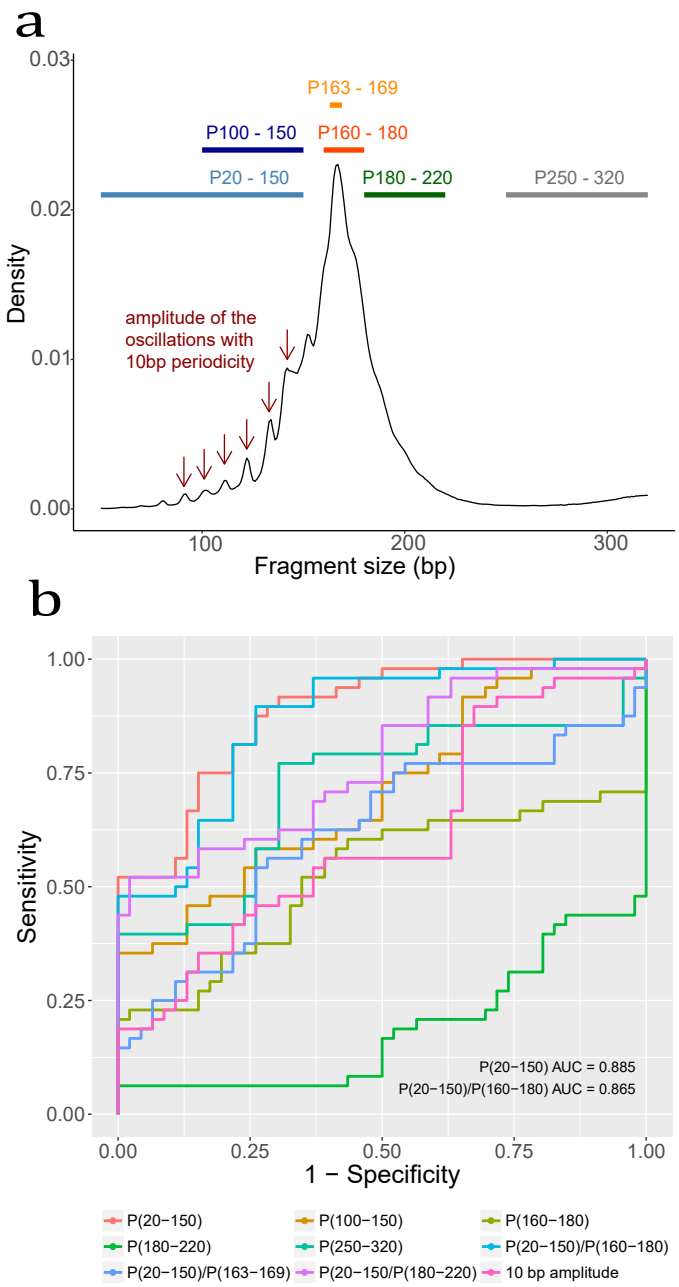


Figure 5.17: **a)** Schematic illustrating the selection of different size ranges and features in the distribution of fragment sizes. For each sample features include the proportion (P) of DNA fragments observed by sWGS in multiple size ranges (P(20–150), P(100–150), P(160–180), P(180–220), P(250–320)), the ratios of proportions of fragments in these size ranges (P(20- 150)/P(160–180), P(20–150)/P(163–169), P(20–150)/P(180–220) and the amplitude of oscillations in fragment size density with 10 bp periodicity observed below 150 bp. **b)** ROC analysis of fragmentation features comparing HGSOC cases to healthy controls.

5.3 Discussion

Due to the low rates of detection of ctDNA in women with newly diagnosed HGSOC using targeted sequencing and sWGS I wanted to investigate methods to improve the sensitivity of detection that could be easily translated for routine clinical use.

In this chapter I have discussed the optimisation of the TAm-Seq V2 method and shown that this method has the potential to detect mutations with a minimum MAF of 0.01–0.02%. This method is potentially more sensitive than other recently published methods that have been applied to newly diagnosed cancers using patient specific assays and multi-region deep sequencing enabling detection of a minimum MAF of 0.05–0.1% (Phallen et al. 2017; Abbosh et al. 2017). However, in this chapter I have only provided proof of principle of the method I have not been able to demonstrate clinical utility due to the limitations outlined below.

The major limitation of the TAm-Seq V2 method is that a large number of DNA molecules need to be input into multiple replicate reactions in order to screen sufficient DNA molecules for the detection of low AF mutations. Using TAm-Seq V2 over 30,000 DNA molecules need to be screened in order to detect a minimum MAF of 0.01%. I have found that plasma samples contain a median of 3590 DNA copies per ml (range 1648–13795). A median of 9 ml of plasma is therefore required for extraction in order to obtain sufficient DNA molecules to screen for the detection of low AF mutations using this method. Currently our standard practice locally is to extract DNA from 4 ml of plasma. In the future it may be important to increase the volume of blood collected from patients and optimise extractions from larger volumes of plasma. Other recently published methods have used 5 ml of plasma and detected a minimum MAF of 0.1% (Bettegowda et al. 2014; Abbosh et al. 2017).

Estimates suggest that being able to detect ctDNA at a MAF of 0.1% would allow detection of tumours with a volume of 10 cm^3 (Abbosh et al. 2017). Improving this detection to a MAF of 0.01% using TAm-Seq V2 could allow detection of tumours with a volume of $<$3.5 cm^3. However, to detect a tumour with a volume of 1 cm^3 would require detection of a minimum MAF of 0.008% (Abbosh et al. 2017). This would require screening of over 30,000 DNA molecules and extraction from even larger volumes of plasma which may not be feasible particularly in future studies looking at a combination of assays and markers all of which will require increased volumes of blood to be collected.

An alternative approach to massively parallel TAm-Seq is selective sequencing of DNA

fragments based on cfDNA size and fragmentation features.

cfDNA fragments are commonly found to be around 167 bp in length corresponding to DNA wrapped around a nucleosome (147 bp) and linker DNA associated with histone H1. DNA fragments of this length are thought to be released into the circulating from cells undergoing apoptosis (Jahr et al. 2001; Lo et al. 2010; Chandrananda et al. 2015; Jiang et al. 2016). However, it has been shown that tumour derived cfDNA fragments can be a different size to non-tumour derived cfDNA fragments (Mouliere et al. 2011; Underhill et al. 2016; Jiang et al. 2015).

In this chapter I have found that tumour-derived cfDNA fragments can be shorter than non-tumour derived cfDNA fragments. The origin of these shorter fragments is still unknown but is likely to be a result of the mechanism of compaction and release into the circulation. One possibility is that shorter fragments could be released as a result of tumour proliferation (Abbosh et al. 2017).

I have shown that selective sequencing for shorter fragments can enrich for ctDNA in a sample measured by targeted sequencing and sWGS. This approach is similar to that applied to non-invasive prenatal testing to improve the sensitivity of detection of fetal cfDNA (Lo et al. 2010; Yu et al. 2014; Lun et al. 2008; Minarik et al. 2015).

A limitation of this analysis is that only samples from relapsed HGSOC were included. It is possible that the size of cfDNA fragments present at the time of newly diagnosed disease will differ due to difference in the biological features of newly diagnosed compared to relapsed disease. It is also possible that the volume of disease or stage of disease may effect the pattern of cfDNA fragments observed meaning that there may not be one size range that can be applied to enrich for tumour-derived cfDNA in all cases. This will be a particular issue if using in-vitro size selection due to the inevitable loss of material.

I have compared in-vitro size selection with in-silico size selection. In-vitro size selection adds marginally to the cost and time required for whole genome library preparation but does not effect the bioinformatics analysis. Although this approach led to a significant enrichment in tumour-derived cfDNA, material is inevitably lost during the size selection process. This could result in the loss of tumour derived cfDNA fragments important for downstream analysis. In comparison in-silico size selection does not add and additional step and does not add to the cost of library preparation but does add an additional step to the bioinformatics analysis. The enrichment observed following in-silico size selection was disappointingly not a large as following in-vitro size selection but there is no irreversible

loss of material meaning that all cfDNA fragments can be interrogated for analysis.

The use of selective sequencing may result in more samples being available for whole exome or whole genome analysis which is currently limited to those with a MAF above 5–10% (Murtaza et al. 2013; Belic et al. 2015). Selective sequencing may be useful to identify clinically actionable mutations thereby allowing entry into clinical trials and the application of targeted therapies. It is also possible that a selective sequencing approach may improve the rate of ctDNA detection in women with newly diagnosed HGSOC. However, it is also possible that size selection may result in some actionable mutations being missed as a result of being present on cfDNA fragments outside of the fragment size range of interest.

Other limitations of this approach are that only double stranded DNA has been sequenced which may result in biased information about DNA fragmentation sizes. The DNA extraction and sequencing methods may also provide additional biases to DNA fragments observed. Finally this investigation was limited to plasma. Different fragmentation patterns may be seen in other sample types including urine and serum.

Improved sensitivity for detection of circulating tumour DNA in newly diagnosed ovarian cancer

6.1 Introduction

Due to the significant enrichment in ctDNA seen in the relapsed HGSOC cases with selective sequencing using DNA fragment size information I wanted to apply this to the cohorts of women with newly diagnosed OC.

In this chapter I will describe the effect of in-silico size selection for DNA fragments on the t-MAD score in the UKOPS and CTCR-OV04 cohorts. I will discuss the effect of in-silico size selection for two different size ranges: 20–150 bp and 90–150 bp. For the UKOPS cohort all reads were downsampled to 3 million to allow direct comparisons between all samples. For the CTCR-OV04 cohort all reads were downsampled to 10 million reads again to allow direct comparisons between samples. Samples with read counts lower than this were excluded. I will also describe the effect of in-vitro size selection in the CTCR-OV04 cohort.

I will also discuss the effect of in-silico size selection on IchorCNA, a tool that has recently been developed by the Broad Institute for quantifying ctDNA from sWGS data (Adalsteinsson et al. 2017) and is therefore a similar metric to the t-MAD score. IchorCNA outputs a tumour fraction. I will discuss the tumour fraction output for two different sets of parameters. The first is recommended for samples with higher ctDNA fractions and I will refer to this as 'IchorCNA default'. The second assumes a lower ctDNA fraction and I will refer

to this as 'IchorCNA customised'.

In this chapter in-silico size selection and calculation of the t-MAD and IchorCNA values has been performed by Dineika Chandrananda.

6.2 Results

6.2.1 UKOPS cohort

In the UKOPS cohort the size selected t-MAD score was highly correlated with the *TP53* MAF determined by targeted sequencing in the HGSOC cases but not the non-HGSOC cases (Figure 6.1).

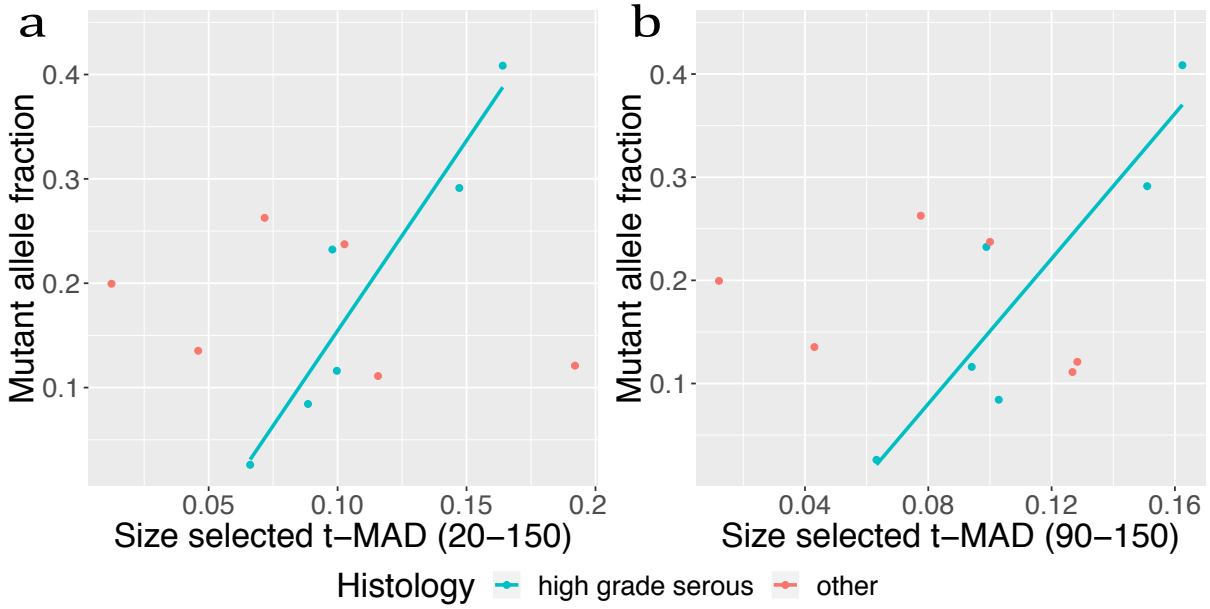


Figure 6.1: **a)** Comparison of t-MAD score following in-silico size selection for DNA fragments between 20–150 bp and *TP53* MAF for HGSOC cases (correlation coefficient 0.94, $p<0.005$) and non-HGSOC cases (correlation coefficient −0.40, $p=0.427$), . **b)** Comparison of t-MAD score following in-silico size selection for DNA fragments between 20–150 bp and *TP53* MAF for HGSOC cases (correlation coefficient 0.92, $p=0.010$) and non-HGSOC cases (correlation coefficient −0.31, $p=0.538$).

Following in-silico size selection copy number amplifications and losses are revealed in the plasma samples that match those seen in the tumour samples (Appendix 12.7). This

suggests that size selection is enriching for ctDNA and that the changes in t-MAD score following size selection are not an artefact of in-silico size selection .

The change in t-MAD score following size selection was variable across all cases in the UKOPS cohort (Figure 6.2). One example where we see a large increase in t-MAD score is TUKO01904 (Appendix 12.7). There are very few copy number changes in the unselected plasma sample however following size selection we see gains in chromosomes 7, 10 and 16 that are also present in the matched tumour sample. This is reflected by a 338% increase in t-MAD score following size selection.

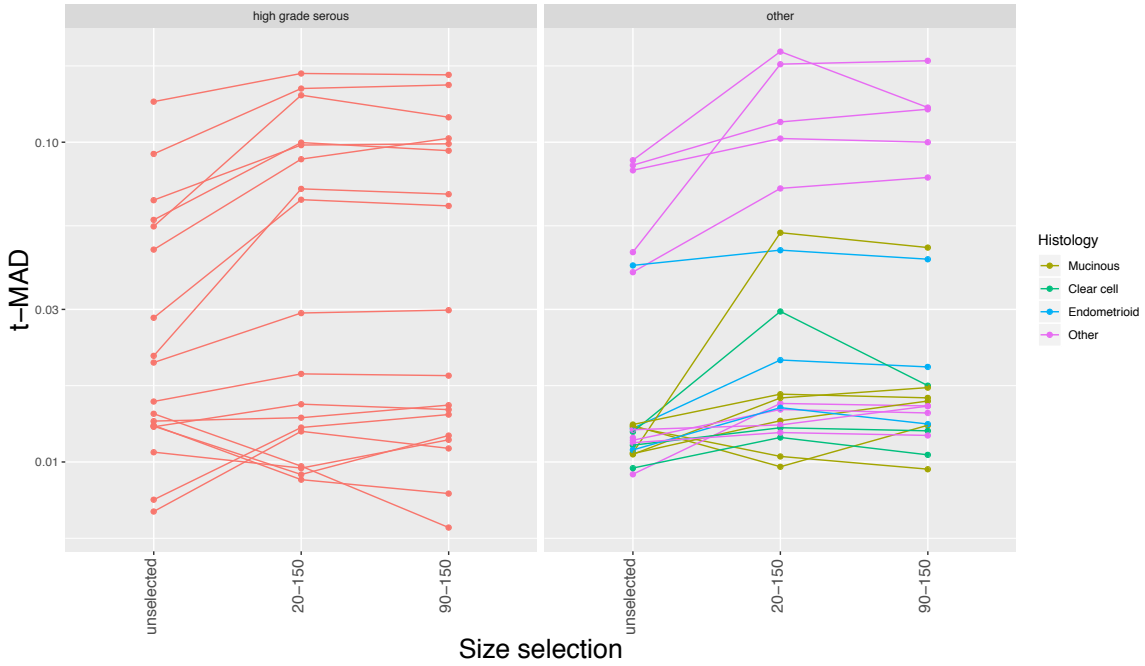


Figure 6.2: Unselected t-MAD score and size selected t-MAD score (logarithmic scale) for UKOPS plasma samples. Panel on the left shows HGSOC cases and panel on the right shows all other histological subtypes.

The median change in t-MAD was 31% (IQR 12–75%) and 30% (IQR 12–64%) for the whole cohort when selecting for DNA fragments between 20–150 bp and 90–150 bp respectively. For the HGSOC cases the median change in t-MAD was higher at 46% (IQR 6–77%) and 48% (IQR 12–80%) when selecting for DNA fragments between 20–150 bp and 90–150 bp respectively (Figure 6.3a).

In relapsed HGSOC we saw a greater increase in t-MAD score for those cases with a low unselected t-MAD score (see section 5.2.2, page 94) . This is not the case in the UKOPS cohort where we see no correlation between the unselected t-MAD score and the percentage change in t-MAD score following size selection (Figure 6.3b, c).

Before size selection 41% of the whole cohort and 56% of the HGSOC cases had a t-MAD score above the highest t-MAD score calculated for 46 healthy controls (Section 4.2.1, page 56). Following size selection the ctDNA detection rate increased to 53% for the whole cohort but did not change for the HGSOC cases (Figure 6.4).

After seeing an improvement in ctDNA detection in women with newly diagnosed OC using the size selected t-MAD score I wanted to investigate whether IchorCNA could improve detection rates further.

The tumour fraction calculated by IchorCNA was moderately correlated with MAF calculated by targeted sequencing in the UKOPS cohort (Figure 6.5).

Depending on the IchorCNA tumour fraction cut off used ctDNA detection ranged from 44–88% for the whole cohort and 44–83% for the HGSOC cases (Table 6.1, Figure 6.6). IchorCNA using the customised settings and the upper limit of the healthy controls as a cut off had the highest rate of detection. This was higher than the rate of detection using targeted sequencing, t-MAD score, or size selected t-MAD score.

IchorCNA settings	Cohort	Cut off	Detection (%)
default	all	10%	44
default	hgsoc	10%	44
default	all	5%	70
default	hgsoc	5%	67
default	all	3%	83
default	hgsoc	3%	83
default	all	healthy control	61
default	hgsoc	healthy control	61
customised	all	healthy control	88
customised	hgsoc	healthy control	83

Table 6.1: ctDNA detection rate for the whole UKOPS cohort and HGSOC cases alone using different cut off values for detection. 10%, 5% and 3% correlate to the tumour fraction. Healthy control is the highest IchorCNA tumour fraction value calculated using the matched parameters from 46 healthy controls.

In-silico size selection was then performed before calculating the IchorCNA tumour fraction to investigate if this would further improve detection rates. As with the unselected IchorCNA the rate of detection was better using the customised settings and was higher

in the HGSOC cases than the cohort as a whole (Table 6.2). However, there was no improvement in the rate of detection using the size selected IchorCNA compared to the unselected IchorCNA.

Size Selection (bp)	IchorCNA Settings	Cohort	Detection (%)
20–150	default	all	45
20–150	default	hgsoc	44
20–150	customised	all	73
20–150	customised	hgsoc	83
90–150	default	all	46
90–150	default	hgsoc	44
90–150	customised	all	69
90–150	customised	hgsoc	72

Table 6.2: Detection using highest IchorCNA tumour fraction calculated using matched parameters in cohort of 46 healthy controls as a cut off.

ROC analysis confirmed that the customised unselected IchorCNA tumour fraction was best at discriminating the OC cases from healthy controls for the whole cohort (Figure 6.7a) with a sensitivity of 88% for a specificity of 100% (AUC 0.996) and HGSOC cases (sensitivity 83%, specificity 100%, AUC 0.996) (Figure 6.7b).

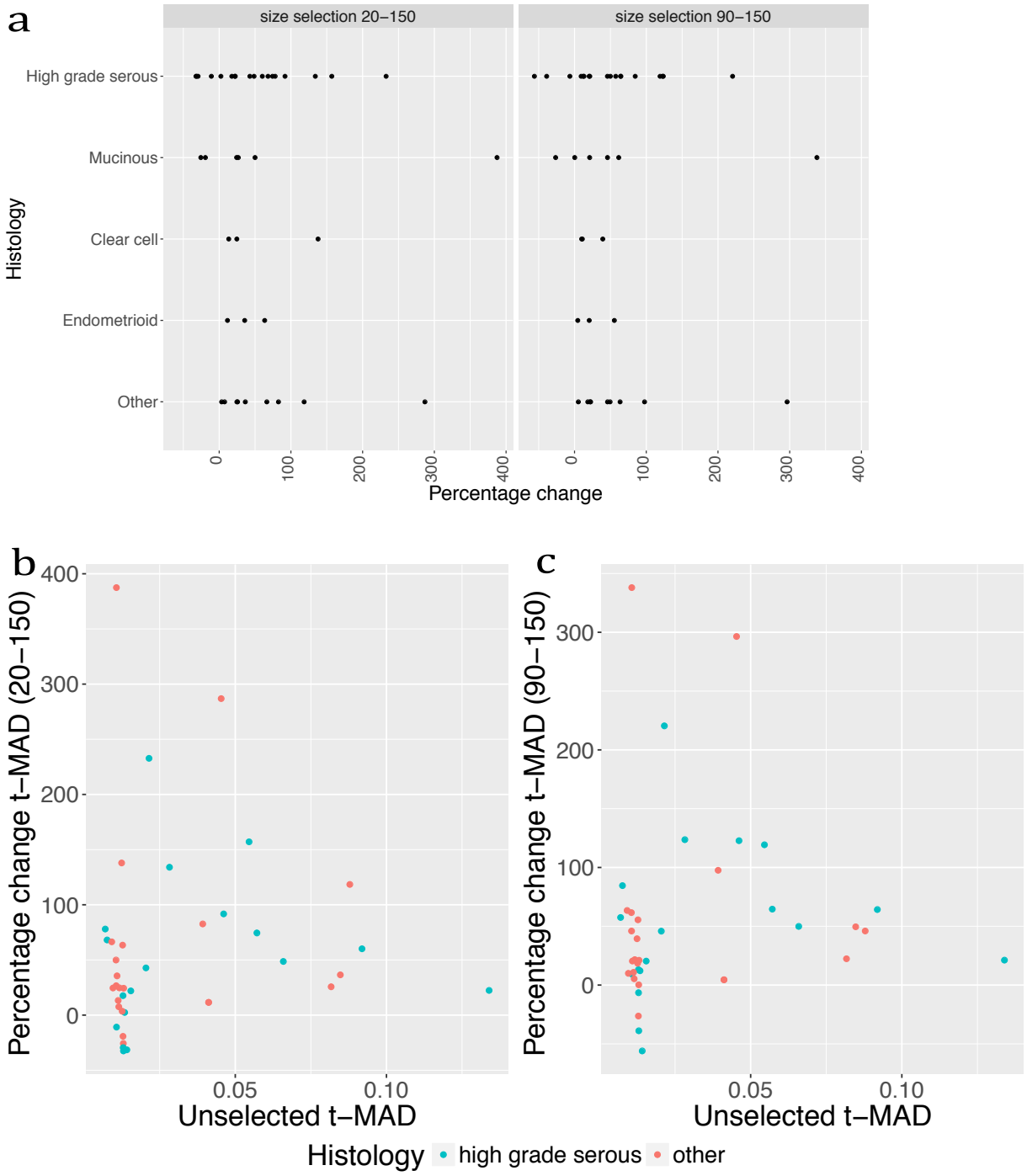


Figure 6.3: **a)** Percentage change in t-MAD score following in-silico size selection for DNA fragments between 20–150 bp and 90–150 bp by histological subtype. **b)** Comparison of unselected t-MAD score with percentage change in t-MAD score following in-silico size selection for DNA fragments between 20–150 bp for HGSOC cases (correlation coefficient 0.14, p=0.582) and non-HGSOC cases (correlation coefficient 0.08, p=0.732). **c)** Comparison of unselected t-MAD score with percentage change in t-MAD score following in-silico size selection for DNA fragments between 90–150 bp for HGSOC cases (correlation coefficient 0.13, p=0.599) and non-HGSOC cases (correlation coefficient 0.05, p=0.817).

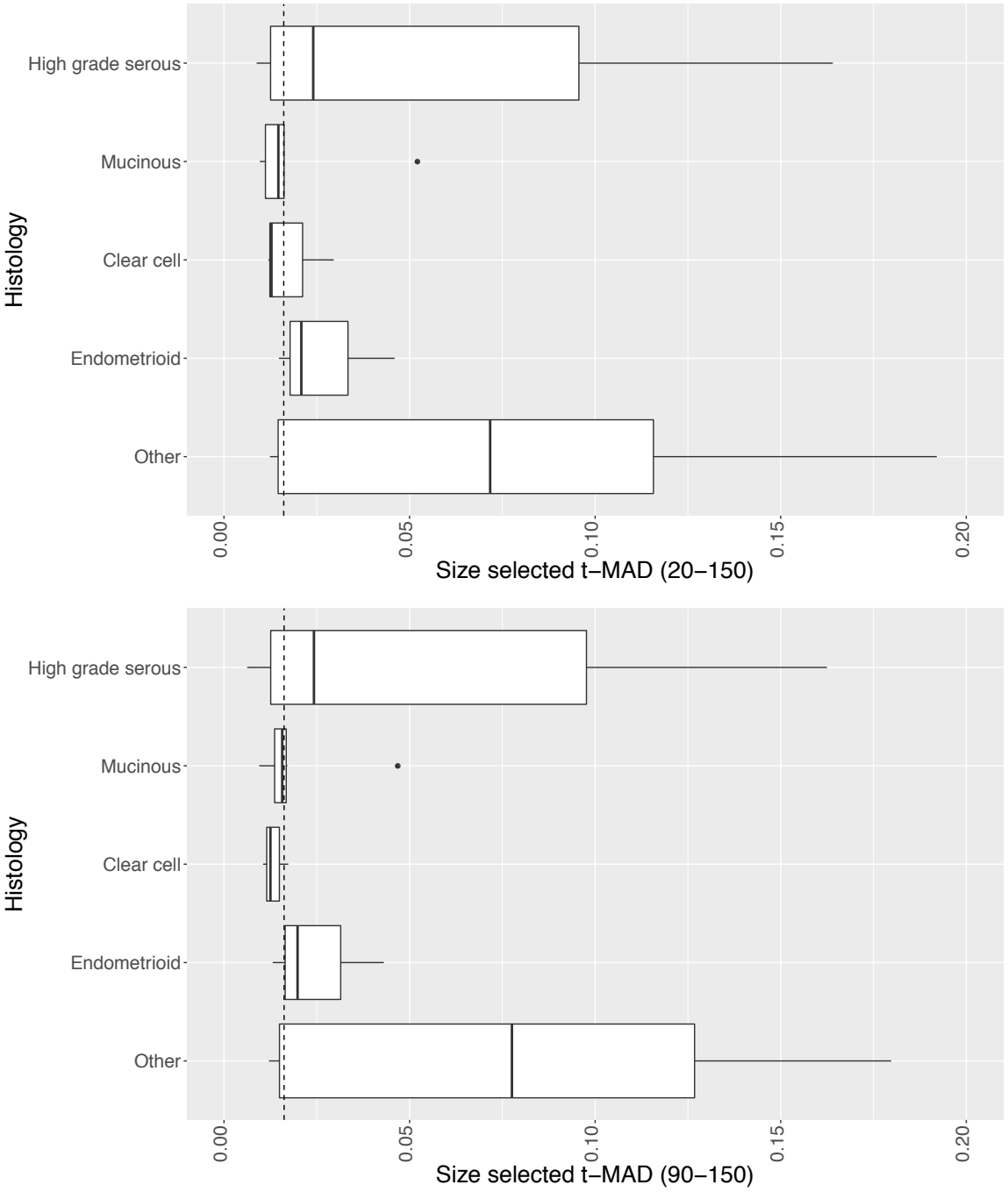


Figure 6.4: Size selected t-MAD score for all UKOPS cases calculated after down sampling to 3 million reads by histological subtype. Dashed line indicates the highest size selected t-MAD score calculated in a cohort of 46 healthy controls. Top panel shows in-silico size selection for fragments between 20–150 bp and bottom panel size selection for fragments between 90–150 bp.

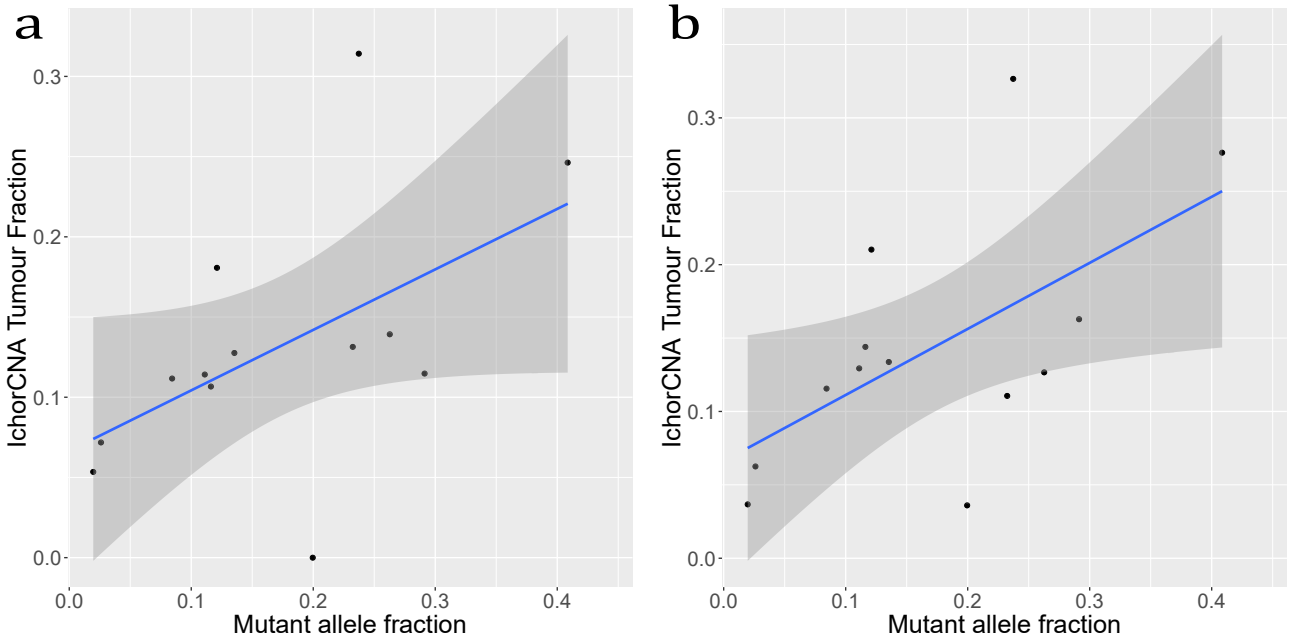


Figure 6.5: **a)** Comparison between MAF measured by targeted sequencing and unselected IchorCNA tumour fraction using default settings (correlation coefficient 0.52). **b)** Comparison between MAF measured by targeted sequencing and unselected IchorCNA tumour fraction using customised settings (correlation coefficient 0.59).

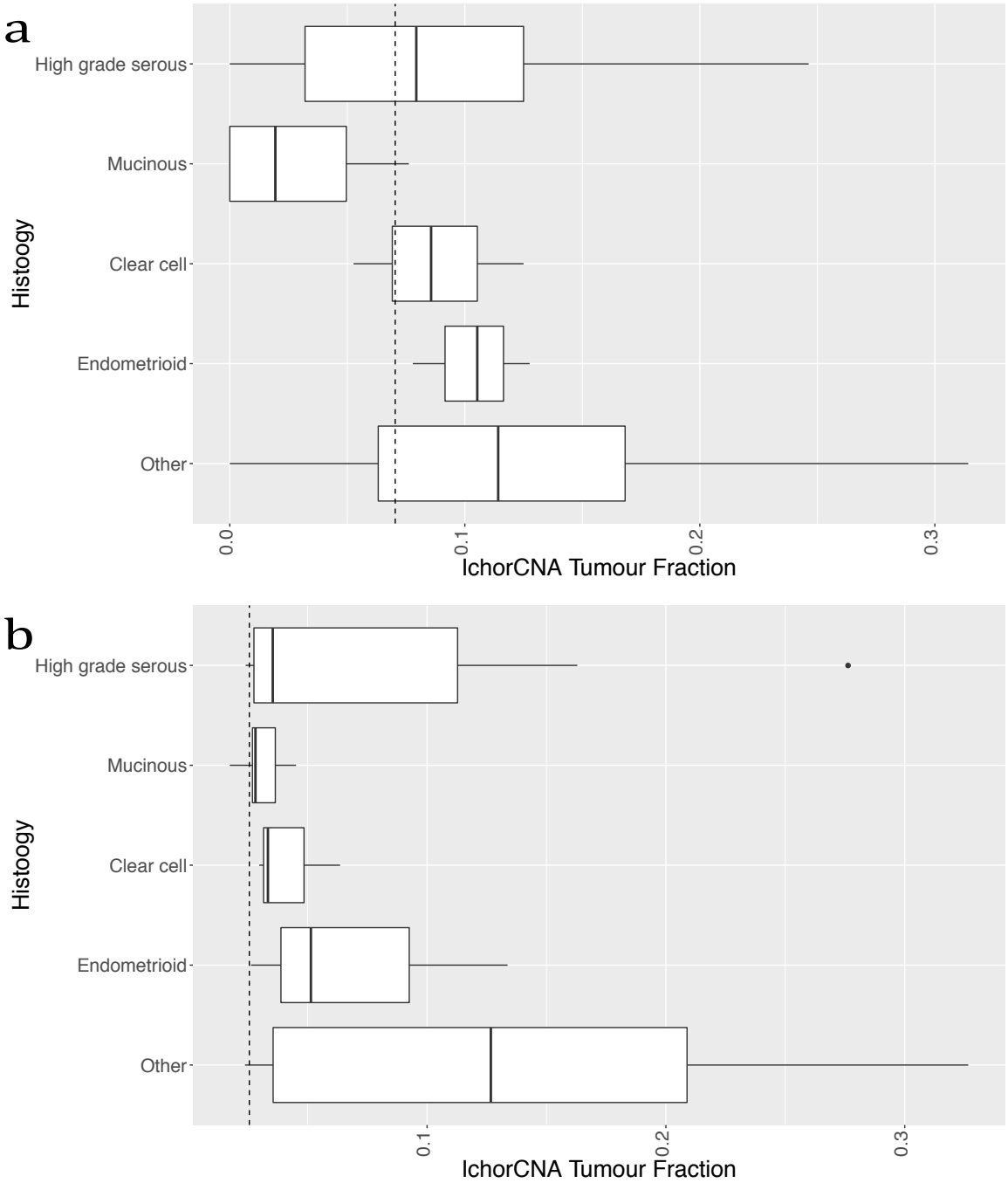
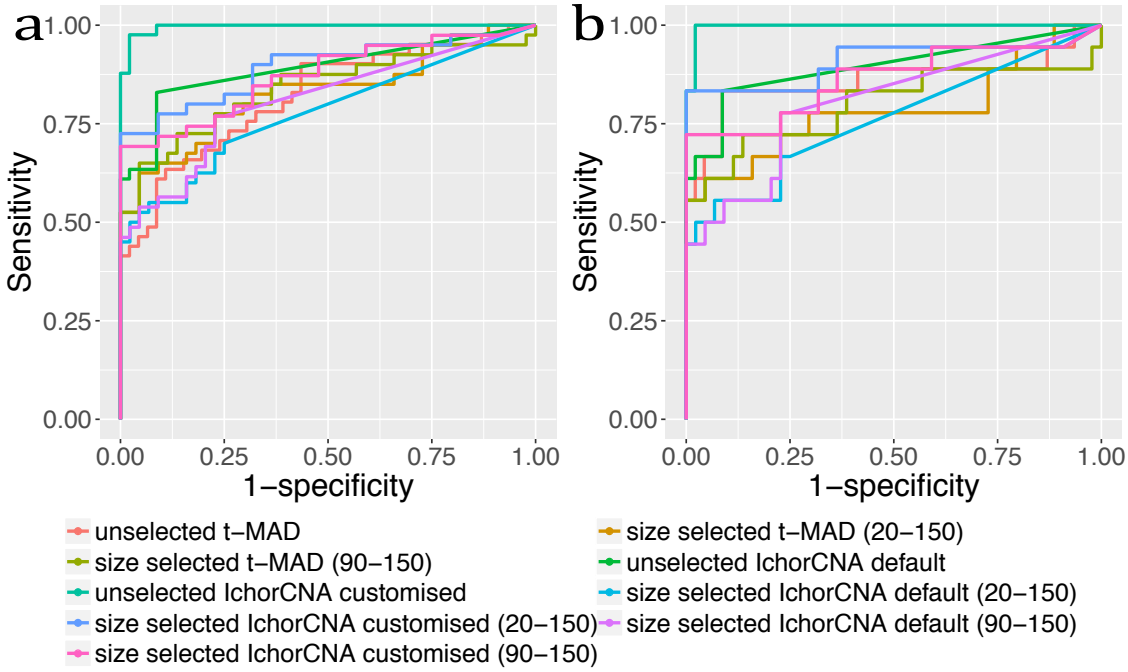


Figure 6.6: **a)** IchorCNA tumour fraction calculated using default settings for all UKOPS cases by histological subtype. Dashed line indicates the highest tumour fraction calculated by default settings in a cohort of 46 healthy controls. **b)** IchorCNA tumour fraction calculated using customised settings to allow for low levels of ctDNA for all UKOPS cases by histological subtype. Dashed line indicates the highest tumour fraction calculated by customised settings in a cohort of 46 healthy controls.



c

Method	AUC Whole Cohort	AUC HGSOC Cases
unselected t−MAD	0.8202545	0.8586957
size selected t−MAD (20−150)	0.8301136	0.7853535
size selected t−MAD (90−150)	0.8397727	0.8005051
unselected IchorCNA default	0.8897137	0.8937198
unselected IchorCNA customised	0.9957582	0.9963768
size selected IchorCNA default (20−150)	0.7784091	0.7613636
size selected IchorCNA customised (20−150)	0.8937500	0.9090909
size selected IchorCNA default (90−150)	0.8097319	0.8042929
size selected IchorCNA customised (90−150)	0.8717949	0.8636364

Figure 6.7: **a)** ROC analysis comparing whole cohort to 46 healthy controls. **b)** ROC analysis comparing HGSOC cases to 46 healthy controls. **c)** AUC for all methods for whole cohort (plot a) and HGSOC cases (plot b).

6.2.2 CTCR-OV04 cohort

To investigate whether the size selected t-MAD score or the customised IchorCNA tumour fraction performed as well in a larger cohort of HGSOC cases I applied these methods to the newly diagnosed CTCR-OV04 cohort.

Firstly I wanted to investigate the effect of in-vitro size selection on the t-MAD score. In-vitro size selection (PippinHT 3% cassette selecting for fragments 90–150 bp) was performed for 34 cases with a starting MAF between 0.8–1%. AF's in this range are within the limits of detection of both targeted sequencing and sWGS but are low enough that an enrichment in ctDNA with size selection could potentially be observed. Whole genome library preparation was performed for the unselected and size selected sample using the Rubicon ThruPlex DNA-Seq kit. In 31/32 (97%) of the samples an increase in t-MAD score was seen following in-vitro size selection of the samples (Figure 6.8). The median change in t-MAD between the unselected sample and size selected sample was 211% (IQR 86% to 283%) suggesting a significant enrichment in ctDNA in these samples.

As in-vitro size selection requires use of further material from the sample I performed size selection of the whole genome library for 22 of the cases (PippinHT 2% cassette selecting for fragments 220–300 bp). The median change in t-MAD score calculated from sWGS of the the unselected sample and sWGS of the size selected library was 4% (IQR −12% to 18%). Whereas size selection of the input DNA led to a significant enrichment this was not so far achieved with size selection of the sequencing library (Figure 6.8a).

Another way of preserving material is to perform in-silico size selection of sWGS data from an unselected sample. I have previously shown that in-silico size selection can increase ctDNA detection both by targeted sequences and sWGS (see section 5.2.2). In the CTCR-OV04 newly diagnosed cohort the median increase in t-MAD following in-silico size selection was 97% (IQR 21–140%). Although this suggests an increase in ctDNA in these samples the enrichment is not as significant as that seen following in-vitro size selection (Figure 6.8b). There is therefore a trade off between a greater increase in the amount of ctDNA accessible in a sample using in-vitro size selection and retaining material for other applications whilst still observing a moderate increase in the amount of ctDNA in a sample using in-silico size selection. In-silico size selection of libraries produced from in-vitro size selected samples did not further increase the t-MAD score (median change −3%, IQR −13% to 18%).

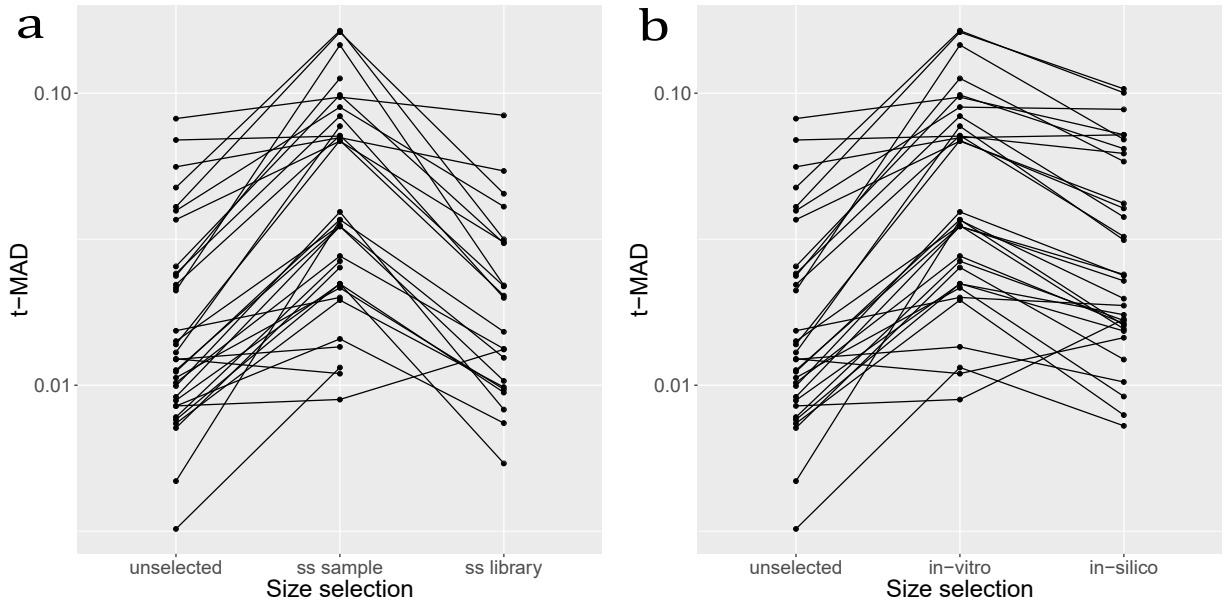


Figure 6.8: **a)** t-MAD score for 32 unselected samples, 32 in-vitro size selected samples (ss sample) and 22 in-vitro size selected libraries (ss library). **b)** t-MAD score for 32 samples calculated following sWGS of unselected sample, in-vitro size selection for DNA fragments between 90–150 bp and in-silico size selection of DNA fragments between 90–150 bp.

In the whole newly diagnosed CTCR-OV04 cohort an increase in t-MAD score was seen in 116/135 (85%) and in 113/130 (87%) of cases following in-silico size selection for DNA fragments 20–150 bp and 90–150 bp respectively. This increase was seen in all stages of disease (Figure 6.9). Interestingly the highest median increase in t-MAD score was seen in the cases with early stage disease following in-silico size selection for DNA fragments between 20–150 bp (Table 6.3). Importantly size selection did not lead to an increase in t-MAD score for the cohort of 46 healthy controls and size selection for DNA fragments between 20–150 bp did not lead to an increase in t-MAD score for the cohort of 32 benign controls. It is therefore possible that size selection for DNA fragments between 20–150 bp might be optimal both in terms of increased rates of detection in early stage disease but also in better discrimination between HGSOC cases and benign controls.

Stage	Size range	Median	Q1	Q3
healthy	20–150	11	−7	26
healthy	90–150	11	−7	26
benign	20–150	24	−9	54
benign	90–150	36	4	78
I	20–150	124	78	150
I	90–150	96	57	133
II	20–150	176	119	193
II	90–150	130	72	169
III	20–150	85	27	137
III	90–150	95	29	153
IV	2–150	77	14	116
IV	90–150	63	14	128

Table 6.3: Median and IQR percentage change in t-MAD following in-silico size selection for DNA fragments between 20–150 bp and 90–150 bp.

The rate of detection of ctDNA using the unselected t-MAD score in the CTCR-OV04 cohort was 39% using the upper limit of the healthy controls as a cut off and 26% using the upper limit of the benign controls as a cut off (see section 4.2.2, page 63). Using the size selected t-MAD score this increased to 67% when using the upper limit of the healthy controls as a cut off and 51% when using the upper limit of the benign controls as a cut off (Figure 6.10a,b). Using in-silico size selection for DNA fragments between 20–150 bp included detection of ctDNA in 88% of the early stage cases using the healthy control cut off and 75% of the early stage cases using the benign control cut off (Figure 6.10a).

Depending on the stage of disease and cut off used between 25–100% of cases that were negative for ctDNA by targeted sequencing for *TP53* were positive for ctDNA using the size selected t-MAD score (Figure 6.10c). In particular 83% of the stage 1 *TP53* negative cases and 100% of the stage II *TP53* negative cases were detected using the size selected t-MAD score (20–150).

The smallest tumour volume detected using the unselected t-MAD score using both the upper limit of the healthy and the upper limit of the benign controls as a cut off was 67.9 cm^3 (see section 4.2.2, page 64). The size selected t-MAD score performed better detecting ctDNA in a case with a tumour volume of 42.3 cm^3 (Figure 6.11).

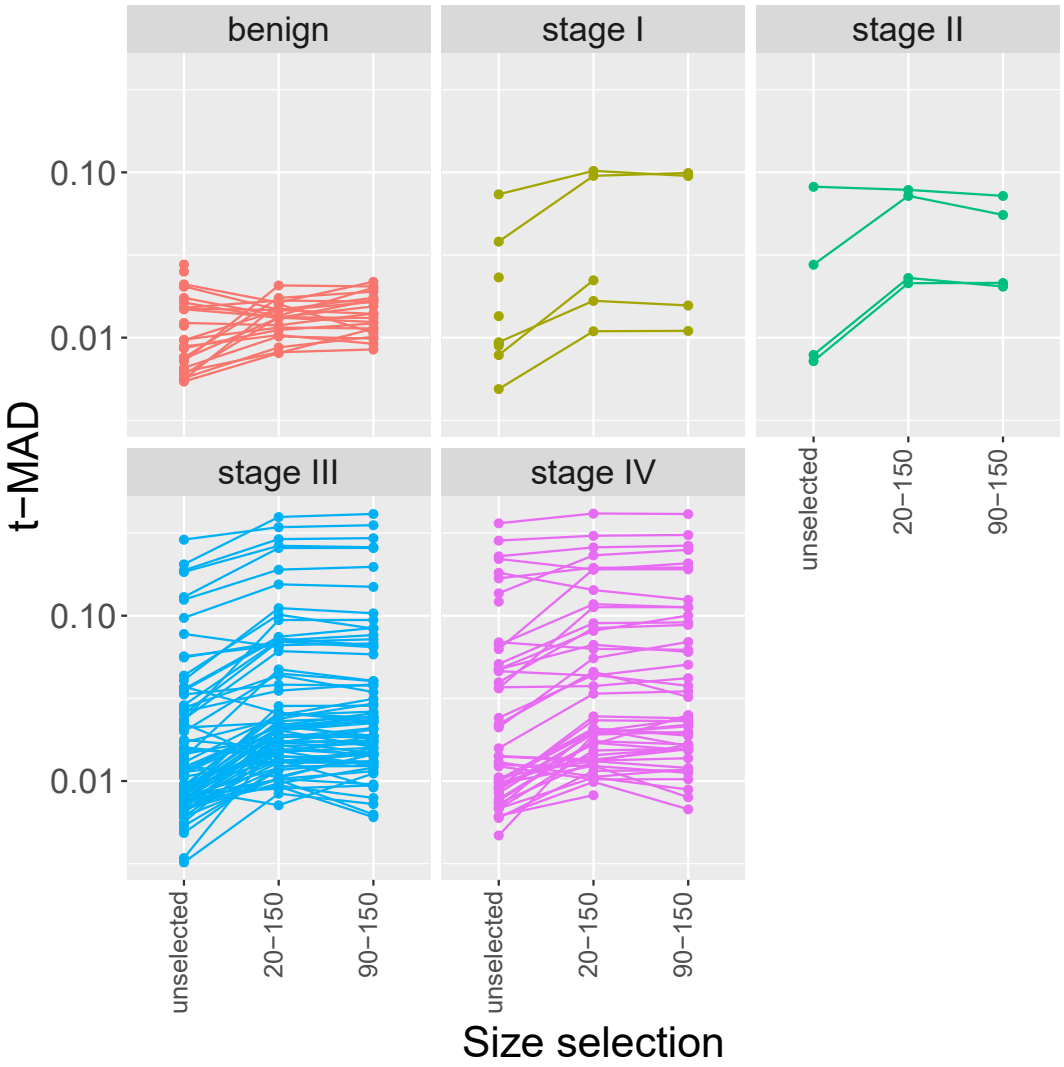


Figure 6.9: t-MAD score calculated from unselected sWGS data and following in-silico size selection for DNA fragments between 20–150 bp and 90–150 bp by stage of disease.

ROC analysis confirmed that the size selected t-MAD score performed better than the unselected t-MAD score in discriminating HGSOC cases from both healthy and benign controls (Figure 6.12). A maximum sensitivity of 65% for a specificity of 100% was observed when comparing the cases to the healthy controls and a maximum sensitivity of 50% for a specificity of 100% when comparing the cases to the benign controls. Importantly a good sensitivity of 88% for a specificity of 100% was observed when comparing the stage I/II HGSOC cases (with the caveat that this is a low number of samples) to the healthy controls and a sensitivity of 77% for a specificity of 100% when comparing to the benign controls (Figure 6.12c, d). The ability to discriminate early stage disease from benign controls with a good sensitivity and specificity needs confirming in larger studies but is promising for the development of a diagnostic biomarker.

In relapsed HGSOC we found that the addition of fragmentation features increased our ability to discriminate between HGSOC cases from healthy controls (in press, Science Translational Medicine). I therefore wanted to investigate whether the same was true in newly diagnosed disease. The proportion (P) of DNA fragments observed by sWGS in multiple size ranges, the ratios of proportions of fragments in these size ranges, the proportion of fragments greater than a certain size and the amplitude of oscillations in fragment size density with 10 bp periodicity observed below 150 bp were calculated for the healthy controls, benign controls and HGSOC cases (Figure 5.17a). There were no differences between the fragmentation features across different stages of disease and there was no feature that was able to clearly discriminate between newly diagnosed HGSOC cases and healthy and/or benign controls (Figure 6.13).

ROC analysis confirmed that the fragmentation features alone were not a good discriminator of newly diagnosed HGSOC cases from controls (Figure 6.14). The proportion of fragments between 250–320 bp (P250–320) was best able to discriminate the HGSOC cases from the healthy controls however the sensitivity was low at 44% for a specificity of 100%. The ratio of the proportion of fragments between 20–150 bp to the proportion of fragments between 180–220 bp (P20–150:P180–220) was best able to discriminate the HGSOC cases from the benign controls however, again the sensitivity was low at 38% for a specificity of 95%.

As the IchorCNA tumour fraction had performed well in the UKOPS cohort I applied it to the newly diagnosed CTCR-OV04 cohort. High correlation was seen between the *TP53* MAF and the IchorCNA tumour fraction (Figure 6.15a, b). There was no correlation between the IchorCNA tumour fraction and tumour volume (Figure 6.15c, d). Depending on the parameters and cut off used ctDNA detection using IchorCNA tumour fraction ranged from 12–57% for the whole cohort (Table 6.4, Figures 6.15e, f). Using the unselected IchorCNA tumour fraction did not improve the rates of ctDNA detection over the size selected t-MAD score.

Unlike in the UKOPS cohort (see section 6.2.1, page 106) the number of cases with a ctDNA fraction above 10%, 5% and 3% increased following in-silico size selection for DNA fragments between 20–150 bp and 90–150 bp using both the default and customised IchorCNA settings (Figure 6.16). ROC analysis showed that the size selected IchorCNA tumour fraction (20–150 bp and 90–150 bp) performed marginally better at distinguishing HGSOC cases from healthy controls (Figure 6.17a) than size selected t-MAD but did not help the discrimination between HGSOC cases and benign controls (Figure 6.17b).

IchorCNA Settings	Cut off	Detection (%)
default	10%	24
customised	10%	22
default	5%	54
customised	5%	34
default	3%	57
customised	3%	45
default	healthy control	37
customised	healthy control	57
default	benign control	16
customised	benign control	12

Table 6.4: ctDNA detection rate for the newly diagnosed CTCR-OV04 cohort using IchorCNA tumour fraction with different cut off values for detection. 10%, 5% and 3% correlate to the tumour fraction. Healthy control is the highest IchorCNA tumour fraction value calculated using the matched parameters from 46 healthy controls. Benign control is the highest IchorCNA tumour fraction value calculated using the matched parameters from 32 benign controls.

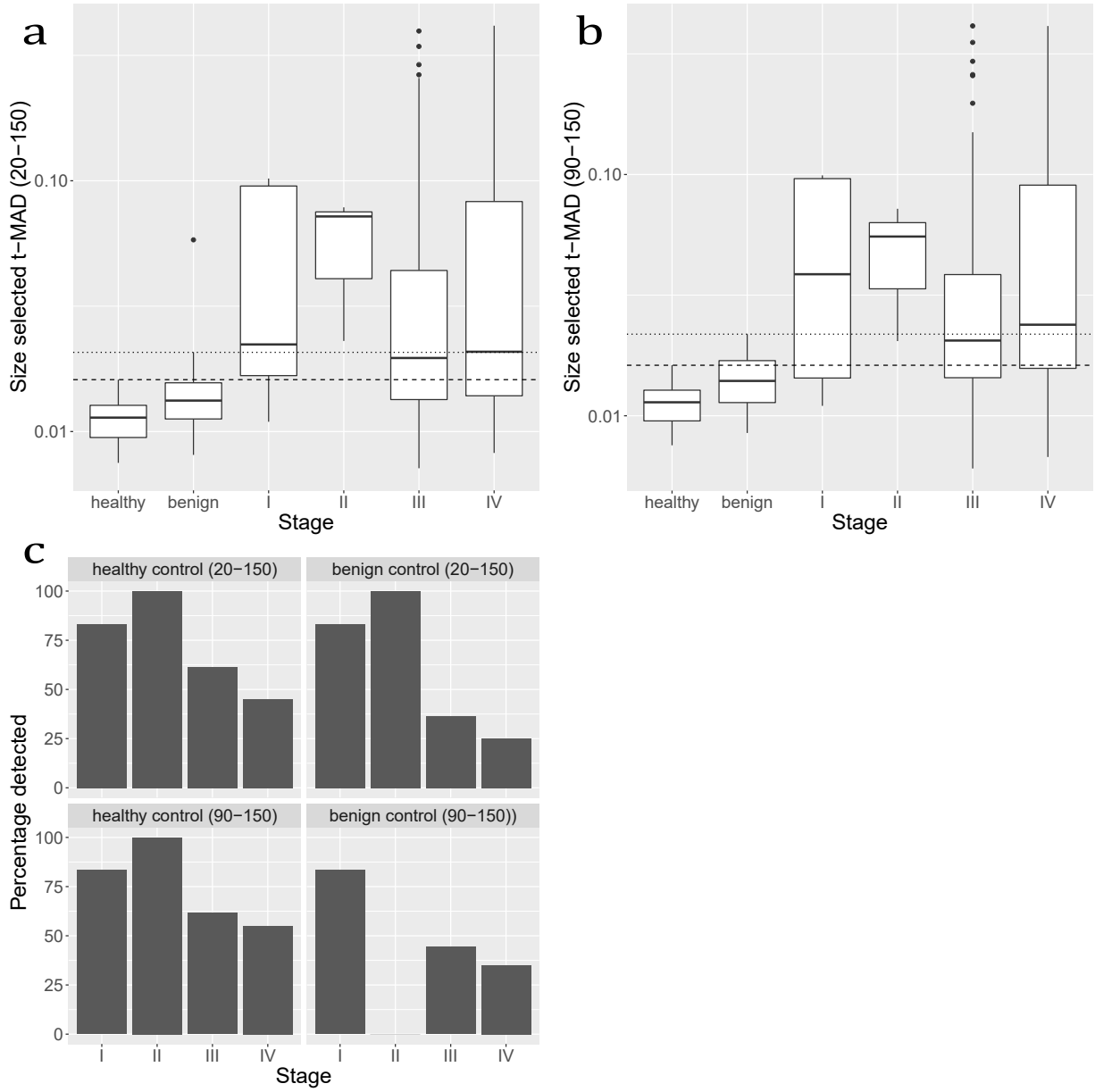


Figure 6.10: **a)** t-MAD score calculated following in-silico size selection for DNA fragments between 20–150 bp from sWGS data for 135 HGSOC cases by stage, 22 benign controls and 44 healthy controls. Dashed line indicates the highest t-MAD score calculated in the cohort of healthy controls. Dotted line indicates the highest t-MAD score calculated in the cohort of benign controls excluding one outlier. **b)** t-MAD score calculated following in-silico size selection for DNA fragments between 90–150 bp from sWGS data for 130 HGSOC cases by stage, 21 benign controls and 44 healthy controls. Dashed line indicates the highest t-MAD score calculated in the cohort of healthy controls. Dotted line indicates the highest t-MAD score calculated in the cohort of benign controls. t-MAD score was not calculated for outlier due to low coverage. **c)** Percentage of ctDNA negative cases by targeted sequencing for *TP53* that are detected using size selected t-MAD score using either highest value in the healthy controls or highest value in the benign controls as a cut off. (Stage I n=6, stage II n=1, stage III n=52, stage IV n=20)

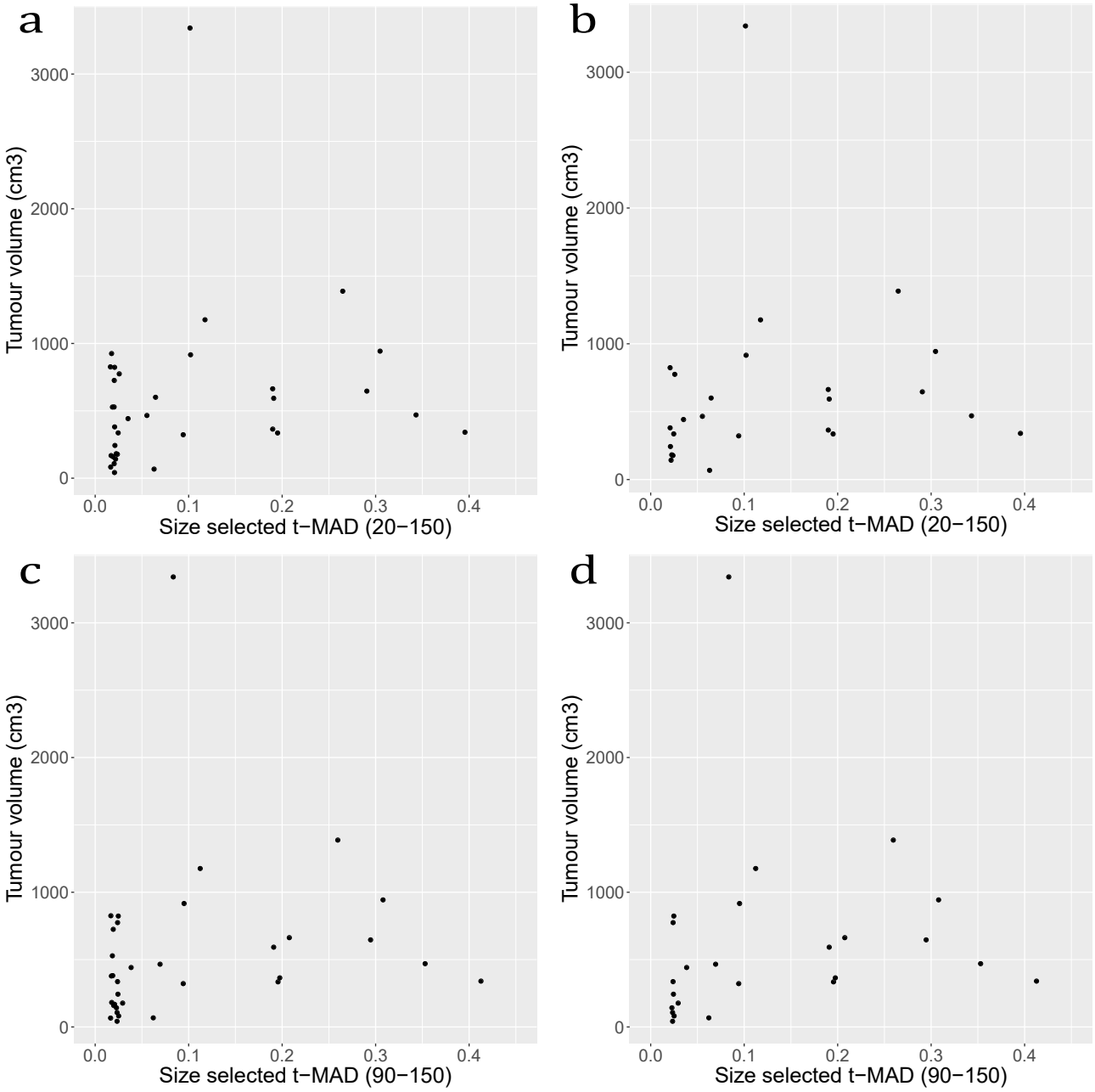


Figure 6.11: Comparison of size selected t-MAD score for detected cases with total tumour volume measured by 3-D CT reconstruction. **a)** 20–150 bp, cut off upper limit healthy controls. **b)** 20–150 bp, cut off upper limit benign controls. **c)** 90–150 bp, cut off upper limit healthy controls. **d)** 90–150 bp, cut off upper limit benign controls.

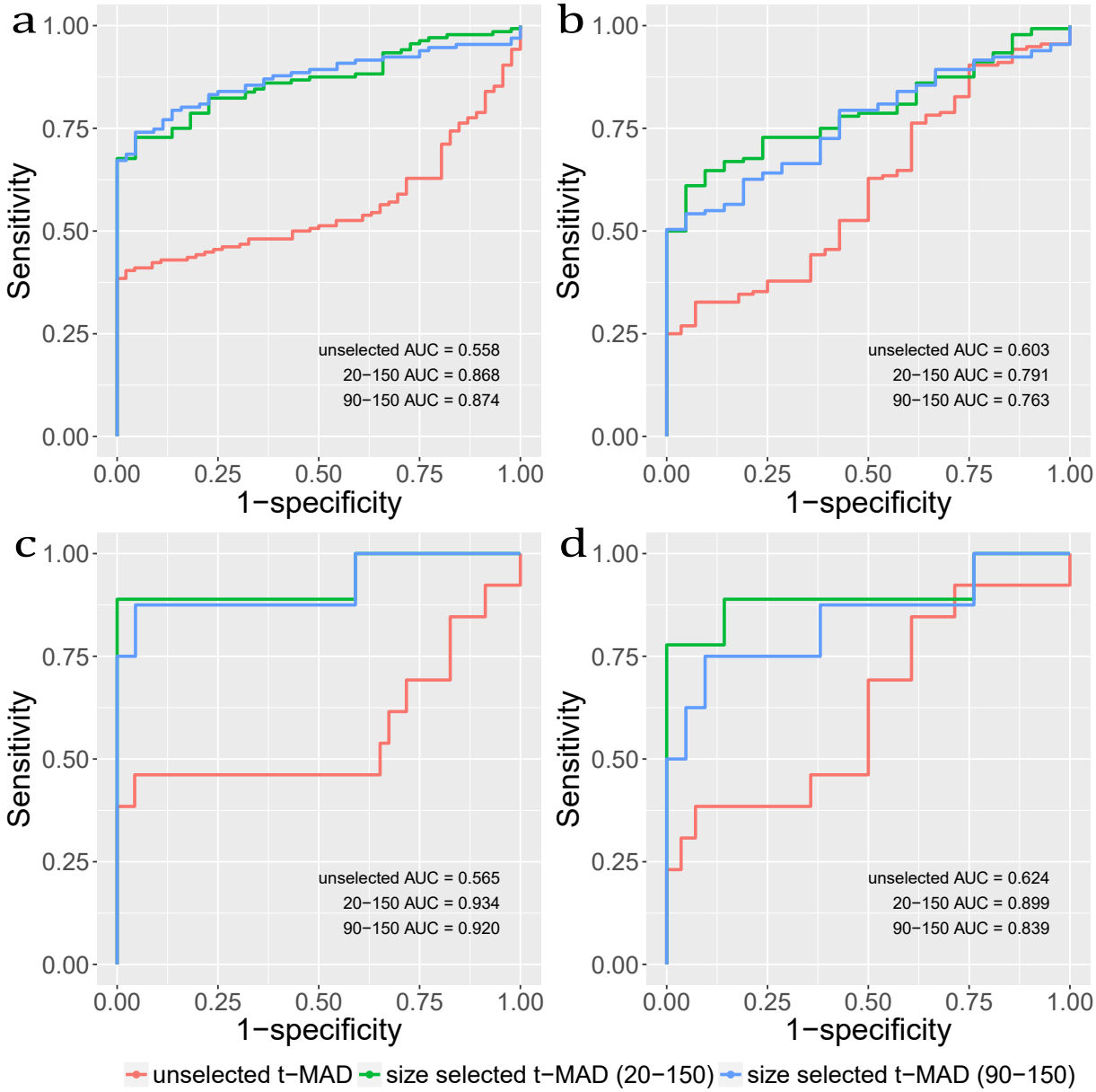


Figure 6.12: ROC analysis of unselected t-MAD score, t-MAD score following in-silico size selection for DNA fragments between 20–150 bp and t-MAD score following in-silico size selection for DNA fragments between 90–150 bp. **a)** Comparing HGSOC cases to healthy controls. **b)** Comparing HGSOC cases to benign controls. **c)** Comparing stage I/II HGSOC cases to healthy controls. **d)** Comparing stage I/II HGSOC cases to benign controls.

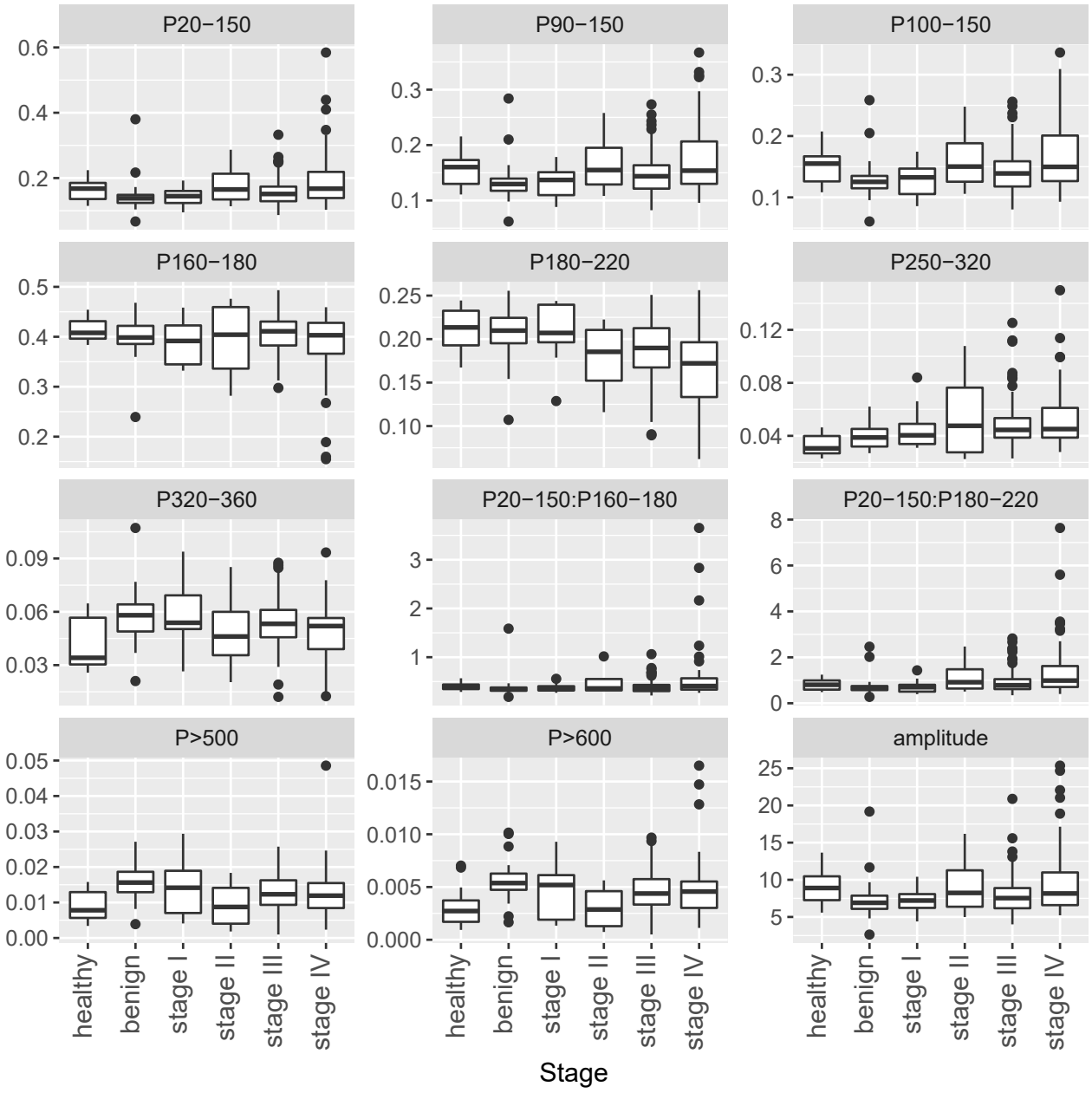


Figure 6.13: The proportion (P) of DNA fragments observed by sWGS in multiple size ranges (P20–150, P90–150, P100–150, P160–180, P180–220, P250–320, P320–360), the ratios of proportions of fragments in these size ranges (P20–150:P160–180, P20–150:P180–220), the proportion of fragments greater than a certain size (P>500, P>600) and the amplitude of oscillations in fragment size density with 10 bp periodicity observed below 150 bp for healthy controls, benign controls and HGSOC cases by stage.

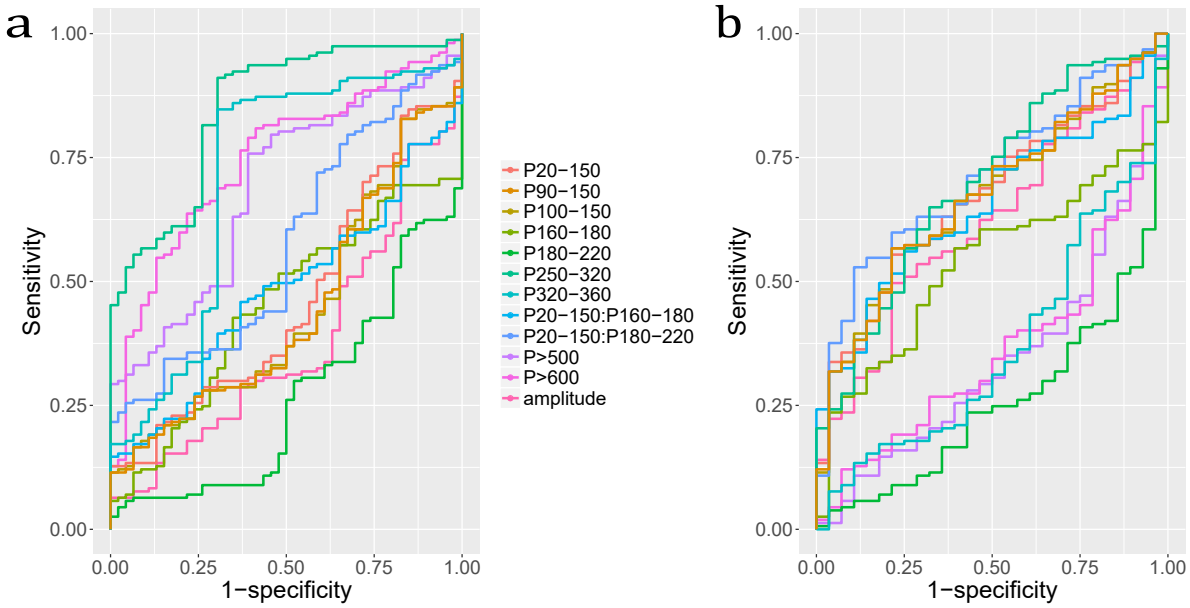


Figure 6.14: ROC analysis of the proportion (P) of DNA fragments observed by sWGS in multiple size ranges (P20–150, P90–150, P100–150, P160–180, P180–220, P250–320, P320–360), the ratios of proportions of fragments in these size ranges(P20–150:P160–180, P20–150:P180–220), the proportion of fragments greater than a certain size (P>500, P>600) and the amplitude of oscillations in fragment size density with 10 bp periodicity observed below 150 bp **a)** Comparing HGSOC cases to healthy controls. **b)** Comparing HGSOC cases to benign controls.

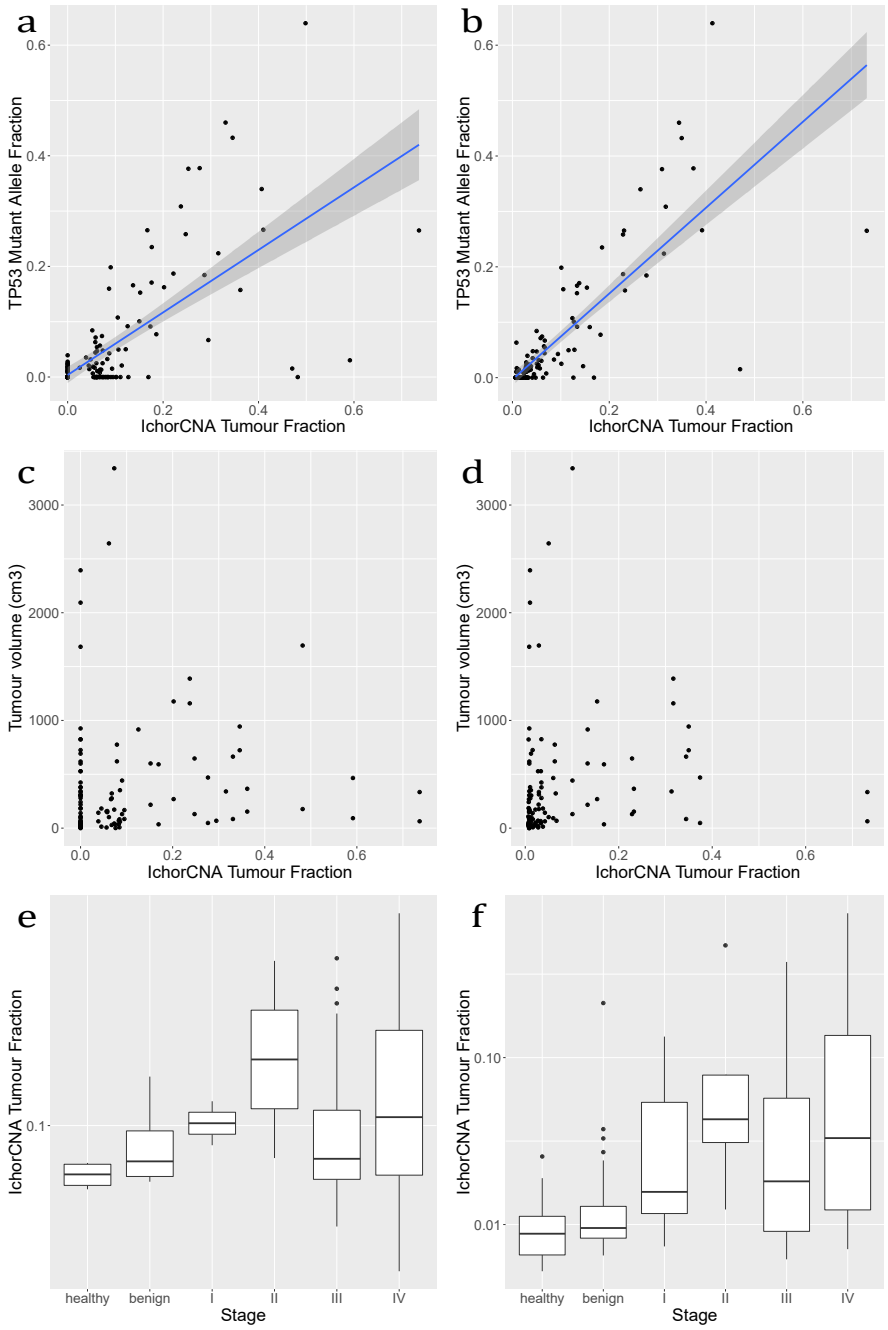


Figure 6.15: **a)** Comparison between unselected IchorCNA tumour fraction calculated with default settings and *TP53* MAF (correlation coefficient 0.69). **b** Comparison between unselected IchorCNA tumour fraction calculated with customised settings and *TP53* MAF (correlation coefficient 0.81). **c)** Comparison between unselected IchorCNA tumour fraction calculated with default settings and tumour volume. **d)** Comparison between unselected IchorCNA tumour fraction calculated with customised settings and tumour volume. **e)** Unselected IchorCNA tumour fraction calculated using default settings by stage of disease. **f)** Unselected IchorCNA tumour fraction calculated using customised settings by stage of disease.

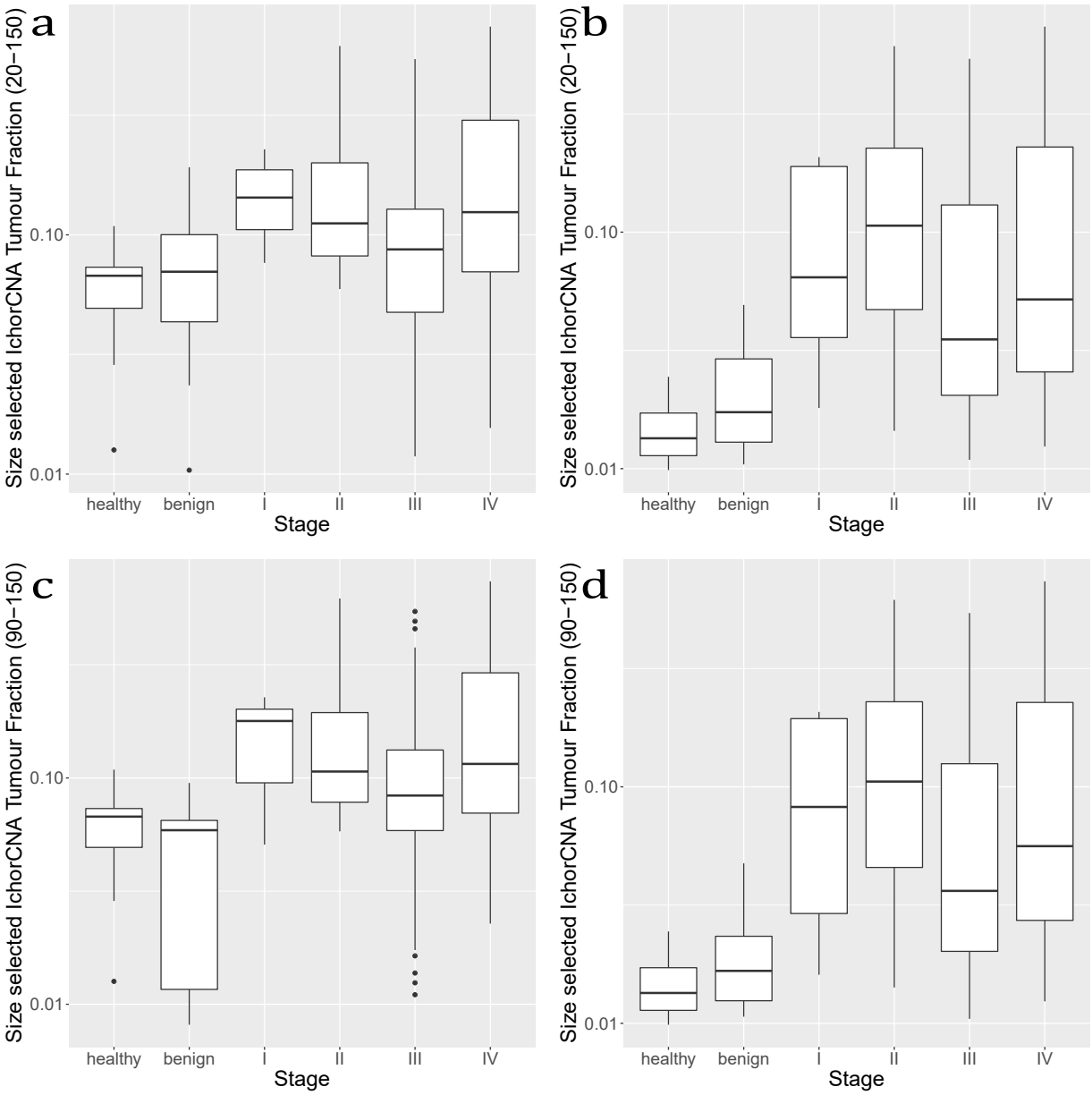


Figure 6.16: IchorCNA tumour fraction calculated following in-silico size selection by stage of disease. **a)** 20–150 bp using default settings. **b)** 20–150 bp using customised settings. **c)** 90–150 bp using default settings. **d)** 90–150 bp using customised settings.

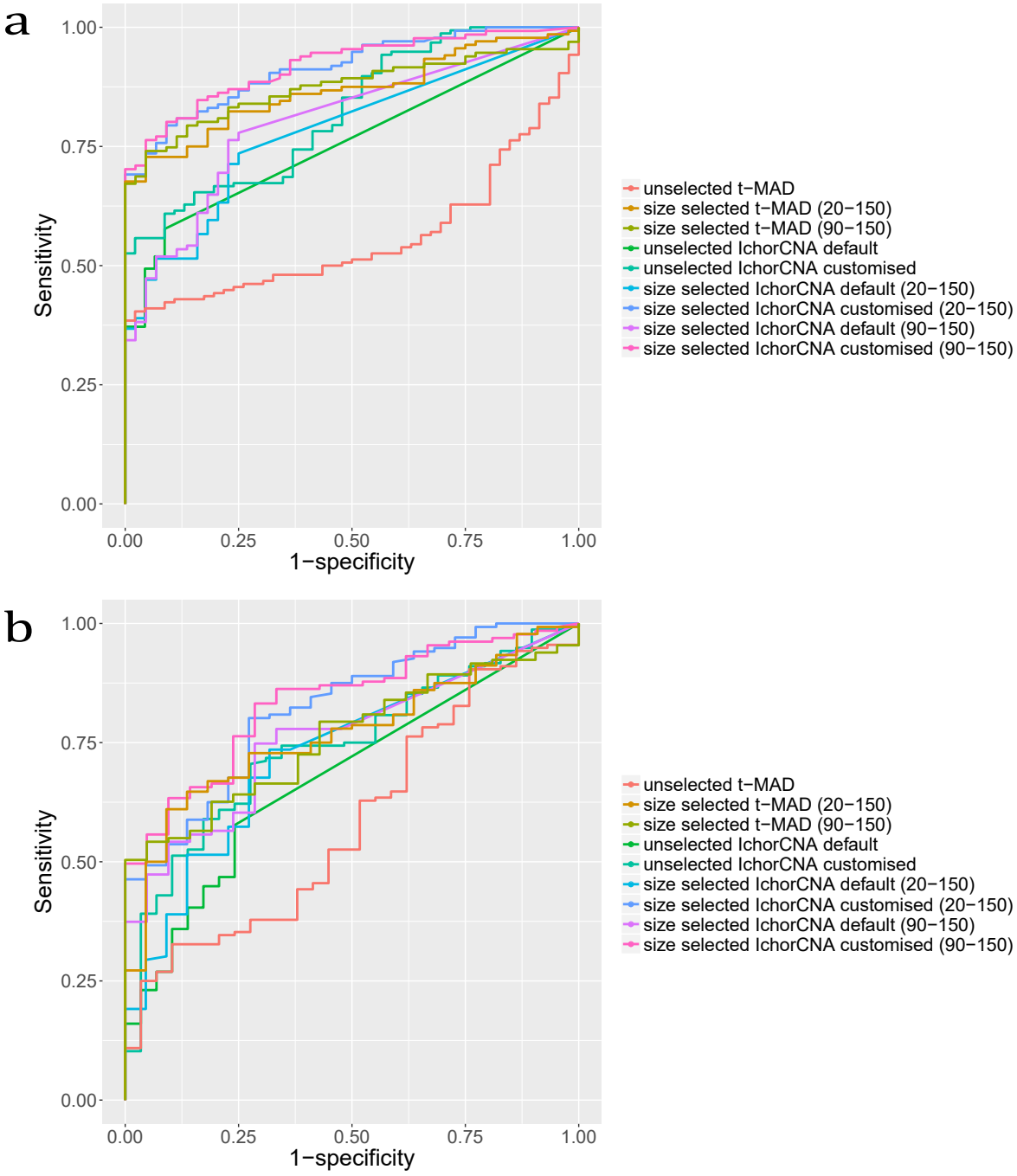


Figure 6.17: **a)** ROC analysis comparing HGSOC cases and healthy controls with t-MAD, size selected t-MAD and IchorCNA values. **b)** ROC analysis comparing HGSOC cases and benign controls with t-MAD, size selected t-MAD and IchorCNA values.

