

Clonal evolution of ovarian high grade serous carcinoma

Zhao Cheng

Imperial College London

Department of Surgery and Cancer

Thesis submitted for the degree of Doctor of Philosophy
Clinical Medicine Research

May, 2022

Declaration of Originality

Except where specifically stated and acknowledged, I hereby declare this thesis is a result of my own work.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

Ovarian high-grade serous carcinoma (HGSC) is usually diagnosed at late stage. I investigated whether late-stage HGSC has unique genomic characteristics consistent with acquisition of evolutionary advantage compared with early-stage tumours and the genomic dynamic changes in the evolution process of HGSC.

Targeted next-generation sequencing and shallow whole-genome sequencing (sWGS) were performed on samples from 43 chemotherapy-naïve patients with FIGO stage I-IIA HGSC to investigate somatic mutations and copy-number (CN) alterations (SCNA). I compared results to pretreatment samples from 52 patients with late stage IIIC/IV HGSC from the BriTROC-1 study. In addition, I also collected p53 signatures, serous tubal intraepithelial carcinomas (STIC), primary and metastatic tumours from five patients as a separate cohort.

Age of diagnosis did not differ between early-stage and late-stage patients (median 61.3 years vs. 62.3 years, respectively). *TP53* mutations were near-universal in both cohorts (89% early-stage, 100% late-stage), and there were no significant differences in the rates of other somatic mutations, including *BRCA1* and *BRCA2*. I also did not observe cohort-specific focal SCNA that could explain biological behavior. However, ploidy was higher in late-stage (median, 3.0) than early-stage (median, 1.9) samples. Seven copy number signatures, recurrent patterns of copy number change that reflect the imprint of underlying mutational process, had previously been derived from late stage high grade serous carcinomas^[1]. CN signature exposures were significantly different between cohorts, with greater relative signature 3 exposure in early-stage and greater signature 4 in late-stage. Signature 3 is associated with *BRCA1/2* homologous recombination deficiency pathway and good overall survival. Signature 4 is associated with abnormalities in the PI3K/AKT signalling pathway as well as

and whole genome duplication. Unsupervised clustering based on CN signatures identified three clusters that were prognostic. In the STIC cohort, all p53 signatures were diploid. Two patients showed no significant difference in ploidies across samples. However, in three patients, ploidy increase was apparent in STIC lesions as well as primary and metastatic lesions.

Early-stage and late-stage HGSCs have highly similar patterns of mutation and focal SCNA. However, CN signature analysis showed that late-stage disease has distinct signature exposures consistent with whole-genome duplication. The STIC cohort provides some evidence that complex genomic alterations can occur very early in the development of HGSC development.

Acknowledgements

Time flies so fast during my PhD at IRDB for more than three years. It has been an unforgettable experience for me to study in Imperial College London and I want to say thank you to lots of people who helped me during my PhD.

First of all, I would like to express my utmost sincere appreciation to my main supervisor Professor Iain McNeish. I am grateful for Iain's guidance and knowledge, which has encouraged me to push through the difficulties during my Ph.D. research. It's full of realistic challenges for an international student to join in a new group in a different university and totally unfamiliar environment. Iain helped me with persistent encouragement and made me adapt to the lab quickly. As a totally wet lab person in my master time, Iain also cultured me bioinformatics knowledge for my PhD project. And Iain also gave me lots of chance to attend conferences and summer school to broaden my horizon and practiced my communication skills to present my work. At the most difficult time of COVID-19 and lockdown in London, Iain's strong support for my life and work gave me confidence to move on and accomplish my project.

Then, I would also like to thank the brilliant colleagues and collaborators that I have been working with during my PhD. I will be great grateful to Dr Darren Ennis for helping me settle down in a new lab and also teaching me hand by hand about sample collection and clinical knowledge. Many thanks to Mr Hasan Mirza for teaching me the bioinformatics techniques and dealing with endless bugs in the commands. Thanks to our lab manager Sophie and Lisa that helping me to order lots of reagents and kits. Also thanks to Gaia and Raffa for their professional gynecological knowledge and provide me lots of suggestions from the view of doctors. To Sarah for her great passion in cancer immunology that also gave

me lots of new knowledge in cancer research. Carmen and Pavlina also helped me a lot in my work. Thanks to Ana, Nayana, MiQi, Dan and Izzy in our lab for their help. Besides, I am also grateful to my collaborators. Dr Jacqueline McDermott, Dr Naveena Singh and Dr Chishimba Sokota kindly helped me with pathology identification of all the slides. Dr Philip Smith, Dr Thomas Bradley and Lena Morrill Gavarró in University of Cambridge were helping me for my bioinformatics analysis. Thanks to Dr Katerina Rekopoulou in Imperial BRC facility that tolerating me hundreds of sequencing questions. And Ms Nadia Rahman in Barts institute and Ms Reema Paudel in Imperial ECMC that helped me cutting block sections.

Finally, I am also deeply grateful for my parents, my close friends' love, support, and help; which are the foundation of this thesis. My parents who have been very open and supportive to my choices, while giving me enough support at the difficult time in COVID-19. At last, thanks to yihuai, for you endless support and kindness to help me in the hardest times and longest of days.

Abbreviations

broad range (BR)

Burrows-Wheeler Alignment (BWA)

Central sequence-specific DNA binding domain (DBD)

coding regions (CDS)

complete absence (CA)

copy number (CN)

copy number alterations (CNA)

CTD: C-terminal domain

cyclin-dependent kinase 2 (CDK2)

double-strand break (DSB)

droplet digital PCR (ddPCR)

Endometrioid ovarian carcinoma (EOVC)

Epithelial-mesenchymal transition (EMT)

Eukaryotic Translation Initiation Factor 4A2 (EIF4A2)

extracellular matrix (ECM)

gain of function (GOF)

Gaussian Mixture Model (GMM)

haematoxylin & eosin (H&E)

high sensitivity (HS)

homologous recombination-deficient (HRD)

immunohistochemistry (IHC)

intraperitoneal (IP)

intravenous (IV)

isometric log-ratio (ILR)

laser capture microdissection (LCM)

loss of heterozygosity (LOH)

maximum likelihood (ML)

multimodal screening (MMS)

Neurofibromatosis type 1 (NF1)

Oligomerization domain (OD)

ovarian surface epithelium (OSE)

overall survival (OS)

Pan-Cancer Analysis of Whole Genomes (PCAWG)

Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha (PIK3CA)

phosphorylated Akt (p-Akt)

polyethylene naphthalate (PEN)

progression-free survival (PFS)

proline-rich domain (PRD)

Sectioning and Extensively Examining the FIMbria (SEE-FIM)

serous tubal intraepithelial carcinoma (STIC)

serous tubal intraepithelial lesion (STIL)

sWGS-absoluteCN (swgs)

tagged amplicon sequencing (TAm-seq)

tandem duplications (TD)

The Cancer Genome Atlas (TCGA)

transcriptional activation domains (TAD)

transvaginal sonography (TVS)

Tubal intraepithelial carcinoma (TIC)

tubal intraepithelial lesion (TIL)

ultrasound screening (USS)

United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)

whole genome duplication (WGD)

whole-genome amplification (WGA)

wild-type (WT)

Table of contents

Declaration of Originality	2
Copyright Declaration.....	2
Abstract	3
Acknowledgements	5
Abbreviations	7
Table of contents	10
List of tables	15
List of figures	16
CHAPTER 1. Introduction.....	21
1.1.General Introduction to ovarian cancer.....	22
1.2 Epidemiology of ovarian cancer	22
1.3 Staging	23
1.4 Grading	24
1.5 Treatment of ovarian cancer	25
1.6 Ovarian cancer classification	27
1.6.1 Endometrioid ovarian carcinoma.....	28
1.6.2 Clear cell carcinoma	29
1.6.3 Mucinous carcinoma.....	29
1.6.4 Low grade serous carcinoma	30
1.6.5 High-grade serous carcinoma	30
1.7 Biology of HGSC.....	31
1.7.1 Frequently mutated genes	31
1.7.1.2 <i>TP53</i> mutation.....	31
1.7.1.3 BRCA1/2 mutations.....	33
1.7.1.4 Other recurrent mutations	34
1.7.2 Copy number abnormalities in HGSC	35
1.7.2.1 Copy Number Landscape of HGSC.....	38
1.7.2.2 Copy number signatures	39
1.8 HGSC development hypothesis	42

1.9 Pathological identification of STIC	43
1.10 Summary and Hypothesis	46
1.10.1 Aims	46
CHAPTER 2. Materials and Methods.....	47
2.1 Study design and workflow	48
2.2 Clinical data and sample collection	49
2.2.1 Hammersmith Hospital, UCLH and Barts archival and BriTROC-1 HGSOc dataset-‘early stage cohort’	49
2.2.2 BriTROC-1 HGSOc datasetlate-‘late-stage cohort’	50
2.2.3 Hammersmith and St Bartholomew’s Hospitals HGSC dataset-‘STIC cohort’	50
2.3 Pathology	51
2.4 DNA extraction and quantification.....	51
2.5 H&E, p53 and Ki67 staining for identification of target lesions	55
2.6 Immunohistochemistry for p53 using PEN Slides.....	56
2.7 Laser capture microdissection by Zeiss microscope.....	58
2.8 Targeted panel sequencing.....	60
2.8.1 Targeted amplicon sequencing for late stage cohort.....	60
2.8.2 AmpliSeq Custom Panel design for early stage cohort	60
2.8.3 Ampliseq library preparation	60
2.8.4 AmpliSeq sequencing	63
2.9 Somatic mutation calling from target sequencing	63
2.9.1 Somatic mutations calling from Ampliseq panel.....	63
2.9.2 Somatic mutations calling from TAm-seq panel	64
2.10 Shallow whole genome library prep and sequencing	64
2.11 Copy number calling.....	68
2.11.1 Relative copy number	68
2.11.2 Absolute copy number	68
2.11.3 Copy number signature calling	69
2.12 Statistical analysis.....	71
2.12.1 Unsupervised hierarchical clustering statistical analysis.....	71
2.12.2 Survival analysis	71
2.13 Materials	72
2.14 Equipment & plasticware.....	72

CHAPTER 3. The genomic landscape of early stage ovarian high grade serous carcinoma ..	74
3.1 Introduction.....	75
3.2 Background.....	75
3.3 Results.....	77
3.3.1 The mutational landscape of early and late-stage cohorts using a custom-made panel of genes	77
3.3.2 SNV/indel mutation rates comparison between early stage and late stage cohort ..	83
3.3.3 <i>TP53</i> hotspots mutations in early stage and late stage cohort	89
3.3.4 Relative copy number profile of early stage and late stage cohort samples by QDNAseq.....	91
3.3.5 Absolute copy number inference of early stage and late stage cohort samples using sWGS-absolute and ACE packages	93
3.3.6 Purity and ploidy comparison of early-stage and late-stage cohorts	95
3.3.7 Exploring focal gene amplifications and deletions in early stage and late stage cohort	96
3.3.8 Focal genes amplification and deletions rate between early stage and late stage cohort	97
3.4 Discussion.....	99
3.4.1 Mutation rates and focal genes amplification and deletions rates between early stage and late stage cohort	99
3.4.2 Differential purity and ploidy between early stage and late stage cohort.....	101
3.5 Summary	101
CHAPTER 4. Copy number signatures in early stage high grade serous carcinoma	103
4.1 Introduction.....	104
4.2 Background.....	104
4.3 Results.....	108
4.3.1 Copy number signatures analysis of early stage and late stage cohort.....	108
4.3.2 Copy number signatures comparison between early stage and late stage cohort ..	112
4.3.2.1 Analysis of zero exposures	112
4.3.2.2. Partial ILR analysis.....	114
4.3.2.3. Summary of copy number signature comparison	115
4.3.3 Unsupervised clustering of copy number signatures from early and late stage cohorts.....	118

4.3.4 Survival analysis of copy number signatures, clusters and ploidy from early and late stage cohort	122
4.4 Discussion	126
4.4.1 Signature 3 and signature 4 significantly change in early and late stage cohort ..	126
4.4.2 Three clusters classification in early and late stage cohort.....	127
4.4.3 High ploidy/whole genome duplication associated with late stage subset population	128
4.5 Summary	130
CHAPTER 5. Genomic evolution during HGSC development.....	131
5.1 Introduction.....	132
5.2 Background	132
5.3 Results.....	134
5.3.1 Pathological identification of p53 signature and STIC lesions.....	134
5.3.2 Laser capture microdissection of p53 signature and STIC lesions	142
5.3.3 Relative copy number profile of clonal evolution process in HGSC.....	151
5.3.4 Absolute copy number profile of evolution process in HGSC	157
5.3.5 Amplifications and deletions in the evolution process of HGSC	160
5.3.6 Ploidy change in the clonal evolution process of HGSC.....	163
5.4 Discussion	164
5.4.1 Whole exome sequencing and shallow whole genome sequencing comparison in the evolution of HGSC.....	164
5.4.2 Ploidy in the evolution process in HGSC	165
5.4.3 High ploidy/whole genome duplication associated with evolution process of HGSC.....	166
5.4.4 Evolution process at anatomic sites of five HGSC patients	167
5.5 Summary	168
CHAPTER 6. General discussion	169
6.1 The genomics of evolution of high grade serous carcinoma	170
6.1.1 The evolution model in HGSC and other cancer types.....	170
6.1.2. The technical challenges and applicable methods to detect HGSC early	173
6.1.3. Different immunohistochemistry staining patterns at STIC lesions.....	175
6.1.4 Microenvironment of the fallopian tube	176
6.2 Technical challenges for sequencing the early lesions of HGSC	177

6.3 Limitations of early stage and STIC projects	179
6.4 Future perspectives	180
CHAPTER 7. References.....	181
CHAPTER 8. Appendix.....	193
Appendix 1. Clinical information of early stage cohort	194
Appendix 2. Clinical information of late stage cohort.....	195
Appendix.3 SNV mutations in early stage by samples by Mutect2, HaplotypeCaller, Strelka and Varscan2 pipelines.....	197
Appendix 4. SNV mutations in late stage by samples by Mutect2, HaplotypeCaller, Strelka pipelines.....	198
Appendix 5. Amplicons for Ampliseq sequencing for Early stage cohort	199
Appendix 6. Indexes for pooling guide of Ampliseq sequencing.....	218
Appendix 7. Tapestation results of Ampliseq libraries preparation from early stage samples	220
Appendix 8. Absolute copy number profiles from swgs-absn pipeline in early stage cohort	223
Appendix 9. Absolute copy number profiles from swgs-absn pipeline in late stage cohorts	224
Appendix 10. DNA extraction result from all the samples in STIC project.....	228
Appendix 11. Absolute copy number profiles in all anatomic sites in five patients	230

List of tables

		Page
Table 1.1	FIGO Ovarian Cancer Staging	23-24
Table 1.2	Classification of ovarian cancer	27-28
Table 3.1	Clinical characteristics of study participants used in the discovery cohort	81
Table 4.1	Fraction of zero exposures for each group/signature combination	113
Table 5.1	Clinical features of STIC cohort. HGSC: high grade serous ovarian cancer	135
Table 5.2	Pathology features of STIC cohort.	136
Appendix 1	Clinical information of early stage cohort.	194
Appendix 2	Clinical information of late stage cohort	195-196
Appendix 3	SNV mutations in early stage by samples by Mutect2, HaplotypeCaller, Strelka and Varscan2 pipelines.	197
Appendix 4	SNV mutations in late stage by samples by Mutect2, HaplotypeCaller, Strelka and Varscan2 pipelines.	198
Appendix 5	Amplicons for Ampliseq sequencing for Early stage cohort.	199-217
Appendix 6	Indexes for pooling guide of Ampliseq sequencing.	218-219
Appendix 10	DNA extraction result from all the samples in STIC project.	228-229

List of figures

		Page
Figure 1.1	p53 domains and hotspot mutations	32
Figure 1.2	Relative copy number profile across the chromosome	36
Figure 1.3	Pan-cancer cancer classification	37
Figure 1.4	Overview of copy number aberrations in 101 high grade serous carcinoma cases	39
Figure 1.5	The features of the seven copy number signatures in HGSC	41
Figure 1.6	The current hypothesis for the development and evolution of HGSC. Normal fallopian tube epithelium develops into p53 signature and STIL. STIC lesions transform into invasive carcinoma	43
Figure 1.7	Diagnostic algorithm for STIC lesions.	45
Figure 2.1	Flowchart of project study design.	49
Figure 3.1	(A) Remark diagram for early-stage and late-stage cohorts by patients	79
	(B) Remark diagram for early-stage and late-stage cohorts by samples	80
Figure 3.2	Time interval between diagnosis and sample analysis by cohort	82
Figure 3.3	Clinical information on early and late-stage cohorts.	82
Figure 3.4	(a) Distribution of unique and duplicated reads of Ampliseq sequencing in early stage cohort samples. (b) Per sequence quality score shows the frequencies of quality scores in a sample. The quality of reads is good quality with the peak shifts to the right side. (c) Per base sequence quality presents the quality score across all bases at each position in the reads.	85-86
Figure 3.5	Adapter content before trimming and after trimming.	86

Figure 3.6	Short variants (SNV and indels) for each patient in early-stage and late-stage cohorts.	87
Figure 3.7	Counts of mutations in each sample in early-stage and late-stage cohorts	88
Figure 3.8	Relative copy number profiles of early stage four samples without TP53 mutations.	88
Figure 3.9	Gene mutation mapper plot of TP53 in early-stage cohort and late-stage cohort.	90
Figure 3.10	Hotspot mutations comparison between early stage and late stage cohorts.	90
Figure 3.11	(a) Distribution of unique and duplicated reads in early stage cohort samples. (b) Per sequence quality score shows the frequencies of quality scores in a sample. The quality of reads is good quality with the peak shifts to the right side. (c) Per base sequence quality presents the quality score across all bases at each position in the reads. (d) Adapter content is essentially 0% in all the samples.	92-93
Figure 3.12	Workflow analyses absolute copy number data by swgs and ACE packages.	95
Figure 3.13	Ploidy and purity comparison between sWGS-absolute and ACE pipelines.	95
Figure 3.14	(a) Purity comparison of early-stage and late-stage cohorts. (b) Ploidy comparison of early-stage and late-stage cohorts.	96
Figure 3.15	Global copy number amplifications, gains and losses in early-stage and late-stage cohorts.	97
Figure 3.16	Estimation of focal amplifications and deletions in 17 genes of interest, determined by sWGS.	98
Figure 3.17	Absolute copy number (y-axis) of 17 individual genes in early and late-stage samples.	99
Figure 4.1	Copy number feature distributions in early and late-stage HGSC.	109

Figure 4.2	Components distribution and comparison between early stage and late stage cohorts.	110
Figure 4.3	Tumour-by-component matrix distribution of 36 copy number components in early stage (A) and late (B) stage cohorts.	111
Figure 4.4	Copy number signature exposures in early-stage and late-stage patients. Note that signature exposures sum to 1 in each sample.	112
Figure 4.5	Results for the Bernoulli model of present/absent signatures.	114
Figure 4.6	Partial ILR analysis for the Analysis of non zero exposures. (a) Density plots for each signature in each group.	115
Figure 4.7	Summary of the results of both models, in which we can see how s4 is the only signature with clear changes both in presence/absence and compositional shift.	116
Figure 4.8	(A) Mean signature exposure proportions across the early-stage and late-stage cohorts. (B) Comparison of signature exposures across early-stage and late-stage cohorts	116
Figure 4.9	Overall survival of combined early and late-stage cohorts by zero vs non-zero exposures to copy number signature 3 (A) and 4 (B).	117
Figure 4.10	Simplex plots representing exposures for CN signature 3 (rest-signature 3 axis), signature 4 (signature 3-signature 4 axis) and the rest of the signatures (1-S3-S4) combined (signature 4 -rest axis) in early (A) and late (B) stage cohorts	118
Figure 4.11	Unsupervised hierarchical clustering in combined early-stage and late-stage cohorts.	120
Figure 4.12	Cluster dendrogram of unsupervised hierarchical clustering of samples by CN signatures.	120
Figure 4.13	Distributions of copy number signature exposures in three clusters	121
Figure 4.14	Early and late-stage samples by cluster; Chi-squared test.	121

Figure 4.15	Overall survival by cluster.	121
Figure 4.16	Forrest plot of hazard ratio estimates on overall survival (OS) for clusters.	122
Figure 4.17	Ploidy distribution of late-stage samples in three clusters.	123
Figure 4.18	Correlation between CN signature 4 exposure and ploidy across both cohorts.	124
Figure 4.19	(A) Fraction of CN segments with absolute copy number ≥ 3 in early and late- stage cohorts. Kolmogorov-Smirnov test. (B) Fraction of CN segments with absolute copy number ≥ 3 in three clusters	124
Figure 4.20	Copy number change point. (A)Graphical depiction of CN change point; (B) distribution of copy number change point $\geq +2$ in the three clusters. Kruskal Wallis test, $p < 0.0001$; (C) fraction changepoint $\geq +2$ density distribution. The three colours represent the three clusters.	125
Figure 4.21	Overall survival of combined early and late-stage cohorts by ploidy. Log-rank for trend analysis.	125
Figure 4.22	Overall survival of late-stage cohort by ploidy.	126
Figure 5.1	(a)-(e). Five representative pathological figures of precursors to high-grade serous carcinoma by staining with p53, Ki67 and H&E. In each patient, the p53 signatures had p53 overexpression, normal morphology, but without high proliferative activity (ie low percentage of Ki67 expression)	137-141
Figure 5.2	(a)-(e) Representative p53, Ki67 and Haematoxylin and Eosin images of five patients with p53 signature, STIC, and primary, metastasis sites.	144-148
Figure 5.3	Laser capture microdissection process for collecting tissue.	149
Figure 5.4	Workflow of sample isolation and next-generation sequencing analyses.	150
Figure 5.5	(a) Distribution of unique and duplicated reads of sWGS sequencing in STIC cohort samples. (b) Per sequence quality score shows the frequencies of quality scores in all sample. The quality of reads is good quality with the peak shifts to the right side. (c) Per base sequence quality presents the	153-154

quality score across all bases at each position in the reads. (d) Adapter content shows a high percentage of adapter contamination in these samples.

Figure 5.6	(a) Distribution of unique and duplicated reads of sWGS sequencing in STIC cohort samples after trimming. (b) Per sequence quality score shows the frequencies of quality scores in all samples after trimming. The quality of reads is good quality with the peak shifts to the right side. (c) Per base sequence quality presents the quality score across all bases at each position in the reads. (d) No samples were found with any adapter contamination >0.1%.	155
Figure 5.7	Represents adapter content distribution before and after trimming in one sample.	156
Figure 5.8	Comparison of copy number profiles of STIC project samples using different bin sizes (30kb, 100kb, 500kb, 1000kb).	159
Figure 5.9	Absolute copy number profile of a representative STIC sample.	159
Figure 5.10	Focal copy number peaks specified for STIC project samples. Each peak's gain (red) and loss (blue) are shown.	161
Figure 5.11	Genomic aberration plot of CNA across all anatomic sites and patients in five patients. Images were analysed using GISTIC_2.0 (v7) beta.	162
Figure 5.12	Ploidy distribution in STIC project.	163
Figure 6.1	Phylogenetic and clonal patterns figure in the tumour evolution process.	172
Appendix 7	Tapestation results of Ampliseq libraries preparation from early stage samples.	220-222
Appendix 8	Absolute copy number profiles from swgs pipeline in early stage cohorts	223
Appendix 9	Absolute copy number profiles from swgs pipeline in late stage cohorts.	224-227
Appendix 11	Absolute copy number profiles in all anatomic sites in five patients	230-231

CHAPTER 1. Introduction

1.1.General Introduction to ovarian cancer

Ovarian cancer occurs in only three percent of cancers in women, but it is the most lethal gynaecological malignancy and accounts for 239,000 new cases and 152,000 deaths worldwide annually^[2, 3]. The majority of cases are found in women older than 50 years, but the diagnosis age can vary, with cases even occurring in infancy^[4]. Despite advances in surgery, chemotherapy and imaging, ovarian cancer mortality has changed little over several decades^[5]. The reason for the poor prognosis is that the majority of patients present with advanced disease (stage IIIC or IV), where treatment is rarely curative; the 5-year survival for late stage patients is less than 30%, even with the addition of maintenance therapy with bevacizumab and PARP inhibitors following surgery and platinum-based chemotherapy^[6].

1.2 Epidemiology of ovarian cancer

The incidence of ovarian cancer presents wide geographic variation^[7]. According to age-adjusted incidence rates, the highest rates are in North America, Central and Eastern Europe, with a frequency of more than 8/100,000 women. Intermediate rates are seen in South America (5.8/100,000), while Asia and Africa have the lowest incidence rates (less than 3/100,000)^[8]. Some non-genetic/environmental factors were underscored when migration from countries with low rates to high rates results in greater risk^[9].

1.3 Staging

Surgical debulking is the most important step in the treatment of all types of ovarian cancer with the aim of removing all visible tumour masses. In addition to the therapeutic effect of debulking, surgery is of fundamental importance, allowing a careful staging of the disease according to the TNM (Tumour-Node-Metastasis) or the FIGO (International Federation of Gynecology and Obstetrics) classification^[10]. The new 2014 version of FIGO classification is shown below (Table 1.1)

STAGE I: Tumour confined to ovaries					
OLD			NEW		
IA	Tumour limited to 1 ovary, capsule intact, no tumour on surface, negative washings/ascites.		IA	Tumour limited to 1 ovary, capsule intact, no tumour on surface, negative washings.	
IB	Tumour involves both ovaries otherwise like IA.		IB	Tumour involves both ovaries otherwise like IA.	
IC	Tumour involves 1 or both ovaries with any of the following: capsule rupture, tumour on surface, positive washings/ascites.		IC Tumour limited to 1 or both ovaries		
			IC1	Surgical spill	
			IC2	Capsule rupture before surgery or tumour on ovarian surface.	
			IC3	Malignant cells in the ascites or peritoneal washings.	
STAGE II: Tumour involves 1 or both ovaries with pelvic extension (below the pelvic brim) or primary peritoneal cancer					
OLD			NEW		
IIA	Extension and/or implant on uterus and/or Fallopian tubes		IIA	Extension and/or implant on uterus and/or Fallopian tubes	
IIB	Extension to other pelvic intraperitoneal tissues		IIB	Extension to other pelvic intraperitoneal tissues	
IIC	IIA or IIB with positive washings/ascites.				
STAGE III: Tumour involves 1 or both ovaries with cytologically or histologically confirmed spread to the peritoneum outside the pelvis and/or metastasis to the retroperitoneal lymph nodes					
OLD			NEW		
IIIA	Microscopic metastasis beyond the pelvis.		IIIA (Positive retroperitoneal lymph nodes and /or microscopic metastasis beyond the pelvis)		
			IIIA1	Positive retroperitoneal lymph nodes only	
			IIIA1(i)	Metastasis ≤ 10 mm	

				IIIA1(ii)	Metastasis > 10 mm
			IIIA2	Microscopic, extrapelvic (above the brim) peritoneal involvement ± positive retroperitoneal lymph nodes	
IIIB	Macroscopic, extrapelvic, peritoneal metastasis ≤ 2 cm in greatest dimension.		IIIB	Macroscopic, extrapelvic, peritoneal metastasis ≤ 2 cm ± positive retroperitoneal lymph nodes. Includes extension to capsule of liver/spleen.	
IIIC	Macroscopic, extrapelvic, peritoneal metastasis > 2 cm in greatest dimension and/or regional lymph node metastasis.		IIIC	Macroscopic, extrapelvic, peritoneal metastasis > 2 cm ± positive retroperitoneal lymph nodes. Includes extension to capsule of liver/spleen.	
STAGE IV: Distant metastasis excluding peritoneal metastasis					
OLD			NEW		
IV	Distant metastasis excluding peritoneal metastasis. Includes hepatic parenchymal metastasis.		IVA	Pleural effusion with positive cytology	
			IVB	Hepatic and/or splenic parenchymal metastasis, metastasis to extra- abdominal organs (including inguinal lymph nodes and lymph nodes outside of the abdominal cavity)	

Table.1.1 FIGO Ovarian Cancer Staging

(https://www.sgo.org/wp-content/uploads/2012/09/FIGO-Ovarian-Cancer-Staging_1.10.14.pdf).

1.4 Grading

Grading describes how abnormal tumour cells appear morphologically and is an indicator of how quickly cancer may grow and spread. Usually, the higher the grade, the faster the tumour grows. Grade 1 or low grade (well-differentiated) tumours tend to grow slowly and are less likely to spread, whilst grade 3 or high grade (poorly-differentiated) refers to the cancer cells that are highly atypical. Grade 3 tumours usually grow quicker and are more likely to spread^[11].

1.5 Treatment of ovarian cancer

In addition to surgical debulking, standard treatment for ovarian cancer includes platinum-based chemotherapy^[12]. As stated above, the aim of surgery for epithelial ovarian cancer is to remove all visible disease, with total abdominal hysterectomy (TAH), bilateral salpingo-oophorectomy (BSO) and omentectomy being essential. In early stage, low grade disease in pre-menopausal women, unilateral salpingo-oophorectomy may be possible to preserve fertility^[13]. However, in late stage patients, evidence clearly indicates that removal of all visible disease (so called complete debulking) is associated with improved outcome compared to those with visible residual disease. In those who are left with residual disease, optimal debulking (<1cm residual) has significantly better survival than sub-optimal (>1cm residual)^[14, 15]. Patients with epithelial ovarian cancer should have germline and somatic testing for *BRCA1/2* as a minimum. This testing is important for prognosis, guidance for future treatment and in order to identify family members at increased risk of breast and epithelial ovarian cancer^[14].

Apart from debulking surgery, adjuvant chemotherapy also plays an important role in treating epithelial ovarian cancer patients. For early stage epithelial ovarian cancer, adjuvant platinum-based chemotherapy has been studied specifically in four randomised trials. The two largest studies were ICON1 and ACTION. The ICON1 study investigated recurrence-free and overall survival (RFS and OS) for women with early-stage epithelial ovarian cancer (ES EOC) who were treated with adjuvant chemotherapy compared to no adjuvant chemotherapy^[16]. In the ACTION study, 448 patients with early epithelial ovarian cancer were randomly assigned into two groups after surgery. One group received adjuvant chemotherapy and another had no further treatment (observation)^[17]. Overall, patients who received adjuvant chemotherapy showed better overall survival (OS) (HR 0.71; 95% CI 0.53

to 0.93) and progression-free survival (PFS) (HR 0.67; 95% CI 0.53 to 0.84) than those who did not. In addition, a meta-analysis that included 772 patients from ICON1 and ACTION provided evidence of overall benefit for the use of adjuvant chemotherapy in early-stage epithelial ovarian cancer, especially in those who had sub-optimal staging^[18]. Long term follow up data from ICON1 suggested that adjuvant chemotherapy was of greatest benefit in those with high risk of relapse (defined as stage IB/IC (FIGO 2008 criteria) with grade 2, any stage I of high grade/grade 3 or clear cell pathology)^[16, 19].

For advanced stage epithelial ovarian cancer, the key first-line chemotherapeutic agent is platinum-based (cisplatin or carboplatin) along with a taxane, either paclitaxel or docetaxel. The evidence for this combination comes from a series of historical randomised trials. ICON2 showed that single agent carboplatin was as effective as the triple combination of cisplatin, doxorubicin and cyclophosphamide and was substantially less toxic^[20], whilst the GOG-111 study was the first to show that addition of a taxane to platinum improved outcome - it showed that patients treated with cisplatin and paclitaxel had improved overall survival compared to those treated with cisplatin and cyclophosphamide^[21]. The GOG-158 trial confirmed that carboplatin/paclitaxel was less toxic, easier to administer and as effective as cisplatin/paclitaxel and has now become the international standard of care^[22]. Weekly dose-dense carboplatin and paclitaxel does not appear to provide any additional benefit over standard three-weekly chemotherapy in Caucasian populations^[23] but does appear to do so in Japanese patients^[24]. The addition of a third agent or a more extended period of the chemotherapy cycle is not beneficial^[23, 25].

More recently, ICON7^[26] and GOG-218^[27] demonstrated that the addition of bevacizumab to carboplatin/paclitaxel extends progression-free survival and may improve overall survival in

high risk populations^[6]. Finally, there is strong evidence that maintenance therapy with a PARP inhibitor extends progression-free survival, especially in patients with germline or somatic mutations in *BRCA1/2*^[28] or whose tumours are classified as having defective homologous recombination^[29].

1.6 Ovarian cancer classification

The classification of epithelial ovarian cancer is shown in Table 1.1. Epithelial ovarian carcinomas are classified according to their histological subtype and grade. According to clinic pathological and molecular genetic studies, emerging evidence shows that the five major ovarian carcinoma subtypes (high-grade serous, clear cell, endometrioid, mucinous and low-grade serous) are different. They have different genomic drivers and may all have separate cells of origin^[30].