6.3 Discussion

I have previously shown that ctDNA can be detected in 56–60% of women with newly diagnosed HGSOC using targeted sequencing for *TP53* and sWGS (Section 4.2).

Other studies have identified ctDNA in 71–100% of cases of newly diagnosed OC (Bettegowda et al. 2014; Phallen et al. 2017). These studies used complex library preparation methods, complex bioinformatics analysis and are therefore costly and not applicable to clinical use currently. These studies also compared results in cases to healthy controls and have not looked at discriminating OC cases from benign controls.

In this chapter I have discussed the effect of in-vitro and in-silico size selection on the t-MAD score. I have also discussed IchorCNA, a tool that has recently been developed by the Broad Institute for quantifying the tumour content of cfDNA from sWGS data (Adalsteinsson et al. 2017). I have also discussed the effect of in-silico size selection on the IchorCNA tumour fraction. These methods only add marginally to the cost and turn around time of analysis which means they are methods that can potentially be translated to routine clinical practice is found to increase rates of ctDNA detection.

However, it is not currently clear if any of these methods will be clinically useful for increases rates of ctDNA detection in women with newly diagnosed disease and therefore whether they will be useful in a diagnostic setting. The main reason for this is the inconsistent findings across the two cohorts.

In the UKOPS cohort the effect of in-silico size selection was variable even within the HGSOC cases. In the non-HGSOC cases which are not driven by copy number changes sWGS, even with size selection, is not the optimal method for detection of ctDNA. The size selected t-MAD score is therefore on its own not a useful metric for triage of symptomatic women with suspected ovarian cancer. Even in some of the HGSOC cases size selection resulted in a reduction in t-MAD score. If size selection were applied to all samples, either in-vitro or in-silico, for the detection of ctDNA this may result in some cases that would have been detected becoming non-detected leading to a false negative test. One solution to this is to calculate the t-MAD score and then apply in-silico size selection to those non-detected cases in an attempt to increase the rates of detection.

The IchorCNA method also performed better in the HGSOC cases compared to the non-HGSOC cases in the UKOPS cohort. This is not unexpected as IchorCNA, as with t-MAD,

relies on detecting copy number changes in plasma. IchorCNA predicts segments of SCNA and estimates of tumour fraction, taking into account subclonality and tumour ploidy. The lower limit of detection using this method is a tumour fraction of 0.03 (Adalsteinsson et al. 2017). The t-MAD score is a quantitative measure of global copy number change. The UKOPS cohort the IchorCNA tumour fraction using the customised settings was best able to discriminate HGSOC cases from healthy controls (sensitivity 83%). However, using the customised IchorCNA reduced the highest tumour fraction in the healthy controls from 7% to 1%. There is therefore a possibility that the higher detection of cases is a result of over-fitting to this set of healthy samples.

In a subset of the CTCR-OV04 cohort I compared in-vitro size selection to in-silico size selection. As with relapsed disease in-vitro size selection resulted in a greater increase in t-MAD score compared to in-silico size selection. The liitations of in-vitro size selection were discussed in Section 5.3 (page 99). In the whole CTCR-OV04 cohort the size selected t-MAD score was best able to discriminate HGSOC cases from healthy controls with a sensitivity of 67%. This sensitivity although approaching that found in previous studies of newly diagnosed OC (Bettegowda et al. 2014; Phallen et al. 2017) remains too low to be clinical useful.

Previous studies although having good rates of detection of ctDNA in women with newly diagnosed disease have only been able to detect 40–50% of stage I cancers (Bettegowda et al. 2014; Phallen et al. 2017). Using the size selected t-MAD score I detected 88% of HGSOC cases with stage I/II disease. Selecting for fragments between 20–150 bp detected more early stage cases than when selecting for fragments between 90–150 bp. This suggests that early stage disease could have a higher proportion of very short fragments which may be related to the tumour proliferation (Abbosh et al. 2017). The limitations of this analysis are the very small number of early stage OC cases included in this analysis and these findings will require validation in much larger cohorts. If it is the case that early stage cancers have a greater proportion of shorter cfDNA fragments implementation of size selection for detection of ctDNA as a diagnostic tool will be challenging. Practically at the time of performing this analysis the stage of disease will be unknown therefore knowing what size range to select for in the absence of this information may be challenging.

In this chapter I have also discussed the cfDNA fragmentation features found in newly diagnosed disease. The sensitivity for all features individually to discriminate HGSOC cases from healthy controls was <50%. In relapsed HGSOC the combination of size selected t-MAD, proportion of fragments between 160–180 bp, proportion of fragments between 180–

220 bp, proportion of fragments between 250–320 bp and the 10 bp amplitude enabled the best discrimination between HGSOC cases and controls (in press, Science Translational Medicine). It is therefore possible that the sensitivity for detection of newly diagnosed disease could be improved using a combination of features. It is likely that a combination of assays, potentially including fragmentation features of cfDNA will be required to optimised detection of OC in symptomatic women in primary care.

In contrast to other studies I have also compared HGSOC cases to benign controls (women who underwent primary surgery for suspected ovarian cancer and were subsequently diagnosed with benign disease). Using the size selected t-MAD score I was able to detect 77% of stage I/II disease from a cohort of 32 benign controls. This sensitivity is higher than that seen in other studies of early stage disease compared to healthy controls (Bettegowda et al. 2014; Phallen et al. 2017). An important limitation is the small number of women with early stage disease included in the the newly diagnosed CTCR-OV04 cohort.

These results provide promising insights into methods that may allow a clinically-meaningful sensitivity of ctDNA detection in women been investigated for suspected OC, including early stage disease. These methods now need to be validated in larger cohorts of symptomatic women to investigate the clinical utility for diagnosis of OC compared to CA 125.

Investigating cervical sampling to increase detection of ovarian cancer

7.1 Introduction

Haematological metastatic spread in OC is uncommon. Detection of ctDNA in plasma samples may therefore not be the most sensitive approach for disease detection, particularly in early stage/low volume disease. Direct metastatic spread within the peritoneal cavity is very common in OC due to the direct communication between the distal end of the fallopian tube (site of origin of HGSOC) and the peritoneal cavity. It is therefore possible that cells and DNA shed from the distal fallopian tube can also pass down the fallopian tube, through the uterus to the external cervical os, and be sampled in the upper vagina.

Studies have shown that tumour DNA and somatic mutations can be identified in cervical cytology samples (Kinde et al. 2013) and vaginal tampons (Erickson et al. 2014) from women with advanced OC. More recently it has been shown that 33% of OCs (n=245) were detected using cervical cytology samples (Wang et al. 2018) including detection of 34% of early stage cancers. Combining cervical cytology and plasma increased the sensitivity of detection to 63%. This suggests that cervical sampling could potentially be used as part of a diagnostic or triage tool for OC.

In this chapter I will discuss the optimisation of DNA extraction, library preparation, and sequencing of cervical cytology samples. I will discuss targeted sequencing of DNA obtained from cervical cytology samples collected through the NHS cervical screening programme through a collaboration with the West Anglia Pathology Service Cervical Cytology Laboratory. I will also discuss targeted sequencing and sWGS of DNA obtained from cer-

vical cytology samples collected at Addenbrookes Hospital from women with newly diagnosed HGSOC (Figure 7.1).

In this chapter the TAm-Seq analysis pipeline was performed by James Morris and the t-MAD scores calculated by Dineika Chandrananda.

ThinPrep

* Optimisation of DNA extraction, library preparation and sequencing

* Targeted sequencing of routine cervical cytology samples with non-cervical glandular abnormalities

* Targeted sequencing and sWGS of cervical cytology samples from women with newly diagnosed HGSOC

Figure 7.1: Overview of approach for investigating cervical sampling for use in the detection of OC.

7.2 Results

7.2.1 Investigation of quality of cervical cytology DNA for sequencing

Routine cervical samples are collected with a brush and stored in 20 ml of PreservCyt (methanol based) fixative before cytological evaluation. The cellularity of cervical cytology samples is variable with a median of 0.4 million cells per ml (IQR 0.2 - 0.8 million cells/ml) (Figure 7.2c) based on measurements of volume of fixative remaining after cytological evaluation and cell count (Figure 7.2a, b, d) in 100 samples from the NHS cervical screening programme. DNA yield following extraction was also variable and was not correlated with the total number of cells in the sample (Figure 7.2e). The targeted sequencing coverage appeared comparable to that observed for FFPE samples (Figure 7.3).

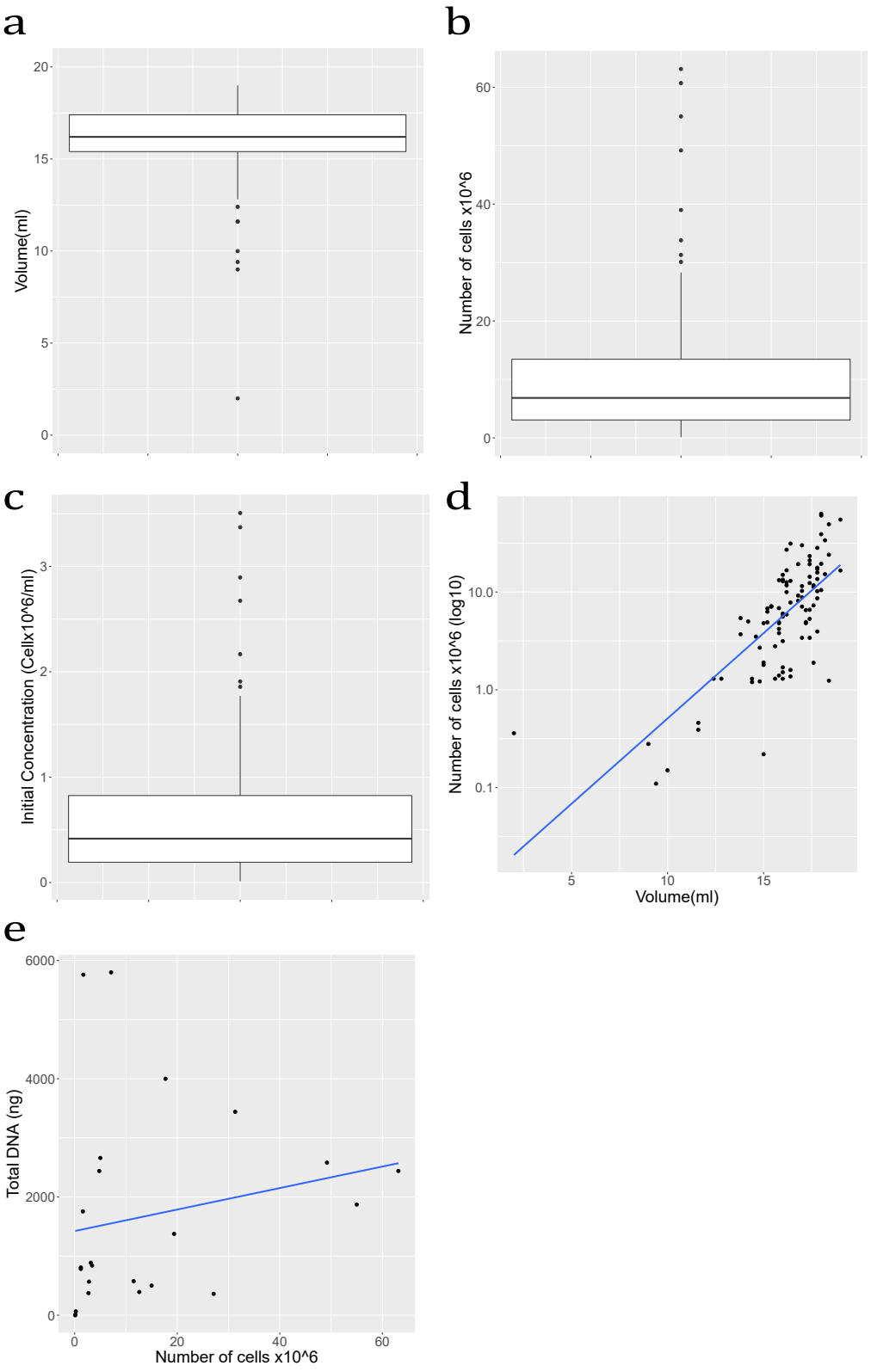


Figure 7.2: **a)** Volume of fixative remaining following cytological evaluation for 100 routine cervical cytology samples. **b)** Total number of cells in remaining volume of fixative. **c)** Number of cells per ml in original sample. **d)** Comparison between remaining volume and total number of cells (correlation coefficient 0.45). **e)** Comparison between number of cells and total DNA extracted in 24 samples (correlation coefficient 0.20).

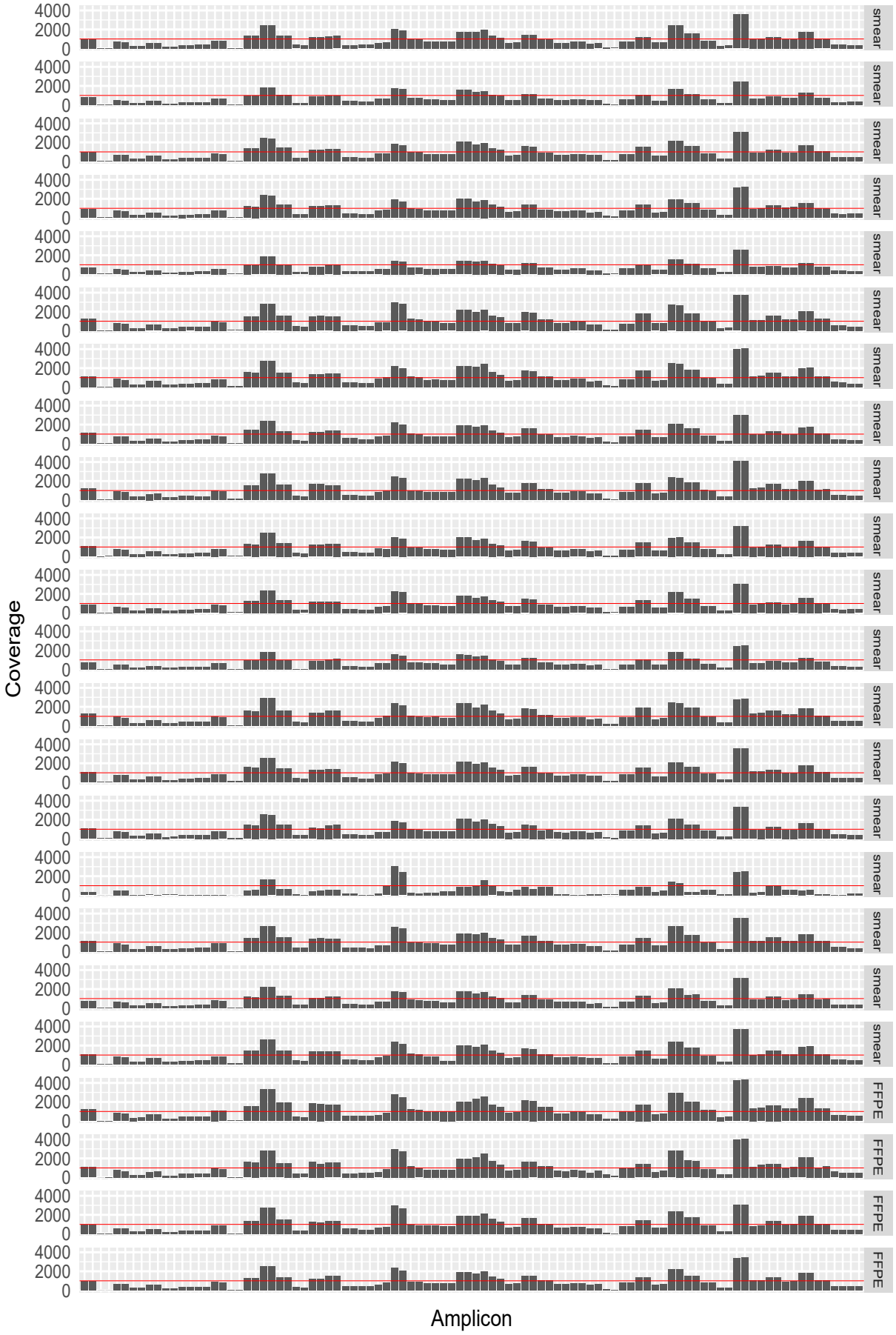


Figure 7.3: Sequencing coverage per amplicon (ordered by panel order from left to right see Appendix 12.3) for 19 cervical cytology samples and four FFPE samples for comparison. The red line indicates a coverage of 1000×.

As cervical cytology samples are stored in fixative it is possible that this may have an impact of the quantity and quality of DNA obtained from samples. By introducing cultured cells into PreservCyt and extracting DNA at different time points I found a large variation in total DNA yield that was not affected by fixation time (Figure 7.4a). The DNA fragmentation pattern (Figure 7.6) and the sequencing coverage also did not appear to vary with fixation time (Figure 7.5). Hela cells (no *TP53* mutation) and OVCAR3 cells (substitution mutation in *TP53* c.G743A, p.R248Q) were mixed in an equal ratio. As such the expected MAF was 50%. The calculated AF by de novo mutation calling remained approximately 50% for all replicates across all time points (Figure 7.4b). This was reassuring that time in fixation did not appear to impact the quantity and quality of DNA.

I then wanted to investigate the sensitivity of mutation detection in cervical cytology samples. Cultured cells (OVCAR3 and SKOV3 (deletion in *TP53* c.C267del, p.P89fs)) were mixed in different proportions with Hela cells in PreservCyt. DNA was extracted after 3 weeks and TAm-Seq using primer panel 10 performed (Appendix 12.4). The observed MAF was similar to the expected MAF with a lower limit of detection of 0.8% for the deletion and 0.4% for the substitution mutation (Figure 7.4c). This is consistent with the lower limit of detection using this method when applied to plasma samples (see section 4.2).

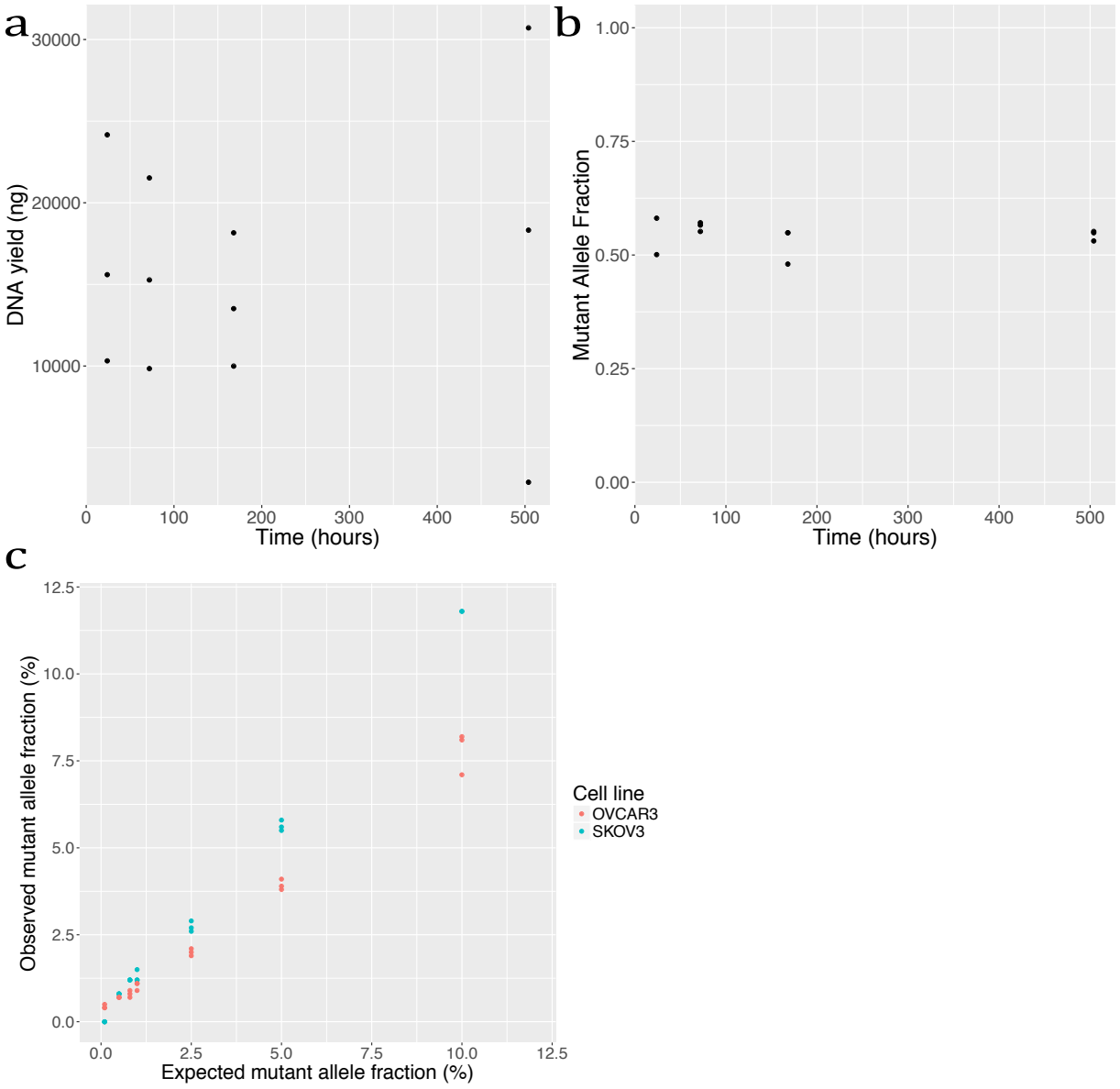


Figure 7.4: **a)** DNA concentration following extraction of cell lines in fixative at 24 hours, 72 hours, 1 week and 3 weeks. **b)** MAF of cervical smear samples by TAm-Seq. Expected MAF was 50% as cells mixed in a ratio of 1:1 OVCAR3:Hela. **c)** Expected MAF based on proportions of cell lines combined and observed MAF measured by standard TAm-Seq.

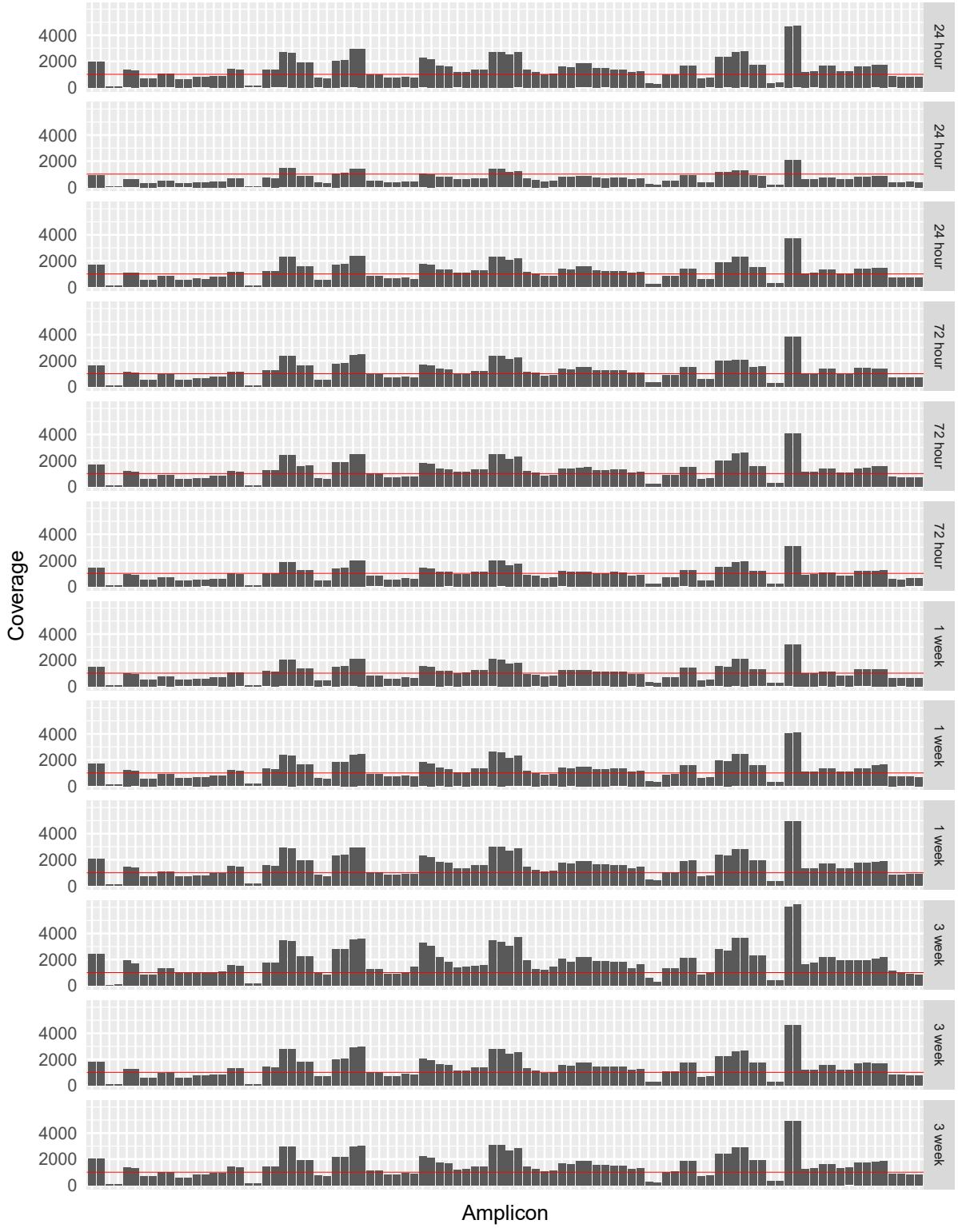


Figure 7.5: Sequencing coverage per amplicon using primer panel 1 (ordered by panel order from left to right see Appendix 12.3) for three replicate smear samples and DNA extraction at 24 hours, 72 hours, 1 week and 3 weeks. The red line indicates a coverage of 1000×.

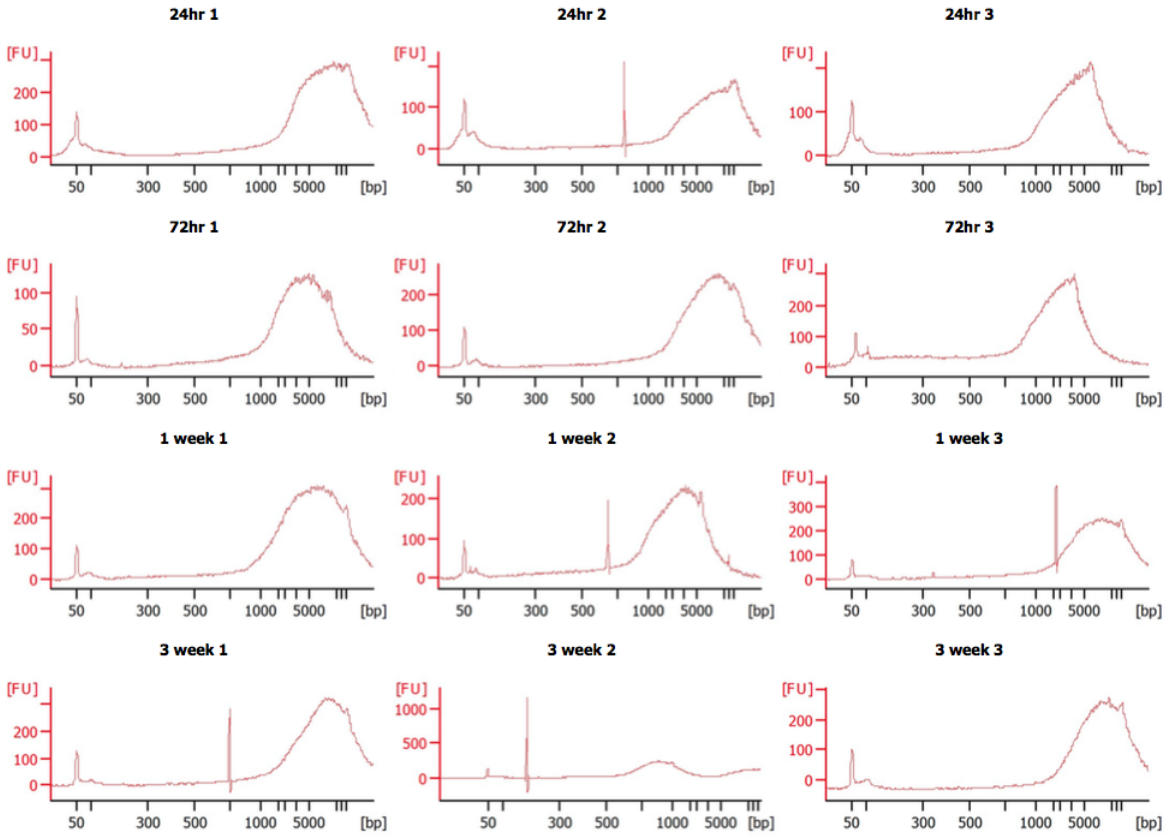


Figure 7.6: Bioanalyzer traces showing fragmentation patterns of cell lines in fixative with DNA extracted at different time points in triplicate.

7.2.2 Application of targeted sequencing to routinely collected cervical cytology samples

Targeted sequencing using TAm-Seq primer panel 1 (Appendix 12.3) was applied to DNA extracted from 71 cervical cytology samples collected through the NHS cervical screening programme (Appendix 12.16). Code 6 samples are samples classified as "?glandular neoplasia of endocervical type" which would have resulted in further investigation for a possible cervical cancer. Code 0 samples are samples classified as "glandular neoplasia (non-cervical)" which would have been investigated for possible endometrial cancer and potentially for OC.

Mutations were detected in 12/28 (43%) of the code 6 samples subsequently diagnosed with a cancer and 4/5 (80%) of the code 0 samples subsequently diagnosed with a cancer (Appendix 12.16). Mutations were also detected in 3 of the code 6 samples subsequently diagnosed with a pre-invasive cervical lesion.

Using these samples I have shown that somatic mutations can be detected in cervical cytology samples collected from women with cervical and endometrial cancers. A limitation of this study is the sparse clinical data. However, this was necessary in order to have ethical approval for access to these samples. Mutations were detected in some samples with no clinical information, these could have a cancer diagnosis, and in some samples that were annotated as having normal large loop excision of the transformation zone (LLETZ) or pipelle samples, this does not preclude a subsequent cancer diagnosis. Unfortunately none of the cervical cytology samples available through the NHS cervical screening programme had been collected from women diagnosed with OC.

7.2.3 Cervical samples in high-grade serous ovarian cancer

I therefore looked at cervical cytology samples collected from 12 women with newly diagnosed HGSOC. De novo calling of targeted sequencing using TAm-Seq primer panel 10 (Appendix 12.4) did not identify a *TP53* mutation in any of the 12 samples. When performing specific variant calling for the known *TP53* mutation identified in the matched FFPE sample the rate of detection increased to 6/12 (50%) (Appendix 12.17).

As sWGS and the t-MAD score had been useful in detecting ctDNA in plasma samples I therefore performed sWGS using the Rubicon ThruPlex DNA-Seq kit (50 ng DNA input, 7

cycles of PCR) for the 12 cervical cytology samples. For comparison I performed the same library preparation and sequencing for cervical cytology samples obtained from three women subsequently diagnosed with benign disease. There was no difference in t-MAD score between the HGSOC cases and the benign controls (Figures 7.7a, c).

Cervical cytology samples were developed to detect pre-malignant changes in ectocervical cells that could progress to cervical cancer if not treated. This sampling technique is not optimised to collect cells shed from the distal fallopian tube. This could explain the low sensitivity for detection of tumour DNA using de novo calling of targeted sequencing and sWGS in the 12 cervical cytology samples from HGSOC cases. It is likely that collecting samples from a more focused location at the cervical os or from within the cervical canal may increase the collection of relevant cells that have been shed from the distal fallopian tube. This would also reduce the number of cells collected from the ectocervix or vagina which will contribute to the background DNA.

I therefore performed targeted sequencing of DNA from cervical mucous aspirate (CMA) samples obtained from 8 of the women with newly diagnosed HGSOC. A *TP53* mutation was detected in 2/8 (25%) of the CMA samples, however only 1 of these matched the mutation identified in the matched FFPE sample. The rate of detection increased to 3/8 (38%) when performing specific variant calling for the know *TP53* mutation identified in the matched FFPE sample (Appendix 12.17).

The t-MAD score was higher in the two cases with a detected *TP53* mutation by de novo calling. There was however no difference in t-MAD score for CMA samples from the HGSOC cases without a *TP53* mutation detected and the benign controls (Figures 7.7a, d). The t-MAD score was lower in the CMA sample compared to the match cervical cytology samples other than in the two cases with a detected *TP53* mutation in the CMA sample (Figures 7.7b).

It is possible that the limited sensitivity of standard TAm-Seq and the t-MAD score without size selection may explain the undetected tumour DNA in these samples. As the total DNA yield from cervical samples is 100–1,000 times greater than plasma samples I investigated the TAm-Seq V2 method (384 wells per sample) for the detection of mutations in cervical cytology and CMA samples. The tumour specific *TP53* mutation was identified in only two of the HGSOC cervical cytology samples (Appendix 12.17). In one sample the tumour specific *TP53* mutation was detected in 2 wells (there were 7 other mutations also called in 2 wells) and in the other sample the tumour specific *TP53* mutation was detected in 4 wells (there was 1 other mutation called in 4 wells, 2 called in 3 wells and 8 called in 2

wells). The tumour specific *TP53* mutation was not called in any of the 6 CMA samples.

Unexpected *TP53* mutations were called in between 2–10 wells in a number of the cervical cytology and CMA samples collected from the HGSOC cases as well as in samples collected from women with benign disease (Appendix 12.17). Figure 7.8 shows the unexpected mutations with the trinucleotide context.

It therefore appears that cervical cytology samples and CMA samples are not optimal for the detection of tumour cells from the ovary/fallopian tube and currently are not sufficient for use as a diagnostic tool in OC.

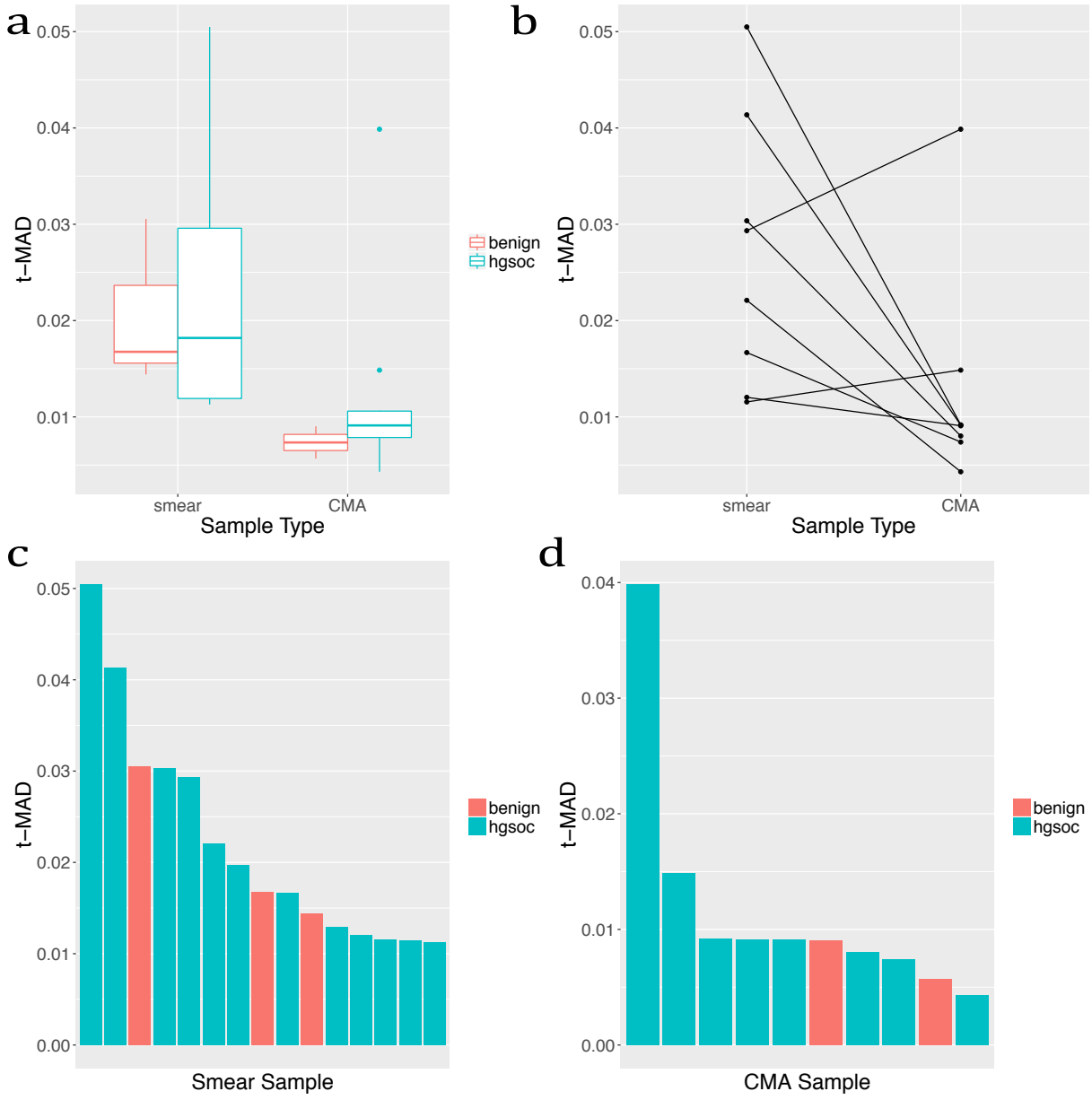


Figure 7.7: **a)** t-MAD calculated from sWGS data from CMA samples and cervical smear samples collected from women with HGSOC and benign gynaecological conditions. **b)** t-MAD score for matched cervical smear and CMA samples from women with HGSOC. **c)** Cervical smear samples by descending t-MAD score. **d)** CMA samples by descending t-MAD score.

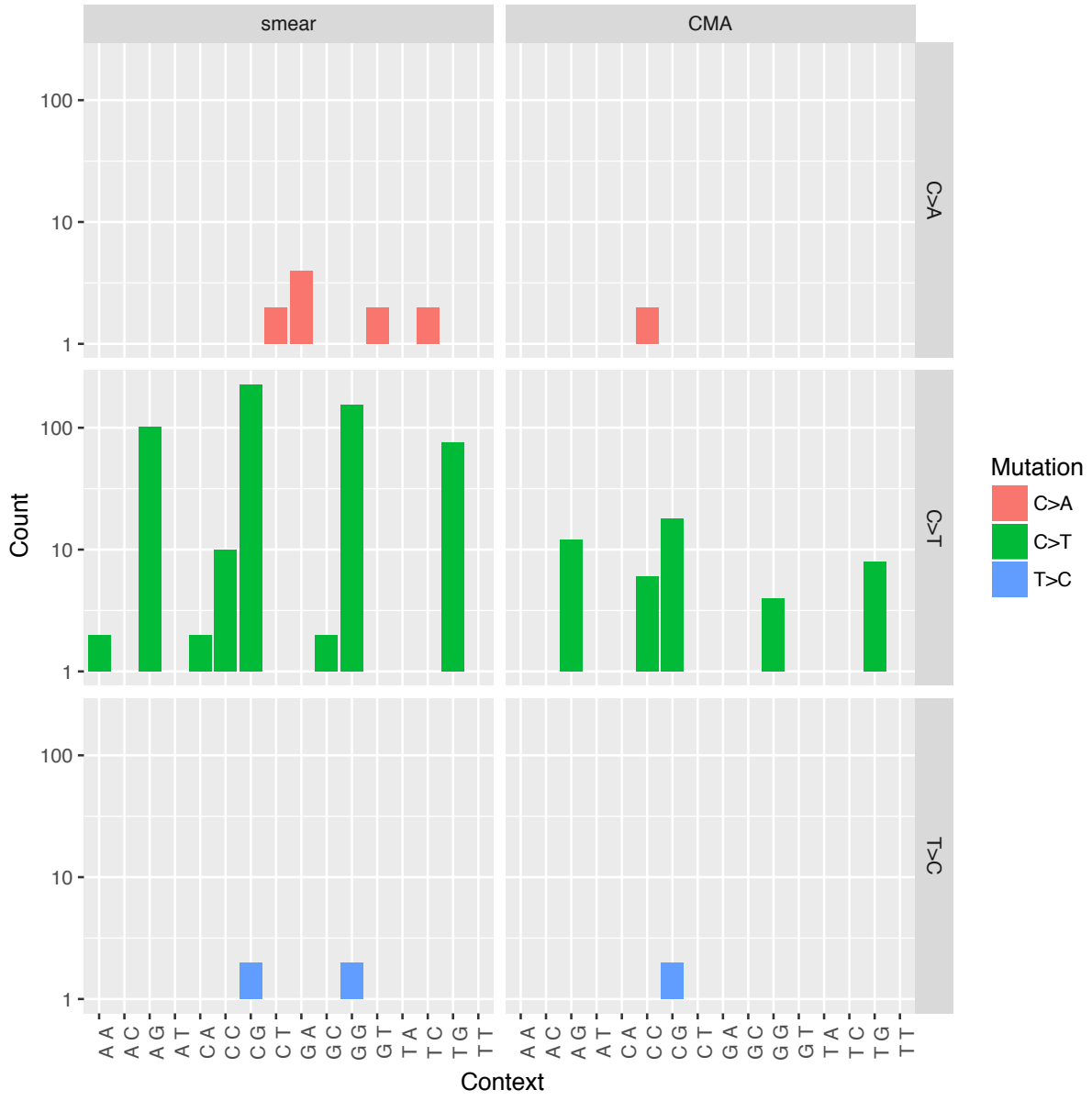


Figure 7.8: All mutations (excluding true positive mutations) called in cervical smear and CMA samples in 2 wells or more with TAm-Seq V2 with trinucleotide context. As sequencing was paired-end the strand that the mutation is on cannot be confirmed. Therefore for ease mutations were merged eg. all reference bases C and G merged and classified as C and all reference bases A and T merged and classified as T.

7.3 Discussion

In this chapter I have shown that DNA can be extracted from cervical cytology samples and sequenced using targeted sequencing and sWGS methods. I have shown that mutations can be detected using targeted sequencing in cervical cytology samples collected through the NHS cervical screening programme from women diagnosed with cervical and endometrial cancer. However, further conclusions cannot be drawn from this study due to the small number of included samples and the lack of clinical data. The targeted sequencing panel and the use of sWGS had been optimised to detect HGSOC which may account for the low rates of detection in the available samples.

Previous studies have shown that somatic mutations can be detected in cervical cytology samples collected from women with OC (Kinde et al. 2013; Wang et al. 2018). Unfortunately none of the cervical cytology samples available through the NHS cervical screening programme had been collected from women diagnosed with OC. In the future a large scale study including all routinely collected cervical cytology samples classified as showing non-cervical glandular abnormalities from across the country could enable more meaningful conclusions to be drawn about whether cervical cytology samples are useful for the detection of OC.

Through CTCR-OV05 I collected cervical cytology samples from women with suspected or confirmed OC. Although 94 cervical cytology samples were collected in total only 12 of these were collected before treatment for HGSOC (others were collected before interval debulking surgery (IDS) or before primary surgery from women not diagnosed with HGSOC). De novo calling of targeted sequencing and sWGS did not detect tumour DNA in any of the 12 samples collected before treatment from women with HGSOC. The minimum MAF detected using targeted sequencing in plasma samples was 0.6% resulting in a low rate of ctDNA detection (Section 4.2). A similarly low rate of detection for ctDNA was seen in plasma samples using sWGS. It is therefore possible that targeted sequencing and sWGS are not sensitive enough for the detection of tumour DNA in cervical cytology samples. In comparison the methods used by Kinde et al. 2013 and Wang et al. 2018 were able to detect a minimum MAF of 0.01% and 0.03% respectively. The sensitivity of the methods used is a limitation of this study particularly since tumour DNA was detected in 50% of the cervical cytology samples when performing specific variant calling for the *TP53* mutation identified in FFPE, therefore suggesting it is present but not detected using de novo calling.

A factor effecting sensitivity is the amount of non-tumour DNA in a sample. Cervical cytology samples were designed to detect pre-malignant changes in ectocervical cells. The majority of cells in cervical cytology samples are therefore from the ectocervix, endocervix and vagina. Samples are also likely to contain inflammatory cell, red blood cells and micro-organisms. These will all therefore contribute DNA to the sample causing a reduction in the signal to noise ratio leading to non-detection of tumour DNA in a sample. As such this method is not optimal for collecting cells/DNA shed from the distal fallopian tube.

Wang et al. 2018 increased the sensitivity of detection from 33% using cervical cytology to 45% by performing direct sampling of the intrauterine cavity using a Tao brush. A procedure such as the Tao brush will not be useful as a triage tool in primary care as it is an invasive procedure that can only be performed by a trained healthcare professional. I therefore wanted to investigate other collection methods that may increase the rate of tumour DNA detection but can potentially in the future be performed in primary care.

A previous study investigated the presence of *TP53* mutations in cervical mucous samples obtained from hysterectomy samples (Lamzabi et al. 2013). Mutations were identified in 2/19 malignant cases. Mutations were also identified in two cases with STIC lesions. CMA samples can be collected at the time of speculum examination and as such could be collected in primary care. I therefore investigated CMA samples collected at the time of primary surgery for HGSOC. *TP53* mutations were detected in two CMA samples collected before treatment from women with HGSOC. This suggests that detection rates can potentially be increased by more focused sampling from the cervical os. The limitation of this collection method is that not all women, particularly post-menopausal women, produce a sufficient quantity of cervical mucous to enable a sample to be collected. although 98 cervical cytology samples were collected through CTCR-OV05 only 62 CMA samples could be collected.

Another way of improving the sensitivity of detection is to apply alternative methods for the detection of ctDNA. Size selection is not applicable in this setting as it is cellular not cell-free DNA that is being investigated. TAm-Seq V2 was therefore applied to the cervical cytology and CMA samples.

Using TAm-Seq V2 the tumour specific *TP53* mutation was identified in two of the HGSOC cervical cytology samples. A large number of unexpected mutations were detected in up to 10 wells in samples from HGSOC cases and benign controls. The majority of the unexpected mutations called in both the cervical smear samples and CMA samples were C>T mutations with a CpG context. CpG sites have a higher mutation rate than other sites, as the C in this context is usually methylated and therefore hypermutable resulting in spon-

taneous deamination of methyl-C to T. This finding is therefore not unexpected and it is possible that in the future the analysis of TAm-Seq V2 calling could incorporate a weight dependent on the type of mutation to aid in the determination of true positive and false positive calls. However, currently the TAm-Seq V2 method results are too 'noisy' with too many false positive mutations called to be a useful analysis method.

It is difficult to comment on the number of unexpected mutations in the smear samples compared to the CMA samples as there were more smear samples included in the analysis. Cervical smear samples are collected and stored in a methanol based fixative. Although methanol based fixatives have been shown to have lower deleterious effects on DNA quantity and quality compared to formalin based fixation for tissue (Piskorz et al. 2016) this has not been studied in cervical samples. It is therefore possible that some of the unexpected mutations seen in the cervical smear samples are a result of methanol fixation inducing chemical modifications and degrading the DNA. Ideally a future cervical sampling device for use in the diagnosis of OC would be collected and stored fresh as is with the current trend for tissue collection for molecular analysis.

In the future if cervical sampling is to be further explored to aid in the triage of symptomatic women from primary care either a more sensitive method for detection of tumour DNA will be required or an alternative sampling method more focused on collecting cells and DNA shed from the distal fallopian tube.

Circulating tumour DNA as a prognostic and predictive biomarker in ovarian cancer

8.1 Introduction

Standard first line therapy for HGSOC is a combination of surgery and chemotherapy (usually carboplatin and paclitaxel). 30% of women do not respond to first line therapy and have a poor outcome. The development of blood based biomarkers to measure and predict response to treatment would be particularly beneficially for identifying those women who are not responding to standard of care treatment enabling early entry into clinical trials that may improve long term outcome.

It has previously been shown that a decrease of $> 60\%$ in levels of ctDNA measured by digital PCR for *TP53* after one cycle of chemotherapy for treatment of relapsed HGSOC is a significant predictor of 6-month time to progression (TTP). CA 125 decrease in the same cohort was not a significant predictor (Parkinson et al. 2016). It is not yet known whether the same is true in newly diagnosed HGSOC treated with neo-adjuvant chemotherapy.

In this chapter I will compare ctDNA levels at baseline, measured by targeted sequencing and sWGS with end of treatment response. I will also compare the change in ctDNA levels after one and two cycles of chemotherapy.

In this chapter radiological assessment and RECIST 1.1 measurements were performed by Helen Addley, the TAm-Seq analysis pipeline was run by James Morris and the t-MAD scores and fragmentation features calculated by Dineika Chandrananda.

8.2 Results

8.2.1 CTCR-OV04 cohort

End of treatment response measured by RECIST 1.1 was determined for 138 of the newly diagnosed CTCR-OV04 cohort (Table 8.1) and compared to ctDNA levels at baseline. 22% of the cohort did not respond to standard first line therapy. More than 50% of the responders had a detectable *TP53* mutation at baseline however, only 39% of the cases with stable disease (SD) and 33% of the cases with progressive disease (PD) had a detectable *TP53* mutation at baseline (Figure 8.1a). There was no difference in *TP53* MAF, unselected t-MAD, or size selected t-MAD score at baseline in clinical responders compared to non-responders (Figure 8.1b-e). There was also no difference in ctDNA fragmentation features at baseline in clinical responders compared to non-responders (Figure 8.2).

Clinical Response	Percentage of Cohort
Complete Response	47
Partial Response	31
Stable Disease	13
Progressive Disease	9

Table 8.1: End of treatment response measured by RECIST 1.1 for CTCR-OV04 cohort

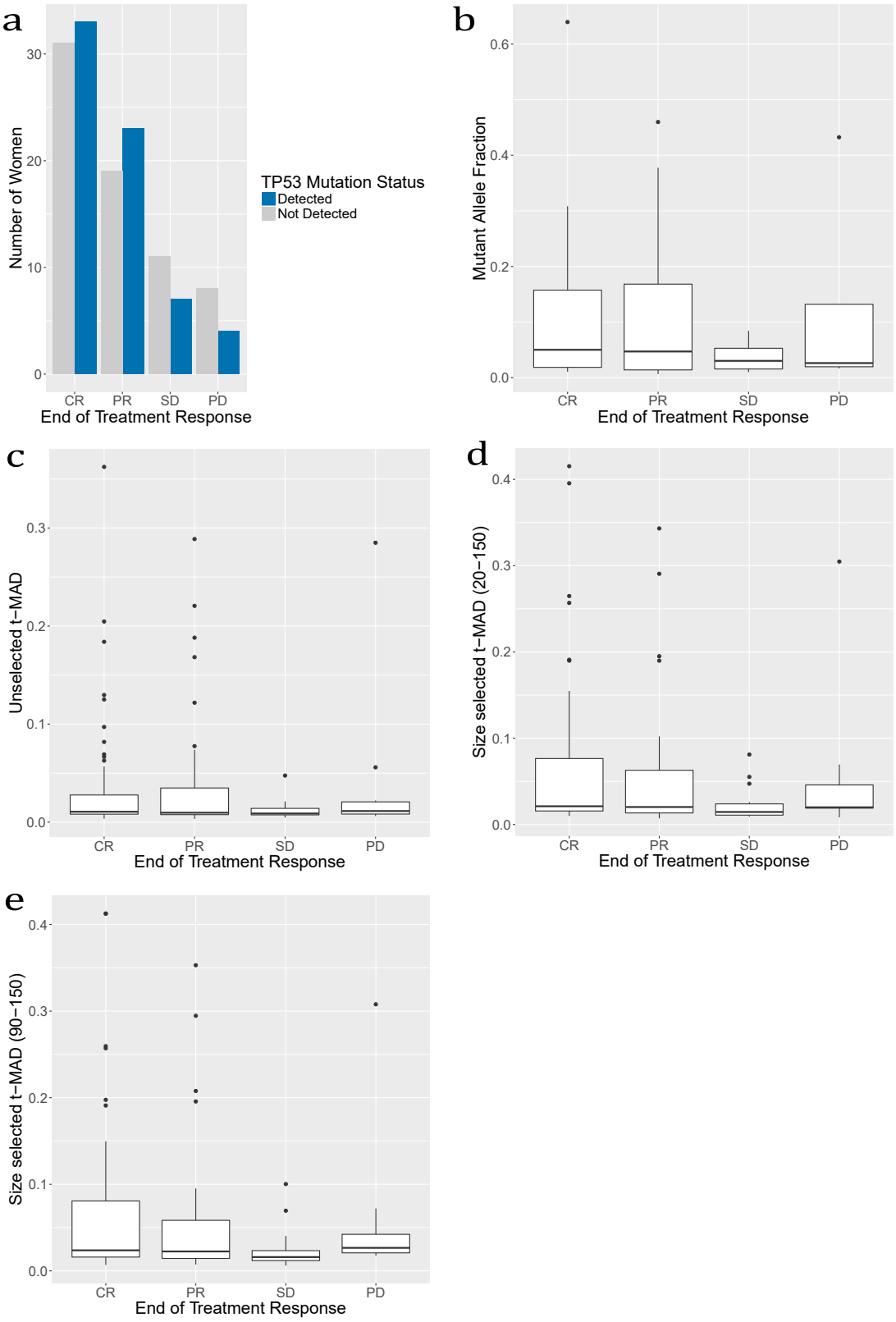


Figure 8.1: **a)** Number of women with a detectable *TP53* mutation at baseline by end of treatment response determined by RECIST 1.1. **b)** *TP53* MAF for detected cases at baseline by end of treatment response determined by RECIST 1.1. **c)** Unselected t-MAD score at baseline by end of treatment response determined by RECIST 1.1. **d)** Size selected t-MAD score (20–150 bp) at baseline by end of treatment response determined by RECIST 1.1. **e)** Size selected t-MAD score (90–150 bp) at baseline by end of treatment response determined by RECIST 1.1.

Clinical group	% with >60% reduction in MAF C1-C2	% with >60% reduction in MAF C1-C3
Responders	88	93
Non-responders	64	80

Table 8.2: Percentage of responders and non-responders with >60% reduction in MAF after 1 (C1-C2) and 2 (C1-C3) cycles of chemotherapy.

The *TP53* MAF dropped after 1 cycle of chemotherapy for the majority of cases (Figure 8.3a). There was an inverse relationship between the median decrease in *TP53* MAF after one cycle of chemotherapy and stage of disease however after two cycles of chemotherapy the median decrease in *TP53* MAF was similar for all stages of disease (Figure 8.3b). The median decrease in *TP53* MAF was higher for clinical responders compared to non-responders after both one and two cycles of chemotherapy (Figure 8.3c) however, there was one case with a partial response (PR) that had an increase in *TP53* MAF after one cycle of chemotherapy. The only cases with an increase in *TP53* MAF after two cycles of chemotherapy had SD. None of the cases with PD had an increase in *TP53* MAF after two or three cycles of chemotherapy. A 60% reduction in *TP53* MAF was not predictive of end of treatment response in a neo-adjuvant setting particularly after two cycles of chemotherapy (Table 8.2).

The change in t-MAD score (unselected and size selected) after one and two cycles of chemotherapy did not correlate with stage of disease (Figure 8.4a, c, e). Neither was it predictive as an increase in t-MAD score after one and two cycles of chemotherapy was seen in both responders and non-responders (Figure 8.4b, d, f).

cfDNA fragmentation features provide information about the mechanism of release of cfDNA. I was therefore interested in whether the DNA fragmentation features changed after one and two cycle of chemotherapy. The proportion of the different fragmentation features did not vary between baseline, pre cycle 2 and pre cycle 3 (Figure 8.3).

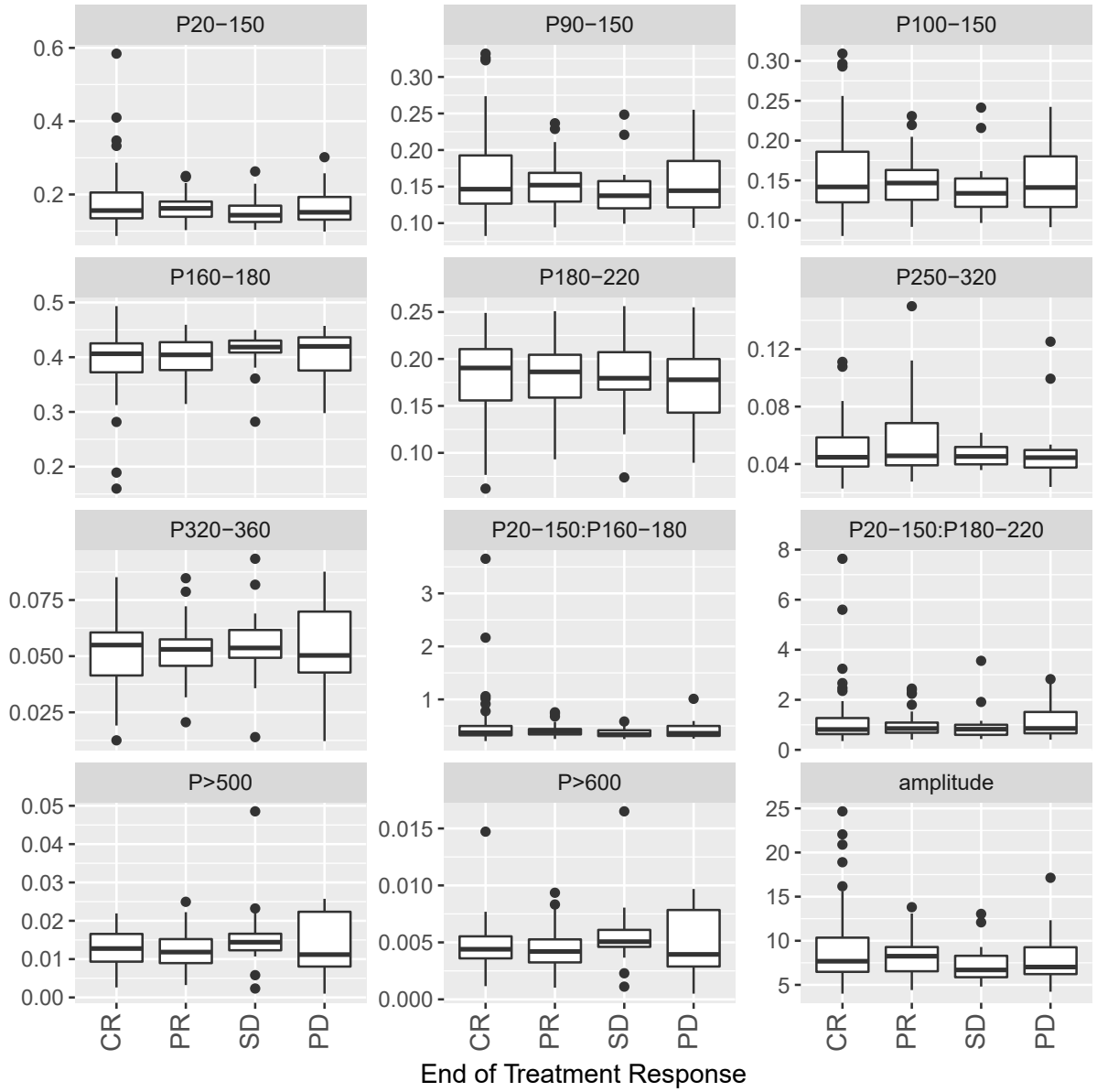


Figure 8.2: The proportion (P) of DNA fragments observed by sWGS in multiple size ranges (P20–150, P90–150, P100–150, P160–180, P180–220, P250–320, P320–360), the ratios of proportions of fragments in these size ranges (P20–150:P160–180, P20–150:P180–220), the proportion of fragments greater than a certain size (P>500, P>600) and the amplitude of oscillations in fragment size density with 10 bp periodicity observed below 150 bp at baseline by end of treatment response determined by RECIST 1.1.

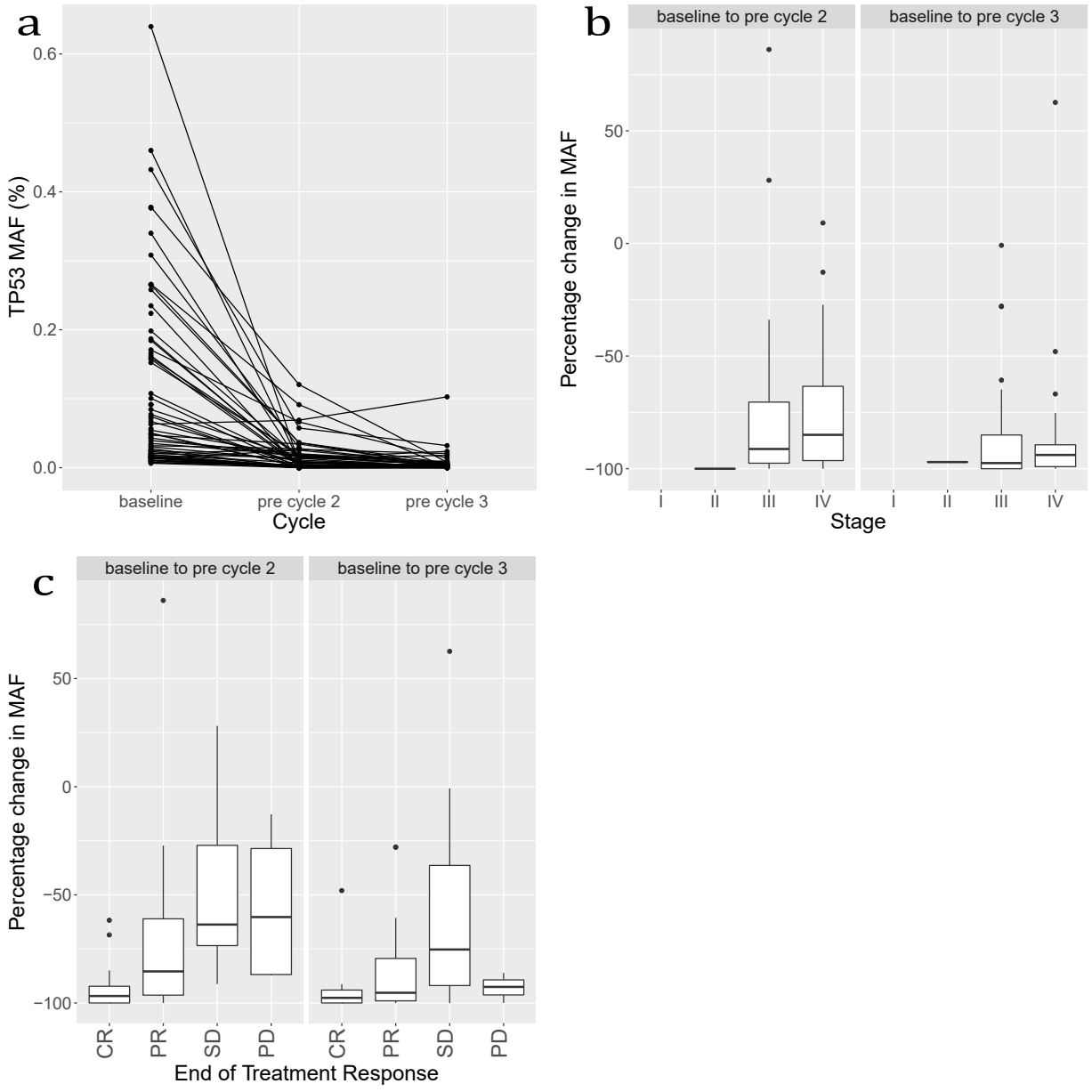


Figure 8.3: **a)** *TP53* MAF at baseline, pre cycle 2 and pre cycle 3 of neo-adjuvant chemotherapy. **b)** Percentage change in MAF between baseline and pre cycle 2 and baseline and pre cycle 3 by stage of disease. **c)** Percentage change in MAF between baseline and pre cycle 2 and baseline and pre cycle 3 by end of treatment clinical response.

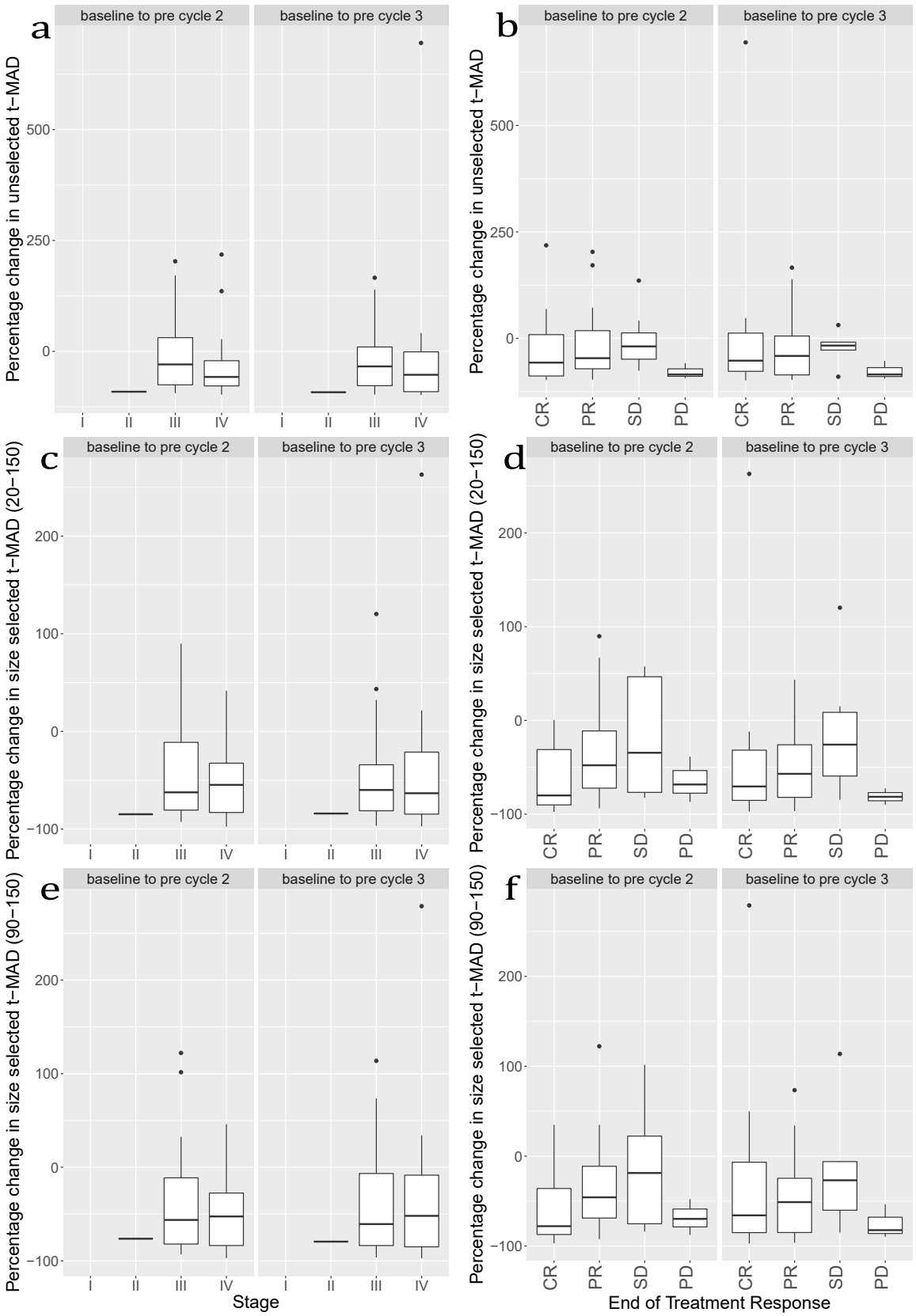


Figure 8.4: Percentage change in t-MAD between baseline and pre cycle 2 and baseline and pre cycle 3 for: **a)** unselected t-MAD by stage, **b)** unselected t-MAD by end of treatment response, **c)** size selected t-MAD (20–150) by stage, **d)** size selected t-MAD (20–150) by end of treatment response, **e)** size selected t-MAD (90–150) by stage, **f)** size selected t-MAD (90–150) by end of treatment response.

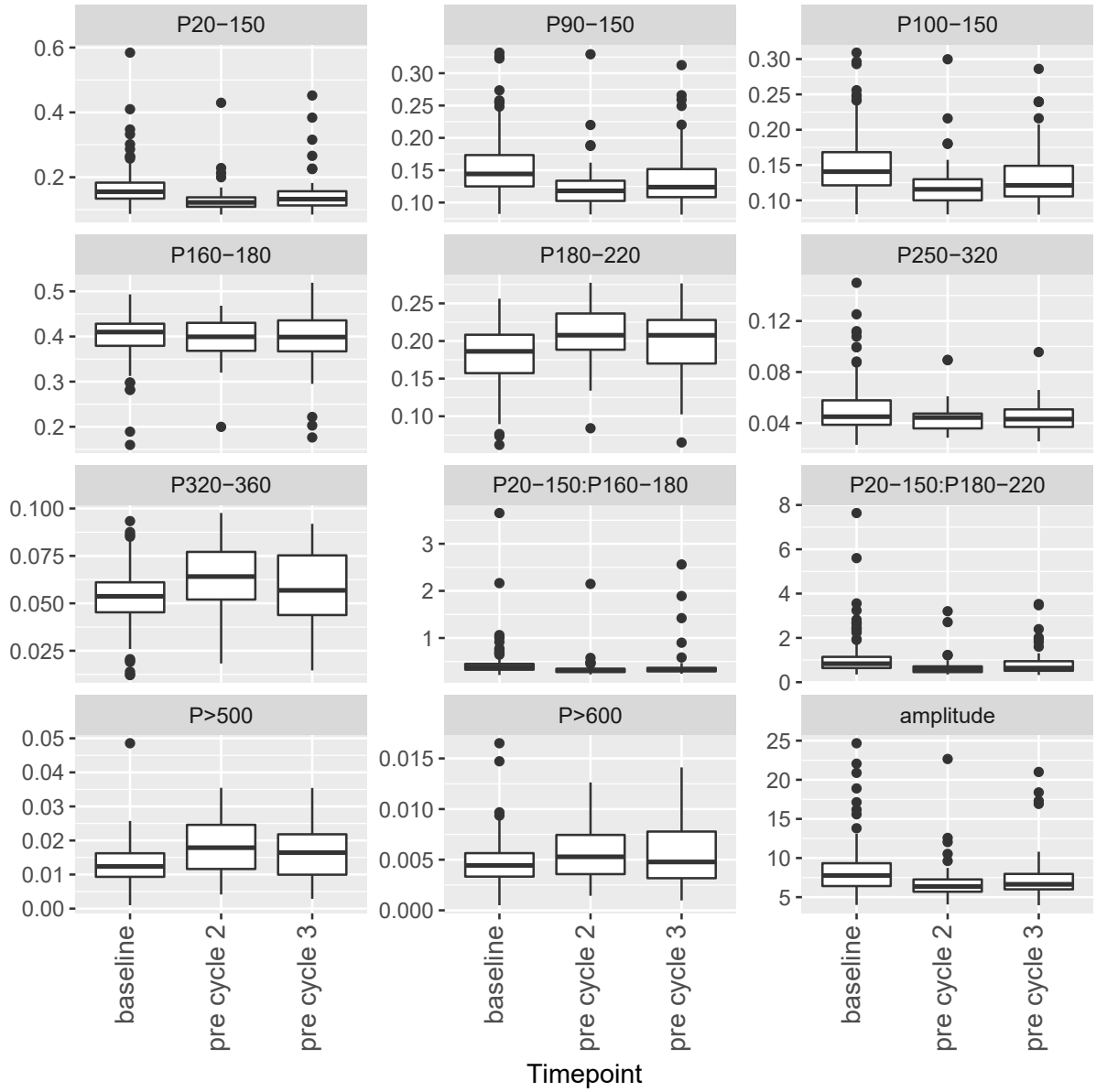


Figure 8.5: The proportion (P) of DNA fragments observed by sWGS in multiple size ranges (P20–150, P90–150, P100–150, P160–180, P180–220, P250–320, P320–360), the ratios of proportions of fragments in these size ranges (P20–150:P160–180, P20–150:P180–220), the proportion of fragments greater than a certain size (P>500, P>600) and the amplitude of oscillations in fragment size density with 10 bp periodicity observed below 150 bp at baseline at baseline, pre cycle 2 and pre cycle 3.

8.3 Discussion

It has previously been shown that ctDNA can outperform CA 153 in metastatic breast cancer (Dawson et al. 2013) and CA 125 in recurrent OC (Parkinson et al. 2016) as a prognostic and predictive biomarker. In recurrent HGSOC baseline *TP53* MAF was a significant predictor of TTP with women with a higher baseline *TP53* MAF having a shorter time to progression. I wanted to investigate whether ctDNA was a prognostic or predictive marker in newly diagnosed HGSOC. There have been no previous studies investigating the change in ctDNA with first line treatment for HGSOC and as the natural progression of the disease is not fully understood it cannot be assumed that newly diagnosed disease responds in the same way as recurrent disease. In fact anecdotal evidence suggests that recurrent HGSOC has a different appearance at surgery to newly diagnosed disease.

In women with newly diagnosed HGSOC I found no relationship between the baseline *TP53* MAF or t-MAD score and end of treatment response suggesting that baseline ctDNA is not prognostic in newly diagnosed disease. This conclusion may be effect by the low rate of ctDNA detection at baseline within the CTCR-OV04 cohort. Interestingly I did find that a higher proportion of responders (53%) had a detectable *TP53* mutation at baseline compared to non-responders (37%). It is possible that if ctDNA release is related to tumour proliferation then highly proliferative tumours that have a better response to chemotherapy have ctDNA detectable at baseline. However, this hypothesis is yet to be investigated as currently there are no reliable measure of tumour proliferation across undiagnosed patients.

In relapsed HGSOC a reduction of >60% in *TP53* MAF was predictive of TTP. This was not the case in newly diagnosed disease. One limitation of this analysis is that 50% of the cohort did not have detectable ctDNA at baseline so this analysis was limited to 55 cases and only 10 of these were non-responders. This analysis was undertaken using specific variant calling for the tumour specific *TP53* mutation identified in the baseline plasma sample and therefore not limited to the lower limit of detection of a MAF of 0.6% that was the case with de novo mutation calling.

After 1 cycle of chemotherapy 88% of the responders and 64% of the non-responders had >60% reduction in *TP53* MAF. After two cycles of chemotherapy this had increased to 93% of the responders and 80% of the non-responders. It is possible in newly diagnosed disease an initial response is seen even in the non-responders as the chemotherapy kills

the sensitive cells. It may be that stable or progressive disease is a result of overgrowth of resistant clones and that this will not be detected after just two cycles of chemotherapy. If this is the case it would have been useful to look at the *TP53* MAF in an end of treatment plasma sample as it is possible it would have increased by this stage mirroring the non-response. It would have also be interesting to compare the baseline ctDNA levels and the change in ctDNA levels to time to progression in the clinical responders as it is possible that this may be predictive. Unfortunately this clinical data was not available for this cohort.

The fragmentation features of cfDNA reflect the mechanism of compaction and release into the circulation. The change in fragmentation features after the start of treatment could therefore be used to distinguish ctDNA from cells killed by treatment compared to those resistant to treatment thereby elucidating the mechanistic effect of treatment on tumour cells. However, I did not identify any change in fragmentation features after one or two cycles of neo-adjuvant chemotherapy. It is however possible that again this is too short a time frame to identify sensitive and resistant cells and it would again be interesting to see if any changes were seen at the end of first line treatment (following the completion of 6 cycles of chemotherapy). This may also be a result of the generally low levels of tumour-derived cfDNA in the samples meaning that the fragmentation features reflect the non-tumour derived cfDNA fragments which are not expected to change with chemotherapy.

Although this analysis was performed on a relatively small cohort it appears that, in contrast to recurrent HGSOC, ctDNA is neither predictive or prognostic in newly diagnosed HGSOC.

Further understanding of circulating tumour DNA and application for future clinical trials

9.1 Introduction

In the future clinical trials investigating therapies for the treatment of HGSOC are likely to involve the collection of blood for the analysis of ctDNA. Further understanding of how ctDNA behaves during standard treatment is therefore necessary to investigate the effects of new treatments. In this chapter I will therefore discuss the change in ctDNA around the time of ascites drainage and surgery to provide further insights into the dynamics of ctDNA.

Trials are likely to be large multi-centre trials. Blood processing at a central location will reduce biases seen in equipment and processing technique. Cell stabilisation tubes (such as STRECK BCT) provide stable ctDNA for up to 72 hours. In this chapter I will discuss the effect of postage of these tubes on sWGS analysis. This work is part of a larger body of work investigating blood processing and cell stabilisation tubes with a manuscript accepted for publication in the Journal of Molecular Diagnostics (Appendix 12.21).

Finally investigation of ctDNA changes longitudinally is currently limited as it is only recently that studies have started to collect plasma and not serum. For example UKCTOCS collected serum for investigation of protein based biomarkers. Serum is not as useful as plasma for ctDNA analysis due to the high amount of background DNA resulting from white blood cell (WBC) lysis. I was therefore interested in whether ctDNA levels could be en-

riched in serum samples using a similar size selection method to that discussed in plasma. In this chapter I will discuss the investigation of the size of cfDNA fragments in serum compared to plasma.

In this chapter the TAm-Seq analysis pipeline was run by James Morris. The t-MAD score and fragmentation features were calculated by Dineika Chandrananda.

9.2 Results

9.2.1 Dynamics of circulating tumour DNA

It has previously been shown that drainage of ascites can result in a decrease in ctDNA measured by *TP53* MAF (Parkinson et al. 2016). However, in this example ctDNA levels were quantified at baseline (day 0) (MAF=7.5%), 8 litres of ascites was drained on day 4 and then ctDNA levels were quantified on day 29 (during which time the patient did not receive any treatment) at which point the TP53 MAF had dropped to 3.3%.

To investigate how quickly ctDNA levels change after ascites drainage I collected plasma samples before ascites drainage and immediately after the drain was removed for 7 patients (Table 9.1). The volume of ascites drained ranged from 2 to 5.5 litres. I quantified ctDNA levels using MAF calculated from targeted sequencing (TAm-Seq with primer panel 10 - appendix 12.4) and t-MAD score calculated from sWGS data. The change in *TP53* MAF and t-MAD score varied considerably between patients and the change in the two measurements was not consistent within patients (Table 9.1).

Patient	Volume ascites drained (L)	MAF pre drainage	MAF post drainage	Percentage change in MAF	t-MAD pre drainage	t-MAD post drainage	Percentage change in t-MAD
788	3	0.020	0.015	−25.00	0.013	0.019	44.53
626	2	0.069	0.069	0.00	0.034	0.032	−6.67
800	2.8	0.059	0.063	6.78	0.029	0.041	42.50
277	5.5	0.008	0.026	225.00	0.012	0.013	1.77
619	2	ND	ND	NA	0.007	0.007	1.89
864	4.1	ND	ND	NA	0.007	0.005	−27.20
156	3.5	0.121	0.118	−2.49	0.006	0.006	−0.42

Table 9.1: 7 patients with plasma collection pre and post ascites drainage. ND=not detected.

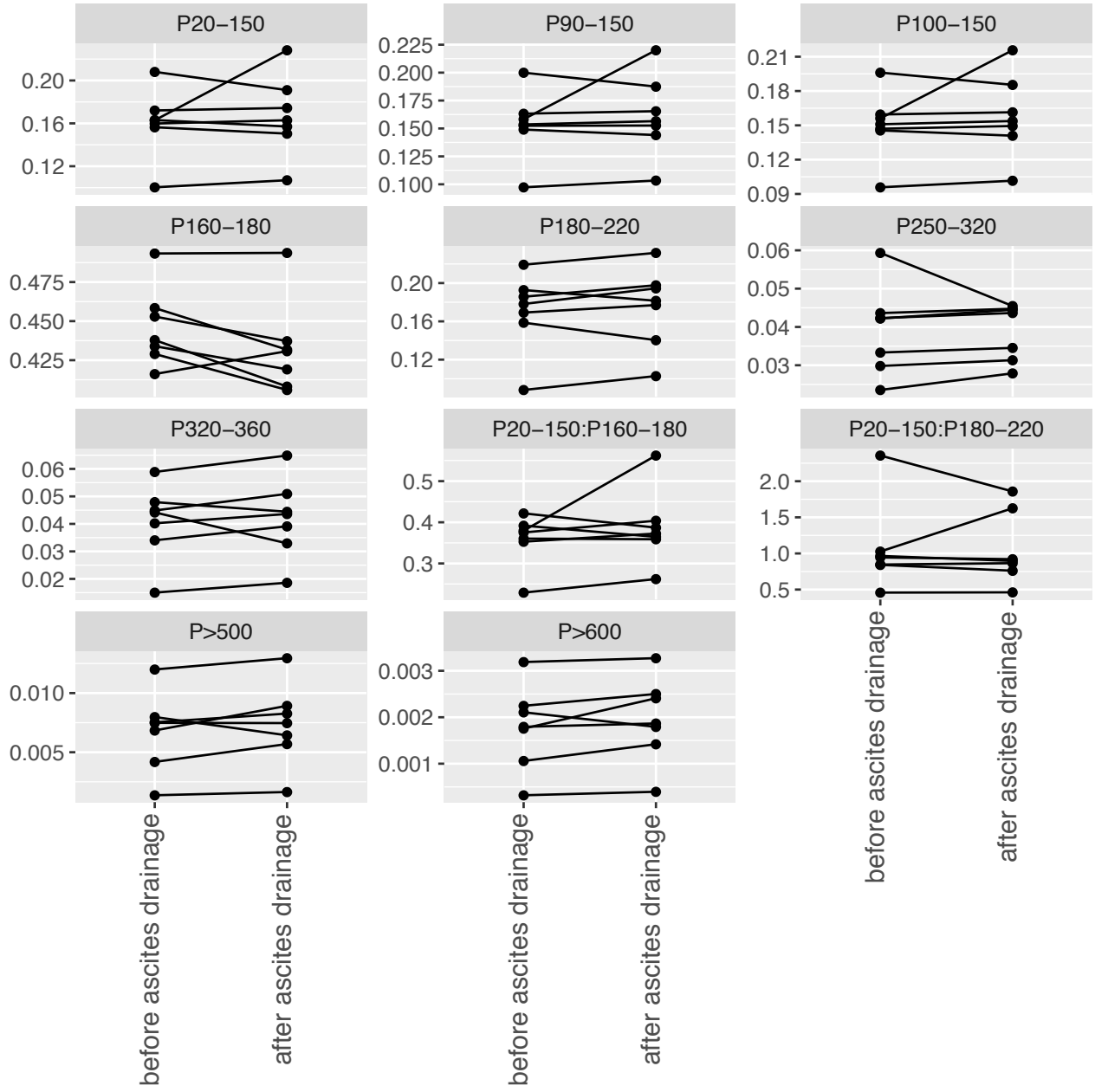


Figure 9.1: The proportion (P) of DNA fragments observed by sWGS in multiple size ranges (P20–150, P90–150, P100–150, P160–180, P180–220, P250–320, P320–360), the ratios of proportions of fragments in these size ranges (P20–150:P160–180, P20–150:P180–220) and the proportion of fragments greater than a certain size (P>500, P>600) before and after ascites drainage.

There was also no consistent change in fragmentation features comparing sWGS data before and after ascites collection (figure 9.1).

I also analysed plasma samples collected around the time of surgery for 6 women with HGSOC. Three of these underwent primary surgery and 3 IDS. Pre surgery, end of surgery, 2 hour post surgery, and daily plasma samples collected up to day 3 post surgery were analysed (Appendix 12.20) using TAm-Seq with primer panel 10 (Appendix 12.4). Analysis was performed using detection calling for the tumour specific *TP53* mutation identified in the patients FFPE sample. Unfortunately *TP53* mutations were not identified in any of the plasma samples.

I therefore identified those cases included in the CTCR-OV04 newly diagnosed cohort that had undergone primary surgery, had detectable ctDNA pre operatively and had a plasma sample available pre cycle 1 of adjuvant chemotherapy (n=5) (Table 9.2). 3/5 of the cases had a complete resection and 2/5 were debulked to <0.5 cm residual disease. I performed TAm-Seq using panel 10 (Appendix 12.4) for the pre-cycle 1 plasma sample. 4/5 of the cases did not have a detectable *TP53* mutation in the pre adjuvant chemotherapy plasma sample. One case did had a *TP53* mutation detected. The MAF had dropped from 0.224 pre surgery to 0.058 pre chemotherapy with debulking to <0.5 cm residual disease.