	High grade serous carcinoma	Clear cell carcinoma	Endometrioid carcinoma	Low grade serous carcinoma	Mucinous carcinoma
% of all ovarian carcinomas	~70%	~10%	~10%	<5%	<5%
Origin	Fallopian tube epithelium	Endometriosis	Endometriosis	Fallopian tube epithelium	?Unknown
Inherited syndromes	BRCA1/2, hereditary breast and	Lynch syndrome	Lynch syndrome	?	?

	ovarian				
Common mutations and molecular aberrations	TP53, BRCA1/2 and HRD, Chromosome instability Aneuploidy	ARID1A, PIK3CA, CTNNB1, PPP2R1A, MSI	PTEN, CTNNB1, ARID1A, PPP2R1A, MSI	KRAS, BRAF	KRAS, HER2 amplification
Potential biomarker	PARP inhibitors, Immune checkpoint inhibitors	Tyrosine kinase inhibitors	mTOR inhibitors	MEK1/2 inhibitors	Trastuzumab

Table 1.1 Classification of ovarian cancer. Modified from Figure 1: Five major types of EOC (<http://www.bccancer.bc.ca/books/ovary-epithelial-carcinoma/histological-classification-of-ovarian-carcinoma>)

1.6.1 Endometrioid ovarian carcinoma

Endometrioid ovarian carcinoma (EOVC) is an uncommon subtype of epithelial ovarian cancer, accounting for 9-11% in all cases^[31], but shares many features with endometrioid endometrial carcinoma. Certainly by immunohistochemistry, both tend to express vimentin, ER and PR, and PAX-8 and CA125^[32]. Traditionally classified according to grade (Grades 1, 2 and 3), whole exome sequencing of 112 cases recently identified that the most frequently mutated genes in endometrioid ovarian carcinoma are *CTNNB1* (43%), *PIK3CA* (43%), *ARID1A* (36%), *PTEN* (29%), *KRAS* (26%), *TP53* (26%) and *SOX8* (19%) and there appear to be three broad categories^[33]- *TP53* mutated, which shares many features with high grade serous carcinoma (see section 1.4.5), *TP53* wild-type/*CTNNB1*-mutated and *TP53* wild-

type/*CTNNB1*-wild-type. Prognosis is best in the *TP53* wild-type/*CTNNB1*-mutated group and worst in the *TP53* mutated group. The *TP53* wild-type/*CTNNB1*-mutated group is marked by low genomic complexity and is almost exclusively found at early stage. The *TP53* wild-type/*CTNNB1*-wild-type group has moderate complexity and is usually, but not always found at early stage.

1.6.2 Clear cell carcinoma

Clear cell carcinoma has distinct biology and clinical characteristics and frequently arises on the background of endometriosis. Clear cell carcinoma tends to be diagnosed in younger women than high grade serous carcinoma. In addition, approximately 60% of cases are diagnosed at stage I-II, in stark contrast to high grade serous carcinoma^[34]. Many sequencing analyses have been performed on clear cell carcinoma and identified that the most frequently mutated genes are *ARID1A* and *PIK3CA*, which are frequently co-mutated^[35].

1.6.3 Mucinous carcinoma

Ovarian mucinous carcinoma accounts for around 3-5% of all epithelial ovarian cancers after revisions to the diagnostic criteria^[36]. Mucinous tumours may show intestinal-type or, less commonly, endocervical-type morphology^[37]. The cell origin of mucinous tumours remains unclear but there does appear to be progression from benign to borderline to carcinoma. Around 65-80% mucinous ovarian cancers are diagnosed at an early stage and have much better overall prognosis for women than other subtypes of epithelial ovarian cancer^[38]. Mucinous ovarian cancers have been classified into six therapeutically relevant subgroups according to the targeted next-sequencing^[39]. Only about 16-52% of mucinous carcinoma was reported to harbour *TP53* mutation and significantly more have *KRAS* mutations (40-

50%)^[40]. Approximately 18-20% mucinous adenocarcinoma have HER2 amplification. Trastuzumab may be a treatment option for patients with HER2 overexpression^[32].

1.6.4 Low grade serous carcinoma

Low grade serous carcinoma is characterised by less aggressive behaviour, less chemotherapy sensitivity and longer survival than high grade serous carcinoma. It accounts for 2% of all epithelial ovarian cancers and 4.7% serous carcinoma^[41]. LGSC is rarely associated with *TP53* or *BRCA1/2* mutations and family history^[42]. Oestrogen and progesterone receptors (ER and PR) are frequently expressed in LGSC, whilst mutations in *KRAS*, *BRAF* and genes within the MAPK pathways are frequent^[43].

1.6.5 High-grade serous carcinoma

Eighty percent of all epithelial ovarian cancers are of the serous subtype, more than 90% of which are classified as high-grade utilising the two-tier WHO classification system^[44]. Unlike low grade serous tumours, HGSC has almost universal *TP53* mutation. By immunohistochemistry, most HGSC exhibit strong Wilm's tumour 1 (WT1) staining (52.2%)^[45]. Other markers include p16^[46] and paired-box gene 8 (PAX8), which is expressed in 79-100% cases of HGSC^[47].

Most HGSC cases are diagnosed at a late stage^[48], with only approximately 10% identified at early stage disease. Early stage HGSC is usually asymptomatic, and no effective screening or early detection methods exist^[49]. Early stage patients have a good prognosis (70 - 80% overall survival at ten years^[16] compared to <30% for advanced disease), but it is unclear whether early stage HGSC is diagnosed purely by chance or has different characteristics preventing rapid dissemination into the peritoneal cavity. This question forms a large basis of

this thesis.

1.7 Biology of HGSC

Despite extensive studies over the past few years^[50, 51], there is still a need to identify new key players with therapeutic potential in HGSC. Mutations and genomic instability are key players which considered the in cancer, they could promote the accumulation of malignant cells and contributing to cancers' ability to evade the immune response. Here are the introduction of some genes and copy number alterations in the development of HGSC.

1.7.1 Frequently mutated genes

The collective somatic mutations observed in cancer result from multiple mutagenic processes operating over a patient's lifetime. Each process leaves a characteristic imprint – a mutational signature on the cancer genome, defined by the type of DNA damage and DNA repair processes that result in base substitutions, insertions and deletions or structural variations^[52], thus potentially revealing the clonal evolution process in tumours. Some genes are frequently mutated in high grade serous carcinoma. Resolving such gene mutation rates could reveal tumour heterogeneity depending on accurately estimating fractions of alleles that harbour somatic mutations. Advances in deep sequencing of tumour cohorts have provided critical insights into tumour aetiology^[53] and unveiled unprecedented degrees of clonal heterogeneity within a single tumour DNA sample.

1.7.1.2 *TP53* mutation

p53 is a well-established studied tumour suppressor gene and plays important roles in the initiation of tumour development. *TP53* is located on chromosome 17p13.1^[54] and consists of 11 exons. p53 protein contains five functional domains: 1. Two N-terminal

transcriptional activation domains (TAD, residues 1-42 and 43-62), 2. A proline-rich domain (PRD, residues 64-92), 3. Central sequence-specific DNA binding domain (DBD, residues 102-292), 4. Oligomerization domain (OD, also known as tetramerization domain; residues 323-356), and 5. C-terminal regulatory domain (residues 363-393). The lollipop figure of p53 protein is shown in Figure 1.2.

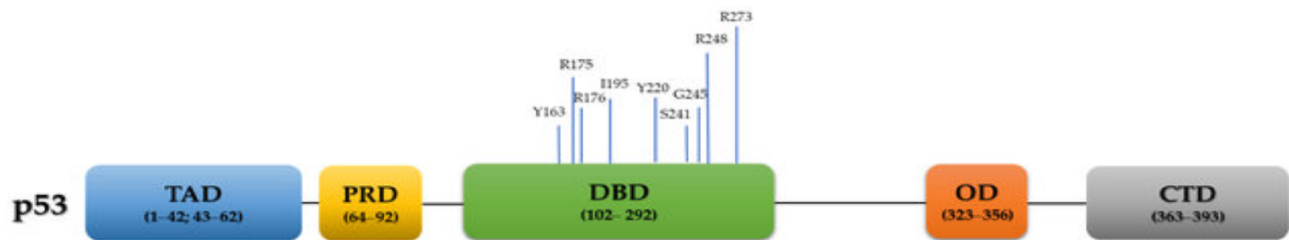


Figure.1.1 p53 domains and hotspot mutations

Representation of the domain structure of p53. The nine most commonly mutated residues in *TP53* in HGSC, based on TCGA data, are highlighted. Key: TAD: transactivation domain; PRD: proline-rich domain; DBD: DNA binding domain; OD: oligomerization domain; CTD: C-terminal domain^[55]

In normal conditions, wild-type p53 has a short half-life and is expressed at low levels in the cell, regulated by its E3 ubiquitin ligase MDM2^[56]. In stress conditions, p53 is stabilised and becomes active in regulating downstream pathways, such as Wnt-beta-catenin, IGF-1-AKT, Rb-E2F, p38 MAP kinase, cyclin-cdk, p14/19 ARF pathways and the cyclin G-PP2A, and p73 gene products^[57]. The most common sites of *TP53* mutations are located in the DNA binding domain (DBD), presenting around 80% of *TP53* mutations^[58]. The mutation types in epithelial ovarian cancers of *TP53* are missense, nonsense, frameshift, in-frame insertions or deletions, and splice site^[55]. The most common mutation type in *TP53* is missense, which accounts for approximately 60% of all *TP53* mutations in HGSC^[59].

The different mutation types may lead to various functional consequences of p53 mutations^[59]. For example, missense mutations tend to produce inactive full-length p53

protein with extended half-life that accumulates within the nucleus. In contrast, frameshift mutations do not typically lead to p53 accumulation, whilst nonsense mutations usually produce a truncated unstable protein^[60]. These different patterns of protein expression can be readily detected by immunohistochemistry, which provides a good surrogate for TP53 mutation status^[61]

1.7.1.3 BRCA1/2 mutations

Apart from *TP53* mutation, which is near-universal^[62], there are few recurrent classical activating oncogenic mutations in HGSC. Currently, clinical classification is limited to defining homologous recombination-deficient (HRD) tumours, usually arising from germline *BRCA1* or *BRCA2* mutation. Germline mutations in *BRCA1* or *BRCA2* confer a well-established increased lifetime risk for developing high-grade serous carcinoma^[63] and are observed in approximately 15% HGSC cases. Somatic mutations are observed in approximately 5% cases. *BRCA1* mutations, both germline and somatic, are typically twice as common as *BRCA2*. *BRCA1* is a known modulator of the cell cycle at multiple points, including the S-phase checkpoint, the G2-M checkpoint, the spindle checkpoint and centrosome duplication through co-activation of p53^[64]. It also regulates the microRNA 155 expression and maintains heterochromatin structure via deacetylation of histones H2A and H3^[65]. Dysfunction of *BRCA1* occurs in patients with HGSC through germline mutations or promoter hyper-methylation and is associated with earlier onset of HGSC^[66]. *BRCA2* mutations are less frequent than *BRCA1*, and are not associated with an earlier age at diagnosis. However, long term data, albeit from prior to the introduction of PARP inhibitor therapy, suggest that overall prognosis is better for those with *BRCA2* mutations than *BRCA1*^[67]. Overall, evaluating the mutation status of *BRCA1/2* in HGSC patients is of great importance, as this can provide information for the treatment for HGSC as well as risk in

other family members where a germline mutation is identified.

Beyond *BRCA1/2* mutations, it is believed that a further 30% HGSC cases demonstrate defective HR, also known as “BRCAness”^[68]. Defective HR can be assessed in HGSC genomes using assays that assess genome-wide LOH, large-scale state transitions and telomere imbalance^[69]. Patients with HRD are more likely to benefit from treatments affecting DNA repair pathways, especially platinum and PARP inhibitors^[70].

1.7.1.4 Other recurrent mutations

Whole genome sequencing has shown frequent loss of *RB1*, *NF1*, *PTEN* and amplification of *MYC*, *MECOM* and *CCNE1*^[71], as well as enrichment of amplification-associated fold-back inversions in non-HRD tumours^[72]. *RB1* is a tumour suppressor^[73] that belongs to the RB protein family and plays a vital role in regulating the advancement of the cell cycle from the G1-S phase through E2F and cyclin-dependent kinases^[74]. *NF1* is also a well-characterised tumour suppressor and encodes a protein of 2818 amino acids that regulates RAS/MAPK signalling^[75]. This large protein has GAP activities via its GTPase-activating protein domain^[76]. Qiao et al. identified that *NF1* expression levels were significantly different in different FIGO stages of epithelial ovarian cancer and suggested that *NF1* loss promotes epithelial ovarian cancer development^[77]. *PTEN* is also a tumour suppressor known for its role in inhibiting the PI3K pathway^[78]. *PTEN* loss appears in pre-invasive lesions in the fallopian tube and could be a putative early driver event in HGSC^[79].

Amplification of Cyclin E1 is an oncogenic driver in a subset of HGSCs. Cyclin E1 binds to cyclin-dependent kinase 2 (CDK2) to drive G1/S cell cycle progression^[80]. *CCNE1* amplification, which is associated with poor prognosis and platinum resistance, usually happens in the early stage of the tumour process and precedes centrosome duplication^[80]

and is mutually exclusive with *BRCA1* mutation.

1.7.2 Copy number abnormalities in HGSC

Normal cells are diploid, with paired chromosomes, one from each parent. Other than human sex cells, all human cells have 22 pairs of autosomes and one pair of sex chromosomes (X and Y)^[81]. Copy number alterations comprise deletions or amplification of genomic fragments.

Sami Amr et al. utilised allele-specific droplet digital PCR (ddPCR) with next-generation sequencing to quantify absolute copy numbers from patients^[82]. As shown in Figure 1.3, the copy number above 2.5 (red dots) in the chromosome defines copy number gains. Also, heterozygous deletion refers to the loss of one of the alleles at a specific position. Furthermore, homozygous deletion is the loss of both alleles identified by the allele-specific analysis.

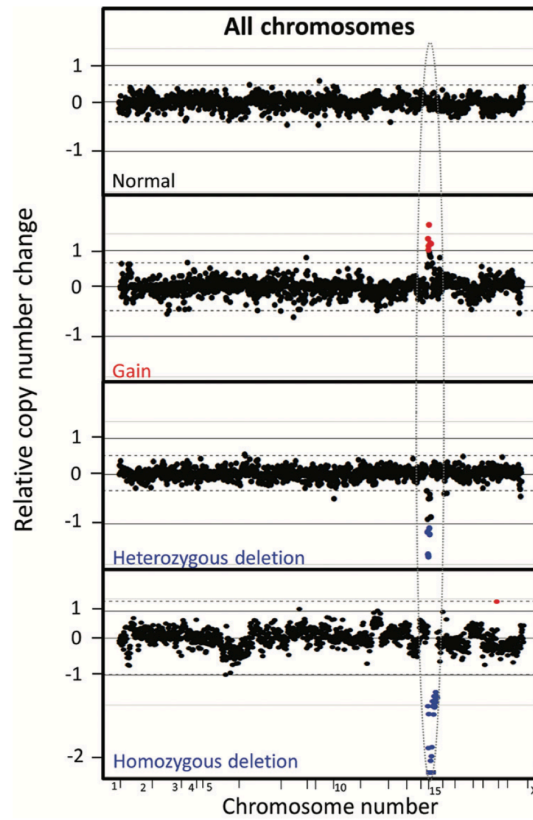


Figure 1.2 Relative copy number profile across the chromosome.

Filled circles represent relative copy number changes, calculated from the $\log_2$ ratio of the fractional coverage value for each banded region divided by the median for the region across the batch. Blue circles represent deletions, while red circles represent copy number gains. This figure comes from modified Figure.2 in Amr et al. (2018)^[82].

Ciriello et al. integrated genomic data from twelve cancer types with 3299 tumour samples from TCGA database. Combining genomic, epigenetic changes, copy number alterations, somatic mutations, and gene DNA methylation events in each cancer study, they identified two main tumour classes using hierarchical classification (Figure 1.4)^[83]. The two-class characteristic events were predominantly somatic mutations (M class) and copy number alterations (C class), and tumours with high frequency of *TP53* mutations and multiple recurrent chromosomal amplifications and deletions were named C class tumours. The tumour type with the highest frequency of C class tumours was ovarian high grade serous carcinoma, indicating that HGSC is marked by extreme copy number alteration and genomic instability.

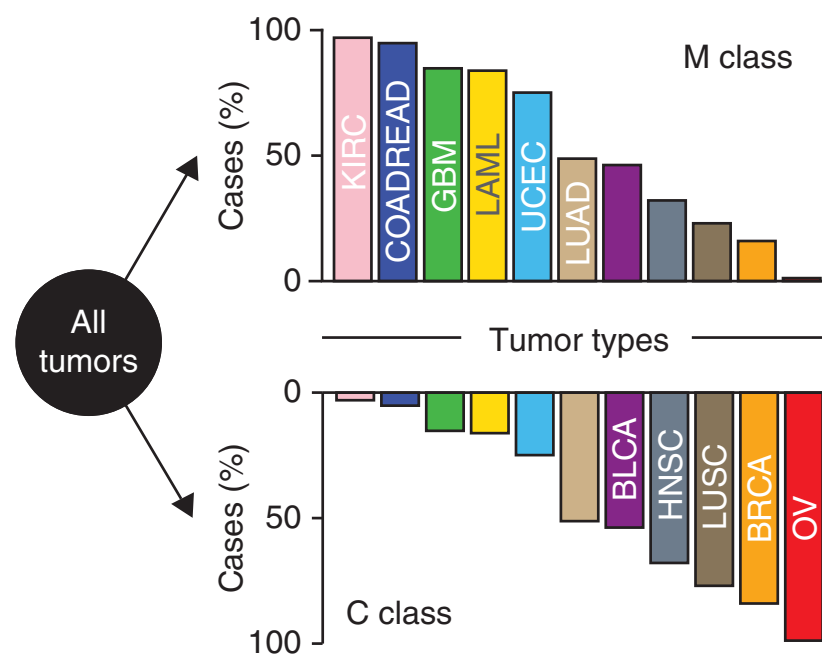


Figure 1.3 Pan-cancer cancer classification

Analysis of pan-cancer data from TCGA identified two main classes of cancers. There were characterised primarily by either recurrent mutations (M class) or recurrent copy number alterations (C class). Each class contains multiple tumour types in different proportions. This figure is from Figure 2a in ^[83].

1.7.2.1 Copy Number Landscape of HGSC

Copy number alterations are very frequent in HGSC. Huang et al. collected 101 serous carcinoma samples from three datasets^[84]. The copy number profile of these samples is shown in Figure 1.5. There are gains and losses throughout the genome and the most common copy number alterations observed were at chromosome arm level. Broadly 3p amplification, 8p deletion, 8q amplifications, 17p deletions and chr20 amplifications were the top alterations^[85].

Some candidate driver genes within the altered regions might contribute to the development of serous carcinoma. There are recurrent focal CNA, including *MYC*, *MECOM*, *CCNE1* and *PIK3CA*. *MYC* is located at chromosome 8q11.21-8q24.3. *MYC* amplification is thought to regulate cellular proliferation, differentiation and apoptosis^[86]. *MECOM* is located at 3q26, which is thought to contribute to the pathophysiology of epithelial ovarian cancers^[87]. Cyclin E1 (*CCNE1*) amplification is associated with primary treatment resistance and poor outcome in HGSC^[88]. Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha (*PIK3CA*) and Eukaryotic Translation Initiation Factor 4A2 (*EIF4A2*) are two important genes that are located in chromosome 3q26.31-3q29. *PIK3CA* amplification was closely associated with increased phosphorylated AKT(p-AKT) level expression and significantly related to cancer-related death^[89]. *EIF4A2* is an ATP-dependent RNA helicase involved in the mRNA translation initiation process. Abnormal expression of *eIF4A2* has been reported as a prognostic factor in various cancer types^[90]. In chromosome 17, Neurofibromatosis type 1 (*NF1*) loss will cause higher RAS pathway activity as *NF1* is a regulatory inhibitor of RAS^[91]. *SS18L1*, which is on chromosome 20, is significantly associated with metastasis^[92].

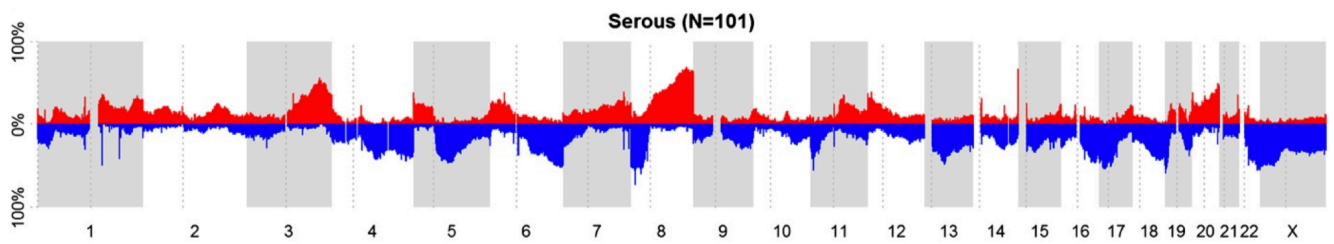


Figure 1.4 Overview of copy number aberrations in 101 high grade serous carcinoma cases

Chromosomes are shown in the alternating grey blocks. The centromere for each chromosome is shown as a dotted line. Y-axis shows the frequency (%) of occurrence of amplification (red regions) and deletion (blue regions) from chromosome 1 to X in serous carcinoma from three datasets. This figure is modified from Ruby Yun-Ju Huang et al. ^[84].

1.7.2.2 Copy number signatures

Macintyre et al. used shallow whole genome sequencing technology to extend the mutational analysis to copy number alterations. They demonstrated seven different copy number (CN) signatures in HGSC genomes^[1]. These signatures were developed using samples from the BriTROC-1 study^[93] and validated on separate cohorts from PanCancer Analysis of Whole Genomes (PCAWG) and The Cancer Genome Atlas (TCGA). More details about how they developed the seven copy number signatures will be presented in Chapter 4. Among the seven copy number (CN) signatures, CN signature 1 correlated with amplification-associated fold-back inversion structural variants and poor overall survival in HGSC. CN signature 1 was also highly enriched in RAS signalling and telomere shortening events. CN signature 2 was strongly correlated with frequent breakpoints per 10 megabases (Mb), copy number single change associated with tandem duplication phenotype. *CDK12* mutation has previously been correlated with tandem duplications (TD) of up to 10Mb^[94], and CN signature 2 was enriched in *CDK12* mutated cases. CN signature 3 represented chromosome break distribution and loss of heterozygosity (LOH) process and was enriched in *BRCA1/BRCA2*-related homologous recombination deficiency (HRD) patients. CN signature 4 was characterised by high copy number and associated with abnormalities in the PI3K-

AKT signalling pathway. CN signature 5 was related to chromothripsis-like events, but there were no significantly associated mutational events. CN signature 6 was associated with genes that affect diverse pathways, including G1/S cell cycle genes and PI3K-AKT signalling genes. Like signature 3, CN signature 7 was also correlated with chromosome break distribution; however, there was no enrichment in *BRCA1* or *BRCA2* mutations, indicating mutational processes with and without the HRD pathway. The summary of the seven copy number signatures is shown in Figure 1.6. The application of CN signatures to epithelial ovarian cancer is novel and significant as previous studies had concentrated on base substitution and rearrangement signatures in revealing cancer classification^[95].

One shortcoming of the original CN signature analysis was that samples were obtained from women with advanced and/or recurrent disease. Therefore, it is not clear when distinct CN signatures emerge during the process of carcinogenesis and how intra-patient heterogeneity is reflected in CN signatures. Chien et al.^[96] undertook WGS on 16 low-stage HGSC and compared their results to whole exome sequencing data from the TCGA cohort of late stage HGSC. They found that early stage and late stage HGSC have a similar mutation and copy number profiles. They also found that deleterious *TP53* mutations were near-universal in early stage HGSC and occurred early in the development process, followed by deletion or loss of heterozygosity (LOH). Inactivation of homologous recombination (HR) was also frequent in early stage HGSC; 8/16 patients had mutations in *BRCA1* or *BRCA2*, and one tumour had aberrant *RAD51C*. This evidence suggested that early stage HGSC has some unique features that differ from late stage HGSC. This will be explored further in Chapters 3

and 4.

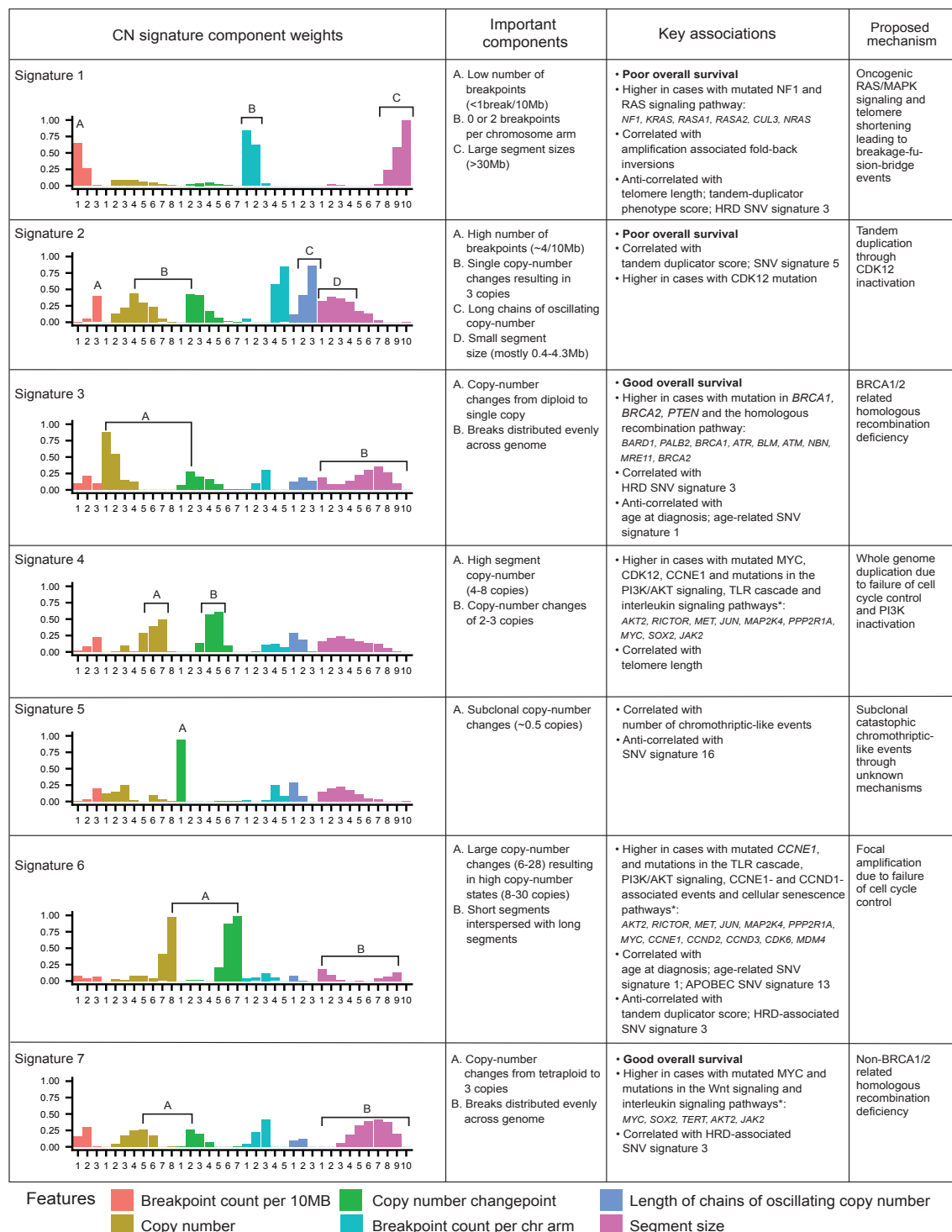


Figure 1.5 The features of the seven copy number signatures in HGSC.

Each copy number signature is described by important components, key association and proposed mechanism^[5]. This figure is from Macintyre et al.^[1]

1.8 HGSC development hypothesis

All cancers are thought to originate from a single cell that behaves abnormally due to the acquired somatic mutations in its genome. The ability of cancer to develop and adapt following treatment can be described in terms of tumour evolution and stems from its intrinsic diversity or heterogeneity. Advances in sequencing technologies and the development of novel mathematical algorithms have allowed the patterns of tumour heterogeneity caused by different mutational processes to be deciphered^[97]. Therefore, many researchers have focussed on tumour heterogeneity of late stage HGSC. Unlike other cancers, such as colon, breast, prostate and pancreas, for which precursor lesions have been identified, the origin lesion of high grade serous carcinoma remained unclear until recently. The current theory suggests HGSC arises in the fallopian tube, starting as focal non-invasive cytological lesions called “p53 signatures”. These signatures contain phenotypically normal secretory cells in the fallopian tube epithelium that are marked either by overexpression or absence of p53 protein on immunohistochemistry (IHC) analysis. p53 signatures progress to lesions named serous tubal intraepithelial lesion (STIL), in which cellular atypia emerges. During this process, the p53 signatures undergo morphologic changes and more somatic mutations, methylation change and telomere shortening^[98]. STIL is differentiated from invasive STIC (serous tubal intraepithelial carcinoma) by the extent of proliferation^[99]. STICs are characterised by nuclear enlargement, loss of polarity, nuclear atypia and increased proliferative activity^[100]. At the genomic level, STIC has more somatic mutations, more LOH events, centrosome amplification, telomere stops shortening, and immune cell infiltration events that cause genomic instability and evolve into invasion HGSC^[98]. The summary process of clonal development of STICs is shown in Figure 1.7.

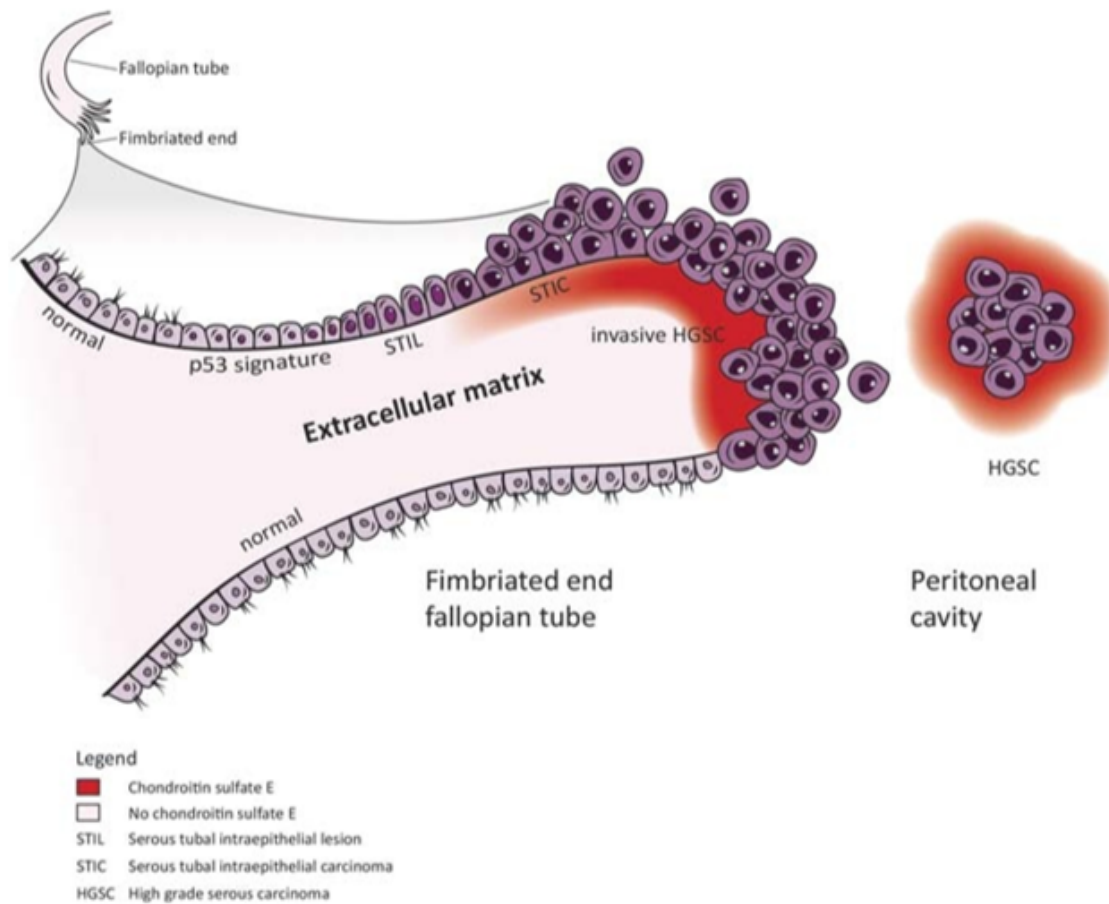


Figure 1.6 The current hypothesis for the development and evolution of HGSC.

Normal fallopian tube epithelium develops into p53 signature and STIL. STIC lesions transform into invasive carcinoma. By influencing multiple cell processes, the invasive HGSC cells contribute to the development and progression of HGSC cancer^[101]. This figure was taken from Steen et al.^[101].

1.9 Pathological identification of STIC

The challenge of investigating STIC lesions is that they are microscopic and are only identified during detailed histopathological analysis of the fimbrial end of fallopian tubes (using so-called SEE-FIM protocols). This protocol requires isolation of fimbria at the infundibulum, followed by longitudinal sectioning to allow maximal exposure, followed by extensive cross-sectioning of the remainder of the tube at 2 – 3mm intervals^[102].

Identification of changes in the epithelium requires morphology examination of these sections following H&E staining. Kuhn et al. described criteria utilising the Ki67 labelling

index as an aid to the diagnosis of STIC^[103], whilst Kito et al. developed a diagnostic algorithm of tubal epithelial lesions that combined morphological analysis (H&E staining) and immunostaining for p53 and Ki67 (Figure 1.8)^[104]. STICs are categorised as cases with either moderate to strong expression of p53 in > 75% of the cells or complete absence of p53 staining (0% cells stained) in foci of histological atypia with high Ki67 staining regions ($\geq$ 10% positively stained cells)^[105]. STIL is identified as strong expression of p53 or complete absence of p53 staining (0% cells stained) without atypia but increased Ki67 labelling index (above 10%). p53 signatures are identified by strong expression of p53 or complete absence of p53 staining (0% cells stained), no atypia and Ki67 labelling index less than 10%^[104]. How these morphological lesions correlate with genomic alterations in the development of HGSC will be explored in Chapter 6.