Patient	Residual disease post surgery	Mutation	MAF pre surgery	MAF pre cycle 1
716	<0.5 cm	7574012	0.224	0.058
559	complete resection	7578384	0.025	ND
650	complete resection	7578382	0.265	ND
658	<0.5 cm	7578534	0.092	ND
870	complete resection	7578455	0.015	ND

Table 9.2: MAF pre primary surgery and pre cycle 1 adjuvant chemotherapy for five CTCR-OV04 HGSOC cases.

sWGS was performed for samples collected around the time of surgery (up to day 3 post surgery) from 7 patients undergoing primary surgery for HGSOC (Table 9.3). The change in t-MAD score was not consistent across different patients (Figure 9.2a). OV04-788 had an increase in t-MAD score at the end and 2 hours after surgery that may be a result of surgical manipulation and release of ctDNA. This increase was seen to a lesser extent for OV04-791 and OV04-829. OV04-811 and OV04-890 had the opposite pattern with a drop in t-MAD score at the end of surgery which may be the result of reduced tumour burden.

Patient	Residual disease (cm)
622	0
788	>1.0
791	0
811	0
829	0
875	<0.5
890	0

Table 9.3: Residual disease for seven women undergoing primary suregry for HGSOC.

The size selected t-MAD scores when selecting for fragments between 20–150 bp (Figure 9.2b) and 90–150 bp (Figure 9.2c) were fairly similar. There was less variability in size selected t-MAD score over time other than in OV04-811 which maintained a large decrease in t-MAD score at the end of surgery.

A decrease in shorter fragments (P20–150, P90–150, P100–150) was seen at the end of surgery that then gradually increased over the next hours and days (Figure 9.3). The proportion of very long fragments (P>500 and P>600) remained relatively stable across time points, however the proportion of fragments between 160–180 (P160–180) increased whilst the proportion of fragments between 320–360 (P320–360) decreased. OV04-788 showed a different pattern of change in fragmentation features to the other cases around the time of surgery.

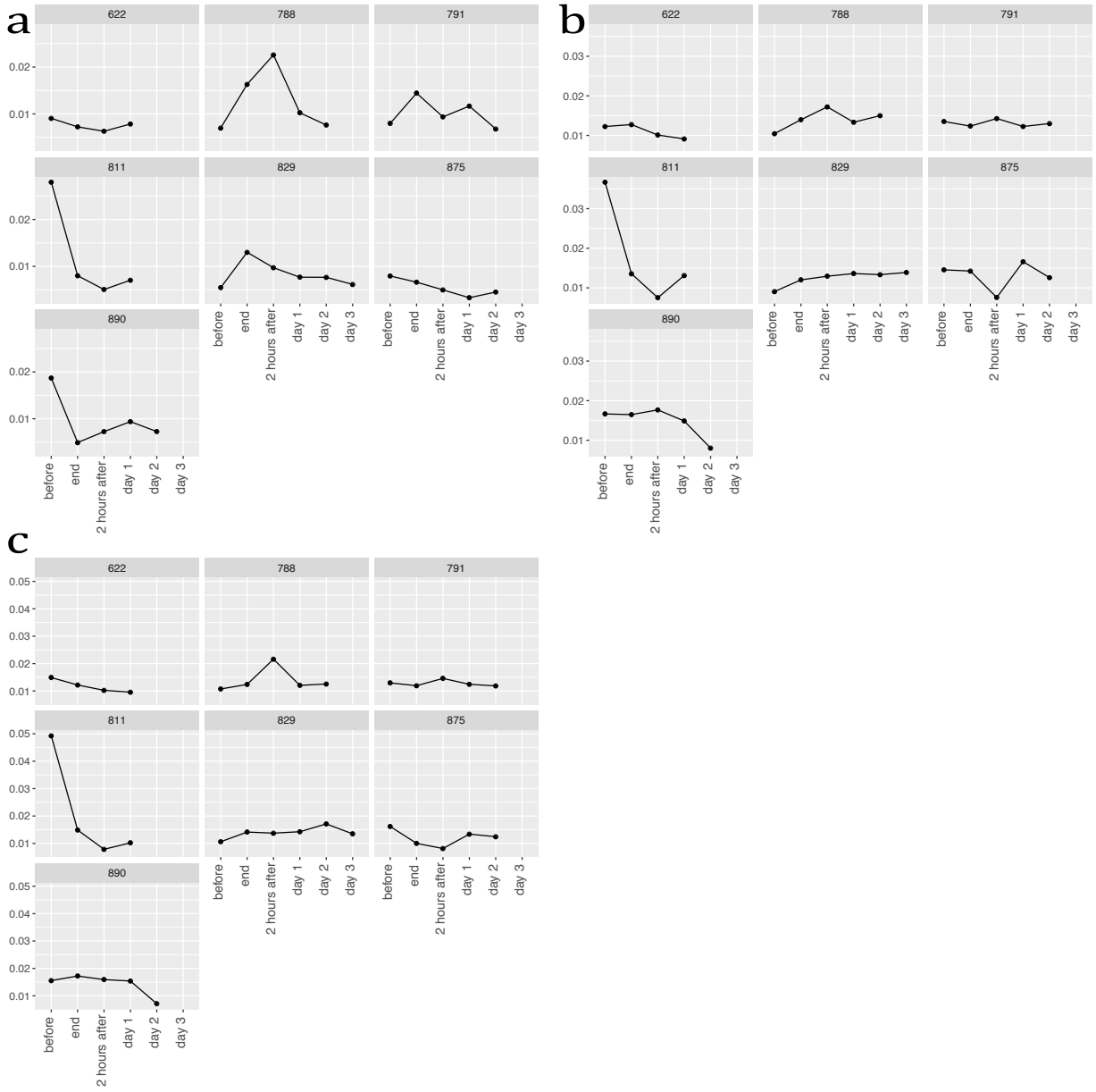


Figure 9.2: t-MAD score at multiple time points around the time of primary surgery for HGSOC for 7 women. **a)** Unselected t-MAD score. **b)** Size selected t-MAD score (20–150 bp). **c)** Size selected t-MAD score (90–150 bp)

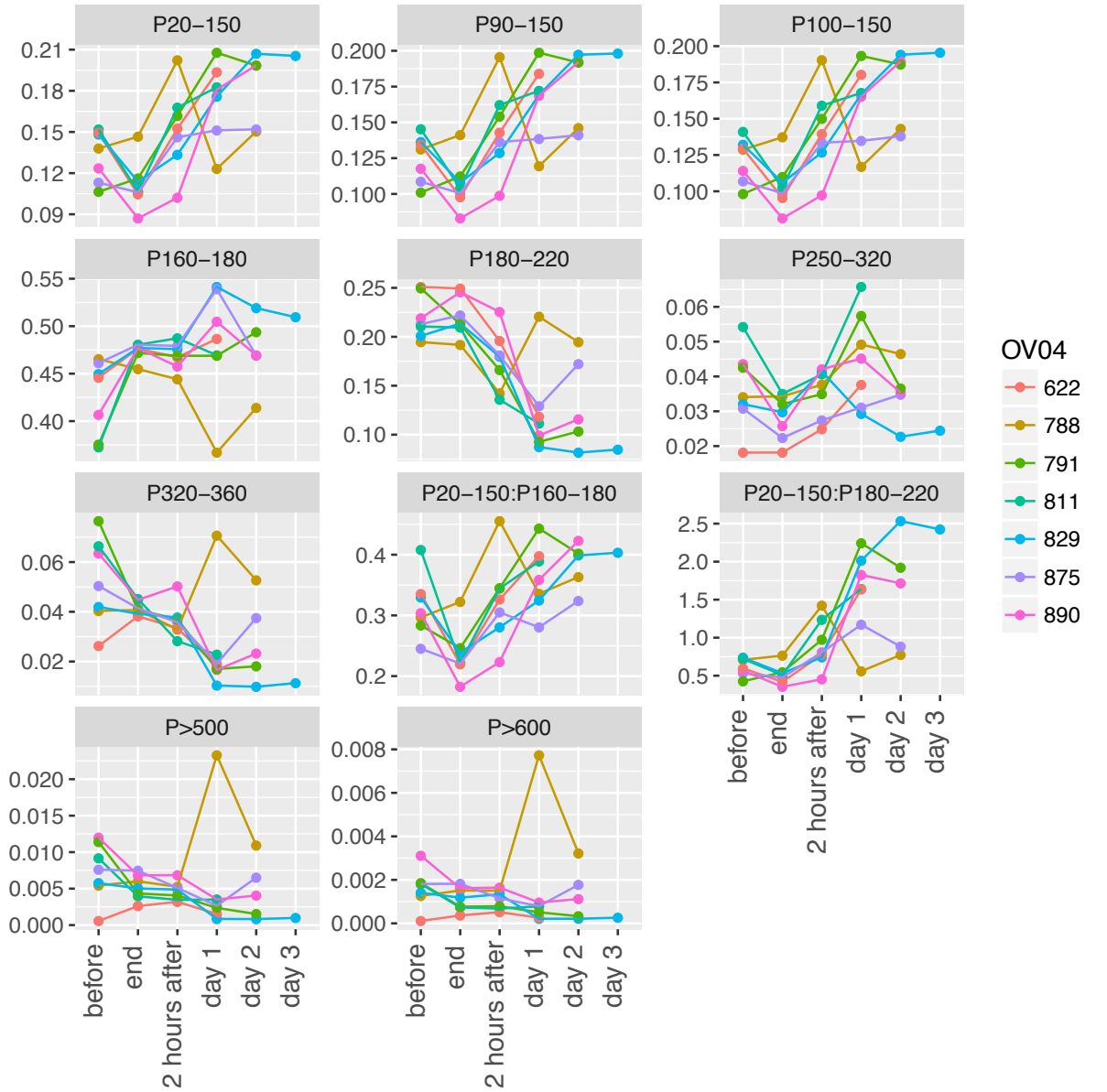


Figure 9.3: The proportion (P) of DNA fragments observed by sWGS in multiple size ranges (P20–150, P90–150, P100–150, P160–180, P180–220, P250–320, P320–360), the ratios of proportions of fragments in these size ranges (P20–150:P160–180, P20–150:P180–220) and the proportion of fragments greater than a certain size (P>500, P>600) around the time of primary surgery.

9.2.2 Collection tubes

Three tubes of blood were drawn from 14 patients with relapsed HGSOC. One EDTA tube and one Cell-free DNA BCT tubes were processed immediately (E.RT.0h), and another Cell-free DNA BCT was shipped via standard UK Royal Mail service back to the same collection centre. All shipped samples were received within 48h from the time of collection . There were no statistical significance in total cfDNA levels between the three collection methods (Figure 9.4a, b). *TP53* mutations were identified by TAm-Seq in 4 patients and the MAF was not statistically significantly across the different collection methods (Figure 9.4c, d). sWGS libraries were prepared using the Rubicon ThruPlex DNA-Seq kit for 8 of the patients (24 samples total) and sequenced on the HiSeq 2500 to further investigate the effects of collection method on global copy number changes. There were no significant difference between the copy number profiles among the three collection methods.

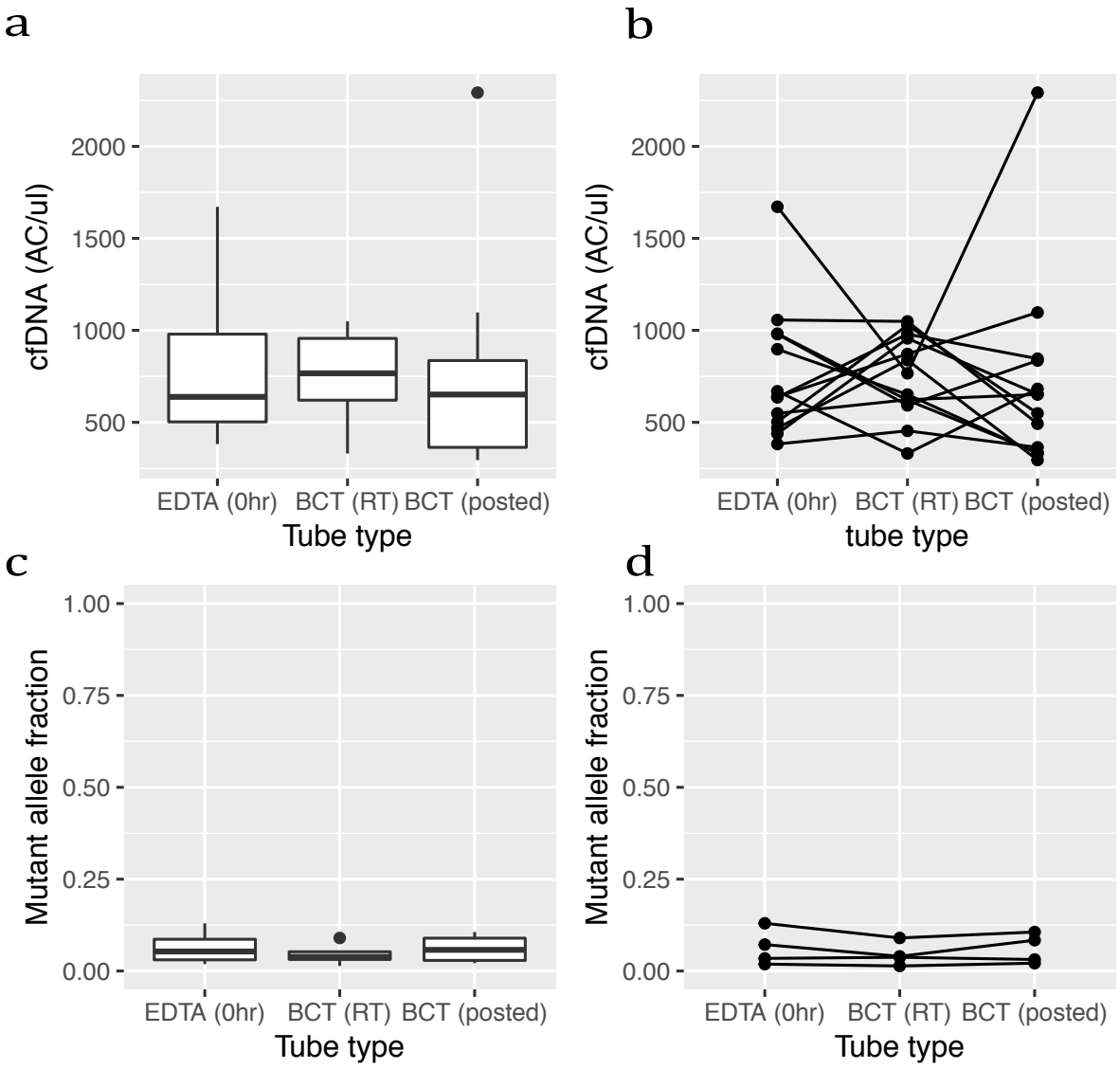


Figure 9.4: **a)** Total cfDNA AC/µl by tube type. **b)** Total cfDNA AC/µl per patient. **c)** MAF by tube tube. **d)** MAF per patient.

9.2.3 Serum

To investigate whether size selection could be used to enrich for tumour DNA in serum samples I performed TAm-Seq using primer panel 10 (Appendix 12.4) for 22 serum samples from women with HGSOC before and after in-vitro size selection (PippinHT, 3% gel, fragments 90–150 bp). There were some limitations with this experiment including the low sequencing coverage for some samples and inconsistent results for replicates of the same sample (Appendix 12.22). In some cases eg. OV04-180 chemotherapy cycle 1, OV04-292 chemotherapy cycle 1 and OV04-297 chemotherapy cycle 1 no *TP53* mutation could be detected in the unselected sample but could be detected in the post size selection sample, suggesting that size selection for fragments between 90–150 bp may enrich for ctDNA in these samples. However, in other cases including OV04-83 chemotherapy cycle 2, OV04-122 chemotherapy cycle 1, OV04-296 chemotherapy cycle 1 and OV04-300 chemotherapy cycle 1 size selecting for DNA fragments between 90–150 bp significantly reduced the TP53 MAF (Appendix 12.22). For this reason I hypothesised that cfDNA, particularly tumour derived cfDNA, in serum samples is of a different size to fragments seen in plasma samples. I therefore proceeded to investigate serum samples collected from xenograft models.

Whole genome libraries were prepared using the Rubicon ThruPlex DNA-Seq for 3 serum samples obtained from the same xenograft models as discussed in section 5.2.2 (Table 5.12). The proportion of reads aligning to the human genome (tumour) and the proportion of reads aligning to the mouse genome (non-tumour) were plotted against fragment length (Figure 9.5). The fragmentation profiles for serum for mouse b (Figure 9.5b) and mouse c (Figure 9.5c) were similar to those seen for the matched plasma samples (Figure 5.9b,c) with the tumour derived DNA fragments shifted towards a shorter size compared to the non-tumour cfDNA fragments. However, in the serum sample for mouse a (Figure 9.5a) a peak was seen at 50 bp that was not present in the plasma sample.

I therefore wanted to investigate if this 50 bp peak was seen in patient serum samples. I compared the fragmentation profiles for DNA from plasma and serum samples from HGSOC cases and controls. Different DNA extraction and library preparation methods lead to the preferential sequencing of different DNA fragments. I therefore also compared different extraction and library preparation methods (Figure 9.6).

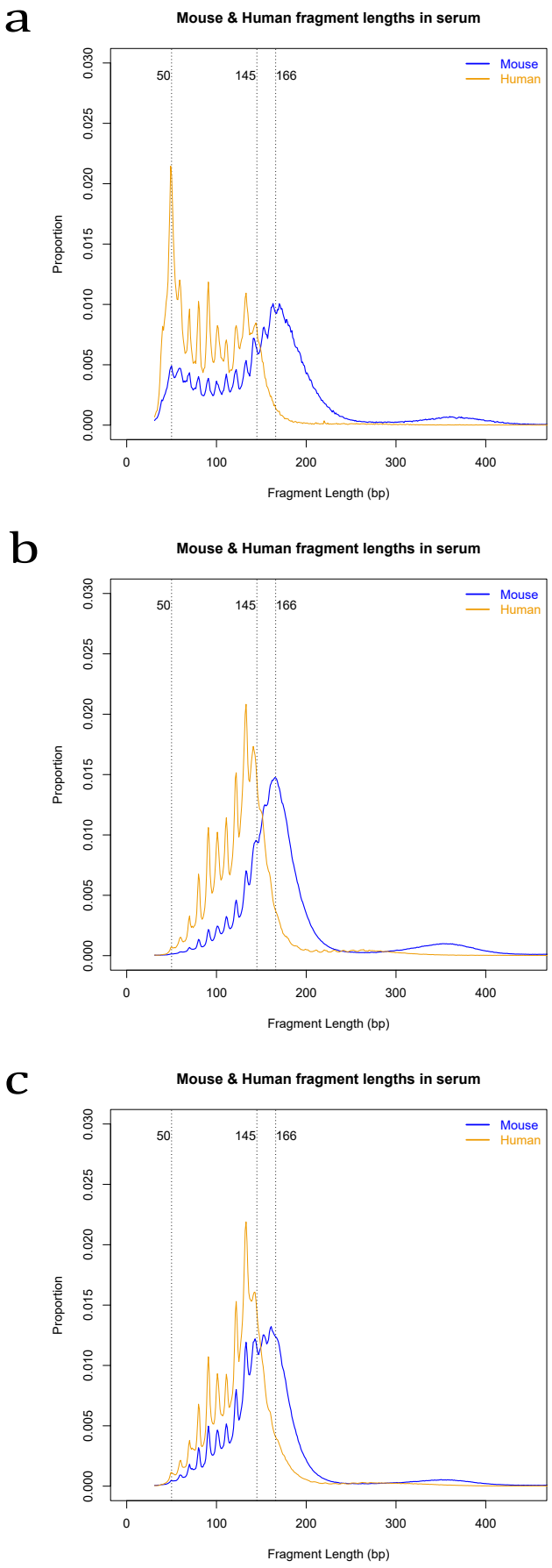


Figure 9.5: Xenograft fragmentation profiles **a)** OVCAR3 **b)** HGSOC ascites **c)** HGSOC solid tumour.

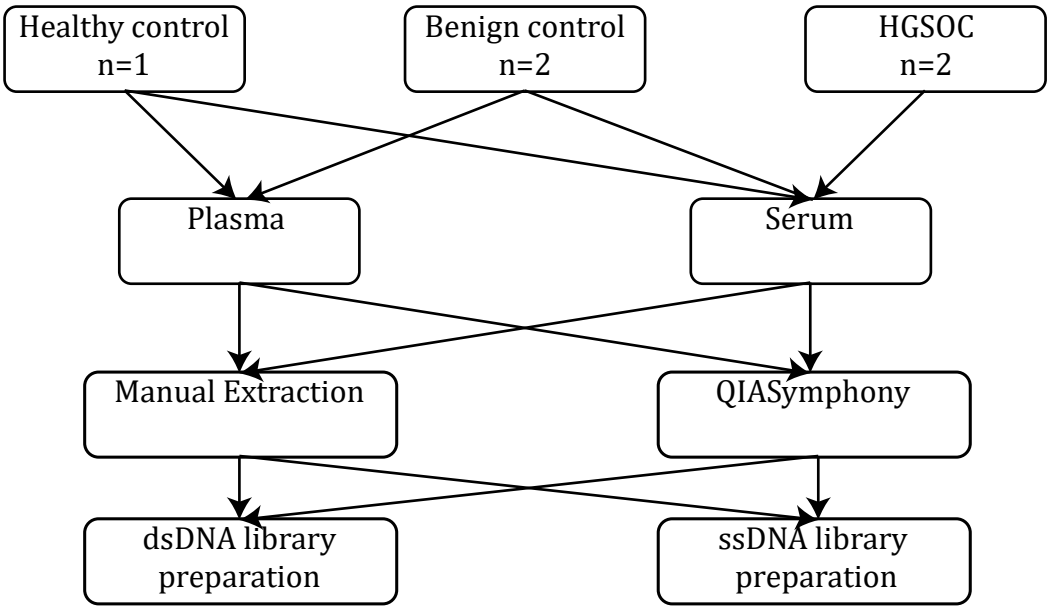


Figure 9.6: Schematic illustrating experimental design for comparing cfDNA fragments in plasma and serum samples. 1ml aliquots of both plasma and serum were extracted using the QIASymphony and the QIAvac 24 Plus vacuum manifold with QIAamp Circulating Nucleic Acid kit (manual extraction). Whole genome libraries were prepared using the Rubicon ThruPlex DNA Seq kit (dsDNA library preparation) and the DNA SMART ChIP-Seq Kit (ssDNA library preparation).

The profiles for the plasma and serum samples with the same extraction and library preparation methods were broadly similar (Figure 9.7). More shorter fragments were seen with the single-stranded library preparation method compared to the standard double stranded library preparation method particularly with the QIAsymphony extraction. HGSOC case 2 plasma samples had a fragmentation profile shifted to smaller sizes using the double stranded library preparation which is consistent with previous findings (unpublished). This same pattern was observed in the serum samples. A higher peak at 167 bp was seen in the healthy control serum sample compared to the other samples that is not seen to such an extent with the matched plasma sample. Using manual extraction and single stranded library preparation showed a slight shift to smaller sizes for the two cases compared to the controls for both the plasma and the serum samples. However using the QIASymphony extraction and single stranded library preparation there is a large 50 bp peak for the healthy and benign control plasma samples that is not so significant for the serum samples.

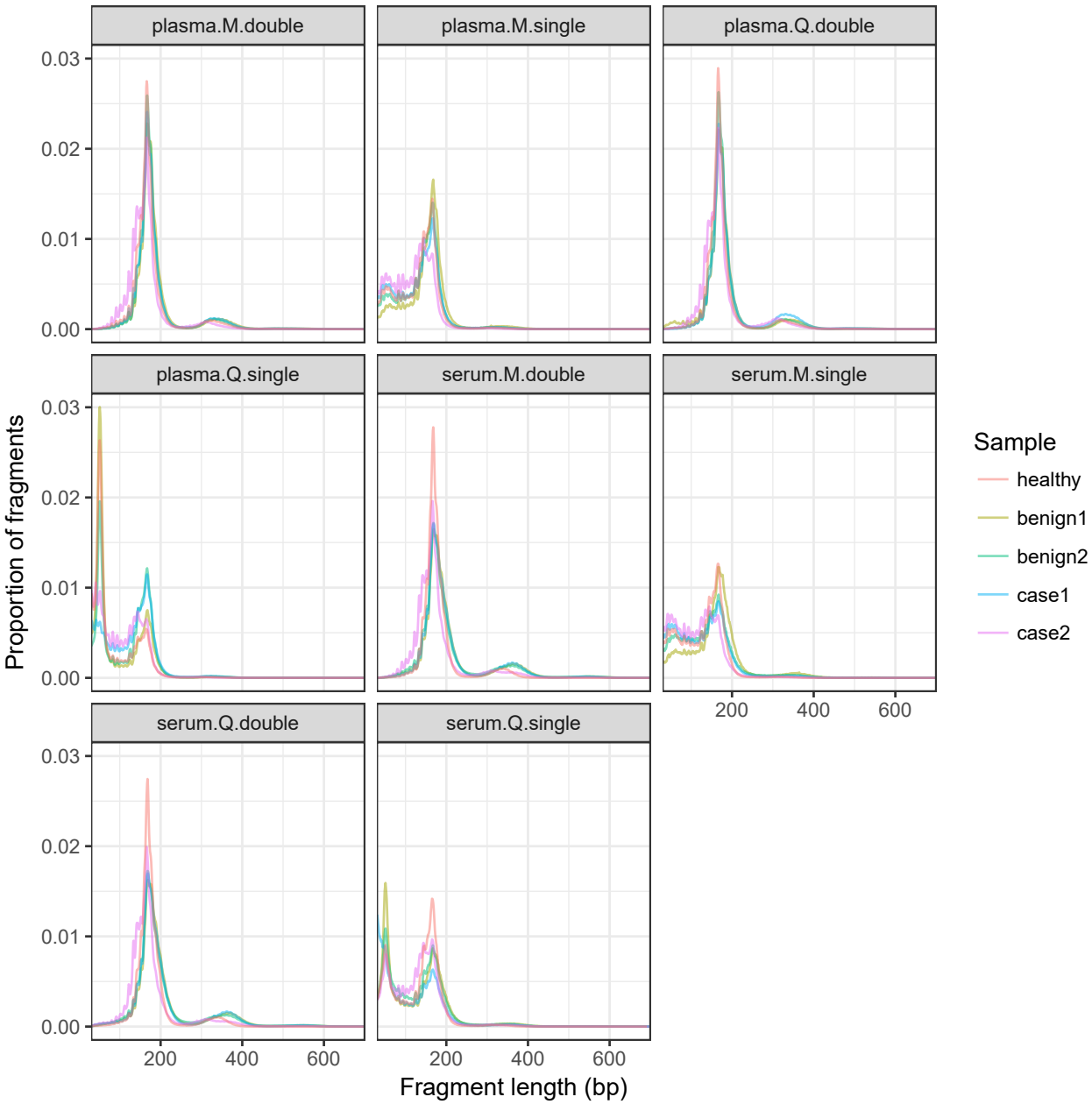


Figure 9.7: Density of reads against fragment size for healthy control, two benign controls and two HGSOC cases (plasma and serum) for different extraction and library preparation methods. Samples with <1 million reads were excluded and the lower limit of fragment size set at 30 bp as below this fragment alignment is not reliable. M=manual extraction (QIAvac 24 Plus vacuum manifold and the QIAamp Circulating Nucleic Acid kit), Q=QIAsymphony extraction, double=double stranded DNA library prep (Rubicon ThruPlex DNA Seq kit), single=single stranded DNA library prep (DNA SMART ChIP-Seq Kit).

9.3 Discussion

It has previously been shown that drainage of ascites can result in a decrease in ctDNA measured by *TP53* MAF (Parkinson et al. 2016). In this study ctDNA levels were measured 3 days before and 25 days after ascites drainage. As the half life of ctDNA is in the region of hours and not days (Diehl et al. 2008) to investigate the dynamics of ctDNA around the time of ascites drainage I collected plasma samples immediately before and immediately after drainage. Investigation of ctDNA around the time of ascites drainage using targeted sequencing and sWGS was inconclusive. It is possible that collecting blood immediately after completion of drainage was too soon to see the effect on ctDNA levels. Ideally blood would have been collect two hours after drainage and/or the following day to properly investigate the dynamics of ctDNA resulting from drainage of ascites. This will be included in future schedules of blood collection for studies of women with HGSOC. Alternatively ascites could have no effect on the level of ctDNA in the circulation and the observation in (Parkinson et al. 2016) could have been unrelated to the ascites drainage.

It is known that women with OC that have complete cytoreduction have improved survival rates compared to women who are sub-optimally debulked (Bristow et al. 2002) . The classification of the degree of cytoreduction is currently a subjective process performed by the operating surgeon and does not take into account micro metastases that may remain. It would therefore be useful to have an objective measure of residual tumour burden particularly if future clinical trials of additional therapy become available for those not sub optimally debulked.

Studies have shown that an increase in ctDNA post-operatively in lung cancer and colorectal cancer is predictive of early recurrence and death (Ng et al. 2017; Sun et al. 2018). The hypothesis being that the presence of ctDNA post-operatively reflects residual tumour volume. Hu et al. 2017 monitored ctDNA more frequently following surgery for lung cancer and found a peak in levels 3 days post operatively. After 3 days ctDNA levels dropped and only remained elevated at 30 days in patients with an early recurrence (within 4 months).

I was therefore interested in monitoring ctDNA levels around the time of surgery for HGSOC to identify 1) if ctDNA could provide a quantitative measure of residual disease and 2) if post operative ctDNA levels were predictive of clinical response. My working hypothesis was that ctDNA levels were likely to increase immediately following surgery due to manipulation of tumour tissue causing release of tumour-derived cfDNA into the circula-

tion. I then expected ctDNA levels to reflect residual tumour volume with women that had been completely debulked with no detectable ctDNA and women with residual disease potentially having detectable ctDNA although the levels may be below the limit of detection with the method applied.

Using targeted sequencing for *TP53* I was unable to identify ctDNA in plasma samples collected pre or post operatively from three women undergoing primary surgery and three women undergoing IDS. This is likely to be a result of the limited sensitivity of the TAm-Seq method for ctDNA detection. In five women undergoing primary surgery with detectable ctDNA pre-operatively only one had ctDNA detectable before commencing adjuvant chemotherapy. The women had post operative residual disease of < 0.5cm, as did one other women with non-detectable ctDNA pre adjuvant chemotherapy. Investigation of ctDNA around the time of surgery was therefore inconclusive. It is likely this is a result of the limitations in sensitivity of the TAm-Seq method. Using this method *TP53* mutations could only be detected in 49% of pre treatment plasma samples. With hindsight detecting *TP53* mutations in plasma pre IDS was optimistic accounting for the significant drop seen in *TP53* MAF following 1 and 2 cycles of chemotherapy (see section 8, page 150). This analysis was also only undertaken in a small number of cases.

sWGS also showed inconsistent results. In some cases the t-MAD score increased at the end of surgery and two hours post-operatively, in others a reduction was seen between pre and post-operative samples and other cases showed no change across time points. OV04-788 had >1 cm residual disease post-operatively and had a large increase in t-MAD score at the end of surgery. Unlike Hu et al. 2017 ctDNA levels starting dropping from day 1 after surgery. Further investigation of larger numbers of samples using more sensitive methods may be informative to better understand the dynamics of ctDNA around the time of surgery.

I wanted to look at the change in fragmentation features over the time of surgery to see if this provided any information about the size of tumour-derived cfDNA fragments or information relating to the mechanism of release of cfDNA fragments. The proportion of shorter cfDNA fragments tended to immediately decrease at the end of surgery and then increase over the following few days. If the shorter cfDNA fragments do reflect those released during proliferation this could indicate a reduction in proliferation during surgery followed by an initial rebound cellular proliferation after surgery. The proportion of very long fragments (>500 bp) remained stable across all time points however, there was a reduction in long fragments (between 180–360 bp) after surgery. The origin of longer cfDNA

fragments remains unknown however, it is interesting that the proportion of these fragments changes with surgery potentially as a result of tissue manipulation or reduction in tumour volume. It is also possible that having an operation can effect the amount, and type, of cfDNA in the circulation. It is possible that pre-medications, starving before a procedure, dehydration and lack of physical exercise can all effect the release of cfDNA although these factors are yet to be studied in detail.

In this chapter I have also shown that postage of cell stabilisation tubes does not have an impact on sWGS ctDNA analysis. This work was part of a larger study of storage and processing of plasma samples as well as the use of cell stabilisation tubes. This work has been accepted for publication in The Journal of Molecular Diagnostics. The result of this work is that in future multi-site studies cell stabilisation tubes will be used for plasma collection to standardise collection and enable central processing of samples to minimise biases.

Plasma samples are preferred for ctDNA analysis due to the large amount of background DNA found in serum samples resulting from WBC lysis. We have found a lower MAF in serum samples compared to matched plasma samples (unpublished). Many large scale studies historically banked serum samples as initial processing is easier and serum is used for protein biomarker analysis. Currently these are precluded from ctDNA studies due to the presumed large amount of wildtype DNA causing a reduction in the signal to noise ration potentially leading to ctDNA levels becoming undetectable. I was therefore interested to investigate if ctDNA could be enriched in serum samples, as it can in plasma samples, to allow use of the rich resource of banked serum samples for ctDNA analysis.

Although only a very small number of samples were investigated in-vitro size selection for DNA fragments between 90–150 bp followed by targeted sequencing did not show an enrichment in ctDNA as it did with plasma samples. It was hypothesised that this was due to a difference in the length of tumour-derived cfDNA fragments in serum, potentially resulting from the processing of the sample and the presence of clotting factors. However, fragmentation profiles of matched plasma and serum samples from two HGSOC cases, two benign controls and one healthy control did not support this hypothesis. It is interesting that different extraction methods and different library preparation methods allow the recovery of different DNA fragments and in future studies it will be important to account for the relevant biases.

Summary

In this thesis I have shown that ctDNA can be detected in women with newly diagnosed OC including in women with early stage disease.

Improvements to the published TAm-Seq method for targeted sequencing (Forshew et al. 2012) including optimised primers and a higher depth of sequencing have led to an improvement in the limit of ctDNA detection down to a MAF of 0.006. Using this method I was able to detect ctDNA in 30–49% of women with newly diagnosed OC.

I have shown that a genome wide score of copy number aberration (t-MAD score) can be used to quantify ctDNA in plasma samples and detect 39–41% of OC cases.

These methods although not as sensitive as other recently published methods (Bettegowda et al. 2014; Phallen et al. 2017; Cohen et al. 2018) have an advantage in that they are low-cost high-throughput methods that do not require prior knowledge of the patients tumour specific mutation and as such are methods that can be directly translated to the clinic.

However, a greater sensitivity will be required for a diagnostic biomarker. I have therefore discussed the development of TAm-Seq V2. This targeted sequencing method performs multiple reactions in parallel, each with a limited number of starting DNA molecules to ensure that the minimum AF is sufficiently above the level of background noise. In optimisation experiments I have shown that this method can detect down to a MAF of 0.001%. The limitation of this method is the availability of sufficient material to ensure that enough molecules are screened to detect low AF mutations. For this to be a viable method to test in populations with OC larger volumes of blood will need to be routinely collected which may not be the most efficient way of increasing sensitivity of detection.

I have also shown that the sensitivity of ctDNA detection can be increased by lever-

aging differences in tumour derived compared to non-tumour derived cfDNA fragments. In particular I have shown that in-vitro and in-silico size selection for shorter DNA fragments (90–150 bp or 20–150 bp) can increase the rates of ctDNA detection using targeted sequencing and sWGS.

Applying the size selected t-MAD score to the cohorts of women with newly diagnosed OC increased the ctDNA detection rate to 53–67%. Importantly the rate of ctDNA detection was high in the cases with early stage disease: 88% when using the upper limit of the healthy controls as a cut off and 77% when using the upper limit of the benign controls as a cut off. A limitation of this analysis is the small number of early stage cases investigated. Validation in larger cohorts is therefore required.

Table 10.1 outlines the percentage of OC cases in the UKOPS and CTCR-OV04 cohorts detected using the various different methods library preparation and analysis methods discussed in this thesis. Detection is variable across the different cohorts and it is not clear currently which parameters will provide the highest rate of detection of OC cases from healthy controls. It is likely that a combination of assays including protein biomarkers as well as ctDNA may provide the optimal detection for further validation in larger cohorts of women who do not already have an OC diagnosis.

Cohort	TAm-Seq	t-MAD	Size selected t-MAD (90–150 bp)	Size selected t-MAD (20–150 bp)	IchorCNA default	IchorCNA customised	Size selected IchorCNA default (90–150 bp)	Size selected IchorCNA default (20–150 bp)	Size selected IchorCNA customised (90–150 bp)	Size selected IchorCNA customised (20–150 bp)
UKOPS										
Whole cohort	30	41	53	53	61	88	46	45	69	73
HGSOC cases	45	56	56	56	61	83	44	44	72	83
CTCR-OV04										
Whole cohort	49	39	67	67	37	53	34	37	70	69
Stage 1	33	33	50	80	22	44	50	40	25	20
Stage 2	67	67	100	100	50	75	50	50	25	25
Stage 3	44	35	64	64	35	45	27	30	66	65
Stage 4	60	45	72	70	47	67	43	45	76	74

Table 10.1: Percentage of women with newly diagnosed OC in different cohorts with detectable ctDNA using different library preparation and analysis methods. Cut off used for the t-MAD score and IchorCNA is the upper limit of the healthy control cohort.

In contrast to relapsed disease (Parkinson et al. 2016) I found no correlation between ctDNA levels and volume of disease in women with newly diagnosed HGSOC. In NSCLC a stronger correlation was identified between tumour proliferation and ctDNA than between tumour volume and ctDNA (Abbosh et al. 2017). The relationship between tumour proliferation and ctDNA therefore needs to be investigated in newly diagnosed HGSOC to further understand the limitations of detection.

In this thesis I have also investigated the use of cervical sampling for use in the diagnosis of OC. I have developed methods for extraction and sequencing of DNA from cervical cytology and CMA samples. Although I have been able to detect tumour DNA in samples collected from women with cervical and endometrial cancers, this has not been the case for OCs. It is however possible that in the future an optimised cervical sampling device may enable the detection of tumour DNA in women with OC.

10.1 Future directions

In the future I plan to continue to develop a panel of assays that could be used in primary care to enable rapid triage of symptomatic women with the aim of diagnosing ovarian cancer at an earlier stage. The combination of protein biomarkers, including CA 125 and HE4, as well as the detection of ctDNA using targeted sequencing and the identification of tumour specific cfDNA fragmentation features may contribute to this clinical advancement.

During my PhD I have investigated ctDNA assays in incident cancer cases. Ongoing work will focus on further bioinformatics analysis of incident cancer cases with a strong focus on combining assays to improve both sensitivity and specificity of OC identification. The combination of assays identified in cohorts of women with a diagnosis of OC will then be validated in larger cohorts of women who do not already have an OC diagnosis.

I am a co-applicant on a CRUK Biomarker Project award led by Sudha Sundar from the University of Birmingham for the Refining Ovarian Cancer Test accuracy Scores – Genomic (ROCkeTS-GEN) study. This is a prospective single arm diagnostic accuracy study study of postmenopausal women with symptoms and either a raised CA 125 or abnormal ultrasound scan or both. 2000 postmenopausal women will be recruited at the time of referral to secondary care. The prevalence of OC in symptomatic postmenopasual women is 8%. ctDNA analysis will therefore be performed on 160 OC cases (of which 100 are likely to be HGSOC) and 200 controls without OC. The sensitivity and specificity of ctDNA will be com-

pared to RMI and CA 125. The combination of ctDNA and CA 125 and other protein based biomarkers identified in ROCkeTS will also be investigated.

I am also a co-applicant on a CRUK Early Detection Programme Award led by Doug Easton and Nitzan Rosenfeld. The purpose of the programme is to study cohorts of women at high genetic risk of ovarian (and breast) cancer in order to investigate the potential of using DNA in blood or cervical samples for the earlier detection of ovarian (and/or breast) cancers. We plan to collect annual plasma samples from over 6,000 women at high risk of ovarian (and breast) cancer. 100 of these women are expected to develop OC over the 5-years this programme will run for. ctDNA analysis using both targeted sequencing and sWGS will be performed on sequential samples from the 100 women with an OC diagnosis and 100 women not diagnosed with a cancer during the 5-years. This will again provide information about the sensitivity and specificity of ctDNA assays in a pre-diagnosis setting. We will also be able to look at serial samples to identify the earliest timepoint that ctDNA could be detected prior to a cancer diagnosis.

We also plan to collect annual cervical samples from women at high risk of OC. This will provide a larger sample set to further investigate whether tumour DNA can be detected in cervical cytology samples from women with OC. Through this study I also plan to locally test other methods for cervical sampling that may be better optimised to collect cells and DNA shed from the distal fallopian tube.

Finally we plan to investigate ctDNA assays in combination with p53 autoantibodies and protein biomarkers in women being followed up after completion of first line treatment for HGSOC. The premise of this study is that the development of assays to detect relapse early are potentially applicable to a diagnostic setting. Findings from this cohort with a known cancer diagnosis can then feed into the studies with sample collections from women without a cancer diagnosis, with the aim of further improving sensitivity of disease detection.

Validation of the combination of the best biomarkers will then be performed in a large study of symptomatic women in primary care.

Bibliography

Abbosh, C. et al. (2017). Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* 545.7655, 446–451.

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* 8.1, 1324.

Ahmed, A. A., D. Etemadmoghadam, J. Temple, A. G. Lynch, M. Riad, R. Sharma, C. Stewart, S. Fereday, C. Caldas, A. Defazio, D. Bowtell and J. D. Brenton (2010). Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *J Pathol* 221.1, 49–56.

Anderson, K. S., J. Wong, A. Vitonis, C. P. Crum, P. M. Sluss, J. Labaer and D. Cramer (2010). p53 autoantibodies as potential detection and prognostic biomarkers in serous ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 19.3, 859–68.

Angelopoulou, K., B. Rosen, M. Stratis, H. Yu, M. Solomou and E. P. Diamandis (1996). Circulating antibodies against p53 protein in patients with ovarian carcinoma. Correlation with clinicopathologic features and survival. *Cancer* 78.10, 2146–52.

Angelopoulou, K., H. Yu, B. Bharaj, M. Giai and E. P. Diamandis (2000). p53 gene mutation, tumor p53 protein overexpression, and serum p53 autoantibody generation in patients with breast cancer. *Clin Biochem* 33 (1), 53–62.

Bast, R. C., T. L. Klug, E. S. John, E. Jenison, J. M. Niloff, H. Lazarus, R. S. Berkowitz, T. Leavitt, C. T. Griffiths, L. Parker, V. R. Zurawski and R. C. Knapp (1983). A Radioimmunoassay Using a Monoclonal Antibody to Monitor the Course of Epithelial Ovarian Cancer. *New England Journal of Medicine* 309.15, 883–887.

Belic, J., M. Koch, P. Ulz, M. Auer, T. Gerhalter, S. Mohan, K. Fischereder, E. Petru, T. Bauernhofer, J. B. Geigl, M. R. Speicher and E. Heitzer (2015). Rapid Identification of Plasma DNA Samples with Increased ctDNA Levels by a Modified FAST-SeqS Approach. *Clin Chem* 61 (6), 838–849.

Bettegowda, C., M. Sausen, R. J. Leary, I. Kinde, Y. X. Wang, N. Agrawal, B. R. Bartlett, H. Wang, B. Luber, R. M. Alani, E. S. Antonarakis, N. S. Azad, A. Bardelli, H. Brem, J. L. Cameron, C. C. Lee, L. A. Fecher, G. L. Gallia, P. Gibbs, D. Le, R. L. Giuntoli, M. Goggins, M. D. Hogarty, M. Holdhoff, S. M. Hong, Y. C. Jiao, H. H. Juhl, J. J. Kim, G. Siravegna, D. A. Laheru, C. Lauricella, M. Lim, E. J. Lipson, S. K. N. Marie, G. J. Netto, K. S. Oliner, A. Olivi, L. Olsson, G. J. Riggins, A. Sartore-Bianchi, K. Schmidt, I. M. Shih, S. M. Oba-Shinjo, S. Siena, D. Theodorescu, J. N. Tie, T. T. Harkins, S. Veronese, T. L. Wang, J. D. Weingart, C. L. Wolfgang, L. D. Wood, D. M. Xing, R. H. Hruban, J. Wu, P. J. Allen, C. M. Schmidt, M. A. Choti, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, N. Papadopoulos and A. J. Luis (2014). Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci Transl Med* 6.224.

Boyraz, G., D. Basaran, M. C. Salman, A. Ibrahimov, S. Onder, O. Akman, N. Ozgul and K. Yuce (2017). Histological Follow-Up in Patients with Atypical Glandular Cells on Pap Smears. *J Cytol* 34.4, 203–207.

Bristow, R. E., R. S. Tomacruz, D. K. Armstrong, E. L. Trimble and F. J. Montz (2002). Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 20 (5), 1248–1259.

Buys, S. S., E. Partridge, A. Black, C. C. Johnson, L. Lamerato, C. Isaacs, D. J. Reding, R. T. Greenlee, L. A. Yokochi, B. Kessel, E. D. Crawford, T. R. Church, G. L. Andriole, J. L. Weissfeld, M. N. Fouad, D. Chia, B. O'Brien, L. R. Ragard, J. D. Clapp, J. M. Rathmell, T. L. Riley, P. Hartge, P. F. Pinsky, C. S. Zhu, G. Izmirlian, B. S. Kramer, A. B. Miller, J. L. Xu, P. C. Prorok, J. K. Gohagan and C. D. Berg (2011). Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA* 305.22, 2295–303.

Carlson, J. W., A. Miron, E. A. Jarboe, M. M. Parast, M. S. Hirsch, Y. Lee, M. G. Muto, D. Kindelberger and C. P. Crum (2008). Serous tubal intraepithelial carcinoma: its potential role in primary peritoneal serous carcinoma and serous cancer prevention. *Journal of clin-*

ical oncology : official journal of the American Society of Clinical Oncology 26 (25), 4160–4165.

Chandrananda, D. (2017). *CNAclinic*. URL: https://github.com/sdchandra/CNAclinic.

Chandrananda, D., N. P. Thorne and M. Bahlo (2015). High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. *BMC Med Genomics* 8, 29.

Ciriello, G., M. L. Miller, B. A. Aksoy, Y. Senbabaoglu, N. Schultz and C. Sander (2013). Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* 45.10, 1127–33.

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, J. Diaz L. A., 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* 359.6378, 926–930.

Coleman, M. P., D. Forman, H. Bryant, J. Butler, B. Rachet, C. Maringe, U. Nur, E. Tracey, M. Coory, J. Hatcher, C. E. McGahan, D. Turner, L. Marrett, M. L. Gjerstorff, T. B. Johannesen, J. Adolfsson, M. Lambe, G. Lawrence, D. Meechan, E. J. Morris, R. Middleton, J. Steward and M. A. Richards (2011). Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* 377.9760, 127–38.

Cramer, D. W., J. Bast R. C., C. D. Berg, E. P. Diamandis, A. K. Godwin, P. Hartge, A. E. Lokshin, K. H. Lu, M. W. McIntosh, G. Mor, C. Patriotis, P. F. Pinsky, M. D. Thornquist, N. Scholler, S. J. Skates, P. M. Sluss, S. Srivastava, D. C. Ward, Z. Zhang, C. S. Zhu and N. Urban (2011). Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. *Cancer Prev Res (Phila)* 4.3, 365–74.

CRUK (2016). *Ovarian cancer statistics*. Ed. by C. R. UK. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer.

Dawson, S.-J., D. W. Y. Tsui, M. Murtaza, H. Biggs, O. M. Rueda, S.-F. Chin, M. J. Dunning, D. Gale, T. Forshew, B. Mahler-Araujo, S. Rajan, S. Humphray, J. Becq, D. Halsall, M. Wallis, D. Bentley, C. Caldas and N. Rosenfeld (2013). Analysis of circulating tumor DNA to monitor metastatic breast cancer. *The New England journal of medicine* 368 (13), 1199–1209.

DeSimone, C. P., M. E. Day, M. M. Tovar, 3. Dietrich C. S., M. L. Eastham and S. C. Modesitt (2006). Rate of pathology from atypical glandular cell Pap tests classified by the Bethesda 2001 nomenclature. *Obstet Gynecol* 107.6, 1285–91.

Diaz Jr., L. A., R. T. Williams, J. Wu, I. Kinde, J. R. Hecht, J. Berlin, B. Allen, I. Bozic, J. G. Reiter, M. A. Nowak, K. W. Kinzler, K. S. Oliner and B. Vogelstein (2012). The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 486.7404, 537–540.

Diehl, F., K. Schmidt, M. A. Choti, K. Romans, S. Goodman, M. Li, K. Thornton, N. Agrawal, L. Sokoll, S. A. Szabo, K. W. Kinzler, B. Vogelstein and L. A. Diaz (2008). Circulating mutant DNA to assess tumor dynamics. *Nat Med* 14 (9), 985–990.

Elliss-Brookes, L., S. McPhail, A. Ives, M. Greenslade, J. Shelton, S. Hiom and M. Richards (2012). Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *Br J Cancer* 107.8, 1220–6.

Erickson, B. K., I. Kinde, Z. C. Dobbin, Y. Wang, J. Y. Martin, R. D. Alvarez, M. G. Conner, W. K. Huh, R. B. S. Roden, K. W. Kinzler, N. Papadopoulos, B. Vogelstein, L. a. Diaz and C. N. Landen (2014). Detection of Somatic TP53 Mutations in Tampons of Patients With High-Grade Serous Ovarian Cancer. *Obstetrics & Gynecology* 124.5, 881–885.

Forshew, T., M. Murtaza, C. Parkinson, D. Gale, D. W. Tsui, F. Kaper, S.-j. Dawson, A. M. Piskorz, M. Jimenez-Linan, D. Bentley, J. Hadfield, A. P. May, C. Caldas, J. D. Brenton and N. Rosenfeld (2012). Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA. *Sci Transl Med* 4.136, 136ra68–136ra68.

Garcia-Murillas, I., G. Schiavon, B. Weigelt, C. Ng, S. Hrebien, R. J. Cutts, M. Cheang, P. Osin, A. Nerurkar, I. Kozarewa, J. A. Garrido, M. Dowsett, J. S. Reis-Filho, I. E. Smith and N. C. Turner (2015). Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Science Translational Medicine* 7.302, 302ra133–302ra133.

Gerstung, M., C. Jolly, I. Leshchiner, S. C. Dentro, S. Gonzalez, T. J. Mitchell, Y. Rubanova, P. Anur, D. Rosebrock, K. Yu, M. Tarabichi, A. Deshwar, J. Wintersinger, K. Kleinheinz, I. Vazquez-Garcia, K. Haase, S. Sengupta, G. Macintyre, S. Malikic, N. Donmez, D. G. Livitz, M. Cmero, J. Demeulemeester, S. Schumacher, Y. Fan, X. Yao, J. Lee, M. Schlesner, P. C. Boutros, D. D. Bowtell, H. Zhu, G. Getz, M. Imielinski, R. Beroukhim, S. C. Sahinalp, Y. Ji, M. Peifer, F. Markowetz, V. Mustonen, K. Yuan, W. Wang, Q. D. Morris, P. T. Spellman, D. C. Wedge, P. V. Loo and and (2017). The evolutionary history of 2,658 cancers.

Giacona, M. B., G. C. Ruben, K. A. Iczkowski, T. B. Roos, D. M. Porter and G. D. Sorenson (1998). Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. *Pancreas* 17.1, 89–97.

Hu, W., Y. Yang, L. Zhang, J. Yin, J. Huang, L. Huang, H. Gu, G. Jiang and J. Fang (2017). Post surgery circulating free tumor DNA is a predictive biomarker for relapse of lung cancer. *Cancer medicine* 6 (5), 962–974.

Jacobs, I. J., U. Menon, A. Ryan, A. Gentry-Maharaj, M. Burnell, J. K. Kalsi, N. N. Amso, S. Apostolidou, E. Benjamin, D. Cruickshank, D. N. Crump, S. K. Davies, A. Dawnay, S. Dobbs, G. Fletcher, J. Ford, K. Godfrey, R. Gunu, M. Habib, R. Hallett, J. Herod, H. Jenkins, C. Karpinskyj, S. Leeson, S. J. Lewis, W. R. Liston, A. Lopes, T. Mould, J. Murdoch, D. Oram, D. J. Rabideau, K. Reynolds, I. Scott, M. W. Seif, A. Sharma, N. Singh, J. Taylor, F. Warburton, M. Widschwendter, K. Williamson, R. Woolas, L. Fallowfield, A. J. McGuire, S. Campbell, M. Parmar and S. J. Skates (2016). Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet* 387.10022, 945–956.

Jacobs, I. and J. Bast R. C. (1989). The CA 125 tumour-associated antigen: a review of the literature. *Hum Reprod* 4.1, 1–12.

Jacobs, I., A. P. Davies, J. Bridges, I. Stabile, T. Fay, A. Lower, J. G. Grudzinskas and D. Oram (1993). Prevalence screening for ovarian cancer in postmenopausal women by CA 125 measurement and ultrasonography. *BMJ* 306.6884, 1030–4.

Jacobs, I., D. Oram, J. Fairbanks, J. Turner, C. Frost and J. G. Grudzinskas (1990). A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *British journal of obstetrics and gynaecology* 97 (10), 922–929.

Jahr, S., H. Hentze, S. Englisch, D. Hardt, F. O. Fackelmayer, R. D. Hesch and R. Knippers (2001). DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* 61.4, 1659–65.

Jiang, P., C. W. Chan, K. C. Chan, S. H. Cheng, J. Wong, V. W. Wong, G. L. Wong, S. L. Chan, T. S. Mok, H. L. Chan, P. B. Lai, R. W. Chiu and Y. M. Lo (2015). Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A* 112.11, E1317–25.

Jiang, P. and Y. M. D. Lo (2016). The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics. *Trends Genet* 32.6, 360–371.

Kim, J., D. M. Coffey, C. J. Creighton, Z. Yu, S. M. Hawkins and M. M. Matzuk (2012). High-grade serous ovarian cancer arises from fallopian tube in a mouse model. *Proceedings of the National Academy of Sciences* 109.10, 3921–3926.

Kinde, I., C. Bettegowda, Y. Wang, J. Wu, N. Agrawal, I.-M. Shih, R. Kurman, F. Dao, D. a. Levine, R. Giuntoli, R. Roden, J. R. Eshleman, J. P. Carvalho, S. K. N. Marie, N. Papadopoulos, K. W.

Kinzler, B. Vogelstein and L. a. Diaz (2013). Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. *Sci Transl Med* 5.167, 167ra4–167ra4.

Kindelberger, D. W., Y. Lee, A. Miron, M. S. Hirsch, C. Feltmate, F. Medeiros, M. J. Callahan, E. O. Garner, R. W. Gordon, C. Birch, R. S. Berkowitz, M. G. Muto and C. P. Crum (2007). Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: Evidence for a causal relationship. *The American journal of surgical pathology* 31 (2), 161–169.

Kobayashi, H., Y. Yamada, T. Sado, M. Sakata, S. Yoshida, R. Kawaguchi, S. Kanayama, H. Shigetomi, S. Haruta, Y. Tsuji, S. Ueda and T. Kitanaka (2008). A randomized study of screening for ovarian cancer: a multicenter study in Japan. *Int J Gynecol Cancer* 18.3, 414–20.

Kobel, M., A. M. Piskorz, S. Lee, S. Lui, C. LePage, F. Marass, N. Rosenfeld, A. M. Mes Masson and J. D. Brenton (2016). Optimized p53 immunohistochemistry is an accurate predictor of TP53 mutation in ovarian carcinoma. *J Pathol Clin Res* 2.4, 247–258.

Kuang, Y., A. Rogers, B. Y. Yeap, L. Wang, M. Makrigiorgos, K. Vetrand, S. Thiede, R. J. Distel and P. A. Janne (2009). Noninvasive Detection of EGFR T790M in Gefitinib or Erlotinib Resistant Non-Small Cell Lung Cancer. *Clinical Cancer Research* 15.8, 2630–2636.

Kuhn, E., R. J. Kurman, R. Vang, A. S. Sehdev, G. Han, R. Soslow, T.-L. Wang and I.-M. Shih (2012). TP53 mutations in serous tubal intraepithelial carcinoma and concurrent pelvic high-grade serous carcinoma–evidence supporting the clonal relationship of the two lesions. *The Journal of pathology* 226 (3), 421–426.

Labidi-Galy, S. I., E. Papp, D. Hallberg, N. Niknafs, V. Adleff, M. Noe, R. Bhattacharya, M. Novak, S. Jones, J. Phallen, C. A. Hruban, M. S. Hirsch, D. I. Lin, L. Schwartz, C. L. Maire, J.-C. Tille, M. Bowden, A. Ayhan, L. D. Wood, R. B. Scharpf, R. Kurman, T.-L. Wang, I.-M. Shih, R. Karchin, R. Drapkin and V. E. Velculescu (2017). High grade serous ovarian carcinomas originate in the fallopian tube. *Nat Commun* 8 (1), 1093.

Lamzabi, I., L. Buckingham, M. Gebrekiristos, R. Jain, P. Gattuso, V. Reddy, A. Guirguis, S. Dewdney, J. Rotmensch and P. Bitterman (2013). Use of cervical mucus to screen for gynecological malignancies: a pilot study. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 26 (11), 1508–1513.

Lee, Y., A. Miron, R. Drapkin, M. R. Nucci, F. Medeiros, A. Saleemuddin, J. Garber, C. Birch, H. Mou, R. W. Gordon, D. W. Cramer, F. D. McKeon and C. P. Crum (2007). A candidate precursor to serous carcinoma that originates in the distal fallopian tube. *The Journal of pathology* 211, 26–35.

Leon, S. A., B. Shapiro, D. M. Sklaroff and M. J. Yaros (1977). Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res* 37.3, 646–50.

Lim, A. W., D. Mesher and P. Sasieni (2014). Estimating the workload associated with symptoms-based ovarian cancer screening in primary care: an audit of electronic medical records. *BMC Fam Pract* 15, 200.

Lo, Y. M., K. C. Chan, H. Sun, E. Z. Chen, P. Jiang, F. M. Lun, Y. W. Zheng, T. Y. Leung, T. K. Lau, C. R. Cantor and R. W. Chiu (2010). Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med* 2.61, 61ra91.

Lu, D., E. Kuhn, R. E. Bristow, R. L. Giuntoli, S. K. Kjær, I.-M. Shih and R. B. Roden (2011). Comparison of candidate serologic markers for type I and type II ovarian cancer. *Gynecol Oncol* 122.3, 560–566.

Lun, F. M., N. B. Tsui, K. C. Chan, T. Y. Leung, T. K. Lau, P. Charoenkwan, K. C. Chow, W. Y. Lo, C. Wanapirak, T. Sanguansermsri, C. R. Cantor, R. W. Chiu and Y. M. Lo (2008). Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. *Proc Natl Acad Sci U S A* 105.50, 19920–5.

Mandel, P. and P. Metais (1948). [Not Available]. *C R Seances Soc Biol Fil* 142.3-4, 241–3.

Maritschnegg, E., Y. Wang, N. Pecha, R. Horvat, E. V. Nieuwenhuysen, I. Vergote, F. Heitz, J. Sehouli, I. Kinde, L. A. Diaz, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, P. Speiser and R. Zeillinger (2015). Lavage of the Uterine Cavity for Molecular Detection of Müllerian Duct Carcinomas: A Proof-of-Concept Study. *Journal of Clinical Oncology* 33.36, 4293–4300.

Menon, U., A. Gentry-Maharaj, R. Hallett, A. Ryan, M. Burnell, A. Sharma, S. Lewis, S. Davies, S. Philpott, A. Lopes, K. Godfrey, D. Oram, J. Herod, K. Williamson, M. W. Seif, I. Scott, T. Mould, R. Woolas, J. Murdoch, S. Dobbs, N. N. Amso, S. Leeson, D. Cruickshank, A. McGuire, S. Campbell, L. Fallowfield, N. Singh, A. Dawnay, S. J. Skates, M. Parmar and I. Jacobs (2009). Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncol* 10.4, 327–40.

Minarik, G., G. Repiska, M. Hyblova, E. Nagyova, K. Soltys, J. Budis, F. Duris, R. Sysak, M. Gerykova Bujalkova, B. Vlkova-Izrael, O. Biro, B. Nagy and T. Szemes (2015). Utilization of Benchtop Next Generation Sequencing Platforms Ion Torrent PGM and MiSeq in Noninvasive Prenatal Testing for Chromosome 21 Trisomy and Testing of Impact of In Silico and Physical Size Selection on Its Analytical Performance. *PLoS One* 10.12, e0144811.

Moore, R. G., M. Jabre-Raughley, A. K. Brown, K. M. Robison, M. C. Miller, W. J. Allard, R. J. Kurman, R. C. Bast and S. J. Skates (2010). Comparison of a novel multiple marker as-

say vs the Risk of Malignancy Index for the prediction of epithelial ovarian cancer in patients with a pelvic mass. *Am J Obstet Gynecol* 203.3, 228.e1–6.

Mouliere, F., S. El Messaoudi, D. Pang, A. Dritschilo and A. R. Thierry (2014). Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol Oncol* 8.5, 927–41.

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* 6.9, e23418.

Murtaza, M., S.-j. Dawson, D. W. Y. Tsui, D. Gale, T. Forshew, A. M. Piskorz, C. Parkinson, S.-f. Chin, Z. Kingsbury, A. S. C. Wong, F. Marass, S. Humphray, J. Hadfield, D. Bentley, T. M. Chin, J. D. Brenton, C. Caldas and N. Rosenfeld (2013). Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* 497.7447, 108–112.

Nagell J. R., J. van, P. D. DePriest, F. R. Ueland, C. P. DeSimone, A. L. Cooper, J. M. McDonald, E. J. Pavlik and R. J. Kryscio (2007). Ovarian cancer screening with annual transvaginal sonography: findings of 25,000 women screened. *Cancer* 109.9, 1887–96.

Ng, S. B., C. Chua, M. Ng, A. Gan, P. S. Poon, M. Teo, C. Fu, W. Q. Leow, K. H. Lim, A. Chung, S.-L. Koo, S. P. Choo, D. Ho, S. Rozen, P. Tan, M. Wong, W. F. Burkholder and I. B. Tan (2017). Individualised multiplexed circulating tumour DNA assays for monitoring of tumour presence in patients after colorectal cancer surgery. *Sci Rep* 7, 40737.

NHS (2016). *Cervical Screening Programme*. Ed. by N. D. Screening & Immunisations Team. URL: https://digital.nhs.uk/data-and-information/publications/statistical/cervical-screening-programme/cervical-screening-programme-england-2015-2016.

NICE (2011). *Ovarian cancer: recognition and initial management (CG122)*. Clinical guideline. NICE.

Otsuka, J., T. Okuda, A. Sekizawa, S. Amemiya, H. Saito, T. Okai and M. Kushima (2004). Detection of p53 mutations in the plasma DNA of patients with ovarian cancer. *International Journal of Gynecological Cancer* 14.3, 459–464.

Parkinson, C. A., D. Gale, A. M. Piskorz, H. Biggs, C. Hodgkin, H. Addley, S. Freeman, P. Moyle, E. Sala, K. Sayal, K. Hosking, I. Gounaris, M. Jimenez-Linan, H. M. Earl, W. Qian, N. Rosenfeld and J. D. Brenton (2016). Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study. *PLoS Med* 13.12, e1002198.

Pedersen, J. W., A. Gentry-Maharaj, E. O. Fourkala, A. Dawnay, M. Burnell, A. Zaikin, A. E. Pedersen, I. Jacobs, U. Menon and H. H. Wandall (2013). Early detection of cancer in the

general population: a blinded case-control study of p53 autoantibodies in colorectal cancer. *Br J Cancer* 108.1, 107–14.

Perets, R., G. Wyant, K. Muto, J. Bijron, B. Poole, K. Chin, J. Chen, A. Ohman, C. Stepule, S. Kwak, A. Karst, M. Hirsch, S. Setlur, C. Crum, D. Dinulescu and R. Drapkin (2013). Transformation of the Fallopian Tube Secretory Epithelium Leads to High-Grade Serous Ovarian Cancer in Brca;Tp53;Pten Models. *Cancer Cell* 24.6, 751–765.

Phallen, J., M. Sausen, V. Adleff, A. Leal, C. Hruban, J. White, V. Anagnostou, J. Fiksel, S. Cristiano, E. Papp, S. Speir, T. Reinert, M. W. Orntoft, B. D. Woodward, D. Murphy, S. Parpart-Li, D. Riley, M. Nesselbush, N. Sengamalay, A. Georgiadis, Q. K. Li, M. R. Madsen, F. V. Mortensen, J. Huiskens, C. Punt, N. van Grieken, R. Fijneman, G. Meijer, H. Husain, R. B. Scharpf, J. Diaz L. A., S. Jones, S. Angiuoli, T. Orntoft, H. J. Nielsen, C. L. Andersen and V. E. Velculescu (2017). Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med* 9.403.

Piskorz, A. M., D. Ennis, G. Macintyre, T. E. Goranova, M. Eldridge, N. Segui-Gracia, M. Valganon, A. Hoyle, C. Orange, L. Moore, M. Jimenez-Linan, D. Millan, I. A. McNeish and J. D. Brenton (2016). Methanol-based fixation is superior to buffered formalin for next-generation sequencing of DNA from clinical cancer samples. *Ann Oncol* 27.3, 532–9.

Przybycin, C. G., R. J. Kurman, B. M. Ronnett, I.-M. Shih and R. Vang (2010). Are all pelvic (nonuterine) serous carcinomas of tubal origin? *The American journal of surgical pathology* 34 (10), 1407–1416.

RCOG (2016). *Ovarain Cysts in Post-menopausal Women (Green-top Guideline Mo. 34)*. Guideline. Royal College of Obstetricians and Gynaecologists.

Rosenfeld, N., T. Forshew, F. Marass and M. Murtaza (2016). 'A method for detecting a genetic variant'. 2016009224.

Stroun, M., P. Anker, P. Maurice, J. Lyautey, C. Lederrey and M. Beljanski (1989). Neoplastic characteristics of the DNA found in the plasma of cancer patients. *Oncology* 46.5, 318–22.

Sun, X., T. Huang, F. Cheng, K. Huang, M. Liu, W. He, M. Li, X. Zhang, M. Xu, S. Chen and L. Xia (2018). Monitoring colorectal cancer following surgery using plasma circulating tumor DNA. *Oncology letters* 15 (4), 4365–4375.

Swisher, E. M., M. Wollan, S. M. Mahtani, J. B. Willner, R. Garcia, B. A. Goff and M.-C. King (2005). Tumor-specific p53 sequences in blood and peritoneal fluid of women with epithelial ovarian cancer. *American Journal of Obstetrics and Gynecology* 193.3, 662–667.

Thorvaldsdottir, H., J. T. Robinson and J. P. Mesirov (2012). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings Bioinf* 14.2, 178–192.

Tsai-Turton, M., A. Santillan, D. Lu, R. E. Bristow, K. C. Chan, M. Shih Ie and R. B. Roden (2009). p53 autoantibodies, cytokine levels and ovarian carcinogenesis. *Gynecol Oncol* 114.1, 12–7.

Tsui, D. W. Y., M. Murtaza, A. S. C. Wong, O. M. Rueda, C. G. Smith, D. Chandrananda, R. A. Soo, H. L. Lim, B. C. Goh, C. Caldas, T. Forshew, D. Gale, W. Liu, J. Morris, F. Marass, T. Eisen, T. M. Chin and N. Rosenfeld (2018). Dynamics of multiple resistance mechanisms in plasma DNA during EGFR-targeted therapies in non-small cell lung cancer. *EMBO Molecular Medicine* 10.6, e7945.

Umetani, N., A. E. Giuliano, S. H. Hiramatsu, F. Amersi, T. Nakagawa, S. Martino and D. S. Hoon (2006). Prediction of breast tumor progression by integrity of free circulating DNA in serum. *J Clin Oncol* 24.26, 4270–6.

Underhill, H. R., J. O. Kitzman, S. Hellwig, N. C. Welker, R. Daza, D. N. Baker, K. M. Gligorich, R. C. Rostomily, M. P. Bronner and J. Shendure (2016). Fragment Length of Circulating Tumor DNA. *PLos Genet* 12.7, e1006162.

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* 17.4, 223–238.

Wang, Y., L. Li, C. Douville, J. D. Cohen, T. T. Yen, I. Kinde, K. Sundfelt, S. K. Kjaer, R. H. Hruban, I. M. Shih, T. L. Wang, R. J. Kurman, S. Springer, J. Ptak, M. Popoli, J. Schaefer, N. Silliman, L. Dobbyn, E. J. Tanner, A. Angarita, M. Lycke, K. Jochumsen, B. Afsari, L. Danilova, D. A. Levine, K. Jardon, X. Zeng, J. Arseneau, L. Fu, J. Diaz L. A., R. Karchin, C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. N. Fader, L. Gilbert and N. Papadopoulos (2018). Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers. *Sci Transl Med* 10.433.

Yang, W. L., A. Gentry-Maharaj, A. Simmons, A. Ryan, E. O. Fourkala, Z. Lu, K. A. Baggerly, Y. Zhao, K. H. Lu, D. Bowtell, I. Jacobs, S. J. Skates, W. W. He, U. Menon and J. Bast R. C. (2017). Elevation of TP53 Autoantibody Before CA125 in Preclinical Invasive Epithelial Ovarian Cancer. *Clin Cancer Res* 23.19, 5912–5922.

Yu, S. C., K. C. Chan, Y. W. Zheng, P. Jiang, G. J. Liao, H. Sun, R. Akolekar, T. Y. Leung, A. T. Go, J. M. van Vugt, R. Minekawa, C. B. Oudejans, K. H. Nicolaides, R. W. Chiu and Y. M. Lo (2014). Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc Natl Acad Sci U S A* 111.23, 8583–8.

Zhu, C. S., P. F. Pinsky, D. W. Cramer, D. F. Ransohoff, P. Hartge, R. M. Pfeiffer, N. Urban, G. Mor, J. Bast R. C., L. E. Moore, A. E. Lokshin, M. W. McIntosh, S. J. Skates, A. Vitonis, Z. Zhang, D. C. Ward, J. T. Symanowski, A. Lomakin, E. T. Fung, P. M. Sluss, N. Scholler, K. H. Lu, A. M. Marrangoni, C. Patriotis, S. Srivastava, S. S. Buys and C. D. Berg (2011). A framework for evaluating biomarkers for early detection: validation of biomarker panels for ovarian cancer. *Cancer Prev Res (Phila)* 4.3, 375–83.

Appendices

12.1 CTCR-OV05 study documents

12.1.1 Protocol

Version 2 December 2014

Study Title: Genetic Biomarkers for Gynaecological Conditions

Ethics Ref:

Date and Version No: Version 2 December 2014

Chief Investigator: Elizabeth Moore
CRUK Cambridge Institute
Li Ka Shing Centre
Robinson Way
Cambridge
CB2 0EE

Investigators: Dr James Brenton
Dr Nitzan Rosenfeld
Mr Robin Crawford
Dr Paul Pharoah

Sponsor: Cambridge University Hospitals NHS Foundation Trust and The University of Cambridge

AMENDMENT HISTORY

Amendment No.	Protocol Version No.	Date issued	Author(s) of changes	Details of Changes made

1. SYNOPSIS

Study Title	**Genetic Biomarkers for Gynaecological Conditions**
Internal ref. no.	CTCR-OV05
Study Design	Pilot/feasibility study
Study Participants	All women referred to Addenbrooke's Hospital with suspected ovarian cancer. All women undergoing surgery for ovarian or endometrial cancer.
Planned Sample Size	230
Follow-up duration	Five years (routine follow up for ovarian cancer) or until the patient is discharged from clinic in cases of relapsed ovarian cancer.
Planned Study Period	3 years
Primary Objective	To determine if somatic mutations can be detected in blood samples obtained from women with ovarian cancer at the time of diagnosis?
Secondary Objectives	To determine if mutations detected in plasma samples at diagnosis can be detected in physician taken and self-taken cervical samples? To determine if changes in the levels of somatic mutations in blood samples during and after surgery can be used as a marker of post-operative residual disease?
Primary Endpoint	Detection and frequency of somatic mutations in blood samples.

2. BACKGROUND AND RATIONALE

OC is the most common cause of death from gynaecologically related cancers[1]. The majority of women (60%) are diagnosed with late (stage III-IV) disease which has a significant impact on survival with 5-year relative survival rates of 6%, 22% and 92% for stage IV, stage II and stage I disease respectively[1]. Difficulties in the diagnosis of OC are related to the non-specific symptoms of the disease resulting in delayed investigations, a significant proportion (>50%) of early stage disease without an elevated CA125 resulting in missed diagnoses[2] and a significant number of benign gynaecological and non-gynaecological conditions, particularly in premenopausal women, with an elevated CA125 leading to inappropriate investigation and referral.

Current guidelines recommend measuring serum CA125 in all women with suspected OC followed by a transvaginal ultrasound scan (TVUS) in those with elevated levels[3]. In combination these are used to calculate a Risk of Malignancy Index (RMI), which guides referral and subsequent management[4].

Despite decades of effort, CA125 remains the single-best biomarker for OC with models derived from the most promising markers failing to show improvement over serum CA125 alone[5,6]. There is therefore a pressing need to explore complementary strategies, prominent among which is the simultaneous detection of circulating tumour DNA (ctDNA). It is known that blood contains free circulating DNA and in cancer patients a small fraction of this includes fragments of tumour DNA. Somatic mutations are highly specific biomarkers of cancer and detection of these via ctDNA in plasma samples this could provide a powerful non-invasive method for OC diagnosis. The gene *TP53* encodes the tumour suppressor protein p53, a transcription factor that regulates the expression of proteins involved in apoptosis and genomic integrity. The Brenton laboratory have identified *TP53* mutations in at least 96% of HGSOC tissue samples[7]. The Rosenfeld and Brenton laboratories have identified mutations in *TP53* in ctDNA from plasma samples from 20/38 (53%) of patients with advanced HGSOC[8].

As mutations in *TP53* are pathognomonic of HGSOC, assays for mutant ctDNA could be used, in combination with other variables in the RMI score, as a non-invasive method to improve specificity of OC diagnosis especially early stage or low volume disease. We therefore propose to collect blood samples from all women referred to Addenbrooke's Hospital with suspected ovarian cancer and sequence DNA for somatic mutations which will be correlated with clinical diagnosis.

It is currently hypothesized that OC originates in the fallopian tube epithelium. Molecular alterations specific to OC have been identified in tubal tissue. It is therefore possible that detection of cells and DNA shed from this site could be utilised in the diagnosis of OC. Recent papers have shown that tumour DNA and common somatic mutations can be detected in liquid based cytology (LBC) samples[9] and cervical mucous samples[10] from patients with both endometrial and ovarian cancers. We propose to collect cervical samples from all women referred to Addenbrooke's Hospital with suspected ovarian cancer and sequence DNA for somatic mutations which will be correlated with clinical diagnosis.

ctDNA has a potential role as a biomarker for monitoring tumour dynamics and response to treatment. Recently it has been demonstrated that that there is a significant correlation between ctDNA levels and tumour burden in 30 women with metastatic breast cancer[11] and that whole exome sequencing of ctDNA can be used to track genomic evolution of cancers in response to systemic chemotherapy[12]. Post-operative levels of circulating tumour DNA in plasma samples from patients with colorectal cancer appear to be related to the degree of resection with patients undergoing more extensive resection having lower post-operative levels[13]. Currently there is no objective measure of post-operative residual disease in OC however, the degree of surgical resection is a strong prognostic factor. We propose to collect normal and abnormal tissue, plasma and cervical samples from women undergoing surgery for ovarian and endometrial cancer to compare mutation status in pre and post-operative samples to investigate the kinetics of ctDNA in response to surgery.

3. OBJECTIVES

3.1 Primary Objective

Can somatic mutations be detected in blood samples obtained from women with ovarian cancer at the time of diagnosis?

3.2 Secondary Objectives

Can mutations detected in plasma samples at diagnosis be detected in physician taken and self taken cervical samples?

Can changes in the levels of somatic mutations in blood samples during and after surgery be used as a marker of marker of post-operative residual disease?

4. STUDY DESIGN

4.1 Summary of Study Design

Feasibility/pilot study

1. Patients referred to Addenbrooke's Hospital with suspected gynaecological cancers will be identified at the time of receipt of the referral letter. They will be sent a cover letter briefly outlining this clinical study with their routine out-patient clinic appointment letter by the clinical care team.

2. At the time of routine outpatient clinic appointment with the gynaecological oncology team patients will be asked by the clinician if they would like to discuss the study further with a member of the research team. This will be documented in the medical notes.

3.The principal investigator or suitably qualified, delegated member of the trial/research/clinical care team will be available at the time of outpatient clinic appointment to discuss the study with the patient and talk through the patient information sheet and answer any questions.

4. Written consent will be taken by the principal investigator or suitably qualified, delegated member of the trial/research/clinical care team from those patients who wish to participate. The principal investigator is an obstetrics and gynaecology trainee but is not a member of the clinical team that will be investigating and treating the patient. The principle investigator has attended good clinical practice and valid informed consent training courses. If appropriate written consent will be taken on the same day as the outpatient clinic. The principal investigator or suitably qualified, delegated member of the trial/research/clinical care team will ensure that the patient has had sufficient time to consider participation, fully understands the study and has had an opportunity to discuss participation with family and friends if required. Some patients will require further time to consider participation in the study therefore for those patients returning for further investigations and/or surgery patients will be given the option of further discussion about the study and/or consenting and collection of samples at future attendance at Addenbrooke's Hospital. Those patients discharged from Addenbrooke's Hospital that require further time to consider participation in the study will be offered the option of returning at a time appropriate for them with reimbursement of travel expenses.

5. Following the consent process the principal investigator or suitably qualified, delegated member of the trial/research/clinical care team will take a blood sample (up to 30ml, approximately 6 teaspoons) and cervical samples from women that have consented to participate. If patients require routine blood tests these will be taken at the same time to minimise risks of discomfort and bruising. A chaperone will be present at the time of collection of the cervical samples.

6. Patients that are required to undergoing surgery for suspected/confirmed ovarian or endometrial cancer will be identified via the MDT.

7. Patients will be seen by the principal investigator or suitably qualified, delegated member of the trial/research/clinical care team at the time of pre-operative assessment clinic. The second part of the study will be discussed with the patient with the second patient information sheet and any questions will be answered.

8. Written consent will be taken by the principal investigator or suitably qualified, delegated member of the trial/research/clinical care team from those patients who wish to participate.

9.Patients will be asked to collect and bring a urine sample with them on the day of surgery.

10. At the time of surgery all patients will routinely have a cannula inserted for intravenous access prior to the start of surgery. A blood sample (up to 30ml, approximately 6 teaspoons) will be taken from this during the anaesthetic before surgery commences, at the end of

surgery and 2 hours after surgery has been completed by the principal investigator or suitably qualified, delegated member of the trial/research/clinical care team.

11. Cervical samples will be taken from the patient following induction of anaesthesia prior to the commencement of surgery by the principal investigator or suitably qualified, delegated member of the trial/research/clinical care team.

12. Patients would normally have as much cancer as possible removed at the time of surgery. Samples of both normal and abnormal tissue surplus to diagnosis will be collected by the operating surgeon.

13. Further blood samples (up to 30ml, approximately 6 teaspoons) will be taken daily whilst the patient is an inpatient in hospital by the principal investigator or suitably qualified, delegated member of the trial/research/clinical care team. These will be taken at the time of routine blood tests where possible.

14. If patients have had previous gynaecological surgery, or proceed to further gynaecological surgery in the future, any stored surplus tissue samples in the Addenbrooke's tissue bank will be requested for comparison to collected blood and cervical, and tissue where applicable, samples collected during the study.

4.2 Primary and Secondary Endpoints/Outcome Measures

Primary Endpoints

-Detection and frequency of somatic mutations in blood samples.

Secondary Endpoints

-Detection and frequency of somatic mutations in cervical samples.

-Correlation of change in frequency of somatic mutations with surgical outcome and clinical response.

4.3 Study Participants

4.3.1 Overall Description of Study Participants

- Women referred to Addenbrooke's Hospital with suspected ovarian cancer.
- Women undergoing surgery for suspected/confirmed ovarian or endometrial cancer.

4.3.2 Inclusion Criteria

- Participant is willing and able to give informed consent for participation in the study.
- Female, aged 18 years or above.
- Referred with suspected ovarian cancer or undergoing surgery for suspected/confirmed ovarian or endometrial cancer.

4.3.3 Exclusion Criteria

The participant may not enter the study if ANY of the following apply:

- Age under 18 years.

4.4 Study Procedures

All participants will have a blood sample up to 30ml (approximately 6 teaspoons) taken at the time of referral for suspected ovarian cancer. At the same time cervical samples will be collected during a speculum examination.

Patients will be asked to collect and bring a urine sample on the day of surgery. At the time of surgery further cervical samples will be collected. Blood samples up to 30ml (approximately 6 teaspoons) will be collected at the start of surgery, at the end of surgery, 2 hours post operatively and daily whilst the patient remains in hospital.

4.4.1 Informed Consent

Patients will be sent a cover letter with the routine clinic appointment letter detailing that a clinical study is ongoing, that they may be approached for participation in the study, the purpose of the study and that the study will involve the collection of blood and cervical samples. The contact details of the research nurse are provided in case the patients requires any further information about the study prior to attending clinic.

At the time of routine outpatient clinic appointment with the gynaecological oncology team patients will be asked by the clinician if they would like to discuss the study further with a member of the research team who will be available at the time of the clinic appointment. Verbal and written information will be available from the principal investigator or suitably

qualified, delegated member of the clinical care team at the time of the outpatient appointment. All patients will be informed of the aims of the study. They will be informed as to the strict confidentiality of their patient data, but that their medical records may be reviewed for research purposes by authorized individuals other than their treating physician. It will be emphasized that the participation is voluntary and that the patient is allowed to refuse further participation in the protocol whenever he/she wants. This will not prejudice the patient's subsequent care. Participants will be given the opportunity to ask any questions and only once all of their queries have been answered will consent be taken by the principal investigator or suitably qualified, delegated member of the clinical care team from those patients willing to participate.

Patients will be given as much time as is required to consider participation in the study. In some cases it will be appropriate to take consent and collect samples on the same day as the initial clinic appointment. This approach as been discussed with patient advocates through Target Ovarian Cancer and has been deemed appropriate if patients feel they have had sufficient time to consider participation. However, in some case this will not be appropriate and will therefore be deferred until further attendance at Addenbrookes Hospital for those patients undergoing further investigation and/or surgery. For those patients that are not required to return for clinical reasons we will provide reimbursement for travel expenses for those patients wishing to return at a later date for consenting to the research study and collection of samples. Documented informed consent will be obtained for all patients included in the study before they are registered in the study. This will be done in accordance with the national and local regulatory requirements.

4.4.2 Study Assessments

Clinical data will be collected and stored on NHS computers. DNA will be extracted and sequenced from all clinical samples. Sequencing data will be correlated with linked anonymised clinical data, including CA125, imaging findings, diagnosis, histology, response to treatment and relapse, at CRUK Cambridge Institute.

4.5 Definition of End of Study

The end of study is the date of the last visit of the last participant.

5. INTERVENTIONS

See study procedures.

6. SAFETY REPORTING

Any serious adverse events (an adverse event that results in death, is life-threatening, requires hospitalisation or prolongation of existing hospitalisation, results in persistent or significant disability or incapacity) that occur during the study will be reported to the sponsors.

7. STATISITICS

7.1 The Number of Participants

To establish if sensitivity differs significantly from 66.7% but to be 80% certain with a probability of 0.05 that a sensitivity as low as 56.1% would be excluded requires 161 subjects with ovarian cancer.

To establish if specificity differs significantly from 90.9% but to be 80% certain with a probability of 0.05 that a specificity as low as 80.0% would be excluded requires 69 subjects without ovarian cancer.

7.2 Analysis of Endpoints

Comparison of the presence and frequency of somatic mutations across clinical samples for each participant. Correlation with clinical diagnosis.

Comparison of mutation status in pre and post operative clinical samples. Correlation with extent of surgical resection and survival.

8. ETHICS

8.1 Participant Confidentiality

All clinical information regarding participants collected during the course of the study will be kept strictly confidential. Each participant will be allocated a unique study number at study entry and will be identified by this number on all study related documentation throughout the course of the trial and data analysis process. This will be kept on an individual computer, access restricted to named individuals and password protection. All the data will be held on an NHS computer and will conform to the Department of Data Security Policy and Department Compliance with the Data Protection Act (1998). Anonymised data may be sent to the laboratory personnel University of Cambridge Computers at CRUK Cambridge Institute, but this will not contain any personal data. Any data transferred will be done according to the NHS Code of Practice on Confidentiality.

8.2 Other Ethical Considerations

This study involves the collection of clinical samples from patients at the time of referral and primary surgery for suspected ovarian cancer.

Patients with suspected ovarian cancer are referred to the hospital as a two week wait referral meaning they are seen in out-patient clinic within two weeks of the referral being received. Not all patients will be aware that they have been referred for suspected ovarian cancer. For this reason ovarian cancer has specifically not been mentioned on any of the documentation (ie. cover letter, patient information sheet, consent form) that potential participants will receive. All patients will be seen, as per routine clinical care, by an experienced oncologist prior to being asked if they would be interested in participating in the research study.