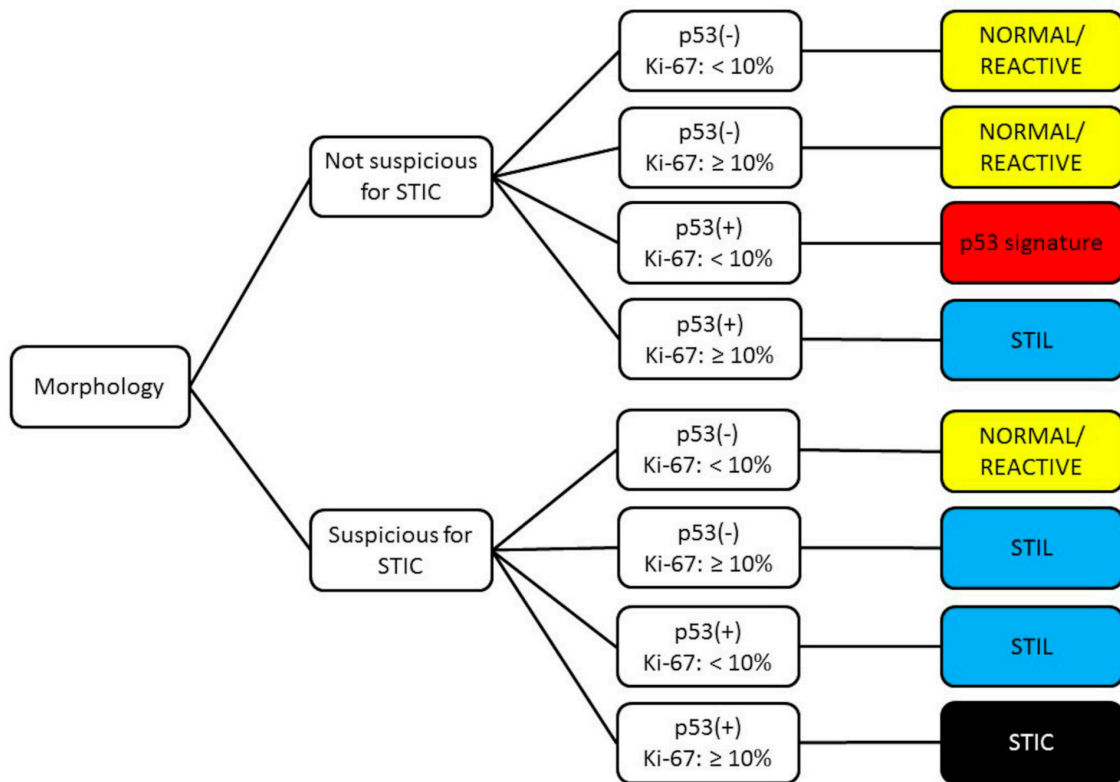


Figure 1.7. Diagnostic algorithm for STIC lesions.

p53 signatures were identified by no suspicious morphology but p53 positive (p53 staining in cancer cells with more than 10% positively^[106]) and Ki67 expression less than 10%. STIL has been identified in three conditions: no morphological atypia, p53 positive but increased Ki67 labelling index (above 10%); suspicious lesions for STIC, p53 negative with Ki67 index above 10%; suspicious lesions for STIC, p53 positive with Ki67 index less than 10%. For STICs lesions, the identification standard is suspicious for STIC in morphological, p53 positive and Ki67 index above 10%^[104]. The figure is taken from Kito et al. ^[104].

1.10 Summary and Hypothesis

I hypothesise that:

- Early stage HGSC has similar genomic characteristics to late stage HGSC and that progression from early- to late-stage disease represents time-dependent evolution.
- Copy number alterations are evident very early in the development of HGSC.

1.10.1 Aims

To investigate the genomic landscape of early stage HGSC, my aims were

- a. To collect early stage HGSC samples from tissue bank and BriTROC-1 cohort.
- b. To undertake sWGS and targeted next generation sequencing on early stage samples.
- c. To compare the genomic landscape of early stage and late stage HGSC.

To evaluate copy number changes during HGSC development, my aims were:

- a. To identify and collect p53, STIC and invasive carcinoma (ovary/fallopian tube and metastasis) samples from patients with HGSC.
- b. To undertake sWGS and targeted next generation sequencing on normal fallopian tube epithelium, p53 signatures, STIC, invasive ovarian/fallopian tube and metastatic samples from each patient.
- c. To identify copy number changes during the evolution of HGSC.
- d. To compare copy number signatures in multi-site samples from the same patient.

CHAPTER 2. Materials and Methods

2.1 Study design and workflow

This thesis focusses on two sub-projects. The early stage project compares the genomic landscape of early stage and late stage HGSC cohorts; The STIC project is about tracking the evolution of HGSC development.

The the early stage project was designed by Iain McNeish. Samples were acquired by Darren Ennis, Jacqueline McDermott, Naveena Singh, Zhao Cheng. The pathological assessment was performed by Baljeet Kaur, Naveena Singh, Chishimba Sokota and Jacqueline McDermott. Data were acquired by Zhao Cheng, Hasan Mirza, Philip Smith, Darren Ennis. Data analysis was done by Zhao Cheng, Hasan Mirza, Philip Smith, Lena Morrill Gavarro, Thomas Broadley, Anna Piskorz.

The STIC project was also designed by Iain McNeish. Samples were acquired by Darren Ennis, Jacqueline McDermott, Naveena Singh, Zhao Cheng. The pathological assessment was performed by Baljeet Kaur, Naveena Singh, Chishimba Sokota and Jacqueline McDermott. Data were acquired by Zhao Cheng, Hasan Mirza, Darren Ennis. Data are under analysis by Zhao Cheng, Bingxin Lu, Hasan Mirza.

The study design is shown in Figure 2.1

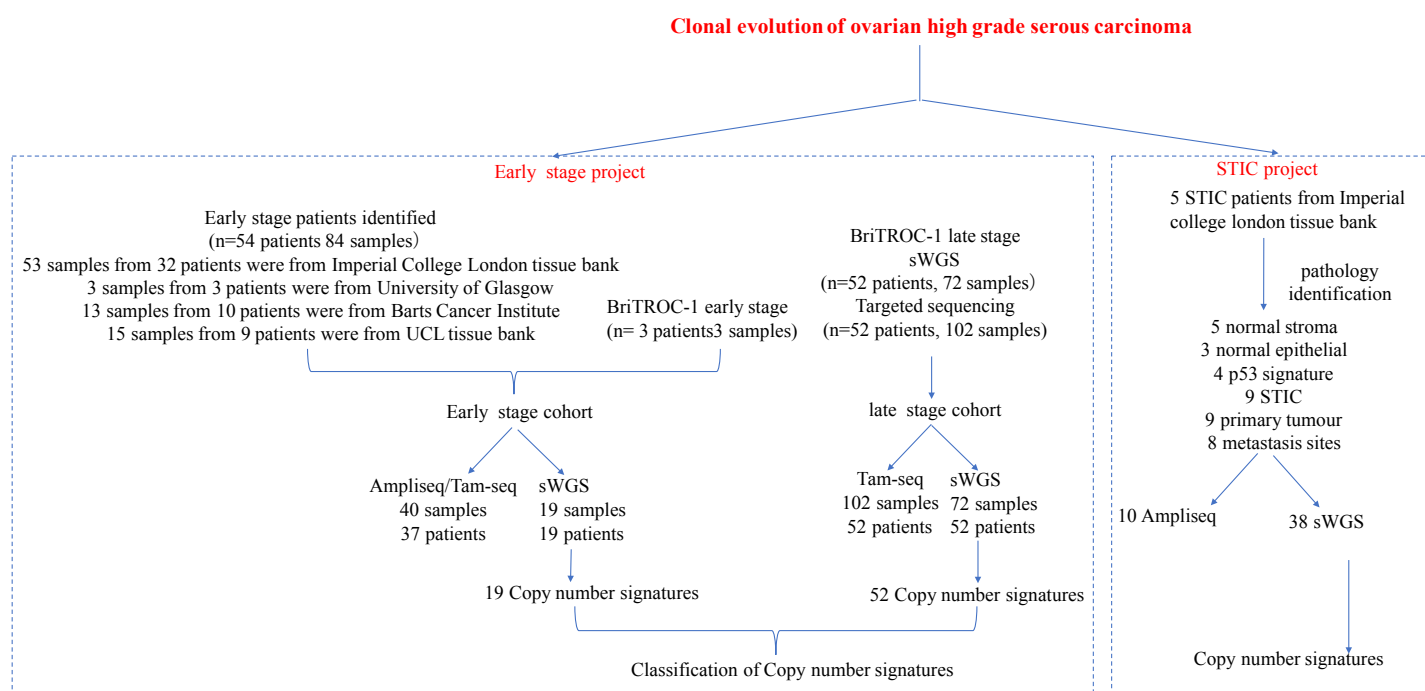


Figure 2.1 Flowchart of project study design

2.2 Clinical data and sample collection

2.2.1 Hammersmith Hospital, UCLH and Barts archival and BriTROc-1 HGSOc dataset-‘early stage cohort’1

The ethics for archival sample collection was approved by Imperial College Healthcare Tissue Bank (HTA licence 12275, Research Ethics Committee number 17/WA/0161, Project ID R18060). The samples used in this project were initially identified from the pathology archives of three large UK gynaecological cancer centres (Imperial College Healthcare, University College London and Barts Health NHS Trusts) as 54 early stage ovarian high grade serous carcinoma patients. ‘Early stage’ was defined as stage IA, IB, IC and IIA using the FIGO classification at the time of diagnosis. All the samples were identified by three pathologists (Dr Chishimba Sokota, Dr Jackie McDermott and Dr Naveena Singh), and clinical data were collected by Dr Darren Ennis, Professor Iain McNeish, Dr Naveena Singh, Dr Chishimba Sokota, Dr Jackie McDermott and Dr Baljeet Kaur. Following pathology review, 21 samples from 13 patients were excluded, whilst two samples from one patient

failed DNA extraction. Additionally, a further cohort of five early stage patients were identified from the BriTROC-1 study^[93], giving a total early stage population of 45.

2.2.2 BriTROC-1 HGSOc dataset-late-‘late stage cohort’

This dataset was selected from archival samples from the BriTROC-1 cohort. The BriTROC-1 study has been described previously^[93]. NHS Greater Glasgow and Clyde sponsored the study, and ethics approval was given by Cambridge Central Research Ethics Committee (12/EE/0349). The study enrolled patients with recurrent ovarian high-grade serous or grade 3 endometrioid carcinoma who had relapsed following at least one line of platinum-based chemotherapy and whose disease was amenable to image-guided biopsy or secondary debulking surgery. All patients provided written informed consent. DNA was initially extracted from 300 samples of 142 patients—158 methanol-fixed relapse biopsies and 142 formalin-fixed paraffin-embedded (FFPE) archival diagnostic tissues. In addition, germline DNA was extracted from blood samples of 137 patients^[93].

2.2.3 Hammersmith and St Bartholomew’s Hospitals HGSC dataset-‘STIC cohort’

The samples used in this project were selected from the pathology archives of Imperial College Healthcare and Barts Healthcare NHS Trust and utilised under the authority of Imperial College Healthcare Tissue Bank (HTA licence 12275, Research Ethics Committee number 17/WA/0161, Project ID R18060). Following review by three expert gynaecological pathologists (Dr Chishimba Sokota, Dr Jackie McDermott and Dr Naveena Singh), five patients were included for the identification of p53 signatures, STIC, primary carcinoma and metastatic tumours.

2.3 Pathology

For early stage archival samples, tumour cellularity was determined on H&E staining. Expert gynaecological pathologists kindly marked the percentage of tumour cells in areas selected for dissection. All the tissue sections per block were macro-dissected. For the STIC project, each fallopian tube (FT) containing p53 signature and STIC lesions was also identified by expert gynaecological pathologists. STIC lesions and p53 signatures were identified using Kito et al.'s algorithm^[104, 107] based on H&E, p53 and Ki67 staining performed for all FFPE blocks. First, the morphology of suspicious lesions was determined for each block followed by coordinate p53/Ki-67 profiles.

At least one STIC and/or p53 signature in each FT was identified and micro-dissected by laser capture microdissection separately. In addition, each FFPE block from STIC project patients, including early lesions, primary carcinomas and metastases sites, was stained with hematoxylin and eosin (H&E). A pathologist assessed primary carcinomas and metastases site slides for tumour cellularity, and DNA was extracted using Qiagen's FFPE DNA Extraction Kit (Qiagen, 166027818).

2.4 DNA extraction and quantification

Early stage project FFPE DNA was extracted from 10 sections (10µm thickness) of NBF-fixed/formalin-fixed tissues using Qiagen's FFPE extraction kit (Qiagen, 166027818) according to manufacturer instructions. Briefly, sections were cut onto a glass slide from an FFPE block using a microtome. The area of tumour selected by the pathologist is used as a guide for macro-dissection of tumour areas. Dr Nadia Rahman, Ms Reema Paudel and Dr Naina Patel from the Imperial ECMC lab helped with section cutting and H&E staining of early stage FFPE blocks. Macro-dissected material was scraped from the slide using a

disposable sterile scalpel and placed into a 1.5ml micro-centrifuge tube. Paraffin wax was removed using Histoclear. In brief, 1ml of Histoclear was added to a 1.5ml microcentrifuge tube with scraped macro-dissected materials. The materials were vortexed for 30s and centrifuged for 5 min at full speed using a micro-centrifuge. The supernatant was removed by pipetting, and 1ml of 100% ethanol was added to remove the rest of Histoclear. Then the micro-centrifuges were vortexed for 30 seconds and centrifuged for 5 min at full speed. Supernatants in the micro-centrifuges were removed by pipetting, and the pellets were dried for 10-15 min at 56 °C in a dry bath. The pellets were re-suspended in 180 µl Buffer ATL and mixed by vortexing, then incubated for 15 min at 95°C. After incubation, samples were allowed to reach room temperature, at which point 20 µl proteinase K was added and mixed by vortexing. The samples were incubated at 56°C for 12-24hr. Then the samples were directly incubated at 90°C for 1hr without agitation to reverse formaldehyde DNA modification. 4 µl RNase A (100mg/ml) was added, and samples were incubated for 2 min at room temperature. Next, 200 µl Buffer AL and 200 µl 100% ethanol were added to the samples and mixed by vortexing for 30s. The samples were then transferred to a QIAamp Mini Elute column (previously kept at room temperature for 30 min) in a 2 ml collection tube and centrifuged for 90 secs at 8500 rpm. The flow-through was discarded. 500 µl Buffer AW1 was added to the column and centrifuged for 90 seconds at 8500 rpm. After discarding the flow-through, 500 µl Buffer AW2 was added, and the column was centrifuged for 90 secs at 8500 rpm. Flow-through was discarded again. The column was centrifuged for 3 min at full speed. Then, the QIAamp Mini Elute column was placed in a new 1.5ml microcentrifuge tube. 20 µl Buffer ATE added to the column and incubated at 5 min at room temperature. The DNA was dissolved in the ATE buffer after centrifuging 1min at full speed. Using dsDNA BR Assay, quantification of DNA was performed by Qubit 2.0 fluorimeter (Life Technologies, Paisley, UK). 200µl of the working solution was prepared for each

sample. Qubit working solution was prepared by diluting the Qubit BR reagent 1:200 in BR Qubit buffer (Q32851, Invitrogen™). 2 extra amounts of working solution were always made to ensure enough working solution could be used for testing. The assay was prepared by adding 198µl of working solution to thin-walled, clear 0.5 ml PCR tubes. A minimum of 2µl of DNA was added to the working solution to give 200 ul. All tubes were vortexed for 2-3 seconds. Tubes were incubated for a minimum of 2 mins at room temperature. Tubes were inserted into the Qubit fluorimeter with broad range (BR) mode and taken the readings. STIC lesions and p53 signatures were laser capture micro-dissected using laser capture microdissection microscope (Carl Zeiss, Germany) (see section 2.7). Primary ovarian masses and metastatic samples were macro-dissected as described above. DNA from p53 signatures and STIC lesions was extracted using Qiagen's Generead DNA FFPE kit (Qiagen, 166011676) according to the manufacturer's instructions. 160 µl Deparaffinization Solution was added to adhesive cap after laser capture microdissection. Cap was vortexed vigorously for 10 s, and the sample was centrifuged briefly to the bottom of the tube. The cap was incubated at 56°C for 3 min, then cooled down to room temperature. Deparaffinization solution may become waxy or solid after cooling if too little deparaffinization solution is used or too much paraffin is carried over with the sample. If this occurred, additional Deparaffinization Solution was added and the 56°C incubation repeated. 55 µl RNase-free water, 25 µl Buffer FTB and 20 µl proteinase K were added to the cap, and the sample was vortexed and briefly centrifuged. Deparaffinization Solution formed a layer above Buffer FTB with the addition of proteinase K. The cap was incubated at 56°C for 1 hour, following 1-hour incubation at 90°C. Tube was briefly centrifuged to remove drops from inside the lid, and the lower, clear phase was transferred into a new micro-centrifuge tube. 115 µl RNase-free water was added to the tube and mixed well. 35 µl UNG was added to the sample, which was vortexed and incubated at 50°C for 1 hour in a thermomixer. The tube was briefly

centrifuged to remove drops from inside the lid. Then 2 µl RNase A (100 mg/ml) was added, mixed and incubated for 2 min at room temperature. 250 µl Buffer AL was added to the sample and mixed thoroughly by vortexing. Then 250 µl ethanol (96 – 100%) was added to each sample and mixed thoroughly by vortexing and centrifuged briefly to remove drops inside the lid. In total, 700 µl lysate was transferred to the QIAamp MinElute column (in a 2ml collection tube) and then centrifuged briefly to discard the flow-through. The collection tube was kept more times centrifuge until the complete lysate was used up. 500 µl Buffer AW1 was added to each spin column and centrifuge followed by 500 µl Buffer AW2 to each spin column and centrifuge. 250 µl ethanol (96 – 100%) was added to the spin column and centrifuge. Each step the flow-through was discarded and the collection tube was reused. The spin column was placed into a new 2 ml collection tube (supplied) and centrifuge to remove residual liquid. Finally, the QIAamp MinElute column was put into a clean 1.5 ml microcentrifuge tube (not provided) and discarded the collection tube containing the flow-through. QIAamp MinElute column was opened and applied with 25 µl Buffer ATE to the centre of the membrane.

Quantification of DNA from STIC samples was using dsDNA high sensitivity (HS) Assay, which were performed by Qubit 2.0 fluorimeter (Life Technologies, Paisley, UK). Qubit working solution was prepared by diluting the Qubit HS reagent 1:200 in HS Qubit buffer (Q32851, InvitrogenTM). In total, 200µl of the working solution was prepared for each sample. Again, 2 extra amounts of working solution were always made to ensure enough working solution could be used for testing. 2µl of DNA after elution was mixed with 198µl of working solution to thin-walled, clear 0.5 ml PCR tubes to give 200 a final total volume of 200 ul. All tubes were vortexed for 2-3 seconds and incubated for a minimum of 2 minutes at

room temperature. Then, the tubes were inserted into the Qubit fluorimeter with HS mode and taken readings.

2.5 H&E, p53 and Ki67 staining for identification of target lesions

Automated immunohistochemistry staining for Hematoxylin and eosin (H&E), p53 and ki67 was used to accurately identify p53 signatures and STIC lesions prior to laser capture microdissection (LCM). This process was performed in Leica Autostainer XL (V2.01) machine and Dr Nadia Rahman provided the protocol from Barts Cancer Institute. For H&E staining, the first step was to place the slides with paraffin sections in a staining basket in the load station and press start button. The slides were deparaffinised for 3 minutes in Xylene (97%, 2.5L, Code: X/0200/17, Fisher) three times. Then the slides were transferred to 100% IMS buffer ((Methylated spirit industrial 74 O.P.), 2.5L, Code: M/445/17, Fisher) for 3 minutes twice. Next, 3 minutes incubation was performed for slides in 90% IMS and 70% individually. After IMS treatment, the slides were washed by running tap water for 2 minutes and then moved to haematoxylin (Surgipath Haematoxylin Gill III, Leica, 1L, Ref: 3801542E) for 3:30 minutes. The slides were rewashed in tap water for 4 minutes. The differentiation step allowed the ability to remove stain from tissues selectively. And 1% acid alcohol (Surgipath, differentiating solution, 5L, Leica) was treated on the slides for 20 seconds. The slides were put into running tap water for 4 minutes. Eosin (Surgipath Eosin Y: alcoholic solution, Leica, 2.5L, Ref: 3801601E) was used to stain the slides for 6 minutes to get the extracellular matrix and cytoplasm pink. Followed by washing with rerunning tap water for 3 minutes, the slides were put into a series of increasing concentration of IMS (70%,90%) once and moved to 100% IMS three times for 1:30 minutes each. Finally, the slides were put into xylene for 2 minutes each three times.

p53 and Ki67 staining on glass slides was also done in a Ventana Discovery ULTRA machine (Roche Diagnostics Ltd). Similar to H&E staining, the tissue samples were cut as 10 µm section and deparaffinized in DISCOVERY WASH solution (950-510) at 75°C for 8 minutes. The slides were then treated by Standard CC1 (Cell Conditioning using Conditioner #1) (950-124, Roche) in 95°C for 44 minutes. Then the slides were blocked with Inhibitor D (760-4157), 37°C for 4 minutes. Without heating, the slides were incubated with primary p53 ((DAKO, M7001), Mouse) or Ki67 (Abcam, ab92742) antibody for 60 minutes. The dilution of p53 antibody was 1:50, and Ki67 antibody dilution was 1:10,000. Next, one drop of Biotinylated goat secondary antibody was applied for slides for 20 minutes at 42°C (anti-mouse IgG, anti-rabbit IgG; Ref: 760-4205). Then one drop of Blocker D (760-4161) was applied for 4 minutes at 42°C. Next, one drop of Streptavidin HRPO was added and incubated for 16 minutes at 42°C. Then one drop of DAB/DAB H₂O₂ (2 dispensers) (DISCOVERY DAB/H₂O₂ D, 760-4159) was used: for each, One Drop incubated the slides for 8 minutes at 42°C, these 2 dispensers are placed side by side on the reagent tray, and the instrument dispenses them consecutively within 6 seconds). One drop of Copper D (760-4160) was then applied to the slides, incubating for 4 minutes at 42°C. Next, haematoxylin (760-2021) was counterstained with slides and incubated for 8 minutes. Next, bluing Reagent (760-2037) was post counterstained with incubating the slides for 8 minutes. Finally, the slides were cleaned in warm tap water with detergent and dehydrated in graded ethanol and xylene. Coverslip the slides in permanent mounting medium.

2.6 Immunohistochemistry for p53 using PEN Slides

LCM-compatible polyethylene naphthalate (PEN) membrane slides (MembraneSlide 1.0 PEN, Carl Zeiss, Germany) ^[108] were pre-treated with UV (254nm) for 30 mins. Subsequently, 500 µl of poly-L-lysine 0.1% w/v (SIGMA, # P8920) was coated onto the slides for 5 mins. This step added extra “stickiness” onto the PEN slides. The slides were

then allowed to air dry at room temperature before use. The IHC staining protocol was also kindly provided by Dr Nadia Rahman from Barts Cancer Institute. FFPE blocks for STIC project were cut into 10 µm sections on a microtome Lecia RM2255. The cut sections were collected on a water bath at 46°C in the middle of pre-treated LCM slides. The deparaffinization of 10 µm thickness sections was done in the Leica AUTOSTAINER XL (12094147, Serial No: 4645), programme 8 as follows:

- 3 times of Xylene x3min,

- 2 times of 100% ethanol x3min,

- 90% ethanol x3min,

- 70% ethanol x2min,

- distilled water x2min

Then the slides were pre-treated by cell conditioning using conditioner #1, standard CC1(950-124, Roche). The step was done by Heat Induced Epitope Retrieval (HIER) in CC1 buffer at 95°C for 32 min. For blocking endogenous peroxides and protein, slides were treated in the Ventana Discovery UILTRA platform (Roche Diagnostics Ltd) using Inhibitor D (760-4157) and Blocker D (760-4161) from Roche. The IHC was optimised and performed by automation in the Ventana Discovery ULTRA system (Serial No: 319602). Reaction buffer (950-300, Roche) is used as a washing buffer between the steps. Automation applied one drop of p53 primary antibody (DAKO, M7001) onto the slides for treating 60 mins. The primary antibody was diluted with 1:50 (DAKO, M7001) by antibody dilution buffer (Discovery Ab Diluent, 100mL, 760-108). And then, one drop of biotinylated goat secondary antibody (universal secondary Ab, anti-mouse IgG, anti-rabbit IgG; Ref: 760-4205) was treated on the slide for 20 minutes at 37°C. Then one drop of Blocker D was applied on the slide for 4 minutes at 37°C. One drop of Streptavidin HRPO (Discovery SA-HRP D, 760-4162) was used and incubated for 16 minutes at 37°C on the slides. For the next step, one

drop of DAB chromogen substrate (DISCOVERY DAB D, 760-4158)/DAB H₂O₂ (DISCOVERY DAB/H₂O₂ D, 760-4159) (2 dispensers) was used for the detection system. For each dispenser, drop was incubated on a slide for 8 minutes at 37°C. These 2 dispensers are placed side by side on the reagent tray. The instrument dispenses them consecutively within 6 seconds. Then one drop of Copper D (DISCOVERY Copper D, 760-4160) was applied on the slides, incubating 4 minutes at 37°C. The slides were counterstained with Haematoxylin (760-2021), incubating for 8 minutes. After counterstaining, post counterstaining was performed with Bluing Reagent (760-2037), incubating for 8 minutes. Finally, the slides were cleaned up with distilled water. The stained slides were microdissected with a Zeiss microscope (Carl Zeiss, Germany). DNA from microdissected FFPE sections was extracted with a Generead kit according to the manufacturer's previously described instructions ^[109].

2.7 Laser capture microdissection by Zeiss microscope

The brief steps for the laser capture microdissection process were as follows: First slides and LCM caps were loaded. Secondly, the tissue of interest was chosen under a microscope. Thirdly, the LCM caps were placed onto the cap stage whilst calibrating the laser. Then a layer of interest was selected and captured; the samples were flipped into the caps for downstream DNA extraction. For the details of laser capture microdissection, the steps are as follows:

Firstly, the PC and the screen monitor were turned on, followed by a control box, power supply box, and microscope/camera. The PALMRobo 4.6 software program was opened by double-clicking on the PALMRobo 4.6 software icon on the desktop. "Moves stage to Load Position" icon was clicked on the SetUp tool pane. The step allowed the adhesive cap and slides to be loaded under the microscope stage moving forward. Three slides were loaded up

on the stage. “Moves stage to Load Position” icon was clicked again to move the stage back to the original position. The “Moves stage to CapCheck and back” icon was clicked to move the adhesive cap to the position above the slides. The objective mode was set in 10X, with capture mode was set as “CenterRoboLPC” and “Tube” individually. Laser capture numbers were set with speed at 12 of Arrow Keys percent, 57 at Cut Energy percent, 20 as LPC energy Delta number, 87 at Focus number and 2 as Focus Delta number. The “Selects a line tool” icon was clicked to mark the specific region when the pencil icon was used for freehand drawing. Freehand drawing objects must be a closed figure, which means the ends of the drawing must touch to enclose the desired area. Some options such as the “Defined circle tool” allows the selection of regions defined of known diameter, while the “Selects a line tool” function could make the microdissection either straight or curved line. “Saves the actual image with a new name” icon was then clicked to set the path regarding the saved location as “before laser capture microdissection images” with marked lines. After ensuring the target lesion, the “Starts laser function” icon in the tool panel was clicked to start laser capture microdissection. “Saves the actual image with a new name” icon was clicked again to set the path regarding the saved location as “after laser capture microdissection images”. When some tissue was outside the cap, the “Select Dot tool” icon was clicked to collect these tissues from flipping into the lid. The steps above were repeated until all the p53 stained tissue was collected. The cap can be removed or placed back on the slide to continue microdissection by clicking on “Moves stage to Capcheck and back” or “Moves stage to Load Position” icons. When no further microdissection was performed with the cap, the cap was carefully removed from the stage and microscope. The cap was kept at room temperature for downstream DNA extraction. After all dissections are completed, check if all slides and caps were removed from the stage. Then the PALMRobo 4.6 software was closed. Zeiss instrument was turned

off in the microscope/camera, power supply box and control box. Finally, the PC and the monitor were turned off.

2.8 Targeted panel sequencing

2.8.1 Targeted amplicon sequencing for late stage cohort

Mutation screening of *PTEN*, *KRAS*, *RBI*, *BRCA2*, *RAD51B*, *FANCM*, *PALB2*, *RAD51D*, *TP53*, *RAD51C*, *BRIP1*, *CDK12*, *NF1*, *BRCA1*, *BARD1* and *PIK3CA* was performed on all 300 samples using tagged-amplicon sequencing as our lab collaborated with Cambridge's previous paper described^[1]. In addition, DNA extracted from matched patients' blood samples were analyzed by tagged-amplicon sequencing for BRCA1 and BRCA2 germline mutations. These samples were sequenced by Hiseq2500, 75PE mode.

2.8.2 AmpliSeq Custom Panel design for early stage cohort

The custom panel of Ampliseq was designed on Illumina website (<https://www.illumina.com/products/by-brand/ampliseq/custom-panels.html>). The selected genes included *PTEN*, *KRAS*, *RBI*, *BRCA2*, *RAD51B*, *FANCM*, *PALB2*, *RAD51D*, *TP53*, *RAD51C*, *BRIP1*, *CDK12*, *NF1*, *BRCA1*, *BARD1* and *PIK3CA*. There are 776 amplicons for covering the genomic coding regions of *PTEN*, *RBI*, *BRCA2*, *RAD51B*, *FANCM*, *PALB2*, *RAD51D*, *TP53*, *RAD51C*, *BRIP1*, *CDK12*, *NF1*, *BRCA1*, *BARD1* genes. Three amplicons covered the *KRAS* hotspot mutation (G12) and four amplicons covered the *PIK3CA* hotspot mutations (E545 and H1047). The sequence details of amplicons for these genes are shown in Appendix 8.

2.8.3 Ampliseq library preparation

Details of the sequencing of BriTROC-1 samples are given elsewhere^[1]. DNA from early stage FFPE tissues was extracted as per 2.4. Following DNA extraction, the sequencing

library was constructed using the AmpliseqTM Custom DNA panel for Illumina (20020495) according to manufacture instructions. First, the concentration of DNA was checked by Qubit as described before. If enough DNA was available, an intermediate concentration was diluted as follows: a. concentration of ~20 – 50 ng/μl of DNA was diluted using Low TE in the pre-PCR lab. b. Diluted DNA was re-quantified by using the same fluorometric quantification method. The desired final concentration was around 4ng/μl. 5 μl desired concentration diluted DNA was prepared and to yield a total of approximately 20 ng. In the first DNA amplification targets step, the mixture was prepared with 5 μl of 5X AmpliSeq HiFi Mix, around 20ng DNA and top up with 12.5 μl Nuclease-free water. Next, 5 μl of the master mix was transferred to one well and the same mixture to a second well. 5 μl 2X Ampliseq DNA Panel Pool1 was added to the first well, and 5 μl 2X Ampliseq DNA Panel pool2 was put to the second well. After pipetting and mixing the master mix and pool, samples were transferred to the post-PCR area and run AMP_DNA program on a thermocycler with the following protocol: Set the reaction volume to 10 μl for two pools; 99°C for 2 minutes; 19 cycles of 99°C for 15 seconds then 60°C for 4 minutes; Hold at 10°C for up to 24 hours.

The sequence of amplicons in the two pools is shown in Appendix 5. After AMP_DNA program, samples were moved on to digest amplicons step partially. Then, the two pools were mixed by pipetting and combined with 2 μl of FuPa enzyme. And the new reaction volume was 22 μl in total for two primer pools. The sample was then incubated in a thermal cycler with the FUPA program: 10 minutes at 50 °C; 10 minutes at 55 °C; 20 minutes at 62 °C; Hold at 10°C for up to one hour.

The entire product after FUPA program was then mixed with ligating index1 and index 2 adapters. The sequence of indexes for the pooling guide is shown in Appendix 6. The samples were mixed thoroughly with 4 μl of switch solution buffer, 2 μl of Ampliseq CD

Indexes (96 Indexes, Illumina, 2127688) and 2 µl of DNA ligase to get 30 µl total volume. Mixture samples were running on the thermal cycler with the LIGATE program: 22 °C for 30 minutes; 68 °C for 5 minutes; 72 °C for 5 minutes; Hold at 10°C for up to 24 hours. After LIGATE program, samples in the library plate were centrifuged briefly to collect contents and then unsealed. Dr Ian Garner made in-house magnetic beads and kindly helped me to clean up the library. The DNA product from the LIGATE program was then purified using 30 µl magnetic beads. The DNA-bead mixes were allowed to incubate for 5 minutes at room temperature and then placed on a magnet plate (Ambion). After the mixture became clear, the plate was unsealed, and the entire supernatant was discarded from each well. Each well was washed twice by 150 µl freshly prepared 70% ethanol, and the plate was sealed immediately and centrifuged briefly. The plate was put back on a magnetic stand and unsealed. All residual ethanol was removed immediately and moved on to the next step of the amplifying library. The samples were mixed thoroughly with 45 µl of 1X Lib Amp Mix, 5 µl 10X Library Amp Primers. The final total volume per reaction reached 50 µl. The samples were then incubated in a thermal cycler as AMP_7 program: 2 minutes at 98 °C; 15 seconds at 98°C; 1 minute at 64°C for 7 cycles; Hold at 10°C for up to 24 hours.

Lastly, DNA was purified by 25 µl magnetic beads by vortexing and incubating at room temperature. The entire supernatant containing the desired amplicon library was transferred to a new plate. 60 µl magnetic beads were added to each well last time and cleaned up with 70 % ethanol as described previously. At the end of library preparation, DNA was eluted in 20 µl 0.1x TE buffer. All samples were then moved on to quality control on the Agilent D5000 Screen Tape System (Agilent technologies, 0201918-98). First, 1 µl DNA samples were mixed with 10 µl D5000 sample buffer each (Agilent, 5067-5589). Similarly, 1 µl the DNA ladder was mixed with 10 µl D5000 sample buffer. Next, the samples and the ladder

were vortexed, spun down and then transferred to an Agilent 2200 TapeStation where they were subjected to automated electrophoresis using D5000 Screen Tapes (Agilent 5067-5588). Finally, DNA was quantified using the Qubit dsDNA high sensitivity assay kit (Thermofisher, Q32854) as described previously.

Samples with tape station electropherograms showing a single DNA peak between 200-300 bp size were then subjected to next-generation sequencing (NGS). The results of electropherograms of all samples are shown in Appendix 7. All samples were pooled to an equal concentration of 2 nM before sequencing.

2.8.4 AmpliSeq sequencing

The pool of early stage libraries was performed sequencing on Imperial College London, BRC facility. HiSeq4000 System (Illumina, Cambridge, UK) was used with paired-end 150 bp protocols. The mean Ampliseq coverage was $>7000\times$. The sequencing data were released on the EGA database under the accession number EGAS00001005567.

2.9 Somatic mutation calling from target sequencing

2.9.1 Somatic mutations calling from Ampliseq panel

The raw sequencing Ampliseq data were transferred from the BRC facility in Imperial College London. FASTQ files were trimmed with adapters and aligned to reference human genome hg19 using Burrows-Wheeler Alignment (BWA-MEM) ^[110] and pre-processed using samtools and Picard to generate sorted BAM files ^[111]. Somatic mutations were called using mutect2 (GATK4.1.4.1) ^[112], VarScan2 (version 2.4.2) ^[113], Strelka2 (version 1.0.14) ^[114] and HaplotypeCaller ^[115] pipelines for single nucleotide variations (SNVs) and small insertions and deletions (Indels) on tumour-only BAM files. Mutations were annotated using Variant

Effect Predictor (VEP) (version 1.5.3)^[116]. Clinical significance filtered Somatic mutations with only “pathogenic” and “likely_pathogenic” mutations retained for downstream analysis. At least two pipelines confirmed single nucleotide variants.

2.9.2 Somatic mutations calling from TAm-seq panel

These sequencing data were obtained from Cambridge University and Mr Hasan Mirza and Dr. Thomas Bradley provided analysis result of TAm-seq data for late stage cohort. FASTQ files were trimmed for adapters and aligned to reference human genome hg19 using Burrows-Wheeler Alignment (BWA-MEM)^[110] and pre-processed using samtools and Picard to generate sorted BAM files^[111]. Somatic mutations were called using Mutect2 (GATK4.1.4.1)^[112], Varscan2 (version 2.4.2)^[113], Strelka2 (version 1.0.14)^[114] by Hasan Mirza, and HaplotypeCaller^[115] pipelines for single nucleotide variations (SNVs) and small insertions and deletions (indels) on tumour-only BAM files (new early stage samples) and tumour-normal pairs for the BriTROC-1 samples, where available, using default parameters. Where multiple samples existed for individual BriTROC-1 patients, sample data in sorted BAM files were merged prior to mutation calling but individual sample identity was retained. Mutations were annotated using Variant Effect Predictor (VEP) (version 1.5.3)^[116]. Somatic mutations were filtered by clinical significance (ClinVar, October 2020) with ‘benign’ and ‘likely_benign’ variants discarded. Variants were further checked for pathogenicity in the COSMIC database (GRCh37, February 2021). Germline mutations for BriTROC-1 samples were detected using similar criteria for pathogenicity using Strelka and HaplotypeCaller.

2.10 Shallow whole genome library prep and sequencing

For late-stage samples, details of the sequencing of BriTROC-1 samples are given previously^[1]. 100ng DNA from BriTROC-1 samples was used for sWGS libraries preparation by modifying TruSeq Nano DNA LT Sample PrepKit (Illumina) protocol. The

libraries were assessed quality and quantity with DNA-7500 kit on Bioanalyzer (Agilent Technologies) and Kapa Library Quantification kit (Kapa Biosystem) according to the manufacturer's instructions. The libraries were pooled by 16 to 20 samples together, and each pool was diluted to equimolar amounts. The final pool was sequenced on Hiseq2500 in SE-50bp mode.

For early stage samples, DNA was extracted from $10 \times 10 \mu\text{m}$ sections using QIAmp DNA FFPE Tissue Kit (Qiagen, UK) according to the manufacturer's protocol in step 2.4. Shallow whole genome libraries preparation and sequencing were performed by the BRC facility in Imperial College London. 50-200ng was sheared with a Covaris LE220 focused-ultrasonicator (Covaris, Woburn, MA) to produce 100-200bp fragments. Libraries were generated using SureSelect XT standard protocol (Agilent Technologies, Santa Clara, CA) for FFPE samples. The quality and quantity of the shallow whole genome sequencing (sWGS) libraries were assessed with DNA-7500 kit on 2100 Bioanalyzer (Agilent Technologies) and with Kapa Library Quantification kit (Kapa Biosystems) according to the manufacturer's protocols. According to the manufacturer's instructions, sequencing was performed on the HiSeq 4000 150 pair end (Illumina Cambridge, UK) platform, using 250-300 ng input DNA. The minimum number of reads per sample was set at 5-10 million (mean coverage of $0.1 \times$). Using previous calculations (https://gmacintyre.shinyapps.io/sWGS_power/), 10 million reads with a bin size of 30kb had 80% power (alpha 0.01) to detect CN change ± 2 at 30% purity assuming ploidy of 2. The sequencing data were placed on EGA database under the accession number EGAS00001005567.

The input DNA for shallow whole genome sequencing for the STIC project is very low. Some of the samples had DNA yield less than 1 ng. NEBNext FFPE DNA Repair Mix (M6630) and NEBNext Ultra II DNA library Prep Kit for Illumina (E7645) were used for STIC project samples library preparation. The BRC facility in Imperial College London performed these sample libraries preparation and sequencing. Dr Katerina Rekopoulou and Dr Benjamin Phillimore from the BRC facility in Imperial College London assisted with Ampliseq sequencing, shallow whole genome library preparation and sequencing. DNA was sheared in 1X TE to a final volume of 50 μ l. For the p53 signatures, 48 μ l SITC fragmented DNA samples were mixed thoroughly with 3.5 μ l FFPE DNA repair buffer, 3.5 μ l NEBNext Ultra II End prep buffer, 2 μ l NEBNext FFPE DNA repair mix. The total volume was 57 μ l, and the mixture was then placed in a thermocycler at 20 °C for 30 minutes (incubation time optimised in advance). After incubation, samples were added 3 μ l NEBNext Ultra II End Prep enzyme mix directly to the FFPE repair reaction mixture. The entire volume was pipetted up and down at least 10 times to mix thoroughly to mix well. The samples were performed a quick spin to collect all liquid from the sides of the tube. Then the samples were placed in a thermocycler, with the heated lid set to 75°C, and run with the following protocol: 30 minutes at 20°C; 30 minutes at 65°C; Hold at 4°C.

To proceed to the Adaptor Ligation step, the NEBNext adaptor was diluted on ice. Next, 5X Adaptor ligation buffer and 2.5 μ l diluted adaptor was added to the End-Prepped DNA plus 1 μ l NEBNext ligation Enhancer and 30 μ l NEBNext Ultra II ligation master mix. The sample was pipetted up and down to mix thoroughly and then incubated at 20 °C for 15 minutes. After this, 3 μ l the USER enzyme was added to make a total volume of 96.5 μ l. The mixture was mixed thoroughly and incubated for another 15 minutes at 37°C with the heated lid set to above 47 °C.

The ligation reaction product was cleaned up one more time using magnetic beads as we mentioned before, with a volume of 87 μ l beads (0.9 x of the sample) and mix well by pipetting up and down at least 10 times. At the end of the purification, DNA was eluted in 17 μ l 0.1 x TE buffer (PH 8.0).

For the final library construction step, DNA was amplified again by PCR enrichment. The adaptor-ligated DNA was mixed with 25 μ l NEBNext Ultra II Q5 master mix, 5 μ l universal PCR primer (i5 primer) and 5 μ l i7 index (barcoded) primer. The entire volume was pipetted up and down at least 10 times to mix thoroughly. Samples were placed in a thermocycler and performed the PCR amplification below: 1 cycle of 30 seconds at 98°C as initial denaturation; 3-12 cycles of 10 seconds at 98°C as denaturation; 75 seconds at 65°C for annealing/extension; 5 minutes at 65°C for the final extension; Hold at 4 °C.

The number for the second cycle was determined based on the amount of input DNA and the following optimisation done before processing the tumour samples of interest. Lastly, DNA was cleaned up again with 1x magnetic beads. Again, 200 μ l 80% freshly prepared ethanol was added twice to remove supernatant. At the end of clean up, DNA was eluted in 20 μ l 0.1x TE buffer. All samples were then subjected to quality control using 1 μ l DNA per sample on the Agilent D5000 high sensitivity Screen Tape System. First, DNA samples were mixed with a 10 μ l high sensitivity D5000 sample buffer (Agilent technologies, 0201485-87). In addition, according to the manufacturer's instruction, 1 μ l the DNA ladder was mixed with 10 μ l D5000 sample buffer. The samples and the ladder were vortexed for 1 min, spun down and then transferred to an Agilent 2200 TapeStation where they were subjected to automated electrophoresis using high sensitivity D5000 Screen Tapes.

Samples with tape station electropherograms showing a single DNA peak between 200-300 bp size were then subjected to next-generation sequencing (NGS). DNA concentration was quantified using the Qubit dsDNA high sensitivity assay kit (Thermofisher, Q32854). All samples were pooled to an equal concentration of 2 nM before sequencing.

The final library was sequenced on the Hiseq4000 (Illumina) with a 150 paired-end sequencing strategy. Each sample was sequenced around 10 million reads.

2.11 Copy number calling

2.11.1 Relative copy number

Reads were aligned using bwa-mem^[110] to the 1000 genomes version of human genome build GRCh37 into SAM file. We used the samtools view command to convert SAM to BAM. SAM format (by default, it expects BAM) using the -S option. And we also used the -b option to get BAM (by default, it produces BAM) file. FASTQ files were aligned randomly to their position in the reference genome. The BAM file is in the order that the sequences occurred in the input FASTQ files. We used the samtools sort command to sort ordered BAM files based on each chromosome's alignment coordinates. QDNAseq^[117] was used to sort BAM files, then segment the genome and identify regions of copy-number change to call relative copy number.

2.11.2 Absolute copy number

The sWGS-absolute (swgs) pipeline implements a method for generating absolute copy number profiles from shallow whole genome sequencing data utilising a multi-stage fitting process to generate a set of read depth-normalised absolute copy number profiles. Copy number fits are generated in read count space using a modified implementation of

QDNAseq^[117] on full read depth shallow whole genome BAM files. After which, a grid search of ploidy and purity values is performed to generate a matrix of potential absolute copy number profiles. The fitting of absolute copy number profiles includes a fit error function and a variant allele fraction anchoring process, which assist in selecting an optimal copy number fit. Minimisation of the clonality error function, expected-to-experimental TP53 distance, and manual review of fits allow for selecting an optimal absolute copy number profile for a given sample. Selected fits are subject to a calculation of power to detect copy number alterations. The read depth of a given sample is assessed against its selected ploidy-purity combination. This process determines if 1) a sample has a sufficient number of reads to support the selected ploidy-purity combination and 2) an optimal target number of reads to perform sample-specific read downsampling. This process normalises the read depth between samples in ploidy and purity-dependent manner. Read variance across segments is consistent, while excluding samples that are not supported by a sufficient number of reads. Samples passing all filtering criteria undergo read-down sampling to the specified target number of reads determined by the previous steps and absolute copy number profiles fitted at the ploidy-purity combination selected prior. This package was developed by Dr Philip Smith, our collaborator in Computational Biology Lab, Cancer Research UK Cambridge Institute. I also used ACE^[118] to calculate the ploidy and purity of each sample as a validation of swgs pipeline with the settings “penalty=0.5” and “penploidy=0.5” and bin size=30kb for the function “loopsquaremodel”.

2.11.3 Copy number signature calling

sWGS reads were aligned to reference human genome hg19. Relative copy numbers were obtained for predefined 30kb bins using QDNASeqmod packages^[117]. I derived absolute copy numbers using the sWGS-absoluteCN (swgs) pipeline - full details are given in chapter 4. Focal amplifications and deletions were defined according to the COSMIC definitions (see

<https://cancer.sanger.ac.uk/cosmic/help/cnv/overview>): amplification was defined as total copy number ≥ 5 if average ploidy ≤ 2.7 , or ≥ 9 if average ploidy > 2.7 . The loss was defined as total copy number 0 if average ploidy ≤ 2.7 or (ploidy minus 2.7) if average ploidy > 2.7 . Copy number signatures were calculated using the R scripts as previously published .

Ms Lena Morrill Gavarró from CRUK Cambridge Institute provided a fixed effects Bernoulli model with an intercept and a coefficient for the change between early and late stage samples to model the presence or absence of signatures. The presence of a signature j in sample i is modelled by a Bernoulli with probability θ_{ij} , where $\theta = x^T \beta$. x has two rows - for the intercept and the difference between the groups - and as many rows as samples. β has two rows and as many columns as signatures ($d = 7$). The change in the differential abundance of non-zero exposures has been modelled similarly. More detailed information will be provided in chapter 4. I used a multivariate normal model based on the isometric log-ratio (ILR)-transformed exposures (also called a logistic-normal model in the literature). The ILR transformation maps a d -dimensional compositional vector (the exposures) to a $d-1$ dimensional vector of real values. To account for absent signatures, the transformation (and subsequently the model) only used the subset of signatures present in each sample.

Covariates are the same as those in the Bernoulli model, but this time $\theta = x^T \beta$ contains the ILR-transformed probabilities. Therefore, it has 6 columns instead of 7, and each row can be transformed back to a 7-dimensional vector of possibilities with the inverse ILR transformation. β will continue to have two rows, but only 6 columns, which indicate changes in log-ratios of signature exposures. The R package TMB^[119] was used for inference. Some more detailed information about copy number signatures comparison will be presented in Chapter 4.

2.12 Statistical analysis

2.12.1 Unsupervised hierarchical clustering statistical analysis

Hierarchical clustering of the exposure vectors of samples (early stage and late stage) used in the survival analysis was performed using the NbClust^[120] package in R 3.3.1. Pearson correlation-based distance and complete linkage were used to obtain the final clusters. The resulting clusters were confirmed by repeating the clustering analysis using Euclidean's distance. Using the cluster labels as covariates, a Cox proportional hazards model was fitted using the R packages survival^[121] and survminer (<https://rpkg.sdatanovia.com/survminer/index.html>).

2.12.2 Survival analysis

Overall survival was estimated using the method of Kaplan and Meier, and univariate survival analysis was performed using the log-rank test and multivariable survival analysis by Cox proportional hazard models adjusting for age, stage, grade and presence of residual disease.

2.12.3 Kolmogorov-Smirnov test

To compare copy number signatures distribution in the early stage and late stage cohorts, we performed Kolmogorov–Smirnov test. Kolmogorov-Smirnov test was also utilised for all variables to test for normality of the data. ($p < 0.05$)

2.13 Materials

Chemicals, kits and buffers

DNA extraction

Histoclear (National diagnostics, 5989-27-5)

Ethanol (VWR, 20821.330)

QIAamp DNA FFPE Tissue Kit (Qiagen, 166027818)

GeneRead DNA FFPE Kit (Qiagen, 166011676)

DNA quantification

Qubit dsDNA HS Assay kit (Invitrogen, 2234958)

Qubit dsDNA BR Assay kit (Invitrogen, 2226331)

Immunocytochemistry

Poly-L-lysine solution 0.1% (w/v) in H₂O (Sigma, P8920)

Cryosprayer freezer spray (Cellpath, 90029384)

Nuclease free water (Invitrogen, 2009041)

Library preparation

Ampliseq library plus for illumina 96 reactions (illumina, 2315210)

Ampliseq CD Indexes, SetA for illumina (96 Indexes, illumina, 2127688)

Ampliseq Custom DNA Panel for illumina (illumina, 85728848)

2.14 Equipment & plasticware

DNA quantification

DNA quantification Nanodrop ND1000 (Thermo Scientific,)

Tapestation 2200 (Agilent,)

D5000 ScreenTape (Agilent technologies, 0201918-98)

High sensitivity D5000 ScreenTape (Agilent technologies, 0201485-87)

Macrodissection

Disposable scalpels 22 (Swann Morton, 6611903)

Disposable premier MB-22 blade (Thermo scientific, 3050822)

Micoscope slides (Thermo scientific, J1800AMNZ)

Library preparation

Thermal cycler (Bio-Rad, ALS1296)

Heraeus multifuge X3R centrifuge (Thermo Scientific, 41021339)

Dry-bath (Flowgen, AS-BSH200-039)

Centrifuge (VWR, 42187805)

Safe-lock 1.5ml tube (Eppendorf, F170264P)

AdhesiveCap 500 opaque (Zeiss, 000810-21)

0.2ml PCR tubes (Starlab UK, 19283)

MicroAmp Optical 8-cap strip (applied Biosystems, 4323032)

Optical Tube, 8x strip (Agilent Technologies, 01301065)

Optical Cap, 8x strip (Agilent Technologies, 01303624)

96 well PCR plate and adhesive covers (Starlab UK)

Microseal 'B' seal (BIO-RAD, MSB1001)

Graduated filter tips of various sizes (Starlab UK)

Stripettes (various sizes) (Corning B.V. Life Sciences)

Falcon tubes (15ml and 50ml) (Corning B.V. Life Science)

20µl Bevelled, filter tip (Tipone, S1120-1810)

Magnet plate (Ambion, L/N:1207056)

CHAPTER 3. The genomic landscape of early stage ovarian high grade serous carcinoma

3.1 Introduction

This chapter explores the rates of genomic short variants (SNV/indel) in a target gene panel and as well as somatic copy number abnormalities (CNA) in early and late stage high grade serous carcinoma cohorts. The SNV/Indel mutations in the early stage cohort are evaluated by Ampliseq sequencing using a custom amplicon panel; the late stage cohort was evaluated using tagged-amplicon sequencing, again using a custom amplicon panel covering the same genes. In addition, shallow whole genome sequencing is used to call absolute copy number using two pipelines (swgs-absn and ACE) as well as to identify focal somatic CNA.

3.2 Background

High grade serous carcinoma (HGSC) accounts for approximately 70% of all epithelial ovarian cancer (O.C.) cases and about 80% of all ovarian cancer deaths. Most patients with HGSC present with advanced (FIGO stage III and IV) disease, where treatment is rarely curative. Despite the addition of anti-angiogenic agents ^[26, 27, 93] and PARP inhibitor therapy ^[122, 123], the majority of patients with advanced disease relapse within 24 months of completion of first-line chemotherapy. By contrast, the small proportion (10 - 15%) of patients who present with early disease (stage I and II) have a much better prognosis and are frequently cured with surgery and platinum-based chemotherapy alone ^[16]. It is unclear whether patients with early-stage HGSC have a distinct subtype of disease that fails to metastasise or whether these cases are genomically similar to late-stage disease but are identified essentially by chance before metastasising. To address this, I undertook genomic analysis a cohort of early-stage HGSC patients identified at three large UK centres, with comparison to late-stage samples from the BriTROC-1 study.

HGSC is marked by near-universal *TP53* mutation^[62, 124]. However, beyond *TP53*, rates of recurrent short variants are low in HGSC. However, SNV/indels are seen in *PTEN*, *KRAS*, *RB1*, *BRCA2*, *RAD51B*, *FANCM*, *PALB2*, *RAD51D*, *TP53*, *RAD51C*, *BRIP1*, *CDK12*, *NF1*, *BRCA1*, *BARD1*, *PIK3CA*, albeit at low frequencies in all genes other than *BRCA1* and *BRCA2*. Therefore, we chose these genes as a custom panel for Ampliseq sequencing.

The genome of HGSC is highly complex, and I estimated relative copy number from shallow whole genome sequencing data from early and late stage cohorts. The absolute copy number was called from the relative copy number and allele frequency of *TP53* by the sWGS-absoluteCN (swgs) pipeline.

3.3 Results

3.3.1 The mutational landscape of early and late-stage cohorts using a custom-made panel of genes

I identified 54 patients with early-stage (defined as stage IA, IB, IC and IIA using the FIGO classification at the time of diagnosis) ovarian high grade serous carcinoma from the pathology archives of large U.K. gynaecological cancer centres (Imperial College Healthcare, University College London and Barts Health NHS Trusts) and collected 84 samples from these patients. Upon pathology review, no tumour or low cellularity was found in 6 samples (3 patients). Nine samples from 7 patients were further identified as having as late stage or recurrent disease, whilst six samples from 3 patients were confirmed as low grade serous carcinoma following pathology review by two pathologists. Overall, 21 samples from 13 patients were excluded, whilst two samples from one patient failed DNA extraction. A summary of the patient and sample workflow is shown in Figure 3.1(A) and Figure 3.1(B). Additionally, we identified a further cohort of three early-stage patients recruited into the BriTROC-1 study ^[93], giving a total early-stage population of 43.

The comparison late-stage cohort consisted of 52 randomly selected patients with stage IIIC/IV disease recruited into the BriTROC-1 study. Table 3.1 shows the clinical characteristics for both cohorts, whilst clinical information is shown in Appendix 1 and Appendix 2.

The early-stage patients were diagnosed more recently than the late stage (early, median 68 months, range 24-177, prior to analysis; late median 101 months, range 60-179; $p=0.0009$ Figure 3.2). Overall survival (OS) was defined as the time from the date of surgery until the date of death or last follow-up. Eighty-five per cent of the patients underwent primary

cytoreductive surgery, and 15% were operated after neoadjuvant chemotherapy but all samples were obtained before any chemotherapy treatment. The BriTROC-1 samples were those obtained at the time of diagnosis rather than relapse.

The median age at diagnosis for early and late-stage cohorts did not differ significantly (early 61.3 years, range 40-84; late 62.3 years, range 34-76, Figure 3.3A), but overall survival was, as expected, significantly longer in the early-stage cohort than for the late-stage (Hazard Ratio 0.11, 95%CI 0.06-0.22, Figure 3.3B).

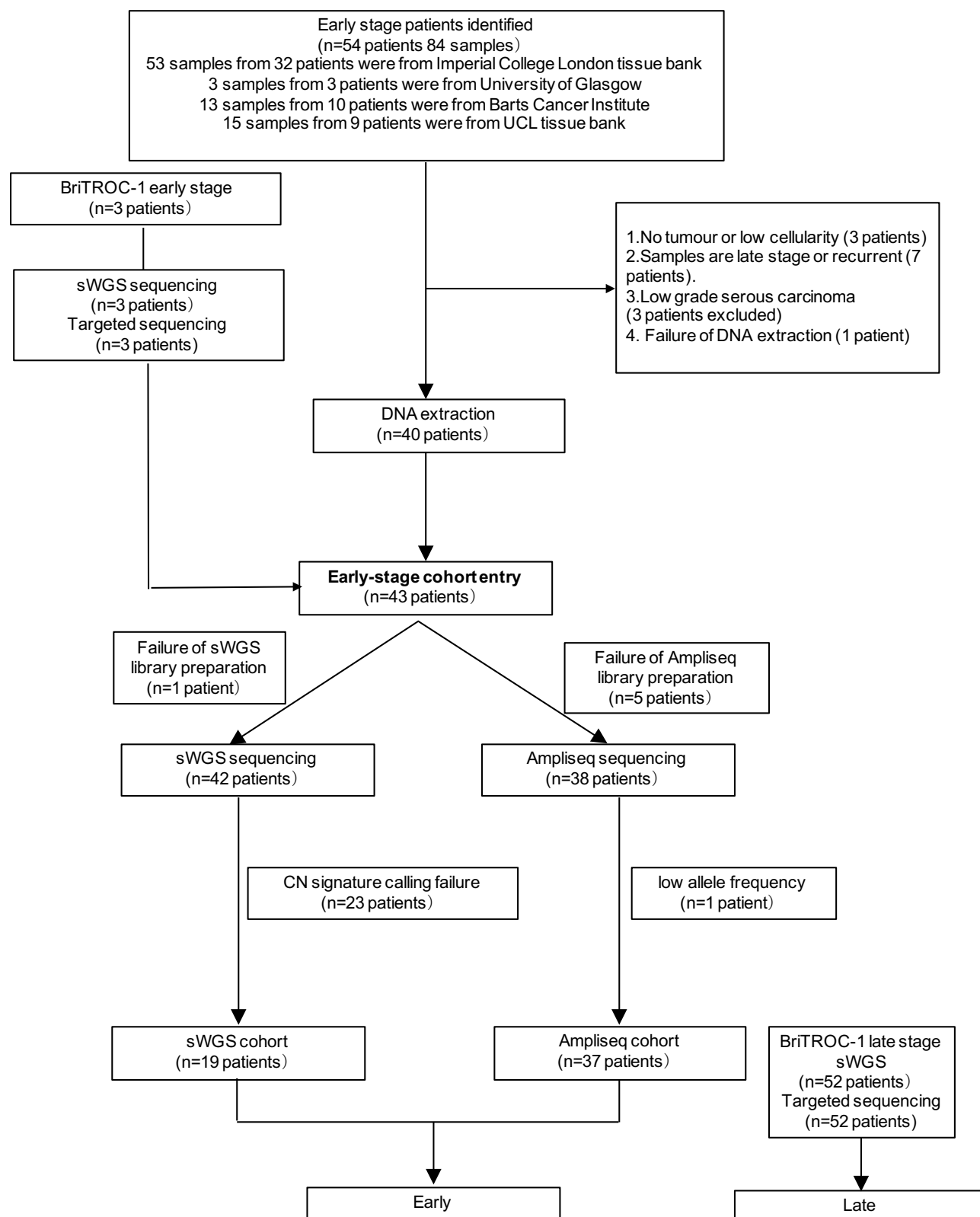


Figure 3.1 (A) Remark diagram for early-stage and late-stage cohorts by patients

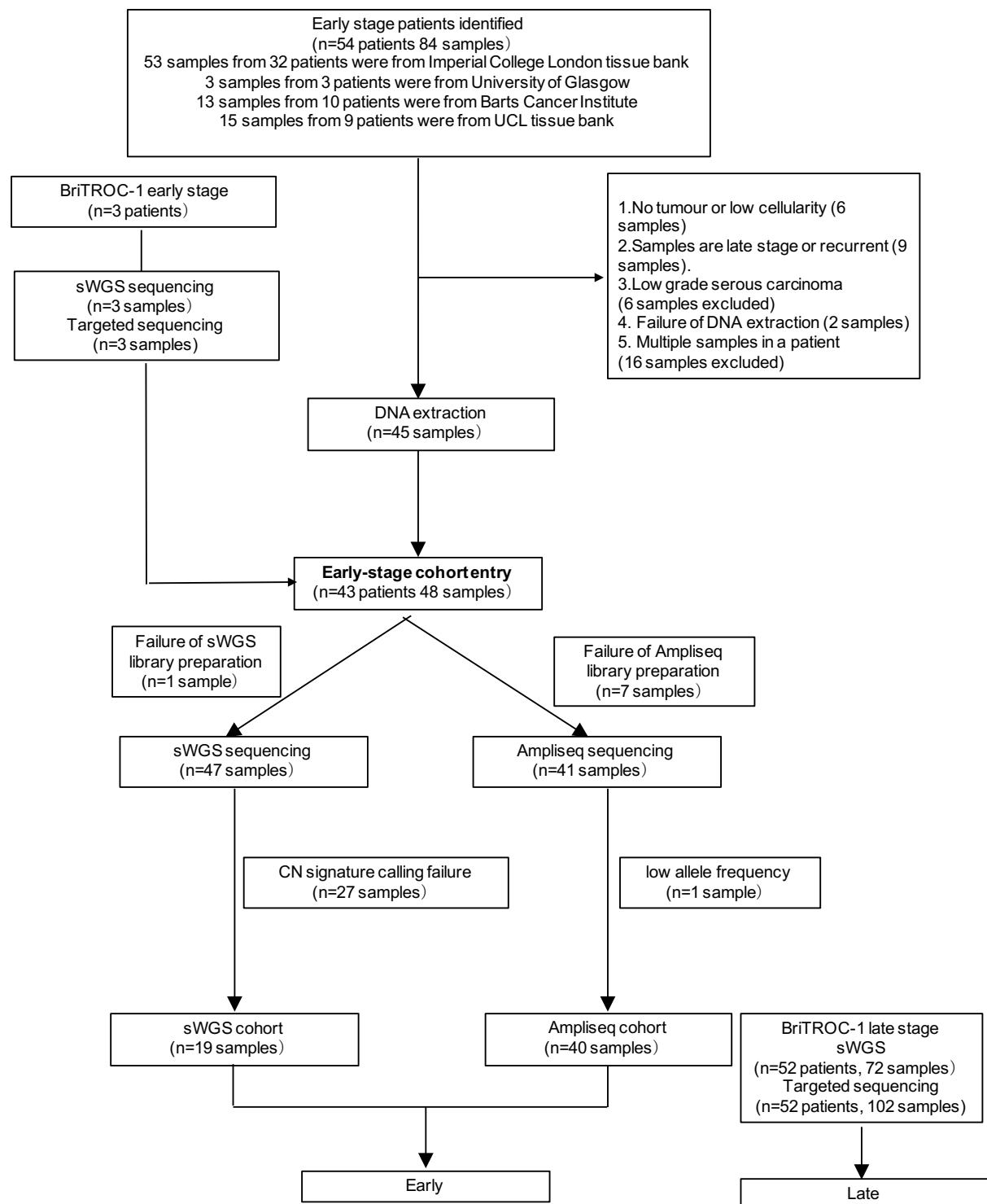


Figure 3.1 (B) Remark diagram for early-stage and late-stage cohorts by samples

Characteristic	early_stage cohort (n=43)	late_stage cohort (n=52)
Age (years)		
Median	61.3	62.3
Range	(40.3-84.6)	(34-76)
Stage		
IA	4	
IB	3	
IC	22	
IIA	14	
IIIC		42
IV		10

Table 3.1 Clinical characteristics of study participants used in the discovery cohort.

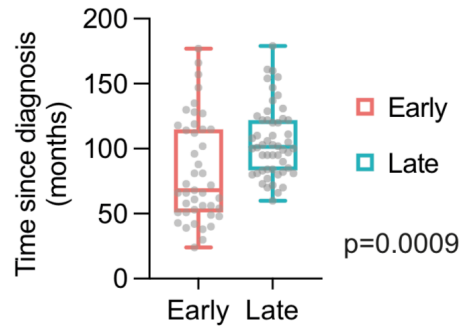


Figure 3.2 Time interval between diagnosis and sample analysis by cohort.
Mann Whitney. $p < 0.05$

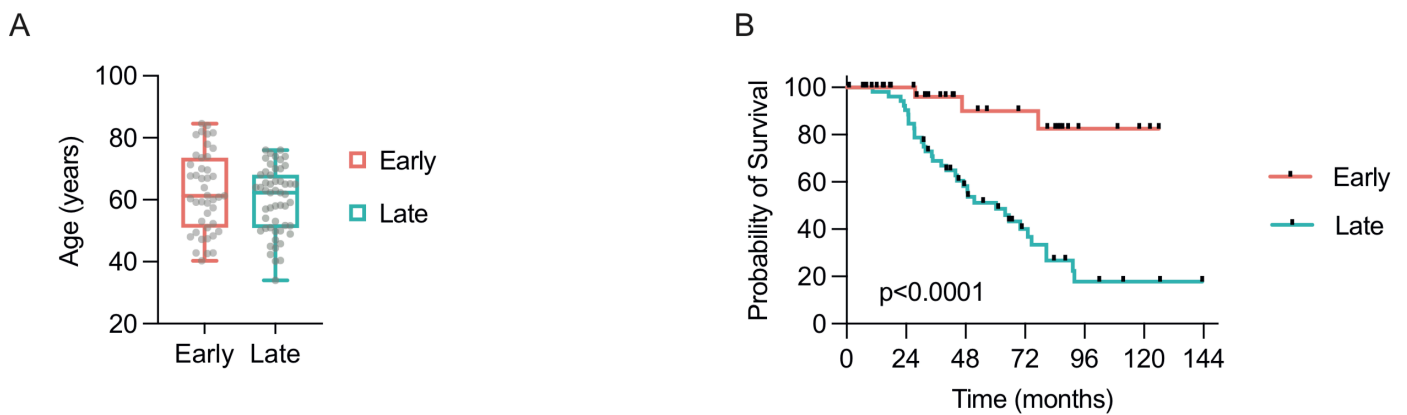


Figure 3.3 Clinical information on early and late-stage cohorts
(A) Age at Diagnosis. Median 61.3 years (early stage), 62.3 years (late stage). $p = \text{NS}$
(B) Overall survival. Median 60.3 months for late stage, and not reached for early stage. Log-rank Hazard Ratio 0.11 (95%CI 0.06-0.2), $p < 0.0001$ (Log-rank Mantel-Cox).