It is unlikely that this study will produce findings of clinical significance for the participants or their relatives. In the unlikely event that clinically significant information does become available this will be referred to the patients treating clinician. In the unlikely event that the

study reveals information that might affect family members this will also be referred to the patients treating clinician. Genetic testing is not currently undertaken routinely in clinical practice. If clinically relevant and the patient consents to undergo genetic testing this is done following appropriate genetic counselling. We therefore feel it is important to know, although this is unlikely, whether the participant would want to be informed if the research reveals any genetic information that might affect family members. We have therefore included this as an optional point in the consent form. If participants would like to be informed of any information that might affect family members this information will be provided to participants after appropriate genetic counselling arranged by the treating clinician. This optional question in the consent form is included in an ethically approved study being undertaken at Addenbrooke's Hospital. Importantly approximately 10% of participants have not ticked this box indicating that they would not wish to be informed of this information should it become available.

DATA HANDLING AND RECORD KEEPING

Patient initials, date of birth and hospital number, NHS number, relevant medical history, disease pathology, treatments and treatment outcomes will be stored in a database on NHS computers. Patients will consent for this data to be stored on their consent forms. This data will be kept on an individual computer, access restricted to named individuals and password protection. Each participant will be allocated a unique study number at study entry and all biological samples and study documentation will be identified by this number throughout the course of the trial and data analysis process.

9. FINANCING AND INSURANCE

The study is being funded by the Medical Research Council and Target Ovarian Cancer. The study will be sponsored jointly by the Cambridge University Hospitals NHS Foundation Trust and The University of Cambridge.

10. ABREVIATIONS

CRUK – cancer research UK

ctDNA – circulating tumour DNA

HGSOC – high grade serous ovarian cancer

LBC – liquid based cytology

MDT – multidisciplinary team

OC – ovarian cancer

RMI – risk of malignany index

TVUS – transvaginal ultrasound scan

11. REFERENCES

1. CancerStats. Cancer Research UK.
http://info.cancerresearchuk.org/cancerstats/types/ovary/ Accessed 01/07/2013

2. Woolas RP, Xu FJ, Jacobs IJ, Yu YH, Daly L, Berchuck A, Soper JT, Clarke- Pearson DL,

Oram DH, Bast RC. 1993 Elevation of multiple serum markers in patients with stage I ovarian cancer. J Natl Cancer Inst:85;1748-51.

3. NICE Guidance CG122. Ovarian Cancer: the recognition and initial management of ovarian cancer. April 2011. http://guidance.nice.org.uk/CG122/Guidance/pdf/English. Accessed 01/07/2013

4. Rufford BD, Jacobs IJ. 2003. Ovarian Cysts in Postmenopausal Women. RCOG Greentop Guideline Number 34.

5. Cramer DW, Bast RC, Berg CD, Diamandis EP, Godwin AK, Hartge P, Lokshin AE, Lu KH, McIntosh MW, Mor G, Patriotis C, Pinsky PF, Thornquist MD, Scholler N, Skates SJ, Sluss PM, Srivastava S, Ward DC, Zhang Z, Zhu CS, Urban N. 2011. Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. Cancer Prev Res:4;365-74.

6. Zhu CS, Pinsky PF, Cramer DW, Ransohoff DF, Hartge P, Pfeiffer RM, Urban N, Mor G, Bast RC Jr, Moore LE, Lokshin AE, McIntosh MW, Skates SJ, Vitonis A, Zhang Z, Ward DC, Symanowski JT, Lomakin A, Fung ET, Sluss PM, Scholler N, Lu KH, Marrangoni AM, Patriotis C, Srivastava S, Buys SS, Berg CD; PLCO Project Team. 2011. A framework for evaluating biomarkers for early detection: validation of biomarker panels for ovarian cancer. Cancer Prev Res:4;375-83.

7. Ahmed AA, Etemadmoghadam D, Temple J, Lynch AG, Riad, M, Sharma R, Stewart C, Fereday S, Caldas C, Defazio A, Bowtell D, Brenton JD. 2010. Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. J Pathol:221;49-56.

8. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, Brenton JD, Rosenfeld N. 2012. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med:4:136-68.

9. Kinde I, Bettegowda C, Wang Y, Wu J, Agrawal N, Shih IeM, Kurman R, Dao F, Levine DA, Giuntoli R, Roden R, Eshleman JR, Carvalho JP, Marie SK, Papadopoulos N, Kinzler KW, Vogelstein B, Diaz LA Jr. Evaluation of DNA from the Papanicolaou Test to Detect Ovarian and Endometrial Cancers. Sci Transl Med 5. 2013 9;5(167):167ra4.

10. Lamzabi I, Buckingham L, Gebrekiristos M, Jain R, Gattuso P, Reddy V, Guirguis A, Dewdney S, Rotmensch J, Bitterman P. 2013. Use of cervical mucous to screen for gynecological malignancies: a pilot study. Mod Pathol: 2013 Jun 14. doi: 10.1038/modpathol.2013.92. [Epub ahead of print]

11. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, Dunning MJ, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C, Rosenfeld N. 2013. Analysis of circulating tumour DNA to monitor metastatic breast cancer. N Engl J Med:368;1199-209.

12. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, Parkinson C, Chin SF, Kingsbury Z, Wong AS, Marass F, Humphray S, Hadfield J, Bentley D, Chin TM, Brenton

JD, Caldas C, Rosenfeld N. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature. 2013 Apr 7. doi: 10.1038/nature12065. [Epub ahead of print]

13. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, Thornton K, Agrawal N, Sokoll L, Szabo SA, Kinzler KW, Vogelstein B, Diaz LA. 2008. Circulating mutant DNA to assess tumor dynamics. Nat Med:14;985-90.

12.1.2 Patient information sheet — referral

CTCR-OV05: Patient Information Sheet

Referral **Version 2: December 2014**

Genetic Biomarkers for Gynaecological Conditions

We would like to invite you to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully and discuss it with friends and relatives if you wish.

- *Part 1* tells you the purpose of this study and what will happen if you take part
- *Part 2* gives you more detailed information about the conduct of the study

Please ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for taking the time to read this information.

Part 1

What is the purpose of the study?
The purpose of this study is to identify new ways of diagnosing ovarian conditions. You may have already had a blood test and an ultrasound scan arranged by your GP and although they give us a lot of information about possible ovarian conditions sometimes it would be useful to have more information to help decide what is the best course of management and treatment for women. We are hoping that by collecting information from blood tests and cervical samples (not currently used in the diagnosis of ovarian conditions) we will be able improve the investigation of women referred by their GP with possible ovarian conditions. We are hoping to be able to identify ovarian conditions at the earliest time point possible to ensure that women receive the best treatment possible.

Why have I been invited?
You have been invited as you have been referred to see a gynaecologist as your GP suspects that you may have an ovarian condition. We would like to collect blood and cervical samples from all women referred to Addenbrooke's Hospital with suspected ovarian conditions. These samples will all be taken at the same time as your routine outpatient clinic appointment.

Do I have to take part?
No. Taking part in this research study is entirely voluntary and it is up to you to decide whether or not to take part. We will describe the study to you and

go through this information sheet, which we will then give to you. We will then ask you to sign a consent form to show you have agreed to take part. You are free to withdraw at any time, without having to give a reason. A decision not to take part or a decision to withdraw at any time will not affect the standard of care that you receive.

What will happen to me if I take part? What will I have to do?
If you agree to take part in this study we would like to ask if we can take one additional blood sample of up to 30ml of blood (approximately 6 teaspoons). This will be taken at the same time as your routine blood tests if possible. We would also like to ask if we can collect some additional cervical samples including samples of cells and mucous from your cervix. The cervical samples will be collected in the same way as a routine cervical smear test.

With your permission we would like to inform your GP that you are participating in this study.

We would like to ask if we can obtain further information about your medical history and progress from your medical records and would like to ask permission to obtain this information from your GP (located via NHS databases) if required. If relevant we would also like to ask permission to update our records using the Eastern Cancer Registration and Information Centre (ECRIC). In order to do this, we will need to use your full name and/or NHS number.

If you have already had surgery for a gynaecological condition, or will be having surgery in the future, there may be stored surplus tissue samples of yours in Addenbrooke's tissue bank (or your local hospital histopathology department). We would like to be able to study these as well for our research.

If you agree to participate in this study you will be asked to sign a consent form. This information would remain strictly confidential ie. all information that could identify you would be removed, before being given to the researchers. Under no circumstances will you be identified in any way in any publication or report arising from this study. Further information is given in Part 2 of this information sheet.

What are the possible disadvantages and risks of taking part?
Risks of blood tests include possible minor discomfort, and occasional bruising or bleeding. Additional blood tests could increase these risks slightly but all blood tests will be taken at the same time as routine blood tests where possible to minimise these risks. The procedure is done under sterile conditions by experienced doctors or nurses.

The cervical samples will be collected in the same way as a routine cervical smear test so may be associated with some minor discomfort but should be painless. Some women can experience a small amount of bleeding (spotting) following collection of cervical samples. The samples will be collected by an

experienced doctor and a chaperone will be present whilst these samples are taken.

What are the possible benefits of taking part?
There are no immediate clinical benefits to you taking part in this study but the information we obtain could be used to improve the investigation of women with possible ovarian conditions in the future.

Expenses and payments
If you choose to come back to Addenbrooke's Hospital for a visit purely related to this research study ie. to consent to participate in the study or for us to collect samples for the research study, your travel expenses will be reimbursed.

Will my taking part in the study be kept confidential?
Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. Full details on this (and details of how any complaints about the way you have been dealt with will be addressed) are included in Part 2 of this information sheet.

This completes Part 1. If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.

Part 2

Will my information be kept confidential?
Yes. The study involves us collecting blood and cervical samples and obtaining information about your relevant medical history and the outcomes of your investigations from your medical records. This information will be collected by a clinical fellow and/or research nurse and will be stored securely and is kept strictly confidential. You will be registered on the study by your NHS number, hospital number, initials and date of birth and assigned a unique code number. Other information on the NHS site will be stored in a separate database and refer to you only by your code number.

Samples will be taken to the Cancer Research UK (CRUK) Cambridge Institute for processing. So that we can link all of your samples, your hospital number, NHS number and date of birth will be securely encoded and linked to the unique sample numbers. At no point will any un-encoded information that could identify you be linked to your samples on the CRUK Cambridge Institute database.

In order retrieve any surplus stored tissue samples from Addenbrooke's Tissue Bank or other hospital histopathology departments we will need to use your full name, and/or NHS number/hospital number and date of birth, however only encoded information will be linked to your actual samples, as above.

All databases are separately password-protected, as are the computers. This information will be stored in accordance with the Data Protection Act of 1998 and is only accessible by authorised personnel. Occasionally we may need to access your medical records in order to update the information. This will be done by clinical staff or designated personnel from the trial office. The sponsor and/or members of the NHS Trust Research and Development Department may require access for audit and monitoring purposes.

With your agreement, we may ask your GP and any other doctors involved in your care to provide information on your progress. If we do need to contact your GP or another hospital (if your main care is not at Addenbrooke's) for any follow-up information, we will need to use your full name in our correspondence if you have agreed to provide this. If you change address and we lose touch with your GP, we may then locate your new GP through the NHS database.

If relevant we would also like to ask for permission to update our records using the Eastern Cancer Registration and Information Centre (ECRIC). In order to do this, we will need to use your full name and/or NHS number.

Under no circumstances will you be identified in any way in any report or publication arising from the study.

What will happen to the samples I give?
All samples will be coded ('anonymised') which means that the laboratory researchers who are carrying out tests on your tumour samples will not be able to trace them back to you. However, the trials office will keep a confidential record of the codes used for your samples and can access your notes should the need arise.

The samples will then be sent for analysis at the CRUK Cambridge Institute on the Addenbrooke's Hospital Site. We will be studying DNA in the blood and cervical samples to look at molecular markers of ovarian conditions that may be useful for diagnosis. Samples will be stored at the CRUK Cambridge Institute and if you agree these may be used in future research that has received full ethics committee approval.

It is unlikely that any new information will be generated from this study that will influence your treatment or reveal genetic information that might affect family members. However, if any relevant information does become available during the course of the study this will be given to the doctor looking after you in the hospital. If you wish to be informed of this information your doctor will arrange this with the appropriate genetic counselling.

What will happen if I don't want to carry on with the study?
You are free to withdraw from the study at any time without giving a reason. A decision not to take part or a decision to withdraw at any time will not affect the standard of care that you receive. Please note that if you do choose to withdraw from the study, we would still like to keep stored blood or tissue for future research purposes. However, should you not wish us to do this, you

can withdraw your consent and you should tell your doctor or nurses of your wishes. Any stored blood or tissue samples that can still be identified as yours will be destroyed if you wish.

What if there is a problem?
We will try to address any concerns you might have about any aspect of this study. If there are any problems, you should ask to speak with the researchers who will do their best to answer your questions (Telephone: Charlotte Hodgkin, Senior Research Nurse, 01223 348084). If you remain unhappy and wish to complain formally, you can do this through the NHS Complaints Procedure. Details on this can be obtained by contacting the Hospital PALS (Patient Advice and Liaison Service) on 01223 216756.

In the unlikely even that something goes wrong and you are harmed as a result of participating in this study, compensation will be available through the NHS Indemnity scheme.

What will happen to the results of the study?
The results of this study will be published in scientific or medical journals and may be presented at scientific or medical meetings. Please be assured it will not be possible to identify you in any report or publication. If you would like to obtain a copy of the published results please contact the research nurse directly who will be able to arrange this for you.

Who is organising (sponsoring) and funding the study?
This study is sponsored by Cambridge University Hospitals NHS Foundation Trust and The University of Cambridge. The study is being funded by the Medical Research Council and Target Ovarian Cancer.

Who has reviewed this study?
All research within the NHS is reviewed by an independent group of people called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by NRES Committee East of England - Cambridge East.

Further information and contact details
If you require any further information about this study please do not hesitate to contact the Research Nurse below. Further advice on research and conduct of studies can be obtained from Addenbrooke's Hospital PALS on 01223 216756.

The Senior Research Nurse is:
Charlotte Hodgkin
Cambridge Trials Centre,
S4 (Box 279)
Addenbrooke's Hospital, Hills Road, Cambridge CB2 OQQ
Telephone: 01223 348084

Thank you for taking the time to read this Patient information Sheet.

12.1.3 Consent form — referral

CTCR-OV05

Patient Consent Form 1

Referral

Version 2: December 2014

Consent for CTCR-OV05

Please initial boxes

1. I confirm that I have read the information sheet dated ….. (version…..) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that relevant sections of my medical notes and data collected during the study may be looked at by responsible individuals from the Cambridge Trials Unit, from regulatory authorities or from the NHS trust where it is relevant to my taking part in research. I give permission for these individuals to have access to my records. ☐

4. I understand that the information collected about me will be used to support other research in the future, and may be shared anonymously with other researchers. ☐

5. I understand that my participation in this study will involve the collection of personal details, information about my treatment and medical history. I understand that all data collected about me will be held under the provisions of the 1998 Data Protection Act and stored in manual and electronic files in secure encoded format. ☐

6. I give permission for my GP to be notified of my participation in this research. If necessary, I give permission for information about my progress to be obtained from my GP (located via NHS databases if required), and if relevant through the Eastern Cancer Registration and Information Centre and understand that this will be done using my full name and/or NHS/hospital number. ☐

7. I agree to the identification and analysis of previously collected tissue, or tissue that will be collected in the future, and understand that this will be done using my full name and/or NHS/hospital number. ☐

8. I agree to the collection and storage of additional blood, urine and cervical samples for future research into ovarian conditions, including DNA analysis. ☐

9. I agree to the linkage of securely encoded encrypted personal details from samples collected on this study, and previous stored samples. ☐

10. I agree to take part in the above study ☐

---

11. OPTIONAL: I wish to be informed if research on my blood reveals genetic information that might affect members of my family (initial box only if you agree) ☐

Patient

Signed ……

Print name …… Date: ……

Doctor/ Person taking consent

Signed ……

Print name …… Date: ……

copies required: **1 copy for researcher; 1 copy for patient; 1 copy to be kept with patient's hospital notes**

12.1.4 Patient information sheet — surgery

OV05: Patient Information Sheet

Surgery **Version 2:December 2014**

Genetic Biomarkers for Gynaecological Conditions

We would like to invite you to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully and discuss it with friends and relatives if you wish.

- *Part 1* tells you the purpose of this study and what will happen if you take part
- *Part 2* gives you more detailed information about the conduct of the study

Please ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for taking the time to read this information.

Part 1

What is the purpose of the study?
This study has two purposes. The first is to identify new ways of diagnosing gynaecological conditions. You are undergoing surgery for a possible gynaecological condition and we are hoping that by collecting information from blood tests and urine and cervical samples we will be able improve the investigation and treatment of women referred to the hospital with possible gynaecological conditions. The second is to see if we can develop a method of telling how successful surgery has been and whether any further treatment will be required.

Why have I been invited?

You have been invited as you are undergoing surgery for a possible gynaecological condition. We would like to collect additional tissue, blood, urine and cervical samples from around the time of the operation from women undergoing surgery at Addenbrooke's Hospital for possible gynaecological conditions. These samples will be taken during the general anaesthetic or at the same time as your routine care where possible.

Do I have to take part?
No. Taking part in this research study is entirely voluntary and it is up to you to decide whether or not to take part. We will describe the study to you and go through this information sheet, which we will then give to you. We will then ask you to sign a consent form to show you have agreed to take part. You

are free to withdraw at any time, without having to give a reason. A decision not to take part or a decision to withdraw at any time will not affect the standard of care that you receive.

What will happen to me if I take part? What will I have to do?
If you agree to take part in this study, you will not need to attend the hospital for any extra visits. Firstly we would like to ask you to bring an additional urine sample with you on the day of surgery.

Normally at surgery if any abnormal tissue is identified the surgeon will try and remove as much of it as possible. They may take some small samples from normal looking areas in the abdomen for comparison. These samples will be sent to the pathology laboratory. With your permission, we would like to take some spare samples of normal and abnormal looking tissue, if present, for our research.

We would like to ask if we can take some additional blood samples from around the time of your operation and after your operation whilst you are still in hospital. We would be asking to take three blood samples of up to 30ml of blood (approximately 6 teaspoons) during your operation and a further blood sample of up to 30ml of blood (approximately 6 teaspoons) each day whilst you remain in hospital after your operation. These will be taken at the same time as your routine blood tests where possible.

We would also like to ask if we can collect some additional cervical samples including samples of cells and mucous from your cervix at the time of your operation. Your operation will be done under a general anaesthetic, this means you will be asleep during the operation. The cervical samples will be collected during this time in the same way as a routine cervical smear test.

With your permission we would like to inform your GP that you are participating in this study.

We would like to ask if we can obtain further information about your medical history and progress from your medical records and would like to ask permission to obtain this information from your GP (located via NHS databases) if required. If relevant we would also like to ask permission to update our records using the Eastern Cancer Registration and Information Centre (ECRIC). In order to do this, we will need to use your full name and/or NHS number.

If you have already had surgery for a gynaecological condition there may be stored surplus tissue samples of yours in Addenbrooke's tissue bank (or your local hospital histopathology department). We would like to be able to study these as well for our research.

If you agree to participate in this study you will be asked to sign a consent form. This information would remain strictly confidential ie. all information that could identify you would be removed, before being given to the researchers. Under no circumstances will you be identified in any way in any publication or

report arising from this study. Further information is given in Part 2 of this information sheet.

What are the possible disadvantages and risks of taking part?
Risks of blood tests include possible minor discomfort, and occasional bruising or bleeding. Additional blood tests could increase these risks slightly but all blood tests will be taken at the same time as routine blood tests where possible to minimise these risks. The procedures are done under sterile conditions by experienced doctors or nurses.

The cervical samples will be collected in the same way as a routine cervical smear test during the general anaesthetic so will not be associated with any increased pain or discomfort. Collecting cervical samples will not affect the surgery in any way.

For patients having additional tissue samples collected during surgery, there will be no significant additional risks to the surgery, and it will not affect the surgery in any way.

What are the possible benefits of taking part?
There are no immediate clinical benefits to you taking part in this study but the information we obtain could be used to improve the investigation and treatment of women with possible gynaecological conditions in the future.

Expenses and payments
There are no expenses or payments associated with this study.

Will my taking part in the study be kept confidential?
Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. Full details on this (and details of how any complaints about the way you have been dealt with will be addressed) are included in Part 2 of this information sheet.

This completes Part 1. If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.

Part 2

Will my information be kept confidential?
Yes. The study involves us collecting blood, urine, cervical and tissue samples and obtaining information about your relevant medical history and the outcomes of your treatment from your medical records. This information will be collected by a clinical fellow and/or research nurse and will be stored securely and is kept strictly confidential. You will be registered on the study by your NHS number, hospital number, initials and date of birth and assigned a unique code number. Other information on the NHS site will be stored in a separate database and refer to you only by your code number.

Samples will be taken to the Cancer Research UK (CRUK) Cambridge Institute for processing. So that we can link all of your samples, your hospital number, NHS number and date of birth will be securely encoded and linked to the unique sample numbers. At no point will any un-encoded information that could identify you be linked to your samples on the CRUK Cambridge Institute database.

In order retrieve any surplus stored tissue samples from Addenbrooke's Tissue Bank or other hospital histopathology departments we will need to use your full name, and/or NHS number/hospital number and date of birth, however only encoded information will be linked to your actual samples, as above.

All databases are separately password-protected, as are the computers. This information will be stored in accordance with the Data Protection Act of 1998 and is only accessible by authorised personnel. Occasionally we may need to access your medical records in order to update the information. This will be done by clinical staff or designated personnel from the trial office. The sponsor and/or members of the NHS Trust Research and Development Department may require access for audit and monitoring purposes.

We would like to collect information about participants for some years to come in order that we can keep track of how you have responded to treatment. With your agreement, we may ask your GP and any other doctors involved in your care to provide information on your progress. If we do need to contact your GP or another hospital (if your main care is not at Addenbrooke's) for any follow-up information, we will need to use your full name in our correspondence if you have agreed to provide this. If you change address and we lose touch with your GP, we may then locate your new GP through the NHS database. With your permission we would also like to be able to keep in contact with you personally, to see how you are getting on and to let you know about any future research we are carrying out.

If relevant we would also like to ask for permission to update our records using the Eastern Cancer Registration and Information Centre (ECRIC). In order to do this, we will need to use your full name and/or NHS number.

Under no circumstances will you be identified in any way in any report or publication arising from the study.

What will happen to the samples I give?
All samples will be coded ('anonymised') which means that the laboratory researchers who are carrying out tests on your samples will not be able to trace them back to you. However, the trials office will keep a confidential record of the codes used for your samples and can access your notes should the need arise.

The samples will then be sent for analysis within Addenbrooke's Hospital, and at the CRUK Cambridge Institute on the Addenbrooke's Hospital Site. We will be studying DNA in the blood, urine, cervical and tissue samples to look at

molecular markers of gynaecological conditions that may be useful for diagnosis and measuring the extent of abnormal tissue removed at the time of surgery. Samples will be stored at the CRUK Cambridge Institute and if you agree these may be used in future research that has received full ethics committee approval.

It is unlikely that any new information will be generated from this study that will influence your treatment or reveal genetic information that might affect family members. However, if any relevant information does become available during the course of the study this will be given to the doctor looking after you in the hospital. If you wish to be informed of this information your doctor will arrange this with the appropriate genetic counselling.

What will happen if I don't want to carry on with the study?
You are free to withdraw from the study at any time without giving a reason. A decision not to take part or a decision to withdraw at any time will not affect the standard of care that you receive. Please note that if you do choose to withdraw from the study, we would still like to keep stored blood or tissue for future research purposes. However, should you not wish us to do this, you can withdraw your consent and you should tell your doctor or nurses of your wishes. Any stored blood or tissue samples that can still be identified as yours will be destroyed if you wish.

What if there is a problem?
We will try to address any concerns you might have about any aspect of this study. If there are any problems, you should ask to speak with the researchers who will do their best to answer your questions (Telephone: Charlotte Hodgkin, Senior Research Nurse, 01223 348084). If you remain unhappy and wish to complain formally, you can do this through the NHS Complaints Procedure. Details on this can be obtained by contacting the Hospital PALS (Patient Advice and Liaison Service) on 01223 216756.

In the unlikely even that something goes wrong and you are harmed as a result of participating in this study, compensation will be available through the NHS Indemnity scheme.

What will happen to the results of the study?
The results of this study will be published in scientific or medical journals and may be presented at scientific or medical meetings. Please be assured it will not be possible to identify you in any report or publication. If you would like to obtain a copy of the published results please contact the research nurse directly who will be able to arrange this for you.

Who is organising (sponsoring) and funding the study?
This study is sponsored by Cambridge University Hospitals NHS Foundation Trust and The University of Cambridge. The study is being funded by the Medical Research Council and Target Ovarian Cancer .

Who has reviewed this study?

All research within the NHS is reviewed by an independent group of people called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by NRES Committee East of England - Cambridge East.

Further information and contact details
If you require any further information about this study please do not hesitate to contact the Research Nurse below. Further advice on research and conduct of studies can be obtained from Addenbrooke's Hospital PALS on 01223 216756.

The Senior Research Nurse is:
Charlotte Hodgkin
Cambridge Trials Centre,
S4 (Box 279)
Addenbrooke's Hospital, Hills Road, Cambridge CB2 OQQ
Telephone: 01223 348084

Thank you for taking the time to read this Patient information Sheet.

12.1.5 Consent form — surgery

CTCR-OV05 Patient Consent Form 1

Surgery Version 2:December 2014

Consent for CTCR-OV05

Please initial boxes

1. I confirm that I have read the information sheet dated ….. (version…..) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐
3. I understand that relevant sections of my medical notes and data collected during the study may be looked at by responsible individuals from the Cambridge Trials Unit, from regulatory authorities or from the NHS trust where it is relevant to my taking part in research. I give permission for these individuals to have access to my records. ☐
4. I understand that the information collected about me will be used to support other research in the future, and may be shared anonymously with other researchers. ☐
5. I understand that my participation in this study will involve the collection of personal details, information about my treatment and medical history. I understand that all data collected about me will be held under the provisions of the 1998 Data Protection Act and stored in manual and electronic files in secure encoded format. ☐
6. I give permission for my GP to be notified of my participation in this research. If necessary, I give permission for information about my progress to be obtained from my GP (located via NHS databases if required), and if relevant through the Eastern Cancer Registration and Information Centre and understand that this will be done using my full name and/or NHS/hospital number. ☐
7. I agree to the identification and analysis of previously collected tissue and understand that this will be done using my full name and/or NHS/hospital number. ☐
8. I agree to the collection and storage of surplus and additional normal and abnormal tissue at the time of operation for a possible gynaecological condition. ☐
9. I agree to the collection and storage of additional blood, urine and cervical samples for future research into ovarian diseases, including DNA analysis. ☐
10. I agree to the linkage of securely encoded encrypted personal details from samples collected on this study, and previous stored samples. ☐
11. **I agree to take part in the above study** ☐

---

12. OPTIONAL: I wish to be informed if research on my blood reveals genetic information that might affect members of my family (initial box only if you agree) ☐

Patient

Signed ……

Print name …… Date: ……

Doctor/ Person taking consent

Signed ……

Print name …… Date: ……

copies required: **1 copy for researcher; 1 copy for patient; 1 copy to be kept with patient's hospital notes**

12.2 CS01 study documents

Study Title: Molecular analysis of cervical smear samples

Ethics Ref:

Date and Version No: Version 1. 25/2/2015

Chief Investigator: Elizabeth Moore
CRUK Cambridge Institute
Li Ka Shing Centre
Robinson Way
Cambridge
CB2 0EE

Investigators: Dr James Brenton
Dr Nitzan Rosenfeld
Dr Paul Pharoah
Dr Brian Rous
Dr Aileen Patterson
Dr Mercedes Jimenez-Linan
Dr Calire Geary
Beverley Haynes
David Cusack

Sponsor: Cambridge University Hospitals NHS Foundation Trust and the University of Cambridge.

AMENDMENT HISTORY

Amendment No.	Protocol Version No.	Date issued	Author(s) of changes	Details of Changes made

List details of all protocol amendments here whenever a new version of the protocol is produced.

1. SYNOPSIS

Study Title	Molecular analysis of cervical smear samples
Internal ref. no.	A093481
Study Design	Pilot/feasibility study
Study Participants	All routine cervical smear samples showing non-cervical glandular abnormalities.
Planned Sample Size	120
Follow-up duration	No follow up
Planned Study Period	3 years
Primary Objective	Can somatic mutations be detected in cervical smear samples?
Secondary Objectives	Is the detection of somatic mutations in cervical smear samples a sensitive method for the diagnosis of ovarian cancer?
Intervention (s)	None

2. BACKGROUND AND RATIONALE

Ovarian cancer (OC) is the most common cause of death in women with gynaecological cancers[1]. Due to the non-specific symptoms of OC the majority of women (60%) are diagnosed with advanced disease which is associated with poor survival. 5-year survival rates are 43% overall and over 90% for early stage tumours[1].

Current guidelines recommend measuring serum CA125 in all women with suspected OC followed by a transvaginal ultrasound scan (TVUS) in those with elevated levels[2]. In combination these are used to calculate a Risk of Malignancy Index (RMI), which guides referral and subsequent management[3]. False positive elevations are a particular problem in premenopausal women, who are the group most likely to present with adnexal masses. Although CA125 levels are elevated in 80% of cases of advanced OC <50% of cases of early stage cancer have an elevated CA125 resulting in missed diagnoses[4].

Despite decades of effort, CA125 remains the single-best biomarker for OC with models derived from the most promising markers failing to show improvement over serum CA125 alone[5,6]. Due to the current difficulties of early diagnosis of OC and poor outcome in patients diagnosed with late stage disease there is therefore a pressing need to explore complementary strategies.

One such strategy is the detection of circulating tumour DNA in blood samples. It is known that blood contains free circulating DNA and in cancer patients a small fraction of this includes fragments of tumour DNA. DNA shed from ovarian tumours contains somatic mutations, which are highly specific biomarkers of cancer. It has been shown that somatic mutations in ctDNA can be detected in blood samples obtained from women with advanced OC[7].

It is currently hypothesized that OC originates in the fallopian tube epithelium. Molecular alterations specific to OC have been identified in tubal tissue. It is therefore possible that detection of cells and DNA shed from this site could be utilised in the diagnosis of OC. It has recently been shown that tumour DNA and common somatic mutations can be detected in

cervical smear samples[8] and cervical mucous samples[9] from patients with both endometrial and ovarian cancers.

Routine cervical screening was introduced as a method of detecting premalignant changes that could progress to cervical cancer. However, a small number of routine cervical smear samples show glandular abnormalities which suggest possible endometrial or ovarian abnormalities. We propose to collect all routine cervical smear samples showing glandular abnormalities to look for a panel of somatic mutations know to be associated with endometrial and ovarian cancer. If such mutations could be detected via tumour DNA in cervical smear samples this could theoretically be utilised particularly in primary care as a minimally invasive method for ovarian cancer diagnosis and screening.

3. OBJECTIVES

3.1 Primary Objective

To determine if somatic mutations can be detected in routine cervical smear samples showing non-cervical glandular abnormalities.

3.2 Secondary Objectives

Correlation of the presence of somatic mutations with clinical diagnosis to determine if the detection of somatic mutations in cervical smear samples is a sensitive method for the diagnosis of ovarian cancer?

4. STUDY DESIGN

4.1 Summary of Study Design

Pilot/feasibility study

1. Liquid-based cervical sample identified by West Anglia Pathology Service Cervical Cytology Laboratory as '?glandular neoplasia (non-cervical)', code 0.

2. Liquid-based cervical sample identified by West Anglia Pathology Service Cervical Cytology Laboratory as '?Glandular neoplasia of endocervical type' (code 6) for inclusion in control group.

3. Patient name, date of birth, NHS number and cytology classification recorded on password protected spreadsheet stored on cytology server network.

4. Sample allocated a unique code number. Recorded on password protect spreadsheet stored on cytology server network.

5. Patient identifiable information removed from liquid-based cervical sample and sample labelled with unique code number.

6. At the time the liquid-based cervical sample would normally be disposed of as waste (ie. 2-3 weeks from time of collection) courier arranged by staff at the West Anglia Pathology Service Cervical Cytology Laboratory for transport of sample to Cancer Research UK Cambridge Institute.

7. Sample collected by courier and transported to Cancer Research UK Cambridge Institute (CRUK CI) for collection by researcher.

8. Researchers at CRUK CI informed by email (elizabeth.moore@cruk.cam.ac.uk) that sample has been sent. Information to be provided: unique code number, cytology classification, date of transfer.

9. Beverley Haynes at Cambridge University Hospitals Tissue Bank informed by email (beverley.haynes@addenbrookes.nhs.uk) of transfer of sample to CRUK CI. Information to be provided: unique code number, date of transfer.

10. DNA will be extracted from cervical smear samples by researchers at CRUK Cambridge Institute and sequenced for a panel of mutations known to be associated with ovarian and endometrial cancer.

11. There will be no change in clinical practice and all women with a cervical smear showing non-cervical glandular abnormalities will be referred via their GP to a gynaecologist for further investigation and treatment as required.

12. Outcome diagnosis is obtained by West Anglia Pathology Service Cervical Cytology Laboratory as per current practice. When available this will be recorded on the password protect database stored on the cytology server network.

13. The unique code number and outcome diagnosis will provided to researchers at CRUK CI (elizabeth.moore@cruk.cam.ac.uk) for correlation with sequencing data.

4.2 Study Participants

4.2.1 Overall Description of Study Participants

-All routine cervical smear samples classified by West Anglia Pathology Service Cervical Cytology Laboratory as showing non-cervical glandular abnormalities.

-Cervical smear samples identified by West Anglia Pathology Service Cervical Cytology Laboratory as showing cervical glandular abnormalities will be used as controls.

4.2.2 Inclusion Criteria

- All routine cervical smear samples showing non-cervical glandular abnormalities.
- Routine cervical smear samples showing cervical glandular abnormalities will be included as controls.

4.2.3 Exclusion Criteria

The participant may not enter the study if ANY of the following apply:

- West Anglia Pathology Service Cervical Cytology Laboratory are required to keep the cervical smear sample for clinical reasons.

4.3 Study Procedures

4.3.1 Informed Consent

Individual women will not be contacted for consent to use their samples in this clinical study. As per routine clinical practice GPs are informed by the cytology service of the identification of a non-cervical glandular smear and referral to a gynaecologist for further investigation and management as clinically appropriate is advised. This procedure will not change. Samples provided to researchers will be annonymised and no researchers at CRUK Cambridge Institute will have access to any patient identifiable data. The clinical significance of a mutation being identified for an individual women is not yet currently understood. Specifically we are not looking for germline mutations so it is extremely unlikely that the research will reveal any genetic information that is of significance to family members. Molecular data produced will therefore not be provided to the woman or treating clinician and will therefore not influence clinical investigation or management in any way.

The West Anglia Pathology Service Cervical Cytology Laboratory analyse approximately 100,000 cervical smear samples per year from six Clinical Commissioning Groups. Of these only 0.04% (40 samples per year) will be classified as showing non-cervical glandular abnormalities. It is therefore not feasible to consent all women undergoing a cervical smear for potential use in this study. Once a non-cervical glandular smear has been identified women are referred via the GP to see a gynaecologist at their local hospital. As referral to a gynaecologist is done by the GP to their local hospital we are unable to access this appointment data. We are therefore unable to contact women to discuss participation in this study at the time of referral to the gynaecologist particularly since the cervical smear samples will be disposed of 2-3 weeks after collection.

4.3.2 Study Assessments

DNA will be extracted from cervical smear samples and sequenced for a panel of mutations known to be associated with endometrial and ovarian cancer. The presence and frequency of somatic mutations detected will be correlated with clinical diagnosis. The sensitivity of molecular analysis of cervical smear samples for the diagnosis of ovarian cancer will be calculated. If promising this study will be extended to include a larger cohort of samples from other cervical cytology laboratories to assess the utility of population based cervical smears in screening for OC.

4.4 Definition of End of Study

The end of study is the date of the last clinical diagnosis of the last participant.

5. STATISITICS

5.1 The Number of Participants

Approximately 0.04% of routine cervical smears are thought to show non-cervical glandular abnormalities. Based on the number of smears the West Anglia Pathology Service Cervical Cytology Laboratory process we would expect to receive approximately 40 samples a year eligible for inclusion. This study will run over 3 years. We would therefore expect to obtain approximately 120 samples in total.

CA125 is currently the first line test in primary care for women with suspected ovarian cancer. Sensitivity and specificity for CA125 with a cut off of 35IU/ml is estimated at 76.2% 95% CI (60.5%, 87.9%) and 81.2% 95% CI (72.2%, 88.3%) respectively. (Jacobs I et al. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer.British Journal of Obstetrics and Gynaecology 1990; 97; 922-929).

To establish if sensitivity differs significantly from 76.2% but to be 80% certain with a probability of 0.05 that a sensitivity as low as 60.5% would be excluded requires 63 subjects with the disease. To establish if specificity differs significantly from 81.2 % but to be 80% certain with a probability of 0.05 that a specificity as low as 72.2% would be excluded requires 162 subjects without the disease. (Arkin CF, Wachtel, MS. How many patients are necessary to assess test performance? JAMA 1990; 263; 275-78.)

As this is a pilot study our sample size is 120 based on the number of samples we expect to be able to collect over three years. If the performance characteristics of molecular analysis are comparable to CA125 (sensitivity not less than 60% and specificity not less than 70%) we would look to extend this project to other sites to reach the required sample size of 225.

6. ETHICS

6.1 Participant Confidentiality

All samples and clinical diagnosis data will be provided to the researchers in an anonymised form, identified only by the unique code number. Patient name, date of birth, NHS number, cytology classification and clinical diagnosis will be stored in a password protected database on the cytology server network. Access to this database will be restricted to named individuals.

7. DATA HANDLING AND RECORD KEEPING

All samples will be anonymised to researchers and will be allocated a unique study number by which all data will be stored at CRUK Cambridge Institute on University of Cambridge computers.

8. FINANCING AND INSURANCE

This study is being sponsored by Cambridge University Hospitals NHS Trust and The University of Cambridge. The University of Cambridge will provide insurance for the

management, conduct and design of the study. The study is being funded by the Medical Research Council and Target Ovarian Cancer.

9. REFERENCES

1. CancerStats. Cancer Research UK
http://info.cancerresearchuk.org/cancerstats/types/ovary/ Accessed 03/11/2014.

2. NICE Guidance CG122. Ovarian Cancer: the recognition and initial management of ovarian cancer. April 2011. http://guidance.nice.org.uk/CG122/Guidance/pdf/English. Accessed 01/07/2013

3. Rufford BD, Jacobs IJ. 2003. Ovarian Cysts in Postmenopausal Women. RCOG Greentop Guideline Number 34.

4. Woolas RP, Xu FJ, Jacobs IJ, Yu YH, Daly L, Berchuck A, Soper JT, Clarke- Pearson DL, Oram DH, Bast RC. 1993 Elevation of multiple serum markers in patients with stage I ovarian cancer. J Natl Cancer Inst:85;1748-51.

5. Cramer DW, Bast RC, Berg CD, Diamandis EP, Godwin AK, Hartge P, Lokshin AE, Lu KH, McIntosh MW, Mor G, Patriotis C, Pinsky PF, Thornquist MD, Scholler N, Skates SJ, Sluss PM, Srivastava S, Ward DC, Zhang Z, Zhu CS, Urban N. 2011. Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. Cancer Prev Res:4;365-74.

6. Zhu CS, Pinsky PF, Cramer DW, Ransohoff DF, Hartge P, Pfeiffer RM, Urban N, Mor G, Bast RC Jr, Moore LE, Lokshin AE, McIntosh MW, Skates SJ, Vitonis A, Zhang Z, Ward DC, Symanowski JT, Lomakin A, Fung ET, Sluss PM, Scholler N, Lu KH, Marrangoni AM, Patriotis C, Srivastava S, Buys SS, Berg CD; PLCO Project Team. 2011. A framework for evaluating biomarkers for early detection: validation of biomarker panels for ovarian cancer. Cancer Prev Res:4;375-83.

7. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, Brenton JD, Rosenfeld N. 2012. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med:4:136-68.

8. Kinde I, Bettegowda C, Wang Y, Wu J, Agrawal N, Shih IeM, Kurman R, Dao F, Levine DA, Giuntoli R, Roden R, Eshleman JR, Carvalho JP, Marie SK, Papadopoulos N, Kinzler KW, Vogelstein B, Diaz LA Jr. Evaluation of DNA from the Papanicolaou Test to Detect Ovarian and Endometrial Cancers. Sci Transl Med 5. 2013 9;5(167):167ra4.

9. Lamzabi I, Buckingham L, Gebrekiristos M, Jain R, Gattuso P, Reddy V, Guirguis A, Dewdney S, Rotmensch J, Bitterman P. 2013. Use of cervical mucous to screen for gynecological malignancies: a pilot study. Mod Pathol: 2013 Jun 14. doi: 10.1038/modpathol.2013.92. [Epub ahead of print]

12.3 TAm-Seq primer panel 1

Plate Name	uid	fwd_seq	rev_seq	strand	chr	fwd_start	fwd_end	rev_start	rev_end	amp_start	amp_end
EGFR	EGFR_E00001601336_1	GCGTCTTCACCTGGAAGGG	CCGGACATAGTCCAGGAGG	+	chr7	55248902	55248920	55249093	55249111	55248902	55249111
EGFR	EGFR_E00001601336_2	GCGTGGACAACCCCCAC	GGCTCCTTATCTCCCCTCC	+	chr7	55249005	55249021	55249195	55249213	55249005	55249213
EGFR	EGFR_E00001631695_1	GTGTCACTCGTAATTAGGTCCA	GGCCTCAGTACAAACTCATTAGC	+	chr7	55260366	55260387	55260553	55260575	55260366	55260575
EGFR	EGFR_E00001681524_1	GGATGCAGAGCTTCTTCCCA	TTCTCTTCCGCACCCAG	+	chr7	55259352	55259371	55259526	55259542	55259352	55259542
EGFR	EGFR_E00001681524_2	GGTCTTCTCTGTTTCAGGGCAT	GCTGACCTAAAGCCACCTCC	+	chr7	55259395	55259416	55259572	55259591	55259395	55259591
EGFR	EGFR_E00001773562_1	TACCCTCCATGAGGCACAC	GGAGAgCTGTAAATTCTGGCTT	+	chr7	55269336	55269354	55269495	55269516	55269336	55269516
EGFR	EGFR_E00001779947_1	TGTTCATTCATGATCCCACTGC	CCACCAGTCACTCACACTTG	+	chr7	55266362	55266383	55266552	55266571	55266362	55266571
EGFR	EGFR_E00001779947_2	TCCCTGCCAGCGAGAT	AGGGATGCAAAGGCCTCA	+	chr7	55266461	55266476	55266624	55266641	55266461	55266641
EGFR	EGFR_E00001790701_1	GCCTTCTTTAAGCAATGCCATCTTTAT	CAATGGAAGCaCAGACTGCAA	+	chr7	55267932	55267958	55268114	55268134	55267932	55268134
EGFR	EGFR_E00001801208_1	CCCCTGCTCCTATAGCCAA	ATGAGGTACTCGTCGGCATC	+	chr7	55268806	55268824	55268968	55268987	55268806	55268987
EGFR	EGFR_E00001801208_2	ACTTCTACCGTGCCCTGA	GTTCAAATGAGTAGACACAGCTT	+	chr7	55268921	55268938	55269079	55269101	55268921	55269101
EGFR	EGFR_Exon19	TCACAATTGCCAGTTAACGTCT	CCACACAGCAAAGCAGAAAC	+	chr7	55242373	55242394	55242518	55242537	55242373	55242537
PTEN_PIK3CA	PTEN_E00001156315_1m	GCAACAGATAACTCAGATTGCCTT	GTTTCCTCTGGTCCTGGTATGA	+	chr10	89720457	89720480	89720685	89720706	89720457	89720706
PTEN_PIK3CA	PTEN_E00001156315_5	AGGACAAAATGTTTCACTTTTGGGTAA	ACTAGATATTCCTTGTCATTATCTGCAC	+	chr10	89720649	89720675	89720772	89720799	89720649	89720799
PTEN_PIK3CA	PTEN_E00001156315_7	CCTCAGAAAAAgTAGAAAATGGAAGTC	ACAAGTCAaCAACCCCCACA	+	chr10	89720706	89720732	89720896	89720915	89720706	89720915
PTEN_PIK3CA	PTEN_E00001156321_1	TGACAGTTTGACAGTTAAAGGCAT	CACACACAGGTAACGGCTGA	+	chr10	89717547	89717570	89717707	89717726	89717547	89717726
PTEN_PIK3CA	PTEN_E00001156321_2	TGTGGTCTGCCAGCTAAAGG	TCTCCCAATGAAAGTAAAGTACAAACC	+	chr10	89717620	89717639	89717776	89717802	89717620	89717802
PTEN_PIK3CA	PTEN_E00001156321_4m	TCCACAAACAGAACAAGATGCT	GGCCTTTTCCTTCAAACAGGATT	+	chr10	89717748	89717769	89717934	89717956	89717748	89717956
PTEN_PIK3CA	PTEN_E00001156327_1	TCTTAAATGGCTACGACCCAG	TCCAGATGATTCTTTAACAGGTAGC	+	chr10	89711775	89711795	89711918	89711942	89711775	89711942
PTEN_PIK3CA	PTEN_E00001156327_4	CAGTCAGAGGCGCTATGTGT	TCTAGATATGGTTAAGAAAACTGTTCCA	+	chr10	89711889	89711908	89712050	89712077	89711889	89712077
PTEN_PIK3CA	PTEN_E00001156330_1	ttCTTATTCTGAGGTTATCTTTTTACCAC	TCATTACACCAGTTCGTCCCT	+	chr10	89692739	89692767	89692899	89692919	89692739	89692919
PTEN_PIK3CA	PTEN_E00001156330_3	TGACCAATGGCTAAGTGAAGATGA	TCCAGGAAGAGGAAAGGAAAAACA	+	chr10	89692840	89692863	89693025	89693048	89692840	89693048
PTEN_PIK3CA	PTEN_E00001156337_4	TATATCACTTTTAAACTTTTCTTTTAGTTGTGC	CTCGATAATCTGGATGACTCATTATTGTT	+	chr10	89690776	89690808	89690912	89690940	89690776	89690940
PTEN_PIK3CA	PTEN_E00001156344_1	AATCTGTCTTTTGGTTTTTCTTGATAGT	AATAGTTGTTTTAGAAGATATTTGCAAGC	+	chr10	89685172	89685199	89685339	89685367	89685172	89685367
PTEN_PIK3CA	PTEN_E00001156351_1	TGCTGCATATTTCAGATATTTCTTTCCTTA	ATGAAAACACAACATGAATATAAACATCAAT	+	chr10	89653738	89653767	89653897	89653927	89653738	89653927
PTEN_PIK3CA	PTEN_E00001456541_1	AGATGAGTcATATTTGTGGGTTTTCA	TCTGGATCAGAGTCAGTGGT	+	chr10	89724997	89725022	89725161	89725180	89724997	89725180
PTEN_PIK3CA	PTEN_E00001456541_2	GTAGAGGAGCCGTCAAATCCA	TTCATGGTGTTTTATCCCTCTTGA	+	chr10	89725068	89725088	89725241	89725264	89725068	89725264
PTEN_PIK3CA	PTEN_E00001456562_1	GCAGCTTCTGCCATCTCTCT	TCCGTCTACTCCCACGTTCT	+	chr10	89624175	89624194	89624353	89624372	89624175	89624372
TP53	TP53_00001404886_13	tctgTATCAGGCAAAGTCATAGAA	GCCTCAAAGACAATGGCTCC	+	chr17	7576584	7576607	7576715	7576734	7576584	7576734
TP53	TP53_E00001255919_1	GAGAAAGCCCCCCTACTGC	AGCATCTTATCCGAGTGGAAGG	+	chr17	7578091	7578109	7578253	7578274	7578091	7578274
TP53	TP53_E00001255919_3	TCCAAATACTCCACACGCAAA	gCTGCCCCCACCATGAG	+	chr17	7578229	7578249	7578390	7578406	7578229	7578406
TP53	TP53_E00001255919_5	AGCTGCTCACCATCGCTA	CCAACTGgCCAAGACCT	+	chr17	7578361	7578378	7578509	7578525	7578361	7578525
TP53	TP53_E00001255919_6	TGTGCTGTGACTGCTTGTAG	TGCCCTGACTTTCAACTCTGT	+	chr17	7578425	7578444	7578574	7578594	7578425	7578594
TP53	TP53_E00001596491_1	TTTCGCTTCCCACAGGTCTC	CAGCCAGACTGCCTTCCG	+	chr17	7579758	7579777	7579923	7579940	7579758	7579940
TP53	TP53_E00001612188_1	ATACGGCCAGGCATTGAAGT	CCTCCTGGCCCCTGTC	+	chr17	7579260	7579279	7579406	7579421	7579260	7579421
TP53	TP53_E00001612188_2	GGAAACCGTAGCTGCCCTG	AAGACCCAGGTCCAGATGAA	+	chr17	7579359	7579377	7579501	7579520	7579359	7579520
TP53	TP53_E00001665758_1	GGGGTCAGaGGCAAGCAG	CTTGGGCCTGTGTTATCTCC	+	chr17	7577432	7577449	7577612	7577631	7577432	7577631
TP53	TP53_E00001728015_1	GGAATCCTATGGCTTTCCAACC	CCCCCTCCTCTGTTGCTG	+	chr17	7573859	7573880	7574037	7574054	7573859	7574054
TP53	TP53_E00001757276_1	GACCCAAAACCCAAAATGGC	TCCCTGCTTCTGTCTCCTAC	+	chr17	7572850	7572869	7573011	7573030	7572850	7573030
TP53	TP53_E00001789298_1	AGAAAACGGCATTTTGAGTGT	AAGGGTGCAGTTATGCCTCA	+	chr17	7576786	7576806	7576964	7576983	7576786	7576983
TP53	TP53_E00001789298_2	CTGGTGtTGTTGGGCAGT	ATCTCCgCAAGAAAGGGGAG	+	chr17	7576908	7576925	7577056	7577075	7576908	7577075
TP53	TP53_E00001789298_3	TGTCCTGCTTGCTTACCTCG	GCCTCTTGCTTCTCTTTTCCT	+	chr17	7577003	7577022	7577167	7577187	7577003	7577187
TF new primers	KRAS_AA_HS	GCCTGCTGAAAATGACTGAA	AGAATGGTCCTGCACCAGTAA	-	chr12	25398163	25398183	25398310	25398329	25398163	25398329
TF new primers	BRAF_AA_HS	TCATAATGCTTGCTCTGATAGGA	CTGATGGGACCCACTCCAT	-	chr7	140453108	140453126	140453234	140453256	140453108	140453256
TF new primers	EXP0116_TP53_E4	CAGCCTCTGGcATTCTGG	cctggtcctctgaCTGCTCT	+	chr17	7579479	7579496	7579607	7579626	7579479	7579626
TF new primers	EXP0116_TP53_E3	TCAAATCATCCATTGCTTGG	ccatgggactgactttctgc	+	chr17	7579557	7579576	7579735	7579754	7579557	7579754
TF new primers	PIK3CA_Ex10_gap	cAGAGGGGAAAAATaTGACAAA	AACAGAGAATCTCCATTTTAGCAC	+	chr3	178935943	178935964	178936127	178936150	178935943	178936150
TF new primers	PIK3CA_Exn21_gap	TGAGCAAGAGGCTTTGGAGT	GGTCTTTGCCTGCTGAGAGT	+	chr3	178952038	178952057	178952208	178952227	178952038	178952227

12.4 TAm-Seq primer panel 10

uid	fwd_seq	rev_seq	strand	chr	fwd_start	fwd_end	rev_start	rev_end	amp_start	amp_end
TP53_D0008_001	GTGGGGAACAAGAAGTGGAGA	CTCCCTGCTTCTGTCTCCTAC	+	chr17	7,572,903	7,572,924	7,573,010	7,573,031	7,572,903	7,573,031
TP53_D0008_002	AGGAAGGGGCTGAGGTCA	GCGCTTCGAGATGTTCCGA	+	chr17	7,573,904	7,573,922	7,574,000	7,574,019	7,573,904	7,574,019
TP53_D0008_005	CATCCAGTGGTTTCTTCTTTGG	TTCACTTTTATCACCTTTCCTTGCC	+	chr17	7,576,873	7,576,895	7,576,936	7,576,961	7,576,873	7,576,961
TP53_D0008_006	CGCTTCTTGTCCTGCTTGC	TGTGCCTGTCCTGGGAGA	+	chr17	7,576,996	7,577,015	7,577,097	7,577,115	7,576,996	7,577,115
TP53_D0008_014	TGTCGTCTCTCCAGCCCC	GGCCATCTACAAGCAGTCACA	+	chr17	7,578,343	7,578,361	7,578,429	7,578,450	7,578,343	7,578,450
TP53_D0008_016	GGAATCAACCCACAGCTGC	AACTCTGTCTCCTTCCTCTTCCTA	+	chr17	7,578,483	7,578,502	7,578,557	7,578,581	7,578,483	7,578,581
TP53_D0008_012	AGACCTCAGGCGGCTCA	CCTCAGCATCTTATCCGAGTGG	+	chr17	7,578,173	7,578,190	7,578,256	7,578,278	7,578,173	7,578,278
TP53_D0008_017	TGGAAGCCAGCCCCTCA	GTCATCTTCTGTCCCTTCCCAG	+	chr17	7,579,284	7,579,301	7,579,386	7,579,408	7,579,284	7,579,408
TP53_D0008_018	CCGTAGCTGCCCTGGTAG	AGCTCCTACACCGGCG	+	chr17	7,579,364	7,579,382	7,579,437	7,579,453	7,579,364	7,579,453
TP53_D0008_007	ATTCTCTTCCTCTGTGCGCC	TTGCTTCTCTTTTCCTATCCTGAGT	+	chr17	7,577,074	7,577,094	7,577,157	7,577,182	7,577,074	7,577,182
TP53_D0008_022	GCCCTTCCAATGGATCCACT	CTTCCGGGTCACTGCCAT	+	chr17	7,579,816	7,579,836	7,579,910	7,579,928	7,579,816	7,579,928
TP53_D0008_021	CCCAGCCCAACCCTTGTC	GAAGCGAAAATTCCATGGGACTG	+	chr17	7,579,677	7,579,695	7,579,743	7,579,766	7,579,677	7,579,766
TP53_D0008_015	GCCTCACAACCTCCGTCAT	AACTGGCCAAGACCTGCC	+	chr17	7,578,407	7,578,426	7,578,505	7,578,523	7,578,407	7,578,523
TP53_D0008_023	CTAGGATCTGACTGCGGCTC	TGGAAGTGTCTCATGCTGGATC	+	chr17	7,579,887	7,579,907	7,579,959	7,579,981	7,579,887	7,579,981
TP53_D0008_009	CAAGTGGCTCCTGACCTGG	CTAGGTTGGCTCTGACTGTACC	+	chr17	7,577,484	7,577,503	7,577,590	7,577,612	7,577,484	7,577,612
TP53_D0008_008	GGTCAGAGGCAAGCAGAGG	CCTCACCATCATCACACTGGAA	+	chr17	7,577,434	7,577,453	7,577,506	7,577,528	7,577,434	7,577,528
TP53_D0008_020	GCATTCTGGGAGCTTCATCTG	TGACTGCTCTTTTCACCCATCT	+	chr17	7,579,488	7,579,509	7,579,594	7,579,616	7,579,488	7,579,616
TP53_D0008_004	AAACGGCATTTTGAGTGTTAGAC	ACAACACCAGCTCCTCTCC	+	chr17	7,576,789	7,576,812	7,576,898	7,576,917	7,576,789	7,576,917
TP53_D0008_003	TTGAGTTCCAAGGCCTCATTCA	GTACTGTGTATATACTTACTTCTCCCCC	+	chr17	7,573,975	7,573,997	7,574,049	7,574,077	7,573,975	7,574,077
TP53_D0008_013	ATCCAAATACTCCACACGCAAATTT	CCTCTGATTCCTCACTGATTGCT	+	chr17	7,578,228	7,578,253	7,578,294	7,578,317	7,578,228	7,578,317
TP53_D0008_011	GGCCACTGACAACCACCC	TCGACATAGTGTGGTGGTGC	+	chr17	7,578,120	7,578,138	7,578,193	7,578,213	7,578,120	7,578,213
TP53_D0008_019	TGACAGGGGCCAGGAG	CAATGGTTCACTGAAGACCCA	+	chr17	7,579,405	7,579,421	7,579,512	7,579,533	7,579,405	7,579,533
TP53_D0008_010	ACTGTTACACATGTAGTTGTAGTGGA	CTTGCCACAGGTCTCCCC	+	chr17	7,577,561	7,577,587	7,577,649	7,577,667	7,577,561	7,577,667

12.5 TAm-Seq UKOPS primer panel

uid	fwd_seq	rev_seq	strand	chr	fwd_start	fwd_end	rev_start	rev_end	amp_start	amp_end
TP53_D0008_001	GTGGGGAACAAGAAGTGGAGA	CTCCCTGCTTCTGTCTCCTAC	+	chr17	7,572,903	7,572,924	7,573,010	7,573,031	7,572,903	7,573,031
TP53_D0008_002	AGGAAGGGGCTGAGGTCA	GCGCTTCGAGATGTTCCGA	+	chr17	7,573,904	7,573,922	7,574,000	7,574,019	7,573,904	7,574,019
TP53_D0008_005	CATCCAGTGGTTTCTTCTTTGG	TTCACTTTTATCACCTTTCCTTGCC	+	chr17	7,576,873	7,576,895	7,576,936	7,576,961	7,576,873	7,576,961
TP53_D0008_006	CGCTTCTTGTCCTGCTTGC	TGTGCCTGTCCTGGGAGA	+	chr17	7,576,996	7,577,015	7,577,097	7,577,115	7,576,996	7,577,115
TP53_D0008_014	TGTCGTCTCTCCAGCCCC	GGCCATCTACAAGCAGTCACA	+	chr17	7,578,343	7,578,361	7,578,429	7,578,450	7,578,343	7,578,450
TP53_D0008_016	GGAATCAACCCACAGCTGC	AACTCTGTCTCCTTCCTCTTCCTA	+	chr17	7,578,483	7,578,502	7,578,557	7,578,581	7,578,483	7,578,581
TP53_D0008_012	AGACCTCAGGCGGCTCA	CCTCAGCATCTTATCCGAGTGG	+	chr17	7,578,173	7,578,190	7,578,256	7,578,278	7,578,173	7,578,278
TP53_D0008_017	TGGAAGCCAGCCCCTCA	GTCATCTCTGTCCCTTCCCAG	+	chr17	7,579,284	7,579,301	7,579,386	7,579,408	7,579,284	7,579,408
TP53_D0008_018	CCGTAGCTGCCCTGGTAG	AGCTCCTACACCGGCG	+	chr17	7,579,364	7,579,382	7,579,437	7,579,453	7,579,364	7,579,453
TP53_D0008_007	ATTCTCTTCCTCTGTGCGCC	TTGCTTCTCTTTTCCTATCCTGAGT	+	chr17	7,577,074	7,577,094	7,577,157	7,577,182	7,577,074	7,577,182
TP53_D0008_022	GCCCTTCCAATGGATCCACT	CTTCCGGGTCACTGCCAT	+	chr17	7,579,816	7,579,836	7,579,910	7,579,928	7,579,816	7,579,928
TP53_D0008_021	CCCAGCCCAACCCTTGTC	GAAGGGAAAATTCCATGGGACTG	+	chr17	7,579,677	7,579,695	7,579,743	7,579,766	7,579,677	7,579,766
TP53_D0008_015	GCCTCACAACCTCCGTCAT	AACTGGCCAAGACCTGCC	+	chr17	7,578,407	7,578,426	7,578,505	7,578,523	7,578,407	7,578,523
TP53_D0008_023	CTAGGATCTGACTGCGGCTC	TGGAAGTGTCTCATGCTGGATC	+	chr17	7,579,887	7,579,907	7,579,959	7,579,981	7,579,887	7,579,981
TP53_D0008_009	CAAGTGGCTCCTGACCTGG	CTAGGTTGGCTCTGACTGTACC	+	chr17	7,577,484	7,577,503	7,577,590	7,577,612	7,577,484	7,577,612
TP53_D0008_008	GGTCAGAGGCAAGCAGAGG	CCTCACCATCATCACACTGGAA	+	chr17	7,577,434	7,577,453	7,577,506	7,577,528	7,577,434	7,577,528
TP53_D0008_020	GCATTCTGGGAGCTTCATCTG	TGACTGCTCTTTTCACCCATCT	+	chr17	7,579,488	7,579,509	7,579,594	7,579,616	7,579,488	7,579,616
TP53_D0008_004	AAACGGCATTTTGAGTGTTAGAC	ACAACACCAGGCTCCTCTCC	+	chr17	7,576,789	7,576,812	7,576,898	7,576,917	7,576,789	7,576,917
TP53_D0008_003	TTGAGTTCCAAGGCCTCATTCA	GTACTGTGTATATACTTACTTCTCCCCC	+	chr17	7,573,975	7,573,997	7,574,049	7,574,077	7,573,975	7,574,077
TP53_D0008_013	ATCCAAATACTCCACACGCAAATTT	CCTCTGATTCCTCACTGATTGCT	+	chr17	7,578,228	7,578,253	7,578,294	7,578,317	7,578,228	7,578,317
TP53_D0008_011	GGCCACTGACAACCACCC	TCGACATAGTGTGGTGGTGC	+	chr17	7,578,120	7,578,138	7,578,193	7,578,213	7,578,120	7,578,213
TP53_D0008_019	TGACAGGGCCAGGAG	CAATGGTTCACTGAAGACCCA	+	chr17	7,579,405	7,579,421	7,579,512	7,579,533	7,579,405	7,579,533
TP53_D0008_010	ACTGTTACACATGTAGTTGTAGTGGA	CTTGCCACAGGTCTCCCC	+	chr17	7,577,561	7,577,587	7,577,649	7,577,667	7,577,561	7,577,667
CTNNB1	TGGCAGCAACAGTCTTACCT	GGTATCCACATCCTCTTCCTCA	+	chr3	41,266,076	41,266,095	41,266,159	41,266,180	41,266,076	41,266,180
KRAS	GGCCTGCTGAAAATGACTGAA	TGTTGGATCATATTCGTCCACAA	+	chr12	25,398,214	25,398,234	25,398,308	25,398,330	25,398,214	25,398,330
BRAF	TCATAATGCTTGCTCTGATAGGA	CTGATGGGACCACTCCAT	+	chr7	140,453,108	140,453,126	140,453,234	140,453,256	140,453,108	140,453,256
PIK3CA_Ex21	TGAGCAAGAGGCTTTGGAGT	TCAGTTCAATGCATGCTGTTT	+	chr3	178,952,038	178,952,057	178,952,132	178,952,152	178,952,038	178,952,152
PIK3CA_Ex10	GAGTAACAGACTAGCTAGAGACAATG	TAGCACTTACCTGTGACTCCA	+	chr3	178,935,997	178,936,022	178,936,112	178,936,132	178,935,997	178,936,132
PTEN_E00001156315_1m	GCAACAGATAACTCAGATTGCCTT	GTTTCCTCTGGTCCTGGTATGA	+	chr10	89,720,457	89,720,480	89,720,685	89,720,706	89,720,457	89,720,706
PTEN_E00001156315_5	AGGACAAAATGTTTCACTTTTGGGTAA	ACTAGATATTCCTTGTCATTATCTGCAC	+	chr10	89,720,649	89,720,675	89,720,772	89,720,799	89,720,649	89,720,799
PTEN_E00001156315_7	CCTCAGAAAAAgTAGAAAATGGAAGTC	ACAAGTCAaCAACCCCACA	+	chr10	89,720,706	89,720,732	89,720,896	89,720,915	89,720,706	89,720,915
PTEN_E00001156321_1	TGACAGTTTGACAGTTAAAGGCAT	CACACACAGGTAACGGCTGA	+	chr10	89,717,547	89,717,570	89,717,707	89,717,726	89,717,547	89,717,726
PTEN_E00001156321_2	TGTGGTCTGCCAGCTAAAGG	TCTCCCAATGAAAGTAAAGTACAAACC	+	chr10	89,717,620	89,717,639	89,717,776	89,717,802	89,717,620	89,717,802
PTEN_E00001156321_4m	TCCACAAACAGAACAAGATGCT	GGCCTTTCCTTCAAACAGGATT	+	chr10	89717748	89717769	89717934	89717956	89717748	89717956
PTEN_E00001156327_1	TCTTAAATGGCTACGACCCAG	TCCAGATGATTCTTTAACAGGTAGC	+	chr10	89711775	89711795	89711918	89711942	89711775	89711942
PTEN_E00001156327_4	CAGTCAGAGGCGCTATGTGT	TCTAGATATGGTTAAGAAAACTGTTCCA	+	chr10	89711889	89711908	89712050	89712077	89711889	89712077
PTEN_E00001156330_1	ttCTTATTCTGAGGTTATCTTTTTACCAC	TCATTACACCAGTTCGTCCCT	+	chr10	89692739	89692767	89692899	89692919	89692739	89692919
PTEN_E00001156330_3	TGACCAATGGCTAAGTGAAGATGA	TCCAGGAAGAGGAAAGGAAAAACA	+	chr10	89692840	89692863	89693025	89693048	89692840	89693048
PTEN_E00001156337_4	TATATCACTTTTAAACTTTCTTTTTAGTTGTGC	CTCGATAATCTGGATGACTCATTATTGTT	+	chr10	89690776	89690808	89690912	89690940	89690776	89690940
PTEN_E00001156344_1	AATCTGTCTTTTGGTTTTTCTTGATAGT	AATAGTTGTTTTAGAAGATATTTGCAAGC	+	chr10	89685172	89685199	89685339	89685367	89685172	89685367
PTEN_E00001156351_1	TGCTGCATATTTCAGATATTTCTTTCCTTA	ATGAAAACACAACATGAATATATAAACATCAAT	+	chr10	89653738	89653767	89653897	89653927	89653738	89653927
PTEN_E00001456541_1	AGATGAGTcATATTTGTGGGTTTTCA	TCTGGATCAGAGTCAGTGGT	+	chr10	89724997	89725022	89725161	89725180	89724997	89725180
PTEN_E00001456541_2	GTAGAGGAGCCGTCAAATCCA	TTCATGGTGTTTTATCCCTCTTGA	+	chr10	89725068	89725088	89725241	89725264	89725068	89725264
PTEN_E00001456562_1	GCAGCTTCTGCCATCTCTCT	TCCGTCTACTCCCAGGTTCT	+	chr10	89624175	89624194	89624353	89624372	89624175	89624372

12.6 Mutation calls for UKOPS FFPE using TAm-Seq

JBLAB number	FFPE Block ID	OTTA ID	Histology	Chromosome	Gene	Position	Effect	Ref	Mutant	Indel	cDNA.effect	Protein.effect
12201	R07-11403-REMA2	TUKO01230	high grade serous									
12202	R07-20333-3A	TUKO01594	endometrioid	17	TP53	7578406	nonsynonymous	C	T		c.G524A	p.R175H
12203	UH07-20421-2	TUKO01595	high grade serous									
12204	R08 05206 10	TUKO01897	high grade serous									
12205	UR08-11153 A1	TUKO01891	high grade serous	17	TP53	7577538	nonsynonymous	C	T		c.G743A	p.R248Q
12206	R08-14573-2	TUKO01892	high grade serous	17	TP53	7578461	nonsynonymous	C	A		c.G469T	p.V157F
12207	077 40703 1H	TUKO01893	unknown	17	TP53	7577028	deletion			TGCTCCCTG	c.902_909delCAGGGAGC	p.P301Hfs*2
12208	H06-18278-F	TUKO01211	mucinous	17	TP53	7578271	nonsynonymous	T	G		c.A578C	p.H193P
12209	H07-18627-1B	TUKO01208	high grade serous	17	TP53	7577108	nonsynonymous	C	A		c.G830T	p.C277F
12210	10758/06 2J	TUKO01894	MMMT	17	TP53	7578271	nonsynonymous	T	C		c.A578G	p.H193R
12211	R07-4984 A	TUKO00579	endometrioid									
12212	R07-12950 A1	No OTTA ID	high grade serous									
12213	R07-18229-A1	TUKO01222	high grade serous	17	TP53	7578433	nonsynonymous	G	A		c.C497T	p.S166L
12214	07/26988 2O	TUKO01552	endometrioid	12	KRAS	25398284	nonsynonymous	C	T		c.G35A	p.G12D
12215	H06-10510-4A	TUKO00193	high grade serous	17	TP53	7578190	nonsynonymous	T	C		c.A659G	p.Y220C
12216	H06-16736-3G	TUKO01133	adenoca NOS	7	EGFR	55259529	nonsynonymous	G	A		c.G2587A	p.G863S
			unknown	17	TP53	7578493	stopgain	C	T		c.G437A	p.W146*
12217	H06-17939-2E	TUKO00876	high grade serous	17	TP53	7578271	nonsynonymous	T	C		c.A578G	p.H193R
12218	H06-19212-3C	TUKO00877	high grade serous									
12219	17028/07 2A	TUKO01510	high grade serous	17	TP53	7577570	nonsynonymous	C	T		c.G711A	p.M237I
12220	R07-6470-RESA11	TUKO00595	endometrioid									
12221	R08-02196	No OTTA ID	mixed clear cell/endometrioid									
12222	07 36137 1G	TUKO01547	clear cell	3	PIK3CA	178952085	nonsynonymous	A	T		c.A3140T	p.H1047L
12223	07 37790 1H	TUKO01468	clear cell	3	PIK3CA	178952085	nonsynonymous	A	G		c.A3140G	p.H1047R
			unknown	10	PTEN	89725077	nonsynonymous	C	T		c.C1060T	p.P354S
12224	H07-8844-1G	TUKO01099	clear cell	17	TP53	7577142	nonsynonymous	C	T		c.G796A	p.G266R
12225	13956/06 1S	TUKO01209	clear cell									
12226	14646-06 1H	TUKO01900	clear cell	7	EGFR	55266458	nonsynonymous	G	A		c.G2750A	p.G917E
			unknown	3	PIK3CA	178952085	nonsynonymous	A	G		c.A3140G	p.H1047R
12227	6865-07 1H	TUKO00763	mucinous	12	KRAS	25398284	nonsynonymous	C	T		c.G35A	p.G12D
			unknown	17	TP53	7578176	insertion				c.672+1G>T	
12228	H07-12969-1Q	TUKO01899	seromucinous	12	KRAS	25398284	nonsynonymous	C	G		c.G35C	p.G12A
12229	H07-14592-2K	TUKO01174	MMMT									
12230	15748-06 1P	TUKO00477	low grade serous	7	BRAF	140453136	nonsynonymous	A	T		c.T1799A	p.V600E
12231	10310/07 1G	TUKO01101	clear cell									
12232	13385/07 1C	TUKO01047	MMMT	17	TP53	7577115	nonsynonymous	A	C		c.T823G	p.C275G
			unknown	3	PIK3CA	178952085	nonsynonymous	A	G		c.A3140G	p.H1047R
12233	R08 3560 A2	TUKO01896	high grade serous	17	TP53	7578271	nonsynonymous	T	C		c.A578G	p.H193R
12234	17308-06 1G	TUKO00438	endometrioid	12	KRAS	25398285	nonsynonymous	C	T		c.G34A	p.G12S
			unknown	3	PIK3CA	178952085	nonsynonymous	A	G		c.A3140G	p.H1047R
12235	R08 03843 A1	TUKO01904	mucinous	17	TP53	7578535	nonsynonymous	T	C		c.A395G	p.K132R
12236	R08 4332 A1	No OTTA ID	mucinous	17	TP53	7578268	nonsynonymous	A	T		c.T581A	p.L194H
12237	R08-16687-2	TUKO01898	mucinous									
12238	07/32430 1J	TUKO01173	mucinous									
12239	08-6484-1M	TUKO01470	mucinous									
12240	H07-477-1E	TUKO01527	low grade serous	7	BRAF	140453136	nonsynonymous	A	T		c.T1799A	p.V600E
12241	H07-4056-1C	TUKO00620	serous grade not known	7	BRAF	140453136	nonsynonymous	A	T		c.T1799A	p.V600E
	H06-16635-2C	TUKO00574	high grade serous	17	TP53	7578406	nonsynonymous	C	T		c.G524A	p.R175H
	H07-803-2D	TUKO01186	unknown	17	TP53	7578469	nonsynonymous	C	A		c.G461T	p.G154V
	H07-16791-1C	TUKO01194	unknown	17	TP53	7577111	nonsynonymous	G	C		c.C827G	p.A276G
	2685/06 1A	TUKO00031	high grade serous	17	TP53	7577141	nonsynonymous	C	T		c.G797A	p.G266E
	14468/06 H	TUKO00520	unknown	17	TP53	7576928	insertion				c.920+1A>G	
	17106-06 1B	TUKO00895	high grade serous	17	TP53	7578210	synonymous	T	C		c.A639G	c.R213R
			unknown	17	TP53	7577094	nonsynonymous	G	A		c.C844T	p.R282W
			unknown	10	PTEN	89717708	stopgain	C	T		c.C733T	p.Q245*
	1567-07 1A	TUKO01032	high grade serous	17	TP53	7578262	nonsynonymous	C	G		c.G587C	p.R196P
	4870-07 1G	TUKO00847	high grade serous	17	TP53	7578188	stopgain	C	A		c.G661T	p.E221*
			unknown	10	PTEN	89711925	synonymous	G	A		c.G543A	p.L181L
	7547-07 1G	TUKO00827	high grade serous	17	TP53	7578271	nonsynonymous	T	C		c.A578G	p.H193R
	9803/07 1L	TUKO01106	high grade serous	12	KRAS	25398284	nonsynonymous	C	A		c.G35T	p.G12V
	10868/07 C	TUKO01084	high grade serous	17	TP53	7577124	nonsynonymous	C	T		c.G814A	p.V272M
	14708/07 2A	TUKO01094	unknown	17	TP53	7577574	nonsynonymous	T	G		c.A707C	p.Y236S
	17801/07 2H	TUKO01493	high grade serous	17	TP53	7574003	stop_gained	G	A		c.C1024T	p.R342*

12.7 De novo mutation calls for UKOPS plasma using TAm-Seq

SampleName	Plasma sample ID	FFPE block ID	OTTA ID	Type	Chr	Position	Ref	Mutant	Indel	Base	MAF	Matches FFPE Mutation
JBLAB_12250	116005	R07 12950 A1	No OTTA ID	sample								
JBLAB_12272	119949	R08 4332 A1	No OTTA ID	sample								
JBLAB_12285	114788	UR07 14789 A2	No OTTA ID	sample								
JBLAB_12292	121196	R08-02196-RES9	no OTTA ID	sample	17	7577091	G	A			0.200	No FFPE mutation
JBLAB_12292	121196	R08-02196-RES9	No OTTA ID	sample	12	25398237	A	G			0.175	No FFPE mutation
JBLAB_12250	116005	R07 12950 A1	No OTTA ID	library								
JBLAB_12272	119949	R08 4332 A1	No OTTA ID	library								
JBLAB_12285	114788	UR07 14789 A2	No OTTA ID	library	10	89717629	C	T			0.145	No FFPE sample
JBLAB_12285	114788	UR07 14789 A2	No OTTA ID	library	10	89653817	G	T			0.150	No FFPE sample
JBLAB_12285	114788	UR07 14789 A2	No OTTA ID	library	10	89692912	T	C			0.076	No FFPE sample
JBLAB_12292	121196	R08-02196-RES9	No OTTA ID	library	10	89720772	G	A			0.107	No FFPE mutation
JBLAB_12292	121196	R08-02196-RES9	No OTTA ID	library	17	7577091	G	A			0.269	No FFPE mutation
JBLAB_12292	121196	R08-02196-RES9	No OTTA ID	library	17	7579425	C	T			0.050	No FFPE mutation
JBLAB_12245	103618	2685/06 1A	TUKO00031	sample								
JBLAB_12245	103618	2685/06 1A	TUKO00031	library	17	7577141	C	T			0.090	Y
JBLAB_12245	103618	2685/06 1A	TUKO00031	library	17	7579580	G	A			0.043	N
JBLAB_12245	103618	2685/06 1A	TUKO00031	library	10	89711899	C	T			0.046	N
JBLAB_12246	102055	H06-10510-4A	TUKO00193	sample								
JBLAB_12246	102055	H06-10510-4A	TUKO00193	library	17	7578190	T	C			0.096	Y
JBLAB_12246	102055	H06-10510-4A	TUKO00193	library	17	7577649			INS	G	0.062	N
JBLAB_12246	102055	H06-10510-4A	TUKO00193	library	17	7578546			INS	G	0.012	N
JBLAB_12246	102055	H06-10510-4A	TUKO00193	library	10	89717718			INS	TG	0.056	N
JBLAB_12266	109163	17308-06 1G	TUKO00438	sample	3	178952085	A	G			0.135	Y
JBLAB_12266	109163	17308-06 1G	TUKO00438	sample	12	25398285	C	T			0.087	Y
JBLAB_12266	109163	17308-06 1G	TUKO00438	library	3	178952085	A	G			0.069	Y
JBLAB_12266	109163	17308-06 1G	TUKO00438	library	12	25398285	C	T			0.071	Y
JBLAB_12266	109163	17308-06 1G	TUKO00438	library	17	7577057			INS	C	0.009	N
JBLAB_12266	109163	17308-06 1G	TUKO00438	library	10	89653820			INS	A	0.017	N
JBLAB_12263	107979	15748-06 1P	TUKO00477	sample								
JBLAB_12263	107979	15748-06 1P	TUKO00477	library								
JBLAB_12258	102174	H06-16635-2C	TUKO00574	sample								
JBLAB_12258	102174	H06-16635-2C	TUKO00574	library	10	89692940			INS	G	0.175	N
JBLAB_12277	112211	R07-4984A	TUKO00579	sample	17	7579546			DEL	G	0.020	No FFPE mutation
JBLAB_12277	112211	R07-4984A	TUKO00579	library	17	7579554	G	A			0.030	No FFPE mutation
JBLAB_12278	112347	R07-6470-RESA11	TUKO00595	sample								
JBLAB_12278	112347	R07-6470-RESA11	TUKO00595	library								
JBLAB_12268	102357	H07-4056-1C	TUKO00620	sample								
JBLAB_12268	102357	H07-4056-1C	TUKO00620	library	12	25398237	A	G			0.226	N
JBLAB_12279	109878	6865-07 1H	TUKO00763	sample								
JBLAB_12279	109878	6865-07 1H	TUKO00763	library	10	89720683	C	G			0.035	N
JBLAB_12280	109898	7547-07 1G	TUKO00827	sample								
JBLAB_12280	109898	7547-07 1G	TUKO00827	library	17	7572968			INS	A	0.051	N
JBLAB_12276	109677	4870-07 1G	TUKO00847	sample	17	7578188	C	A			0.116	Y
JBLAB_12276	109677	4870-07 1G	TUKO00847	library	17	7578188	C	A			0.100	Y
JBLAB_12260	102197	H06-17939-2E	TUKO00876	sample	17	7578271	T	C			0.409	Y
JBLAB_12260	102197	H06-17939-2E	TUKO00876	library	17	7578271	T	C			0.531	Y
JBLAB_12260	102197	H06-17939-2E	TUKO00876	library	10	89711903			INS	TG	0.028	N
JBLAB_12264	102235	H06-19212-3C	TUKO00877	sample								
JBLAB_12264	102235	H06-19212-3C	TUKO00877	library	17	7573936			INS	C	0.020	No FFPE mutation
JBLAB_12265	109130	17106-06 1B	TUKO00895	sample								
JBLAB_12265	109130	17106-06 1B	TUKO00895	library	17	7577094	G	A			0.041	Y
JBLAB_12265	109130	17106-06 1B	TUKO00895	library	12	25398234			INS	A	0.009	N
JBLAB_12265	109130	17106-06 1B	TUKO00895	library	10	89692782			INS	T	0.077	N
JBLAB_12269	109350	1567-07 1A	TUKO01032	sample	17	7578262	C	G			0.232	Y
JBLAB_12269	109350	1567-07 1A	TUKO01032	library	17	7578262	C	G			0.217	Y
JBLAB_12269	109350	1567-07 1A	TUKO01032	library	3	178936051			INS	A	0.020	N
JBLAB_12251	114379	13385/07 1C	TUKO01047	sample	17	7577115	A	C			0.263	Y
JBLAB_12251	114379	13385/07 1C	TUKO01047	library								
JBLAB_12290	114186	10868/07 C	TUKO01084	sample	12	25398237	A	G			0.236	N
JBLAB_12290	114186	10868/07 C	TUKO01084	library	17	7579546			INS	G	0.033	N
JBLAB_12248	113029	H07-8844-1G	TUKO01099	sample								
JBLAB_12248	113029	H07-8844-1G	TUKO01099	library								
JBLAB_12282	114081	10310/07 1G	TUKO01101	sample								
JBLAB_12282	114081	10310/07 1G	TUKO01101	library								
JBLAB_12281	114085	9803/07 1L	TUKO01106	sample								
JBLAB_12281	114085	9803/07 1L	TUKO01106	library								
JBLAB_12261	102184	H06-16736-3G	TUKO01133	sample								
JBLAB_12261	102184	H06-16736-3G	TUKO01133	library								
JBLAB_12294	111798	07/32430 1J	TUKO01173	sample								
JBLAB_12294	111798	07/32430 1J	TUKO01173	library								
JBLAB_12284	113109	H07-14592-2K	TUKO01174	sample	17	7578290	C	T			0.111	No FFPE mutation