3.3.2 SNV/indel mutation rates comparison between early stage and late stage cohort

We designed a custom panel of genes using Illumina's online basespace sequence hub to include the coding regions (CDS) of the following genes: *PTEN*, *RBI*, *BRCA2*, *RAD51B*, *FANCM*, *PALB2*, *RAD51D*, *TP53*, *RAD51C*, *BRIP1*, *CDK12*, *NF1*, *BRCA1* and *BARD1*. Amplicons covered these regions of interest genes. For *KRAS*, we selected hotspot mutation G12 and mutation E545 and H1047 for *PIK3CA*. Amplicon libraries were made from early stage samples and sequencing was performed on the HiSeq4000 Sequencing System (Illumina, Cambridge, UK) using paired-end 150 bp protocols described in Chapter 2. For the late stage cohort, the same genes were sequenced by ^[125] Tagged-Amplicon deep sequencing (TAM-seq) using a HiSeq2500 as described in Chapter 2.

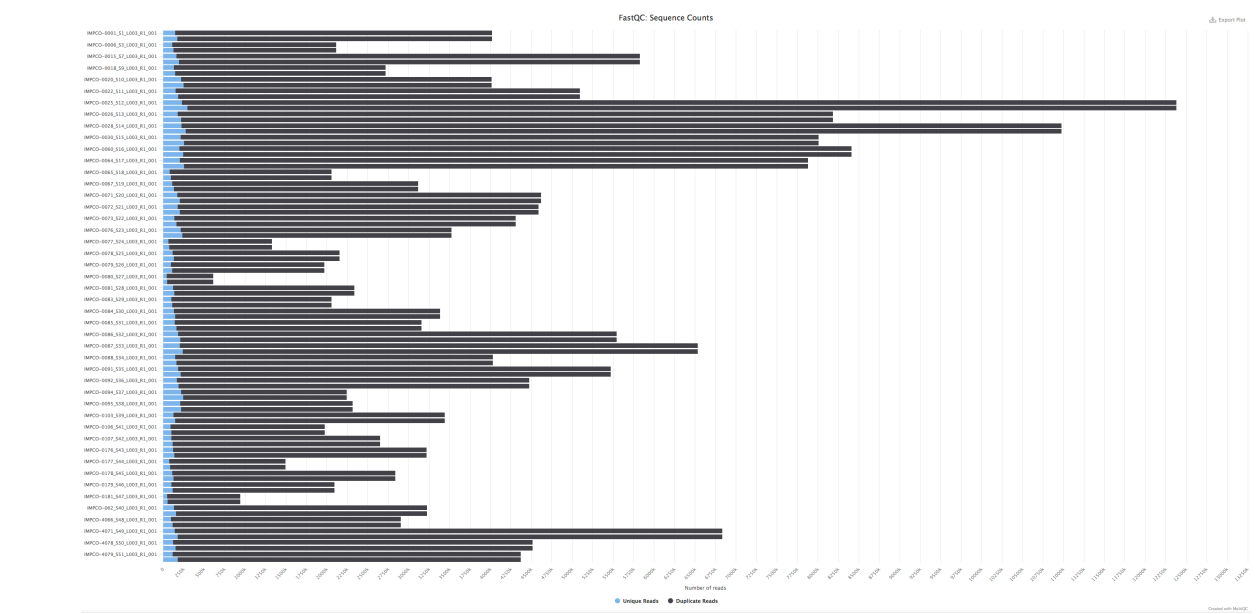
I used MultiQC to combine all the HTML web pages and generated a quality control (QC) report. The representative figures are shown in Figure 3.4. The properties of raw reads (sequence counts (Figure 3.4a), quality scores (Figure 3.4b), mean quality score (Figure 3.4c), adapter content (Figure 3.4d) are shown to assess data quality. Data were of acceptable quality with all the read scores were between 30 to 40 quality value, and all the properties were at reasonable variations. After MultiQC step, I applied the BWA (Burrows-Wheeler Alignment) tool to efficiently align short sequencing reads against the hg19 human genome, allowing mismatches and gaps^[110]. Adapter contamination was found at the 3' ends, and these were removed using Trimmomatic tool by illumina adapter (CTGTCTCTTATACACATCT). The representative multiQC report and adapter content before and after trimmomatic figures are shown in Figure 3.5.

After alignment with the reference genome hg19, BWA unified the result in a single mapping file in the SAM (Sequence Alignment Map) format^[126]. Then the generated SAM file was

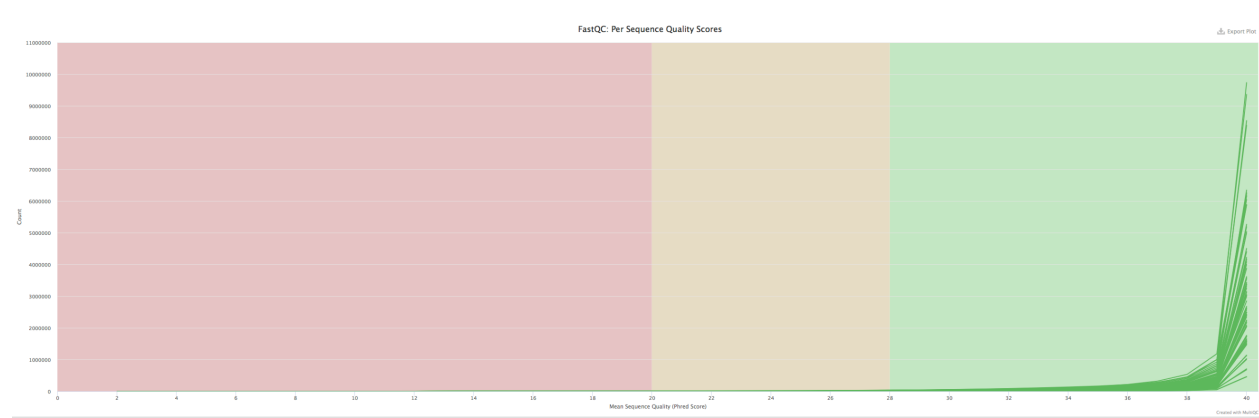
converted to binary Alignment/Map format (BAM) by SAMtools^[126]. BAM format is a much more stable file type and ensures efficient storing and access to alignment information.

SAMtools are used to sort the reads by chromosomal coordinates and merge the paired and unpaired BAM files to the same BAM sample. I then analysed short variants (SNV, indels) in both cohorts using GATK Mutect2 pipeline with “tumour-only” mode^[127]. Mr Hasan Mirza provided the analysis results by using Varscan2^[113], Strelka^[114] and HaplotypeCaller^[128] pipelines.

(a) FastQC: Sequence Counts



(b) FastQC: Mean Quality Scores



(c) FastQC: Per Sequence Quality Scores

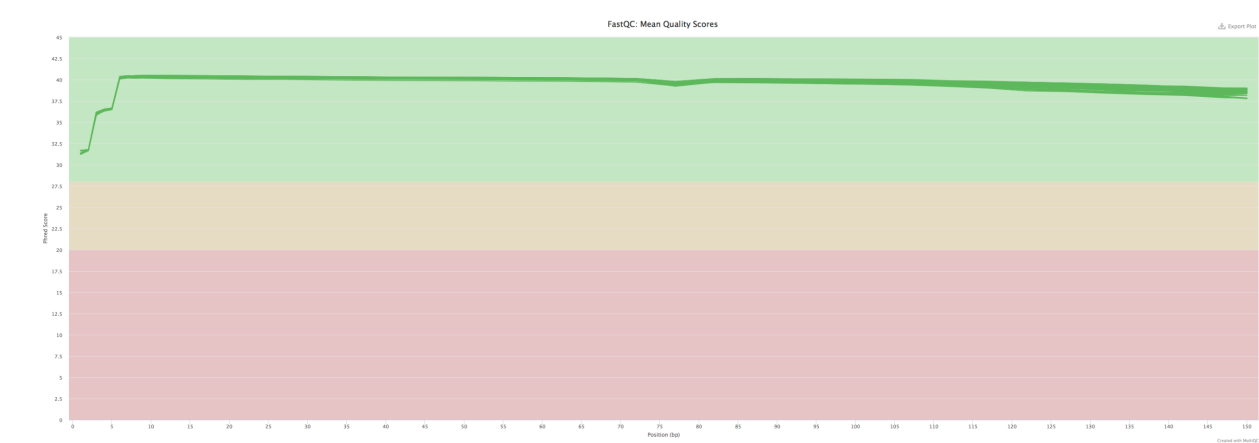
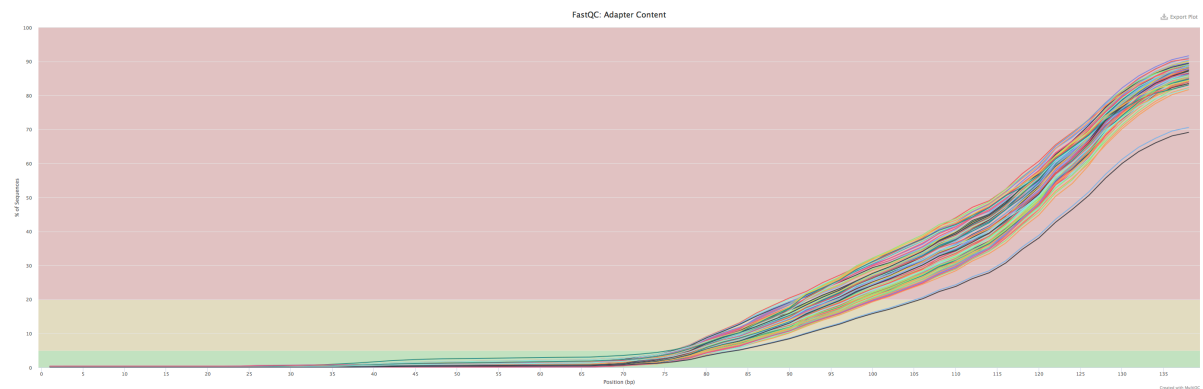


Figure 3.4 (a) Distribution of unique and duplicated reads of Ampliseq sequencing in early stage cohort samples. (b) Per sequence quality score shows the frequencies of quality scores in a sample. The quality of reads is good quality with the peak shifts to the right side. (c) Per base sequence quality presents the quality score across all bases at each position in the reads.

(A) FastQC: Adapter Content before trimming



(B) FastQC: Adapter Content after trimming

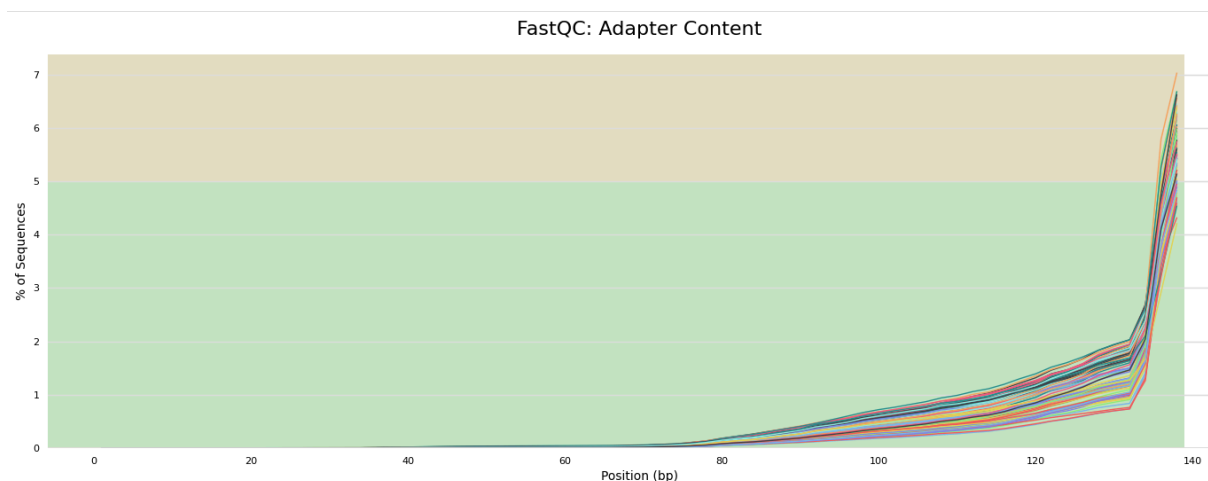


Figure 3.5 Ampliseq quality control and removal adapter contamination. The Figures show cumulative percentage count of the library which has been contaminated by adapter. (A) adapter content before trimming (B) after trimming. This step is performed by Trommatic tool^[129].

Mutations in *TP53* were near-universal (100% late-stage: 97/97 samples, 51/51 patients. 92% early-stage: 45/49 samples, 35/39 patients). The oncoprints showing the distribution of mutation rates in early and late stage samples and patients individually are presented in Figure 3.6 and Figure 3.7 respectively. The four early-stage samples in which *TP53*

mutations were not identified underwent pathology re-review; all were confirmed to be HGSC. Two had copy number profiles consistent with HGSC, whilst two showed no CN abnormalities, suggesting very low tumour cellularity (Figure 3.8). There was no difference in the rates of mutations in the other analysed genes. Specifically, the rates of pathological mutations in *BRCA1* and *BRCA2* were 10% and 2%, respectively, in the early-stage and 14% and 2% in the late-stage. These combined *BRCA1/2* mutation rates in the late stage cohort are very consistent with combined germline and somatic mutation rates of 16.7 – 28% reported in previous studies^[130].

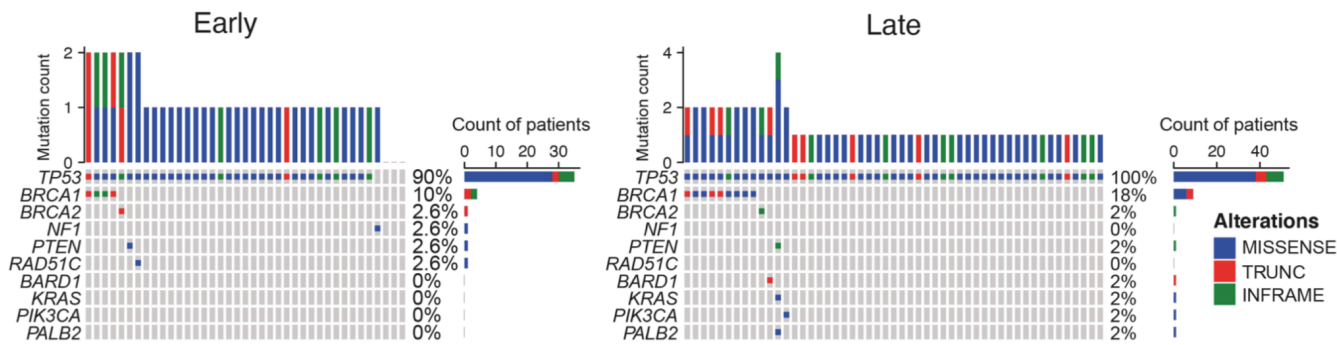


Figure 3.6 Short variants (SNV and indels) for each patient in early-stage and late-stage cohorts. The upper plot shows the counts of mutations per patient.

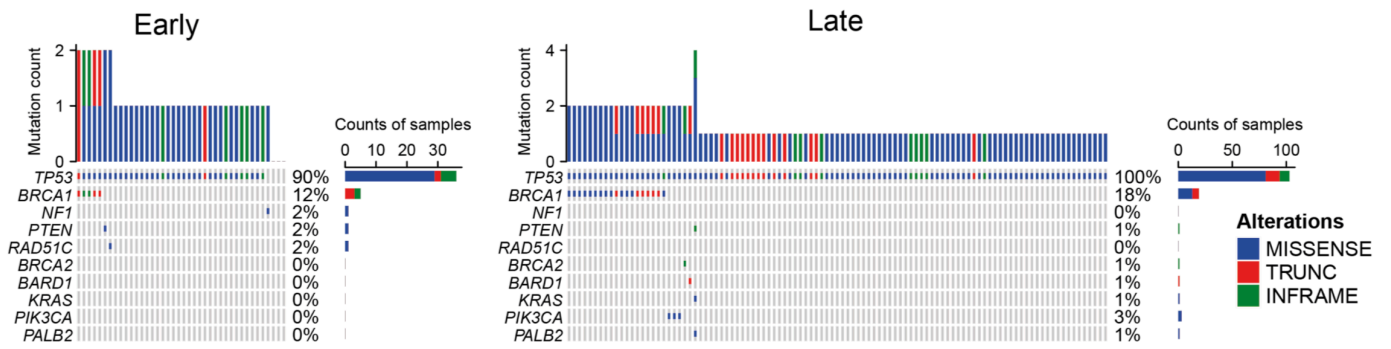


Figure 3.7 Counts of mutations in each sample in early-stage and late-stage cohorts. Mutations include somatic and germline variants. Germline mutations are only available in late stage cohort. The upper plot shows the number of mutations in per tumour sample.

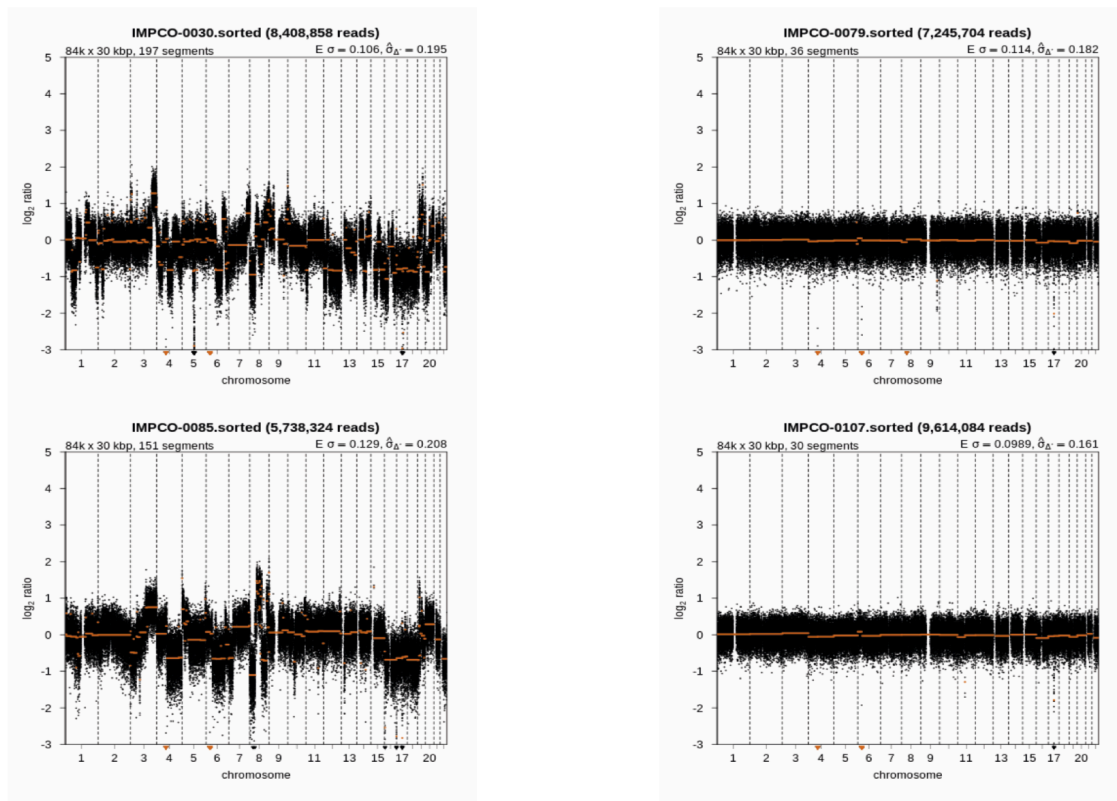


Figure 3.8 Relative copy number profiles of early stage four samples without *TP53* mutations. Raw bin values are shown in black and segmented copy number values are shown in red.

3.3.3 *TP53* hotspots mutations in early stage and late stage cohort

Around 50-60 % of all human cancers contain homozygous mutations in *TP53* ^[131].

According to COSMIC database, the ten most common mutated codons of *TP53* are R175, R196, R213, Y220, G245, R248, R273, R282, R306 and R342^[132]. Eight of these account for ~28% of the total. These alleles appear to be selected preferentially in human cancers of many tissue types. Previous research identified hotspot mutations at codons R273, R248 and R175, constituting 19.7% of all *TP53* mutations in HGSC^[133]. A recent study investigated the clinical relevance of *TP53* hotspot mutations in HGSC, and found that R273 (20.6%, 26/126), R248 (16.7%, 21/126), R175 (14.3%, 18/126), Y220 (9.5%, 12/126) were top four mutations in TCGA HGSC dataset^[59]. Therefore, I choose these four codons as my key hotspots to investigate the mutation frequency in early and late stage cohorts.

I observed 15 hotspot mutations in 40 early stage samples and 13 in 102 late stage samples.

The mutation mapper plot of *TP53* in early stage and late stage cohorts is shown in Figure 3.9. The frequency of four key *TP53* hotspot mutations was significantly greater in the early-stage cohort than late (Figure 3.10. Fisher's exact test; $p=0.0075$).

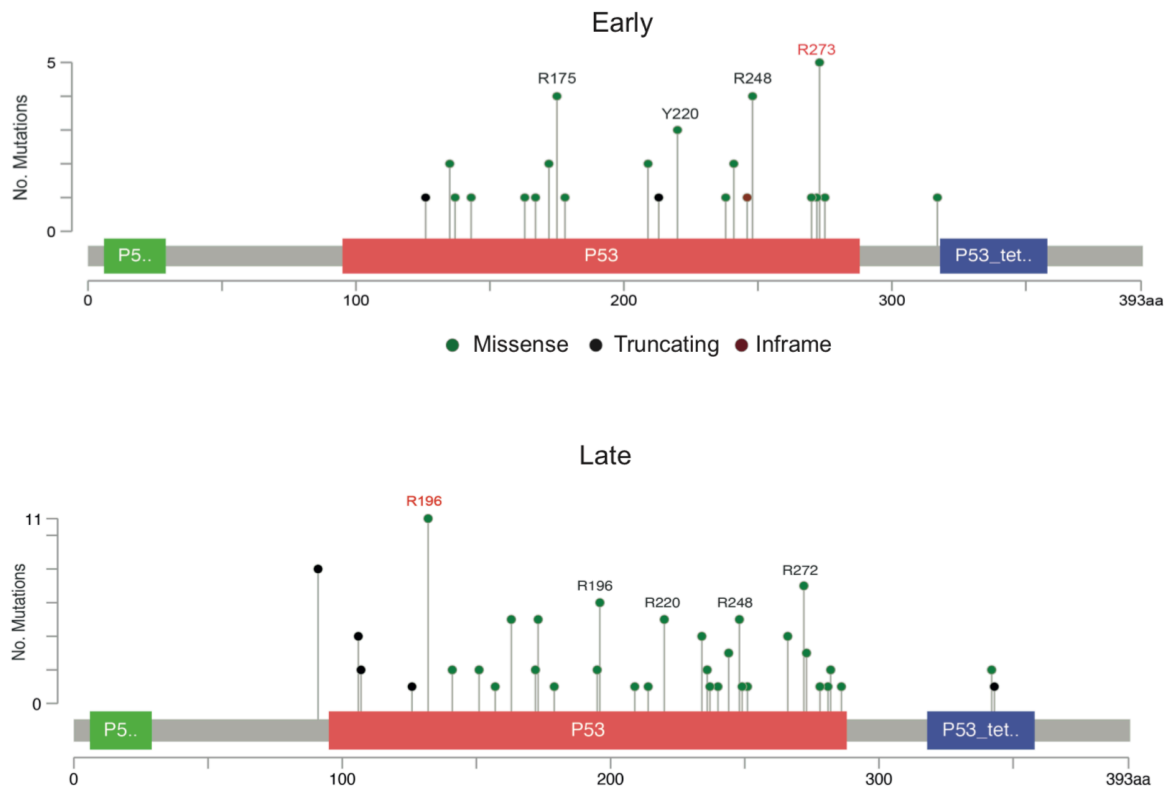


Figure 3.9 Gene mutation mapper plot of p53 in early-stage cohort and (E) late-stage cohort. Key hotspot residues are marked. The commonest residue mutations in each cohort are marked in red.

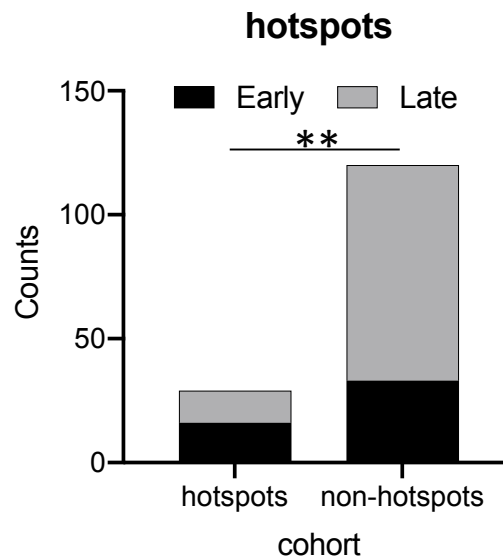
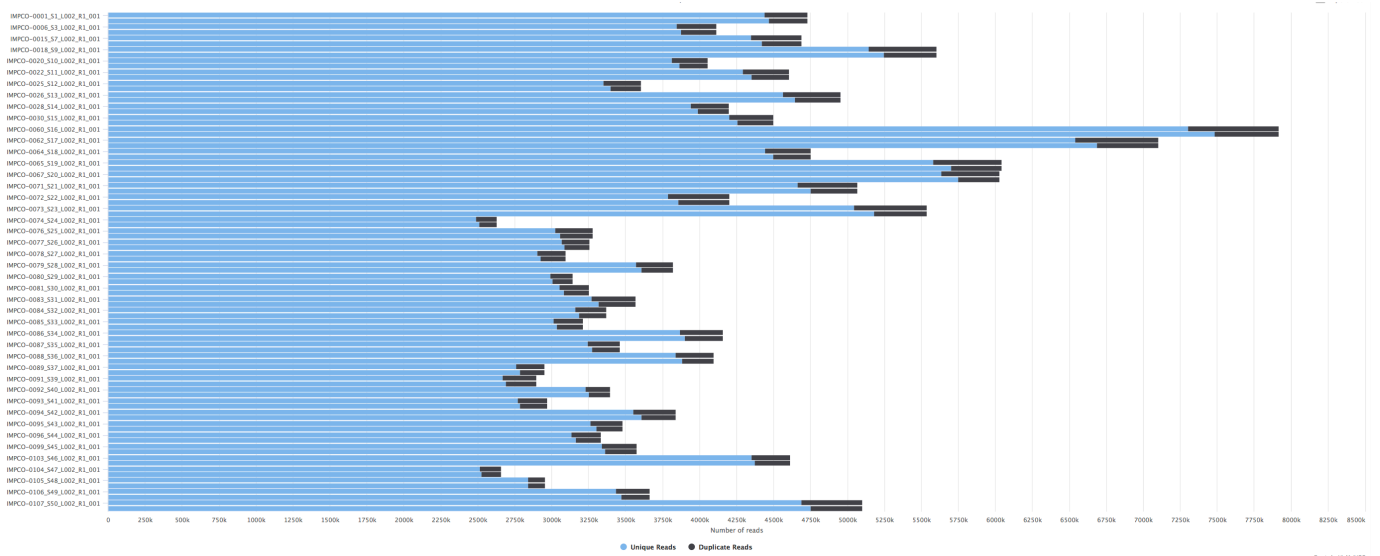


Figure 3.10 Hotspot mutations comparison between early stage and late stage cohorts. Fisher's exact test; $p=0.0075$

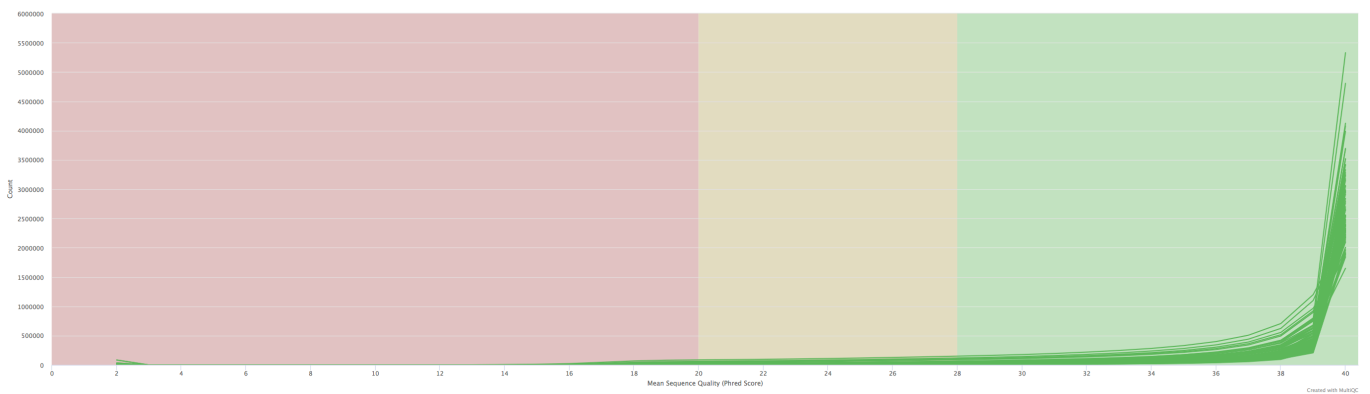
3.3.4 Relative copy number profile of early stage and late stage cohort samples by QDNAseq

Following shallow WGS (sWGS), sequencing data were uploaded into the FastQC package, generating an HTML web page with different graphs of detailed QC results. MultiQC combined all the HTML web pages, which generated a QC report. The properties of raw reads (sequence counts, quality scores, mean quality score, adapter content etc.) are shown to assess data quality^[134]; The representative figures are shown in Figure 3.11. The data were of good quality based on the QC scores and per base sequence quality, which presents the quality score across all bases (Figure 3.11b). Acceptable quality control scores were between 30 to 40 quality values and all the properties were at reasonable variations. After MultiQC was performed, the BWA (Burrows-Wheeler Alignment) pipeline was applied to efficiently align short sequencing reads against the hg19 human genome, allowing for mismatches and gaps^[110]. As with the Ampliseq analysis steps, sorted BAM files were also derived from sWGS data. The package QDNAseq uses BAM files as input and outputs as graphs that represent read counts per bin, which have been corrected, filtered, normalised, and optionally log2-transformed in 30kb bins^[117]

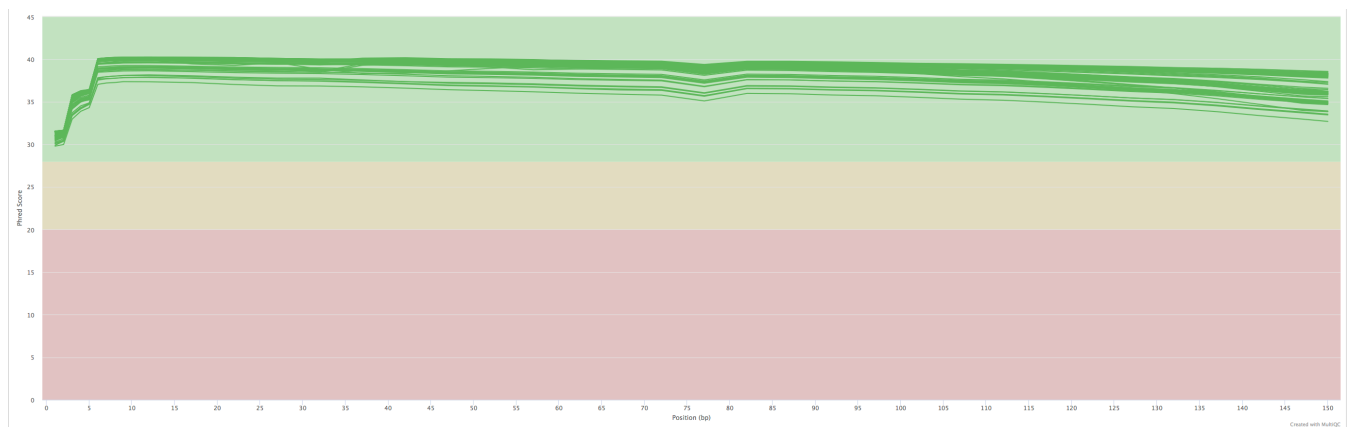
(a) FastQC: Sequence Counts



(b) FastQC: Per Sequence Quality Scores



(c) FastQC: Mean Quality Scores



(d) FastQC: Adapter Content

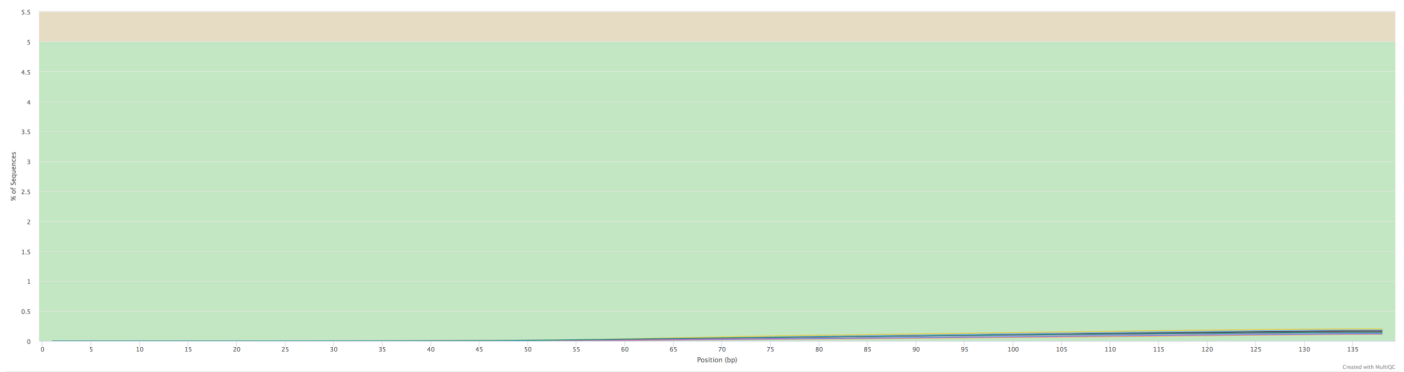


Figure 3.11 (a) Distribution of unique and duplicated reads in early stage cohort samples. (b) Per sequence quality score shows the frequencies of quality scores in a sample. The quality of reads is good quality with the peak shifts to the right side. (c) Per base sequence quality presents the quality score across all bases at each position in the reads. (d) Adapter content is essentially 0% in all the samples.

3.3.5 Absolute copy number inference of early stage and late stage cohort samples using sWGS-absolute and ACE packages

Two packages, sWGS-absoluteCN(swgs) and ACE, were used to calculate the absolute copy number of the early stage and late stage cohorts from sWGS data as previously described in chapter 2. The swgs package implements a method for generating absolute copy number profiles utilising a multi-stage fitting process to create a set of read depth-normalised absolute copy numbers profiles. This package is kindly provided by Dr Philip Smith in Cambridge University. Copy number fits are generated in read count space using a modified implementation of QDNAseq on full read depth shallow whole genome BAM files. After this, a grid search of ploidy and purity values is performed to generate a matrix of potential absolute copy number profiles. The fitting of absolute copy number profiles includes a fit error function and a variant allele fraction anchoring process, which assist in selecting an optimal copy number fit.

Minimisation of the clonality error function, expected-to-experimental *TP53* distance, and

manual review of fits allow for selecting an optimal absolute copy number profile for a given sample. Selected fits are subject to power calculation to detect copy number alterations. The read depth of a given sample is assessed against its selected ploidy-purity combination. This process determines if 1) a sample has a sufficient number of reads to support the selected ploidy-purity combination and 2) an optimal target number of reads to perform sample-specific read downsampling. This process normalises the read depth between samples in ploidy and purity-dependent manner. Read variance across segments is consistent, while excluding samples that are not supported by a sufficient number of reads. Samples passing all filtering criteria undergo read-down sampling to the specified target number of reads determined by the previous steps and absolute copy number profiles fitted at the ploidy-purity combination selected prior.

I also used ACE^[118], which calculates the ploidy and purity of each sample as a validation of swgs pipeline with the settings "penalty=0.5" and "penploidy=0.5" and bin size=30kb. The workflow of analysis absolute copy number in swgs and ACE pipelines are shown in Figure 3.12. The results of comparing purity and ploidy from sWGS-absolute and ACE packages are presented in Figure 3.13. The ploidy correlation R is 0.94 between ACE and sWGS-absolute packages, and the p-value is 1e-44. The purity correlation rho is 0.81 between ACE and sWGS-absolute packages, and the p-value is 6e-23. These correlation results indicate that the sWGS-absolute pipeline performs well in inferring purity and ploidy of tumour samples compared to the ACE pipeline. All the absolute copy number figures of early stage and late stage cohorts from sWGS-absolute pipelines are shown in Appendix 8 and Appendix 9.

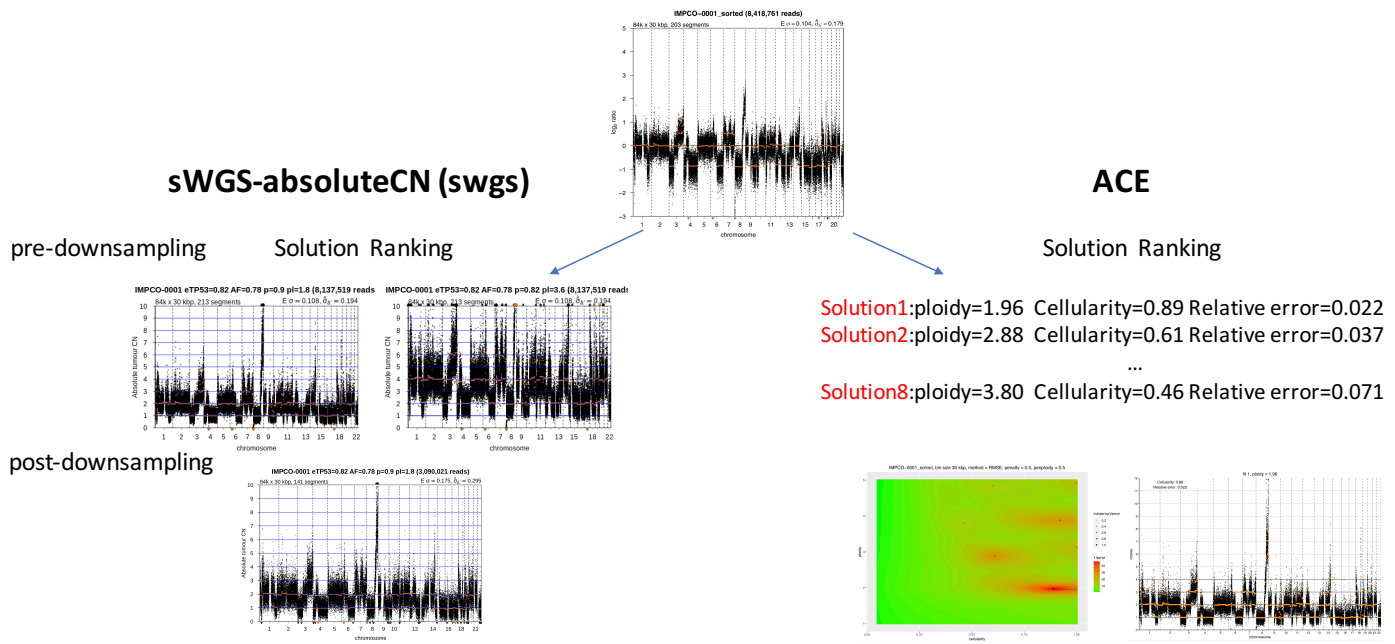


Figure 3.12 Workflow analyses absolute copy number data by swgs and ACE packages.

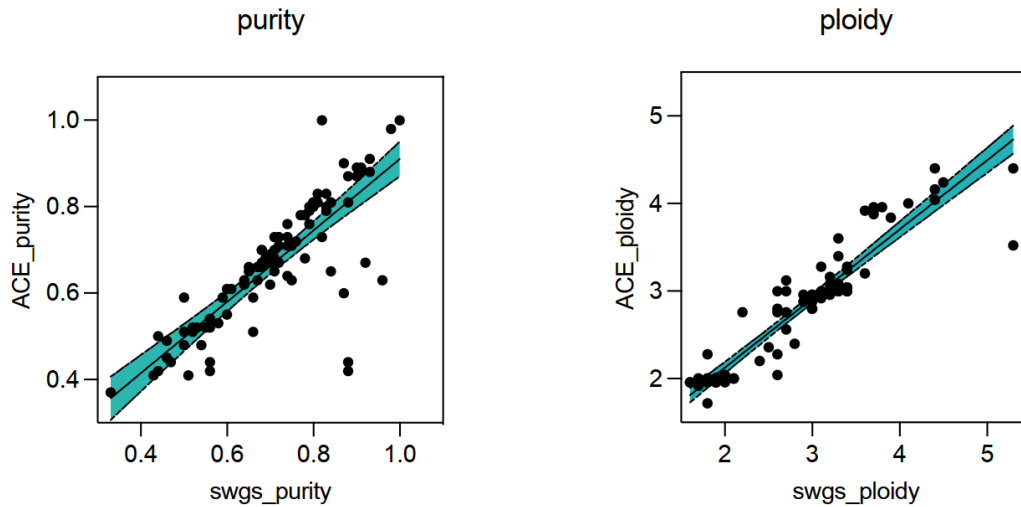


Figure 3.13. Ploidy and purity comparison between sWGS-absolute and ACE pipelines. Spearman rank correlation test, $p < 0.0001$ for both correlations.

3.3.6 Purity and ploidy comparison of early-stage and late-stage cohorts

I then compared the purity and ploidy between early stage and late stage cohorts. The ploidy of cancer cells refers to the number of sets of chromosome amount of DNA they contain.

Purity is defined as the proportion of cancer cells in the tumour tissue^[135]. In our results, we show that there was no statistically significant difference in purity between the cohorts (Fig. 3.14a), but median ploidy was significantly greater in late-stage samples compared to early (Figure 3.14b). Median early 1.9; median late 3.0; $p < 0.0001$, Mann-Whitney test).

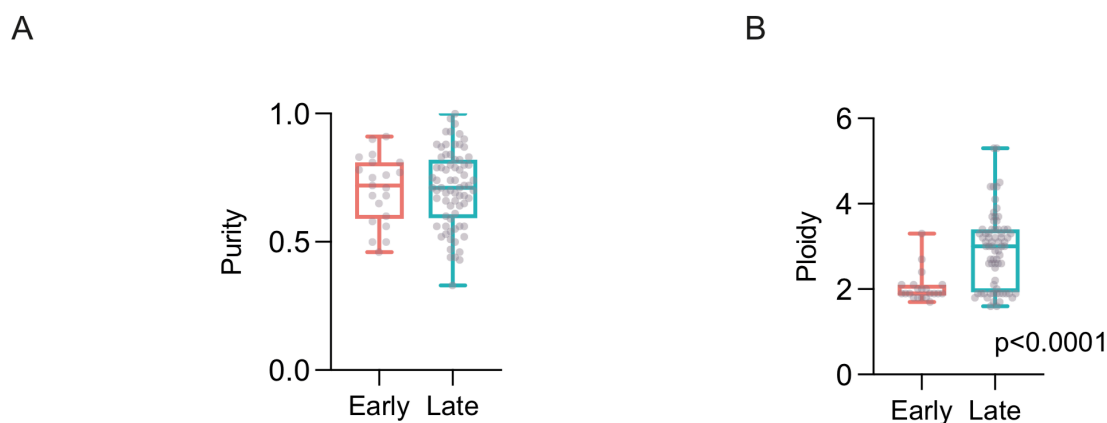


Figure 3.14 (a) Purity comparison of early-stage and late-stage cohorts. (b) Ploidy comparison of early-stage and late-stage cohorts; Mann-Whitney test.

3.3.7 Exploring focal gene amplifications and deletions in early stage and late stage cohort

Global copy number gains/losses are shown in Figure 3.15 - there were generally more gains and amplifications in late-stage samples, and regions of CN loss in the early-stage samples, in keeping with the differential ploidy between the cohorts. There were several regions of differential gain in the late-stage cohort (e.g. chromosome 4, 6, 9, 11, 12) and losses in the early-stage cohort (e.g. chromosome 4, 9, 12, 17). However, there were no regions that were uniquely amplified or deleted in either cohort.

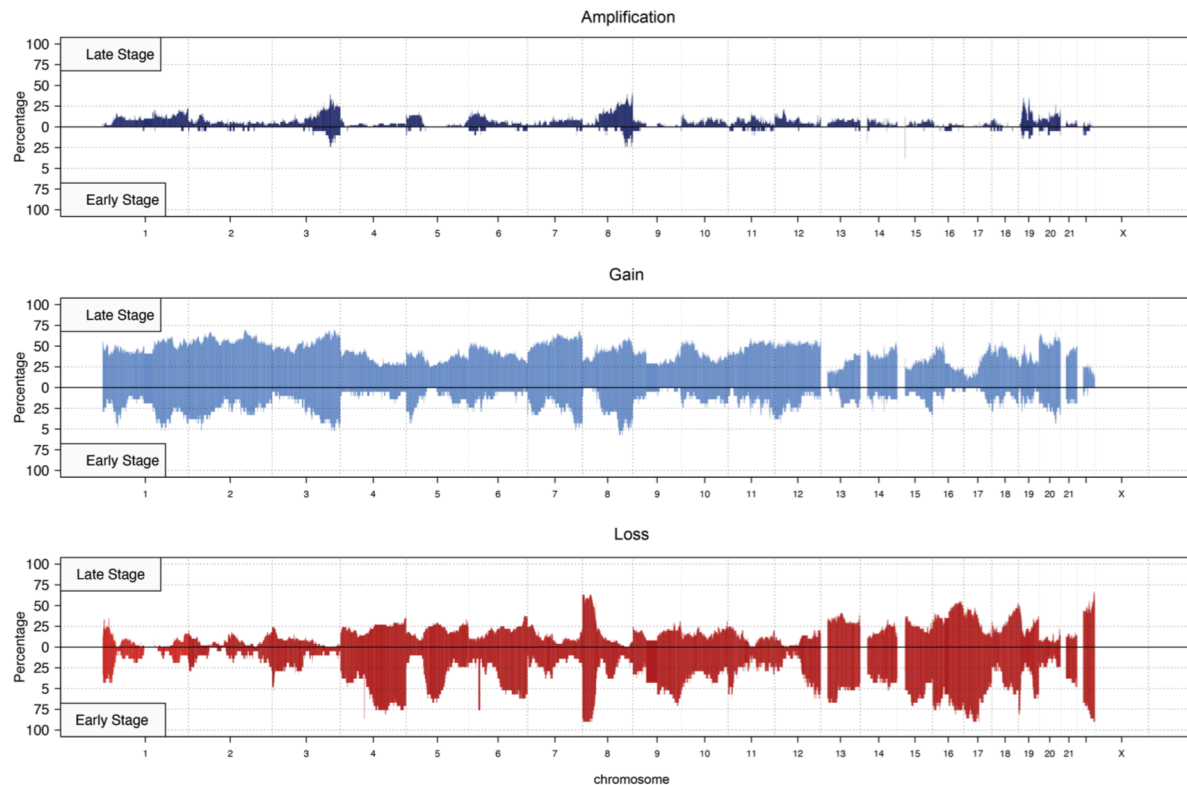


Figure 3.15. Global copy number amplifications, gains and losses in early-stage and late-stage cohorts.

3.3.8 Focal genes amplification and deletions rate between early stage and late stage cohort

Widespread copy number alteration plays a vital role in HGSC: indeed, HGSC is the archetypal C class, copy number-driven malignancy^[83]. The most recurrently amplified genes such as *MYC*, *MECOM*, *PIK3CA* and *CCNE1*^[124] are all essential for cancer cell growth. *CCNE1* amplification is associated with poor response to platinum-based chemotherapy^[136]. Commonly deleted genes are *NF1* and *PTEN*. NF1 defects lead to Ras pathway activation^[91], whilst PTEN inhibits cell proliferation induced by the PI3K pathway^[137]. PTEN loss defines a subgroup in HGSC with a worse prognosis^[138].

Focal amplifications and deletions were defined according to the COSMIC definitions (see <https://cancer.sanger.ac.uk/cosmic/help/cnv/overview>): amplification was defined as total

copy number ≥ 5 if average ploidy ≤ 2.7 , or ≥ 9 if average ploidy > 2.7 . Loss was defined as total copy number 0 if average ploidy ≤ 2.7 or (ploidy minus 2.7) if average ploidy > 2.7 .

I compared the absolute copy number of *AKT1*, *AKT2*, *AKT3*, *PIK3CA*, *PTEN*, *TP53*, *CCND1*, *CCND2*, *CCND3*, *CDKN2A*, *CDKN2B*, *RBI*, *CCNE1*, *MYC*, *MECOM*, *KRAS* and *NF1*. The commonest amplifications were in *MYC* (25% in the early-stage and 19% in the late-stage) and *MECOM* (20% in the early-stage and 14% in the late-stage), with no significant difference between the cohorts. There were also no difference in amplification and deletion rates of the other genes (Figure 3.16). The absolute copy number of *AKT1*, *AKT3*, *PIK3CA*, *CCND1*, *CCND2*, *CCND3*, *CDKN2A*, *CDKN2B*, *MYC*, *MECOM*, *KRAS* and *NF1* was greater in the late stage cohort (Figure 3.17), which is consistent with the ploidy increase in late stage samples, while amplification/deletion rates are no significant difference in the two cohorts.

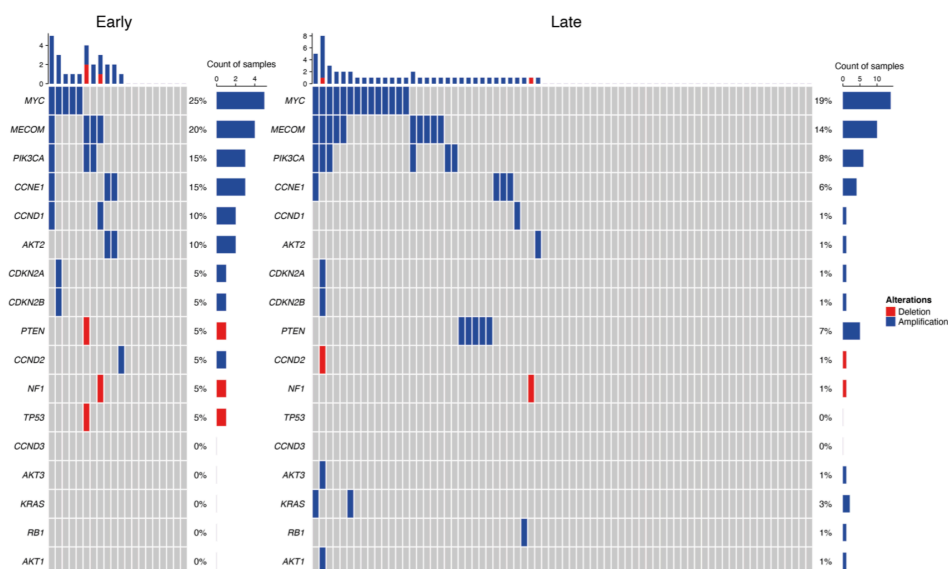


Figure 3.16 Estimation of focal amplifications and deletions in 17 genes of interest, determined by sWGS. The upper plot shows the number of amplifications and deletions in each tumour sample.

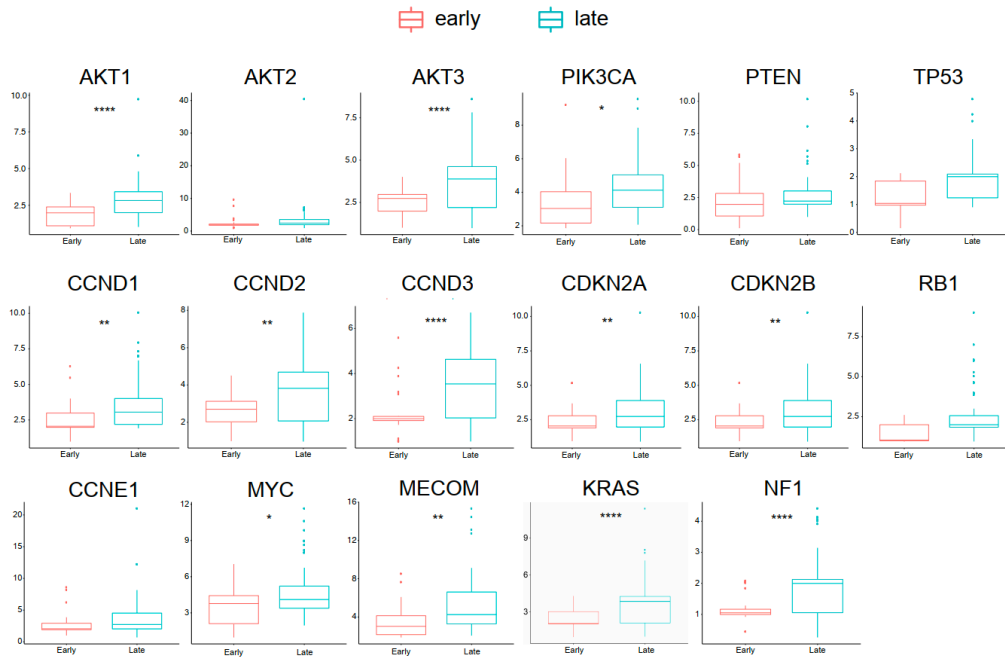


Figure 3.17 Absolute copy number (y-axis) of 17 individual genes in early and late- stage samples. *, $p<0.05$, **, $p<0.01$, ***, $p<0.001$, ****, $p<0.0001$.

3.4 Discussion

The analysis described in this chapter shows no significant differences in somatic mutations or focal copy number alterations between early-stage and late-stage cohorts. However, I did identify that ploidy is significantly higher in late stage than early stage cohort.

3.4.1 Mutation rates and focal genes amplification and deletions rates between early stage and late stage cohort

As expected, we found that *TP53* mutations were near-universal in both early stage and late stage cohorts (Figure 3.6), in keeping with previous studies^[62]. However, four early-stage samples were *TP53* wild type (Figure 3.8), possibly due to a combination of low tumour cellularity and poor DNA quality from FFPE preservation. A previous study of 16 early-stage HGSC cases^[96] also identified that *TP53* mutations were very frequent but not universal. Two of our *TP53* wild type samples certainly had CN profiles that were highly consistent with

HGSC. Together, these data suggest that a small proportion of early stage HGSC cases may be truly *TP53* wild type.

The rate of missense *TP53* mutations here was higher (75%) than in previous HGSC cohorts^[50], and hotspot mutations (defined here as mutation at the four most commonly mutated codons, R248, R273, R175 and Y220) were more prevalent in our early-stage cohort than late-stage. However, a recent analysis of nearly 800 HGSC cases (78 stages I/II, 709 stages III/IV) has suggested no overall difference in mutation type between early and late samples^[59].

Crucially, I found no differences in the rates of *BRCA1/2* mutations between early and late cohorts. The absence of germline DNA from most of our early-stage cohort meant that I could not verify the germline mutation status. Still, our *BRCA1/2* mutation results are broadly in line with previous cohorts^[139] and reflect that these samples were obtained from routine practice rather than risk-reducing surgery.

When examining global copy number change, there were no regions uniquely lost or gained in either cohort, suggesting that the process of dissemination is not driven by specific amplifications or deletions, whilst I also found no significant difference in copy number of 17 key genes. Together, the SNV/indel and focal CNA data corroborate findings by Köbel et al. that early and late-stage HGSC appear identical by immunohistochemistry^[140]. The previous WGS analysis of early stage HGSC^[96] also identified high levels of genomic instability with few, if any, recurrent focal differences between early and late-stage HGSC and no unique mutation or focal CN alterations in the early-stage patients.

3.4.2 Differential purity and ploidy between early stage and late stage cohort

My most striking observation was the difference in overall ploidy between early and late stage cases. Unlike focal gene copy number alterations, which only resulted from a single or a small number of genes changes, leading to specific phenotypes^[141], ploidy change causes a total number of sets of chromosome alteration, and the origin of ploidy phenotypes are more complex.

One mechanism to generate high ploidy in cancer is via whole genome duplication (WGD). WGD generates polyploid cells possessing more than two copies of the genome and is among the most common genetic abnormalities in cancer^[142]. WGD has been described in many solid malignancies^[143] and is generally associated with poor prognosis^[144]. It is thought to arise from aberrant cell division^[145] and potentially may mitigate the effects of mutations that would otherwise be deleterious, thus preventing cancer cell attrition^[146]. In a pan-cancer analysis, HGSC has been shown to have one of the highest rates of WGD at approximately 40%^[143]. An increase in ploidy in the late stage cohort suggests that WGD could promote the survival and metastatic spread of HGSC. The relationship between copy number genomic changes and whole genome duplication in late stage HGSC will be explored in the next chapter.

3.5 Summary

In summary, I collected samples from 43 patients with stage IA-IIA HGSC to identify potential differences in short genomic variants and copy number aberrations and compared them to a cohort of 52 late-stage (stage IIIC-IV) cases to determine whether early-stage ovarian high grade serous carcinoma (HGSC) represents a distinct genomic entity. No significant differences were identified in our selected panel of somatic mutations or focal

copy number alterations between early-stage and late-stage cohorts, and there were no unique regions lost or gained in either cohort in the genome-wide copy number analysis.

Interestingly, the *TP53* hotspot mutation rate is significantly higher in early stage than late stage cohort, which needs further investigation for a larger sample set. In addition, the ploidy is markedly higher in late stage than early stage cohort, while there is no difference in purity between these two cohorts. Overall, the higher ploidy in late stage cohort could provide evidence for the tumour progression in HGSC.

CHAPTER 4. Copy number signatures in early stage high grade serous carcinoma

4.1 Introduction

This chapter focuses on using copy number features and copy number signatures to identify possible differences in the genomic profile of late stage HGSC cohort compared to early stage cohort. In the previous chapter, I showed that there is no significant difference in SNV and focal copy number alterations between early and late stage cohorts. Using the input data from absolute copy number, further analysis of copy number features between the two cohorts and copy number signatures provide more evidence about the ploidy changes between the two cohorts.

4.2 Background

Genome instability is dynamic and prevalent in high grade serous carcinoma^[147].

Chromosomal rearrangements result in the gain or loss of multiple chromosomal locations and translocations in the HGSC development and will cause specific copy number features across the whole genome. Six fundamental copy number features were universal in genomics: the breakpoint count per 10MB, the copy-number of segments, the breakpoint count per chromosome arm, copy number change point, the lengths of oscillating CN segment chains and the size of segments^[1].

A breakpoint is the point or region of breakage when chromosomes break^[148]. Breakpoint count per 10MB indicates the breakpoint counts were calculated for the window of 10 megabases length on the chromosome. Similarly, breakpoint count per arm represents the number of breakpoint counts derived from each chromosome arm. Copy number change point refers to the difference in CN between adjacent segments. As a result of the massive DNA alterations occurring, chromosomes affected by chromothripsis show a characteristic pattern of copy-number "oscillations," whereby typically only two (or occasionally three)

copy-number states are detectable along the chromosome in the context of a large number of rearrangements^[149]. Chromosomal rearrangements determine segment sizes within the genome. The mean length of segments was used to estimate the number of chromosomal rearrangements within cells^[150]. Copy number of segments represents the copy number state of the segments on a chromosome. Each copy-number segment was produced using the CBS (Circular Binary Segmentation) algorithm^[151] by the genomic extents of the segment (chromosome, start, and end), the number of SNP6 probes contained within that segment, and the estimated mean copy-number value for that segment.

Previous work from the host laboratory and collaborators used absolute copy number data from 117 BriTROC cohort samples to calculate distributions of these six copy number features. The distribution of copy number features may be calculated by either Poisson or Gaussian distributions^[1]. Poisson distribution is a theoretical discrete probability known as the Poisson distribution probability mass function. It is used to find the probability of an independent event occurring in a fixed interval of time and has a constant mean rate. Using Poisson distribution, they identified the total counts' distribution in each optimal number of components in breakpoint per 10MB, oscillating copy number, breakpoint per arm.

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. Similarly, we used Gaussian distributions to calculate the copy number distribution, copy number change point and segment size in each cohort. Selecting the number of components for the mixture model from each feature is an important issue, and determining the appropriate number of components is critical for adequately representing the empirical distribution characteristics of the data^[152]. Too many components indicate the mixed model may over-fit the observed data and restrict

the application of the components. On the other hand, too few components may not be flexible enough to reflect the actual characteristics of the distribution^[153].

Using R package flexmix, Macintyre et al^[1] employed an expectation-maximisation (EM) algorithm that uses maximum likelihood to estimate the model parameters and calculates the probability of belonging to each component for every observation. They determined the component number of each feature through the package to check the model "goodness-of-fit" properties by Bayesian Information Criterion (BIC)^[154]. They got 3 components for breakpoint count per 10Mb feature, 8 components for copy number feature, 7 components for copy number change point feature, 5 components for breakpoint count per chromosome arm feature, 3 components for length of chains of oscillating copy number feature and segment size feature has 10 components. In total, this yields 36 components.

Once the components were identified, they generated a sample-by-component matrix representing the sum of posterior probabilities of each copy-number event assigned to each component^[1]. They used the Non-negative matrix factorisation (NMF) package to factorise the sample-by-component matrix into a signature-by-sample matrix and component-by-signature matrix. NMF has been widely used as a general method for high dimensional reduction and feature extraction on non-negative data^[155]. Previous studies suggest that there should be a minimum of roughly four detectable signatures representing mutagenic processes that give rise to distinct patterns of copy-number change: breakage-fusion-bridge, duplicator phenotype, chromothripsis, and amplifier phenotype. Their ability to detect these depends on whether each pattern is sufficiently represented in our data. As NMF demands the upper bound on the number of signatures be far less than both the sample and component number, they chose a signature search interval between 3-12. They ran the matrix factorisation 1000

times for each number of signatures, each with a different random seed, and compared the cophenetic, dispersion, silhouette, and sparseness scores for the signature-feature matrix (basis), patient-signature matrix (coefficients) and a consensus matrix of patient-by-patient across the 1000 runs. In addition, they performed 1000 random shuffles of the input matrix to get a null estimate of each of the scores^[1]. The results of cophenetic, dispersion, silhouette, sparseness analysis show that the optimal number of signatures is 7. A signature number higher than this would force the sparseness in the signature by feature matrix (basis) to be greater than that which could be obtained by randomly shuffling the input matrix.

After performing NMF, they got two matrixes which consist by tumour-by-signature matrix and signature-by-component matrix. They used an R package called *yet another package for signature analysis* (YAPSA) to deconvolute each copy number signature contributions to tumour genome with LCD (linear combination decomposition) function^[156]. The different compositions of components represent specific pathways and gene changes as I described in chapter 1. Therefore, each signature has a specific and distinct pattern of the 36 components.

In this chapter, I will investigate the copy number features, 36 components, and 7 copy number signatures exposures in early and late stage cohorts.

4.3 Results

4.3.1 Copy number signatures analysis of early stage and late stage cohort

The absolute copy numbers of early stage and late stage cohort were obtained by sWGS and allele frequency of *TP53* was obtained as described in Chapter 3. I used absolute copy numbers as input data and applied mixture modelling to separate the copy number feature distributions. The distributions of the six specific CN features - segment length, segment copy number, number of breakpoints per chromosome arm, number of breakpoints per 10Mb, copy number change point and length of chains of oscillating copy number are shown in Figure 4.1.

The number of counts is significantly higher in the late stage than in the early stage cohort in each component. I overlapped the early and late stage distributions together and did further analysis by normalising six feature counts per patient in each cohort. I found more breakpoint per 10MB in late stage samples than early stage (Figure 4.1 and Figure 4.2). Late stage samples were significantly more likely to have 5 or more breakpoints per chromosome arm than early stage, and oscillating chain length was higher in late stage than early stage samples. Late stage samples also a higher copy number change point and copy number per segment than early stage, while early stage samples had a more extended segment size (Figure 4.2). In addition to the six copy number features, I also calculated the tumour-by-component matrix in early stage and late stage cohorts (Figure 4.3). The profiles of components in early stage and late stage are very similar, and the distribution of each component is consistent with copy number features analysis. Overall, these comparisons indicate that late stage cohort is more fragmented than early stage cohort.

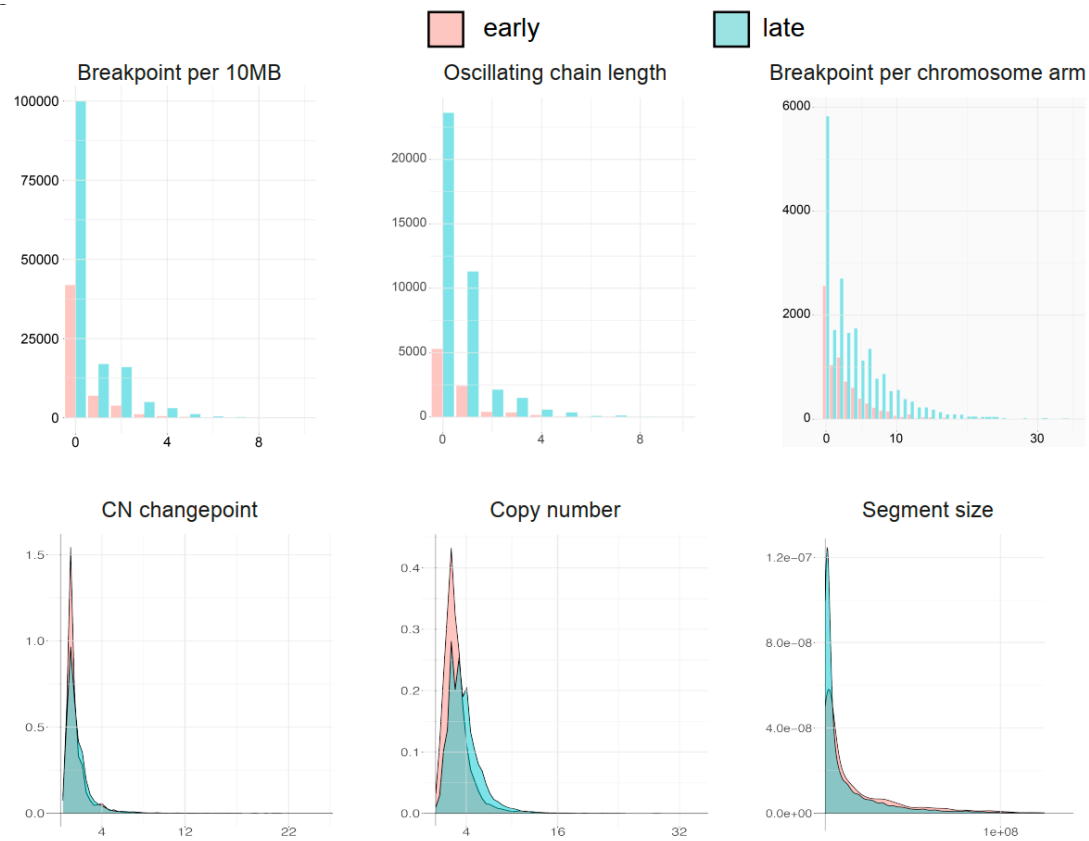


Figure 4.1 Copy number feature distributions in early and late-stage HGSC. Total count distributions in copy number features - Breakpoint per 10MB, Oscillating chain length and Breakpoint per chromosome arm by Poisson distribution analysis; Density distributions in copy number features - Breakpoint per 10MB, Oscillating chain length and Breakpoint per chromosome arm by Gaussian distribution analysis.