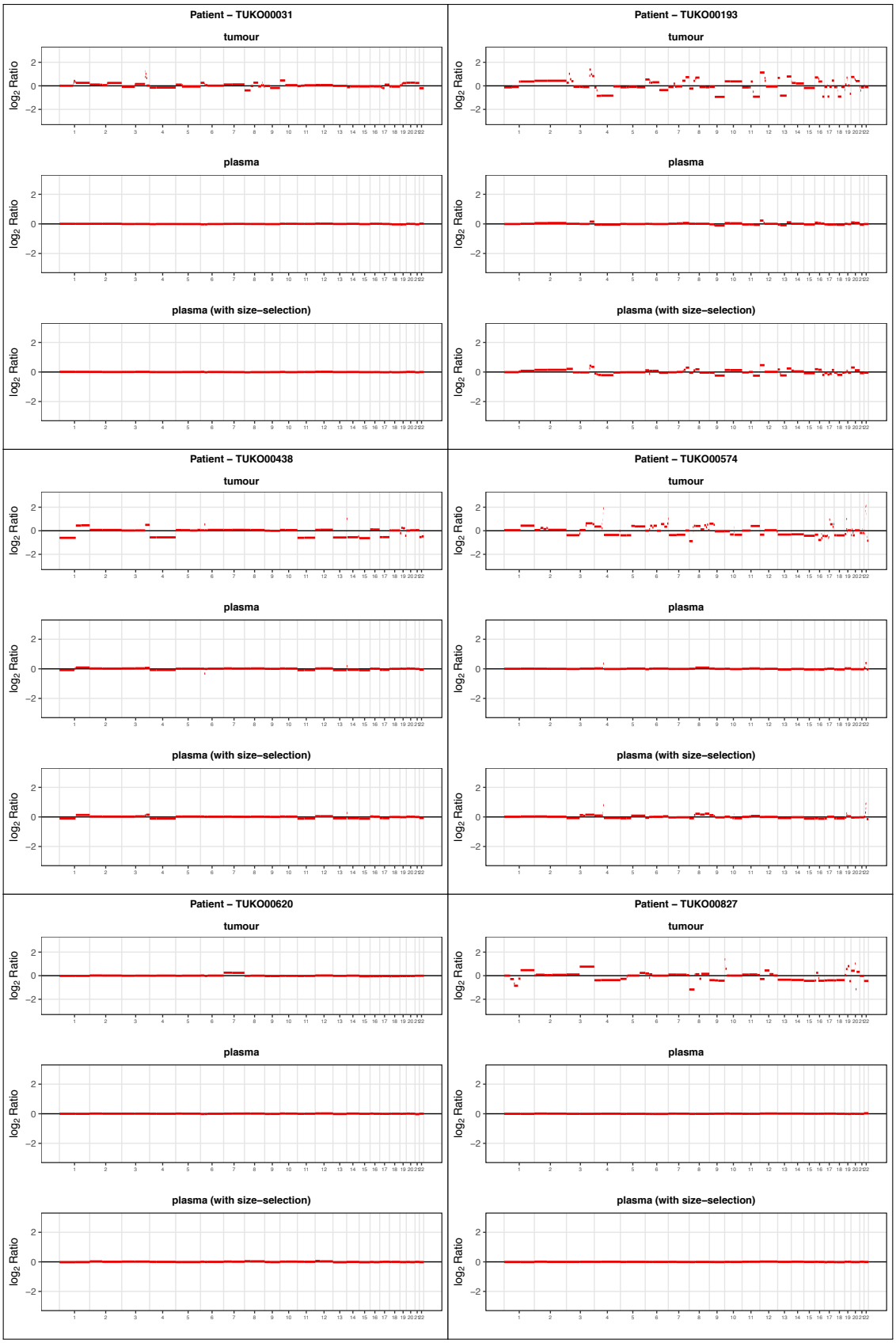
SampleName	Plasma sample ID	FFPE block ID	OTTA ID	Type	Chr	Position	Ref	Mutant	Indel	Base	MAF	Matches FFPE Mutation
JBLAB_12284	113109	H07-14592-2K	TUKO01174	library	17	7578290	C	T			0.075	No FFPE mutation
JBLAB_12284	113109	H07-14592-2K	TUKO01174	library	17	7573966			INS	G	0.034	No FFPE mutation
JBLAB_12293	113157	H07-18627-1B	TUKO01208	sample	17	7577108	C	A			0.084	Y
JBLAB_12293	113157	H07-18627-1B	TUKO01208	library	17	7577108	C	A			0.240	Y
JBLAB_12257	107793	13956/06 1S	TUKO01209	sample								
JBLAB_12257	107793	13956/06 1S	TUKO01209	library	12	25398262			DEL	A	0.016	No FFPE mutation
JBLAB_12259	102194	H06-18278-F	TUKO01211	sample								
JBLAB_12259	102194	H06-18278-F	TUKO01211	library	17	7578464	G	A			0.079	N
JBLAB_12259	102194	H06-18278-F	TUKO01211	library	10	89725078	C	T			0.062	N
JBLAB_12259	102194	H06-18278-F	TUKO01211	library	10	89624308			DEL	T	0.057	N
JBLAB_12255	116478	R07-18229-A1	TUKO01222	sample	10	89717628	G	A			0.441	N
JBLAB_12255	116478	R07-18229-A1	TUKO01222	library								
JBLAB_12249	112196	R07-11403-REMA2	TUKO01230	sample								
JBLAB_12249	112196	R07-11403-REMA2	TUKO01230	library								
JBLAB_12286	118518	07 37790 1H	TUKO01468	sample	12	25398237	A	G			0.232	N
JBLAB_12286	118518	07 37790 1H	TUKO01468	library	10	89692844			DEL	AA	0.267	N
JBLAB_12253	118607	08-6484-1M	TUKO01470	sample								
JBLAB_12253	118607	08-6484-1M	TUKO01470	library	17	7577140			INS	C	0.075	No FFPE mutation
JBLAB_12287	117956	17801/07 2H	TUKO01493	sample								
JBLAB_12287	117956	17801/07 2H	TUKO01493	library								
JBLAB_12256	117896	17028/07 2A	TUKO01510	sample								
JBLAB_12256	117896	17028/07 2A	TUKO01510	library	17	7579850			DEL	T	0.044	N
JBLAB_12267	102274	H07-477-1E	TUKO01527	sample								
JBLAB_12267	102274	H07-477-1E	TUKO01527	library	10	89653851			INS	TGA	0.055	N
JBLAB_12254	118512	07 36137 1G	TUKO01547	sample								
JBLAB_12254	118512	07 36137 1G	TUKO01547	library	17	7579409			DEL	G	0.009	N
JBLAB_12252	111751	07/26988 2O	TUKO01552	sample								
JBLAB_12252	111751	07/26988 2O	TUKO01552	library	10	89717630	C	T			0.220	N
JBLAB_12252	111751	07/26988 2O	TUKO01552	library	17	7577583			DEL	G	0.013	N
JBLAB_12252	111751	07/26988 2O	TUKO01552	library	10	89685226			INS	T	0.098	N
JBLAB_12252	111751	07/26988 2O	TUKO01552	library	10	89720680			DEL	T	0.056	N
JBLAB_12289	116355	R07-20333-3A	TUKO01594	sample								
JBLAB_12289	116355	R07-20333-3A	TUKO01594	library	10	89711901	C	T			0.523	N
JBLAB_12289	116355	R07-20333-3A	TUKO01594	library	10	89690895	C	T			0.168	N
JBLAB_12288	116379	UH07-20421-2	TUKO01595	sample	17	7578413	C	A			0.146	No FFPE mutation
JBLAB_12288	116379	UH07-20421-2	TUKO01595	library	17	7578413	C	A			0.104	No FFPE mutation
JBLAB_12283	123230	UR08-11153 A 1	TUKO01891	sample	17	7577538	C	T			0.121	Y
JBLAB_12283	123230	UR08-11153 A 1	TUKO01891	library								
JBLAB_12271	123253	R08-14573-2	TUKO01892	sample								
JBLAB_12271	123253	R08-14573-2	TUKO01892	library								
JBLAB_12247	107615	10758/06 2J	TUKO01894	sample	17	7578271	T	C			0.237	Y
JBLAB_12247	107615	10758/06 2J	TUKO01894	library	17	7578271	T	C			0.195	Y
JBLAB_12275	119975	R08 3560 A2	TUKO01896	sample	10	89725053	T	C			0.026	N
JBLAB_12275	119975	R08 3560 A2	TUKO01896	library								
JBLAB_12274	123195	R08 05206 10	TUKO01897	sample	17	7577498	C	A			0.291	No FFPE mutation
JBLAB_12274	123195	R08 05206 10	TUKO01897	library	17	7577498	C	A			0.390	No FFPE mutation
JBLAB_12274	123195	R08 05206 10	TUKO01897	library	17	7572995			DEL	G	0.040	No FFPE mutation
JBLAB_12274	123195	R08 05206 10	TUKO01897	library	10	89692981			INS	G	0.523	No FFPE mutation
JBLAB_12270	123277	R08-16687-2	TUKO01898	sample	12	25398237	A	G			0.170	No FFPE mutation
JBLAB_12270	123277	R08-16687-2	TUKO01898	library								
JBLAB_12291	113084	H07-12969-1Q	TUKO01899	sample								
JBLAB_12291	113084	H07-12969-1Q	TUKO01899	library	17	7576835	G	A			0.047	N
JBLAB_12291	113084	H07-12969-1Q	TUKO01899	library	10	89692797			INS	C	0.030	N
JBLAB_12262	107895	14646-06 1H	TUKO01900	sample								
JBLAB_12262	107895	14646-06 1H	TUKO01900	library	10	89720775	C	T			0.465	N
JBLAB_12262	107895	14646-06 1H	TUKO01900	library	17	7577118			DEL	A	0.018	N
JBLAB_12273	122982	R08 03843 A1	TUKO01904	sample								
JBLAB_12273	122982	R08 03843 A1	TUKO01904	library	17	7572990			INS	T	0.082	N

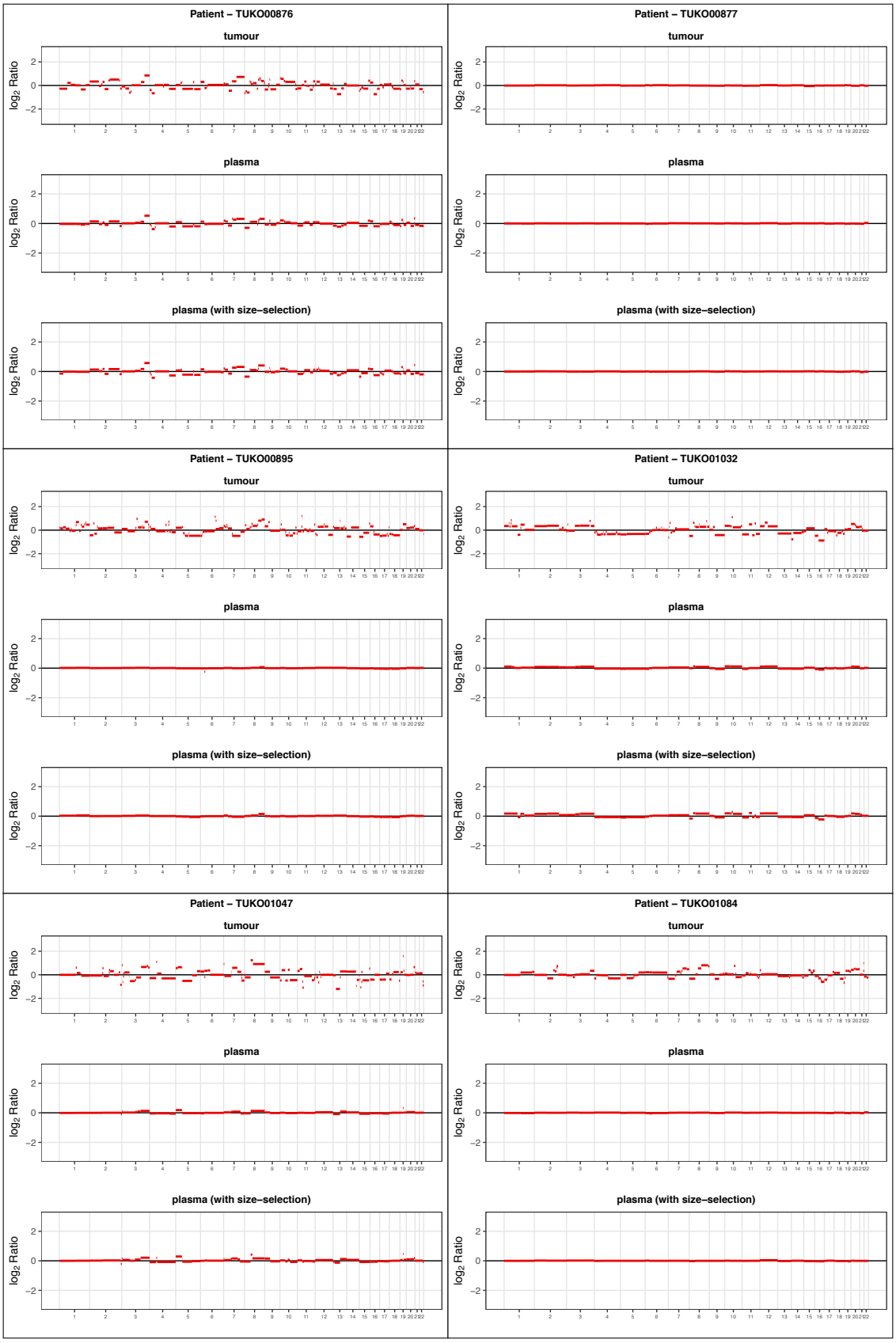
12.8 Specific variant mutation calls for UKOPS plasma using TAm-Seq

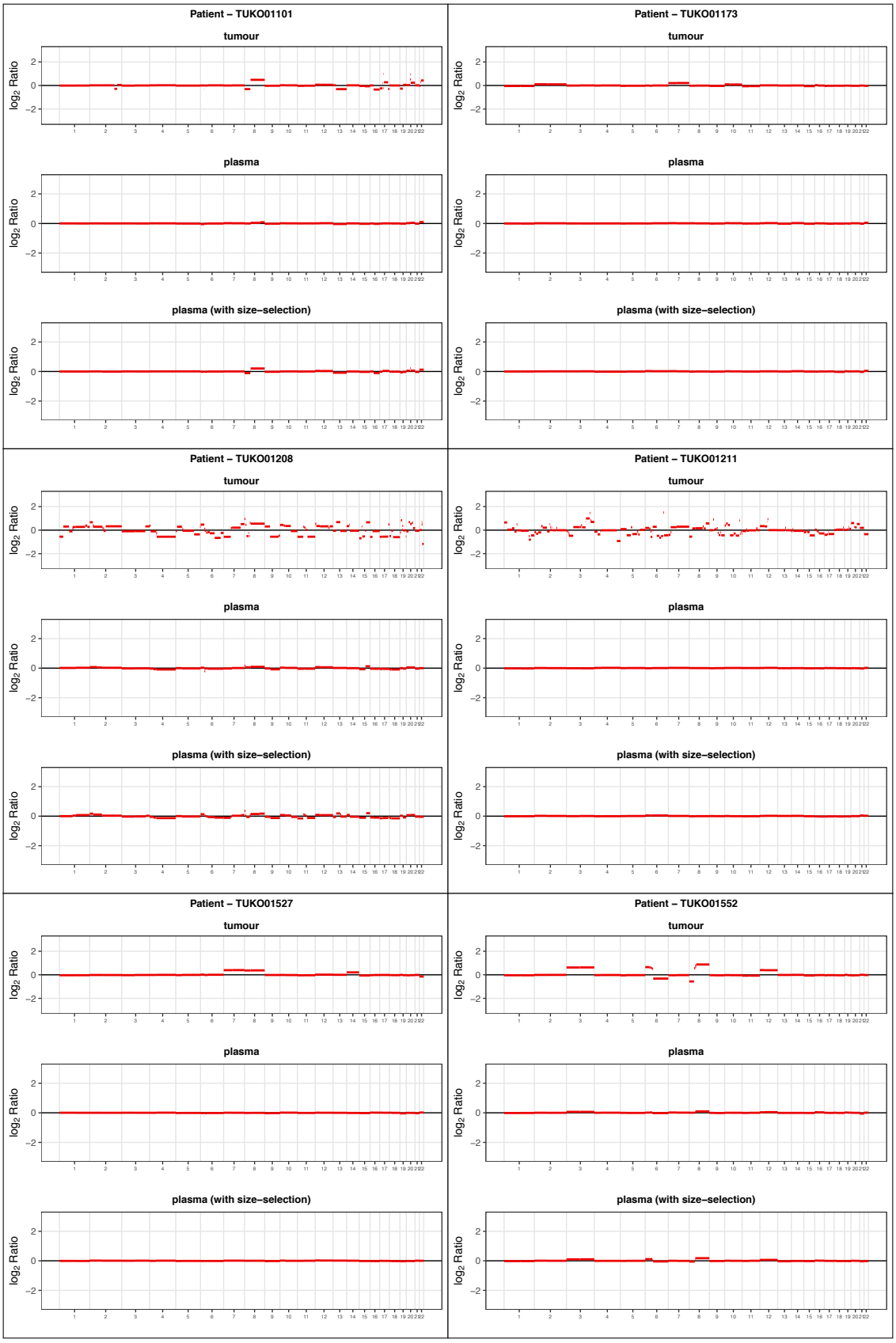
Sample Name	Patient Id	Type	Chr	Position	Ref	Mutant	Barcode_1	Barcode_2	MAF	Error	Detected
JBLAB_12245	TUKO00031	sample	17	7577141	C	T	FLD0097	FLD0121	0.001361	0.001381789	FALSE
JBLAB_12245	TUKO00031	library	17	7577141	C	T	FLD0203	FLD0227	0.090944	0.001138821	TRUE
JBLAB_12246	TUKO00193	sample	17	7578190	T	C	FLD0098	FLD0122	0.1164595	0.002834168	TRUE
JBLAB_12246	TUKO00193	library	17	7578190	T	C	FLD0204	FLD0228	0.098622	0.002855301	TRUE
JBLAB_12247	TUKO01894	sample	17	7578271	T	C	FLD0099	FLD0123	0.23786425	0.000175222	TRUE
JBLAB_12247	TUKO01894	library	17	7578271	T	C	FLD0205	FLD0229	0.194914	0.00017485	TRUE
JBLAB_12248	TUKO01099	sample	17	7577142	C	T	FLD0100	FLD0124	0.01486025	0.001600774	TRUE
JBLAB_12248	TUKO01099	library	17	7577142	C	T	FLD0206	FLD0230	0.010095	0.001638061	FALSE
JBLAB_12251	TUKO01047	sample	17	7577115	A	C	FLD0103	FLD0127	0.26347	0.000227617	TRUE
JBLAB_12251	TUKO01047	sample	3	178952085	A	G	FLD0103	FLD0127	0.26347	0.000541479	TRUE
JBLAB_12251	TUKO01047	library	17	7577115	A	C	FLD0209	FLD0233	0.18018	0.000238161	TRUE
JBLAB_12251	TUKO01047	library	3	178952085	A	G	FLD0209	FLD0233	0.18018	0.000325236	TRUE
JBLAB_12252	TUKO01552	sample	12	25398284	C	T	FLD0104	FLD0128	0.0491965	0.001408396	TRUE
JBLAB_12252	TUKO01552	library	12	25398284	C	T	FLD0210	FLD0234	0.007888	0.001359644	TRUE
JBLAB_12254	TUKO01547	sample	3	178952085	A	T	FLD0106	FLD0130	0.0002495	0.000541479	FALSE
JBLAB_12254	TUKO01547	library	3	178952085	A	T	FLD0212	FLD0236	0.000183	0.000325236	FALSE
JBLAB_12255	TUKO01222	sample	17	7578433	G	A	FLD0107	FLD0131	0.00039725	0.001572223	FALSE
JBLAB_12255	TUKO01222	library	17	7578433	G	A	FLD0213	FLD0237	0.000743	0.001031628	FALSE
JBLAB_12256	TUKO01510	sample	17	7577570	C	T	FLD0108	FLD0132	0.03659475	0.00123582	TRUE
JBLAB_12256	TUKO01510	library	17	7577570	C	T	FLD0214	FLD0238	0.000872	0.001076814	FALSE
JBLAB_12258	TUKO00574	sample	17	7578406	C	T	FLD0110	FLD0134	0.01329525	0.002254892	TRUE
JBLAB_12258	TUKO00574	library	17	7578406	C	T	FLD0241	FLD0265	0.00298	0.003091521	FALSE
JBLAB_12259	TUKO01211	sample	17	7578271	T	G	FLD0111	FLD0135	0.00005775	0.000175222	FALSE
JBLAB_12259	TUKO01211	library	17	7578271	T	G	FLD0242	FLD0266	0.000044	0.00017485	FALSE
JBLAB_12260	TUKO00876	sample	17	7578271	T	C	FLD0112	FLD0136	0.40903625	0.000175222	TRUE
JBLAB_12260	TUKO00876	library	17	7578271	T	C	FLD0243	FLD0267	0.545642	0.00017485	TRUE
JBLAB_12261	TUKO01133	sample	7	55259529	G	A	FLD0113	FLD0137	0.0474065	0	TRUE
JBLAB_12261	TUKO01133	sample	17	7578493	C	T	FLD0113	FLD0137	0.0474065	0.001108214	TRUE
JBLAB_12261	TUKO01133	library	7	55259529	G	A	FLD0244	FLD0268	0.035563	0	TRUE
JBLAB_12261	TUKO01133	library	17	7578493	C	T	FLD0244	FLD0268	0.035563	0.001110341	TRUE
JBLAB_12262	TUKO01900	sample	7	55266458	G	A	FLD0114	FLD0138	0.016153	0	TRUE
JBLAB_12262	TUKO01900	sample	3	178952085	A	G	FLD0114	FLD0138	0.016153	0.000541479	TRUE
JBLAB_12262	TUKO01900	library	7	55266458	G	A	FLD0245	FLD0269	0.004137	0	TRUE
JBLAB_12262	TUKO01900	library	3	178952085	A	G	FLD0245	FLD0269	0.004137	0.000325236	TRUE
JBLAB_12263	TUKO00477	sample	7	140453136	A	T	FLD0115	FLD0139	0.00036825	0.000247788	TRUE
JBLAB_12263	TUKO00477	library	7	140453136	A	T	FLD0246	FLD0270	0.00021	0.00018023	TRUE
JBLAB_12265	TUKO00895	sample	17	7578210	T	C	FLD0117	FLD0141	0.52998175	0.004167561	TRUE
JBLAB_12265	TUKO00895	sample	17	7577094	G	A	FLD0117	FLD0141	0.52998175	0.003816157	TRUE
JBLAB_12265	TUKO00895	sample	10	89717708	C	T	FLD0117	FLD0141	0.52998175	0.001065201	TRUE
JBLAB_12265	TUKO00895	library	17	7578210	T	C	FLD0248	FLD0272	0.523769	0.003644538	TRUE
JBLAB_12265	TUKO00895	library	17	7577094	G	A	FLD0248	FLD0272	0.523769	0.003610661	TRUE
JBLAB_12265	TUKO00895	library	10	89717708	C	T	FLD0248	FLD0272	0.523769	0.000938832	TRUE
JBLAB_12266	TUKO00438	sample	12	25398285	C	T	FLD0118	FLD0142	0.087477	0.001765522	TRUE
JBLAB_12266	TUKO00438	sample	3	178952085	A	G	FLD0118	FLD0142	0.087477	0.000541479	TRUE

Sample Name	Patient Id	Type	Chr	Position	Ref	Mutant	Barcode_1	Barcode_2	MAF	Error	Detected
JBLAB_12266	TUKO00438	library	12	25398285	C	T	FLD0249	FLD0273	0.065399	0.002083602	TRUE
JBLAB_12266	TUKO00438	library	3	178952085	A	G	FLD0249	FLD0273	0.065399	0.000325236	TRUE
JBLAB_12267	TUKO01527	sample	7	140453136	A	T	FLD0119	FLD0143	0.00009475	0.000247788	FALSE
JBLAB_12267	TUKO01527	library	7	140453136	A	T	FLD0250	FLD0274	0.000179	0.00018023	FALSE
JBLAB_12268	TUKO00620	sample	7	140453136	A	T	FLD0145	FLD0169	0.00036825	0.000247788	TRUE
JBLAB_12268	TUKO00620	library	7	140453136	A	T	FLD0251	FLD0275	0.000676	0.00018023	TRUE
JBLAB_12269	TUKO01032	sample	17	7578262	C	G	FLD0146	FLD0170	0.23182725	0.000202268	TRUE
JBLAB_12269	TUKO01032	library	17	7578262	C	G	FLD0252	FLD0276	0.21737	0.000147171	TRUE
JBLAB_12271	TUKO01892	sample	17	7578461	C	A	FLD0148	FLD0172	0	3.68701E-05	FALSE
JBLAB_12271	TUKO01892	library	17	7578461	C	A	FLD0254	FLD0278	0.000326	3.30116E-05	TRUE
JBLAB_12272	R08 4332 A1	sample	17	7578268	A	T	FLD0149	FLD0173	0.00048325	0.000651232	FALSE
JBLAB_12272	R08 4332 A1	library	17	7578268	A	T	FLD0255	FLD0279	0.000787	0.000548336	TRUE
JBLAB_12273	TUKO01904	sample	17	7578535	T	C	FLD0150	FLD0174	0.01515	0.005395687	TRUE
JBLAB_12273	TUKO01904	library	17	7578535	T	C	FLD0256	FLD0280	0.003603	0.004871951	FALSE
JBLAB_12275	TUKO01896	sample	17	7578271	T	C	FLD0152	FLD0176	0.00617375	0.000175222	TRUE
JBLAB_12275	TUKO01896	library	17	7578271	T	C	FLD0258	FLD0282	0.021145	0.00017485	TRUE
JBLAB_12276	TUKO00847	sample	17	7578188	C	A	FLD0153	FLD0177	0.1169525	5.62495E-05	TRUE
JBLAB_12276	TUKO00847	sample	10	89711925	G	A	FLD0153	FLD0177	0.1169525	0.002322851	TRUE
JBLAB_12276	TUKO00847	library	17	7578188	C	A	FLD0259	FLD0283	0.098876	4.58496E-05	TRUE
JBLAB_12276	TUKO00847	library	10	89711925	G	A	FLD0259	FLD0283	0.098876	0.002104682	TRUE
JBLAB_12279	TUKO00763	sample	12	25398284	C	T	FLD0156	FLD0180	0.00158875	0.001408396	TRUE
JBLAB_12279	TUKO00763	sample	17	7578176			FLD0156	FLD0180	0.00158875	#N/A	FALSE
JBLAB_12279	TUKO00763	sample	17	7578176	INS		1				FALSE
JBLAB_12279	TUKO00763	library	12	25398284	C	T	FLD0262	FLD0286	0.001071	0.001359644	FALSE
JBLAB_12279	TUKO00763	library	17	7578176			FLD0262	FLD0286	0.001071	0	TRUE
JBLAB_12279	TUKO00763	library	17	7578176	INS		1				FALSE
JBLAB_12280	TUKO00827	sample	17	7578271	T	C	FLD0157	FLD0181	0.0045595	0.000175222	TRUE
JBLAB_12280	TUKO00827	library	17	7578271	T	C	FLD0263	FLD0287	0.003685	0.00017485	TRUE
JBLAB_12281	TUKO01106	sample	12	25398284	C	A	FLD0158	FLD0182	0.000099	0.001408396	FALSE
JBLAB_12281	TUKO01106	library	12	25398284	C	A	FLD0289	FLD0313	0.000581	0.001359644	FALSE
JBLAB_12283	TUKO01891	sample	17	7577538	C	T	FLD0160	FLD0184	0.12103725	0.002176205	TRUE
JBLAB_12283	TUKO01891	library	17	7577538	C	T	FLD0291	FLD0315	0.064318	0.002722134	TRUE
JBLAB_12286	TUKO01468	sample	3	178952085	A	G	FLD0163	FLD0187	0.01175625	0.000541479	TRUE
JBLAB_12286	TUKO01468	sample	10	89725077	C	T	FLD0163	FLD0187	0.01175625	0.001736959	TRUE
JBLAB_12286	TUKO01468	library	3	178952085	A	G	FLD0294	FLD0318	0.004957	0.000325236	TRUE
JBLAB_12286	TUKO01468	library	10	89725077	C	T	FLD0294	FLD0318	0.004957	0.001097507	TRUE
JBLAB_12287	TUKO01493	sample	17	7574003	G	A	FLD0164	FLD0188	0.03223925	0.002236027	TRUE
JBLAB_12287	TUKO01493	library	17	7574003	G	A	FLD0295	FLD0319	0.035589	0.001918542	TRUE
JBLAB_12289	TUKO01594	sample	17	7578406	C	T	FLD0166	FLD0190	0.00654775	0.002254892	TRUE
JBLAB_12289	TUKO01594	library	17	7578406	C	T	FLD0297	FLD0321	0.001732	0.003091521	FALSE
JBLAB_12290	TUKO01084	sample	17	7577124	C	T	FLD0167	FLD0191	0.002137	0.001586276	TRUE
JBLAB_12290	TUKO01084	library	17	7577124	C	T	FLD0298	FLD0322	0.002153	0.001595454	TRUE
JBLAB_12291	TUKO01899	sample	12	25398284	C	G	FLD0194	FLD0218	0.00191025	0.001408396	TRUE
JBLAB_12291	TUKO01899	library	12	25398284	C	G	FLD0299	FLD0323	0.000106	0.001359644	FALSE
JBLAB_12293	TUKO01208	sample	17	7577108	C	A	FLD0196	FLD0220	0.0846355	0.000117562	TRUE
JBLAB_12293	TUKO01208	library	17	7577108	C	A	FLD0301	FLD0325	0.262262	8.96445E-05	TRUE

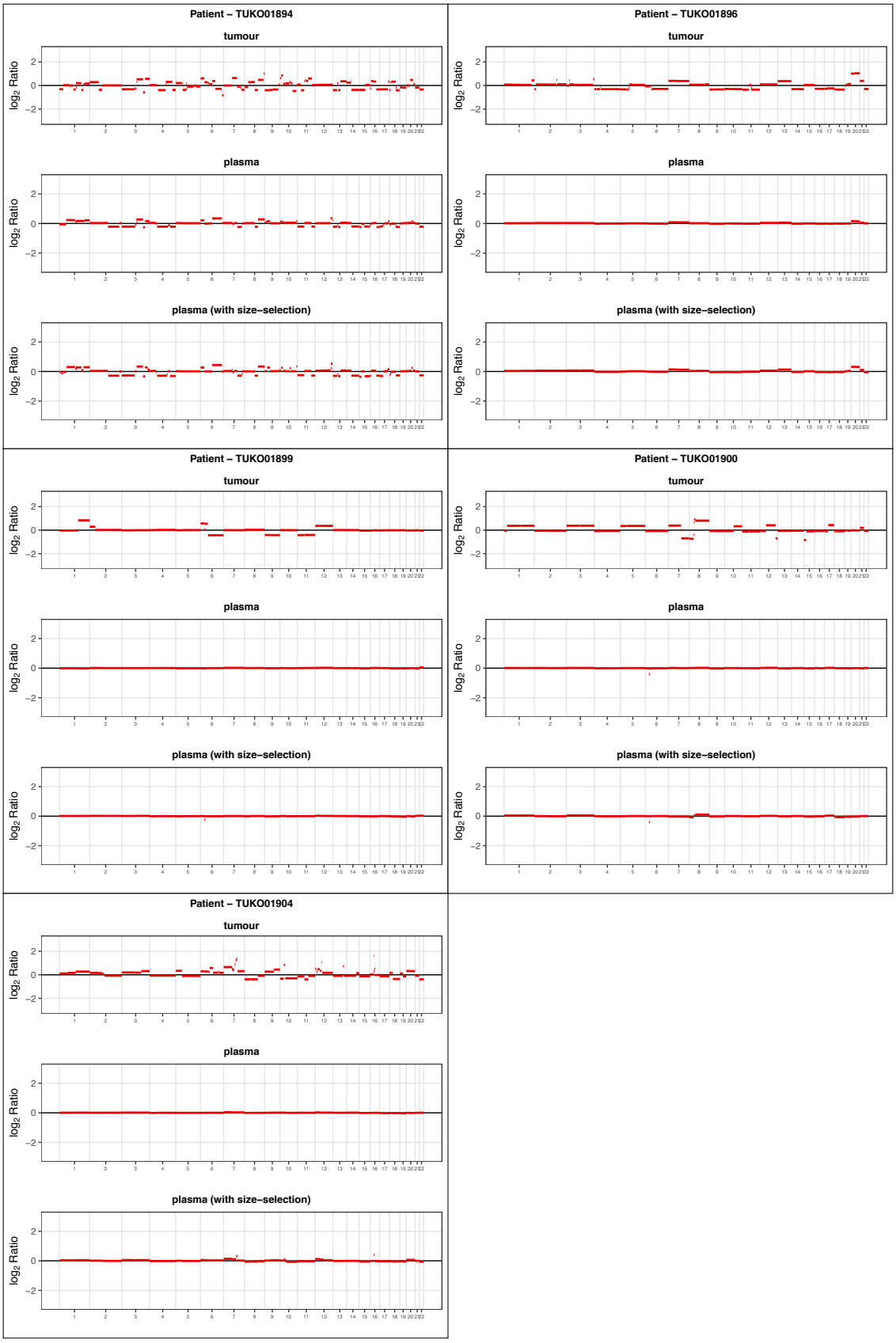
12.9 UKOPS copy number profiles: tumour, unselected plasma and size selected plasma




Patient – TUKO00876
tumour
plasma
plasma (with size–selection)
Patient – TUKO00877
tumour
plasma
plasma (with size–selection)
Patient – TUKO00895
tumour
plasma
plasma (with size–selection)
Patient – TUKO01032
tumour
plasma
plasma (with size–selection)
Patient – TUKO01047
tumour
plasma
plasma (with size–selection)
Patient – TUKO01084
tumour
plasma
plasma (with size–selection)
log2 Ratio


Patient – TUKO01101
tumour
plasma
plasma (with size–selection)
Patient – TUKO01173
tumour
plasma
plasma (with size–selection)
Patient – TUKO01208
tumour
plasma
plasma (with size–selection)
Patient – TUKO01211
tumour
plasma
plasma (with size–selection)
Patient – TUKO01527
tumour
plasma
plasma (with size–selection)
Patient – TUKO01552
tumour
plasma
plasma (with size–selection)
log2 Ratio

lv


Patient – TUKO01894
tumour
plasma
plasma (with size–selection)
Patient – TUKO01896
tumour
plasma
plasma (with size–selection)
Patient – TUKO01899
tumour
plasma
plasma (with size–selection)
Patient – TUKO01900
tumour
plasma
plasma (with size–selection)
Patient – TUKO01904
tumour
plasma
plasma (with size–selection)
log2 Ratio

12.10 De novo mutation calls for CTCR-OV04 plasma using TAm-Seq

OV04	Treatment	Stage	Plasma sample	Volume plasma (ml)	Extraction method	Chromosome	Gene	Position	Effect	Ref	Mutant	Indel	cDNA effect	Protein effect	MAF	Matches FFPE mutation
34	neo-adjuvant chemotherapy	3C	JBLAB-16614	1.5	QIASymphony Circulating DNA (2ml FIX)											
38	neo-adjuvant chemotherapy	4	JBLAB-16671	3.1	QIASymphony Circulating DNA (3ml FIX)											
72	neo-adjuvant chemotherapy	4	JBLAB-16616	1.4	QIASymphony Circulating DNA (2ml FIX)											
74	neo-adjuvant chemotherapy	3C	JBLAB-16621	1.1	QIASymphony Circulating DNA (2ml FIX)	17	TP53	7578526	nonsynonymous	C	T		c.G404A	p.C135Y	0.071	Y
75	neo-adjuvant chemotherapy	3C	JBLAB-16622	0.9	QIASymphony Circulating DNA (2ml FIX)											
95	neo-adjuvant chemotherapy	1C_1	JBLAB-16625	0.7	QIASymphony Circulating DNA (2ml FIX)											
103	neo-adjuvant chemotherapy	3	JBLAB-16627	1.5	QIASymphony Circulating DNA (2ml FIX)											
127	neo-adjuvant chemotherapy	3C	JBLAB-16628	0.9	QIASymphony Circulating DNA (2ml FIX)											
133	neo-adjuvant chemotherapy	1C_1	JBLAB-16629	0.7	QIASymphony Circulating DNA (2ml FIX)											
143	neo-adjuvant chemotherapy	3C	JBLAB-5583	1.9	QIASymphony Circulating DNA (3ml)	17	TP53	7574034	splicing	C	G				0.015	No FFPE sample
167	neo-adjuvant chemotherapy	3C	JBLAB-16676	2.55	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7577539	nonsynonymous	G	A		c.C742T	p.R248W	0.008	Y
176	primary surgery	2B	JBLAB-5374	2.2	manual											
178	neo-adjuvant chemotherapy	4	JBLAB-5591	3	QiaSymphony Circulating DNA (3ml)	17	TP53	7578517	nonsynonymous	G	A		c.C413T	p.A138V	0.011	No FFPE sample
188	neo-adjuvant chemotherapy	3C	JBLAB-16680	3.1	QIASymphony Circulating DNA (3ml FIX)											
191	neo-adjuvant chemotherapy	3C	JBLAB-16632	2.1	QIASymphony Circulating DNA (2ml FIX)											
194	neo-adjuvant chemotherapy	3C	JBLAB-16635	0.7	QIASymphony Circulating DNA (2ml FIX)											
197	primary surgery	3c	JBLAB-18153	2.6	QIASymphony Circulating DNA (3ml FIX)											
217	neo-adjuvant chemotherapy	4	JBLAB-15643	3	QIASymphony Circulating DNA (3ml)	17	TP53	7578204	nonsynonymous	A	C		c.T645G	p.S215R	0.028	Y
249	neo-adjuvant chemotherapy	4	JBLAB-16683	2.55	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7578474	DEL			GGGG	c.453_456_del	p.151_152del	0.171	No FFPE mutation
254	neo-adjuvant chemotherapy	3C	JBLAB-16640	1.15	QIASymphony Circulating DNA (2ml FIX)											
259	neo-adjuvant chemotherapy	3C	JBLAB-16642	1.7	QIASymphony Circulating DNA (2ml FIX)											
260	neo-adjuvant chemotherapy	3C	JBLAB-16645	1.8	QIASymphony Circulating DNA (2ml FIX)	17	TP53	7577106	nonsynonymous	G	A		c.C832T	p.P278S	0.160	Y
263	neo-adjuvant chemotherapy	3C	JBLAB-16687	3.1	QIASymphony Circulating DNA (3ml FIX)			7578454	nonsynonymous	G	A		c.C476T	p.A159V	0.008	N
263	neo-adjuvant chemotherapy	3C	JBLAB-16687	3.1	QIASymphony Circulating DNA (3ml FIX)			7577557	DEL			G	c.724delT	p.242fs	0.039	N
266	neo-adjuvant chemotherapy	4	JBLAB-15661	3	QIASymphony Circulating DNA (3ml)	17	TP53	7578542	nonsynonymous	G	C		c.C388G	p.L130V	0.077	Y
278	neo-adjuvant chemotherapy	4	JBLAB-16690	2.55	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7579426	DEL			G	c.261delA	p.P87fs	0.017	No FFPE sample
280	neo-adjuvant chemotherapy	3	JBLAB-5638	3	QIASymphony Circulating DNA (3ml)	17	TP53	7579471	DEL			C	c.216delC	p.P72fs	0.025	N
294	neo-adjuvant chemotherapy	4	JBLAB-5744	2.9	QIASymphony Circulating DNA (3ml)	17	TP53	7577539	nonsynonymous	G	A		c.C742T	p.R248W	0.015	N
295	neo-adjuvant chemotherapy	3C	JBLAB-16648	0.8	QIASymphony Circulating DNA (2ml FIX)											
314	neo-adjuvant chemotherapy	3C	JBLAB-16651	1.85	QIASymphony Circulating DNA (2ml FIX)	17	TP53	7578190	nonsynonymous	T	C		c.A659G	p.Y220C	0.021	Y
316	neo-adjuvant chemotherapy	4	JBLAB-5756	3	QIASymphony Circulating DNA (3ml)											
317	neo-adjuvant chemotherapy	3C	JBLAB-5762	3	QIASymphony Circulating DNA (3ml)											
318	neo-adjuvant chemotherapy	4A	JBLAB-16692	2.6	QIASymphony Circulating DNA (3ml FIX)											
319	neo-adjuvant chemotherapy	3C	JBLAB-16695	3.1	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7576852	splicing	C	T				0.035	Y
328	neo-adjuvant chemotherapy	3C	JBLAB-5520	2.2	manual	17	TP53	7578239	DEL			CA	c.609_610del	p.203_204del	0.010	Y
333	neo-adjuvant chemotherapy	4	JBLAB-16697	3	QIASymphony Circulating DNA (3ml FIX)											
334	neo-adjuvant chemotherapy	4	JBLAB-16655	0.85	QIASymphony Circulating DNA (2ml FIX)											
338	neo-adjuvant chemotherapy	4	JBLAB-16706	3	QIASymphony Circulating DNA (3ml FIX)											
341	neo-adjuvant chemotherapy	3	JBLAB-5070	1.6	manual											
345	neo-adjuvant chemotherapy	3C	JBLAB-16703	3.1	QIASymphony Circulating DNA (3ml FIX)											
346	neo-adjuvant chemotherapy	4	JBLAB-5041	2.5	manual											
348	neo-adjuvant chemotherapy	4	JBLAB-5104	0.75	manual											
358	neo-adjuvant chemotherapy	3C	JBLAB-5055	2.4	manual	17	TP53	7577114	nonsynonymous	C	T		c.G824A	p.C275Y	0.025	N
359	neo-adjuvant chemotherapy	4	JBLAB-5087	1.6	manual											
362	neo-adjuvant chemotherapy	3C	JBLAB-16710	2.8	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7578413	nonsynonymous	C	A		c.G517T	p.V173L	0.006	Y
364	neo-adjuvant chemotherapy	4	JBLAB-5802	3	QIASymphony Circulating DNA (3ml)	17	TP53	7578266	nonsynonymous	T	A		c.A583T	p.I195F	0.030	Y
367	neo-adjuvant chemotherapy	3C	JBLAB-16657	1.7	QIASymphony Circulating DNA (2ml FIX)											
375	neo-adjuvant chemotherapy	4	JBLAB-15711	3	QIASymphony Circulating DNA (3ml)											
382	primary surgery	4	JBLAB-18145	2.5	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7578279	INS			G	c.	p.P190fs	0.166	Y
392	neo-adjuvant chemotherapy	3C	JBLAB-16716	3.1	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7573999	INS			CTCGGAACATCTCGAAGCG	c.1027_1028insCTCGGAACATCTCGAAGCG	p.E343fs	0.021	Y
394	neo-adjuvant chemotherapy	3C	JBLAB-16722	2.55	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7578404	nonsynonymous	A	T		c.T526A	p.C176S	0.017	No FFPE sample
396	neo-adjuvant chemotherapy	4	JBLAB-15549	3	QIASymphony Circulating DNA (3ml)	17	TP53	7574023	DEL			G	c.1004delG	p.R335fs	0.017	Y
399	neo-adjuvant chemotherapy	4	JBLAB-16724	3.1	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7578263	stopgain	G	A		c.C586T	p.R196X	0.054	Y
400	neo-adjuvant chemotherapy	3A	JBLAB-16728	2.7	QIASymphony Circulating DNA (3ml FIX)											
402	neo-adjuvant chemotherapy	3B	JBLAB-16721	2.55	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7579358	nonsynonymous	C	G		c.G329C	p.R110P	0.101	N
409	neo-adjuvant chemotherapy	4	JBLAB-5787	3	QIASymphony Circulating DNA (3ml)	17	TP53	7577610	splicing	T	G				0.432	Y
410	neo-adjuvant chemotherapy	3C	JBLAB-16730	3.1	QIASymphony Circulating DNA (3ml FIX)											
414	neo-adjuvant chemotherapy	3C	JBLAB-16735	2.6	QIASymphony Circulating DNA (3ml FIX)											
416	neo-adjuvant chemotherapy	3C	JBLAB-16739	2.55	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7577022	stopgain	G	A		c.C916T	p.R306X	0.157	Y
427	neo-adjuvant chemotherapy	3B	JBLAB-16744	3.1	QIASymphony Circulating DNA (3ml FIX)											
428	neo-adjuvant chemotherapy	3C	JBLAB-16746	2.5	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7578403	nonsynonymous	C	G		c.G527C	p.C176S	0.084	Y
430	neo-adjuvant chemotherapy	3C	JBLAB-16659	2.1	QIASymphony Circulating DNA (2ml FIX)											
431	neo-adjuvant chemotherapy	4	JBLAB-16751	2.6	QIASymphony Circulating DNA (3ml FIX)											
434	neo-adjuvant chemotherapy	3	JBLAB-16662	0.8	QIASymphony Circulating DNA (2ml FIX)											
441	neo-adjuvant chemotherapy	3C	JBLAB-16755	3.1	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7579466	nonsynonymous	G	A		c.C221T	p.A74V	0.013	No FFPE mutation
444	neo-adjuvant chemotherapy	3C	JBLAB-16758	2.55	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7578203	nonsynonymous	C	T		c.G646A	p.V216M	0.198	Y
450	neo-adjuvant chemotherapy	3C	JBLAB-16663	1.7	QIASymphony Circulating DNA (2ml FIX)											
451	neo-adjuvant chemotherapy	4	JBLAB-16764	3.1	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7579583	DEL			AGGGG	c.100_104del	p.34_35del	0.187	No FFPE sample
452	neo-adjuvant chemotherapy	4	JBLAB-16767	3.1	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7578406	nonsynonymous	C	T		c.G524A	p.R175H	0.067	Y
453	neo-adjuvant chemotherapy	4	JBLAB-16763	3.1	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7577118	nonsynonymous	C	A		c.G820T	p.V274F	0.266	No FFPE sample
461	neo-adjuvant chemotherapy	3C	JBLAB-16773	2.45	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7576852	splicing	C	T				0.050	Y
466	neo-adjuvant chemotherapy	3C	JBLAB-16776	3.1	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7577114	nonsynonymous	C	T		c.G824A	p.C275Y	0.019	No FFPE mutation
467	neo-adjuvant chemotherapy	3C	JBLAB-16772	2.6	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7577114	nonsynonymous	C	T		c.G824A	p.C275Y	0.024	Y
469	neo-adjuvant chemotherapy	3C	JBLAB-16778	2.65	QIASymphony Circulating DNA (3ml FIX)											
470	neo-adjuvant chemotherapy	3C	JBLAB-16783	2.45	QIASymphony Circulating DNA (3ml FIX)											
473	neo-adjuvant chemotherapy	4	JBLAB-16780	3.1	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7578442	nonsynonymous	T	C		c.A488G	p.Y163C	0.107	Y
475	primary surgery	4	JBLAB-18158	1.7	QIASymphony Circulating DNA (3ml FIX)											
483	neo-adjuvant chemotherapy	3	JBLAB-16667	2.1	QIASymphony Circulating DNA (2ml FIX)											

OV04	Treatment	Stage	Plasma sample	Volume plasma (ml)	Extraction method	Chromosome	Gene	Position	Effect	Ref	Mutant	Indel	cDNA effect	Protein effect	MAF	Matches FFPE mutation
485	neo-adjuvant chemotherapy	4	JBLAB-15561	3	QIASymphony Circulating DNA (3ml)	17	TP53	7576897	stopgain	G	A		c.C949T	p.Q317X	0.640	Y
487	neo-adjuvant chemotherapy	4	JBLAB-16794	3.1	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7577120	nonsynonymous	C	T		c.G818A	p.R273H	0.340	No FFPE sample
488	neo-adjuvant chemotherapy	3C	JBLAB-16795	3	QIASymphony Circulating DNA (3ml FIX)	17	TP53	7578190	nonsynonymous	T	C		c.A659G	p.Y220C	0.010	Y
489	neo-adjuvant chemotherapy	3C	JBLAB-15562	3	QIASymphony Circulating DNA (3ml)											
495	neo-adjuvant chemotherapy	3C	JBLAB-16797	3.1	QIASymphony Circulating DNA (3ml FIX)											
497	neo-adjuvant chemotherapy	4	JBLAB-16798	2.55	QIASymphony Circulating DNA (3ml FIX)											
509	primary surgery	3C	JBLAB-16804	2.6	QIASymphony Circulating DNA (3ml FIX)											
515	neo-adjuvant chemotherapy	4B	JBLAB-14722	3.5	QIASymphony Circulating DNA (4ml)	17	TP53	7578190	nonsynonymous	T	C		c.A659G	p.Y220C	0.011	No FFPE sample
516	neo-adjuvant chemotherapy	3C	JBLAB-5845	3	QIASymphony Circulating DNA (3ml)											
523	neo-adjuvant chemotherapy	3C	JBLAB-5856	3	QIASymphony Circulating DNA (3ml)											
524	neo-adjuvant chemotherapy	3C	JBLAB-5855	3	QIASymphony Circulating DNA (3ml)											
525	neo-adjuvant chemotherapy	4A	JBLAB-5871	3.7	QIASymphony Circulating DNA (4ml)	17	TP53	7577517	nonsynonymous	A	G		c.T764C	p.I255T	0.019	Y
527	neo-adjuvant chemotherapy	4	JBLAB-5893	3.3	QIASymphony Circulating DNA (4ml)	17	TP53	7577581	nonsynonymous	A	T		c.T700A	p.Y234N	0.265	Y
531	neo-adjuvant chemotherapy	3C	JBLAB-5982	4.3	QIASymphony Circulating DNA (4ml)											
535	primary surgery	1C	JBLAB-5929	4	QIASymphony Circulating DNA (4ml)											
539	neo-adjuvant chemotherapy	3C	JBLAB-5962	3.9	QIASymphony Circulating DNA (4ml)	17	TP53	7577508	DEL			CCA		p.258del	0.011	N
544	neo-adjuvant chemotherapy	4	JBLAB-14714	3.5	QIASymphony Circulating DNA (4ml)	17	TP53	7578212	stopgain	G	A		c.C637T	p.R213X	0.063	No FFPE sample
547	primary surgery	3C	JBLAB-14726	2.6	QIASymphony Circulating DNA (4ml)											
551	neo-adjuvant chemotherapy	3C	JBLAB-14752	2.75	QIASymphony Circulating DNA (3ml)											
559	primary surgery	3C	JBLAB-14815	3	QIASymphony Circulating DNA (3ml)	17	TP53	7578384	DEL			C	c.546delC	p.C182fs	0.025	No FFPE mutation
588	neo-adjuvant chemotherapy	3C	JBLAB-15042	2.6	QIASymphony Circulating DNA (3ml)	17	TP53	7577574	nonsynonymous	T	C		c.A707G	p.Y236C	0.050	Y
600	primary surgery	4	JBLAB-15012	3	QIASymphony Circulating DNA (3ml)											
606	neo-adjuvant chemotherapy	3C	JBLAB-15067	3	QIASymphony Circulating DNA (3ml)											
610	neo-adjuvant chemotherapy	3C	JBLAB-15108	2.75	QIASymphony Circulating DNA (3ml)											
611	neo-adjuvant chemotherapy	3C	JBLAB-15105	2.75	QIASymphony Circulating DNA (3ml)	17	TP53	7577120	nonsynonymous	C	A		c.G818T	p.R273L	0.235	Y
612	primary surgery	4	JBLAB-15115	3	QIASymphony Circulating DNA (3ml)											
617	neo-adjuvant chemotherapy	3B	JBLAB-15131	3	QIASymphony Circulating DNA (3ml)											
622	primary surgery	1A	JBLAB-15147	3	QIASymphony Circulating DNA (3ml)											
627	neo-adjuvant chemotherapy	3C	JBLAB-15179	3	QIASymphony Circulating DNA (3ml)											
628	neo-adjuvant chemotherapy	4B	JBLAB-15258	3	QIASymphony Circulating DNA (3ml)	17	TP53	7577123	nonsynonymous	A	C		c.T815G	p.V272G	0.162	No FFPE sample
629	neo-adjuvant chemotherapy	4	JBLAB-15221	3	QIASymphony Circulating DNA (3ml)	17	TP53	7577121	nonsynonymous	G	A		c.C817T	p.R273C	0.039	Y
635	neo-adjuvant chemotherapy	3C	JBLAB-15239	3	QIASymphony Circulating DNA (3ml)	17	TP53	7577040	nonsynonymous	G	T		c.C898A	p.P300T	0.014	No FFPE sample
644	neo-adjuvant chemotherapy	4B	JBLAB-15270	3	QIASymphony Circulating DNA (3ml)	17	TP53	7577120	0	C	T		c.G818A	p.R273H	0.460	No FFPE sample
645	neo-adjuvant chemotherapy	4B	JBLAB-15295	3	QIASymphony Circulating DNA (3ml)	17	TP53	7578508	nonsynonymous	C	T		c.G422A	p.C141Y	0.047	Y
647	primary surgery	1C	JBLAB-15297	3	QIASymphony Circulating DNA (3ml)											
648	neo-adjuvant chemotherapy	3C	JBLAB-15306	2.75	QIASymphony Circulating DNA (3ml)	17	TP53	7577539	nonsynonymous	G	C		c.C742G	p.R248G	0.258	Y
650	primary surgery	3B	JBLAB-15332	3	QIASymphony Circulating DNA (3ml)	17	TP53	7578382	stopgain	G	C		c.C548G	p.S183X	0.265	Y
658	primary surgery	1C_1	JBLAB-15379	3	QIASymphony Circulating DNA (3ml)	17	TP53	7578534	DEL			TTG	c.394_396del	p.132_132del	0.092	Y
659	neo-adjuvant chemotherapy	3C	JBLAB-15441	3	QIASymphony Circulating DNA (3ml)											
668	neo-adjuvant chemotherapy	3C	JBLAB-15492	3	QIASymphony Circulating DNA (3ml)	17	TP53	7577541	DEL			TCATGCCGCC	c.731_740del	p.244_247del	0.017	Y
670	neo-adjuvant chemotherapy	3C	JBLAB-17841	3.8	QIASymphony Circulating DNA (4ml)											
683	neo-adjuvant chemotherapy	4B	JBLAB-15489	3	QIASymphony Circulating DNA (3ml)	17	TP53	7579329	nonsynonymous	T	C		c.A358G	p.K120E	0.032	No FFPE sample
686	neo-adjuvant chemotherapy	3C	JBLAB-17850	4.5	QIASymphony Circulating DNA (4ml)	17	TP53	7577497	splicing	A	C		-		0.184	Y
696	neo-adjuvant chemotherapy	4B	JBLAB-15798	3	QIASymphony Circulating DNA (3ml)	17	TP53	7578406	0	C	T		c.G524A	p.R175H	0.020	No FFPE sample
701	neo-adjuvant chemotherapy	3C	JBLAB-15898	3	QIASymphony Circulating DNA (3ml)	17	TP53	7578406	nonsynonymous	C	T		c.G524A	p.R175H	0.308	Y
706	neo-adjuvant chemotherapy	3C	JBLAB-15877	3	QIASymphony Circulating DNA (3ml)											
707	neo-adjuvant chemotherapy	3C	JBLAB-15919	3	QIASymphony Circulating DNA (3ml)											
709	neo-adjuvant chemotherapy	3C	JBLAB-15901	3	QIASymphony Circulating DNA (3ml)											
711	neo-adjuvant chemotherapy	3C	JBLAB-15904	3	QIASymphony Circulating DNA (3ml)	17	TP53	7578190	nonsynonymous	T	C		c.A659G	p.Y220C	0.153	Y
713	neo-adjuvant chemotherapy	3C	JBLAB-15945	3	QIASymphony Circulating DNA (3ml)	17	TP53	7574003	stopgain	G	A		c.C1024T	p.R342X	0.378	Y
714	primary surgery	1C_3	JBLAB-16024	4.5	QIASymphony Circulating DNA (4ml)	17	TP53	7578554	nonsynonymous	A	T		c.T376A	p.Y126N	0.033	Y
715	neo-adjuvant chemotherapy	3C	JBLAB-15924	3	QIASymphony Circulating DNA (3ml)	17	TP53	7577079	stopgain	C	A		c.G859T	p.E287X	0.074	Y
716	primary surgery	3C	JBLAB-15925	3	QIASymphony Circulating DNA (3ml)	17	TP53	7574012	stopgain	C	A		c.G1015T	p.E339X	0.224	Y
730	neo-adjuvant chemotherapy	4B	JBLAB-16083	1.65	QIASymphony Circulating DNA (4ml)	17	TP53	7577554	DEL			GCAG	c.724_727del	p.242_243del	0.091	No FFPE sample
740	primary surgery	3B	JBLAB-16146	4.4	QIASymphony Circulating DNA (4ml)	17	TP53	7578496	nonsynonymous	A	T		c.T434A	p.L145Q	0.044	Y
747	neo-adjuvant chemotherapy	3C	JBLAB-16231	4.5	QIASymphony Circulating DNA (4ml)	17	TP53	7578203	nonsynonymous	C	T		c.G646A	p.V216M	0.014	No FFPE sample
749	neo-adjuvant chemotherapy	4B	JBLAB-16242	4.5	QIASymphony Circulating DNA (4ml)											
751	neo-adjuvant chemotherapy	3C	JBLAB-16237	4.5	QIASymphony Circulating DNA (4ml)											
758	neo-adjuvant chemotherapy	3C	JBLAB-16265	4.5	QIASymphony Circulating DNA (4ml)											
759	neo-adjuvant chemotherapy	4	JBLAB-16273	4.5	QIASymphony Circulating DNA (4ml)											
776	neo-adjuvant chemotherapy	4	JBLAB-16403	4.5	QIASymphony Circulating DNA (4ml)											
777	neo-adjuvant chemotherapy	3C	JBLAB-16408	4.5	QIASymphony Circulating DNA (4ml)											
786	neo-adjuvant chemotherapy	3C	JBLAB-16518	4.5	QIASymphony Circulating DNA (4ml)											
788	neo-adjuvant chemotherapy	4B	JBLAB-16503	4.5	QIASymphony Circulating DNA (4ml)	17	TP53	7577509	stopgain	C	A		c.G772T	p.E258X	0.043	Y
791	primary surgery	1C_2	JBLAB-16526	4.5	QIASymphony Circulating DNA (4ml)											
792	neo-adjuvant chemotherapy	4B	JBLAB-16539	4.5	QIASymphony Circulating DNA (4ml)											
805	primary surgery	1B	JBLAB-16585	4.5	QIASymphony Circulating DNA (4ml)	17	TP53	7578176	splicing	C	A				0.023	Y
810	primary surgery	3C	JBLAB-16881	4.5	QIASymphony Circulating DNA (4ml)											
811	primary surgery	2B	JBLAB-16859	4.5	QIASymphony Circulating DNA (4ml)	17	TP53	7579358	nonsynonymous	C	G		c.G329C	p.R110P	0.048	Y
822	primary surgery	3C	JBLAB-16930	4.5	QIASymphony Circulating DNA (4ml)											
831	primary surgery	4B	JBLAB-16986	4.5	QIASymphony Circulating DNA (4ml)	17	TP53	7578265	nonsynonymous	A	G		c.T584C	p.I195T	0.376	Y
847	neo-adjuvant chemotherapy	3C	JBLAB-17626	4.5	QIASymphony Circulating DNA (4ml)											
853	primary surgery	3C	JBLAB-17627	4.5	QIASymphony Circulating DNA (4ml)	17	TP53	7577529	nonsynonymous	A	T		c.T752A	p.I251N	0.025	Y
870	primary surgery	3B	JBLAB-17805	4.5	QIASymphony Circulating DNA (4ml)	17	TP53	7578455	DEL			GCGGACGCG	c.467_475del	p.156_159del	0.015	Y

12.11 Benign controls

OV04	Histology	Plasma sample	Volume plasma (ml)	Extraction method
563	benign fibroma, benign leiomyomata, benign endometrial polyp	JBLAB-14880	3	QIASymphony Circulating DNA (3ml)
578	benign serous cystadenofibroma	JBLAB-14951	3	QIASymphony Circulating DNA (3ml)
582	benign serous cystadenoma	JBLAB-14964	3	QIASymphony Circulating DNA (3ml)
589	benign mucinous cystadenoma	JBLAB-14985	3	QIASymphony Circulating DNA (3ml)
595	benign mucinous cystadenoma, benign leiomyomata	JBLAB-15005	3	QIASymphony Circulating DNA (3ml)
594	torted benign serous cystadenofibroma, benign leiomyoma, benign endometrial polyp	JBLAB-15006	3	QIASymphony Circulating DNA (3ml)
593	benign serous cystadenofibroma	JBLAB-15008	3	QIASymphony Circulating DNA (3ml)
609	mild complex atypical hyperplasia, benign seromucinous cystadenofibroma, benign serous cystadenofibroma	JBLAB-15096	3	QIASymphony Circulating DNA (3ml)
608	bilateral serous cystadenofibroma	JBLAB-15097	3	QIASymphony Circulating DNA (3ml)
613	torted ovarian cyst of uncertain aetiology (no malignancy), adenomyosis, benign serous cystadenofibroma	JBLAB-15113	3	QIASymphony Circulating DNA (3ml)
630	benign serous cyst	JBLAB-15199	3	QIASymphony Circulating DNA (3ml)
638	benign seromucinous cystadenoma, adenomyosis, benign endometrial polyp	JBLAB-15259	3	QIASymphony Circulating DNA (3ml)
639	benign serous cystadenofibroma	JBLAB-15260	3	QIASymphony Circulating DNA (3ml)
641	bilateral benign serous cystadenomas	JBLAB-15261	3	QIASymphony Circulating DNA (3ml)
642	benign leiomyoma, adenomyosis, benign haemorrhagic cyst	JBLAB-15263	3	QIASymphony Circulating DNA (3ml)
653	benign brenner tumour and mucinous cystadenoma, bilateral benign leiomyomas, benign endometrial polyp	JBLAB-15359	3	QIASymphony Circulating DNA (3ml)
657	benign ovarian fibroma, benign endometrial polyp, adenomyosis, simple ovarian cyst	JBLAB-15361	3	QIASymphony Circulating DNA (3ml)
661	benign fibroma, focal simple hyperplasia, adenomyosis, benign leiomyoma	JBLAB-15382	3	QIASymphony Circulating DNA (3ml)
672	fibrothecoma, ovarian fibroma, benign endometrial polyp, adenomyosis, leiomyoma	JBLAB-15450	3	QIASymphony Circulating DNA (3ml)
679	benign mucinous cystadenoma with Brenner tumour	JBLAB-15469	3	QIASymphony Circulating DNA (3ml)
682	benign mucinous cystadenoma	JBLAB-15490	3	QIASymphony Circulating DNA (3ml)
795	bilateral benign fibromasa, benign simple cyst	JBLAB-16543	4.5	QIASymphony Circulating DNA (4ml)
799	bilateral serous cystadenofibromas and endosalpingosis	JBLAB-16562	4.5	QIASymphony Circulating DNA (4ml)
804	benign serous cystadenofibroma, benign serous cystadenoma, benign leiomyoma	JBLAB-16584	4.5	QIASymphony Circulating DNA (4ml)
814	benign fibroma, benign serous cystadenoma, benign endometrial polyp, adenomyosis, submucosal leiomyoma	JBLAB-16861	4.5	QIASymphony Circulating DNA (4ml)
821	bilateral benign ovarian fibromas	JBLAB-16929	4.15	QIASymphony Circulating DNA (4ml)
823	tubo-ovarian abscess	JBLAB-16931	4.5	QIASymphony Circulating DNA (4ml)
537	endometriosis, adenomyosis, leiomyoma	JBLAB-5945	3.4	QIASymphony Circulating DNA (4ml)

12.12 Specific variant calls for CTCR-OV04 plasma using TAm-Seq

OV04	Chr	Position	Ref	Mutant	Type	Length	Base	MAF	Background error	Mutation detected
34	17	7577570	C	T				0.003	0.001	Y
72	17	7578275	G	A				0.004	0.002	Y
74	17	7578526	C	T				0.072	0.002	Y
75	17	7576855	G	A				0.006	0.001	Y
95	17	7578290	C	T				0.001	0.002	N
127	17	7577120	C	T				0.008	0.003	Y
133	17	7578493	C	T				0.001	0.001	N
167	17	7577539	G	A				0.008	0.000	Y
176	17	7577536	T	C				0.007	0.005	Y
191	17	7577538	C	T				0.003	0.003	N
194	17	7577095	G	C				0.000	0.000	N
217	17	7578204	A	C				0.028	0.000	Y
254	17	7578555	C	T				0.001	0.001	N
259	17	7577560	A	C				0.002	0.000	Y
260	17	7577106	G	A				0.159	0.001	Y
266	17	7578542	G	C				0.077	0.000	Y
280	17	7577578			INS	1	C	0.000	NA	N
294	17	7578423	C	T				0.001	0.001	T
314	17	7578190	T	C				0.021	0.004	Y
316	17	7578442	T	C				0.003	0.002	Y
318	17	7577120	C	T				0.010	0.003	Y
319	17	7576852	C	T				0.035	0.001	Y
333	17	7577114	C	T				0.001	0.001	T
334	17	7577570	C	T				0.001	0.001	F
338	17	7578212	G	A				0.011	0.003	Y
348	17	7578212	G	A				0.008	0.003	Y
362	17	7578413	C	A				0.006	0.000	Y
364	17	7578266	T	A				0.030	0.000	Y
375	17	7578190	T	C				0.008	0.004	Y
382	17	7578279			INS	1	G	0.000	NA	N
392	17	7573999			INS	19	CTCGGAACATCTCGAAGCG	0.000	NA	N
399	17	7578263	G	A				0.055	0.003	Y
402	17	7579511	C	T				0.001	0.002	N
409	17	7577610	T	G				0.433	0.000	Y
410	17	7573926	C	T				0.001	0.001	N
414	17	7577539	G	A				0.003	0.000	Y
416	17	7577022	G	A				0.161	0.002	Y
428	17	7578403	C	G				0.084	0.002	Y
434	17	7577118			INS	1	T	0.000	NA	N
444	17	7578203	C	T				0.198	0.001	Y
452	17	7578406	C	T				0.067	0.002	Y
461	17	7576852	C	T				0.050	0.001	Y
467	17	7577114	C	T				0.029	0.001	Y
473	17	7578442	T	C				0.107	0.002	Y
475	17	7577120	C	A				0.013	0.003	Y
485	17	7576897	G	A				0.580	0.002	Y
488	17	7578190	T	C				0.013	0.004	Y
489	17	7578280	G	A				0.005	0.002	Y
495	17	7578406	C	T				0.008	0.002	Y
516	17	7577539	G	A				0.007	0.000	Y
525	17	7577517	A	G				0.019	0.003	Y
527	17	7577581	A	T				0.605	0.000	Y
539	17	7577157	T	G				0.000	0.001	N
551	17	7578265	A	G				0.005	0.004	Y
588	17	7577574	T	C				0.050	0.002	Y
606	17	7577556	C	T				0.005	0.002	Y
611	17	7577120	C	A				0.235	0.003	Y
622	17	7578433	G	C				0.000	0.000	N
627	17	7577121	G	A				0.009	0.003	Y
629	17	7577121	G	A				0.030	0.003	Y
645	17	7578508	C	T				0.047	0.002	Y
648	17	7577539	G	C				0.259	0.000	Y
670	17	7577142	C	T				0.005	0.002	Y
686	17	7577497	A	C				0.184	0.000	Y
701	17	7578406	C	T				0.310	0.002	Y
707	17	7577517	A	C				0.001	0.003	N
709	17	7578263	G	A				0.003	0.003	T
711	17	7578190	T	C				0.152	0.004	Y
713	17	7574003	G	A				0.377	0.003	Y
714	17	7578554	A	T				0.033	0.000	Y
715	17	7577079	C	A				0.075	0.000	Y
716	17	7574012	C	A				0.224	0.001	Y
740	17	7578496	A	T				0.044	0.000	Y
749	17	7577559	G	A				0.003	0.001	Y
751	17	7577539	G	A				0.003	0.002	T
758	17	7578535	T	C				0.010	NA	T
776	17	7578461	C	A				0.001	NA	T
788	17	7577509	C	A				0.043	NA	T
791	17	7574003	G	A				0.006	0.003	T
805	17	7578176	C	A				0.023	NA	T
847	17	7577094	G	A				0.005	0.004	Y
853	17	7577529	A	T				0.024	0.000	Y
870	17	7578455			DEL	9	GCGGACGCG	0.013	NA	Y

12.13 Manuscript: Science Translational Medicine

Enhanced detection of circulating tumor DNA by fragment size analysis

One sentence summary: Selective sequencing or in silico analysis for differences in DNA fragment size can improve the detection of circulating tumor DNA

Authors: Florent Mouliere[1, 2, †], Dineika Chandrananda[1, 2, †], Anna M. Piskorz[1, 2, †], Elizabeth K. Moore[1, 2, 3, †], James Morris[1, 2], Lise Barlebo Ahlborn[4, 5], Richard Mair[1, 2, 6], Teodora Goranova[1, 2], Francesco Marass[1, 2, 7, 8], Katrin Heider[1, 2], Jonathan C. M. Wan[1, 2], Anna Supernat[1, 2, 9], Irena Hudecova[1, 2], Ioannis Gounaris[1, 2, 3], Susana Ros[1, 2], Mercedes Jimenez-Linan[2, 3], Javier Garcia-Corbacho[10], Keval Patel[1, 2], Olga Østrup[5], Suzanne Murphy[1, 2], Matthew D. Eldridge[1, 2], Davina Gale[1, 2], Grant D. Stewart[2, 11], Johanna Burge[2, 11], Wendy N. Cooper[1, 2], Michiel S. van der Heijden[12, 13], Charles E. Massie[1, 2], Colin Watts[14], Pippa Corrie[3], Simon Pacey[3, 15], Kevin Brindle[1, 2, 16], Richard D. Baird[17], Morten Mau-Sørensen[4], Christine A. Parkinson[1, 2, 3, 18, 19], Christopher G. Smith[1, 2], James D. Brenton[1, 2, 3, 18, 19, #, *], Nitzan Rosenfeld[1, 2, #, *].

Affiliations:

1. Cancer Research UK Cambridge Institute, University of Cambridge, CB2 0RE, Cambridge, UK.
2. Cancer Research UK Major Centre – Cambridge, Cancer Research UK Cambridge Institute, CB2 0RE, Cambridge, UK.
3. Cambridge University Hospitals NHS Foundation Trust, CB2 0QQ, Cambridge, UK.
4. Department of Oncology, Rigshospitalet, Copenhagen University Hospital, DK-2100, Denmark.
5. Centre for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, DK-2100, Denmark.
6. Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, CB2 0QQ, Cambridge, UK.
7. Department of Biosystems Science and Engineering, ETH Zurich, 4058, Basel, Switzerland.
8. Swiss Institute of Bioinformatics, 4058, Basel, Switzerland.
9. Department of Medical Biotechnology, Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, 80-211, Poland.
10. Clinical Trials Unit, Clinic Institute of Haematological and Oncological Diseases, Hospital Clinic de Barcelona, 170 08036, Barcelona, Spain.
11. Academic Urology Group, Department of Surgery, University of Cambridge, CB2 0QQ, Cambridge, UK.
12. Division of Molecular Carcinogenesis, Netherlands Cancer Institute, Amsterdam, 1066 CX, Netherlands.
13. Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam, 1066 CX, Netherlands.
14. Institute of Cancer Genomics Science, University of Birmingham, B15 2TT, Birmingham, UK.
15. Department of Oncology, University of Cambridge, CB2 0XZ, Cambridge, UK.
16. Department of Biochemistry, University of Cambridge, CB2 1QW, Cambridge, UK.
17. Early Phase Clinical Trials and Breast Cancer Research Teams, Cancer Research UK Cambridge Centre, CB2 0QQ, Cambridge, UK.
18. Department of Oncology, Hutchison/MRC Research Centre, University of Cambridge, CB2 0XZ, Cambridge, UK.
19. NIHR Cambridge Biomedical Research Centre, CB2 0QQ, Cambridge, UK.

[*]Correspondence to: James D. Brenton (james.brenton@cruk.cam.ac.uk) and Nitzan Rosenfeld (nitzan.rosenfeld@cruk.cam.ac.uk).
[†] co-first authors; [#] co-senior authors

Abstract: Existing methods to improve detection of ctDNA have focused on sensitivity of mutation detection and not the biological properties of plasma cell-free DNA (cfDNA). We hypothesized that differences in fragment lengths of circulating DNA could be exploited to enhance sensitivity for detecting the presence of ctDNA and for non-invasive genomic analysis of cancer. We surveyed ctDNA fragment sizes in 344 plasma samples from 200 cancer patients using low-pass whole-genome sequencing (0.4×). To establish the size distribution of mutant ctDNA, tumor-guided personalized deep sequencing was performed in in 19 patients. We detected enrichment of ctDNA in fragment sizes between 90–150 bp, and developed methods for in vitro and in silico size selection of these fragments. Selecting fragments between 90–150 bp improved detection of tumor DNA by more than 2-fold median enrichment in >95% of cases, and by more than 4-fold enrichment in >10% of cases. Analysis of size-selected cfDNA identified clinically actionable mutations and copy number alterations that were otherwise not detected. Identification of patients with advanced cancer was improved by predictive models integrating fragment length and copy number analysis of cfDNA with AUC>0.99 compared to AUC<0.80 without fragmentation features. Increased detection of cfDNA from patients with glioma, renal and pancreatic cancer patients was achieved with AUC>0.91, compared to AUC<0.5 without fragmentation features. Fragment-size analysis and selective sequencing of specific fragment sizes can boost ctDNA detection, and could be an alternative to deeper mutation sequencing for clinical applications, earlier diagnosis and to study tumor biology.

Introduction:

Blood plasma of cancer patients contains circulating tumor DNA (ctDNA), but this valuable source of information is diluted by much larger quantities of DNA of non-cancerous origins: ctDNA therefore represents only a small fraction of the total cell-free DNA (cfDNA) *(1, 2)*. High-depth targeted sequencing of selected genomic regions can be used to detect low levels of ctDNA, but broader analysis with methods such as whole exome sequencing (WES) and shallow whole genome sequencing (sWGS) are only generally informative when ctDNA levels are ~10% or greater *(3–5)*. The concentration of ctDNA can exceed 10% of the total cfDNA in patients with advanced-stage cancers *(6–8)*, but is much lower in patients with low tumor burden *(9–12)* and in patients with some cancer types such as gliomas and renal cancers *(6)*. Current strategies to improve ctDNA detection rely on increasing depth of sequencing coupled with various error-correction methods *(2, 13, 14)*. However, approaches that focus only on mutation analysis do not take advantage of the potential differences in chromatin organization or fragment size in ctDNA *(15–17)*. Results of ever-deeper sequencing are also confounded by the likelihood of false positive results from detection of mutations from non-cancerous cells or clonal expansions in normal epithelia, or clonal hematopoiesis of indeterminate potential (CHIP) *(13, 18, 19)*.

The cell of origin and the mechanism of cfDNA release into blood can mark cfDNA with specific fragmentation signatures, potentially providing precise information about cell type, gene expression, oncogenic potential or action of treatment *(15, 16, 20)*. cfDNA fragments

commonly show a prominent mode at 167 bp, suggesting release from apoptotic caspase-dependent cleavage *(21–24)* (**Fig. 1A**). Circulating fetal DNA has been shown to be shorter than maternal DNA in plasma, and these size differences have been used to improve sensitivity of non-invasive prenatal diagnosis *(22, 25–27)*. The size distribution of tumor-derived cfDNA has only been investigated in a few studies, encompassing a small number of cancer types and patients, and shows conflicting results *(28–33)*. A limitation of previous studies is that determining the specific sizes of tumor-derived DNA fragments requires detailed characterization of matched tumor-derived alterations *(30, 33)*, and the broader understanding and implications of potential biological differences have not previously been explored.

We hypothesized we could improve sensitivity for non-invasive cancer genomics by selective sequencing of ctDNA fragments, and leveraging differences in the biology that determines DNA fragmentation. As a proof of concept, we established a pan-cancer catalogue of cfDNA fragmentation features in plasma samples from patients with different cancer types and healthy individuals to identify biological features enriched in tumor-derived DNA. We developed methods for selecting specific sizes of cfDNA fragments prior to sequencing, and investigated the impact of combining cfDNA size selection with genome-wide sequencing to improve the detection of ctDNA and the identification of clinically actionable genomic alterations.

Results

Surveying the fragmentation features of tumor cfDNA.

We generated a catalogue of cfDNA fragmentation features (**Fig. 1A**) from 344 plasma samples from 200 patients with 18 different cancer types, and additional 65 plasma samples from healthy controls (**Fig. 1B, Suppl. Fig. 1, Suppl. Table 1** and **Suppl. Table 2**). The size distribution of cfDNA fragments in cancer patients differed in the size ranges of 90–150 bp, 180–220 bp and 250–320 bp compared to healthy individuals (**Fig. 1B** and **Suppl. Fig. 2**). cfDNA fragment sizes in plasma of healthy individuals, and in plasma of patients with late stage glioma, renal, pancreatic and bladder cancers, were significantly longer than in other late stage cancer types including breast, ovarian, lung, melanoma, colorectal and cholangiocarcinoma ($p<0.001$, Kruskal-Wallis; **Fig. 1C**). Sorting the 18 cancer types according to the proportion of cfDNA fragments in the size range 20–150 bp was very similar to ordering by Bettegowda et al. based on the concentrations of ctDNA measured by individual mutation assays (**Fig. 1D**) *(6)*. In contrast to previous reports *(6, 34)*, this sorting analysis was performed without any prior knowledge of the presence of mutations or somatic copy number alterations (SCNAs), yet allowed the investigation of ctDNA content in different cancers.

Sizing up mutant ctDNA.

We determined the size profile of mutant ctDNA in plasma using two high specificity approaches. First, we inferred the specific size profile of ctDNA and non-tumor cfDNA with sWGS from the plasma of mice bearing human ovarian cancer xenografts (**Fig. 2A**). We observed a shift in ctDNA fragment sizes to less than 167 bp (**Fig. 2B**). Second, the size profile of mutant ctDNA was determined in plasma from 19 cancer patients, using deep sequencing with patient-specific hybrid-capture panels developed from whole-exome profiling of matched tumor samples (**Fig. 2C**). By sequencing hundreds of mutations at a depth >300× in cfDNA, allele-specific reads from mutant and normal DNA were obtained. Enrichment of DNA fragments carrying tumor-mutated alleles was observed in fragments ~20–40 bp shorter than nucleosomal DNA sizes (multiples of 167 bp) (**Fig. 2D**). We determined that mutant ctDNA is generally more fragmented than non-mutant cfDNA, with a maximum enrichment of ctDNA in fragments between 90 and 150 bp (**Suppl. Fig. 3**), as well as enrichment in the size range 250–320 bp. These data also indicated that mutant DNA in plasma of patients with advanced cancer (pre-treatment) is consistently shorter than predicted mono-, and di-nucleosomal DNA fragment lengths (**Fig. 2D**).

Selecting tumor-derived DNA fragments.

These data indicated that ctDNA is shorter than non-tumor cfDNA and suggested that biological differences in fragment lengths could be harnessed to improve ctDNA detection. We determined the feasibility of selective sequencing of shorter fragments using in vitro size selection with a bench-top microfluidic device followed by sWGS, in 48 plasma samples from 35 patients with high-grade serous ovarian cancer (HGSOC) (**Fig. 3A**) (**Suppl. Fig. 4** and **Suppl. Fig. 5**). We assessed the accuracy and quality of the size selection with the plasma from 20 healthy individuals (**Fig. 3B** and **Suppl. Fig. 6**). We also explored the utility of in silico size selection of fragmented DNA using read-pair positioning from unprocessed sWGS data (**Fig. 3A**). In silico size selection was performed once reads were aligned to the genome reference, by selecting the paired-end reads that corresponded to the fragments lengths in a 90–150 bp size range. **Fig. 3C, Fig. 3D** and **Fig. 3E** illustrate the effect of in vitro size selection for one HGSOC case (see all 5 samples in **Suppl. Fig. 7** and **Suppl. Fig. 8**). First, we identified SCNAs in plasma cfDNA before treatment, when the concentration of ctDNA was high (**Fig. 3C**). Only a small number of focal SCNAs were observed in the subsequent plasma sample collected 3 weeks after initiation of chemotherapy (without size selection, **Fig. 3D**). In vitro size selection of the same post-treatment plasma sample showed a median increase of 6.4 times in the amplitude of detectable SCNAs without size selection. Selective sequencing of shorter fragments in this sample resulted in the detection of multiple other SCNAs that were not observed without size selection (**Fig. 3E**), and a genome-wide copy-number profile that was similar to that obtained before treatment when ctDNA levels were 4 times higher (**Fig. 3C**). In silico size selection also enriched ctDNA but to a lower extent than using in vitro size selection (**Suppl. Fig. 7**). We concluded that selecting short DNA fragments in plasma can enrich tumor content on a genome-wide scale.

Quantifying the impact of size selection.

To quantitatively assess the enrichment after size selection on a genome-wide scale, we developed a metric from sWGS data (<0.4× coverage) called t-MAD (trimmed Median Absolute Deviation from copy-number neutrality, see **Fig. 4A**). All sWGS data were downsampled to 10 million sequencing reads for comparison. To define the detection threshold, we measured the t-MAD score for sWGS data from 65 plasma samples from 46 healthy individuals and took the maximal value (median=0.01, range 0.004–0.015). We compared t-MAD to the mutant allele fraction (MAF) in the high ctDNA cancer types assessed by digital PCR (dPCR) or WES in 97 samples. We observed a high correlation (Pearson correlation, r=0.80) (**Fig. 4B**) between t-MAD and MAF, for samples with t-MAD greater than the detection threshold (0.015), or with MAF>0.025. **Suppl. Fig. 9** shows that the slope of t-MAD versus MAF fit lines differed between cancer types (range 0.17–1.12) reflecting likely differences in the extent of SCNAs. We estimated the sensitivity of t-MAD for detecting low ctDNA levels using a spike-in dilution of DNA from a patient with a *TP53* mutation into DNA from a pool of 7 healthy individuals (**Suppl. Fig. 10**) which confirmed that the t-MAD score was linear with ctDNA levels down to MAF of ~0.01. In addition, t-MAD scores greater than the detection threshold (0.015) for samples were present even in samples with a MAF as low as 0.004. t-MAD was also strongly correlated with tumor volume determined by RECIST1.1 (Pearson correlation, r=0.6, p<0.0001, n=35) (**Suppl. Fig. 11**).

Using t-MAD we detected ctDNA from 69% (130/189) of the samples from cancer types where ctDNA levels have been shown to be high (**Fig. 4C**). From cancer types for which ctDNA levels are suspected to be low (glioma, renal, bladder, pancreatic), we detected ctDNA in 17% (10/57) of the cases (**Fig. 4C**). We used in silico size selection of the DNA fragments between 90–150 bp from the high ctDNA cancers (n=189) and healthy controls (n=65) to improve the sensitivity for detecting t-MAD (**Fig. 4D**). Receiver operating characteristic (ROC) analysis comparing the t-MAD score for the samples revealed an area under the curve (AUC) of 0.90 after in silico size selection, against an AUC of 0.69 without size selection (**Fig. 4D**).

We explored whether size selected sequencing could improve the detection of response or disease progression. We used sWGS of longitudinal plasma samples from six cancer patients (**Fig. 4E, F**) and in silico size selection of the cfDNA fragments between 90–150bp. In two patients, size selected samples indicated tumor progression 60 and 87 days before detection by imaging or unselected t-MAD analysis (**Fig. 4E, F)**. Other longitudinal samples exhibited improvements in the detection of ctDNA with t-MAD and size selection (**Fig. 4F**). Confirmation in large clinical studies will be necessary to determine the potential of selective sequencing of ctDNA for clinical applications.

Identifying more clinically relevant mutations with size selection.

We next tested whether size selection could increase the sensitivity for detecting new mutations in cfDNA. To test effects on copy number aberrations, we studied 35 patients with HGSOC as this is the archetypal copy-number driven cancer *(35)*. t-MAD was used to quantify the enrichment of ctDNA with in vitro size selection in 48 plasma samples, including samples collected before and after initiation of chemotherapy treatment. In vitro size selection resulted in an increase in the calculated t-MAD score from the sWGS data for

47/48 of the plasma samples (98%, t-test, p=0.06) with a mean 2.5 and median 2.1-fold increase (**Fig. 5A** and **Suppl. Table 6**). We compared the t-MAD scores against those obtained by sWGS for the plasma samples from healthy individuals. 44 of the 48 size-selected HGSOC plasma samples (92%) had a t-MAD score greater than the highest t-MAD value determined in the in vitro size selected healthy plasma samples (**Fig. 5A**, **Suppl. Fig. 12** and **Suppl. Fig. 6**), compared to only 24 out of 48 without size selection (50%). ROC analysis comparing the t-MAD score for the samples from the cancer patients (pre- and post-treatment initiation, n=48) and healthy controls (n=46) revealed an AUC of 0.97 after in vitro size selection, with maximal sensitivity and specificity of 90% and 98%, respectively. This was significantly superior to detection by sWGS without size selection (AUC=0.64) (**Fig 5B**).

We then determined if this improved sensitivity resulted in the detection of SCNAs with potential clinical value. Across the genome, t-MAD scores evaluating SCNAs were higher after size selection in 33/35 (94%) HGSOC patients, and the absolute level of the copy number (log_2 ratio) values significantly increased after in vitro size selection (t-test for the means, p=0.003) (**Fig. 5C**). We compared the relative copy number values for 15 genes frequently altered in HGSOC (**Suppl. Table 3)**. Analysis of plasma cfDNA after size selection revealed a large number of SCNAs that were not observed in the same samples without size selection (**Fig. 5D**), including amplifications in key genes such as *NF1*, *TERT*, and *MYC* (**Suppl. Fig. 13**).

We also tested whether similar enrichment was seen for substitutions, to exclude the possiblity that size selection might only increase the sensitivity for sWGS analysis. We performed whole exome sequencing of plasma cfDNA from 23 patients with 7 cancer types (**Suppl. Fig. 1**). We used the WES data to compare the size distributions of fragments carrying mutant or non-mutant alleles (**Fig. 6A**), and to test whether size selection could identify additional mutations. We first selected 6 patients with HGSOC and performed WES of plasma DNA with and without in vitro size selection in the 90–150 bp range, analysing time-points before and after initiation of treatment *(36)*. In addition, in silico size selection for the same range of fragment sizes was performed (**Fig. 6A**). Analysis of the MAF of SNVs revealed statistically significant enrichment of the tumor fraction with both in vitro size selection (mean 4.19-fold, median 4.27-fold increase, t-test, p<0.001) and in silico size selection (mean 2.20-fold, median 2.25-fold increase, t-test, p<0.001) (**Fig. 6A** and **Suppl. Fig. 14**). Three weeks after initiation of treatment, ctDNA levels are often lower *(36)*, we therefore further analyzed post-treatment plasma samples using Tagged-Amplicon Deep Sequencing (TAm-Seq) *(37)*. We observed enrichment of MAFs by in vitro size selection between 0.9 and 118 times (mean 2.1 times, median 1.5 times) compared to the same samples without size selection (**Suppl. Fig. 15**).

Size selection with both in vitro and in silico methods increased the number of mutations detected by WES by an average of 53% compared to no size selection (**Fig. 6B**). We identified a total of 1023 mutations in the non-size-selected samples. An additional 260 mutations were detected by in vitro size selection, and an additional 310 mutations were called after in silico size selection (**Fig. 6B** and **Suppl. Table 4**). To exclude the possibility that the improved sensitivity for mutation detection was a result of sequencing artefacts, we validated whether new mutations were also detectable in tumor specimens. We used in silico size selection in an independent cohort of 16 patients where matched tumor tissue DNA was available (**Suppl. Table 5**). In silico size selection enriched the MAF for nearly all mutations

(2061/2133, 97%), with an average increase of MAF of ×1.7 (**Fig. 6C**). For 13 of 16 patients (81%) we identified additional mutations in plasma after in silico size selection. Of these 82 additional mutations, 23 (28%) were confirmed to be present in the matched tumor tissue DNA (**Fig. 6D**). Notably, this included mutations in key cancer genes including *BRAF*, *ARID1A*, and *NF1* (**Suppl. Fig. 16**).

Detecting cancer by supervised machine learning combining cfDNA fragmentation and somatic alteration analysis.

It is important to note that although in vitro and in silico size selection increase the sensitivity of detection, they also result in a loss of cfDNA for analysis. Regions of the cancer genome which are not altered by mutation also excluded and cannot contribute to the analysis (**Suppl. Fig. 17**). We hypothesized that leveraging other biological properties of the cfDNA fragmentation profile could enhance the detection of ctDNA.

We defined other cfDNA fragmentation features from sWGS data including (1) the proportion of fragments in multiple size ranges, (2) the ratios of proportions of fragments in different sizes and (3) the amplitude of oscillations in fragment-size density with 10 bp periodicity (see Methods and **Fig. 7A**). These fragmentation features were compared between cancer patients and healthy individuals (**Suppl. Fig. 18**) and the feature representing the proportion (P) of fragments between 20–150 bp exhibited the highest AUC (0.819). Principal component analysis (PCA) of the samples represented by t-MAD and fragmentation features showed a separation between healthy and cancerous samples and that fragment features clustered with t-MAD scores (**Fig. 7B**).

We next explored the potential of fragmentation features to enhance the detection of tumor DNA in plasma samples. A predictive analysis was performed using the t-MAD score and 9 fragmentation features across 304 samples (239 from cancers patients and 65 from healthy controls) (**Fig. 7C**) (**Suppl. Fig. 19** and **Suppl. Table 2**). The 9 fragmentation features determined from sWGS included five features based on the proportion (P) of fragments in defined size ranges: P(20–150), P(100–150), P(160–180), P(180–220), P(250–320); three features based on ratios of those proportions: P(20–150)/P(160–180), P(100–150)/P(163–169), P(20–150)/P(180–220); and a further feature based on the amplitude of the oscillations having 10 bp periodicity observed below 150 bp.

Variable selection and the classification of samples as “healthy” or “cancer” were performed using logistic regression (LR) and random forests (RF) trained on 153 samples, and validated on two datasets of 94 and 83 independent samples (**Fig. 7C**). The best feature set for the LR model included t-MAD, 10 bp amplitude, P(160–180), P(180–220) and P(250–320). The same five variables were independently identified using the RF model (with some differences in their ranking). **Suppl. Fig. 20** shows performance metrics for the different algorithms on training set data using cross-validation. Using t-MAD alone in the validation pan-cancer dataset (**Fig. 7D** and **Suppl. Fig. 19**), we could distinguish cancer samples from healthy individuals with AUC=0.764. Using the LR model improved the classification of the

samples to AUC=0.908. The RF model (trained on the 153-sample training set) could distinguish cancer from healthy individuals even more accurately in the validation data set (n=94) with AUC=0.994. On the second validation dataset containing low-ctDNA cancer samples (n=83) (**Fig. 7E**), t-MAD alone or the LR performed less well, with AUC values of 0.421 and 0.532 respectively. However, the RF model was still able to distinguish samples from low-ctDNA cancer samples from healthy controls with AUC=0.914. At a specificity of 95%, the RF model correctly classified as cancer 64/68 (94%) of the samples from high-ctDNA cancers (colorectal, cholangiocarcinoma, ovarian, breast, melanoma), and 37/57 (65%) of the samples from low-ctDNA cancers (pancreatic, renal, glioma) (**Fig. 7F**). In a second iteration of model training, we omitted t-MAD, using only the 4 fragmentation features (**Suppl. Fig. 21**). The RF model could still distinguish cancer from healthy controls albeit with slightly reduced AUCs (0.989 for cancer types with high levels of ctDNA and 0.891 for cancer types with low levels of ctDNA), suggesting that the cfDNA fragmentation pattern is most important predictive component.

Discussion:

Our results indicate that exploiting fundamental properties of cfDNA with fragment specific analyses can provide more sensitive analysis of ctDNA. We based the selection criteria on a biological observation that ctDNA fragment size distribution is shifted from normal cfDNA. Our work builds on a comprehensive survey of plasma cfDNA fragmentation patterns across 200 patients with multiple cancer types and 65 healthy individuals. We identified features that could determine the presence and amount of ctDNA in plasma samples, without *a priori* knowledge of somatic aberrations. Although this catalogue is the first of its kind, we caution that it is limited to double-stranded DNA from plasma samples, and is subject to potential biases incurred by the DNA extraction and sequencing methods we used. Additional biological effects could contribute to further selective analysis of cfDNA. Other bodily fluids (urine, cerebrospinal fluid, saliva), different nucleic acids and structures, altered mechanisms of release into circulation, or sample processing methods could exhibit varying fragment size signatures and could offer additional exploitable biological patterns for selective sequencing.

Previous work has reported the size distributions of mutant ctDNA, but only considered limited genomic loci, cancer types, or cases *(30, 32, 33)*. We identified the size differences between mutant and non-mutant DNA on a genome-wide and pan-cancer scale. We developed a method to size mutant ctDNA without using high-depth WGS. By sequencing >150 mutations per patient at high depth we obtained large numbers of reads that could be unequivocally identified as tumor-derived, and thus determined the size distribution of mutant ctDNA and non-mutant cfDNA in cancer patients. A potential limitation of our approach is that capture-based sequencing is biased by probe capture efficiency and therefore our data may not accurately reflect ctDNA fragments <100bp or >300bp.

Our work provides strong evidence that the modal size of ctDNA for many cancer types is less than 167bp, which is the length of DNA wrapped around the chromatosome. In addition, our work also shows that there is a high level of enrichment of mutant DNA fragments at sizes greater than 167 bp, notably in the range 250–320 bp. These longer fragments may explain previous observations that longer ctDNA can be detected in the plasma of cancer

patients *(29, 32)*. The origin of these long fragments is still unknown, and their observation could be linked to technical factors. However, it is likely that mechanisms of compaction and release of cfDNA into circulation, which may differ depending on its origin, will be reflected by different fragment sizes *(38)*. Improving the characterization of these fragments will be important, especially for future work combining ctDNA analysis with other entities in blood such as microvesicles and tumor-educated platelets *(39, 40)*. Fragment specific analyses not only increase the sensitivity for detection of rare mutations, but could be used to track modifications in the size distribution of ctDNA. Future work should address whether this approach could be used to elucidate mechanistic effects of treatment on tumor cells, for example by distinguishing between necrosis and apoptosis based on fragment size *(41)*.

Genome-wide and exome sequencing of plasma DNA at multiple time-points during cancer treatment have been proposed as non-invasive means to study cancer evolution and for the identification of possible resistance mechanisms to treatment *(3)*. However, WGS and WES approaches are costly and have thus far been applicable only in samples for which the tumor DNA fraction was >5–10% *(3–5, 42)*. We demonstrated that we could exploit the differences in fragment lengths using in vitro and in silico size selection to enrich for tumor content in plasma samples which improved mutation and SCNA detection in sWGS and WES data. We demonstrated that size selection improved the detection of mutations that are present in plasma at low allelic fractions, while maintaining low sequencing depth by sWGS and WES. Size selection can be achieved with simple means and at low cost, and is compatible with a wide range of downstream genome-wide and targeted genomic analyses, greatly increasing the potential value and utility of liquid biopsies.

Size selection can be applied in silico, which incurs no added costs, or in vitro, which adds a simple and low-cost intermediate step that can be applied to either the extracted DNA or the libraries created from it. This approach, applied prospectively to new studies, could boost the clinical utility of ctDNA detection and analysis, and creates an opportunity for re-analysis of large volumes of existing data *(4, 34, 43)*. The limitation of this technique is a potential loss of material and information, since some of the informative fragments may be found in size ranges that are filtered out or de-prioritized in the analysis. This may be particularly problematic if only a few copies of the fragments of interest are present in plasma. Despite potential loss of material, we demonstrated that classification algorithms can learn from cfDNA fragmentation features and SCNAs analysis and improve the detection of ctDNA with a cheap sequencing approach (**Fig. 7**). Moreover, the cfDNA fragmentation features alone can be leveraged to classify cancer and healthy samples with a high accuracy (AUC=0.989 for high ctDNA cancers, and AUC=0.891 for low ctDNA cancers) (**Suppl. Fig. 21**).

Analysis of fragment sizes could provide improvements in other applications. Introducing fragment size information on each read could enhance mutation-calling algorithms from high depth sequencing, to identify tumor-derived mutations from other sources such as somatic variants or background sequencing noise. In addition, cfDNA analysis in patients with CHIP is likely to be structurally different from ctDNA released during tumor cell proliferation *(18, 19)*. Thus, fragmentation analysis or selective sequencing strategies could be applied to distinguish clinically relevant tumor mutations from those present in clonal expansions of normal cells. This will be critical for the development of cfDNA-based methods for identification of patients with early stage cancer.

Size selection could also have an impact on the detection of other types of DNA in body fluids or to enrich signals for circulating bacterial or pathogen DNA and mitochondrial DNA. These DNA fragments are not associated with nucleosomes and are often highly fragmented below 100bp. Filtering such fragments may prove to be important in light of the recently established link between the microbiome and treatment efficiency *(17, 44)*. Moreover, recent work highlights a stronger correlation between ctDNA detection and cellular proliferation, rather than cell-death *(45)*. We hypothesize that the mode of the distribution of ctDNA fragment sizes at 145bp could reflect cfDNA released during cell proliferation, and the fragments at 167bp may reflect cfDNA released by apoptosis or maturation/turnover of blood cells. The effect of other cancer hallmarks *(46)* on ctDNA biology, structure, concentration and release is yet unknown.

In summary, ctDNA fragment size analysis, via size selection and machine learning approaches, boosts non-invasive genomic analysis of tumor DNA. Size selection of shorter plasma DNA fragments enriches ctDNA, and leads to the identification of a greater number of genomic alterations with both targeted and untargeted sequencing at a minimal additional cost. Combining cfDNA fragment size analysis and the detection of SCNAs with a non-linear classification algorithm improved the discrimination between samples from cancer patients and healthy individuals. As the analysis of fragment sizes is based on the structural property of ctDNA, size selection could be used with any downstream sequencing applications. Our work could help overcome current limitations of sensitivity for liquid biopsy, supporting expanded clinical and research applications. Our results indicate that exploiting the endogenous biological properties of cfDNA provides an alternative paradigm to deeper sequencing of ctDNA.

Materials and Methods:

Study design. 344 plasma samples from 200 patients with multiple cancer types were collected along with plasma from 65 healthy controls. Among the patients, 172 individuals were recruited through prospective clinical studies at Addenbrooke's Hospital, Cambridge, UK, approved by the local research ethics committee (REC reference numbers: 07/Q0106/63; and NRES Committee East of England - Cambridge Central 03/018). Written informed consent was obtained from all patients and blood samples were collected before and after initiation of treatment with surgery or chemotherapeutic agents. DNA was extracted from 2 mL of plasma using the QIAamp circulating nucleic acid kit (Qiagen) or QIAsymphony (Qiagen) according to the manufacturer's instructions. In addition, 28 patients were recruited as part of the Copenhagen Prospective Personalized Oncology (CoPPO) program (Ref: PMID: 25046202) at Rigshospitalet, Copenhagen, Denmark, approved by the local research ethics committee. Baseline tumor tissue biopsies were available from all 28 patients, together with re-biopsies collected at relapse from two patients, and matched plasma samples. Brain tumor patients were recruited at the Addenbrooke's Hospital, Cambridge, UK, as part of the BLING study (REC – 15/EE/0094). Bladder cancer patients were recruited at the Netherlands Cancer Institute, Amsterdam, The Netherlands, and approval according to national guidelines was obtained (N13KCM/CFMPB250) *(47)*. 65 plasma samples were obtained from healthy control individuals using a similar protocol (Seralab). Plasma samples

have not been freezed-thawed more than 2 times to reduce artifactual fragmentation of cfDNA. Study flowchart of the work is described in **Suppl. Fig. 1**.

In vitro size selection. Between 8-20 ng of DNA were loaded into a 3% agarose cassette (HTC3010, Sage Bioscience) and size selection was performed on a PippinHT (Sage Bioscience) according to the manufacturer's protocol. Quality controls of in vitro size selection were performed on 20 healthy controls samples and detailed in **Suppl. Fig. 6**. We observed an increase in duplicate reads with in vitro size selection, and therefore duplicate reads have been removed for any downstream size selection analysis in the manuscript. To determine the sequencing noise, we used a QC metric called the median absolute pairwise difference (MAPD) algorithm. MAPD measures the absolute difference between the log2 CN ratios of every pair of neighboring bins and then takes the median across all bins. Higher MAPD scores reflect greater noise, typically associated with poor-quality samples. All samples exhibit a MAPD score of 0.01 (+-0.01), irrespective of the size selection condition.

TAm-Seq. Tagged-Amplicon Deep Sequencing libraries were prepared as previously described (*34*), using primers designed to assess single nucleotide variants (SNV) and small indels across selected hotspots and the entire coding regions of *TP53*. Libraries were sequenced using MiSeq or HiSeq 4000 (Illumina).

sWGS. Indexed sequencing libraries were prepared using commercially available kits (ThruPLEX-Plasma Seq and/or Tag-Seq, Rubicon Genomics). Libraries were pooled in equimolar amounts and sequenced to <0.4x depth of coverage on a HiSeq 4000 (Illumina) generating 150-bp paired-end reads. Sequence data were analyzed using an in-house pipeline that consists of the following; Paired end sequence reads were aligned to the human reference genome (GRCh37) using BWA-mem following the removal of contaminating adapter sequences *(48)*. PCR and optical duplicates were marked using MarkDuplicates (Picard Tools) feature and these were excluded from downstream analysis along with reads of low mapping quality and supplementary alignments. When necessary, reads were down-sampled to 10 million in all samples for comparison purposes.

Somatic copy number aberration analysis: The analysis was performed in R using a software suite for shallow Whole Genome Sequencing copy number analysis named CNAclinic (https://github.com/sdchandra/CNAclinic) as well as the QDNAseq pipeline *(49)*. Sequencing reads were randomly sampled to 10 million reads per dataset and allocated into equally sized (30 Kbp) non-overlapping bins throughout the length of the genome. Read counts in each bin were corrected to account for sequence GC content and mappability, and bins overlapping 'blacklisted' regions (derived from the ENCODE project + 1000 Genomes database) prone to artefacts were excluded from downstream analysis. Read counts in test samples were normalized by the counts from an identically processed healthy individual and $\log_2$ transformed to obtained copy number ratio values per genomic bin. Read counts in healthy controls were normalized by their median genome-wide count. Next, bins were segmented using both Circular Binary Segmentation and Hidden Markov Model algorithms, and an averaged $\log_2$R value per bin was calculated.

An in-house empirical blacklist of aberrant read count regions was constructed as follows: 65 sWGS datasets from healthy plasma were used to calculate median read counts per 30 Kbp genomic bin as a function of GC content and mappability. A 2D LOESS surface was applied

and the difference between the actual count and the LOESS fitted values were calculated. The median of these residual values across the 65 controls were calculated per genomic bin. Regions with median residuals greater than 4 standard deviations were blacklisted. The averaged segmental $\log_2$R values in each test sample that overlap this cfDNA blacklist were trimmed and the median absolute value was calculated. This score is defined as t-MAD or the trimmed median absolute deviation from $\log_2$R = 0. The R code to reproduce this analysis is provided in https://github.com/sdchandra/tMAD.

WES. Indexed sequencing libraries were prepared as described above (see Methods, sWGS). Plasma DNA libraries from each sample were made and pooled together for exome capture (TruSeq Exome Enrichment Kit, Illumina). Pools were concentrated using a SpeedVac vacuum concentrator (Eppendorf). Exome enrichment was performed following the manufacturer's protocol. Enriched libraries were quantified using quantitative PCR (KAPA library quantification, KAPA Biosystems), and DNA fragments sizes observed by Bioanalyzer (2100 Bioanalyzer, Agilent Genomics) and pooled in equimolar ratio for paired-end next generation sequencing on a HiSeq4000 (Illumina). Sequencing reads were de-multiplexed allowing zero mismatches in barcodes. Paired-end alignment to the GRCh37 reference genome was performed using BWA-mem for all exome sequencing data including germline, plasma and tumor tissue DNA where generated. PCR duplicates were marked using Picard. Base quality score recalibration and local realignment were performed using Genome Analysis Tool Kit (GATK).

Mutation calling. MAFs for each single-base locus were calculated with MuTect2 for all bases with PHRED quality ≥30. After MuTect2, we applied filtering parameters so that a mutation was called if no mutant reads for an allele were observed in germline DNA at a locus that was covered at least 10x, and if at least 4 reads supporting the mutant were found in the plasma data with at least 1 read on each strand (forward and reverse). At loci with <10x coverage in normal DNA and no mutant reads, mutations were called in plasma if a prior plasma sample showed no evidence of a mutation and was covered adequately (10x or more). We aggregated mutations called before and after size selection with a method called Integrated Signal Amplification for Non-invasive Interrogation of Tumors. This method combines different subsets of mutations called from the same plasma DNA sample using different processing approaches. The mutation aggregation as used in this study is formalized as follows: aggregated mutations = mutations detected without size selection **U** (mutations detected with in vitro size selection **U** mutations detected with in silico size selection).

In silico size selection: Paired-end reads are generated by sequencing DNA from both ends of the fragments present in the library. The original length of the DNA can be inferred using the mapping locations of the read ends in the genome. Once alignment is complete, Samtools software is used to select paired reads that correspond to fragment lengths in a specific range. Mutect2 is used to call mutations from this in silico size selected data as described in the previous section.

Tumor-guided capture sequencing. Matched tumor tissue DNA and plasma DNA samples of 19 patients collected from the RigsHospitalet (Copenhagen, Denmark) with advanced

cancer were sequenced by WES. Variants were called from these samples as previously described (see Methods, Mutation calling). Hybrid-based capture for longitudinal plasma samples analysis were designed to cover these variants for each patient using SureDesign (Agilent). A median of 160 variants were included per patient, and in addition, 41 common genes of interest for pan-cancer analysis were included in the tumor-guided sequencing panel. Indexed sequencing libraries were prepared as described above (see Methods, sWGS). Plasma DNA libraries from each sample were made and pooled together for tumor-guided capture sequencing (SureSelect, Agilent). Pools were concentrated using a SpeedVac vacuum concentrator (Eppendorf). Capture enrichment was performed following the manufacturer's protocol. Enriched libraries were quantified using quantitative PCR (KAPA library quantification, KAPA Biosystems), and DNA fragments sizes controlled by Bioanalyzer (2100 Bioanalyzer, Agilent Genomics) and pooled in equimolar ratio for paired-end next generation sequencing on a HiSeq4000 (Illumina). Sequencing reads were de-multiplexed allowing zero mismatches in barcodes. Paired-end alignment to the GRCh37 reference genome was performed using BWA-mem for all exome sequencing data including germline, plasma and tumor tissue DNA where generated. PCR duplicates were marked using Picard. Base quality score recalibration and local realignment were performed using Genome Analysis Tool Kit (GATK).

Classification analysis. The preliminary analysis was carried out on 304 samples (182 high ctDNA cancer samples, 57 low ctDNA cancer samples and 65 healthy controls). For each sample the following features were calculated from sWGS data: t-MAD, amplitude_10bp, P(20-150), P(160-180), P(20-150)/P(160-180), P(100-150), P(100-150)/P(163-169), P(180-220), P(250-320), P(20-150)/P(180-220). The data was arranged in a matrix where the rows represent each sample and the columns held the aforementioned features with an extra "class" column with the binary labels of "cancer"/"healthy". The following analysis was carried out in R utilising *RandomForest, caret*, and *pROC* packages. The pairwise correlations between the features were calculated to assess multi-collinearity in the dataset (**Suppl. Fig. 19**). A single variable was selected for removal from pairs with Pearson correlation > 0.75. Highly correlated fragmentation features that were composite of individual variables already in the dataset such as P(20-150)/P(180-220), were prioritized for removal. The features were also assessed for zero variance and linear dependencies but none were flagged. After this pre-processing the following 5 variables were selected for further analysis: t-MAD, amplitude_10bp, P(160-180), P(180-220) and P(250-320). All 57 low ctDNA samples were set aside for validation of the models. The data matrix for the remaining high ctDNA cancer samples and healthy controls (n = 247) were randomly partitioned in a 60:40 split into 1 training and 1 validation dataset with the different cancer types and healthy samples represented in similar proportions. Hence, the training data contained 153 samples (cancer=114, healthy=39) while the first validation set of high ctDNA cancers contained 94 samples (cancer=68, healthy=26). This validation dataset was only utilized for final assessment of the classifiers.

Classification of samples as healthy or cancer was performed using one linear and one non-linear machine learning algorithm, namely logistic regression (LR), and random forest (RF). Each algorithm was paired with recursive feature selection in order to identify the best predictor variables. This analysis was carried out with *caret* within the framework of 5 repeats of 10-fold cross-validation on the training set. The algorithm was configured to explore all possible subsets of the features. The optimal model for each classifier was

selected using ROC metric. Separately, a logistic regression model was trained only using t-MAD as a predictor in order to assess the difference in performance without the addition of fragmentation features. Finally, the 68 high ctDNA cancer samples, 57 low ctDNA cancer samples and 26 healthy controls set aside for validation were used to test the classifiers, utilizing area under the curve in a ROC analysis to quantify their performance.

A secondary analysis was carried out on the same training and validation cohorts with the only difference being the features used in the model. Here, we tested predictive ability of fragmentation features without the addition of information from SCNAs (i.e. t-MAD). Hence the features utilized were: amplitude_10bp, P(160-180), P(180-220) and P(250-320).

Quantification of the 10bp periodic oscillation. The amplitude of the 10 bp periodic oscillation observed in the size distribution of cfDNA samples was determined from the sWGS data as follows. Local maxima and minima in the range 75 bp to 150 bp were identified. The average of their positions across the samples was calculated (for minima: 84, 96, 106, 116, 126, 137, 148, and maxima: 81, 92, 102, 112, 122, 134, 144). To compute the amplitude of the oscillations with 10 bp periodicity observed below 150 bp, we calculated the sum of the heights of the maxima and subtracted the sum of the minima. The larger this difference, the more distinct the peaks. The height of the x bp peak is defined as the number of fragments with length x divided by the total number of fragments. To define local maxima, we selected the positions y such that y was the largest value in the interval [y-2, y+2]. The same rationale was used to pick minima.

Supplementary Materials:

Materials and Methods

Figure S1: flowchart of the study design.

Figure S2: size distribution of cfDNA determined by sWGS depending on the cancer type.

Figure S3: insert size distribution of mutant cfDNA determined with hybrid-capture sequencing for 19 patients.

Figure S4: distribution of inserts sizes determined by sWGS for each plasma samples from the 13 ovarian cancer patients of the OV04 cohort collected before and after treatment.

Figure S5: quality control assessment of the in vitro size selection was estimated with sWGS and targeted sequencing.

Figure S6: quality control assessment of the in vitro and in silico size selection on a batch of 20 healthy controls

Figure S7: SCNA analysis without and with size selection of the segmental log2ratio determined after sWGS (<0.4x coverage) for the patient OV04-83.

Figure S8: SCNA analysis of the segmental log2ratio determined after sWGS using on genome wide scale (<0.4x coverage) for the OV04 samples.

Figure S9: Comparison of the MAF and t-MAD score depending on the cancer type for available matched data.

Figure S10: spike-in dilution of cfDNA extracted from the plasma of a breast cancer patient in DNA from a healthy control.

Figure S11: comparison of the available RECIST volume (in mm) determined by CT-scan to the tMAD score and fragmentation features.

Figure S12: t-MAD analysis of the 48 plasma samples collected from 35 ovarian patients was analyzed with and without size selection.
Figure S13: SCNA analysis of the segmental log2ratio determined after sWGS using a list of 29 genes frequently mutated in recurrent ovarian cancer from the plasma samples collected at baseline and after treatment for 13 patients.
Figure S14: MAF for each single nucleotide variants (SNVs) called by WES on the OV04 samples without and with size selection.
Figure S15: analysis of the effect of size selection on ovarian cancer samples with TAm-Seq.
Figure S16: Mutations detected for 9 genes of clinical importance by WES with and without size selection of the short DNA fragments.
Figure S17: Size distribution of non-mutant DNA is affected by the concentration in ctDNA.
Figure S18: ROC curve comparing the individual fragmentation features on the cohort of cancer samples collected from high ctDNA cancer-types.
Figure S19: comparison of t-MAD score to the 9 fragmentation features.
Figure S20: performance metrics for the 50 random sub-samples of cross-validation for the logistic regression on t-MAD score and fragmentation features; random forest (RF) model on training set data from sWGS (n=153; 114 cancer samples, and 39 healthy controls).
Figure S21: LR and RF models detect cancer from healthy controls using only fragmentation features.
Tables S1: summary table of the patients and samples included in the study.
Tables S2: table of the level of the 9 fragmentation features determined by sWGS on the samples included in the study.
Tables S3: table of the log_2ratio levels observed by sWGS of the plasma samples from the OV04 cohort.
Tables S4: mutations called by WES of the 6 patients selected from the OV04 cohort.
Tables S5: mutations called by WES of the 16 patients from the CoPPO cohort.
Tables S6: t-MAD score for the 48 plasma samples of the OV04 cohort before and after in vitro size selection.

References and Notes:

1. G. Siravegna, S. Marsoni, S. Siena, A. Bardelli, Integrating liquid biopsies into the management of cancer, *Nat. Rev. Clin. Oncol.* (2017), doi:10.1038/nrclinonc.2017.14.

2. J. C. M. Wan, C. Massie, J. Garcia-Corbacho, F. Mouliere, J. D. Brenton, C. Caldas, S. Pacey, R. Baird, N. Rosenfeld, Liquid biopsies come of age: towards implementation of circulating tumour DNA, *Nat. Rev. Cancer* **17**, 223–238 (2017).

3. M. Murtaza, S.-J. Dawson, D. W. Y. Tsui, D. Gale, T. Forshew, A. M. Piskorz, C. Parkinson, S.-F. Chin, Z. Kingsbury, A. S. C. Wong, F. Marass, S. Humphray, J. Hadfield, D. Bentley, T. M. Chin, J. D. Brenton, C. Caldas, N. Rosenfeld, Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA, *Nature* **497**, 108–112 (2013).

4. V. A. Adalsteinsson, 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, M. Meyerson, Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors, *Nat. Commun.* **8**, 1324 (2017).

5. E. Heitzer, P. Ulz, J. Belic, S. Gutschi, F. Quehenberger, K. Fischereder, T. Benezeder, M. Auer, C. Pischler, S. Mannweiler, M. Pichler, F. Eisner, M. Haeusler, S. Riethdorf, K. Pantel, H. Samonigg, G. Hoefler, H. Augustin, J. B. Geigl, M. R. Speicher, Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing, *Genome Med.* **5**, 30 (2013).

6. C. Bettegowda, M. Sausen, R. J. Leary, I. Kinde, Y. Wang, N. Agrawal, B. R. Bartlett, H. Wang, B. Luber, R. M. Alani, E. S. Antonarakis, N. S. Azad, A. Bardelli, H. Brem, J. L. Cameron, C. C. Lee, L. A. Fecher, G. L. Gallia, P. Gibbs, D. Le, R. L. Giuntoli, M. Goggins, M. D. Hogarty, M. Holdhoff, S.-M. Hong, Y. Jiao, H. H. Juhl, J. J. Kim, G. Siravegna, D. A. Laheru, C. Lauricella, M. Lim, E. J. Lipson, S. K. N. Marie, G. J. Netto, K. S. Oliner, A. Olivi, L. Olsson, G. J. Riggins, A. Sartore-Bianchi, K. Schmidt, I.-M. Shih, S. M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T. T. Harkins, S. Veronese, T.-L. Wang, J. D. Weingart, C. L. Wolfgang, L. D. Wood, D. Xing, R. H. Hruban, J. Wu, P. J. Allen, C. M. Schmidt, M. A. Choti, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, N. Papadopoulos, L. A. Diaz, Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies, *Sci. Transl. Med.* **6**, 224ra24-224ra24 (2014).

7. F. Diehl, M. Li, D. Dressman, Y. He, D. Shen, S. Szabo, L. A. Diaz, S. N. Goodman, K. A. David, H. Juhl, K. W. Kinzler, B. Vogelstein, Detection and quantification of mutations in the plasma of patients with colorectal tumors, *Proc. Natl. Acad. Sci.* **102**, 16368–16373 (2005).

8. S.-J. Dawson, D. W. Y. Tsui, M. Murtaza, H. Biggs, O. M. Rueda, S.-F. Chin, M. J. Dunning, D. Gale, T. Forshew, B. Mahler-Araujo, S. Rajan, S. Humphray, J. Becq, D. Halsall, M. Wallis, D. Bentley, C. Caldas, N. Rosenfeld, Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer, *N. Engl. J. Med.* **368**, 1199–1209 (2013).

9. F. Diehl, K. Schmidt, M. A. Choti, K. Romans, S. Goodman, M. Li, K. Thornton, N. Agrawal, L. Sokoll, S. A. Szabo, K. W. Kinzler, B. Vogelstein, L. A. Diaz, Circulating mutant DNA to assess tumor dynamics., *Nat. Med.* **14**, 985–90 (2008).

10. J. Tie, Y. Wang, C. Tomasetti, L. Li, S. Springer, I. Kinde, N. Silliman, M. Tacey, H.-L. Wong, M. Christie, S. Kosmider, I. Skinner, R. Wong, M. Steel, B. Tran, J. Desai, I. Jones, A. Haydon, T. Hayes, T. J. Price, R. L. Strausberg, L. A. Diaz, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, P. Gibbs, Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer., *Sci. Transl. Med.* **8**, 346ra92 (2016).

11. A. A. Chaudhuri, J. J. Chabon, A. F. Lovejoy, A. M. Newman, H. Stehr, T. D. Azad, M. S. Khodadoust, M. S. Esfahani, C. L. Liu, L. Zhou, F. Scherer, D. M. Kurtz, C. Say, J. N. Carter, D. J. Merriott, J. C. Dudley, M. S. Binkley, L. Modlin, S. K. Padda, M. F. Gensheimer, R. B. West, J. B. Shrager, J. W. Neal, H. A. Wakelee, B. W. Loo, A. A. Alizadeh, M. Diehn, Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling., *Cancer Discov.* **7**, 1394–1403 (2017).

12. J. D. Cohen, 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, C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. M. Lennon, N. Papadopoulos, Detection and localization of surgically resectable cancers with a multi-analyte blood test., *Science* **359**, 926–930 (2018).

13. I. S. Haque, O. Elemento, Challenges in Using ctDNA to Achieve Early Detection of Cancer, *bioRxiv* , 237578 (2017).

14. A. M. Newman, 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, J. W. Neal, H. A. Wakelee, M. Diehn, A. A. Alizadeh, Integrated digital error suppression for improved detection of circulating tumor DNA, *Nat. Biotechnol.* **34**, 547–555 (2016).

15. P. Ulz, G. G. Thallinger, M. Auer, R. Graf, K. Kashofer, S. W. Jahn, L. Abete, G. Pristauz, E. Petru, J. B. Geigl, E. Heitzer, M. R. Speicher, Inferring expressed genes by whole-genome sequencing of plasma DNA, *Nat. Genet.* **48**, 1273–1278 (2016).

16. M. W. Snyder, M. Kircher, A. J. Hill, R. M. Daza, J. Shendure, Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin., *Cell* **164**, 57–68 (2016).

17. P. Burnham, M. S. Kim, S. Agbor-Enoh, H. Luikart, H. A. Valantine, K. K. Khush, I. De Vlaminck, Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma, *Sci. Rep.* **6**, 27859 (2016).

18. G. Genovese, A. K. Kähler, 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. Landén, M. Höglund, S. Lehmann, S. B. Gabriel, J. L. Moran, E. S. Lander, P. F. Sullivan, P. Sklar, H. Grönberg, C. M. Hultman, S. A. McCarroll, Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence, *N. Engl. J. Med.* **371**, 2477–2487 (2014).

19. Y. Hu, B. Ulrich, J. Supplee, Y. Kuang, P. H. Lizotte, N. Feeney, N. Guibert, M. M. Awad, K.-K. Wong, P. A. Janne, C. P. Paweletz, G. R. Oxnard, False positive plasma genotyping due to clonal hematopoiesis., *Clin. Cancer Res.* , clincanres.0143.2018 (2018).

20. A. J. Bronkhorst, J. F. Wentzel, J. Aucamp, E. van Dyk, L. du Plessis, P. J. Pretorius, Characterization of the cell-free DNA released by cultured cancer cells, *Biochim. Biophys. Acta - Mol. Cell Res.* **1863**, 157–165 (2016).

21. S. Jahr, H. Hentze, S. Englisch, D. Hardt, F. O. Fackelmayer, R. D. Hesch, R. Knippers, DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells., *Cancer Res.* **61**, 1659–65 (2001).

22. Y. M. D. Lo, K. C. A. Chan, H. Sun, E. Z. Chen, P. Jiang, F. M. F. Lun, Y. W. Zheng, T. Y. Leung, T. K. Lau, C. R. Cantor, R. W. K. Chiu, Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus., *Sci. Transl. Med.* **2**, 61ra91 (2010).

23. D. Chandrananda, N. P. Thorne, M. Bahlo, L.-S. Tam, G. Liao, E. Li, High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA, *BMC Med. Genomics* **8**, 29 (2015).

24. P. Jiang, Y. M. D. Lo, The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics, *Trends Genet.* **32**, 360–371 (2016).

25. S. C. Y. Yu, K. C. A. Chan, Y. W. L. Zheng, P. Jiang, G. J. W. Liao, H. Sun, R. Akolekar, T. Y. Leung, A. T. J. I. Go, J. M. G. van Vugt, R. Minekawa, C. B. M. Oudejans, K. H. Nicolaides, R. W. K. Chiu, Y. M. D. Lo, Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing., *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8583–8 (2014).

26. F. M. F. Lun, N. B. Y. Tsui, K. C. A. Chan, T. Y. Leung, T. K. Lau, P. Charoenkwan, K. C. K. Chow, W. Y. W. Lo, C. Wanapirak, T. Sanguansermsri, C. R. Cantor, R. W. K. Chiu, Y. M. D. Lo, Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma., *Proc. Natl. Acad. Sci. U. S. A.* **105**, 19920–5 (2008).

27. G. Minarik, G. Repiska, M. Hyblova, E. Nagyova, K. Soltys, J. Budis, F. Duris, R. Sysak,

M. Gerykova Bujalkova, B. Vlkova-Izrael, O. Biro, B. Nagy, T. Szemes, Utilization of Benchtop Next Generation Sequencing Platforms Ion Torrent PGM and MiSeq in Noninvasive Prenatal Testing for Chromosome 21 Trisomy and Testing of Impact of In Silico and Physical Size Selection on Its Analytical Performance., *PLoS One* **10**, e0144811 (2015).

28. M. B. Giacona, G. C. Ruben, K. A. Iczkowski, T. B. Roos, D. M. Porter, G. D. Sorenson, Cell-Free DNA in Human Blood Plasma, *Pancreas* **17**, 89–97 (1998).

29. N. Umetani, A. E. Giuliano, S. H. Hiramatsu, F. Amersi, T. Nakagawa, S. Martino, D. S. B. Hoon, Prediction of breast tumor progression by integrity of free circulating DNA in serum., *J. Clin. Oncol.* **24**, 4270–6 (2006).

30. F. Mouliere, B. Robert, E. Arnau Peyrotte, M. Del Rio, M. Ychou, F. Molina, C. Gongora, A. R. Thierry, T. Lee, Ed. High Fragmentation Characterizes Tumour-Derived Circulating DNA, *PLoS One* **6**, e23418 (2011).

31. F. Mouliere, S. El Messaoudi, D. Pang, A. Dritschilo, A. R. Thierry, Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer, *Mol. Oncol.* **8**, 927–941 (2014).

32. P. Jiang, C. W. M. Chan, K. C. A. Chan, S. H. Cheng, J. Wong, V. W.-S. Wong, G. L. H. Wong, S. L. Chan, T. S. K. Mok, H. L. Y. Chan, P. B. S. Lai, R. W. K. Chiu, Y. M. D. Lo, Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients., *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1317-25 (2015).

33. H. R. Underhill, J. O. Kitzman, S. Hellwig, N. C. Welker, R. Daza, D. N. Baker, K. M. Gligorich, R. C. Rostomily, M. P. Bronner, J. Shendure, D. J. Kwiatkowski, Ed. Fragment Length of Circulating Tumor DNA, *PLOS Genet.* **12**, e1006162 (2016).

34. O. A. Zill, K. C. Banks, S. R. Fairclough, S. A. Mortimer, J. V Vowles, R. Mokhtari, D. R. Gandara, P. C. Mack, J. I. Odegaard, R. J. Nagy, A. M. Baca, H. Eltoukhy, D. I. Chudova, R. B. Lanman, A. Talasaz, The Landscape of Actionable Genomic Alterations in Cell-Free Circulating Tumor DNA from 21,807 Advanced Cancer Patients., *Clin. Cancer Res.* , clincanres.3837.2017 (2018).

35. G. Macintyre, T. E. Goranova, D. De Silva, D. Ennis, A. M. Piskorz, M. Eldridge, D. Sie, L.-A. Lewsley, A. Hanif, C. Wilson, S. Dowson, R. M. Glasspool, M. Lockley, E. Brockbank, A. Montes, A. Walther, S. Sundar, R. Edmondson, G. D. Hall, A. Clamp, C. Gourley, M. Hall, C. Fotopoulou, H. Gabra, J. Paul, A. Supernat, D. Millan, A. Hoyle, G. Bryson, C. Nourse, L. Mincarelli, L. N. Sanchez, B. Ylstra, M. Jimenez-Linan, L. Moore, O. Hofmann, F. Markowetz, I. A. McNeish, J. D. Brenton, Copy number signatures and mutational processes in ovarian carcinoma, *Nat. Genet.* , 1 (2018).

36. C. A. Parkinson, D. Gale, A. M. Piskorz, H. Biggs, C. Hodgkin, H. Addley, S. Freeman, P. Moyle, E. Sala, K. Sayal, K. Hosking, I. Gounaris, M. Jimenez-Linan, H. M. Earl, W. Qian, N. Rosenfeld, J. D. Brenton, E. R. Mardis, Ed. Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study, *PLOS Med.* **13**, e1002198 (2016).

37. T. Forshew, M. Murtaza, C. Parkinson, D. Gale, D. W. Y. Tsui, F. Kaper, S.-J. Dawson, A. M. Piskorz, M. Jimenez-Linan, D. Bentley, J. Hadfield, A. P. May, C. Caldas, J. D. Brenton, N. Rosenfeld, Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA., *Sci. Transl. Med.* **4**, 136ra68 (2012).

38. A. R. Thierry, S. El Messaoudi, P. B. Gahan, P. Anker, M. Stroun, Origins, structures, and functions of circulating DNA in oncology, *Cancer Metastasis Rev.* **35**, 347–376 (2016).

39. M. G. Best, N. Sol, B. A. Tannous, P. Wesseling, T. Wurdinger, RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics, *Cancer Cell* **28**, 666–676 (2015).

40. M. G. Best, N. Sol, S. G. J. G. In ’t Veld, A. Vancura, M. Muller, A.-L. N. Niemeijer, A. V Fejes, L.-A. Tjon Kon Fat, A. E. Huis In ’t Veld, C. Leurs, T. Y. Le Large, L. L. Meijer, I. E. Kooi, F. Rustenburg, P. Schellen, H. Verschueren, E. Post, L. E. Wedekind, J. Bracht, M. Esenkbrink, L. Wils, F. Favaro, J. D. Schoonhoven, J. Tannous, H. Meijers-Heijboer, G. Kazemier, E. Giovannetti, J. C. Reijneveld, S. Idema, J. Killestein, M. Heger, S. C. de Jager, R. T. Urbanus, I. E. Hoefer, G. Pasterkamp, C. Mannhalter, J. Gomez-Arroyo, H.-J. Bogaard, D. P. Noske, W. P. Vandertop, D. van den Broek, B. Ylstra, R. J. A. Nilsson, P. Wesseling, N. Karachaliou, R. Rosell, E. Lee-Lewandrowski, K. B. Lewandrowski, B. A. Tannous, A. J. de Langen, E. F. Smit, M. M. van den Heuvel, T. Wurdinger, Swarm Intelligence-Enhanced Detection of Non-Small-Cell Lung Cancer Using Tumor-Educated Platelets., *Cancer Cell* **32**, 238–252.e9 (2017).

41. A. L. Riediger, S. Dietz, U. Schirmer, M. Meister, I. Heinzmann-Groth, M. Schneider, T. Muley, M. Thomas, H. Sültmann, Mutation analysis of circulating plasma DNA to determine response to EGFR tyrosine kinase inhibitor therapy of lung adenocarcinoma patients, *Sci. Rep.* **6**, 33505 (2016).

42. J. Belic, M. Koch, P. Ulz, M. Auer, T. Gerhalter, S. Mohan, K. Fischereder, E. Petru, T. Bauernhofer, J. B. Geigl, M. R. Speicher, E. Heitzer, Rapid Identification of Plasma DNA Samples with Increased ctDNA Levels by a Modified FAST-SeqS Approach, *Clin. Chem.* **61**, 838–849 (2015).

43. D. G. Stover, H. A. Parsons, G. Ha, S. S. Freeman, W. T. Barry, H. Guo, A. D. Choudhury, G. Gydush, S. C. Reed, J. Rhoades, D. Rotem, M. E. Hughes, D. A. Dillon, A. H. Partridge, N. Wagle, I. E. Krop, G. Getz, T. R. Golub, J. C. Love, E. P. Winer, S. M. Tolaney, N. U. Lin, V. A. Adalsteinsson, Association of Cell-Free DNA Tumor Fraction and Somatic Copy Number Alterations With Survival in Metastatic Triple-Negative Breast Cancer., *J. Clin. Oncol.* **36**, 543–553 (2018).

44. B. Routy, E. Le Chatelier, L. Derosa, C. P. M. Duong, M. T. Alou, R. Daillère, A. Fluckiger, M. Messaoudene, C. Rauber, M. P. Roberti, M. Fidelle, C. Flament, V. Poirier-Colame, P. Opolon, C. Klein, K. Iribarren, L. Mondragón, N. Jacquelot, B. Qu, G. Ferrere, C. Clémenson, L. Mezquita, J. R. Masip, C. Naltet, S. Brosseau, C. Kaderbhai, C. Richard, H. Rizvi, F. Levenez, N. Galleron, B. Quinquis, N. Pons, B. Ryffel, V. Minard-Colin, P. Gonin, J.-C. Soria, E. Deutsch, Y. Loriot, F. Ghiringhelli, G. Zalcman, F. Goldwasser, B. Escudier, M. D. Hellmann, A. Eggermont, D. Raoult, L. Albiges, G. Kroemer, L. Zitvogel, Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors., *Science* **359**, 91–97 (2018).

45. C. Abbosh, 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. Grönroos, 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, H. J. W. L. Aerts, A. Hackshaw, J. A. Shaw, B. G. Zimmermann, TRACERx consortium, PEACE consortium, C. Swanton, Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution., *Nature* **545**, 446–451 (2017).

46. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation., *Cell* **144**, 646–

74 (2011).

47. K. M. Patel, K. E. van der Vos, C. G. Smith, F. Mouliere, D. Tsui, J. Morris, D. Chandrananda, F. Marass, D. van den Broek, D. E. Neal, V. J. Gnanapragasam, T. Forshew, B. W. van Rhijn, C. E. Massie, N. Rosenfeld, M. S. van der Heijden, Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer, *Sci. Rep.* **7**, 5554 (2017).

48. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics* **25**, 1754–1760 (2009).

49. I. Scheinin, 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, B. Ylstra, 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**, 2022–2032 (2014).

Acknowledgments: The authors would like to thank all members of the Rosenfeld Lab and Brenton Lab for their help and constructive discussion, in particular Mareike Thompson, Andrea Ruiz-Valdepanas, Jenny P.Y. Chan, and Anja Lisa Riediger. The authors would like to also thank the Cancer Research UK Cambridge Institute core facilities for their support, in particular the genomics, bioinformatics and biorepository facilities. Support is also acknowledged from the Cancer Research UK Cambridge Cancer Centre, the Cambridge Experimental Cancer Medicine Centre (ECMC), Cancer Molecular Diagnostics Laboratory (CMDL) and NIHR Biomedical Research Centre (BRC). We would like to acknowledge our patients and caregivers, and the help and support of the research nurses, trial staff and the staff at Addenbrooke's Hospital and Rigshospitalet. In particular, we would like to acknowledge Charlotte Hodgkin, Heather Biggs and Karen Hosking. We would like to thank Hedley Carr and AstraZeneca for support for the CALIBRATE study.

Funding: We would also like to acknowledge the support of The University of Cambridge, and Cancer Research UK (grant number A11906, A20240). The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 337905. This research is also supported by Target Ovarian Cancer and the Medical Research Council through their Joint Clinical Research Training Fellowship for Dr Moore. The CALIBRATE study was supported by funding from AstraZeneca. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions: FM, AMP, DC, EM, JDB and NR conceptualized and designed the study; FM, AMP, EM, LBA, KH, CGS, JCMW, DG, RM, TG, AS, IG, OO, CAP, MMS, IH, KP, WNC performed experiments and collected data; FM, AMP, DC, EM and CGS conceptualized the size selection approach; FM, AMP and EM designed and performed in vitro size selection; FM and DC conceptualized and designed the fragmentation feature analysis, with input from F. Marass and NR; DC conceptualized and designed the t-MAD index with input from FM; FM and DC carried out bioinformatics analysis of SCNAs from sWGS; JM performed bioinformatics analysis of TAm-Seq; FM and LBA designed the tailored captured sequencing and performed WES; FM and JM performed bioinformatics

analysis of the capture sequencing and WES; ME developed and optimized mutation calling algorithms; RM, KB and SR designed the animal model; JCG, SP, RB, MMS, GDS, JB, SM, PC, CW, RM, MvdH have collected human samples; MJL and J. Burge performed histopathology revision; FM, DC, AMP, EM, JDB and NR wrote the manuscript; all co-authors have critically reviewed the manuscript; FM, AMP, DC, JDB and NR supervised the study; FM coordinated the study.

Competing interests: NR, JDB and DG are cofounders, shareholders and officers/consultants of Inivata Ltd, a cancer genomics company that commercializes ctDNA analysis. Inivata Ltd had no role in the conceptualization, study design, data collection and analysis, decision to publish or preparation of the manuscript. FM, NR and F. Marass are co-inventors of patent applications that describe methods for the analysis of DNA fragments and applications of circulating tumor DNA. Other co-authors have declared no conflict of interests.

Data and materials availability: Sequencing data for this study are deposited in the EGA database, accession number ega-box-1048.

Figures

Figure 1: Survey of plasma DNA fragmentation with genome-wide sequencing on a pan-cancer scale. A, The size profile of cfDNA can be determined from paired-end sequencing of plasma samples and reflects its organization around the nucleosome. cfDNA is released in the blood circulation by various means, each of which leaves a signature on the fragment sizes. We inferred the size profile of cfDNA by analyzing with sWGS (n=344 plasma samples from 65 healthy controls and 200 cancer patients), and the size profile of mutant ctDNA by personalized capture sequencing (n=18 plasma samples). **B**, Fragment size distributions of 344 plasma samples from 200 cancer patients. Patients are split into two groups based on previous literature (*3*), orange representing cancer samples previously observed to have low levels of ctDNA (renal, bladder, pancreatic, and glioma) and blue representing cancer samples observed to have higher ctDNA levels (breast, melanoma, ovarian, lung, colorectal, cholangiocarcinoma, and others, see Suppl. Table 1). **C**, Proportion of cfDNA fragments below 150bp by cancer grouping defined in **B.** The Kruskal-Wallis test for difference in size distributions indicated a significant difference between the group of cancer types releasing high amounts of ctDNA, and the group releasing low amounts as well as the group of healthy individuals ($p<0.001$). **D**, Proportion of cfDNA fragments below 150 bp by cancer type (all samples). Cancer types represented by fewer than 4 individuals are grouped in the “other” category. The red line indicates the median proportion per cancer type.

Figure 2: Determining the size profile of mutant ctDNA with animal models and personalized capture sequencing. A, A mouse model with xenografted human tumor cells enabled the discrimination of DNA fragments released by cancer cells (reads aligning to the human genome) from the DNA released by healthy cells (reads aligning to the mouse

genome), with the use of sWGS. **B**, Fragment size distribution, from the plasma extracted from a mouse xenografted with a human ovarian tumor, showing ctDNA originating from tumor cells (red) and cfDNA from non-cancerous cells (blue). Two vertical lines indicate 145bp and 167bp. The fraction of reads shorter than 150 bp is indicated. **C**, Design of personalized hybrid-capture sequencing panels developed to specifically determine the size profiles of mutant DNA and non-mutant DNA in plasma from 19 patients with late stage cancers. Capture panels included somatic mutations identified in tumor tissue by WES. A mean of 165 mutations per patient were then analyzed from matched plasma samples. Reads were aligned and separated into fragments that carry either the reference or the mutant sequence. Fragment sizes for paired-end reads were calculated. **D**, Size profiles of mutant DNA and non-mutant DNA in plasma from 19 patients with late stage cancers were determined by tumor-guided capture sequencing. The fraction of reads shorter than 150 bp is indicated.

Figure 3: Enhancing the tumor fraction from plasma sequencing with size selection. **A**, Plasma samples collected from ovarian cancer patients were analyzed in parallel without size selection, or using either in silico and in vitro size selection. **B**, accuracy of the in vitro and in silico size selection determined on a cohort of 20 healthy controls. The size distribution before size selection is shown in green, after in silico size selection (with sharp cutoff at 90 and 150 bp) in blue and after in vitro size selection in orange. **C**, SCNA analysis with sWGS from plasma DNA of an ovarian cancer patient collected before initiation of treatment, when ctDNA MAF was 0.271 for a TP53 mutation as determined by TAm-Seq. Inferred amplifications are shown in blue and deletions in orange. Copy number neutral regions are in grey. **D**, SCNA analysis of a plasma sample from the same patient as panel **C** collected three weeks after treatment start. The MAF for the TP53 mutation was 0.068, and ctDNA was not detected at this time-point by sWGS (before size selection). **E**, Analysis of the same plasma sample as **D** after in vitro size selection of fragments between 90 bp and 150 bp in length. The MAF for the TP53 mutation increased to 0.402 after in vitro size selection, and SCNAs were clearly apparent by sWGS. More SCNAs are detected in comparison to C and D (e.g. in chr2, chr9, chr10). SCNAs were also detected in this sample after in silico size selection (**Suppl. Fig. 7**).

Figure 4: Quantifying the ctDNA enrichment by sWGS with in silico size selection and t-MAD. A, Workflow to quantify tumor fraction from SCNA as a genome-wide score named t-MAD. **B**, Correlation between the MAF of SNVs determined by digital PCR or hybrid-capture sequencing and t-MAD score determined by sWGS. Data included 97 samples from cancer patients of multiples cancer types with matched MAF measurements and t-MAD scores. Pearson correlation (coefficient *r*) between MAF and t-MAD scores was calculated for all cases with MAF>0.025 and t-MAD>0.015. Linear regression indicated a fit with a slope of 0.44 (purple solid line). **C**, Comparison of t-MAD scores determined from sWGS between healthy samples, samples collected from patients with cancer types that exhibit low amounts of ctDNA in circulation and from patients with cancer types that exihibit high amounts of ctDNA in circulation (as in Fig. 1). All samples for which t-MAD could be calculated have been included. **D**, ROC analysis comparing the classification of these plasma samples from high ctDNA cancer samples (n=189) and plasma samples from healthy controls (n=65) using t-MAD had an area under curve (AUC) of 0.69 without size selection (black solid curve).

After applying in silico size selection to the samples from the cancer patients, we observed an AUC of 0.90 (black dashed curve). **E**, Determination of t-MAD from longitudinal plasma samples of a colorectal cancer patient. t-MAD was analyzed before and after in silico size selection of the DNA fragments 90-150bp, and then compared to the RECIST status for this patient. **F**, Application of in silico size selection to 6 patients with long follow-up. t-MAD score was determined before and after in silico size selection of the short DNA fragments. Dark blue circles indicate samples in which ctDNA was detected both with and without in silico size selection. Light blue circles indicate samples where ctDNA was detected only after in silico size selection. Empty circles indicate samples where ctDNA was not detected by either analysis. Times when RECIST status was assessed are indicated by a red bar for progression, or an orange bar for regression or stable disease.

Figure 5: Quantifying the ctDNA enrichment by sWGS with in vitro size selection. A, The effect of in vitro size selection on the t-MAD score. For each of 48 plasma samples collected from 35 patients, the t-MAD score was determined from the sWGS after in vitro size selection (y axis) and without size selection (x axis). In vitro size selection increased the t-MAD score for nearly all samples, with a median increase of 2.1-fold (range from 1.1 to 6.4 fold). t-MAD scores determined from sWGS for 46 samples from healthy individuals were all <0.015 both before and after in vitro size selection. **B**, ROC analysis comparing the classification of these plasma samples from cancer samples (n=48) and plasma samples from healthy controls (n=46) using t-MAD had an area under curve (AUC) of 0.64 without size selection (green curve). After applying in silico size selection to the samples from the cancerous and healthy patients, we observed an AUC of 0.78 (blue curve), and after in vitro size selection, an AUC of 0.97 (orange curve). **C**, Comparison of t-MAD scores determined from sWGS between matched ovarian cancer samples with and without in vitro size selection. The t-test for the difference in means indicate a significant increase in tumor fraction (measured by t-MAD) with in vitro size selection ($p<0.0001$). **D**, Detection of SCNAs across 15 genes frequently mutated in recurrent ovarian cancer, measured in plasma samples collected during treatment for 35 patients. Patients were ranked from left to right by increasing tumor fraction as quantified by tMAD (before in vitro size selection). SCNAs are labelled as detected for a gene if the relative copy number in that region was greater than 0.05. Empty squares represent copy number neutral regions, bottom left triangles in light blue indicate that SCNAs were detected without size selection and top right triangles in dark blue represent SCNAs detected after in vitro size selection.

Figure 6: Improving the detection of somatic alterations by WES in multiple cancer types with size selection. A, Analysis of the MAF of mutations detected by WES in 6 patients with HGSOC without size selection and with in vitro and in silico size selection. **B**, Comparison of size-selected WES data with non-selected WES data to assess the number of mutations detected in plasma samples from 6 patients with HGSOC. For each patient, the first bar in light blue shows the number of mutations called without size selection, the second bar quantifies the number of mutations called after the addition of those identified with in silico size selection, and the third, dark blue bar shows the number of mutations called after addition of mutations called after in vitro size selection. **C**, Patients (n=16) were retrospectively selected from a cohort with different cancer types (colorectal, cholangiocarcinoma, pancreatic, prostate) enrolled in early phase clinical trials. Matched

tumor tissue DNA was available for each plasma sample, and 2 patients also had a biopsy collected at relapse. WES was performed on tumor tissue DNA and plasma DNA samples, and in silico size selection was applied to the data. 2061/2133, 97% of the shared mutations detected by WES showed higher MAF after in silico size selection. **D**, Mutations detected only after in silico selection of WES data from 16 patients (as in **C**) compared to mutations called by WES of the matched tumor tissue. Three of 16 patients had no additional mutations identified after in silico size selection. Of the 82 mutations detected in plasma after in silico size selection, 23 (28%) had low signal levels in tumor WES data and were not initially identified in those samples.

Figure 7: Enhancing the potential for ctDNA detection by combining SCNAs and fragment-size features. A, Schematic illustrating the selection of different size ranges and features in the distribution of fragment sizes. For each sample, fragmentation features included the proportion (P) of fragments in specific size ranges, the ratio between certain ranges and a quantification of the amplitude of the 10bp oscillations in the 90-145 size bp range calculated from the periodic "peaks" and "valleys" (see **Methods**). **B**, Principal Component Analysis (PCA) comparing cancer and healthy samples using data from t-MAD scores and the fragmentation features. Fragmentation features shown in grey are not included in the following steps. **C**, Workflow for the predictive analysis combining SCNAs and fragment size features (see **Methods**). Plasma DNA sWGS data from healthy controls was split into a training set (60% of samples) and a validation set (used in both Validation data 1 and Validation set 2). sWGS data from plasma samples from a pan-cancer cohort of 182 samples from patients with cancer types with high levels of ctDNA (colorectal, cholangiocarcinoma, lung, ovarian, breast) was split into a training set (60% of samples) and a validation set (Validation data 1, together with the healthy individual validation set). A further dataset of sWGS from 57 samples from cancer types exhibiting low levels of ctDNA (glioma, renal, pancreatic) was used as Validation data 2, together with the healthy individual validation set. **D**, ROC curves for Validation data 1 (samples from cancer patients with high ctDNA levels=68, healthy=26) for 3 predictive models built on the pan-cancer training cohort (cancer=114, healthy=39). The beige curve represents the ROC curve for classification with t-MAD only, the long dashed green line represents the logistic regression model combining the top 5 features based on recursive feature elimination (t-MAD score, 10bp amplitude, P(160-180), P(180-220) and P(250-320)), and the dashed red line shows the result for a random forest classifier trained on the combination of the same 5 features, independently chosen for the best RF predictive model. **E,** ROC curves for Validation data 2 (samples from cancer patients with low ctDNA levels=57, healthy=26) for the same 3 classifiers as **D**. The beige curve represents the model using t-MAD only, the long-dashed green represents the logistic regression model combining the top 5 features (t-MAD score, 10bp amplitude, P(160-180), P(180-220), and P(250-320)), and the dashed red shows the result for a random forest classifier trained on the combination of same 5 predictive features. **F**, Plot representing the probability of classification as cancer with the RF model for all samples in both validation datasets. Samples are separated by cancer type and sorted within each by the RF probability of classification as cancer. The dashed horizontal line indicates 50% probability and the light long-dashed line indicates 33% probability.

Figure 1

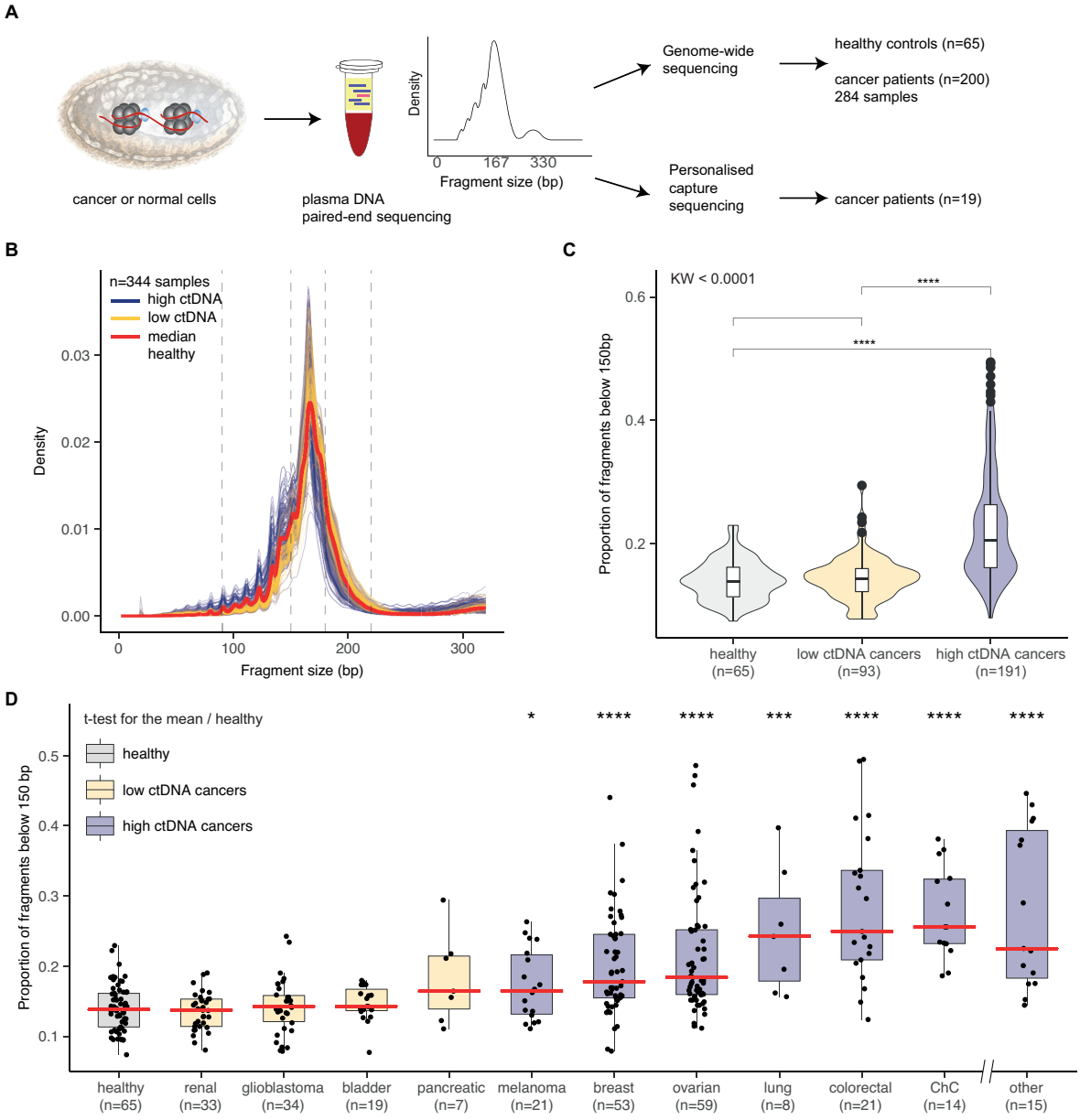

A
cancer or normal cells
plasma DNA
paired-end sequencing
Density
0
167
330
Fragment size (bp)
Genome-wide
sequencing
healthy controls (n=65)
cancer patients (n=200)
284 samples
Personalised
capture
sequencing
cancer patients (n=19)
B
n=344 samples
high ctDNA
low ctDNA
median
healthy
Density
Fragment size (bp)
C
KW < 0.0001
****
****
Proportion of fragments below 150bp
healthy
(n=65)
low ctDNA cancers
(n=93)
high ctDNA cancers
(n=191)
D
t-test for the mean / healthy
healthy
low ctDNA cancers
high ctDNA cancers
Proportion of fragments below 150 bp
healthy
(n=65)
renal
(n=33)
glioblastoma
(n=34)
bladder
(n=19)
pancreatic
(n=7)
melanoma
(n=21)
breast
(n=53)
ovarian
(n=59)
lung
(n=8)
colorectal
(n=21)
ChC
(n=14)
other
(n=15)

Figure 2

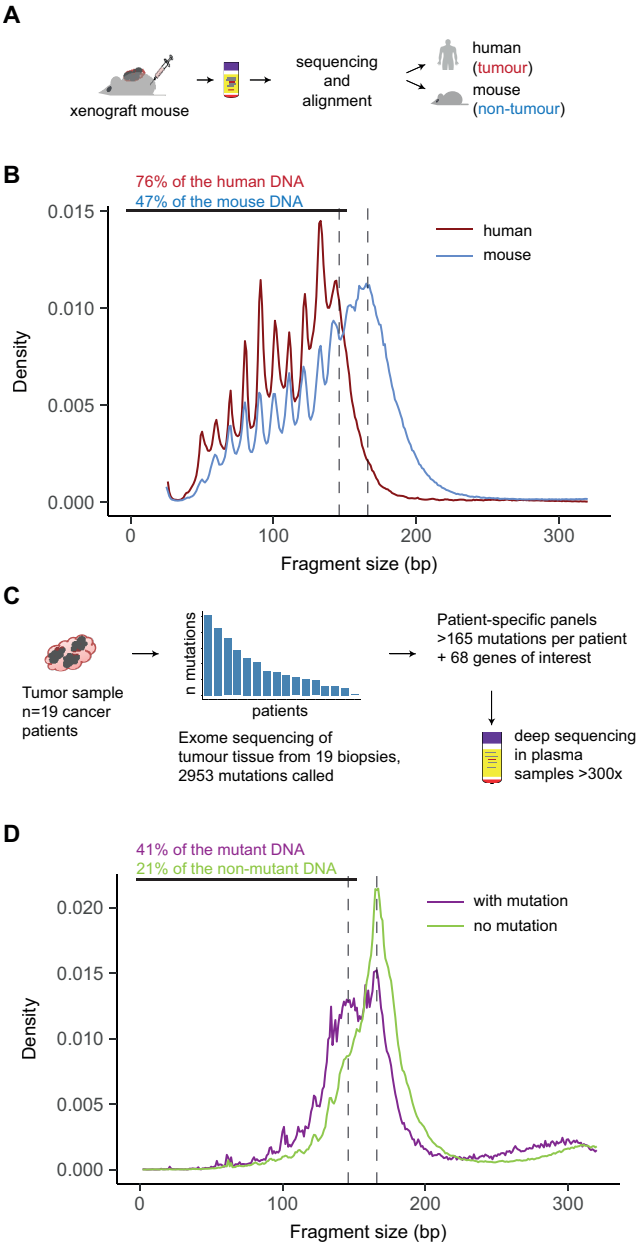

A
xenograft mouse
sequencing
and
alignment
human
(tumour)
mouse
(non-tumour)
B
76% of the human DNA
47% of the mouse DNA
human
mouse
Density
0.015
0.010
0.005
0.000
0
100
200
300
Fragment size (bp)
C
Tumor sample
n=19 cancer
patients
n mutations
patients
Exome sequencing of
tumour tissue from 19 biopsies,
2953 mutations called
Patient-specific panels
>165 mutations per patient
+ 68 genes of interest
deep sequencing
in plasma
samples >300x
D
41% of the mutant DNA
21% of the non-mutant DNA
with mutation
no mutation
Density
0.020
0.015
0.010
0.005
0.000
0
100
200
300
Fragment size (bp)

Figure 3

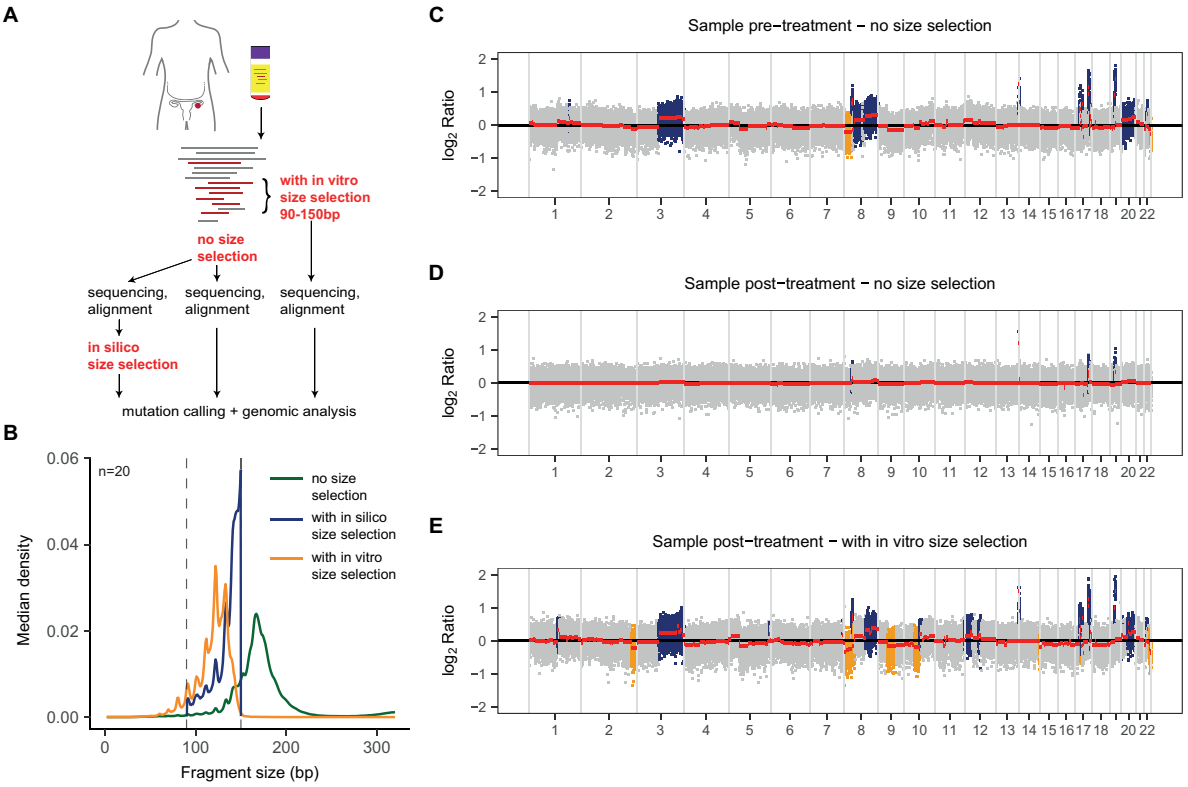

A
with in vitro
size selection
90-150bp
no size
selection
sequencing,
alignment
sequencing,
alignment
sequencing,
alignment
in silico
size selection
mutation calling + genomic analysis
B
n=20
no size
selection
with in silico
size selection
with in vitro
size selection
Median density
Fragment size (bp)
C
Sample pre−treatment − no size selection
D
Sample post−treatment − no size selection
E
Sample post−treatment − with in vitro size selection
log2 Ratio

Figure 4

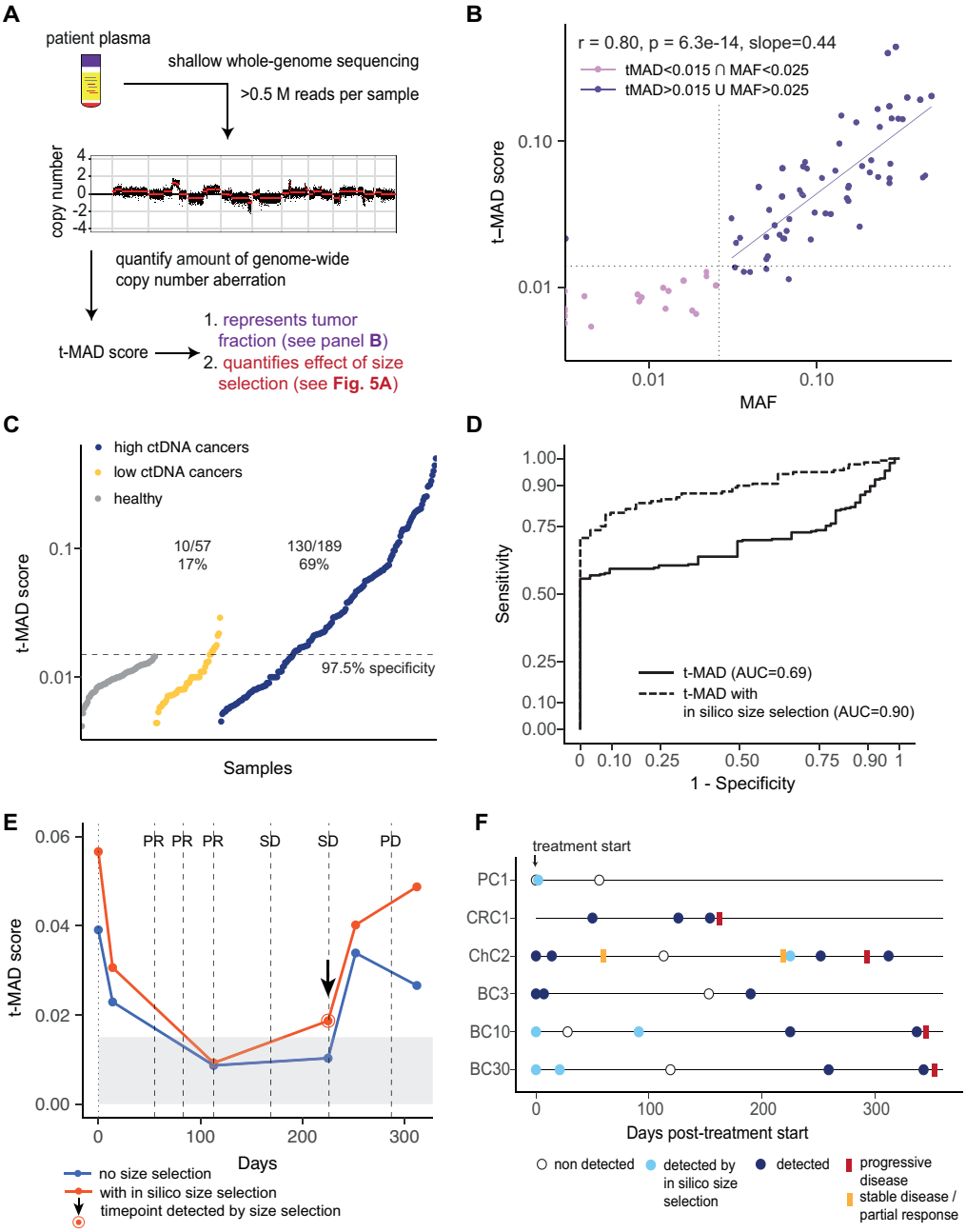

A
patient plasma
shallow whole-genome sequencing
>0.5 M reads per sample
copy number
quantify amount of genome-wide copy number aberration
t-MAD score
1. represents tumor fraction (see panel B)
2. quantifies effect of size selection (see Fig. 5A)
B
r = 0.80, p = 6.3e-14, slope=0.44
tMAD<0.015 ∩ MAF<0.025
tMAD>0.015 U MAF>0.025
t-MAD score
MAF
C
high ctDNA cancers
low ctDNA cancers
healthy
10/57
17%
130/189
69%
97.5% specificity
t-MAD score
Samples
D
Sensitivity
t-MAD (AUC=0.69)
t-MAD with
in silico size selection (AUC=0.90)
1 - Specificity
E
PR PR PR SD SD PD
t-MAD score
Days
no size selection
with in silico size selection
timepoint detected by size selection
F
treatment start
PC1
CRC1
ChC2
BC3
BC10
BC30
Days post-treatment start
non detected
detected by in silico size selection
detected
progressive disease
stable disease / partial response

Figure 5

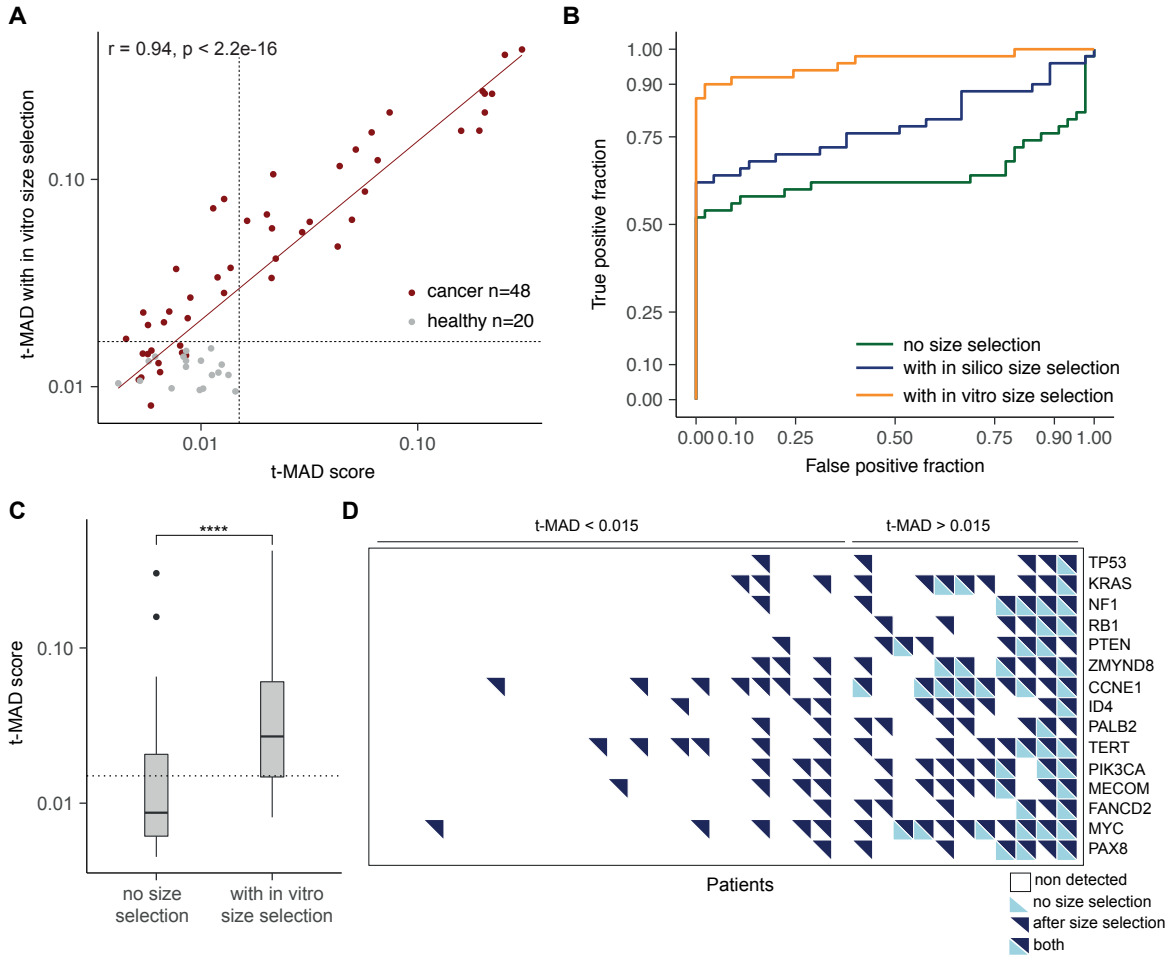

A
r = 0.94, p < 2.2e-16
t-MAD with in vitro size selection
t-MAD score
cancer n=48
healthy n=20
B
True positive fraction
False positive fraction
no size selection
with in silico size selection
with in vitro size selection
C
****
t-MAD score
no size selection
with in vitro size selection
D
t-MAD < 0.015
t-MAD > 0.015
TP53
KRAS
NF1
RB1
PTEN
ZMYND8
CCNE1
ID4
PALB2
TERT
PIK3CA
MECOM
FANCD2
MYC
PAX8
Patients
non detected
no size selection
after size selection
both

Figure 6

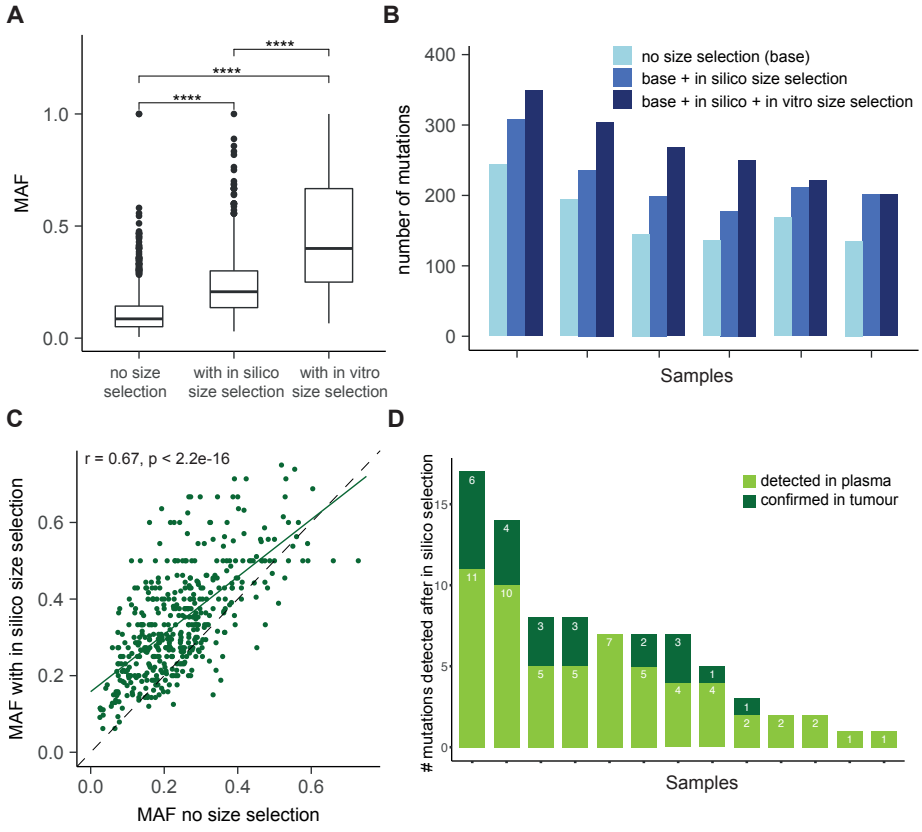
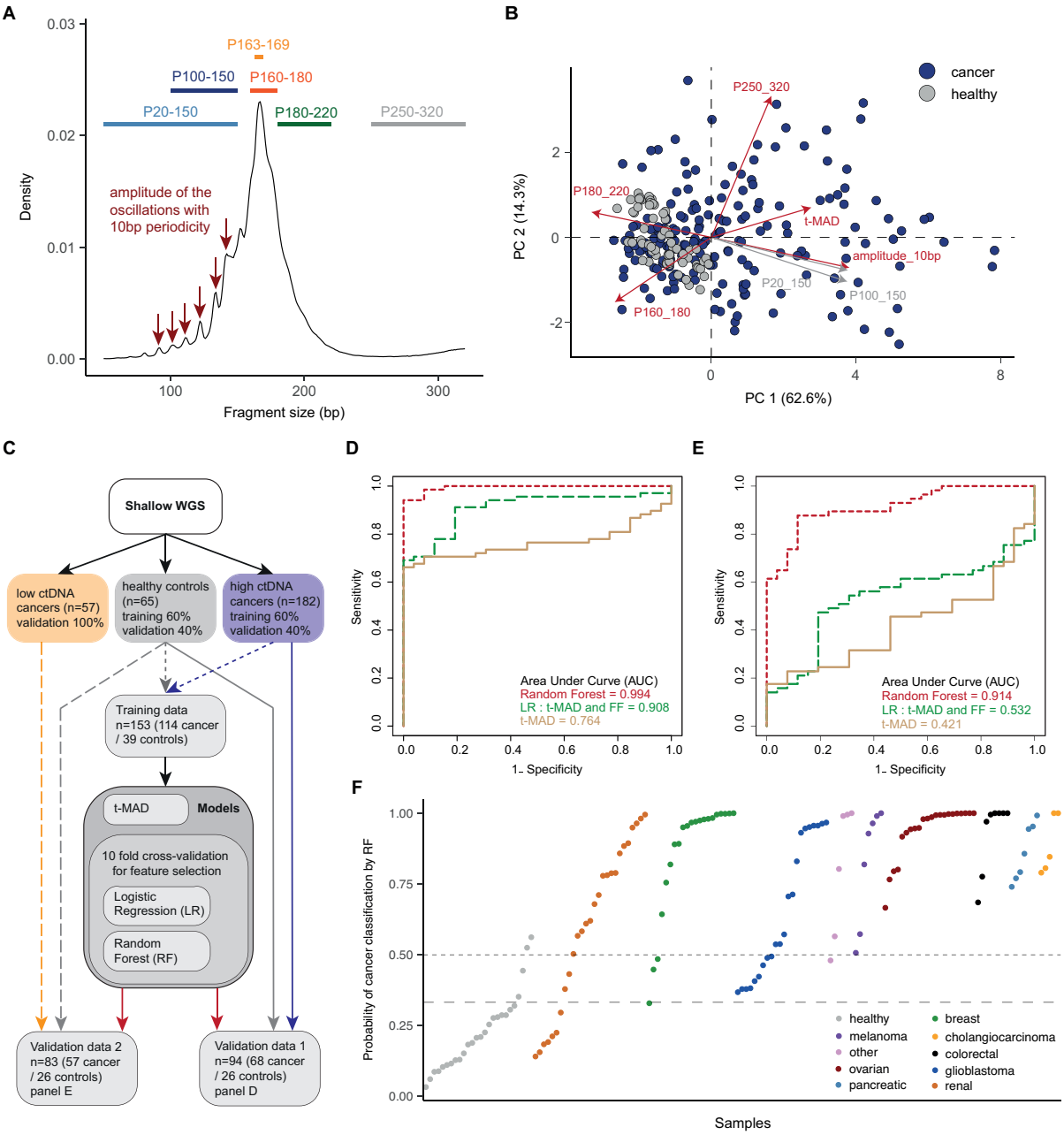
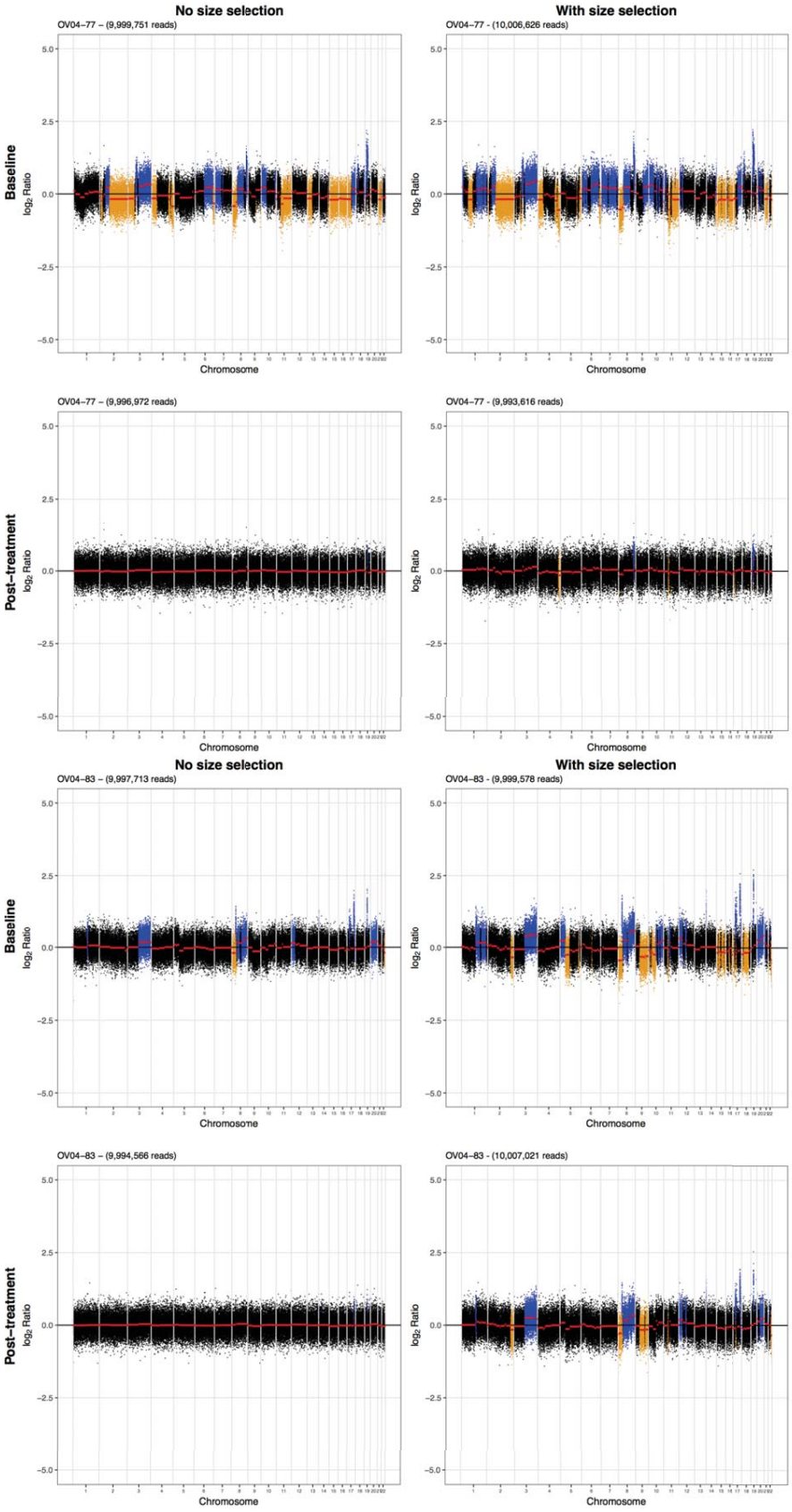
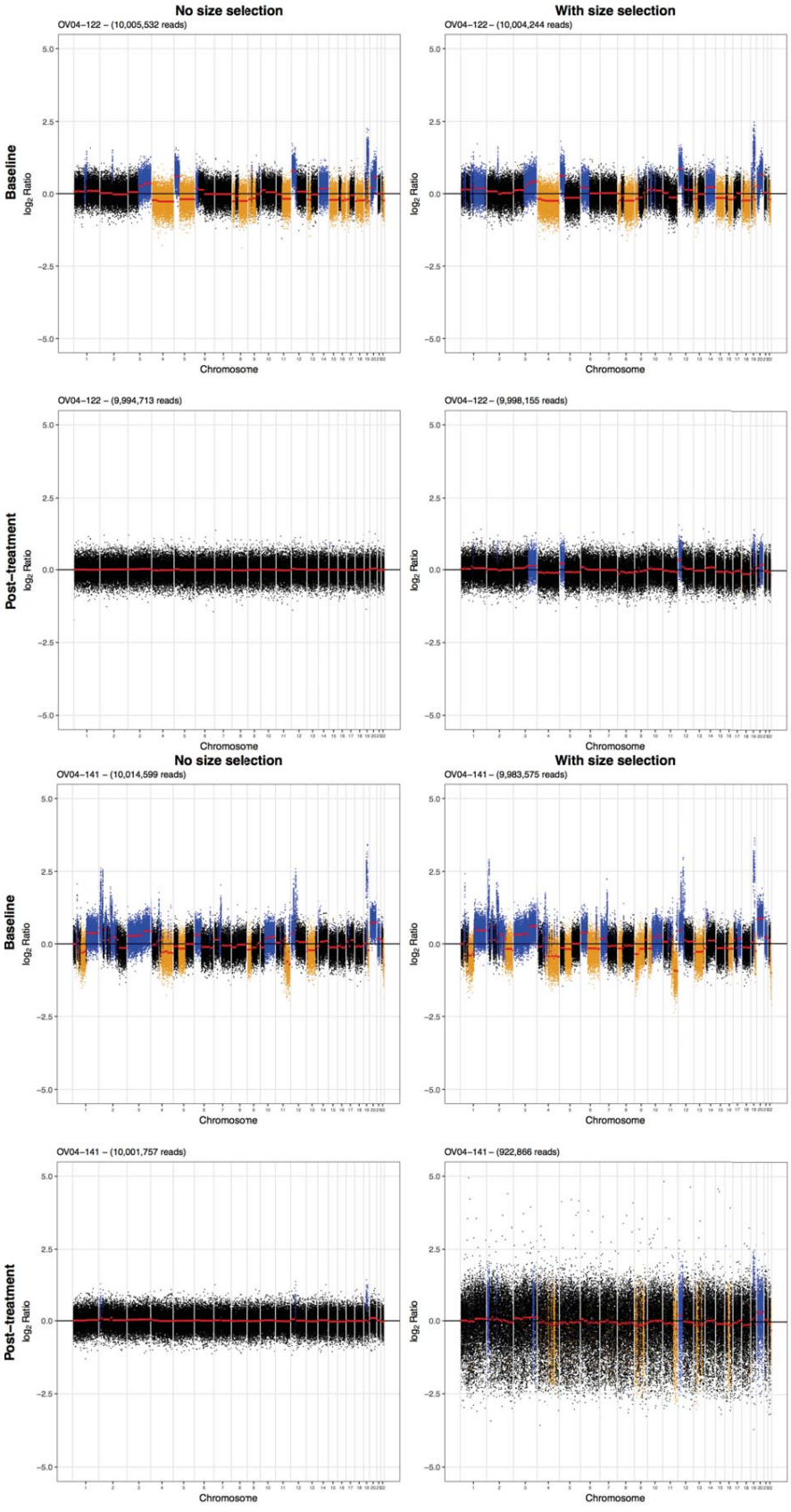