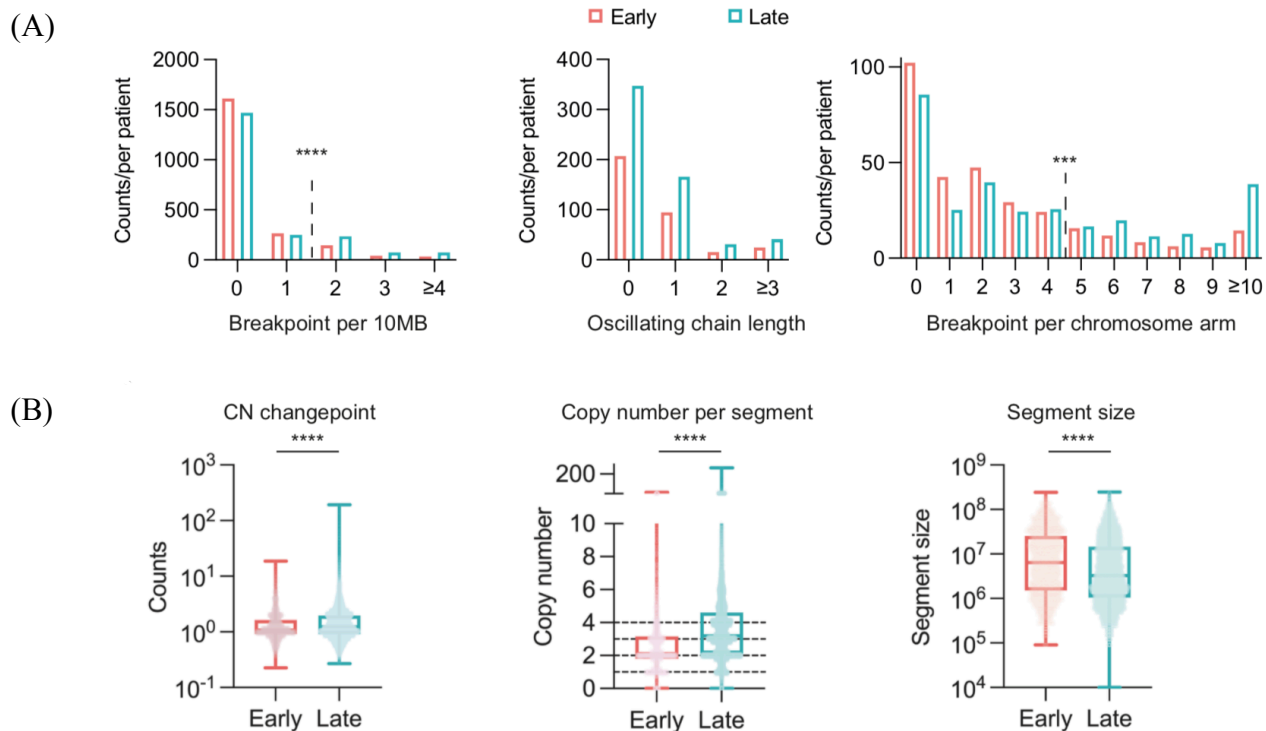


Figure 4.2 Components distribution and comparison between early stage and late stage cohorts. (A) Counts per patient in copy number features - Breakpoint per 10MB, Oscillating chain length and Breakpoint per chromosome arm; Fisher's exact test (B) Counts in copy number features - copy number change point per segment, absolute copy number per segment, segment size); Kolmogorov-Smirnov test ***; $p < 0.001$ ****; $p < 0.0001$

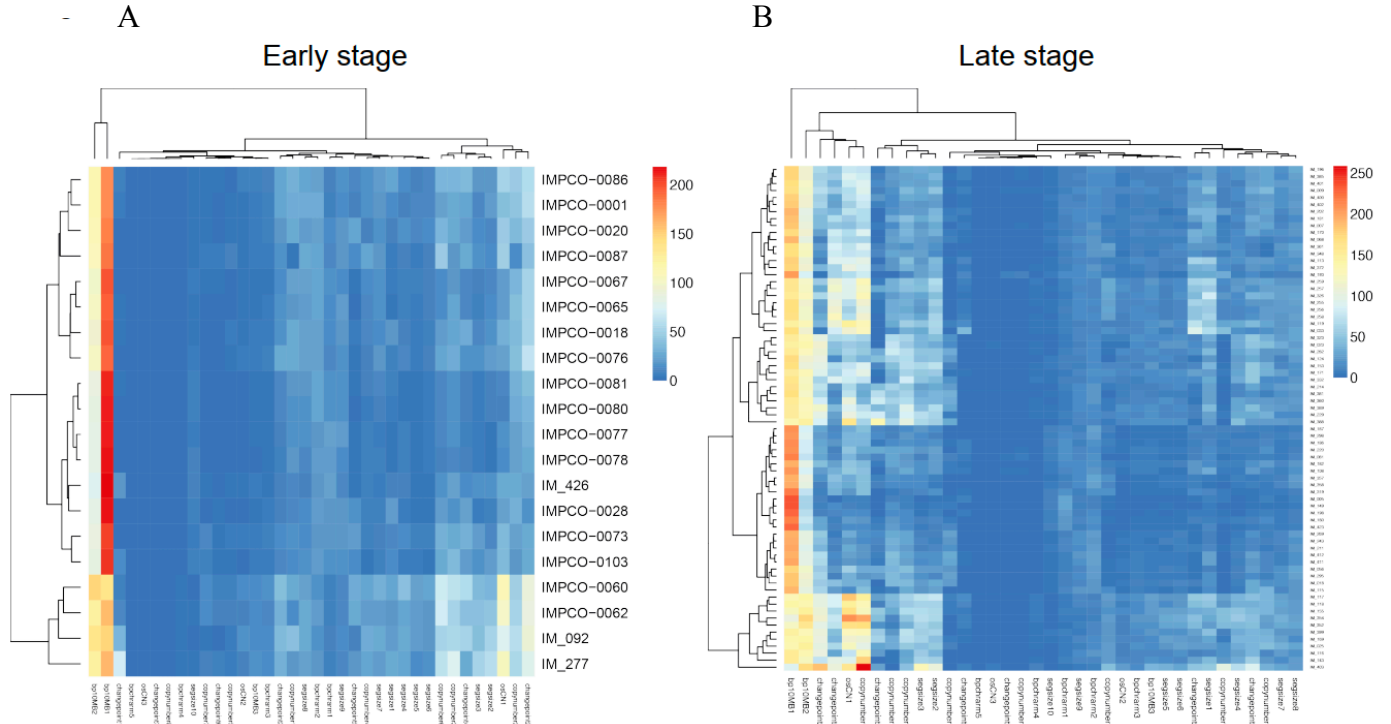


Figure 4.3 Tumour-by-component matrix distribution of 36 copy number components in early stage (A) and late (B) stage cohorts.

I then derived copy number signatures from absolute copy number of early stage and late stage cohorts. The component distributions are defined in our previous analysis, and seven copy number signatures represent a specific pattern of components^[1]. I applied the pattern of components to our tumour-by-component matrices in early stage and late stage cohorts to obtain the seven copy number signature exposures of the two cohorts. The sum of copy number signature exposures is 1 in each sample, and the proportion of each signature varies from sample to sample (Figure 4.4). Most samples only contain a subset of signatures, and zero exposures to specific signatures are observed in both early-stage and late-stage cohorts.

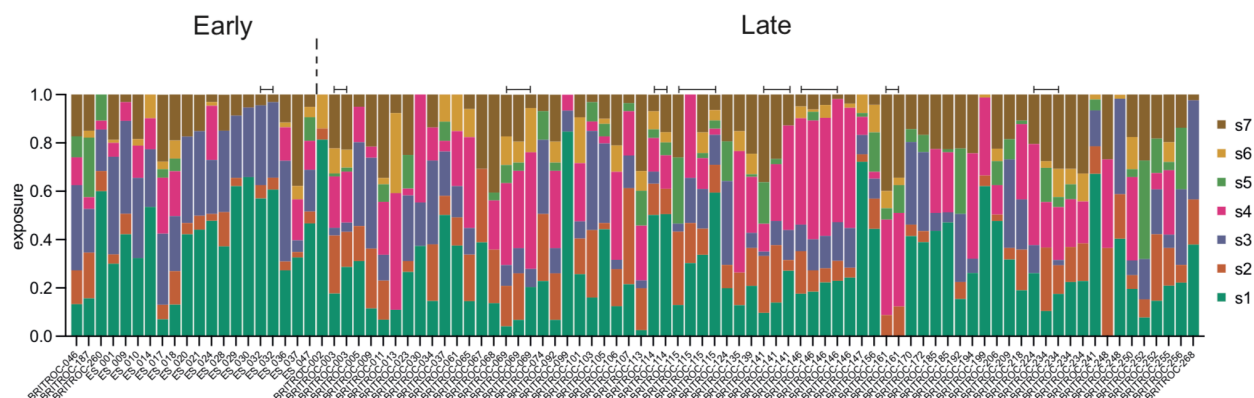


Figure 4.4 Copy number signature exposures in early-stage and late-stage patients. Note that signature exposures sum to 1 in each sample. The bars above signatures indicate adjacent samples derived from the same patient.

4.3.2 Copy number signatures comparison between early stage and late stage cohort

To compare copy number signatures differences between early stage and late stage cohort, a fixed-effects (Bernoulli) analysis was used to model the presence or absence of signatures and a fixed-effects multivariate normal distribution model, based on isometric log ratio (ILR)-transformation, to compare the two cohorts. A fixed-effects model is a statistical model whose parameters are fixed or non-random quantities. This model is different to random effects models and mixed models in which all or some of the model parameters are random variables^[157]. This comparison analysis is kindly provided by Lena Morrill Gavarró in the University of Cambridge based on copy number signatures that I provided.

4.3.2.1 Analysis of zero exposures

Signature 1 and signature 3 are always present in early-stage samples but sometimes absent in late stage samples. The change is particularly noteworthy for s3, which goes from being present in all samples of the early group to be absent in 24% of samples of the late group (Figure 4.4, Table 4.1)

Table 4.1: Fraction of zero exposures for each group/signature combination

	early	late
s1	0.00	0.04
s2	0.12	0.13
s3	0.00	0.24
s4	0.32	0.19
s5	0.60	0.46
s6	0.52	0.49
s7	0.12	0.13

S1 and s3 are removed from the following analysis of zero exposures. The model treats each signature independently, so excluding them is not problematic. The only potential result that is not reported in this analysis is s3. Whereas in s1, only three samples have zero exposure in the late stage group, there are 16 samples with inactive s3. As the model is not suitable because the two groups show a perfect separation, another way of assessing whether s3 is less present in the late stage cohort is required. Wald test is typically based on the asymptotic normality of the maximum likelihood (ML) estimation, and the Wald statistic is tested using the asymptotic chi-square distribution^[158]. Using a generalised Wald test for the global presence and absence of signatures, the results are not statistically significant. However, from the estimates of the slopes (right figure in the Figure 4.5), there appears to be a good indication that there are fewer zero s4 exposures and fewer s5 exposures in the late stage cohort.

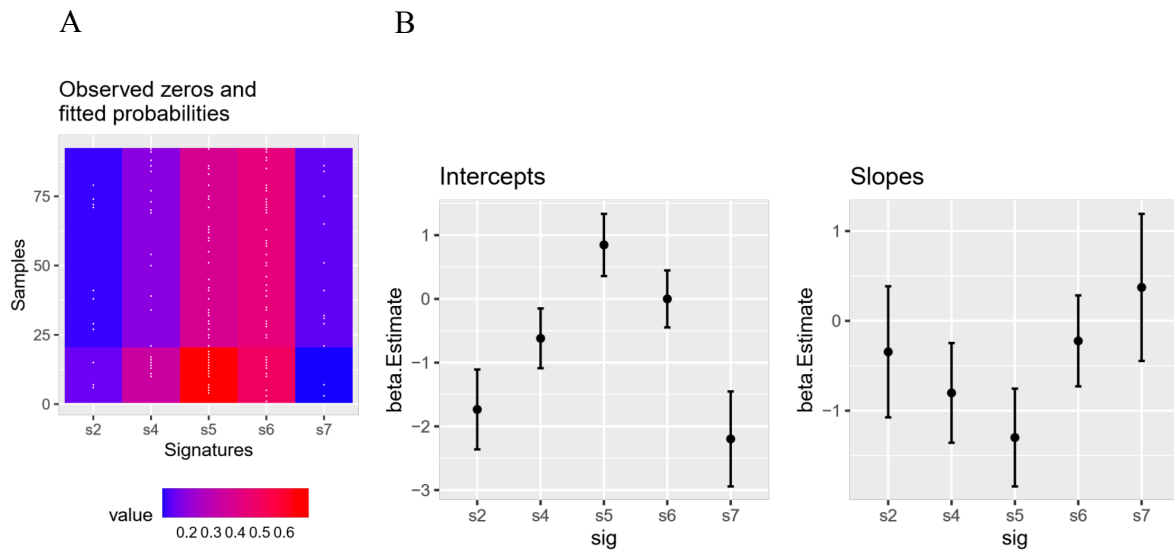


Figure 4.5: Results for the Bernoulli model of present/absent signatures. (A) Observed zeros (white dots) and fitted probabilities of a zero exposure (colour-coded). (B) estimates and standard errors for the intercepts, and estimates and standard errors for the slope.

4.3.2.2. Partial ILR analysis

Here, Lena analysed non-zero exposures to compare early and late-stage cohorts. Isometric Log-Ratio transformation (ILR) was proposed by Egozcue et al ^[159] to transform geometric elements into actual coordinates while preserving all metric properties. An important result is the decomposition of the simplex, as a vector space, into orthogonal subspaces associated with non-overlapping sub compositions^[160]. The density plots and the slope coefficients can be summarised from ILR1, ILR2 and ILR3 (Figure4.6 (a)). There is a relative increase of s2 in the late stage cohort (see ILR1) for s1. S3 is relatively decreased in the late stage cohort (see ILR2, Figure4.6(b)). Conversely, late stage cohort has a relative increase of s4 (see ILR3, Figure4.6(c)). A generalised Wald test here indicates that the difference between the two groups is statistically significant ($p=7.234e-10$).

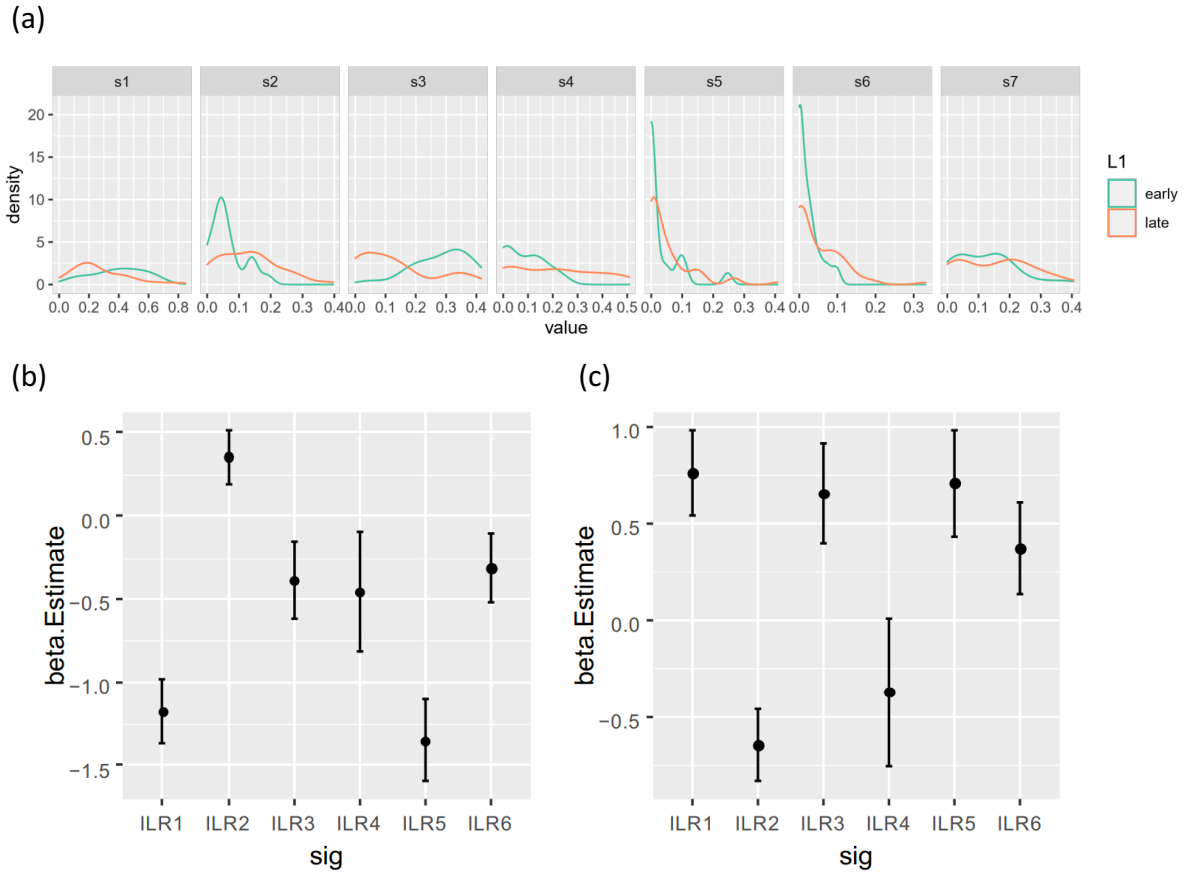


Figure 4.6: Partial ILR analysis for the Analysis of non zero exposures. (a) Density plots for each signature in each group. (b) Estimates and standard errors for the intercepts. (c) estimates and standard errors for the slope (i.e. the change between the two groups).

4.3.2.3. Summary of copy number signature comparison

A fixed-effects Bernoulli was used to model the presence/absence of signatures and a fixed-effects multivariate normal distribution on the transformed data to model subsets of non-zero exposures. The transformation used is a variant of the ILR transformation which is used in Compositional Data Analysis. The findings indicate that s3 has higher exposures in the early stage cohort, analysing zero and nonzero exposures. S4 seems to behave in the opposite direction. See Figure 4.7 for a comparison of the coefficients of each model.

Overall, based on the ILR analysis, there was a highly significant difference between cohorts (generalised Wald test; $p=7.234e-10$, Figure 4.8B), with greater signature three exposure in the early-stage cohort and more signature 4 in the late-stage cohort (Figure 4.8A). In keeping with this, the presence of signature 3 and the absence of signature 4 were both associated with improved overall survival across all patients (Figure 4.9)

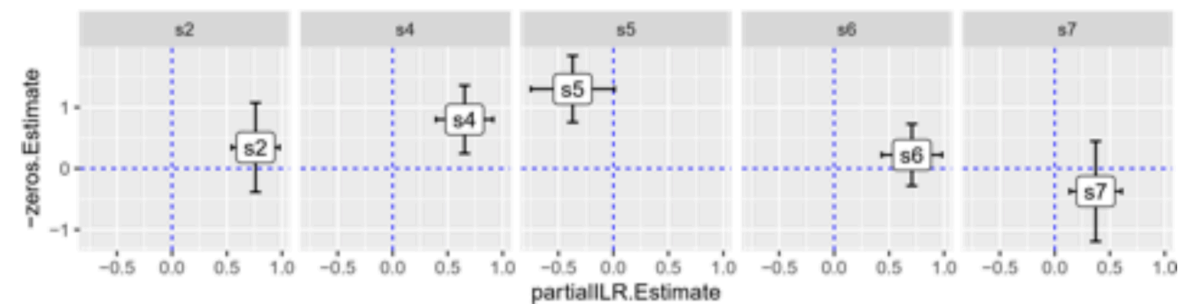


Figure 4.7: Summary of the results of both models, in which we can see how s4 is the only signature with clear changes both in presence/absence and compositional shift

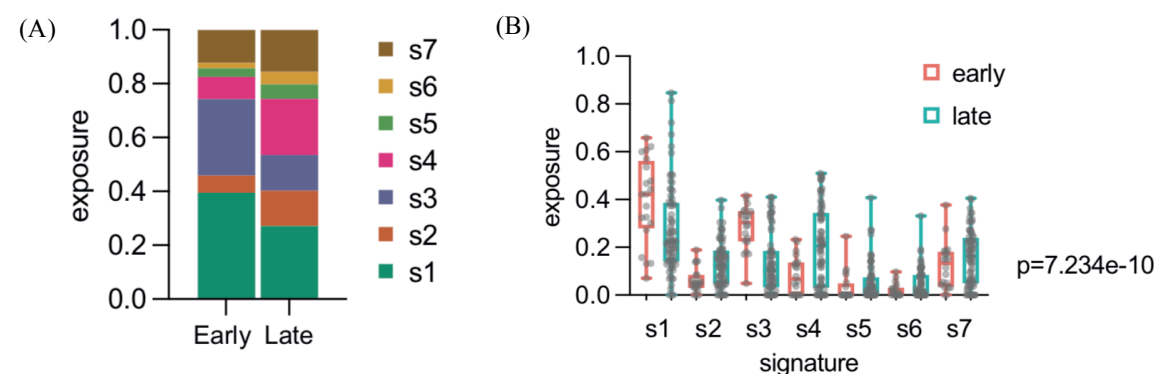


Figure 4.8 (A) Mean signature exposure proportions across the early-stage and late-stage cohorts. (B) Comparison of signature exposures across early-stage and late-stage cohorts; Wald test. $p=7.234e-10$

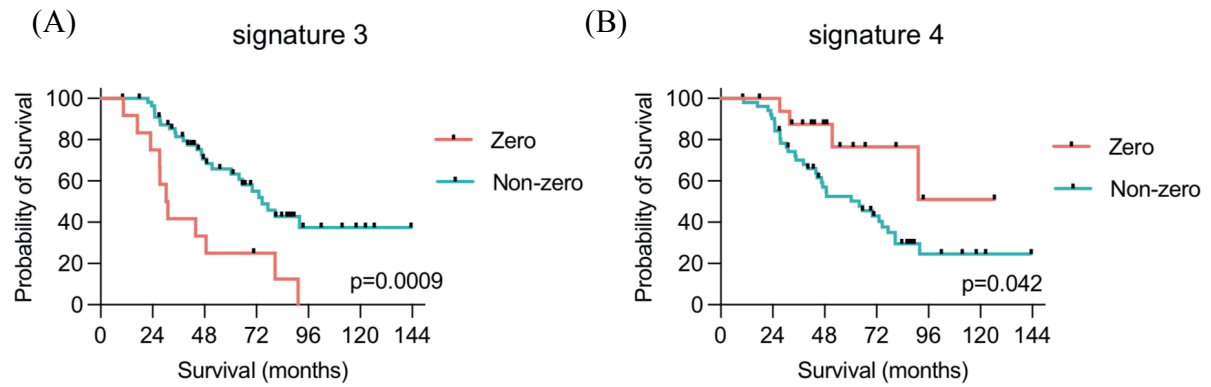


Figure 4.9 Overall survival of combined early and late-stage cohorts by zero vs non-zero exposures to copy number signature 3 (A) and 4 (B). Log-rank (Mantel-Cox) analysis.

Next, CN signature exposures were visualised using simplex plots (Figure 4.10) comparing signature 3 (S3), signature 4 (S4) and all other signatures (1-S3-S4) for the validation of our previous analysis. In the early cohort, the sample observations (red dots) cluster towards the top of the right side of the simplex, in keeping with low or zero signature 4 exposure. For the late group, although some of the observations remain in the same place, many are located towards the left of the simplex, indicating that they have non-zero exposure to signature 4, with a relative decrease in the amount of signature 3. The relative contribution of the other signatures does not change between early and late cohorts - the distance from the observations to the top apex of the plot remains similar. Together, this suggests that signature 3 decreases in relative intensity and signature 4 increases in relative intensity in the late-stage samples, whilst the rest of the signatures remain approximately constant. However, the observations that stay towards the top right of the simplex suggest that a subset of late-stage samples have genomic features more reminiscent of early-stage. In keeping with this, signature 3 remained significantly associated with improved overall survival in the late-stage cases, whilst there was a trend for poorer survival with any exposure to signature 4 (Figure 4.10). This analysis was undertaken by Lena Morrill Gavarró, who also provided the simplex figures.

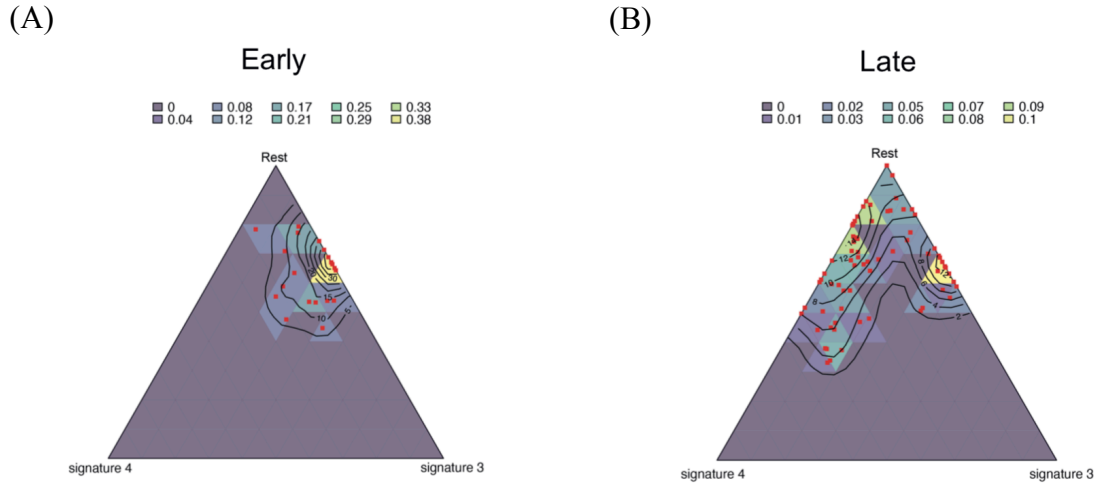


Figure 4.10 Simplex plots representing exposures for CN signature 3 (rest-signature 3 axis), signature 4 (signature 3-signature 4 axis) and the rest of the signatures (1 - S3 - S4) combined (signature 4 -rest axis) in early (A) and late (B) stage cohorts. Each red dot represents a single sample, and the contours represent the density of observed samples. The distribution of red dots moved towards to signature4-rest axis in late stage than early stage cohort, which indicated a trend for poorer survival with higher exposures to signature 4 in late stage cases.

4.3.3 Unsupervised clustering of copy number signatures from early and late stage cohorts

Hierarchical clustering analysis is a technique often applied to gene expression studies to classify expression patterns. Unsupervised means no assumptions are made in hierarchical clustering analysis. This method is determined independently of clinical or pathological factors^[161]. I performed unsupervised hierarchical clustering of the copy number signature exposures across both cohorts and identified three clusters (Figure 4.11). Figure 4.12 also shows the samples and patients distribution in the three clusters. In the 12 patients with >1 sample, samples clustered within the same cluster group in 11 cases. The same cluster classification of samples from the same patient indicated the robust duplication of our sample collection.

In the three clusters, cluster 1 had the highest exposure to CN signature 3, cluster 2 was

dominated by genomes with high signature 1 exposure, whilst cluster 3 showed the highest signature 4 exposure (Figure 4.13). There was a significant difference in sample distribution between clusters ($p < 0.0001$, Chi-squared) with nearly all early-stage samples in clusters 1 and 2, whilst late-stage samples were spread across all three clusters, with the majority in cluster 3 (Figure 4.14). Early and late stage samples distribution shows genomic changes in the two cohorts. To further investigate the relationship between cluster and prognostic significance, I used survival analysis in the three clusters. In Figure 4.15, the clusters were prognostic, with a significant trend for reduced survival across the clusters.

The evaluation of the proportional hazards (PH) assumption in survival analysis is an important issue when Hazard Ratio (HR) is chosen as a summary measure^[162]. The method assesses the appropriateness of statistical methods based on the PH assumption in oncological trials. I chose the standard Cox model from the Survival R package to do analysis in cluster, and the presence/absence of cluster 3 was proven to be of prognostic significance. This result is also consistent with survival analysis, as we described previously. (Figure 4.16).

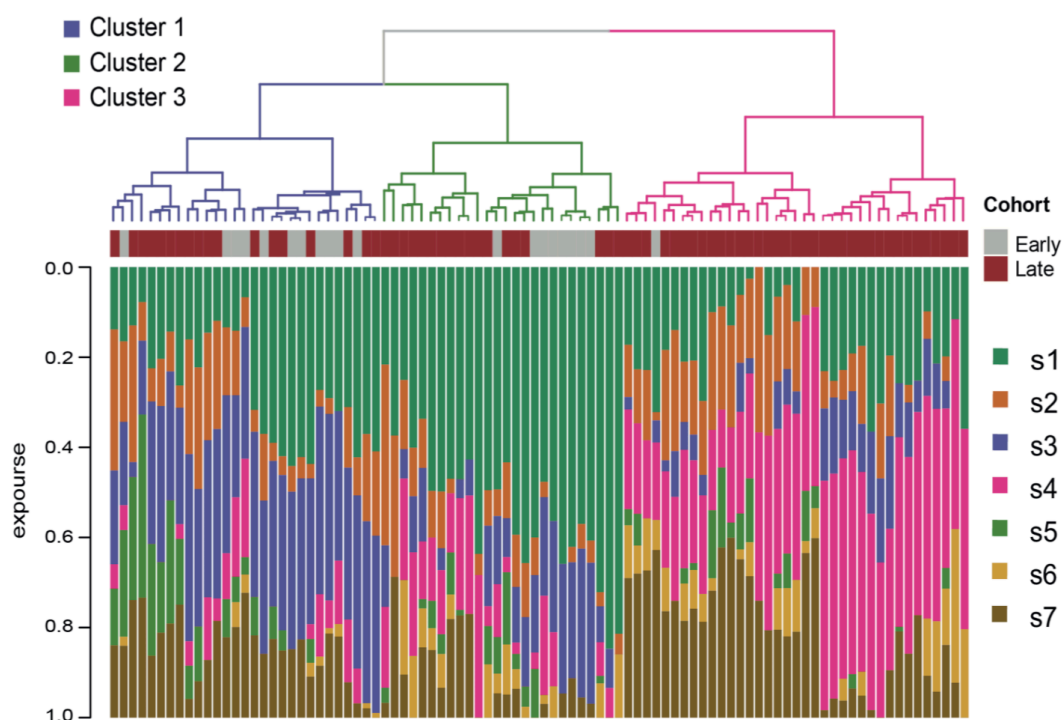


Figure 4.11 Unsupervised hierarchical clustering in combined early-stage and late-stage cohorts.

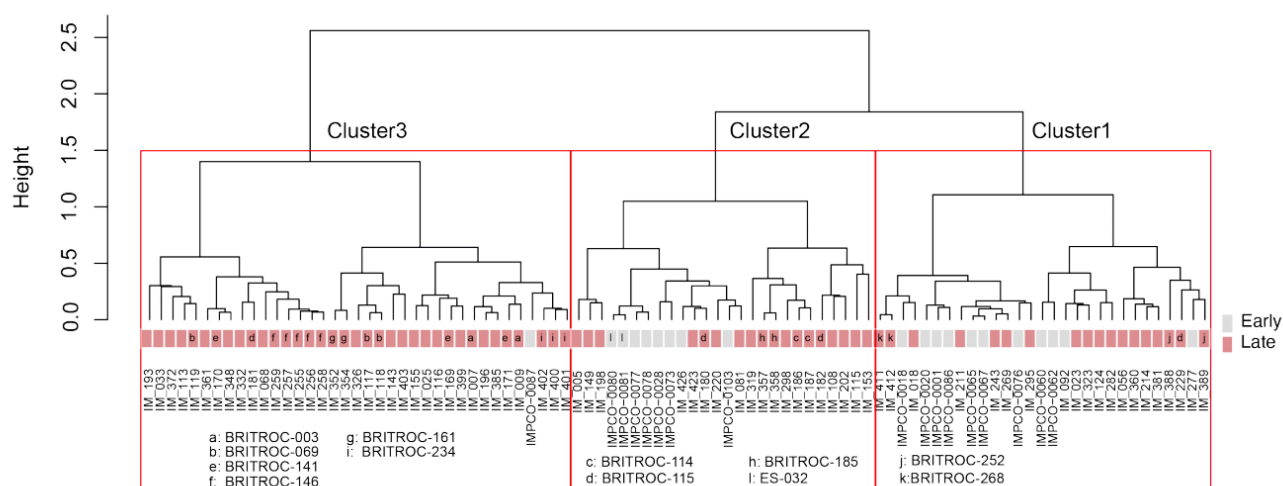


Figure 4.12 Cluster dendrogram of unsupervised hierarchical clustering of samples by CN signatures. Patients who generated more than one sample are indicated.