A
B
C
D
MAF
1.0
0.5
0.0
****
****
****
no size selection
with in silico size selection
with in vitro size selection
number of mutations
400
300
200
100
0
no size selection (base)
base + in silico size selection
base + in silico + in vitro size selection
Samples
r = 0.67, p < 2.2e-16
MAF with in silico size selection
MAF no size selection
0.0
0.2
0.4
0.6
mutations detected after in silico selection
detected in plasma
confirmed in tumour
Samples

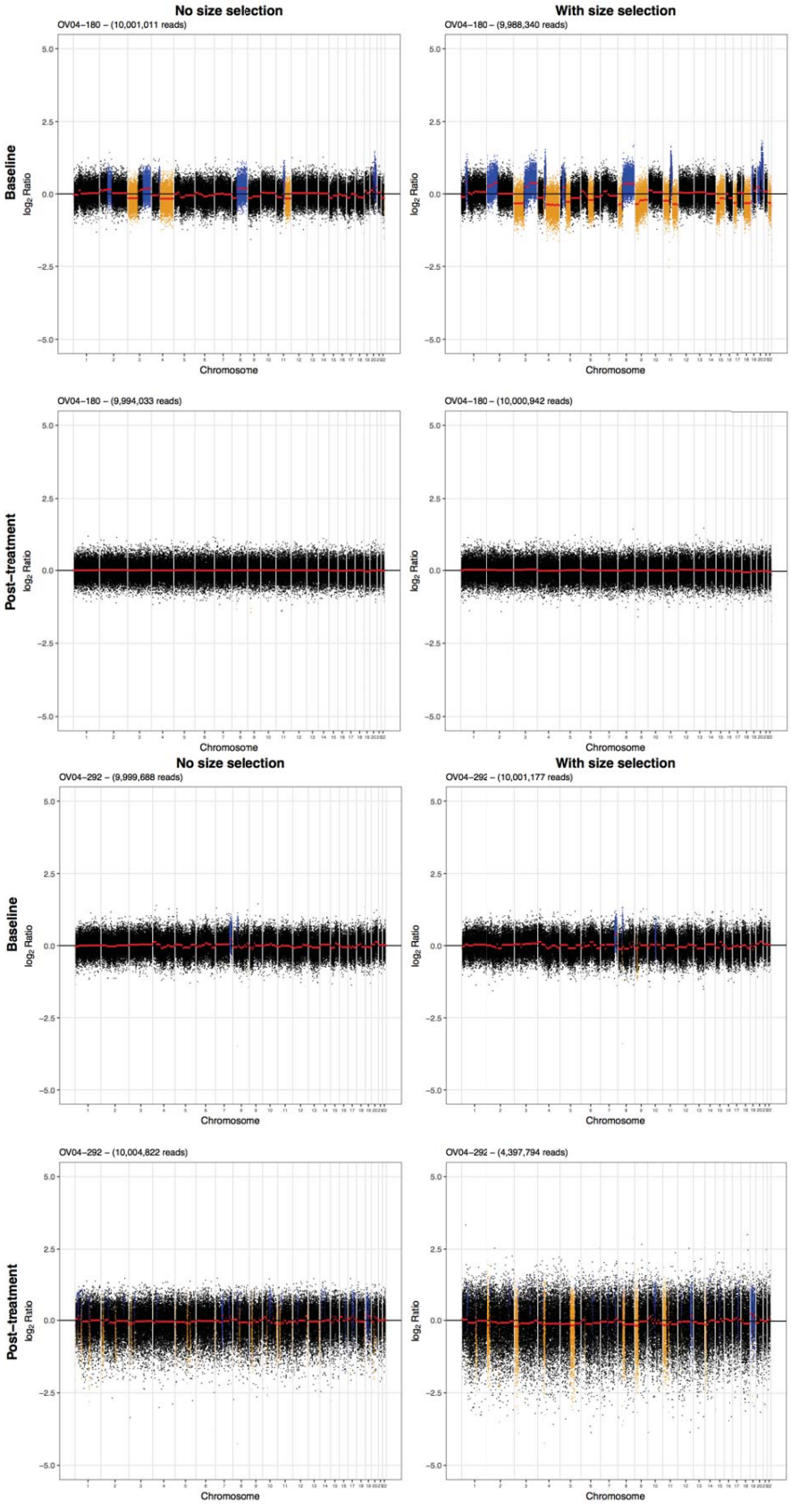
Figure 7

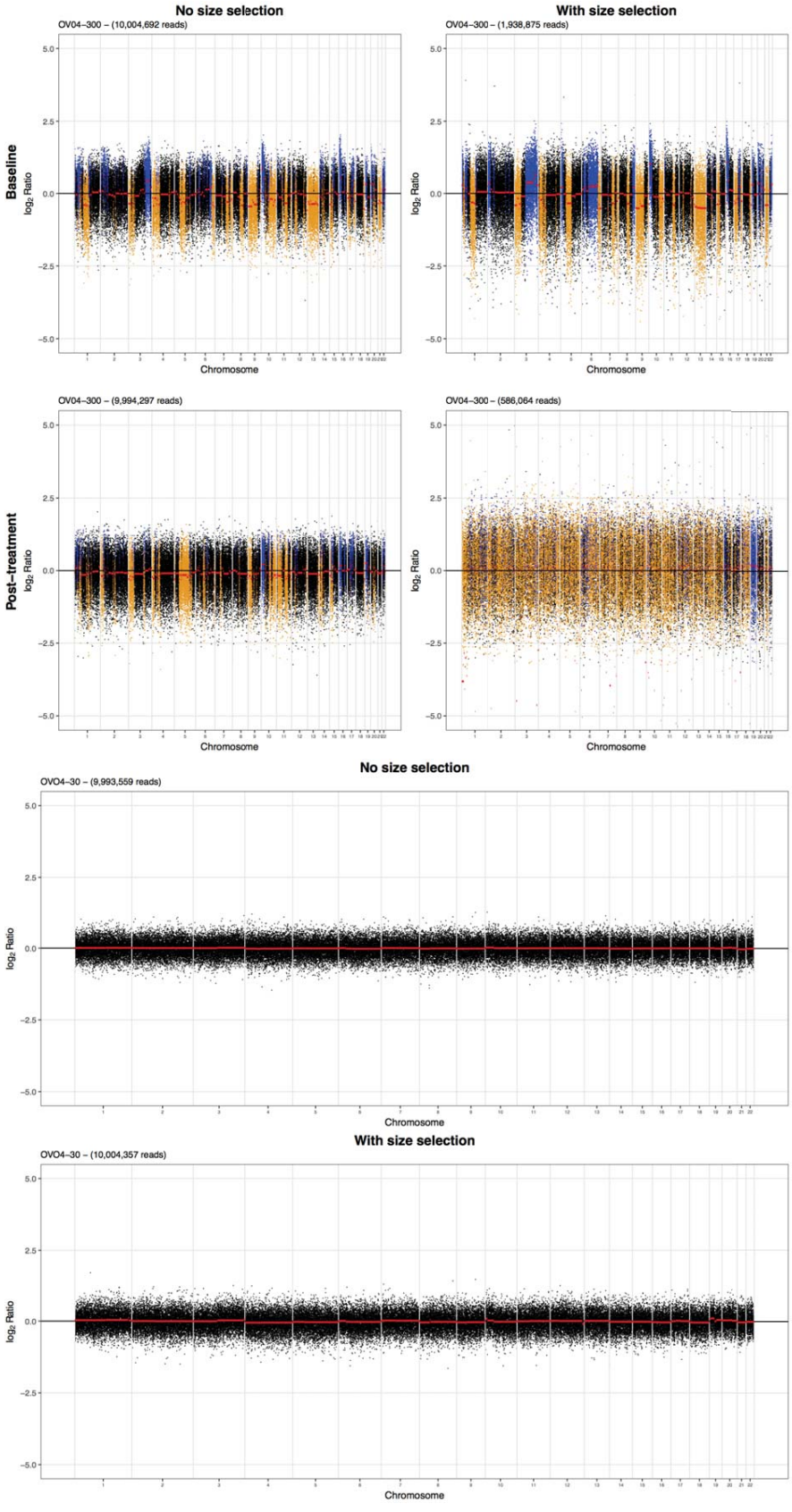

A
P163-169
P100-150
P160-180
P20-150
P180-220
P250-320
amplitude of the oscillations with 10bp periodicity
Density
Fragment size (bp)
B
cancer
healthy
P250_320
P180_220
t-MAD
amplitude_10bp
P20_150
P100_150
P160_180
PC 2 (14.3%)
PC 1 (62.6%)
C
Shallow WGS
low ctDNA cancers (n=57) validation 100%
healthy controls (n=65) training 60% validation 40%
high ctDNA cancers (n=182) training 60% validation 40%
Training data n=153 (114 cancer / 39 controls)
t-MAD
Models
10 fold cross-validation for feature selection
Logistic Regression (LR)
Random Forest (RF)
Validation data 2 n=83 (57 cancer / 26 controls) panel E
Validation data 1 n=94 (68 cancer / 26 controls) panel D
D
Sensitivity
1- Specificity
Area Under Curve (AUC)
Random Forest = 0.994
LR : t-MAD and FF = 0.908
t-MAD = 0.764
E
Sensitivity
1- Specificity
Area Under Curve (AUC)
Random Forest = 0.914
LR : t-MAD and FF = 0.532
t-MAD = 0.421
F
Probability of cancer classification by RF
Samples
healthy
melanoma
other
ovarian
pancreatic
breast
cholangiocarcinoma
colorectal
glioblastoma
renal

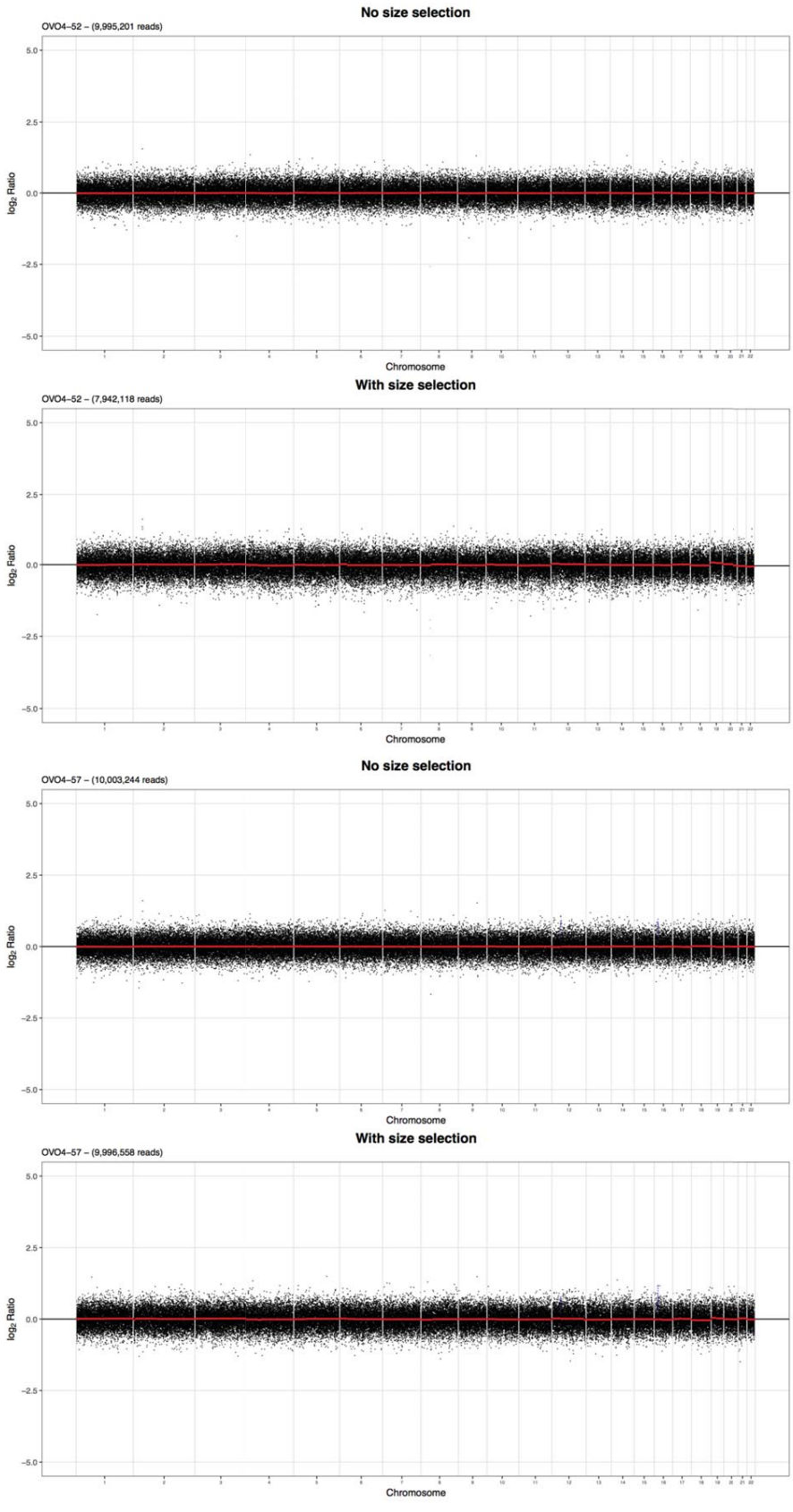
12.14 Copy number profiles for relapsed HGSOC cases with and without size selection

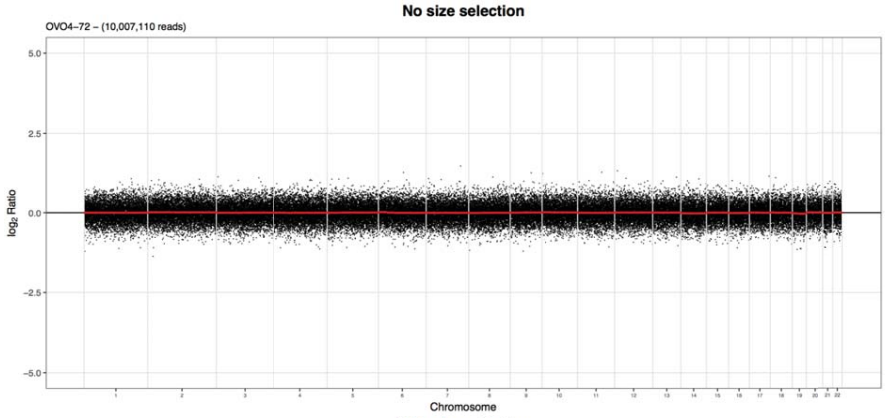


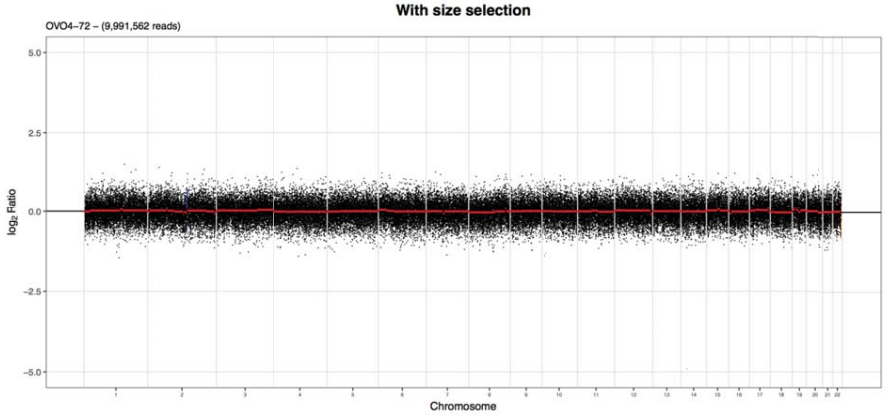

No size selection
With size selection
OV04–122 – (10,005,532 reads)
OV04–122 – (10,004,244 reads)
OV04–122 – (9,994,713 reads)
OV04–122 – (9,998,155 reads)
OV04–141 – (10,014,599 reads)
OV04–141 – (9,983,575 reads)
OV04–141 – (10,001,757 reads)
OV04–141 – (922,866 reads)
Baseline
Post–treatment
log2 Ratio
Chromosome

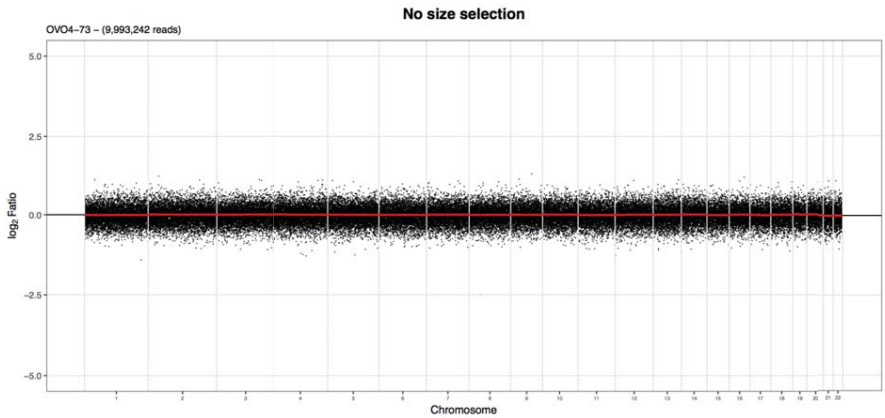

No size selection
With size selection
OV04−180 − (10,001,011 reads)
OV04−180 − (9,988,340 reads)
OV04−180 − (9,994,033 reads)
OV04−180 − (10,000,942 reads)
OV04−292 − (9,999,688 reads)
OV04−292 − (10,001,177 reads)
OV04−292 − (10,004,822 reads)
OV04−292 − (4,397,794 reads)
Baseline
Post−treatment
log2 Ratio
Chromosome

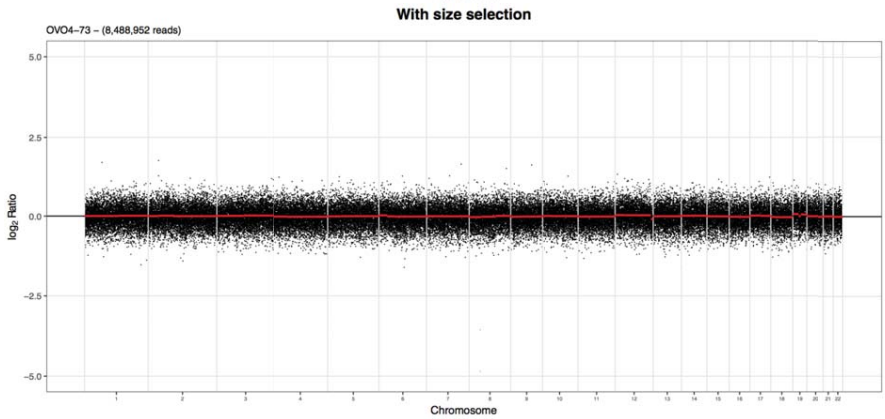

No size selection
With size selection
OV04−300 − (10,004,692 reads)
OV04−300 − (1,938,875 reads)
OV04−300 − (9,994,297 reads)
OV04−300 − (586,064 reads)
Baseline
Post−treatment
log2 Ratio
Chromosome
No size selection
With size selection
OVO4−30 − (9,993,559 reads)
OVO4−30 − (10,004,357 reads)

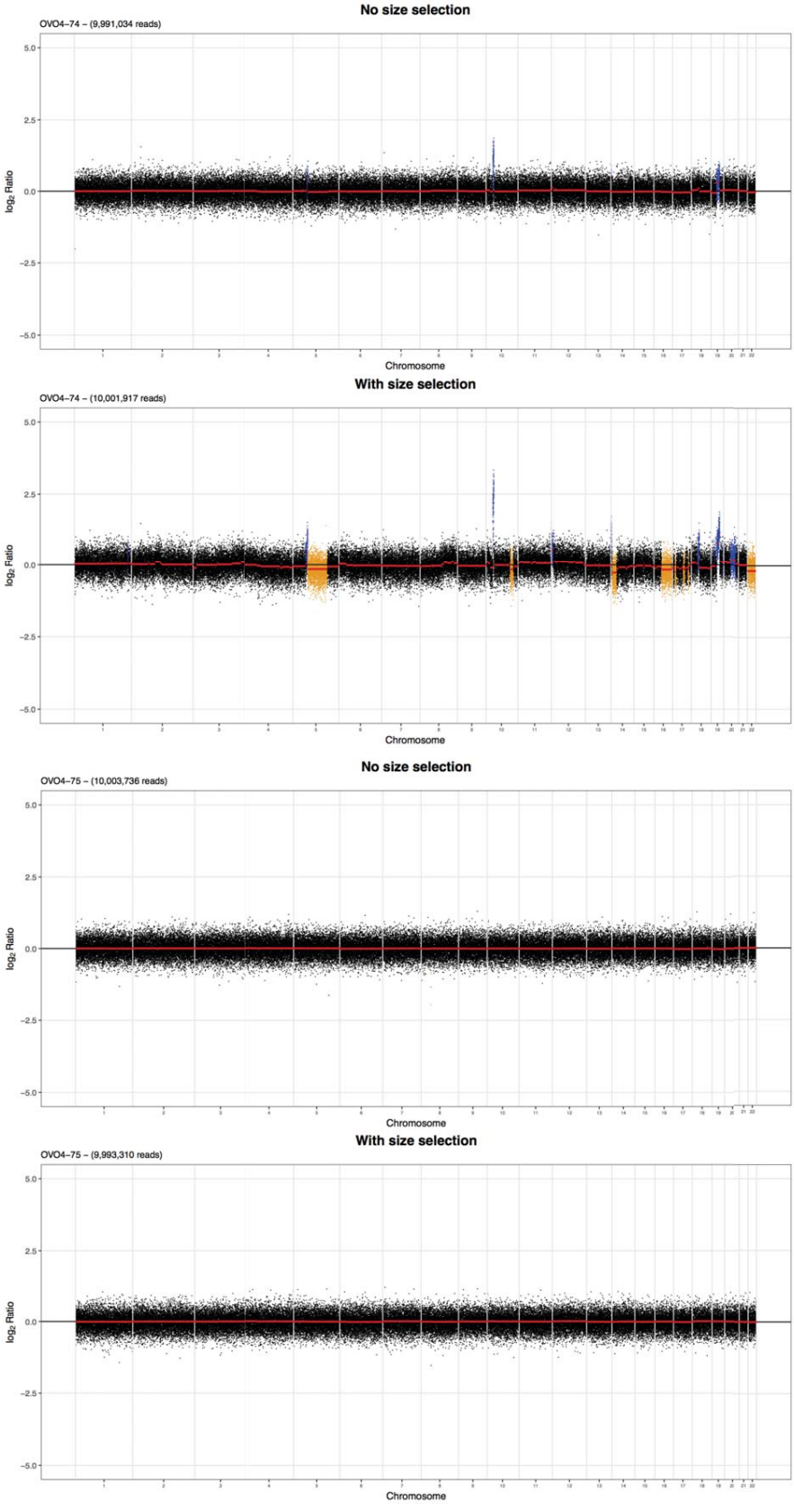

No size selection
OVO4–52 – (9,995,201 reads)
With size selection
OVO4–52 – (7,942,118 reads)
No size selection
OVO4–57 – (10,003,244 reads)
With size selection
OVO4–57 – (9,996,558 reads)
log2 Ratio
Chromosome


No size selection
OVO4–72 – (10,007,110 reads)
log2 Ratio
5.0
2.5
0.0
−2.5
−5.0
Chromosome

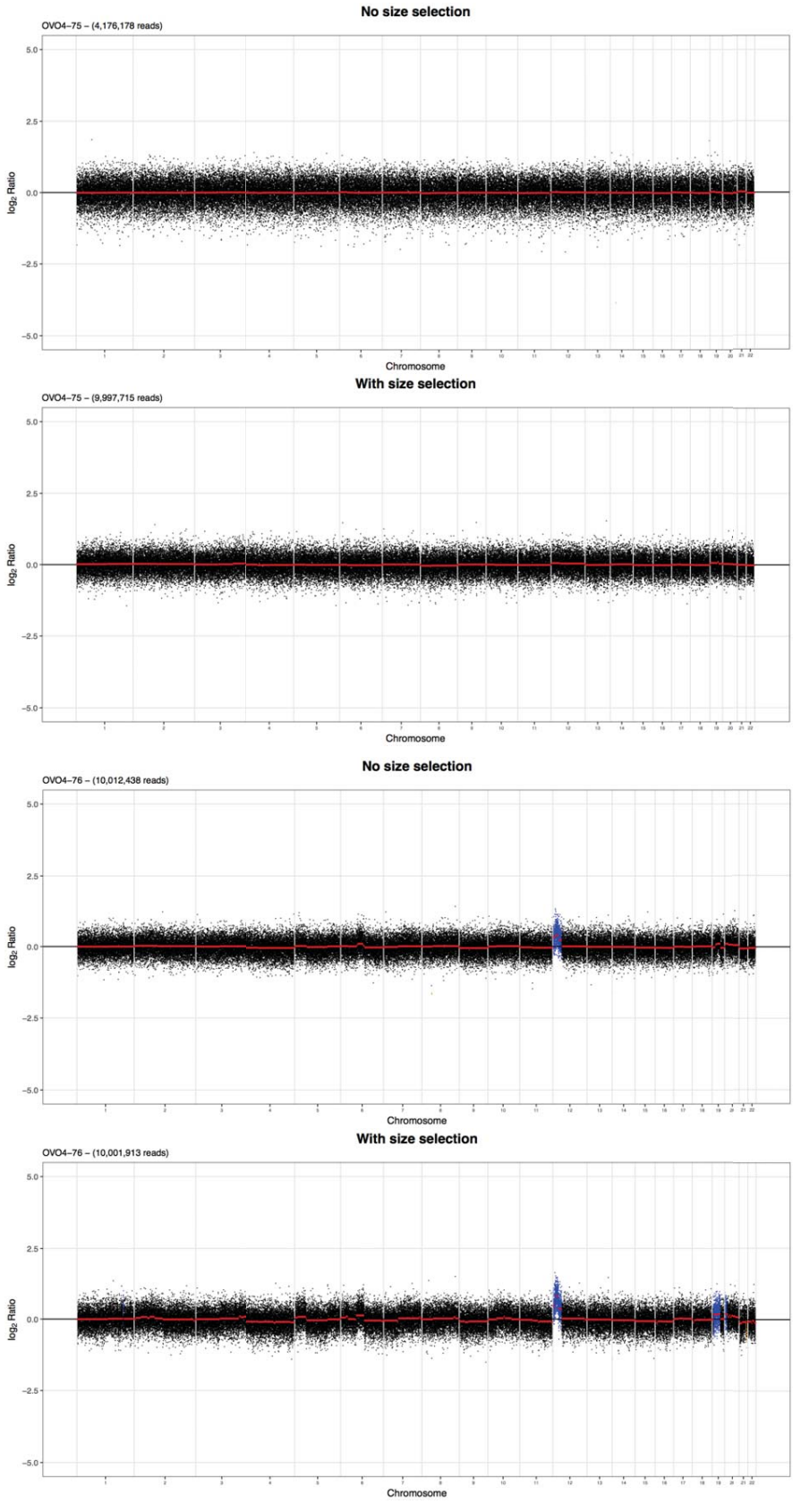

With size selection
OVO4–72 – (9,991,562 reads)
log2 Ratio
5.0
2.5
0.0
−2.5
−5.0
Chromosome


No size selection
OVO4–73 – (9,993,242 reads)
log2 Ratio
5.0
2.5
0.0
−2.5
−5.0
Chromosome

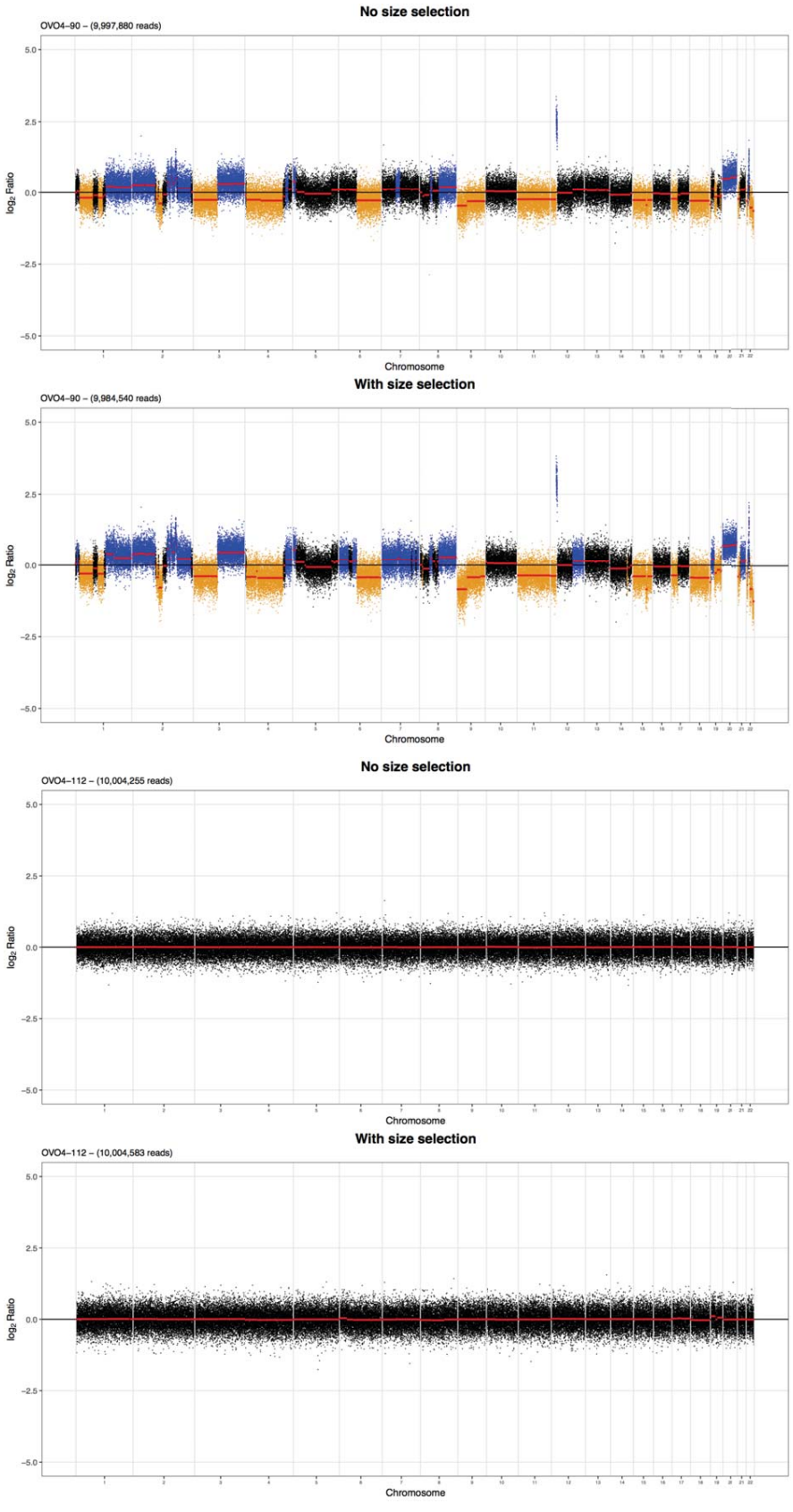

With size selection
OVO4–73 – (8,488,952 reads)
log2 Ratio
5.0
2.5
0.0
−2.5
−5.0
Chromosome

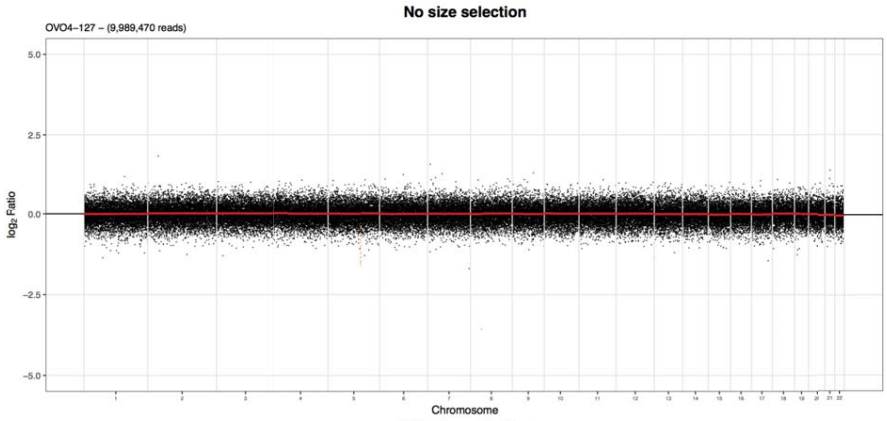

No size selection
OVO4–74 – (9,991,034 reads)
With size selection
OVO4–74 – (10,001,917 reads)
No size selection
OVO4–75 – (10,003,736 reads)
With size selection
OVO4–75 – (9,993,310 reads)
log2 Ratio
Chromosome

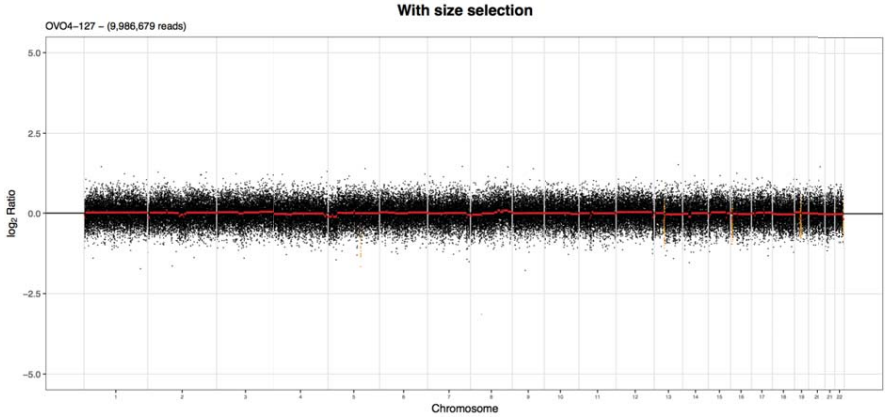
C

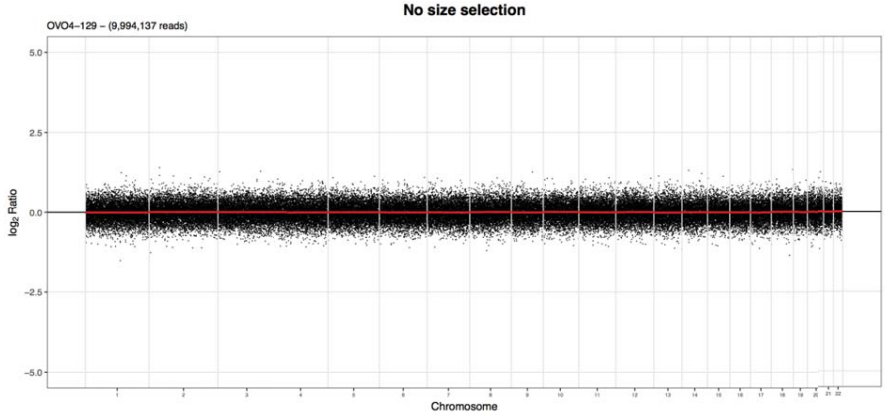

No size selection
OVO4–75 – (4,176,178 reads)
log2 Ratio
Chromosome
With size selection
OVO4–75 – (9,997,715 reads)
log2 Ratio
Chromosome
No size selection
OVO4–76 – (10,012,438 reads)
log2 Ratio
Chromosome
With size selection
OVO4–76 – (10,001,913 reads)
log2 Ratio
Chromosome

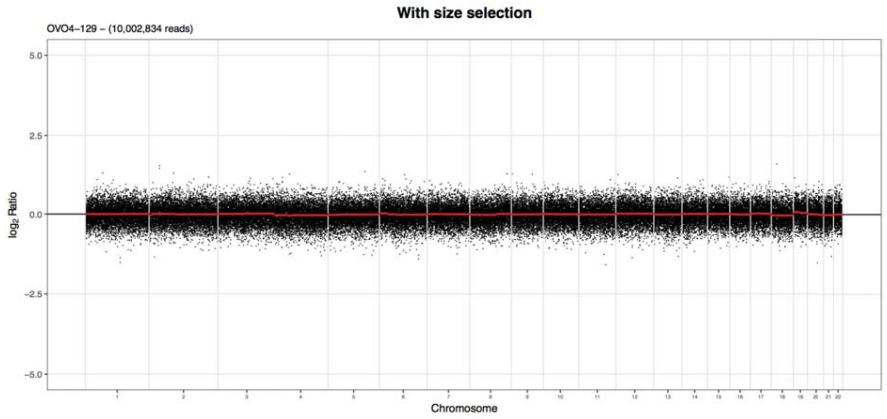
ci

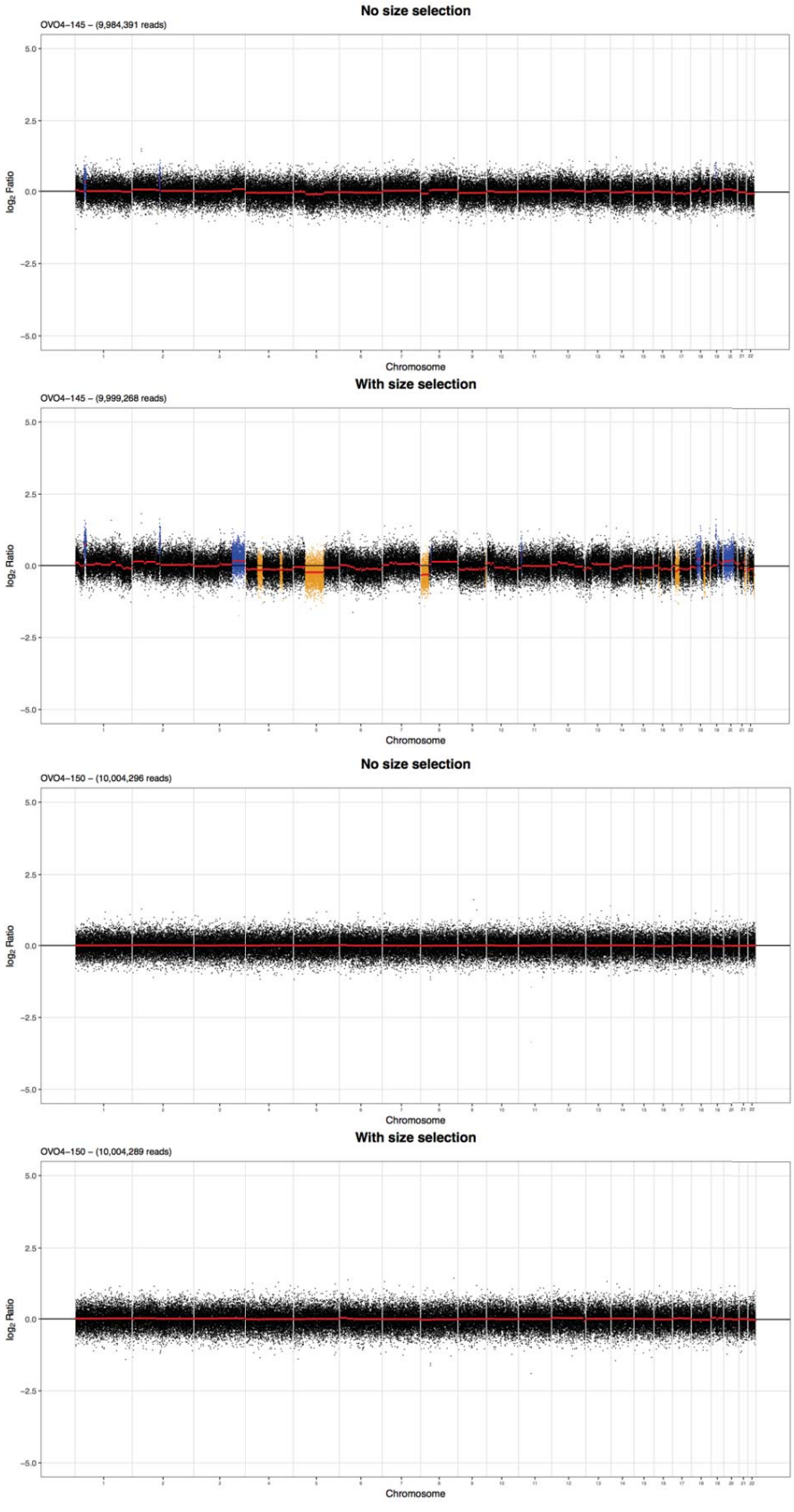

No size selection
OVO4–90 – (9,997,880 reads)
With size selection
OVO4–90 – (9,984,540 reads)
No size selection
OVO4–112 – (10,004,255 reads)
With size selection
OVO4–112 – (10,004,583 reads)
log2 Ratio
Chromosome

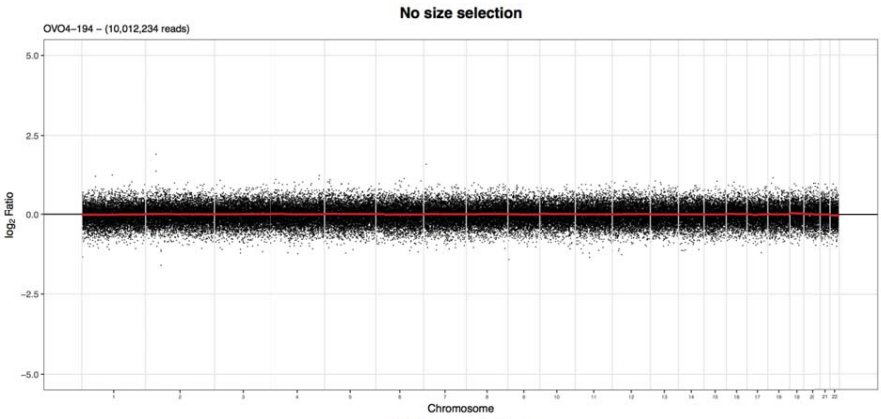

No size selection
OVO4–127 – (9,989,470 reads)
5.0
2.5
0.0
−2.5
−5.0
log2 Ratio
Chromosome

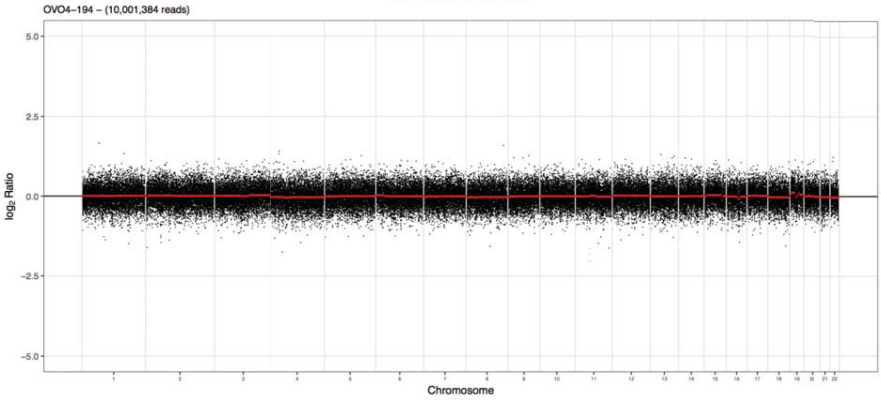

With size selection
OVO4–127 – (9,986,679 reads)
5.0
2.5
0.0
−2.5
−5.0
log2 Ratio
Chromosome

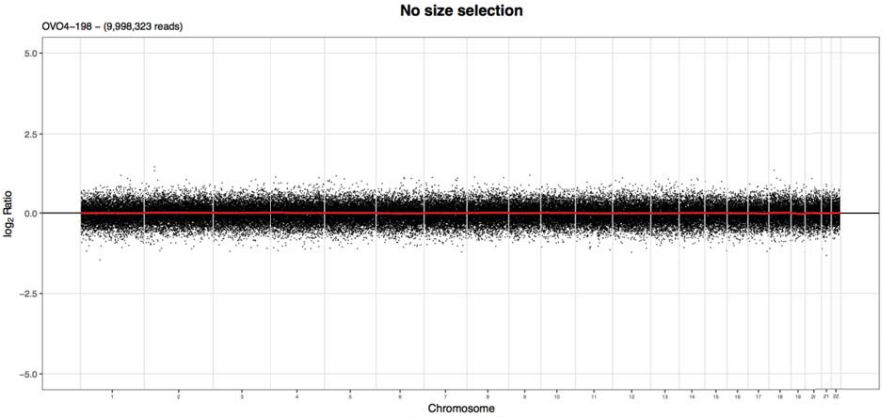

No size selection
OVO4–129 – (9,994,137 reads)
5.0
2.5
0.0
−2.5
−5.0
log2 Ratio
Chromosome

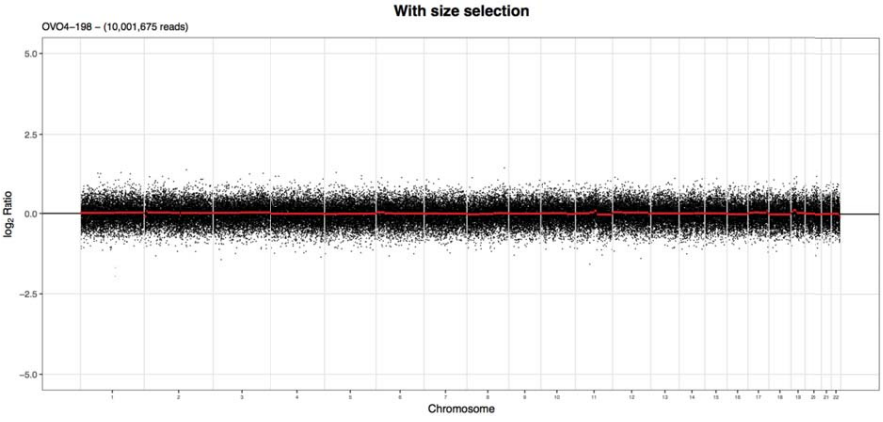

With size selection
OVO4–129 – (10,002,834 reads)
5.0
2.5
0.0
−2.5
−5.0
log2 Ratio
Chromosome

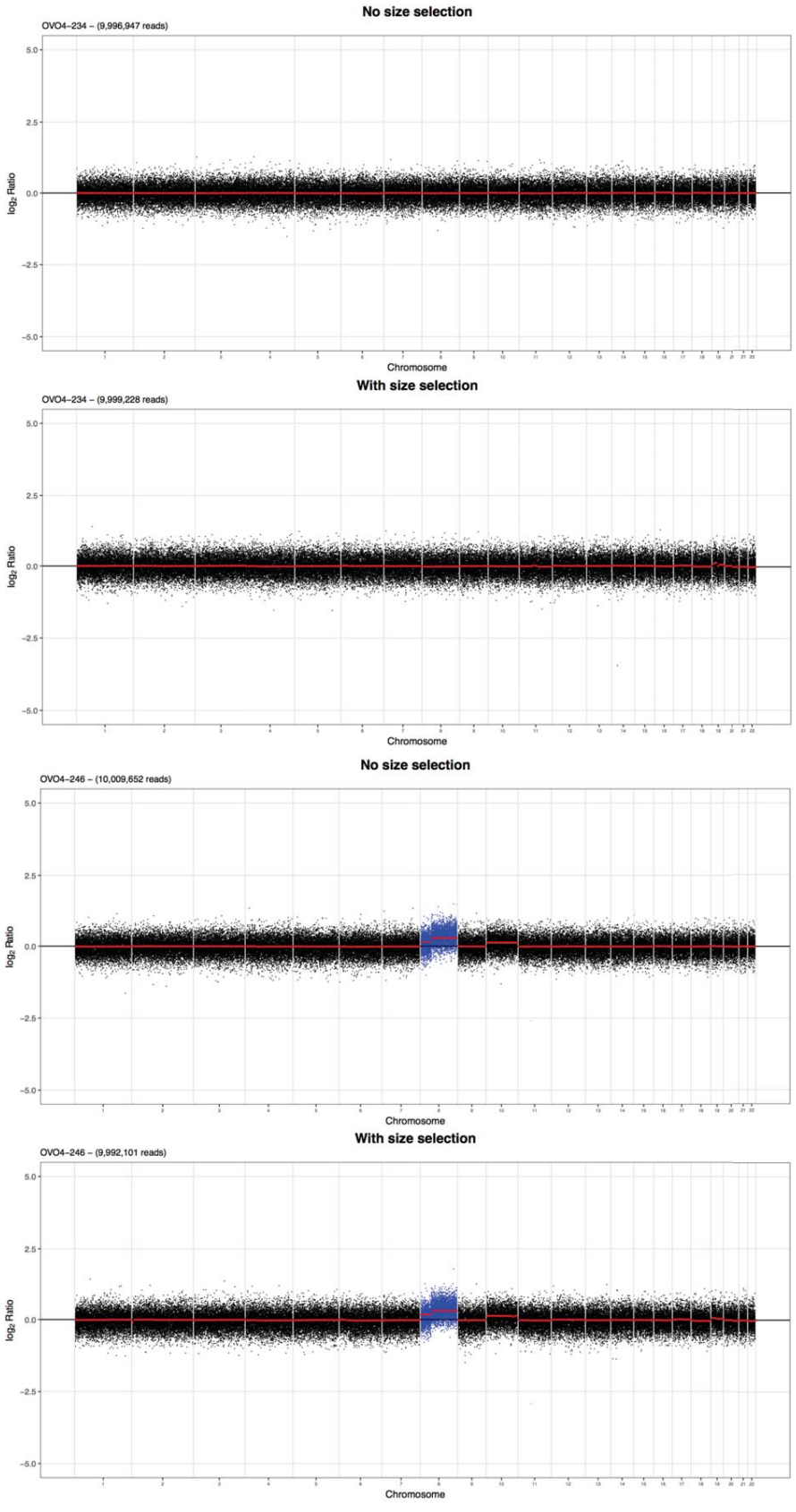

No size selection
OVO4–145 – (9,984,391 reads)
With size selection
OVO4–145 – (9,999,268 reads)
No size selection
OVO4–150 – (10,004,296 reads)
With size selection
OVO4–150 – (10,004,289 reads)
log2 Ratio
Chromosome

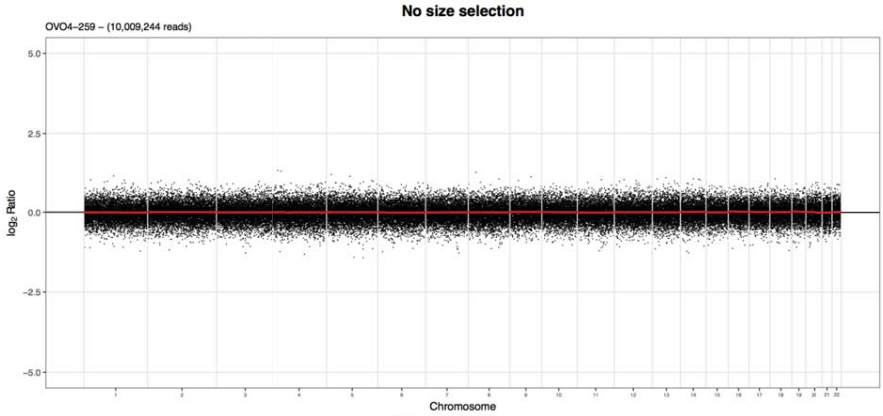

No size selection
OVO4–194 – (10,012,234 reads)
log2 Ratio
5.0
2.5
0.0
−2.5
−5.0
Chromosome

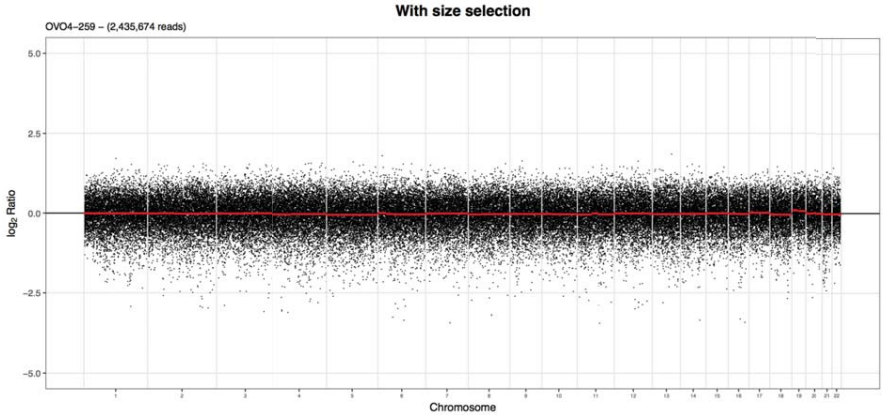

With size selection
OVO4–194 – (10,001,384 reads)
log2 Ratio
5.0
2.5
0.0
−2.5
−5.0
Chromosome

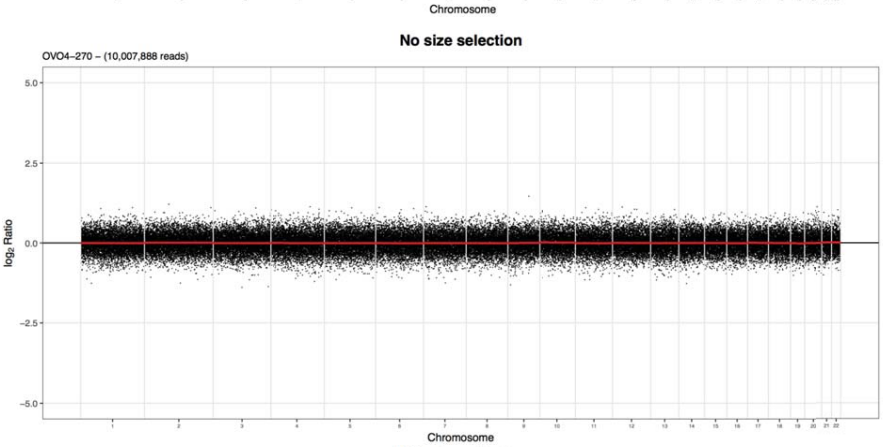

No size selection
OVO4–198 – (9,998,323 reads)
log2 Ratio
5.0
2.5
0.0
−2.5
−5.0
Chromosome

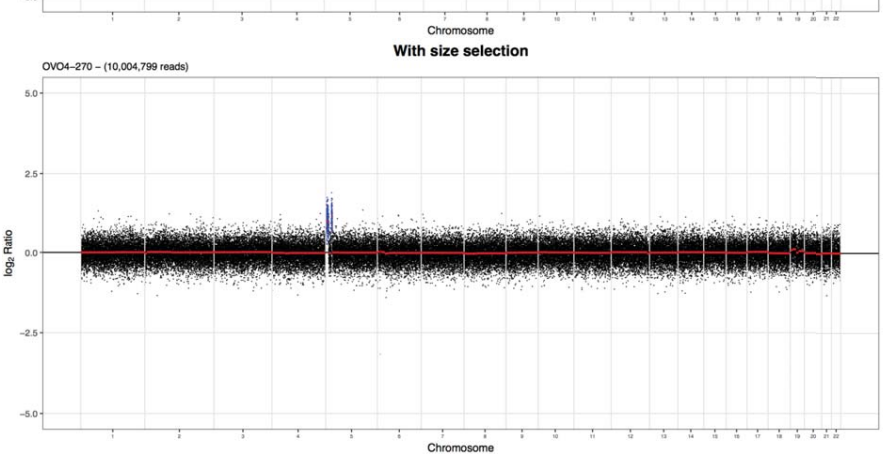

With size selection
OVO4–198 – (10,001,675 reads)
log2 Ratio
5.0
2.5
0.0
−2.5
−5.0
Chromosome


No size selection
OVO4–234 – (9,996,947 reads)
log2 Ratio
Chromosome
With size selection
OVO4–234 – (9,999,228 reads)
log2 Ratio
Chromosome
No size selection
OVO4–246 – (10,009,652 reads)
log2 Ratio
Chromosome
With size selection
OVO4–246 – (9,992,101 reads)
log2 Ratio
Chromosome


No size selection
OVO4–259 – (10,009,244 reads)
log2 Ratio
Chromosome


With size selection
OVO4–259 – (2,435,674 reads)
log2 Ratio
Chromosome


No size selection
OVO4–270 – (10,007,888 reads)
log2 Ratio
Chromosome


With size selection
OVO4–270 – (10,004,799 reads)
log2 Ratio
Chromosome

12.15 TAm-Seq V2 optimisations

12.15.1 Enzyme

Q5 HS (NEB M0493), Ultra II (NEB M0544), FastStart High Fidelity Enzyme Blend (Roche). Master mix was prepared for each primer pool as outlined below to give 10 µl total volume per reaction and PCR performed in 96 well plates.

Master mix for Q5 HS per reaction		
Q5 HS	0.1 µl	Final conc 0.02 U/µl
dNTPs	0.2 µl	Final conc 200 µM
F and R primer mix (2 µM)	2.5 µl	Final conc 0.5 µM
5× Q5 buffer	2.0 µl	Final conc 1x
DNA	1.0 µl	50 copies
H_2O	4.2 µl	

Master mix for Ultra II per reaction		
Ultra II master mix	5.0 µl	Final conc 1× Unit/U/µl
F and R primer mix (2 µM)	2.5 µl	Final conc 0.5 µM
DNA	1.0 µl	50 copies
H_2O	1.5 µl	

Master mix for Roche per reaction	
10× FastStart High Fidelity Reaction Buffer without $MgCl_2$ (Roche)	0.50 µl
25 mM $MgCl_2$ (Roche)	0.90 µl
DMSO (Roche)	0.25 µl
10 mM Grade Nucleotide Mix (Roche)	0.10 µl
5 U/µl FastStart High Fidelity Enzyme Blend (Roche)	0.05 µl
20× Access Array Loading Reagent (Fluidigm)	0.25 µl
DNA (50 copies/µl)	1.00 µl
H_2O	1.45 µl

12.15.2 Annealing temperature and extension time

Based on previous protocols in the lab and manufacturer recommendations different PCR cycling parameters were tested. For the Q5 HS enzyme 7 different PCR protocols were performed as outlined below.

98 °C	30 sec	
98 °C 63 °C	10 sec 4 min	30 cycles

98 °C	30 sec	
98 °C 63/65/67 °C 72 °C	10 sec 15 sec 5/10 sec	30 cycles
72 °C	2 min	

For the Ultra II enzyme three different PCR protocols, including a touchdown protocol, were performed as outlined below.

98 °C	30 sec	
98 °C	10 sec	
63/65 °C	30 sec	30 cycles
65 °C	45 sec	
65 °C	5 min	

98 °C	30 sec	
98 °C	10 sec	
68 °C	30 sec	4 cycles
65 °C	45 sec	
98 °C	10 sec	
67 °C	30 sec	4 cycles
65 °C	45 sec	
98 °C	10 sec	
66 °C	30 sec	4 cycles
65 °C	45 sec	
98 °C	10 sec	
65 °C	30 sec	30 cycles
65 °C	45 sec	
65 °C	5 min	

For the Roche enzyme one PCR protocol was performed as outlined below.

Protocol for Roche		
50 °C	2 min	
70 °C	20 min	
95 °C	10 min	
95 °C	15 sec	
60 °C	30 sec	10 cycles
72 °C	60 sec	
95 °C	15 sec	
80 °C	30 sec	2 cycles
60 °C	30 sec	
72 °C	60 sec	
95 °C	15 sec	
60 °C	30 sec	8 cycles
72 °C	60 sec	
95 °C	15 sec	
80 °C	30 sec	2 cycles
60 °C	30 sec	
72 °C	60 sec	
95 °C	15 sec	
60 °C	30 sec	8 cycles
72 °C	60 sec	
95 °C	15 sec	
80 °C	30 sec	5 cycles
60 °C	30 sec	
72 °C	60 sec	

12.15.3 Primer concentrations

Two different primer concentrations were investigated: 100 nM, 500 nM. Master mix outlined below.

Master mix for 500 nM primer concentration		
Ultra II master mix	2.50 µl	Final conc 1× U/µl
F and R primer mix (2 µM)	1.25 µl	Final conc 500 nM
DNA	1.00 µl	50 copies
H_2O	0.25 µl	

Master mix for 100 nM primer concentration		
Ultra II master mix	2.50 µl	Final conc 1× U/µl
F and R primer mix (2 µM)	0.25 µl	Final conc 100 nM
DNA	1.00 µl	50 copies
H_2O	1.00 µl	

12.15.4 Purification first round PCR product

PippinHT 3% agarose gel selecting for: 140–190 bp, 140–200 bp and 140–250 bp.

12.15.5 Barcoding

Three different barcoding reactions using the Fluidigm barcodes were investigated as outlined below.

Master mix for Roche barcoding per reaction	
10× FastStart High Fidelity Reaction Buffer without $MgCl_2$ (Roche)	1.00 µl
25 mM $MgCl_2$ (Roche)	1.80 µl
DMSO (Roche)	0.50 µl
10 mM PCR Grade Nucleotide Mix (Roche)	0.20 µl
5 U/µl FastStart High Fidelity Enzyme Blend (Roche)	0.10 µl
H_2O	3.40 µl
2 µM barcode	2.00 µl
DNA (direct from Pippin)	1.00 µl

Protocol for Roche barcoding		
95 °C	10 min	
95 °C	15 sec	
60 °C	30 sec	15 cycles
72 °C	1 min	
72 °C	3 min	

Master mix for Q5 HS barcoding per reaction	
Q5 HS	0.1 µl
dNTPs	0.2 µl
5× Q5 buffer	2.0 µl
H_2O	4.2 µl
2 µM barcode	2.5 µl
DNA (direct from Pippin)	1.0 µl

Protocol for Q5 HS barcoding		
98 °C	30 sec	
98 °C	10 sec	
67 °C	30 sec	15 cycles
72 °C	8 sec	
72 °C	2 min	

Master mix for Ultra II barcoding per reaction	
Ultra II master mix	5 µl
2 µM barcode	4 µl
DNA (direct from Pippin)	1 µl

Protocol for Ultra II barcoding		
98 °C	30 sec	
98 °C	10 sec	15 cycles
65 °C	75 sec	
65 °C	5 min	

12.15.6 Purification of final PCR product

AMPure bead purification 8 µl of PCR product + 12 µl H_2O. Sample:bead ratio 1:1.8. Eluted in 20 µl H_2O.

PippinHT 8 µl of PCR product + 12 µl H_2O. 3% agarose gel selecting for 190–250 bp.

12.15.7 Primer pooling

Coverage in some amplicons particularly 004 and 006 remained poor across all experimental conditions. Six different primer pooling strategies were therefore investigated (Figure 12.1).

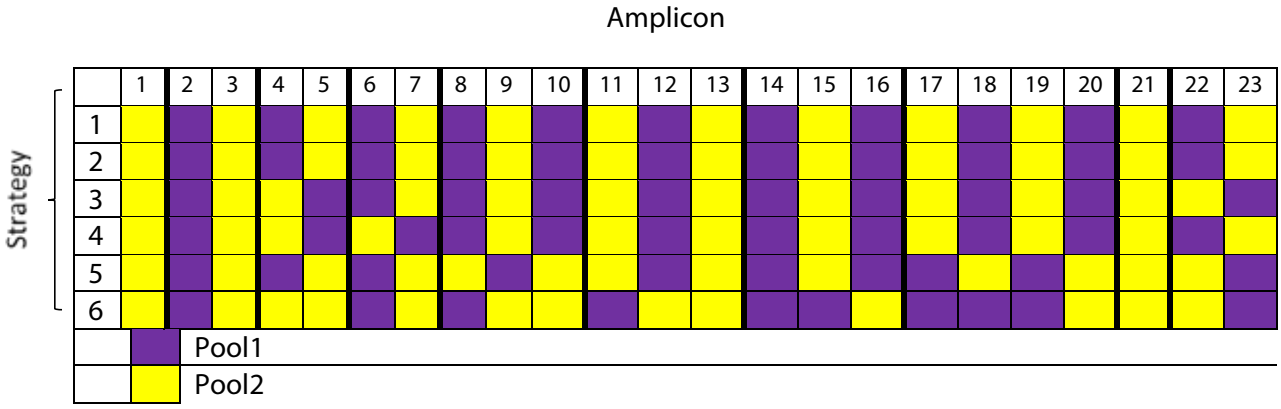


Figure 12.1: Primer pooling strategies. Darker lines indicate ends of exons. Strategy 1: Standard. Strategy 2: Standard strategy mixing pools in ratio of 4:1 pool1:pool2. Strategy 3: Swapping amplicon 4 with 5 and amplicon 22 with 23. Strategy 4: Swapping amplicon 4 with 5 and amplicon 6 with 7. Strategy 5: Manual improvements to in house pooling pipeline. Strategy 6: In house pooling pipeline with distance penalty < 500 bp

12.16 Mutation calls in cervical cytology samples

Sample	JBLAB ID	Cytology Grade	Mutation Analysis	Clinical Outcome
CRUK-CYT1	9201	6	EGFR mutation	cervical adenocarcinoma
CRUK-CYT2	9202	6	no mutation	CIN2 and high grade CGIN
CRUK-CYT3	9203	6	EGFR mutation	cervical adenocarcinoma
CRUK-CYT4	9204	6	EGFR mutation	high grade CGIN
CYT003c	9205	6	PIK3CA mutation	CIN1 and high grade CGIN
CYT004d	9206	6	no mutation	CIN3
CRUK-CYT5	9207	6	no mutation	CIN1 and high grade CGIN
CRUK-CYT6	9208	0	PIK3CA, TP53 mutation	high-grade serous endometrial carcinoma
CRUK-CYT7	9209	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT8	9210	0	KRAS mutation	Outcome know - non-cervical
CRUK-CYT9	9211	6	no mutation	CIN3 and high grade CGIN
CRUK-CYT10	9212	0	no mutation	normal endometrium (pipelle)
CRUK-CYT11	9213	6	no mutation	cervical adenocarcinoma
CRUK-CYT12	9214	6	no mutation	CGIN
CRUK-CYT13	9215	6	no mutation	CIN1 and high grade CGIN
CRUK-CYT14	9216	6	KRAS mutation	high grade CGIN
CRUK-CYT15	9217	6	no mutation	CIN1 and high grade CGIN
CRUK-CYT16	9218	6	no mutation	normal LLETZ
CRUK-CYT17	9219	6	PTEN mutation	normal LLETZ
CRUK-CYT18	9220	0	PTEN mutation	Endometrial adenocarcinoma
CRUK-CYT19	9221	6	no mutation	cervical adenocarcinoma
CRUK-CYT20	9222	0	KRAS mutation	Outcome know - non-cervical
CRUK-CYT21	9223	6	no mutation	CIN3 and high grade CGIN
CRUK-CYT22	9224	6	no mutation	CIN3 and high grade CGIN
CRUK-CYT23	9225	6	no mutation	CIN1 and CGIN
CRUK-CYT24	9226	0	no mutation	Outcome know - non-cervical
CRUK-CYT25	9227	6	no mutation	CIN1 and CGIN
CRUK-CYT26	9228	6	no mutation	high grade CGIN
CRUK-CYT27	9229	6	no mutation	CIN2
CRUK-CYT28	9230	6	EGFR and TP53 mutation	cervical adenocarcinoma and CGIN
CRUK-CYT29	9231	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT30	9232	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT31	9233	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT32	9234	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT33	9235	0	no mutation	normal endometrium (pipelle)
CRUK-CYT34	9236	6	no mutation	normal endoemtrium (pipelle)
CRUK-CYT35	9237	6	KRAS mutation	Stage 1B endometrial cancer
CRUK-CYT36	9238	6	no mutation	normal endometrium (pipelle)
CRUK-CYT37	9239	6	EGFR mutation	normal endoemtrium (pipelle)
CRUK-CYT38	9240	0	no mutation	normal endoemtrium (pipelle)
CRUK-CYT39	9244	0	KRAS mutation	Stage 1B endometrial cancer
CRUK-CYT40	9242	6	no mutation	normal endometrium (pipelle)
CRUK-CYT41	9241	0	no mutation	normal endoemtrium (pipelle)
CRUK-CYT42	9245	6	TP53 and PIK3CA mutation	Non-cervical cancer detected
CRUK-CYT43	9243	6	no mutation	Outcome known - none of the above
CRUK-CYT44	9246	6	no mutation	Endometroid carcinoma FIGO grade 1
CRUK-CYT45	9247	6	TP53 mutation	Outcome known - none of the above
CRUK-CYT46	9248	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT47	9249	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT48	9250	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT49	9251	0	PTEN mutation	normal endoemtrium (pipelle)
CRUK-CYT50	9252	6	TP53 mutation	Non-cervical cancer detected
CRUK-CYT51	9253	0	TP53 and PIK3CA mutation	Non-cervical cancer detected
CRUK-CYT52	9254	6	TP53 and PIK3CA mutation	Endometroid carcinoma FIGO grade 1
CRUK-CYT53	9255	6	no mutation	Outcome known - none of the above
CRUK-CYT64	9256	0	TP53 mutation	Outcome known - none of the above
CRUK-CYT65	9257	6	no mutation	CIN2
CRUK-CYT56	9258	6	no mutation	CIN3
CRUK-CYT58	9259	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT55	9260	6	TP53 mutation	cervical adenocarcinoma and CGIN
CRUK-CYT93	9261	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT67	9262	6	TP53 and PIK3CA mutation	cervical adenocarcinoma and CGIN
CRUK-CYT68	9263	6	TP53 mutation	cervical adenocarcinoma and CGIN
CRUK-CYT61	9264	0	no mutation	Endometroid carcinoma FIGO grade 1
CRUK-CYT66	9265	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT62	9266	6	PTEN mutation	cervical adenocarcinoma and CGIN
CRUK-CYT54	9267	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT59	9268	6	no mutation	cervical adenocarcinoma and CGIN
CRUK-CYT60	9269	6	TP53 mutation	cervical adenocarcinoma and CGIN
CRUK-CYT57	9270	6	no mutation	CIN3
CRUK-CYT69	9271	0	no mutation	Outcome known - none of the above

12.17 Specific variant calls for cervical samples using TAm-Seq

Patient	Sample type	Position	Reference	Mutation	MAF	Background error	Detected
612	cytology	7578375		DEL4	0.00000	NA	FALSE
612	CMA	7578375		DEL4	0.00000	NA	FALSE
622	cytology	7578433	G	C	0.00003	0.00011	FALSE
622	CMA	7578433	G	C	0.00007	0.00006	TRUE
791	cytology	7574003	G	A	0.00332	0.00265	TRUE
791	CMA	7574003	G	A	0.00255	0.00228	TRUE
617	cytology	7579378	G	T	0.00006	0.00007	FALSE
617	CMA	7579378	G	T	0.00000	0.00007	FALSE
810	cytology	7578407	G	C	0.00168	0.00070	TRUE
810	CMA	7578407	G	C	0.02669	0.00046	TRUE
822	cytology	7578259	A	C	0.00007	0.00033	FALSE
822	CMA	7578259	A	C	0.00000	0.00033	FALSE
831	cytology	7578265	A	G	0.00000	0.00307	FALSE
847	cytology	7577094	G	A	0.00681	0.00382	TRUE
853	cytology	7577529	A	T	0.00019	0.00016	TRUE
870	cytology	7578455		DEL9	0.00000	NA	FALSE
870	CMA	7578455		DEL9	0.00000	NA	FALSE
890	cytology	7578269	G	A	0.00122	0.00111	TRUE
897	cytology	7578394	T	C	0.00430	0.00376	TRUE
897	CMA	7578394	T	C	0.00338	0.00364	FALSE

12.18 Mutation calls using TAm-Seq V2 in cervical samples

Patient	Sample	Type	Histology	Chromosome	Position	Reference	Mutant	Wells	Matched FFPE Mutation
OV04 617	9605	smear	hgsoc	chr17	7574029	C	T	2	N
OV04 617	9605	smear	hgsoc	chr17	7577539	G	A	2	N
OV04 617	9605	smear	hgsoc	chr17	7578245	G	A	4	N
OV04 617	9605	smear	hgsoc	chr17	7578263	G	A	3	N
OV04 617	9605	smear	hgsoc	chr17	7578420	C	T	3	N
OV04 617	9605	smear	hgsoc	chr17	7578471	G	A	2	N
OV04 617	9605	smear	hgsoc	chr17	7578475	G	A	3	N
OV04 617	9605	smear	hgsoc	chr17	7579473	G	A	2	N
OV04 617	9605	smear	hgsoc	chr17	7579547	G	A	2	N
OV04 622	9613	smear	hgsoc	chr17	7574002	C	T	2	N
OV04 622	9613	smear	hgsoc	chr17	7574013	G	A	2	N
OV04 622	9613	smear	hgsoc	chr17	7574018	G	A	2	N
OV04 622	9613	smear	hgsoc	chr17	7577121	G	A	2	N
OV04 622	9613	smear	hgsoc	chr17	7578375	G	A	2	N
OV04 622	9613	smear	hgsoc	chr17	7578388	C	T	2	N
OV04 622	9613	smear	hgsoc	chr17	7578420	C	T	2	N
OV04 622	9613	smear	hgsoc	chr17	7579312	C	T	2	N
OV04 622	9613	smear	hgsoc	chr17	7579365	C	T	2	N
OV04 622	9613	smear	hgsoc	chr17	7579543	G	A	2	N
OV04 622	9613	smear	hgsoc	chr17	7579901	C	T	2	N
OV04 791	9623	smear	hgsoc	chr17	7572978	G	T	2	N
OV04 791	9623	smear	hgsoc	chr17	7574017	C	T	3	N
OV04 791	9623	smear	hgsoc	chr17	7578406	C	T	2	N
OV04 791	9623	smear	hgsoc	chr17	7579312	C	T	4	N
OV04 791	9623	smear	hgsoc	chr17	7579901	C	T	3	N
OV04 791	9623	smear	hgsoc	chr17	7579924	G	A	2	N
OV04 810	9635	smear	hgsoc	chr17	7577070	G	A	2	N
OV04 810	9635	smear	hgsoc	chr17	7577094	G	A	2	N
OV04 810	9635	smear	hgsoc	chr17	7577548	C	T	2	N
OV04 810	9635	smear	hgsoc	chr17	7578244	C	T	2	N
OV04 810	9635	smear	hgsoc	chr17	7578262	C	T	2	N
OV04 810	9635	smear	hgsoc	chr17	7578375	G	A	2	N
OV04 810	9635	smear	hgsoc	chr17	7578407	G	C	2	Y
OV04 810	9635	smear	hgsoc	chr17	7579885	C	T	2	N
OV04 810	9651	smear	hgsoc	chr17	7574002	C	T	2	N
OV04 810	9651	smear	hgsoc	chr17	7574018	G	A	3	N
OV04 810	9651	smear	hgsoc	chr17	7574023	C	T	3	N
OV04 810	9651	smear	hgsoc	chr17	7574023	C	T	2	N
OV04 810	9651	smear	hgsoc	chr17	7577094	G	A	4	Y
OV04 810	9651	smear	hgsoc	chr17	7577481	T	C	2	N
OV04 810	9651	smear	hgsoc	chr17	7577539	G	A	3	N
OV04 810	9651	smear	hgsoc	chr17	7577549	G	A	2	N
OV04 810	9651	smear	hgsoc	chr17	7578212	G	A	4	N
OV04 810	9651	smear	hgsoc	chr17	7578262	C	T	2	N
OV04 810	9651	smear	hgsoc	chr17	7578461	C	T	2	N
OV04 810	9651	smear	hgsoc	chr17	7578475	G	A	2	N
OV04 810	9651	smear	hgsoc	chr17	7579312	C	T	2	N
OV04 853	9652	smear	hgsoc	chr17	7576908	C	A	2	N
OV04 853	9652	smear	hgsoc	chr17	7577539	G	A	2	N
OV04 853	9652	smear	hgsoc	chr17	7577645	G	A	2	N
OV04 853	9652	smear	hgsoc	chr17	7578388	C	T	2	N
OV04 853	9652	smear	hgsoc	chr17	7578420	C	T	3	N
OV04 853	9652	smear	hgsoc	chr17	7578463	C	T	3	N
OV04 853	9652	smear	hgsoc	chr17	7578474	C	T	2	N
OV04 853	9652	smear	hgsoc	chr17	7579437	C	T	2	N
OV04 853	9652	smear	hgsoc	chr17	7579883	G	A	2	N
OV05 96	9685	smear	benign	chr17	7572974	G	A	2	N

Patient	Sample	Type	Histology	Chromosome	Position	Reference	Mutant	Wells	Matched FFPE Mutation
OV05 96	9685	smear	benign	chr17	7574002	C	T	4	N
OV05 96	9685	smear	benign	chr17	7574002	C	T	3	N
OV05 96	9685	smear	benign	chr17	7574003	G	A	3	N
OV05 96	9685	smear	benign	chr17	7574012	C	T	4	N
OV05 96	9685	smear	benign	chr17	7574013	G	A	4	N
OV05 96	9685	smear	benign	chr17	7574017	C	T	2	N
OV05 96	9685	smear	benign	chr17	7574023	C	T	6	N
OV05 96	9685	smear	benign	chr17	7574024	G	A	4	N
OV05 96	9685	smear	benign	chr17	7574024	G	A	3	N
OV05 96	9685	smear	benign	chr17	7574029	C	T	5	N
OV05 96	9685	smear	benign	chr17	7574030	G	A	2	N
OV05 96	9685	smear	benign	chr17	7577094	G	A	2	N
OV05 96	9685	smear	benign	chr17	7577094	G	A	3	N
OV05 96	9685	smear	benign	chr17	7577121	G	A	2	N
OV05 96	9685	smear	benign	chr17	7577139	G	A	2	N
OV05 96	9685	smear	benign	chr17	7577538	C	T	3	N
OV05 96	9685	smear	benign	chr17	7577539	G	A	2	N
OV05 96	9685	smear	benign	chr17	7577548	C	T	4	N
OV05 96	9685	smear	benign	chr17	7577549	G	A	2	N
OV05 96	9685	smear	benign	chr17	7577645	G	A	4	N
OV05 96	9685	smear	benign	chr17	7578184	G	A	3	N
OV05 96	9685	smear	benign	chr17	7578211	C	T	2	N
OV05 96	9685	smear	benign	chr17	7578212	G	A	5	N
OV05 96	9685	smear	benign	chr17	7578244	C	T	3	N
OV05 96	9685	smear	benign	chr17	7578245	G	A	3	N
OV05 96	9685	smear	benign	chr17	7578263	G	A	2	N
OV05 96	9685	smear	benign	chr17	7578374	C	T	2	N
OV05 96	9685	smear	benign	chr17	7578388	C	T	3	N
OV05 96	9685	smear	benign	chr17	7578389	G	A	3	N
OV05 96	9685	smear	benign	chr17	7578406	C	T	3	N
OV05 96	9685	smear	benign	chr17	7578421	G	A	2	N
OV05 96	9685	smear	benign	chr17	7578457	C	T	4	N
OV05 96	9685	smear	benign	chr17	7578458	G	A	8	N
OV05 96	9685	smear	benign	chr17	7578461	C	T	3	N
OV05 96	9685	smear	benign	chr17	7578462	G	A	3	N
OV05 96	9685	smear	benign	chr17	7578463	C	T	4	N
OV05 96	9685	smear	benign	chr17	7578464	G	A	2	N
OV05 96	9685	smear	benign	chr17	7578470	C	T	2	N
OV05 96	9685	smear	benign	chr17	7578471	G	A	10	N
OV05 96	9685	smear	benign	chr17	7578474	C	T	5	N
OV05 96	9685	smear	benign	chr17	7578475	G	A	6	N
OV05 96	9685	smear	benign	chr17	7579312	C	T	2	N
OV05 96	9685	smear	benign	chr17	7579313	G	A	2	N
OV05 96	9685	smear	benign	chr17	7579358	C	T	8	N
OV05 96	9685	smear	benign	chr17	7579359	G	A	2	N
OV05 96	9685	smear	benign	chr17	7579366	G	A	3	N
OV05 96	9685	smear	benign	chr17	7579470	C	T	2	N
OV05 96	9685	smear	benign	chr17	7579542	C	T	2	N
OV05 96	9685	smear	benign	chr17	7579547	G	A	4	N
OV05 96	9685	smear	benign	chr17	7579552	C	T	2	N
OV05 96	9685	smear	benign	chr17	7579580	G	A	2	N
OV05 96	9685	smear	benign	chr17	7579706	G	A	3	N
OV05 96	9685	smear	benign	chr17	7579883	G	A	7	N
OV05 96	9685	smear	benign	chr17	7579885	C	T	2	N
OV05 96	9685	smear	benign	chr17	7579886	G	A	2	N
OV05 96	9685	smear	benign	chr17	7579901	C	T	4	N
OV05 96	9685	smear	benign	chr17	7579902	G	A	3	N

Patient	Sample	Type	Histology	Chromosome	Position	Reference	Mutant	Wells	Matched FFPE Mutation
OV05 96	9685	smear	benign	chr17	7579924	G	A	4	N
OV05 102	9688	smear	benign	chr17	7572973	C	T	3	N
OV05 102	9688	smear	benign	chr17	7574002	C	T	7	N
OV05 102	9688	smear	benign	chr17	7574003	G	A	4	N
OV05 102	9688	smear	benign	chr17	7574012	C	T	6	N
OV05 102	9688	smear	benign	chr17	7574013	G	A	6	N
OV05 102	9688	smear	benign	chr17	7574018	G	A	2	N
OV05 102	9688	smear	benign	chr17	7574024	G	A	7	N
OV05 102	9688	smear	benign	chr17	7574029	C	T	3	N
OV05 102	9688	smear	benign	chr17	7574030	G	A	8	N
OV05 102	9688	smear	benign	chr17	7577022	G	A	2	N
OV05 102	9688	smear	benign	chr17	7577070	G	A	2	N
OV05 102	9688	smear	benign	chr17	7577090	C	T	2	N
OV05 102	9688	smear	benign	chr17	7577139	G	A	4	N
OV05 102	9688	smear	benign	chr17	7577144	A	G	2	N
OV05 102	9688	smear	benign	chr17	7577484	C	A	2	N
OV05 102	9688	smear	benign	chr17	7577538	C	T	5	N
OV05 102	9688	smear	benign	chr17	7577539	G	A	4	N
OV05 102	9688	smear	benign	chr17	7577548	C	T	8	N
OV05 102	9688	smear	benign	chr17	7577549	G	A	2	N
OV05 102	9688	smear	benign	chr17	7577645	G	A	2	N
OV05 102	9688	smear	benign	chr17	7578184	G	A	5	N
OV05 102	9688	smear	benign	chr17	7578211	C	T	5	N
OV05 102	9688	smear	benign	chr17	7578212	G	A	5	N
OV05 102	9688	smear	benign	chr17	7578239	C	A	2	N
OV05 102	9688	smear	benign	chr17	7578244	C	T	6	N
OV05 102	9688	smear	benign	chr17	7578245	G	A	3	N
OV05 102	9688	smear	benign	chr17	7578262	C	T	3	N
OV05 102	9688	smear	benign	chr17	7578263	G	A	5	N
OV05 102	9688	smear	benign	chr17	7578374	C	T	6	N
OV05 102	9688	smear	benign	chr17	7578388	C	T	2	N
OV05 102	9688	smear	benign	chr17	7578389	G	A	2	N
OV05 102	9688	smear	benign	chr17	7578406	C	T	2	N
OV05 102	9688	smear	benign	chr17	7578420	C	T	3	N
OV05 102	9688	smear	benign	chr17	7578421	G	A	4	N
OV05 102	9688	smear	benign	chr17	7578455	C	T	2	N
OV05 102	9688	smear	benign	chr17	7578461	C	T	6	N
OV05 102	9688	smear	benign	chr17	7578462	G	A	4	N
OV05 102	9688	smear	benign	chr17	7578463	C	T	2	N
OV05 102	9688	smear	benign	chr17	7578464	G	A	3	N
OV05 102	9688	smear	benign	chr17	7578470	C	T	3	N
OV05 102	9688	smear	benign	chr17	7578471	G	A	4	N
OV05 102	9688	smear	benign	chr17	7578474	C	T	3	N
OV05 102	9688	smear	benign	chr17	7578475	G	A	4	N
OV05 102	9688	smear	benign	chr17	7579312	C	T	4	N
OV05 102	9688	smear	benign	chr17	7579313	G	A	3	N
OV05 102	9688	smear	benign	chr17	7579358	C	T	4	N
OV05 102	9688	smear	benign	chr17	7579365	C	T	2	N
OV05 102	9688	smear	benign	chr17	7579366	G	A	2	N
OV05 102	9688	smear	benign	chr17	7579423	G	A	2	N
OV05 102	9688	smear	benign	chr17	7579438	C	T	3	N
OV05 102	9688	smear	benign	chr17	7579438	C	T	4	N
OV05 102	9688	smear	benign	chr17	7579441	C	T	4	N
OV05 102	9688	smear	benign	chr17	7579441	C	T	3	N
OV05 102	9688	smear	benign	chr17	7579442	G	A	5	N
OV05 102	9688	smear	benign	chr17	7579543	G	A	2	N
OV05 102	9688	smear	benign	chr17	7579547	G	A	3	N

Patient	Sample	Type	Histology	Chromosome	Position	Reference	Mutant	Wells	Matched FFPE Mutation
OV05 102	9688	smear	benign	chr17	7579883	G	A	7	N
OV05 102	9688	smear	benign	chr17	7579885	C	T	4	N
OV05 102	9688	smear	benign	chr17	7579886	G	A	5	N
OV05 102	9688	smear	benign	chr17	7579902	G	A	3	N
OV05 102	9688	smear	benign	chr17	7579923	C	T	4	N
OV04 897	9691	smear	hgsoc	chr17	7572973	C	T	3	N
OV04 897	9691	smear	hgsoc	chr17	7572973	C	T	2	N
OV04 897	9691	smear	hgsoc	chr17	7572974	G	A	5	N
OV04 897	9691	smear	hgsoc	chr17	7574002	C	T	9	N
OV04 897	9691	smear	hgsoc	chr17	7574003	G	A	6	N
OV04 897	9691	smear	hgsoc	chr17	7574012	C	T	3	N
OV04 897	9691	smear	hgsoc	chr17	7574013	G	A	7	N
OV04 897	9691	smear	hgsoc	chr17	7574017	C	T	3	N
OV04 897	9691	smear	hgsoc	chr17	7574023	C	T	6	N
OV04 897	9691	smear	hgsoc	chr17	7574023	C	T	5	N
OV04 897	9691	smear	hgsoc	chr17	7574029	C	T	6	N
OV04 897	9691	smear	hgsoc	chr17	7574030	G	A	4	N
OV04 897	9691	smear	hgsoc	chr17	7574030	G	A	5	N
OV04 897	9691	smear	hgsoc	chr17	7577047	G	A	2	N
OV04 897	9691	smear	hgsoc	chr17	7577070	G	A	2	N
OV04 897	9691	smear	hgsoc	chr17	7577091	G	A	3	N
OV04 897	9691	smear	hgsoc	chr17	7577093	C	T	5	N
OV04 897	9691	smear	hgsoc	chr17	7577093	C	T	4	N
OV04 897	9691	smear	hgsoc	chr17	7577094	G	A	4	N
OV04 897	9691	smear	hgsoc	chr17	7577094	G	A	5	N
OV04 897	9691	smear	hgsoc	chr17	7577121	G	A	3	N
OV04 897	9691	smear	hgsoc	chr17	7577138	C	T	2	N
OV04 897	9691	smear	hgsoc	chr17	7577139	G	A	8	N
OV04 897	9691	smear	hgsoc	chr17	7577538	C	T	3	N
OV04 897	9691	smear	hgsoc	chr17	7577539	G	A	2	N
OV04 897	9691	smear	hgsoc	chr17	7577548	C	T	3	N
OV04 897	9691	smear	hgsoc	chr17	7577645	G	A	4	N
OV04 897	9691	smear	hgsoc	chr17	7578184	G	A	2	N
OV04 897	9691	smear	hgsoc	chr17	7578211	C	T	3	N
OV04 897	9691	smear	hgsoc	chr17	7578212	G	A	6	N
OV04 897	9691	smear	hgsoc	chr17	7578244	C	T	4	N
OV04 897	9691	smear	hgsoc	chr17	7578245	G	A	5	N
OV04 897	9691	smear	hgsoc	chr17	7578245	G	A	4	N
OV04 897	9691	smear	hgsoc	chr17	7578263	G	A	6	N
OV04 897	9691	smear	hgsoc	chr17	7578374	C	T	4	N
OV04 897	9691	smear	hgsoc	chr17	7578375	G	A	4	N
OV04 897	9691	smear	hgsoc	chr17	7578389	G	A	5	N
OV04 897	9691	smear	hgsoc	chr17	7578406	C	T	2	N
OV04 897	9691	smear	hgsoc	chr17	7578407	G	A	7	N
OV04 897	9691	smear	hgsoc	chr17	7578420	C	T	3	N
OV04 897	9691	smear	hgsoc	chr17	7578421	G	A	2	N
OV04 897	9691	smear	hgsoc	chr17	7578455	C	T	3	N
OV04 897	9691	smear	hgsoc	chr17	7578458	G	A	3	N
OV04 897	9691	smear	hgsoc	chr17	7578461	C	T	2	N
OV04 897	9691	smear	hgsoc	chr17	7578462	G	A	4	N
OV04 897	9691	smear	hgsoc	chr17	7578463	C	T	2	N
OV04 897	9691	smear	hgsoc	chr17	7578470	C	T	4	N
OV04 897	9691	smear	hgsoc	chr17	7578474	C	T	3	N
OV04 897	9691	smear	hgsoc	chr17	7578475	G	A	6	N
OV04 897	9691	smear	hgsoc	chr17	7578516	G	A	2	N
OV04 897	9691	smear	hgsoc	chr17	7579312	C	T	4	N
OV04 897	9691	smear	hgsoc	chr17	7579313	G	A	6	N

Patient	Sample	Type	Histology	Chromosome	Position	Reference	Mutant	Wells	Matched FFPE Mutation
OV04 897	9691	smear	hgsoc	chr17	7579358	C	T	2	N
OV04 897	9691	smear	hgsoc	chr17	7579358	C	T	3	N
OV04 897	9691	smear	hgsoc	chr17	7579359	G	A	3	N
OV04 897	9691	smear	hgsoc	chr17	7579365	C	T	2	N
OV04 897	9691	smear	hgsoc	chr17	7579366	G	A	3	N
OV04 897	9691	smear	hgsoc	chr17	7579438	C	T	5	N
OV04 897	9691	smear	hgsoc	chr17	7579441	C	T	5	N
OV04 897	9691	smear	hgsoc	chr17	7579441	C	T	6	N
OV04 897	9691	smear	hgsoc	chr17	7579442	G	A	2	N
OV04 897	9691	smear	hgsoc	chr17	7579470	C	T	5	N
OV04 897	9691	smear	hgsoc	chr17	7579470	C	T	3	N
OV04 897	9691	smear	hgsoc	chr17	7579473	G	A	3	N
OV04 897	9691	smear	hgsoc	chr17	7579473	G	A	4	N
OV04 897	9691	smear	hgsoc	chr17	7579542	C	T	5	N
OV04 897	9691	smear	hgsoc	chr17	7579543	G	A	4	N
OV04 897	9691	smear	hgsoc	chr17	7579547	G	A	2	N
OV04 897	9691	smear	hgsoc	chr17	7579580	G	A	3	N
OV04 897	9691	smear	hgsoc	chr17	7579706	G	A	4	N
OV04 897	9691	smear	hgsoc	chr17	7579885	C	T	5	N
OV04 897	9691	smear	hgsoc	chr17	7579886	G	A	3	N
OV04 897	9691	smear	hgsoc	chr17	7579901	C	T	5	N
OV04 897	9691	smear	hgsoc	chr17	7579902	G	A	3	N
OV04 897	9691	smear	hgsoc	chr17	7579923	C	T	4	N
OV04 897	9691	smear	hgsoc	chr17	7579924	G	A	5	N
OV05 108	9692	smear	benign	chr17	7572973	C	T	3	N
OV05 108	9692	smear	benign	chr17	7574002	C	T	9	N
OV05 108	9692	smear	benign	chr17	7574003	G	A	3	N
OV05 108	9692	smear	benign	chr17	7574012	C	T	6	N
OV05 108	9692	smear	benign	chr17	7574013	G	A	7	N
OV05 108	9692	smear	benign	chr17	7574017	C	T	6	N
OV05 108	9692	smear	benign	chr17	7574018	G	A	4	N
OV05 108	9692	smear	benign	chr17	7574023	C	T	4	N
OV05 108	9692	smear	benign	chr17	7574024	G	A	2	N
OV05 108	9692	smear	benign	chr17	7574029	C	T	3	N
OV05 108	9692	smear	benign	chr17	7574030	G	A	3	N
OV05 108	9692	smear	benign	chr17	7577070	G	A	3	N
OV05 108	9692	smear	benign	chr17	7577094	G	A	2	N
OV05 108	9692	smear	benign	chr17	7577139	G	A	3	N
OV05 108	9692	smear	benign	chr17	7577538	C	T	3	N
OV05 108	9692	smear	benign	chr17	7577539	G	A	5	N
OV05 108	9692	smear	benign	chr17	7577548	C	T	6	N
OV05 108	9692	smear	benign	chr17	7578184	G	A	4	N
OV05 108	9692	smear	benign	chr17	7578211	C	T	2	N
OV05 108	9692	smear	benign	chr17	7578244	C	T	5	N
OV05 108	9692	smear	benign	chr17	7578262	C	T	4	N
OV05 108	9692	smear	benign	chr17	7578263	G	A	2	N
OV05 108	9692	smear	benign	chr17	7578374	C	T	2	N
OV05 108	9692	smear	benign	chr17	7578388	C	T	3	N
OV05 108	9692	smear	benign	chr17	7578389	G	A	2	N
OV05 108	9692	smear	benign	chr17	7578406	C	T	2	N
OV05 108	9692	smear	benign	chr17	7578420	C	T	3	N
OV05 108	9692	smear	benign	chr17	7578421	G	A	2	N
OV05 108	9692	smear	benign	chr17	7578455	C	T	4	N
OV05 108	9692	smear	benign	chr17	7578455	C	T	3	N
OV05 108	9692	smear	benign	chr17	7578457	C	T	3	N
OV05 108	9692	smear	benign	chr17	7578458	G	A	3	N
OV05 108	9692	smear	benign	chr17	7578461	C	T	8	N

Patient	Sample	Type	Histology	Chromosome	Position	Reference	Mutant	Wells	Matched FFPE Mutation
OV05 108	9692	smear	benign	chr17	7578462	G	A	4	N
OV05 108	9692	smear	benign	chr17	7578463	C	T	5	N
OV05 108	9692	smear	benign	chr17	7578464	G	A	6	N
OV05 108	9692	smear	benign	chr17	7578470	C	T	5	N
OV05 108	9692	smear	benign	chr17	7578471	G	A	3	N
OV05 108	9692	smear	benign	chr17	7578474	C	T	4	N
OV05 108	9692	smear	benign	chr17	7578475	G	A	3	N
OV05 108	9692	smear	benign	chr17	7579312	C	T	4	N
OV05 108	9692	smear	benign	chr17	7579312	C	T	5	N
OV05 108	9692	smear	benign	chr17	7579358	C	T	5	N
OV05 108	9692	smear	benign	chr17	7579359	G	A	4	N
OV05 108	9692	smear	benign	chr17	7579365	C	T	2	N
OV05 108	9692	smear	benign	chr17	7579366	G	A	4	N
OV05 108	9692	smear	benign	chr17	7579423	G	A	2	N
OV05 108	9692	smear	benign	chr17	7579431	C	A	2	N
OV05 108	9692	smear	benign	chr17	7579441	C	T	6	N
OV05 108	9692	smear	benign	chr17	7579442	G	A	2	N
OV05 108	9692	smear	benign	chr17	7579470	C	T	2	N
OV05 108	9692	smear	benign	chr17	7579473	G	A	3	N
OV05 108	9692	smear	benign	chr17	7579542	C	T	2	N
OV05 108	9692	smear	benign	chr17	7579547	G	A	5	N
OV05 108	9692	smear	benign	chr17	7579883	G	A	5	N
OV05 108	9692	smear	benign	chr17	7579886	G	A	4	N
OV05 108	9692	smear	benign	chr17	7579901	C	T	5	N
OV05 108	9692	smear	benign	chr17	7579902	G	A	3	N
OV05 108	9692	smear	benign	chr17	7579923	C	T	6	N
OV05 108	9692	smear	benign	chr17	7579923	C	T	4	N
OV05 108	9692	smear	benign	chr17	7579924	G	A	4	N
OV05 12	9903	CMA	benign	chr17	7579423	G	A	2	N
OV05 26	9904	CMA	benign	chr17	7578184	G	A	2	N
OV05 26	9904	CMA	benign	chr17	7578262	C	T	2	N
OV05 26	9904	CMA	benign	chr17	7579423	G	A	3	N
OV05 26	9904	CMA	benign	chr17	7579470	C	T	2	N
OV05 26	9904	CMA	benign	chr17	7579924	G	A	2	N
OV04 612	9906	CMA	hgsoc	chr17	7578389	G	A	3	N
OV04 612	9906	CMA	hgsoc	chr17	7578475	G	A	2	N
OV04 791	9907	CMA	hgsoc	chr17	7578263	G	A	2	N
OV04 897	9923	CMA	hgsoc	chr17	7574012	C	T	2	N
OV04 897	9923	CMA	hgsoc	chr17	7576899	G	T	2	N
OV04 897	9923	CMA	hgsoc	chr17	7578211	C	T	2	N
OV04 897	9923	CMA	hgsoc	chr17	7578386	A	G	2	N
OV04 897	9923	CMA	hgsoc	chr17	7578420	C	T	2	N
OV04 897	9923	CMA	hgsoc	chr17	7578421	G	A	2	N
OV04 897	9923	CMA	hgsoc	chr17	7578463	C	T	2	N
OV04 897	9923	CMA	hgsoc	chr17	7579312	C	T	10	N
OV04 897	9923	CMA	hgsoc	chr17	7579313	G	A	2	N
OV04 897	9923	CMA	hgsoc	chr17	7579365	C	T	2	N
OV04 897	9923	CMA	hgsoc	chr17	7579366	G	A	2	N
OV04 897	9923	CMA	hgsoc	chr17	7579423	G	A	2	N
OV04 897	9923	CMA	hgsoc	chr17	7579543	G	A	2	N
OV04 897	9923	CMA	hgsoc	chr17	7579706	G	A	2	N
OV04 897	9923	CMA	hgsoc	chr17	7579883	G	A	2	N
OV04 897	9923	CMA	hgsoc	chr17	7579923	C	T	2	N
OV04 897	9923	CMA	hgsoc	chr17	7579924	G	A	2	N

12.19 p53 autoantibodies and ctDNA

OV04	Mutation	Protein effect	Mutation type	MAF	p53 autoantibody status
20	c.G646T	p.V216L	nonsynonymous	0.0333	Strong Positive
22	c.G824A	p.C275Y	nonsynonymous	0.0196	Negative
37	c.C298T	p.Q100X	stopgain	ND	Negative
47	c.T428A	p.V143E	nonsynonymous	ND	Negative
66	c.C742T	p.R248W	nonsynonymous	ND	Strong Positive
68	c.G396C	p.K132N	nonsynonymous	0.2373	Negative
81	c.T584A	p.I195N	nonsynonymous	0.0094	Strong Positive
83	c.C833T	p.P278L	nonsynonymous	0.2700	Negative
88	c.A774T	p.E258D	nonsynonymous	ND	Negative
98	c.681dupT	p.D228_C229delinsX	indel	0.3506	Negative
127	c.G818A	p.R273H	nonsynonymous	0.0081	Negative
135	c.583dupA	p.I195fs	indel	0.0635	Negative
139	c.C742T	p.R248W	nonsynonymous	ND	Positive
141	c.C722T	p.S241F	nonsynonymous	0.0300	Negative
168	c.T526A	p.C176S	nonsynonymous	ND	Negative
170	c.G524A	p.R175H	nonsynonymous	ND	Negative
174	c.C380T	p.S127F	nonsynonymous	0.0023	Negative
176	c.A745G	p.R249G	nonsynonymous	ND	Negative
189	c.G422A	p.C141Y	nonsynonymous	0.0313	Negative
190	c.139_148del	p.47_50del	indel	0.1073	Negative
191	c.G743A	p.R248Q	nonsynonymous	0.0028	Negative
193	c.G733A	p.G245S	nonsynonymous	ND	Negative
194	c.C843G	p.D281E	nonsynonymous	0.0001	Negative
197	c.C206G	p.A69G	nonsynonymous	0.0001	Negative
203	c.G469T	p.V157F	nonsynonymous	ND	Strong Positive
211	c.611delA	p.E204fs	indel	ND	Negative
212	c.T815G	p.V272G	nonsynonymous	0.0134	Negative
219	c.1013dupT	p.F338fs	indel	ND	Positive
220	c.G743A	p.R248Q	nonsynonymous	ND	Negative
223	c.T470G	p.V157G	nonsynonymous	0.0001	Negative
224	c.G743A	p.R248Q	nonsynonymous	0.0026	Negative
225	c.G427A	p.V143M	nonsynonymous	0.0180	Strong Positive
227	c.C215T	p.P72L	nonsynonymous	0.0600	Negative
229	c.G524A	p.R175H	nonsynonymous	ND	Strong Positive
235	c.del626-627GA	p.209del	indel	ND	Negative
238	c.G1044T	p.L348F	nonsynonymous	ND	Negative
242	c.G859T	p.E287X	stopgain	ND	Negative
244	c.A838T	p.R280X	stopgain	ND	Negative
247	c.G713A	p.C238Y	nonsynonymous	0.0016	Strong Positive
254	c.C376T	p.Y126T	splicing	ND	Negative
256	c.A659G	p.Y220C	nonsynonymous	0.1805	Strong Positive
259	c.A122G	p.D41G	nonsynonymous	ND	Negative
260	c.C832T	p.P278S	nonsynonymous	0.1600	Negative
261	c.T584C	p.I195T	nonsynonymous	0.0044	Positive
262	c.T584C	p.I195T	nonsynonymous	ND	Strong Positive
263	c.del723C	p.241del	indel	0.0076	Negative
267	c.G1015T	p.E339X	stopgain	0.0790	Negative

ND - Not detected

OV04	Mutation	Protein effect	Mutation type	MAF	p53 autoantibody status
270	c.del308-309AC	p.103del	indel	ND	Positive
271	c.G527T	p.C176F	nonsynonymous	0.0001	Positive
274	c.G673A	p.V225I	splicing	0.0025	Positive
275	c.A701G	p.Y234C	nonsynonymous	0.0032	Negative
276	c.C577T	p.H193Y	nonsynonymous	0.0018	Strong Positive
277	c.T581G	p.L194R	nonsynonymous	0.0408	Strong Positive
279	c.C380T	p.S127F	nonsynonymous	0.0048	Strong Positive
280	c.702-703insG	p.235del	indel	0.0290	Negative
281	C.189delT	p.64del	indel	ND	Negative
283	c.G818T	p.R273L	nonsynonymous	ND	Negative
284	c.383delC	p.128del	indel	0.2198	Negative
293	c.275-281delCCCTGT	p.86del	indel	0.0047	Negative
295	c.T613C	p.Y205H	nonsynonymous	0.0034	Negative
297	c.G661T	p.E221X	stopgain	0.0360	Positive
301	c.G713T	p.C238F	nonsynonymous	ND	Negative
303	c.G841C	p.D281H	nonsynonymous	ND	Positive
304	c.G892T	p.E298X	stopgain	ND	Negative
305	c.C376T	p.Y126T	splicing	ND	Strong Positive
307	c.G824T	p.C275F	nonsynonymous	ND	Strong Positive
309	c.C310T	p.Q104X	stopgain	0.0084	Positive
311	c.G524A	p.R175H	nonsynonymous	ND	Strong Positive
312	c.C742T	p.R248W	nonsynonymous	ND	Negative
314	c.A659G	p.Y220C	nonsynonymous	0.0207	Strong Positive
315	c.C1024T	p.R342X	stopgain	0.0029	Negative
316	c.A488G	p.Y163C	nonsynonymous	ND	Negative
323	c.585-590delCCGAGT	p.197del	indel	ND	Negative
326	c.G524A	p.R175H	nonsynonymous	ND	Negative
330	c.403delT	p.135del	indel	0.0026	Negative
331	c.T809G	p.F270C	nonsynonymous	ND	Negative
334	c.G711A	p.M237I	nonsynonymous	ND	Positive
337	c.C742T	p.R248W	nonsynonymous	ND	Positive
341	c.C430T	p.Q144X	stopgain	0.0017	Negative
346	c.G524A	p.R175H	nonsynonymous	0.0039	Negative
348	c.C637T	p.R213X	stopgain	0.0311	Negative
354	c.761delT	p.I254fs	indel	ND	Negative
356	c.G713T	p.C238F	nonsynonymous	0.0006	Strong Positive
358	c.1014delC	p.F338fs	indel	0.0284	Negative
359	c.565delG	p.189del	indel	0.1872	Negative
362	c.G517T	p.V173L	nonsynonymous	0.0065	Negative
363	c.A394G	p.K132E	nonsynonymous	0.0298	Strong Positive
367	c.C376T	p.Y126T	splicing	ND	Negative
373	c.G824A	p.C275Y	nonsynonymous	0.0013	Strong Positive
378	c.G1036T	p.E346X	stopgain	ND	Positive
379	c.T363G	p.S121S	synonymous	0.0012	Positive
381	c.G151T	p.E51X	stopgain	0.0001	Negative
382	c.569-570insG	p.190ins	indel	0.1700	Negative
383	c.G747T	p.R249S	nonsynonymous	ND	Negative
385	c.A659G	p.Y220C	nonsynonymous	0.0032	Strong Positive

ND - Not detected

12.20 Mutation calling using TAm-Seq for plasma samples around surgery

OV04	OV05	Surgery	FFPE mutation	Effect	Ref	Mutant	cDNA effect	Protein effect	Pre surgery	End of surgery	2hr post surgery	Day 1 post op	Day 2 post op	Day 3 post op
621	17	IDS	7578555	splicing	G	T			ND	ND	ND	ND	ND	no sample
790	32	IDS	7578499	nonsynonymous	T	G	c.A431C	p.Q144P	ND	ND	ND	ND	ND	no sample
776	41	IDS	7578461	nonsynonymous	C	A	c.G469T	p.V157F	ND	ND	ND	ND	ND	ND
612	6	Primary	7578375	indel			c.551_554delATAG	p.D184Afs	ND	ND	ND	ND	ND	no sample
622	18	Primary	7578433	stopgain	G	C	c.C497G	p.S166*	ND	ND	ND	ND	no sample	no sample
791	33	Primary	7574003	stopgain	G	A	c.C1024T	p.R342*	ND	ND	ND	ND	ND	no sample

ND - Not detected

12.21 Manuscript: Journal of Molecular Diagnostics

Effects of collection and processing procedures on plasma circulating cell-free DNA from cancer patients

Bente Risberg[1,3,7,*], Dana WY Tsui[1,2*,±,@], Heather Biggs[4,*], Andrea Ruiz-Valdepenas Martin de Almagro[1,2], Sarah-Jane Dawson[1,4,5,#], Charlotte Hodgkin[5], Linda Jones[4], Christine Parkinson[5], Anna Piskorz[1,2], Francesco Marass[1,2,^], Dineika Chandrananda[1,2], Elizabeth Moore[1,2], James Morris[1,2], Vincent Plagnol[6,§], Nitzan Rosenfeld[1,2], Carlos Caldas[1,2,4,5], James D. Brenton[1,2,5], Davina Gale[1,2,@]

[1] Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Robinson Way, University of Cambridge, Cambridge, CB2 0RE

[2] Cancer Research UK Major Centre, Cambridge, CB2 0RE, UK

[3] Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Norway

[4] The Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Cambridge, UK

[5] Department of Oncology, Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Cambridge, UK

[6] UCL Genetics Institute, University College London, London, UK

[7] Department of Pathology, Oslo University Hospital, Norway

[±] Current affiliation: Department of Pathology, Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, USA

[#] Current affiliation: Division of Research and Cancer Medicine, Peter MacCallum Cancer Centre, Melbourne, Victoria, 3000, Australia.

[^] Current affiliations: Department of Biosystems Science and Engineering ETH Zurich; 4058 Basel; Switzerland; SIB Swiss Institute of Bioinformatics 4058 Basel; Switzerland

[§] Current affiliation: Inivata Ltd, The Portway Building, Granta Park, Cambridge, CB21 6GS,

UK

* These authors contribute equally to this work

@ Correspondence to: Dana Tsui (Email: tsuiw@mskcc.org), or Davina Gale (Email: davina.gale@cruk.cam.ac.uk; Cancer Research UK Cambridge Institute, Robinson Way, Cambridge, CB2 0RE, UK; Telephone: +44 1223 769500; Fax: +44 1223 769890)

Number of text pages: 25

Number of Tables: 1; Supplemental Tables: 3

Number of Figures: 5; Supplemental Figures: 5

Short Running Title: Effects of processing on cell-free DNA

Funding

This work was supported by Cancer Research UK, University of Cambridge [grant numbers A15601 (JDB), A11906 (NR), A20240 (NR), A18072 (JDB]; National Institute for Health Research Cambridge Biomedical Research Centre and Cambridge Experimental Cancer Medicine Centre (JDB); Cambridge Experimental Cancer Medicine Centre (JDB); European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no 337905 (NR). CP was supported by NIHR and Academy of Medical Sciences, the Wellcome Trust, British Heart Foundation and Arthritis Research UK. S-J.D. was supported by an Australian National Breast Cancer Foundation and Victorian Cancer Agency Early Career Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest disclosure statement

NR, DG, JDB and VP are co-founders, share-holders, consultants and/or employees of Inivata Ltd., a cancer genomics company focused on ctDNA analysis. DT, FM, DG, NR, JB and CC are co-inventors or contributors on a patent describing methods for analysis of rare DNA fragments. NR has received research support from AstraZeneca. The other authors declare no potential conflicts of interest.

Abstract

Circulating tumour DNA (ctDNA) offers new opportunities for non-invasive cancer management. Detecting ctDNA in plasma is challenging since it constitutes only a minor fraction of the total cell-free DNA (cfDNA). Pre-analytical factors affect cfDNA levels contributed from leukocyte lysis, hence the ability to detect low-frequency mutant alleles. This study investigates the effects of (i) delay in processing (ii) storage temperatures, (iii) different blood collection tubes, (iv) centrifugation protocols and (v) sample shipment on cfDNA levels. Peripheral blood (n=231) from cancer patients (n=62) were collected into K_3EDTA or Cell-free DNA BCT® (BCT) tubes and analysed by digital PCR (dPCR), targeted amplicon or shallow whole-genome (sWGS) sequencing. To assess pre-analytic effects, plasma was processed under different conditions after 0h, 6h, 24h, 48h, 96h and 1 week at room temperature (RT) or 4°C, or using different centrifugation protocols. dPCR showed that cfDNA levels increased gradually with time in K_3EDTA tubes, but were stable in BCT tubes. K_3EDTA samples stored at 4°C showed less variation than RT storage, but levels were elevated compared to BCT. A second centrifugation at 3000g gave similar cfDNA yields compared to higher speed centrifugation. Next-generation sequencing showed negligible differences in background error or copy number changes between K_3EDTA and BCT, or following shipment in BCT. This study provides insights into the effects of sample processing on ctDNA analysis.

Introduction

Circulating tumour DNA (ctDNA) in plasma offers new opportunities for non-invasive cancer management. Recent studies have demonstrated its potential for molecular stratification, monitoring tumour response, identifying resistance mutations and patients at risk of relapse [1, 2] . Detecting ctDNA in plasma is challenging since it constitutes only a minor fraction of the total cell-free DNA (cfDNA), particularly in early stage cancers and in the minimal residual disease setting [3, 4]. A proportion of background wild-type DNA is believed to originate from lysis of white blood cells [5]. Previous studies have highlighted the pre-analytic effects of different processing and collection protocols on plasma ctDNA levels from cancer patients and pregnant women [6-9]. Based on these results, it is recommended to process whole blood samples for retrieval of plasma as soon as possible after collection, prior to *in vitro* cell lysis. At the same time, a double-centrifugation protocol has been recommended to obtain cell-free plasma, using an initial slow centrifugation speed to separate plasma, then fast centrifugation to clear cellular material [7]. However, some of these procedures may be difficult to carry out in a clinical setting due to lack of appropriate personnel or equipment. To circumvent this, cell-stabilising blood collection tubes have become available to stabilise cell-free DNA, enabling a delay in processing which may be carried out under more controlled conditions and within centralised laboratories. This study aimed to perform a systematic comparison of the effects of different processing protocols and collection tubes on the levels of cfDNA and ctDNA from cancer patients using digital PCR (dPCR). With the growing use of next-generation sequencing (NGS) for the analysis of ctDNA, we also aimed to investigate the effect of different protocols and collection tubes on the performance of targeted amplicon and shallow whole genome sequencing (sWGS) for quantification of plasma DNA.

Materials and Methods

Analysis Modules

The study was designed to include 5 different Modules: Module 1 investigated the effects of delayed processing on the levels of circulating DNA (cfDNA and ctDNA) in plasma collected in K_3EDTA tubes (9mL S-Monovette®, Sarstedt, Germany). The separation of plasma was delayed for different durations: 0h, 6h, 24h, 48h, 96h, and 1 week at room temperature (19-25°C, RT). Module 2 investigated the effects of storage temperature on the levels of circulating DNA in plasma collected in K_3EDTA tubes. Samples were stored at RT or at 4°C prior to processing at the following hours post-collection: 0h, 24h, 48h and 96h. Module 3 investigated the effects of collection devices on the levels of circulating DNA. Blood samples from each patient were collected at the same time point into K_3EDTA tubes and cell-stabilization blood collection tubes (10mL Cell-Free DNA BCT®, Streck, US) respectively. BCT's contain a proprietary formaldehyde-free preservative that stabilises nucleated blood cells preventing the release of genomic DNA [10, 11]. The processing was performed at the following times post-collection: 0h, 96h and 1 week. Module 4 investigated the effects of different centrifugation protocols on the levels of circulating DNA. Module 5 investigated the effects of shipment on samples collected in BCT tubes at ambient temperature. For Modules 1, 2, 3 and 5, plasma was separated from blood using a double-centrifugation protocol (Protocol A): a first centrifugation at 820g for 10min in a mega-centrifuge (Thermo Sorvall Legend RT, Fisher Scientific, US), then subjected to a second centrifugation step of the plasma supernatant at 14,000g for 10min in a bench top micro-centrifuge (Heraeus Fresco 21, Thermo Scientific, US). For Module 4, blood aliquots from the same patients were processed with three different protocols: Protocol A as above, Protocol B with the first centrifugation performed at 1600g and the second centrifugation at 14000g for 10 min in a bench top micro-centrifuge, and Protocol C with both first and second centrifugations performed in the same mega-centrifuge, initially at 1600g for 10min, then at 3000g for 10min.

Patient samples and DNA extractions

Peripheral whole blood was collected from 62 patients in total: 47 patients with high-grade serous ovarian cancer and 15 patients with metastatic breast cancer. Informed consent was obtained from each patient with protocols approved by an institutional ethics committee. 15-30mL blood from each patient was processed according to each analysis Module. DNA from all samples, except Module 5, was extracted from average 1.4 mL (range: 0.3 mL - 2.76 mL) plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Germany) according to manufacturer's protocol, except that 6.2 µg carrier RNA was added per sample. DNA was eluted twice through the column to maximise yield. A non-human spike-in PCR product was added to each sample as an internal quality control to assess extraction efficiency [12]. In Module 5, DNA was extracted from plasma on a QIAsymphony robot (Qiagen, Germany) using a 2mL extraction protocol. Eluted DNA was stored at -20 °C until analysis.

A total of 231 blood samples aliquots were analysed in this study. Table 1 summarises the number of plasma samples collected for each Module. Note that the collection was designed in such a way that each sample from every processing condition (temperature; collection tube; delayed processing duration) had a matched sample that was collected in K_3EDTA and processed immediately (denoted E.RT.0h) using centrifugation Protocol A, and was assigned as the reference sample for each condition. The levels of circulating DNA (either cfDNA or ctDNA), were expressed as a ratio of the respective data with the reference sample (E.RT.0h). Therefore, data collected under the same processing conditions could be grouped together to evaluate the effect of the processing even though they were collected from different patients. A more detailed summary of the distribution of samples involved in each Module is given in Supplemental Table S1.

Quantification of circulating plasma DNA by dPCR and targeted amplicon sequencing

Plasma samples from ovarian and breast cancer patients were first quantified by dPCR (using the Biomark microfluidic system (Fluidigm, US) as previously described [13], using an assay which targets a 65-bp amplicon in *RPP30*, a non-amplified region in the genome, to estimate cfDNA levels [12, 14]. ctDNA levels were then determined by dPCR using dual-labelled patient-specific TaqMan assays designed to mutant and wild-type sequences in *TP53* or *PIK3CA,* or deletions in chromosome 8, 11 or 17. A summary of the samples analysed is provided in Supplemental Table S1, and sequences of primers and fluorescent probes, amplicon sizes and amplification conditions used in dPCR are detailed in Supplemental Table S2.

The levels of cfDNA and ctDNA were calculated from the number of observed amplifications above a set threshold, and Poisson statistics were used to convert the number of observed amplifications to estimated targets, assuming independent segregation of DNA molecules into the microfluidic reaction chambers. The total number of amplifiable copies of DNA molecules per mL of plasma (copies/mL) were calculated, taking into account the relative fraction of the extracted DNA loaded and the proportion of sample lost during the loading process through the microfluidic channels. The levels of ctDNA were calculated as mutant allele fraction (i.e. the fraction of mutant DNA copies divided by the total cfDNA copies) expressed as a percentage or as mutant copies/mL plasma. For the purpose of comparing different protocols in the Modules, we expressed the data at each processing condition as a ratio from the E.RT.0h reference sample that was collected in K_3EDTA and immediately processed according to Protocol A, unless otherwise specified.

To investigate the effects of different collection devices and processing protocols on the performance of NGS, plasma samples from all Modules were analysed by TAm-Seq (Tagged Amplicon deep sequencing), as previously described [13]. TAm-Seq is a targeted amplicon sequencing method that allows identification and quantification of low frequency mutant alleles in plasma across sizable genomic regions. Sequencing was performed using an Illumina HiSeq 2500 to an average of greater than 1000x sequencing depth. Mutations were identified

and quantified as previously described [13]. To assess the effect of collection and processing procedures on the background error rates during NGS, we generated the allelic read ratio (reference/alternative) at each position within R (v3.1.2) from the BAM files, using the Bioconductor packages Rsamtools and Biostrings. All positions flagged as polymorphic by the 1000 Genomes Project or the COSMIC database, were filtered out.

To investigate the effects of shipping on global somatic copy number alterations (sCNAs), samples in Module 5 were also subjected to sWGS [15]. Briefly, a DNA library was prepared from 2-10ng of cfDNA from each sample using the ThruPLEX DNA-seq Kit (Rubicon Genomics, US) and sequenced on an Illumina HiSeq 4000 to 0.1x average depth using single-end sequencing. Sequence data was analysed using a pipeline that involved the following: single-end sequence reads were aligned to the human reference genome (GRCh37) using BWA-mem v0.7.17 [16] after removing any contaminant adapter sequences. SAMtools v1.7 was used to convert files to BAM format. PCR and optical duplicates were marked using Picard-Tools' 'MarkDuplicates' feature v2.17.6 and these were excluded from downstream analysis along with reads of low mapping quality and supplementary alignments. Reads in each sample were down-sampled to approximately 3 million reads in order to have similar coverage between patients and conditions. Subsequently, copy number analysis was performed in R [17] using the R package CNAclinic v1.0 (https://github.com/sdchandra/CNAclinic; manuscript under review), a software suite which allows for robust copy number analysis of sWGS data. Briefly, sequence reads were allocated into equally sized (100Mb) non-overlapping bins throughout the length of the genome. Read counts in each bin were corrected to account for sequence GC content and mappability, and regions corresponding to artefacts and probable germline changes were excluded from downstream analysis utilising a cohort of 45 healthy controls. After median normalization, binned counts were segmented using both the 'Circular Binary Segmentation' and Hidden-Markov Model based algorithms and an averaged log2R value per bin was calculated.

Statistical analysis

The difference in circulating DNA levels between different sub-groups in each Module was analysed using non-parametric Mann Whitney Rank Sum Test unless specified and a p-value of <0.05 was considered statistically significant. To assess the noise of sWGS data we calculated values corresponding to the `Median of the Absolute values of all Pairwise Differences' (MAPD) between log2R copy numbers. This metric provides a measure of the noise of the sample that is less dependent on true biological copy number variation and more on technical variation[18]. To compare the 3 collection methods in all patients, we calculated pairwise Spearman correlations between the binned copy number segments of the 3 collection methods. Furthermore, we applied a non-parametric Wilcoxon signed rank test on these values to test the similarity of the copy number profiles between all pairwise samples.

Results

Module 1: The effects of delayed processing on the levels of circulating DNA in plasma collected in EDTA tubes

In this Module, all samples (n=26) were collected in K_3EDTA tubes. One tube from each collection was processed immediately. The other tubes were stored at RT and processed at different prolonged time points: 6h, 24h, 48h, 96h and 1 week. Analysis by dPCR showed that the levels of cfDNA in the plasma samples increased gradually with increasing delay in the processing (Figure 1A), whereas the fraction of ctDNA decreased (Figure 1B). In particular, the levels of cfDNA increased significantly after 48h, 96h and 1 week of delay, whereas the mutant allele fraction of ctDNA decreased significantly after 96h and 1 week of delay (Mann Whitney Rank Sum Test, p-value<0.05). Previous reports have indicated that in analysis of circulating cell-free DNA from maternal plasma, despite changes in total cfDNA, the levels of fetal DNA are relatively stable in different storage and processing conditions [8, 19]. Indeed,

our results confirm that the numbers of mutant molecules, expressed as copies/mL of plasma, were relatively stable across the different processing time points with no statistically significant difference observed compared to samples that were processed immediately (Figure 1C, Supplemental Figure S1A and S1B)

Module 2: The effects of storage temperature on the levels of circulating DNA in plasma collected in K_3*EDTA tubes*

In this Module, all samples (n=26) were collected in K_3EDTA tubes, and either processed to plasma immediately or after 24h, 48h and 96h. The individual tubes were stored in two conditions: at RT (19-25°C), or at 4°C. If kept at RT, dPCR showed that the levels of cfDNA significantly increased after 48h. If kept at 4°C, the levels increased after 48h but were significantly lower than those observed at RT (Figure 2A). If delayed for 96h, samples kept at RT and 4°C all increased significantly. The changes in mutant allele fraction showed an inverted similar trend, although the amount of available data was too low for statistical analysis (Figure 2B).

Module 3: The effects of collection devices (K_3*EDTA versus Cell-free DNA BCT) on the levels of circulating DNA*

In this Module, one K_3EDTA tube for each collection was processed immediately (E.RT.0h) and served as a reference sample (n=20). The other K_3EDTA tubes were stored for 96h (n=5) and one week (n=5) at RT. Cell-free DNA BCT's were stored at RT and processed immediately (n=5) or delayed for 96h (n=10) and 1 week (n=15). The cfDNA levels increased significantly after 1 week if kept in K_3EDTA tubes, but remained at similar levels if kept in BCT (Figure 3A). The changes in the mutant allele fraction showed an inverted similar trend, but the amount of data available was too low for statistical analysis (Figure 3B). We compared the mutant allele fraction from 6 patients that were collected in K_3EDTA and processed immediately, versus the

matched samples that were collected in BCT and processed after 1 week's delay. The levels of ctDNA were similar for 4 patients but decreased two-fold for 2 patient samples (Supplemental Figure S2). There was no statistically significant difference in the numbers of mutant copies/mL plasma between storage in the two tube types (Supplemental Figure S1A and S1B).

We next assessed the effects of collection and processing procedures on the background error rates during NGS analysis using targeted amplicon sequencing. As previously described, different A/C/G/T base substitutions are associated with different error rates [13]. We plotted the distribution of the ratio of non-reference:reference alleles as box plots, shown according to mutation types. We did not observe a difference using different collection devices and processing conditions (Figure 3C).

Module 4: The effects of different centrifugation speed on the levels of circulating DNA

In this Module, all samples (n=13) were collected in K_3EDTA tubes and processed immediately. Aliquots from the same patients were processed using three different centrifugation protocols (A to C) as defined in Methods. There were no statistically significant differences across the three protocols on the total circulating DNA levels as measured by dPCR (Figure 4A and B), or in mutant allele fraction as measured by targeted amplicon sequencing (Figure 4C and D).

Module 5: The effects of shipment of Cell-free DNA BCT on mutant allele fraction and global copy number changes.

In this Module, three tubes of blood were drawn from each patient (n=13). K_3EDTA tubes were processed immediately (E.RT.0h), one Cell-free DNA BCT collected and stored at RT within the same centralised processing lab while the other BCT was packaged and shipped back to the same lab. All shipped samples, apart from three, were received and processed within 48h

from the time of collection. Of these, two BCT's were processed after 96h and one was processed after 5 days. The stored BCT's were processed at the same time as the matched shipped sample. There was no statistically significant difference in cfDNA levels between the three collection methods (Figure 5A and B). *TP53* mutations were identified by amplicon sequencing in 4 patients, and there were no statistically significant differences in mutant allele fraction using the different collection methods (Figure 5C and D).

To further investigate the effects of collection methods on global copy number changes, we performed sWGS analysis on 4 patients that had *TP53* mutations detected (P161, P227, P479, P488) and 4 without (P615, P489, P464, P450). Data from 1 patient (P464) was excluded from further analysis as the total read counts generated for one of the collection methods was below 1 million. This is below the threshold recommended for inference when analysing shallow coverage [20]. The segmental copy number profiles among the 3 collection methods were highly similar showing an average Spearman correlation of 0.76, range = 0.44 - 0.98 (Supplemental Figure S3, Supplemental Table S3). The paired Wilcoxon test p-values indicated no significant differences in all 21 copy number distributions comparisons (p-values > 0.001). Supplemental Figure S4 shows an example of the copy number alterations in plasma samples processed with and without shipping. The same gains and losses in chromosomal arms were identified in all three protocols. Supplemental Figure S5 depicts the estimation of noise in the sWGS data using MAPD values. All patients showed very similar noise levels between the different tubes and protocols.

Discussion

Multiple research studies have demonstrated the potential of using plasma as a tool for non-invasive cancer management. There is increasing interest in incorporating ctDNA as a 'liquid biopsy' in both clinical and research settings. As the frequency of mutant alleles in plasma may be low, particularly in early stage disease, it is crucial to optimise and standardise pre-analytic

sample processing procedures to maintain the quality of samples for accurate quantification of rare mutant molecules. In this study, we examined the pre-analytic effects of blood sample processing procedures, including the use of different blood collection tubes, storage conditions and centrifugation speeds, on downstream analysis of cfDNA using different molecular technologies including dPCR, targeted amplicon and genome-wide sequencing. Our results show that levels of cfDNA are stable in K_3EDTA tubes at room temperature for up to 24 hours. If delayed beyond 24 hours, storage of K_3EDTA blood at 4°C appeared to delay the increase in background cfDNA. It is worth noting that a recent study demonstrated that storing the samples in K_2EDTA tubes at 4°C kept the cfDNA levels stable for a course of 3 days [21]. This agrees with our observations that storing K_3EDTA tubes at 4°C improved the stability of cfDNA compared to room temperature storage. Alternatively, collection into Cell-free DNA BCT tubes at room temperature maintained stable cfDNA levels for at least a week. These tubes can facilitate delayed and centralised blood processing, circumventing issues arising with delayed plasma processing. Other researchers have evaluated alternative cell-stabilisation tubes such as CellSave (CellSearch) and PAXgene Blood ccfDNA tubes (Qiagen) and demonstrated similar stability when sample processing was delayed[9, 22]. New cell-free stabilization tubes have recently become available (eg. Cell-free DNA Collection tube, Roche; cf-DNA Preservation tube, Norgen; Blood STASIS 21-ccfDNA, Magbio and LBgard Blood tubes, Biomatrica) and it will be important to test these thoroughly to assess their performance for optimal sample processing procedures prior to next-generation sequencing and dPCR analysis of ctDNA.

Our findings have addressed a few of the practical challenges in the ‘blood to plasma’ sample processing workflow in a hospital setting. For example, in the clinic, processing may be delayed due to shortage of staff to enable immediate processing, or outside office-hours collection. In some scenarios, when conducting multi-centre clinical trials, many individual centres do not have access to the full spectrum of centrifuges with the higher second centrifugation speeds required to carry out the recommended double-centrifugation

procedures. The ability to delay processing by collecting into cell-stabilisation tubes, or the flexibility to perform the centrifugation in a range of different type of centrifuges or storing at 4°C after collection for a short period, will greatly improve the feasibility of collecting high-quality specimens. For samples collected across a wide geographical area, shipment may be necessary prior to central processing to standardise pre-analytic factors and maximize cost-effectiveness. Our study showed no statistically significant difference in NGS background noise with or without shipment. However, other studies have shown that the shipping temperature of Cell-free DNA BCT was deemed to be a critical factor to ensure delivery of high quality specimens for downstream ctDNA analysis[23]. In these studies, variable results were observed at extreme temperatures, at ≤10°C and 40°C, which affected the cellular interface, resulted in an elevated ratio of long:short genomic DNA fragments, and a decrease in plasma volume. These studies indicate that shipment temperature should be carefully controlled by use of insulated packages, gel blocks or temperature logging devices to maintain stability.

Previous studies have mainly focused on locus-specific analysis using quantitative PCR or dPCR which examined one locus at a time. With technology advances, an increasing number of molecular profiling strategies have been developed using NGS [24], which provides a higher resolution and larger genomic coverage than a locus-specific approach. It is therefore important to also understand the effects of cfDNA sample processing on the analytical performance of NGS-based analysis. It is particularly important to test whether using a collection tube containing a preservative has the potential to introduce DNA sequence modifications, which may be misinterpreted as true patient-specific genomic alterations. A recent study examined the influence of sample collection in CellSave tubes on the analysis of global copy number variations using NGS technology, and did not find differences in allele frequencies compared to EDTA blood [9]. In this study with BCT and K_3EDTA tubes, we evaluated the effects of processing on the background error rates during targeted amplicon sequencing and sWGS. As expected, we observed different error rates in different base

substitutions, but there was no difference in background error rate regardless of the type of collection device and sample processing schedule. Our sWGS analysis results agreed with previous findings in that copy number data was consistent across conditions.

All of these findings provide important insights for the potential incorporation of routine NGS technology in plasma-based molecular diagnostics. Beyond the analysis of ctDNA, it is crucial to also understand the impact of pre-analytical factors on other nucleic acids or genomic variants, such as tumour-specific RNA (ctRNA), microRNA or DNA methylation, some of which have been studied [25] but more evidence is required. Their quantification would likely be affected by the levels of total RNA or methylated DNA that is derived from the blood cells. It is important to understand whether the effects of sample processing procedures could be addressed in a similar manner to the effects on circulating DNA.

With the increasing understanding of genomic alterations and matched targeted treatment options, the demand for a non-invasive molecular profiling tool is growing. Analysing cell-free nucleic acids presents a unique opportunity for longitudinal follow-up during treatment of cancer patients. Initiatives have begun to pursue the standardization of methods for cell-free DNA analysis. Understanding the impact of different pre-analytic factors will help accelerate the process and drive large-scale cross-centre validation studies to provide robust evidence for clinical utility of circulating tumour DNA and its integration into routine clinical practice.

Acknowledgements

We thank the patients for consenting to participate, and the Human Research Tissue Bank at Addenbrooke's Hospital, which is supported by the NIHR Cambridge Biomedical Research Centre. We acknowledge the Genomics Core of the Cancer Research UK Cambridge Institute for sequencing support. We thank Frank Diehl for discussions about method standardisation, and Irena Hudecova for help preparing figures.

Authors' contribution

D.W.Y.T., S-J.D., H.B., A.P., C.P., N.R., J.B., C.C. and D.G. conceived and designed the study. D.W.Y.T. H.B., C.H., L.J., A.P., C.P., S-J.D. contributed to sample processing, collection of clinical data and sample management. D.W.Y.T., B.R., A.R.V.M.A., E.M. D.G and S-J.D. performed NGS and dPCR experiments. F.M., J.M., D.C., and V.P. analysed the NGS data. D.W.Y.T., B.R., DC and DG wrote the manuscript and all authors approved the final version.

References

[1] Wan JC, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nature reviews Cancer. 2017 Apr;17(4):223-238

[2] Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. Nature reviews Clinical oncology. 2017 Sep;14(9):531-548

[3] Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Science translational medicine. 2015; Aug 26;7(302):302ra133

[4] Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Science translational medicine. 2014;6:224ra24.

[5] Lui YY, Chik KW, Chiu RW, Ho CY, Lam CW, Lo YM. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. Clin Chem. 2002;48:421-7.

[6] El Messaoudi S, Rolet F, Mouliere F, Thierry AR. Circulating cell free DNA: Preanalytical considerations. Clin Chim Acta. 2013;424:222-30.

[7] Chiu RWK, Poon LLM, Lau TK, Leung TN, Wong EM, Lo YMD. Effects of blood-processing protocols on fetal and total DNA quantification in maternal plasma. Clin Chem. 2001;47:1607-13.

[8] Barrett AN, Zimmermann BG, Wang D, Holloway A, Chitty LS. Implementing prenatal diagnosis based on cell-free fetal DNA: accurate identification of factors affecting fetal DNA yield. PLoS ONE 6(10): e25202.

[9] Rothwell DG, Smith N, Morris D, Leong HS, Li Y, Hollebecque A, et al. Genetic profiling of tumours using both circulating free DNA and circulating tumour cells isolated from the same preserved whole blood sample. Molecular oncology. 2016;10:566-74.

[10] Fernando MR, Chen K, Norton S, Krzyzanowski G, Bourne D, Hunsley B, et al. A new methodology to preserve the original proportion and integrity of cell-free fetal DNA in maternal plasma during sample processing and storage. Prenat Diagn. 2010;30:418-24.

[11] Norton SE, Lechner JM, Williams T, Fernando MR. A stabilizing reagent prevents cell-free DNA contamination by cellular DNA in plasma during blood sample storage and shipping as determined by digital PCR. Clin Biochem. 2013;46:1561-5.

[12] Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med. 2013;368:1199-209.

[13] Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Science translational medicine. 2012;4:136ra68.

[14] Wang J, Ramakrishnan R, Tang Z, Fan W, Kluge A, Dowlati A, et al. Quantifying EGFR alterations in the lung cancer genome with nanofluidic digital PCR arrays. Clin Chem. 2010;56:623-32.

[15] Heitzer E, Ulz P, Belic J, Gutschi S, Quehenberger F, Fischereder K, et al. Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. Genome Med. 2013;5:30.

[16] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-60.

[17] R-Core-TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015.

[18] Affymetrix. Median of the Absolute Values of all Pairwise Differences and
Quality Control on Affymetrix Genome-Wide Human SNP Array 6.0. White Paper. 2008.

[19] Wong D, Moturi S, Angkachatchai V, Mueller R, DeSantis G, van den Boom D, et al. Optimizing blood collection, transport and storage conditions for cell free DNA increases access to prenatal testing. Clin Biochem. 2013;46:1099-104.

[20] Gusnanto A, Wood HM, Pawitan Y, Rabbitts P, Berri S. Correcting for cancer genome size and tumour cell content enables better estimation of copy number alterations from next-generation sequence data. Bioinformatics. 2012;28:40-7.

[21] Parpart-Li S, Bartlett B, Popoli M, Adleff V, Tucker L, Steinberg R, et al. The effect of preservative and temperature on the analysis of circulating tumor DNA. Clin Cancer Res. 2016.

[22] Warton K, Yuwono NL, Cowley MJ, McCabe MJ, So A, Ford CE. Evaluation of Streck BCT and PAXgene Stabilised Blood Collection Tubes for Cell-Free Circulating DNA Studies in Plasma. Mol Diagn Ther. 2017 Oct;21(5):563-570

[23] Medina Diaz I, Nocon A, Mehnert DH, Fredebohm J, Diehl F, Holtrup F. Performance of Streck cfDNA Blood Collection Tubes for Liquid Biopsy Testing. PloS one. 2016;11:e0166354.

[24] Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. Nat Rev Genet. 2013;14:415-26.

[25] Page K, Guttery DS, Zahra N, Primrose L, Elshaw SR, Pringle JH, et al. Influence of plasma processing on recovery and analysis of circulating nucleic acids. PLoS ONE 8(10): e77963

LEGENDS

Table 1. Summary of number of samples analysed in each Module.

Summary of collection devices, temperatures, number of samples and delay in processing for each analysis Module:

Module 1 Effects of delayed processing

Module 2 Effects of storage temperature

Module 3 Effects of collection devices (EDTA versus BCT)

Module 4 Effects of different centrifugation speed

Module 5 Effects of shipment in BCT

For Module 2, samples were stored both in room temperature (RT) and at 4°C. 20/11 indicates that 20 tubes were stored at room temperature and 11 at 4°C, etc.

Figure 1. The effects of delayed processing on the levels of circulating DNA in plasma collected in K_3EDTA tubes

Blood samples were collected into K_3EDTA tubes and stored at room temperature (RT) for 0h, 6h, 24h, 48h, 96h and 1week before plasma separation.

(A) cfDNA copies/mL plasma and (B) mutant allele fraction and (C) ctDNA copies/mL plasma in samples processed at different time of delay. All data were expressed as ratio from E.RT.0h (each patient's immediately processed K_3EDTA sample). Asterisk indicates statistical significant difference compared with E.RT.0h (Mann Whitney Rank Sum Test, p-value<0.05). The bottom and top of the box represent the first and third quartiles, and the band inside the box represents the median.

Figure 2. The effects of storage temperature on the levels of circulating DNA in plasma collected in K_3EDTA tubes

Blood samples collected into K_3EDTA tubes were stored at room temperature and at 4°C for 24h, 48h, 96h and 1 week before plasma was separated.

(A) cfDNA copies/mL plasma and (B) mutant allele fraction. All data were expressed as ratio from E.RT.0h (each patient's immediately processed K_3EDTA sample). Asterisk indicates statistical significant difference compared with E.RT.0h (Mann Whitney Rank Sum Test, p-value<0.05). The bottom and top of the box represent the first and third quartiles, and the band inside the box represents the median.

Figure 3. The effects of collection device (K_3EDTA versus BCT) on the levels of circulating DNA

Blood samples collected into K_3EDTA tubes were processed immediately, after 96h or one week at room temperature. Blood samples in BCT were stored at room temperature for 96h and 1 week before plasma separation.

(A) cfDNA copies/mL plasma and (B) mutant allele fraction. All data were expressed as ratio from E.RT.0h (each patient's immediately processed K_3EDTA sample). Asterisk indicates statistical significant difference compared with E.RT.0h (Mann Whitney Rank Sum Test, p-value<0.05) (C) The distributions of the ratio of non-reference:reference alleles as generated by targeted amplicon sequencing shown in boxplots. The bottom and top of the box represent the first and third quartiles, and the band inside the box represents the median. The data are represented in log (10) scale.

Figure 4. The effects of different centrifugation speed on the levels of circulating DNA

Blood samples were collected into K_3EDTA tubes and processed to plasma with three different protocols. All protocols included two 10 min centrifugation steps, the first on whole blood, and the second on plasma aliquots. Protocol A (820g and 14 000g), Protocol B (1600g and 14000g), Protocol C (1600g and 3000g). (A and B) cfDNA copies/mL plasma and (C and D) mutant allele fractions (%) in samples processed by different protocols. The bottom and top of the box represent the first and third quartiles, and the band inside the box represents the median.

Figure 5. The effects of shipping using Cell-free DNA BCT on the levels of circulating DNA

Blood samples were collected in K_3EDTA tubes and Cell-free DNA BCT and processed immediately except for one Cell-free DNA BCT from each collection that was shipped by mail back to the same collection centre. (A and B) cfDNA levels (AC/µl) and (C and D) mutant allele fractions. The bottom and top of the box represent the first and third quartiles and the band inside represent the median.

Supplemental Figure S1

The numbers of ctDNA copies, expressed as copies/mL plasma, in (A) different storage and processing conditions and (B) the ratio from the sample collected in K_3EDTA tubes and processed immediately (E.RT.0h).

Supplemental Figure S2

Mutant allele fractions of matched samples from 6 patients, those collected in K_3EDTA tubes and processed immediately versus those collected in BCT and processed after 1 week of delay.

Supplemental Figure S3 Pairwise comparison between copy number profiles of plasma samples processed with different protocols.

"EDTA" indicates the sample was collected in K_3EDTA tube and processed immediately; "BCT post" indicates the sample was collected in BCT and delivered in the post and processed upon arrival; "BCT" indicates the sample was collected in BCT and stored at room temperature until processing with the posted samples. The copy number data were expressed as log2ratio of the segmented bin counts. The pairwise comparisons were assessed by Spearman correlation, and the correlation coefficients were indicated on top of each panel.

Supplemental Figure S4. The effects of shipping using Cell-free DNA BCT on global copy number alterations from sWGS data (0.175x fold coverage).

Shallow whole genome sequencing data of an ovarian cancer patient's (P227) plasma sample processed with three different protocols: (A) collection in K_3EDTA tube and processed immediately; (B) collection in BCT, stored at room temperature, not processed until the posted sample arrived; (C) collection in BCT, delivered in the post and processed upon arrival. The data were expressed as log2ratio of segmented bin counts across the 23 chromosomes. Blue indicates a statistically significant gain, and orange indicates a significant loss.

Supplemental Figure S5. Estimating noise from the copy number profiles generated by sWGS data

Median of all absolute pairwise differences (MAPD) between coordinate-sorted copy number values for the different protocols. The data points are coloured by patient ID.

Supplemental Table S1. Summary of samples analysed in each Module

Table details the number of ovarian and breast cancer samples analysed in each Module, at different timepoints, temperatures or using different methods of analysis (dPCR or TAm-Seq). ND = Not detected; + = Sample analysed; RT = Room temperature

Supplemental Table S2. dPCR Assays

Table details the sequences of the primer and probes sequences of the dPCR assays targeting specific mutant and wild-type sequences, the annealing temperature and number of PCR cycles used to amplify sequences in the dPCR assay, and the amplicon sizes of each of the dPCR assays.

Supplemental Table S3. Comparison of the segmental copy number profiles among the different collection methods

Table details the values for the Wilcoxon Rank Sum Test, adjusted p-value and Spearman correlation for the different comparison of segmental copy number profiles using EDTA, BCT RT or Posted BCT collection tubes.

Figure 1

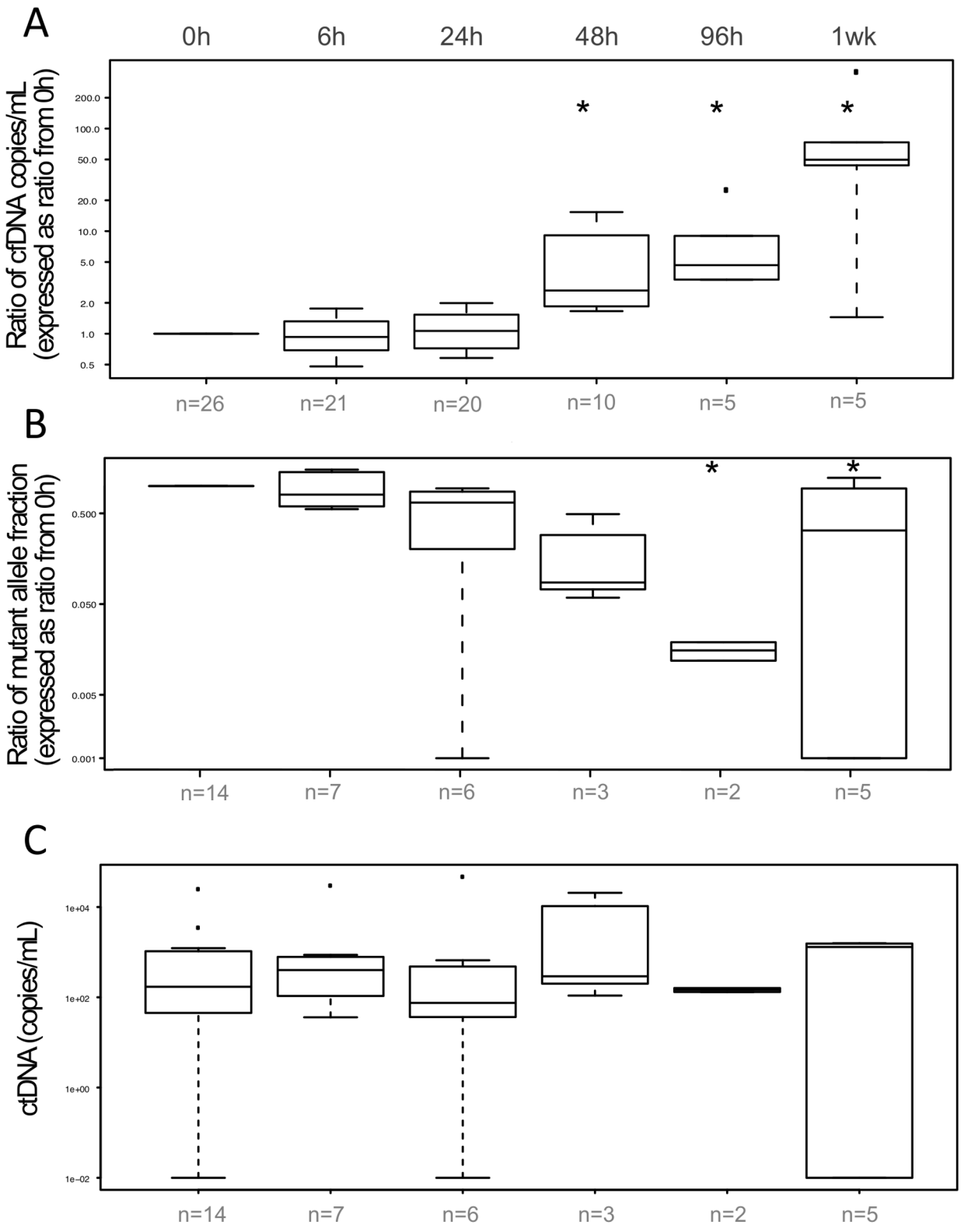

A
0h
6h
24h
48h
96h
1wk
Ratio of cfDNA copies/mL (expressed as ratio from 0h)
200.0
100.0
50.0
20.0
10.0
5.0
2.0
1.0
0.5
n=26
n=21
n=20
n=10
n=5
n=5
B
Ratio of mutant allele fraction (expressed as ratio from 0h)
0.500
0.050
0.005
0.001
n=14
n=7
n=6
n=3
n=2
n=5
C
ctDNA (copies/mL)
1e+04
1e+02
1e+00
1e−02
n=14
n=7
n=6
n=3
n=2
n=5

Figure 2

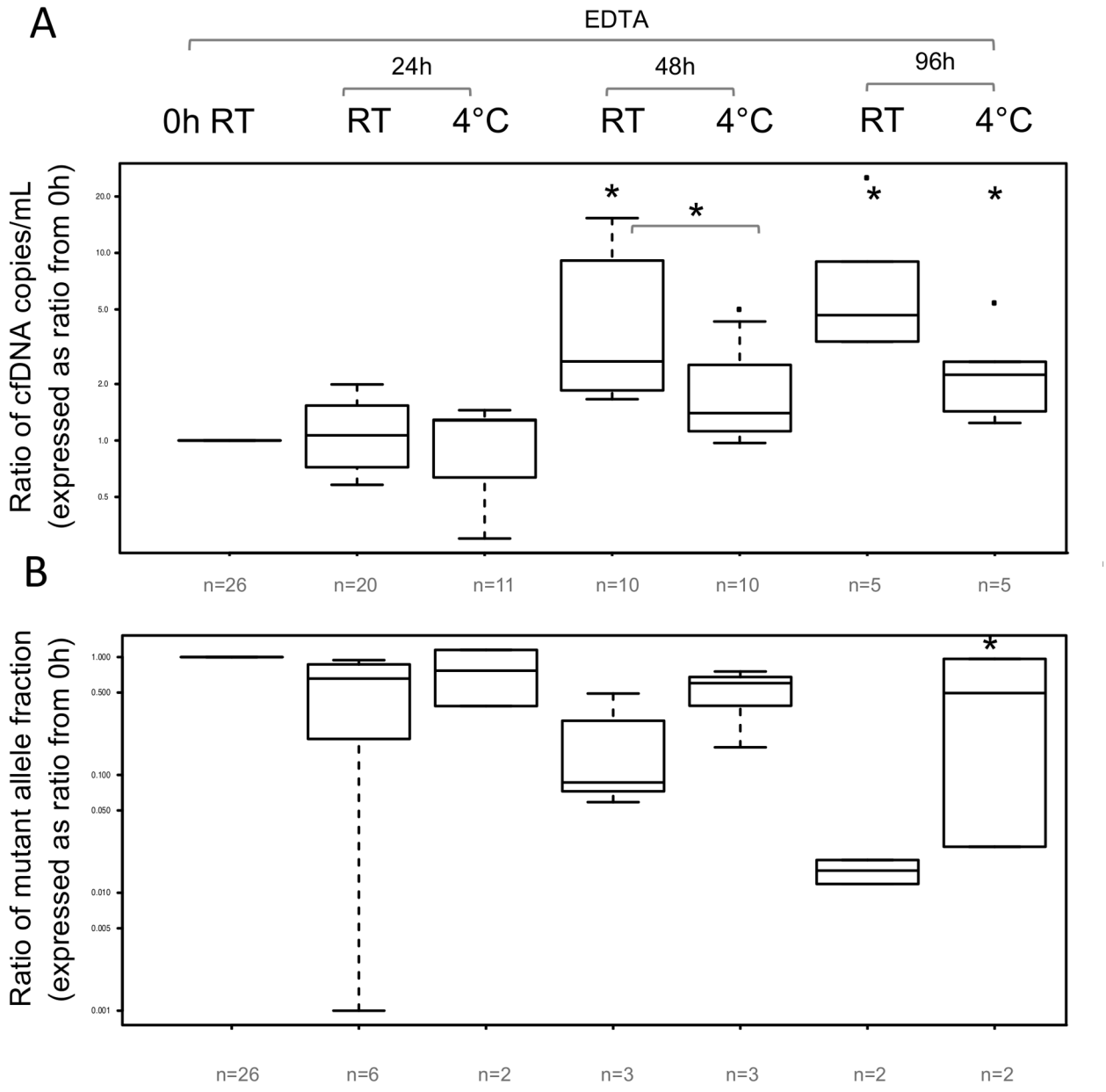

A
EDTA
24h
48h
96h
0h RT
RT
4°C
RT
4°C
RT
4°C
Ratio of cfDNA copies/mL
(expressed as ratio from 0h)
20.0
10.0
5.0
2.0
1.0
0.5
n=26
n=20
n=11
n=10
n=10
n=5
n=5
B
Ratio of mutant allele fraction
(expressed as ratio from 0h)
1.000
0.500
0.100
0.050
0.010
0.005
0.001
n=26
n=6
n=2
n=3
n=3
n=2
n=2

Figure 3

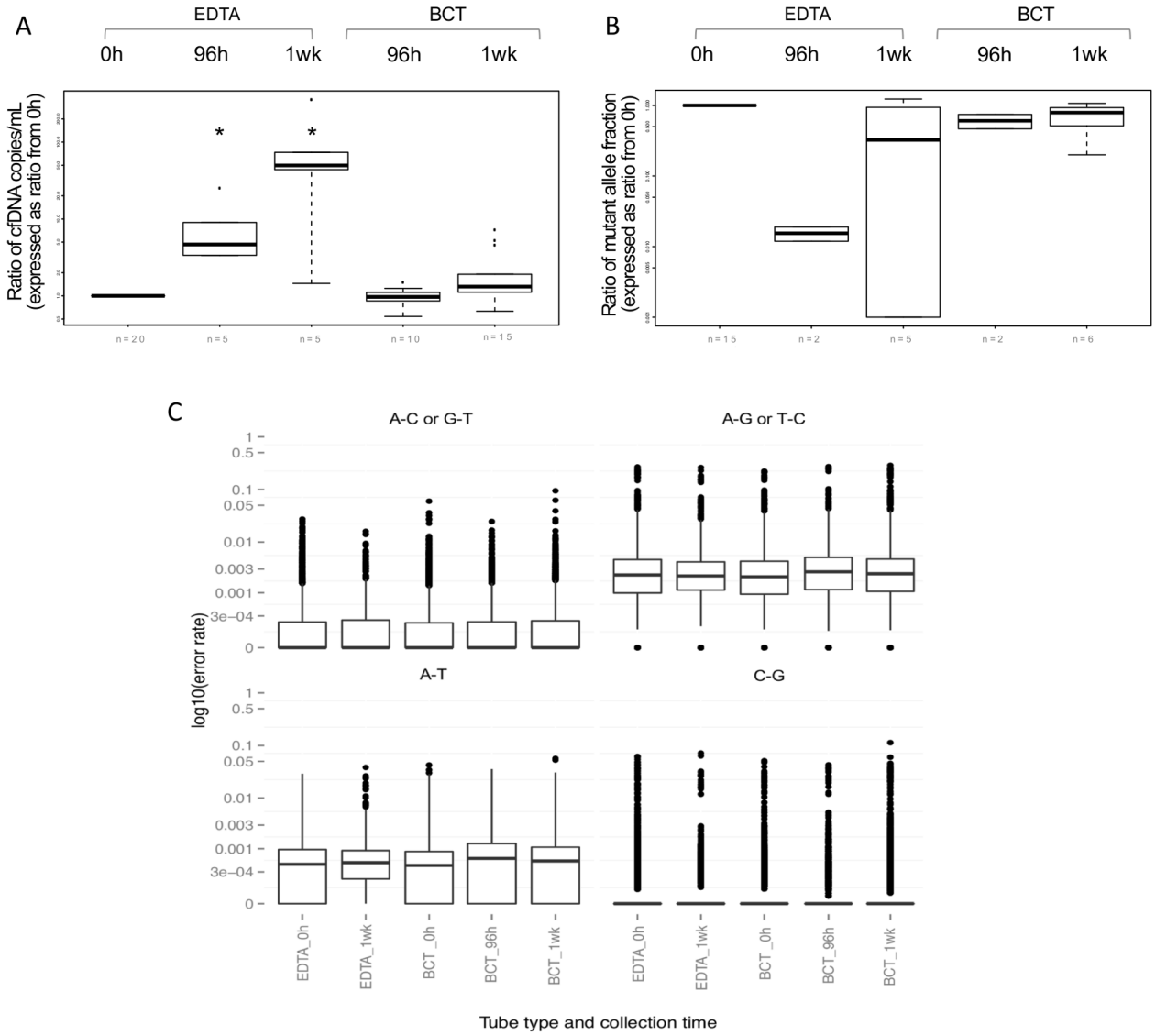

A
EDTA
BCT
0h
96h
1wk
96h
1wk
Ratio of cfDNA copies/mL (expressed as ratio from 0h)
n = 20
n = 5
n = 5
n = 10
n = 15
B
EDTA
BCT
0h
96h
1wk
96h
1wk
Ratio of mutant allele fraction (expressed as ratio from 0h)
n = 15
n = 2
n = 5
n = 2
n = 6
C
A-C or G-T
A-G or T-C
A-T
C-G
log10(error rate)
1
0.5
0.1
0.05
0.01
0.003
0.001
3e−04
0
EDTA_0h
EDTA_1wk
BCT_0h
BCT_96h
BCT_1wk
Tube type and collection time

Figure 4

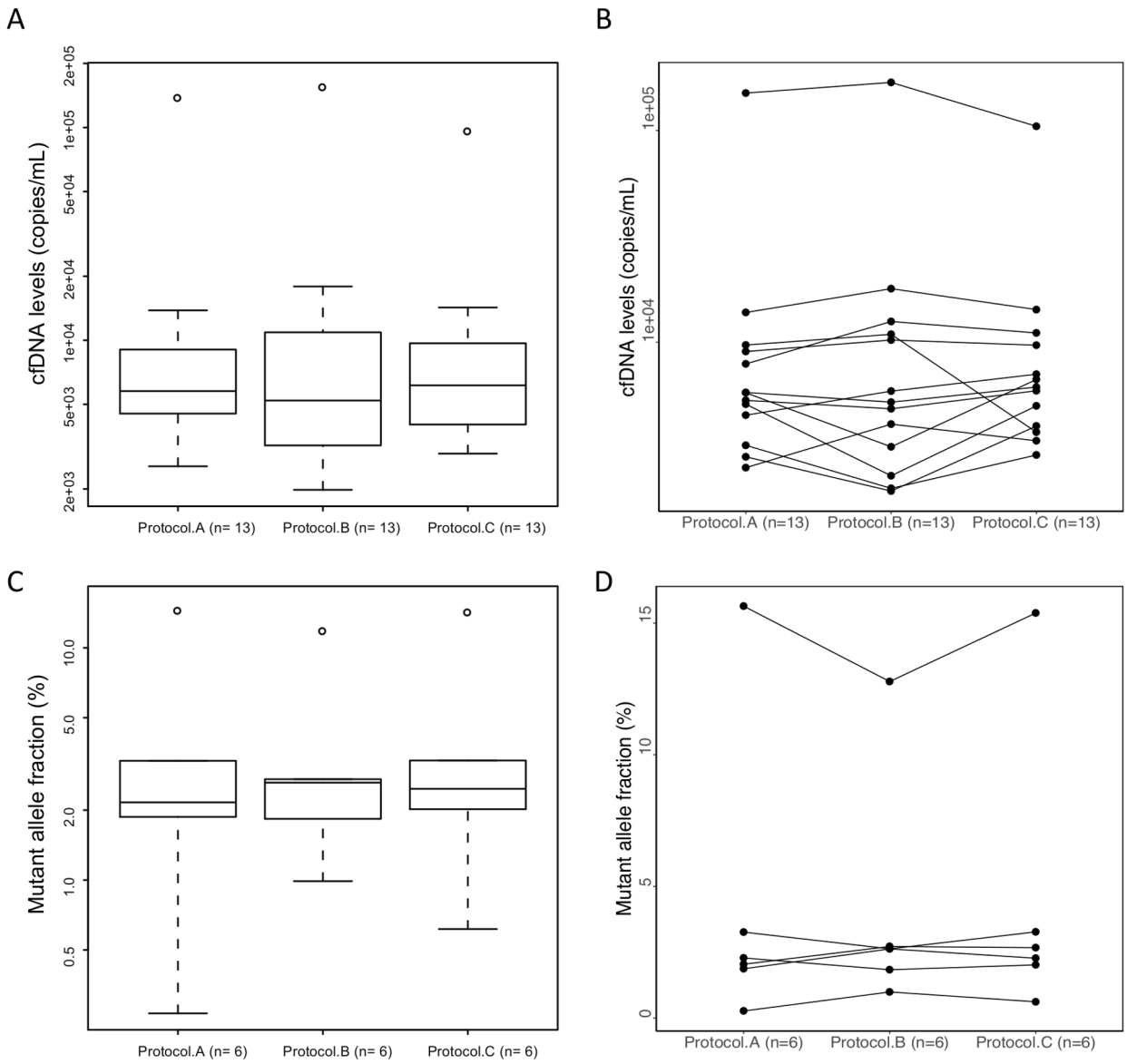
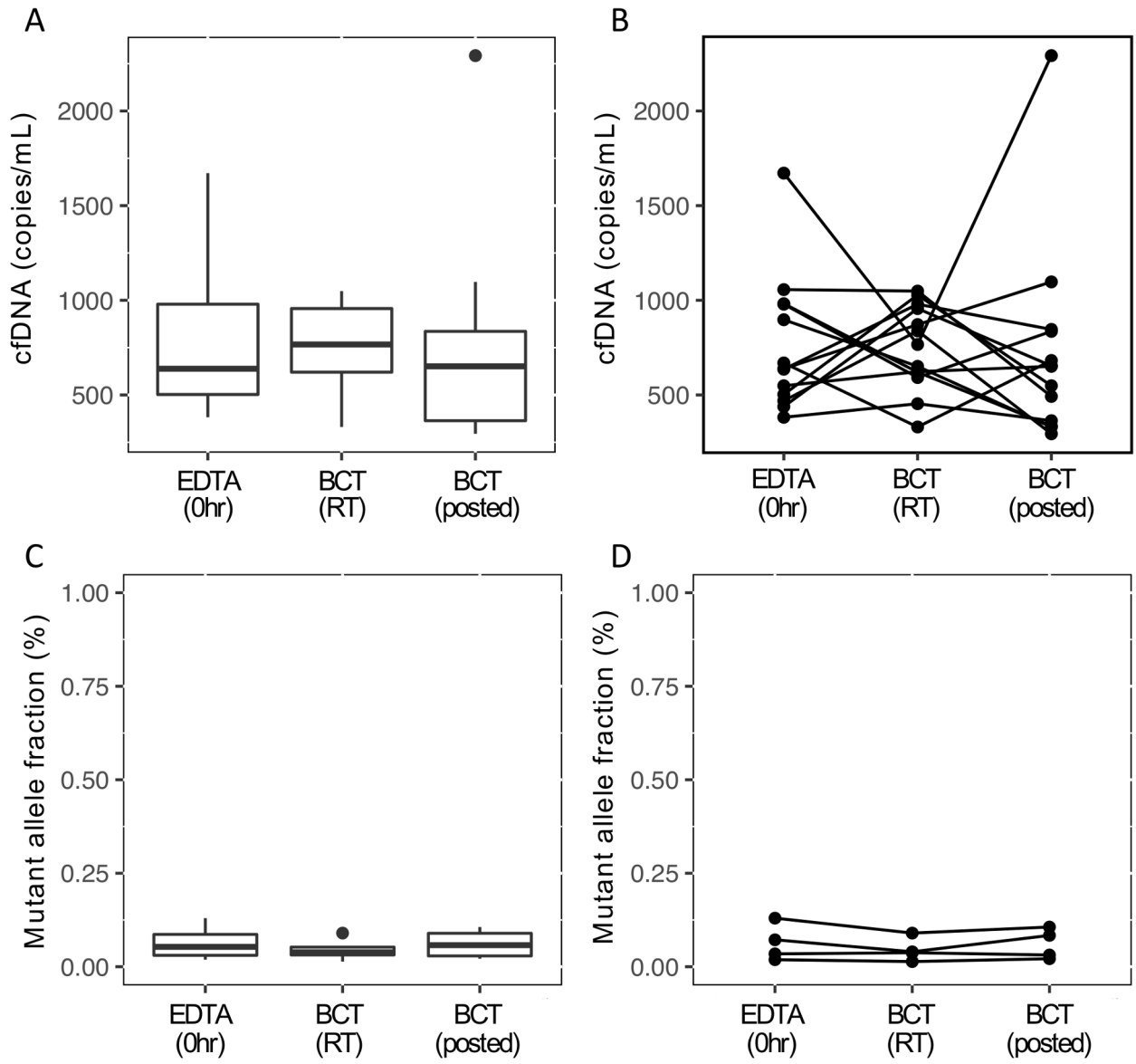

A
cfDNA levels (copies/mL)
2e+03
5e+03
1e+04
2e+04
5e+04
1e+05
2e+05
Protocol.A (n= 13)
Protocol.B (n= 13)
Protocol.C (n= 13)
B
cfDNA levels (copies/mL)
1e+04
1e+05
Protocol.A (n=13)
Protocol.B (n=13)
Protocol.C (n=13)
C
Mutant allele fraction (%)
0.5
1.0
2.0
5.0
10.0
Protocol.A (n= 6)
Protocol.B (n= 6)
Protocol.C (n= 6)
D
Mutant allele fraction (%)
0
5
10
15
Protocol.A (n=6)
Protocol.B (n=6)
Protocol.C (n=6)

Figure 5


A
B
C
D
cfDNA (copies/mL)
2000
1500
1000
500
Mutant allele fraction (%)
1.00
0.75
0.50
0.25
0.00
EDTA (0hr)
BCT (RT)
BCT (posted)

12.22 Mutation calling using TAm-Seq for serum samples

OV04	Chemotheraphy cycle	Mutation	MAF pre size selection	MAF post size selection
77	1	c.C916T, p.306X	ND	0.24
77	2	c.C916T, p.306X	ND	0.01
83	1	c.C833T, p.P278L	ND	0.05
83	2	c.C833T, p.P278L	0.33	0.04
122	1	c.C817T, p.R273C	0.58	0.22
122	2	c.C817T, p.R273C	ND	ND
141	1	c.T584C, p.I95T	ND-low coverage	0.41
141	2	c.T584C, p.I95T	ND-low coverage	0.04
180	1	c.574_578del, p.192_193del	ND	0.1
180	1	c.574_578del, p.192_193del	ND	0.01
211	1	c.611delA, p.E204fs	0.04	0.05
211	2	c.611delA, p.E204fs	ND	ND
226	1	c.225delT, p.P75fs	ND-low coverage	0.03
226	2	c.225delT, p.P75fs	ND-low coverage	ND-low coverage
292	1	c.A578G, p.H193R	ND	0.08
292	2	c.A578G, p.H193R	ND	ND
295	1	c.T613C, p.Y205H	0.63	0.13
295	2	c.T613C, p.Y205H	ND	ND
297	1	c.G661T, p.E221X	ND	0.07
297	2	c.G661T, p.E221X	ND-low coverage	0.09
300	1	c.G560T, p.G187V	0.41	0.04
300	2	c.G560T, p.G187V	ND	0.02

ND - Not detected