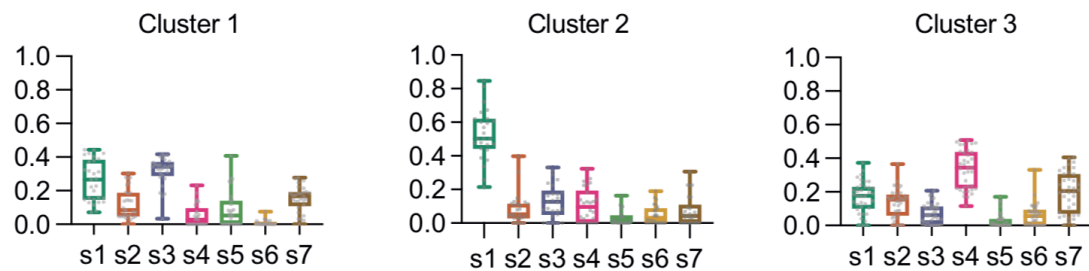


Figure 4.13 Distributions of seven copy number signature exposures in three clusters

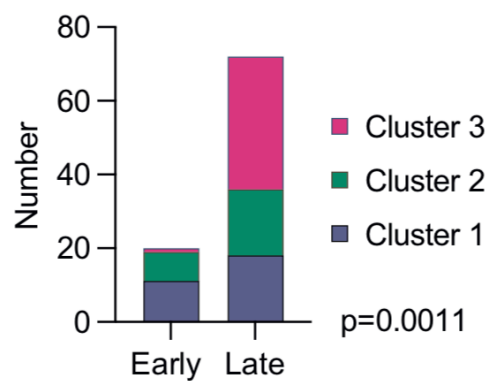


Figure 4.14 Early and late-stage samples by cluster; Chi-squared test.

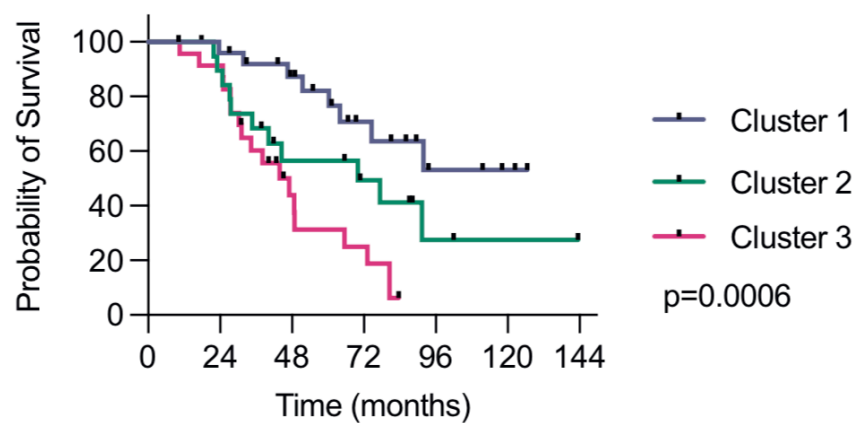


Figure 4.15 Overall survival by cluster; Log-rank for trend

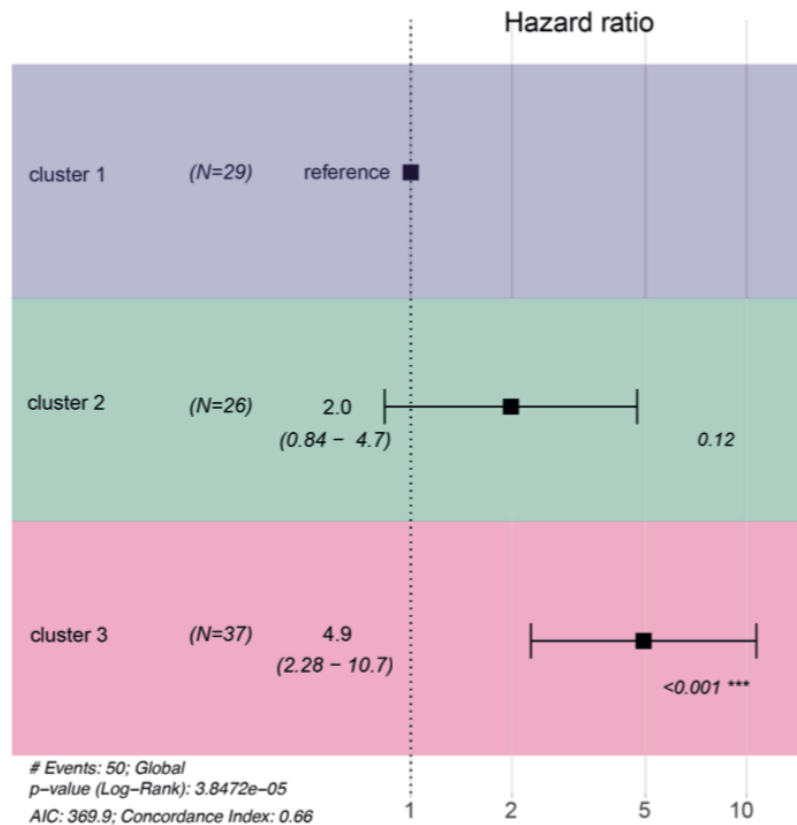


Figure 4.16 Forrest plot of hazard ratio estimates on overall survival (OS) for clusters. Cox proportional hazards.

4.3.4 Survival analysis of copy number signatures, clusters and ploidy from early and late stage cohort

I next quantified ploidy for the three clusters and found highly significant differences across the whole cohort (Figure 4.17) with median ploidy of 1.9, 2.6 and 3.3 in clusters 1, 2 and 3, respectively. High ploidy can reflect whole genome duplication (WGD), which is frequent in HGSC^[143]. Those mutations that occur before genome duplication will be present at multiple copies, whereas the mutation will only be present at one copy after the duplication process^[146]. Therefore, accurately identifying WGD requires assessing allele-specific copy numbers, which provide evidence about the copies of specific mutations. However, this is not possible using sWGS or the limited next generation sequencing panel that I utilised.

Nonetheless, a robust statistical association between CN signature 4 exposure and WGD^[1] was previously identified. Here, I found a solid correlation (Spearman rho = 0.73, $p < 0.0001$) between ploidy and signature 4 exposure across both cohorts (Figure 4.18), implying that WGD may indeed drive the high ploidy. In addition, WGD is likely to generate high ploidy states across the entire genome, and I found a significantly greater proportion of segments with ploidy > 3.0 in late-stage compared to early-stage samples (Figure 4.19(A)) as well as in clusters 2 and 3 compared to cluster 1 (Figure 4.19(B)). Analysis of copy number change point (the change in CN state between adjacent segments) also indicated that clusters 2 and 3 had a significantly greater fraction of segments with change point of at least +2 than cluster 1, again in keeping with WGD^[163] (Figure 4.20). Finally, it was previously shown that median ploidy for cancers with and without WGD were 3.3 and 2.1, respectively^[143]. Using these delineators, we analysed the overall survival of the whole cohort (Figure 4.21) and late-stage patients (Figure 4.22) and identified significant differences in both analyses.

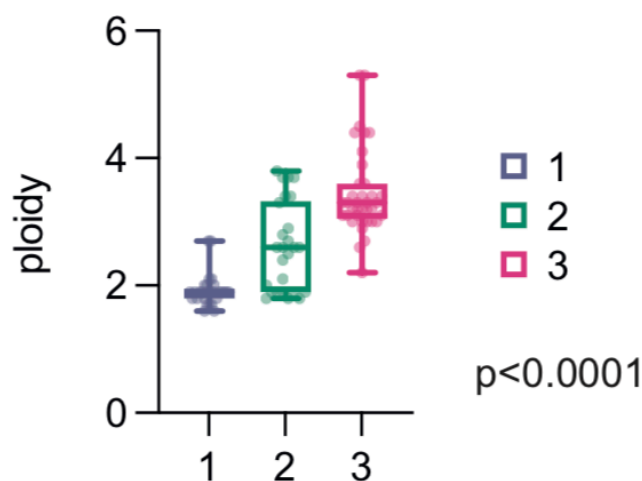


Figure 4.17 Ploidy distribution of late-stage samples in three clusters. Kruskal Wallis test. $p < 0.0001$.

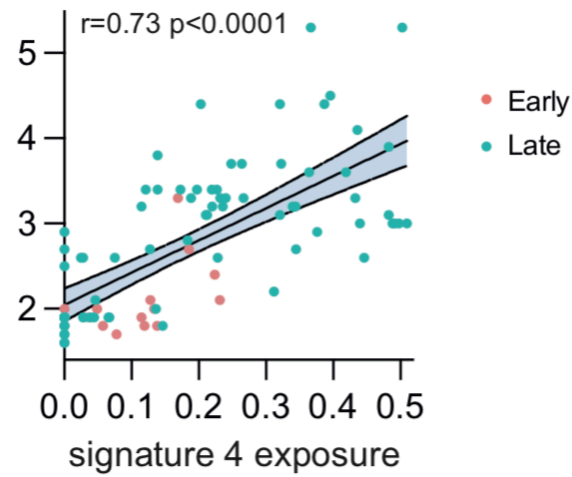


Figure 4.18 Correlation between CN signature 4 exposure and ploidy across both cohorts. Spearman rank correlation. $r=0.73$, $p<0.0001$.

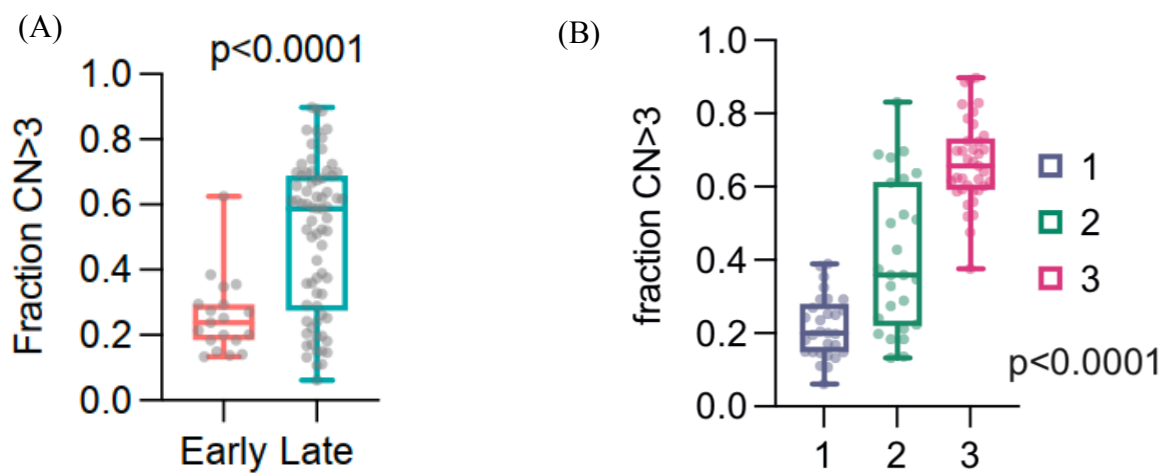


Figure 4.19 (A) Fraction of CN segments with absolute copy number ≥ 3 in early and late-stage cohorts. Kolmogorov-Smirnov test. (B) Fraction of CN segments with absolute copy number ≥ 3 in three clusters. Kruskal Wallis test. $p<0.0001$.

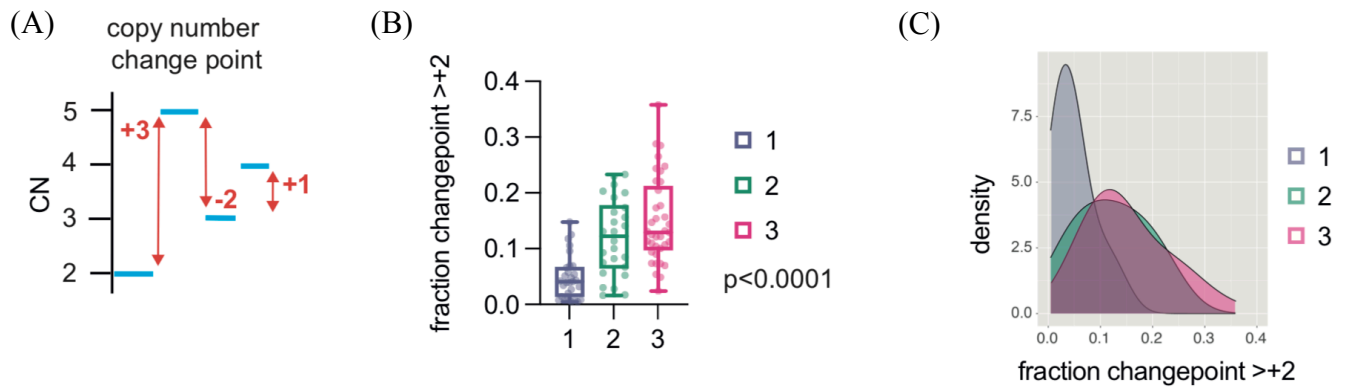


Figure 4.20 Copy number change point. (A) Graphical depiction of CN change point; (B) distribution of copy number change point $\geq +2$ in the three clusters. Kruskal Wallis test, $p < 0.0001$; (C) fraction changepoint $\geq +2$ density distribution. The three colours represent the three clusters.

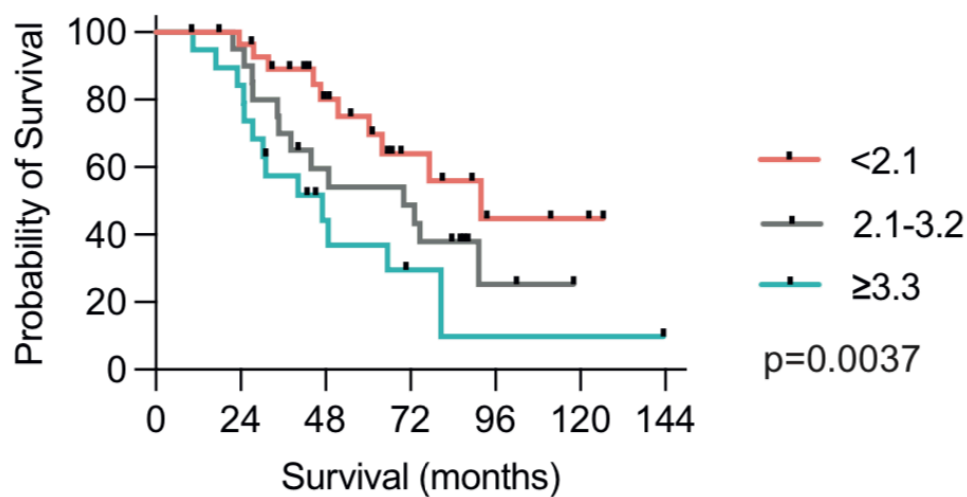


Figure 4.21 Overall survival of combined early and late-stage cohorts by ploidy. Log-rank for trend analysis. $p = 0.0037$.

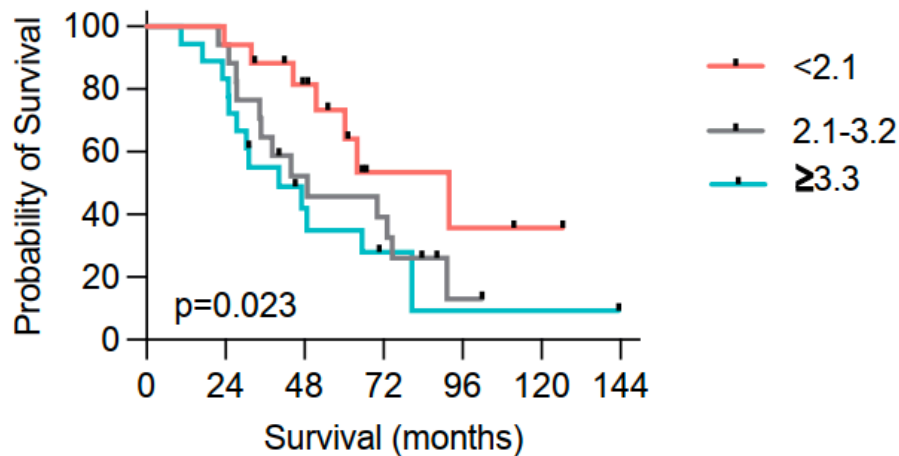


Figure 4.22 Overall survival of late-stage cohort by ploidy. Log-rank for trend analysis. $p=0.023$.

4.4 Discussion

4.4.1 Signature 3 and signature 4 significantly change in early and late stage cohort

My most striking observation here was the difference in overall ploidy between early and late cases, which was further reflected in CN signature exposures in Figure 4.13. Comparison of CN signatures between samples and cohorts is complex because signature exposures are compositional (i.e. they sum to 1 in each sample) and are thus not independent variables: any decrease in one signature will, by definition, be mirrored by an increase in at least one other. In addition, classical statistical methods for analysing compositional data are poor at dealing with zero proportions and many samples have zero exposure to at least one signature.

However, using isometric-log ratio analysis of non-zero signature exposures, it was possible to identify a significant difference between the cohorts overall, driven by signatures 3 (higher in early-stage) and 4 (higher in late-stage). The features that define signature 4 are high segment copy numbers and high copy number changes, both of which are greater in the late cohort. The simplex analysis indicates that the late-stage genomes have increased signature 4,

although a proportion remains 'early-like' with prominent CN signature 3.

According to the previous paper from the host lab, signature 4 is related to high segment copy number and copy number changes of 2-3 copies^[1]. High signature 4 exposure is also a presence in cases with mutated *MYC*, *CDK12*, *CCNE1* and PI3K-AKT signalling pathway, TLR cascade and interleukin signalling pathways, whilst high signature 3 is associated with *PALB2*, *BRCA1/2* and *ATM* mutations in the homologous recombination pathway. Investigating further the top pathways changed between the two cohorts remains important for future analyses.

4.4.2 Three clusters classification in early and late stage cohort

The unsupervised hierarchical clustering identified three patterns in the signatures. Clusters 1 and 2 contained most early-stage samples, whilst the late-stage samples were divided between the clusters. However, cluster 3 had almost exclusively late-stage samples and was associated with higher ploidy and worse survival, as we described in Figure 4.15 and Figure 4.21.

The signature distributions are different in three clusters. Cluster1 has a higher signature3 while cluster 3 has a higher signature 4. It was interesting to note that signature 1 was an independent variable with the highest percentage in cluster2, and is consistent with overall survival analysis in three clusters. The three important components of signature 1 are a low number of breakpoints, 0 or 2 breakpoints per chromosome arm and large segment sizes. This signature is associated with *NFI* mutation and the RAS signalling pathway and is also correlated with amplification-associated fold back inversion events^[1]. A previous study in our lab has established *Trp53*^{-/-}; *Nfi*^{-/-} and *Trp53*^{-/-}; *Brca1/2*^{-/-} ID8 cell lines in mice model and

identified the survival difference between tumours with defects in NF1 compared to Brca1/2 pathway^[164]. Loss of NF1 expression increases intraperitoneal growth in ID8 cells and is related to poor mice survival^[12]. This result provides biological support for our finding that cluster 1 with higher signature 3 (related to abnormalities BRCA1/2/HRD pathway) has better survival than cluster 2 with higher signature 1 (related to abnormalities in the NF1 and RAS pathway). It suggests that a more profound investigation into the role of NF1 and RAS pathways in the development of HGSC is required. The higher copy number, greater overall ploidy and signature 4 exposure in cluster3 suggested that a proportion of the late-stage tumours had undergone whole genome duplication (WGD).

4.4.3 High ploidy/whole genome duplication associated with late stage subset population

Whole genome duplication (WGD) has been described in many solid malignancies^[143] and is generally associated with poor prognosis, as we described in chapter 3. Definitive demonstration of WGD is challenging and requires assessment of both ploidy and extent of LOH^[163], which is not possible using sWGS and the targeted sequencing panel employed here. In practical terms, the analysis of early stage HGSC requires FFPE material that precludes deep WGS analysis. This result is a limitation of my findings. However, the original analysis using deep whole genome sequencing analysis of HGSC specimens showed that CN signature 4 was significantly associated with WGD^[1]. Here I observed a strong correlation between signature 4 and ploidy, suggesting that the high ploidy samples in cluster 3 may have undergone WGD. I also confirmed other features suggestive of WGD, including that cluster 3 samples were more likely to have high CN across the whole genome and high CN changepoint between segments. Lastly, the median ploidy of early and late cohorts was similar to pan-cancer analyses showing that the median ploidy of WGD tumours is 3.3, compared to 2.1 in those lacking WGD^[143]. The commonest genomic correlate of WGD in

previous analyses was mutation in *TP53*, which usually precedes the duplication, and *CCNE1* amplification and loss of *RBI* ^[163]. Although the rates of *CCNE1* amplification here did not differ significantly between our early and late-stage cohorts, my data support the idea that high ploidy is associated with advanced HGSC and poorer prognosis, a concept first explored over twenty years ago^[165]. HGSC has profound levels of segregation error during cell division, an essential precursor of aneuploidy and WGD^[166], and recent data suggest that WGD can emerge in hTERT-immortalised human fallopian tube epithelial cells with loss of wild-type p53 function in the presence of *BRCA1* mutation and *MYC* over-expression, but not with *TP53* mutation alone ^[73].

The critical outstanding question is whether the processes that generate high ploidy are the primary drivers of rapid dissemination in HGSC, or whether high ploidy/WGD are simply time-related markers of evolutionary fitness and are thus more likely to be observed in late-stage than early-stage disease. The absence of an age difference between the two cohorts may suggest that the genomic differences are not time-related, consistent with the finding that WGD is an early event in colorectal^[145] and non-small cell lung carcinomas^[146]. In addition, there was close clustering of samples in patients with >1 sample, again suggesting that WGD does not appear as a late event. However, definitive assessments of accurate rates of WGD in early HGSC compared to advanced disease will require WGS analysis of prospectively collected samples and detailed comparison between fallopian tube primary site and multi-site examination of large cohorts of disseminated late-stage HGSC as well as *in vitro* models.

4.5 Summary

In summary, I found a significant difference in both ploidy and copy number signature exposure between early and late-stage samples, with higher ploidy and signature 4 exposure in late-stage cases. Higher signature 4 exposure is consistent with whole genome duplication. Unsupervised hierarchical clustering revealed three prognostic clusters. Together, my data in chapters 3 and 4 suggest that early and late-stage HGSC share fundamental genomic features but that late-stage disease appears distinct from early-stage, with evidence of whole genome duplication that may provide evolutionary benefit. Further analyses will be required to ascertain whether these differences reflect genuine biological differences between early and late-stage or simply time-related markers of evolutionary fitness.

CHAPTER 5. Genomic evolution during HGSC development

5.1 Introduction

Traditionally epithelial ovarian cancer has been thought to arise from ovarian surface epithelium. However, with the development of SEE-FIM protocols, recent data provided clear evidence that HGSC can originate in the epithelium of fallopian tube^[167]. This chapter will focus on the genomic changes in the development and evolution of high grade serous carcinoma.

5.2 Background

The origin of high-grade serous carcinomas (HGSC) is still controversial, and there are three potential sites; the ovarian surface epithelium (OSE), fallopian tube and mesothelium-lined peritoneal cavity^[168]. In 2001, Jorgen et al described the occurrence of preneoplastic lesions in the normal fallopian tube in women at high familial risk of HGSC. In 2006, Medeiros et al^[169] developed the protocol for Sectioning and Extensively Examining the FIMbria (SEE-FIM) on BRCA-positive women following an index case of fimbrial serous carcinoma. SEE-FIM methodology is now used to carefully evaluate the fallopian tube and maximises examination of the fimbriated end to detect these "early carcinomas." In addition, SEE-FIM mandates that ovaries should be sectioned at 2- to 3-mm intervals and submitted for histological examination. This protocol identified secretory-type cells exhibiting strong p53 staining but non-proliferative on histology. These cells were identified as 'p53 signatures', and were suggested to be a potential precursor in the secretory epithelium of the fallopian tube fimbriae^[170]. p53 signatures are thought to transform into serous tubal intraepithelial carcinomas (STICs) via serous tubal intraepithelial lesion (STIL)^[107], although the relationship between STIL and STIC has not clearly established.

At the genomic level, STIC lesions have been found to have a clonal relationship with established HGSCs based on shared *TP53* mutations and integrated molecular analyses^[171]. Integrative whole-exome sequencing (WES) of STIC and metastatic lesions from Eckert et al^[172] and Labidy-Galy et al^[48] indicated that p53 signatures and STICs are precursors of ovarian carcinoma. Mutation rate and mutation signature and the CNA rates of STIC and HGSC were consistent and unique to each patient. Every patient had very different genomic changes. These alterations were consistent between anatomic sites within each patient, suggesting the biological processes underlying genomic instability during the evolution of HGSC. They also inferred a window of at least 7 years between developing a STIC and initiation of ovarian carcinoma, with metastases following rapidly thereafter.

However, current research only focuses on the whole-exome level, and it is not clear how the genomic-wide changes alter during the evolution of HGSC. Further analyses will be required to ascertain whether the STIC, primary and metastasis tumours have genomic biological differences or the genomic changes were simply time-related markers of evolutionary fitness.

In this chapter, I have characterised the dynamics of copy number and ploidy changes during HGSC dissemination by shallow whole genome sequencing, using a genomic characterisation of five patients with advanced HGSC, who also had identifiable p53 signatures and STIC.

5.3 Results

5.3.1 Pathological identification of p53 signature and STIC lesions

I initially identified 30 ovarian high grade serous carcinoma patients from the pathology archives of Imperial College Healthcare and Barts Health NHS Trusts. The study was approved by the Imperial College Tissue Bank and the REC number is 17/WA/0161 (Project ID: R18060). The patients were selected by reading through pathology reports with the record of “STIC lesions” and “high grade serous carcinoma” from 2015 to 2020. As previously described, the fallopian tube blocks were derived from SEE-FIM techniques. I picked out all the possible blocks recorded with “STIC” in the pathology reports, and each block was cut into three 10µm sections. For further confirmation, all the sections were stained with H&E, p53 and Ki67. Following a review by three expert gynaecological pathologists (Dr Jackie Mcdermott, Dr Naveena Singh, Dr Chish Sokota), eight patients were excluded because of failure to identify STIC lesions. In the remaining 22 patients, four patients (HS17-533, HS15-8221, HS15-6498, 100103) were chosen for further analysis because it was possible to identify p53 signatures, STIL, STICs using Russell et al’s^[173] algorithm within the same samples. The algorithm for identifying STIC lesions is presented in chapter 1. STICs were categorized as having either moderate to strong expression of p53 in > 75% of the cells or complete absence of p53 staining (0% cells stained) in foci of histological atypia and with high Ki67 staining regions ($\geq 10\%$ positively stained cells). STILs were identified as strong expression of p53 or complete absence of p53 staining (0% cells stained) with increased Ki67 labelling index (above 10%) but no atypia. P53 signatures were identified as strong expression of p53 or complete absence of p53 staining (0% cells stained) without atypia or increased Ki67 staining^[105]. One patient, HS18-2994 had STIL and STIC lesions only, and I also included this patient in the STIC cohort. Five patients were kept for downstream analysis and their clinical and pathology information is shown in Table

5.1 and Table 5.2. Four patients had late stage (stage III and IV) HGSC and one early stage patient. The presentative figures of p53 signature, STIL and STIC lesions from five patients are shown in Figure 5.1 (a)-(e). All the IHC was carried out by Nadia Rahman at Barts Institute.

Patient	Dx Age	FIGO Stage	Surgery Type	Operation	Chemotherapy	BRCA1/2 mut	OS	Status
HS17-533	68	IV	interval debulking	TAH,BSO	Carboplatin+Paclitaxel	No mutations confirmed	18.4	AwD
HS15-8221	72	IV	primary debulking	TAH, BSO, Omentectomy	Carboplatin+Paclitaxel		3.1	DoD
HS15-6498	73	IIIC	primary debulking	TAH, BSO, Omentectomy, transverse coloectomy, rectosigmoid anastomosis			11.3	AwD
HS18-2994	43	IIIC	primary debulking	TAH, BSO, Omentectomy	Carboplatin+Paclitaxel	No mutations confirmed	30	DoD
100103	52	IC	primary debulking	TAH, BSO, Omentectomy	Carboplatin+AUC5 +Paclitaxel		95.6	AwD

Table 5.1 Clinical features of STIC cohort. HGSC: high grade serous ovarian cancer. FIGO: International Federation of Gynecologic and Obstetrics.

Dx Age: diagnosis age.

BRCA mut: germline *BRCA1* or *BRCA2* mutation status.

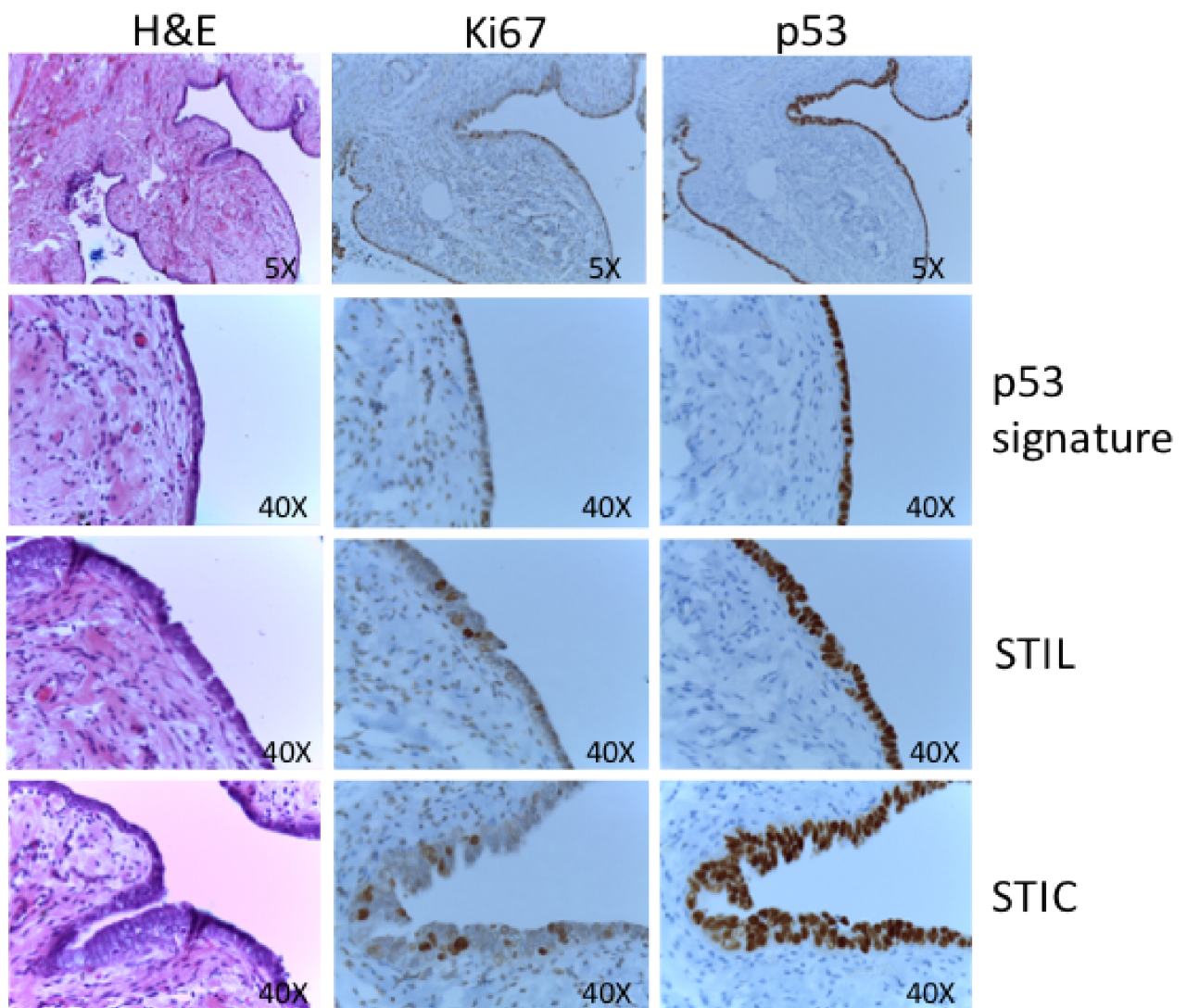
OS: Overall survival.

DoD: Died of disease

AwD: Alive with disease

Patient ID	Sample ID	Anatomical site	Histology	Pathologic Specimen	p53 pattern
HS17-533	HS17-533_N	left fallopian tube	stroma	Block A7	
	HS17-533_N_1	left fallopian tube	Normal fallopian tube epithelium	Block A7	
	HS17-533_T_1	left fallopian tube	p53 signature	Block A7	overexpression
	HS17-533_T_2	left fallopian tube	STIC	Block A7	overexpression
	HS17-533_T_3	left fallopian tube	primary tumour1	Block A7	overexpression
	HS17-533_T_4	left fallopian tube	primary tumour2	Block A7	overexpression
	HS17-533_T_5	right ovary	primary tumour	Block A3	overexpression
	HS17-533_T_6	Omentum	metastasis tumour	Block B2	overexpression
HS15-8221	HS15-8221_N	left fallopian tube	stroma	Block C12	
	HS15-8221_N_1	left fallopian tube	Normal fallopian tube epithelium	Block C12	
	HS15-8221_T_1	left fallopian tube	p53 signature	Block C12	overexpression
	HS15-8221_T_2	left fallopian tube	STIC1	Block C12	overexpression
	HS15-8221_T_3	left fallopian tube	STIC2	Block C12	overexpression
	HS15-8221_T_4	left fallopian tube	STIC3	Block C12	overexpression
	HS15-8221_T_7	right ovary	primary tumour	Block C19	overexpression
	HS15-8221_T_8	Omentum	metastasis tumour1	Block B3	overexpression
	HS15-8221_T_9	left paracolic peritoneum	metastasis tumour2	Block E1	overexpression
HS15-6498	HS15-6498_N	right fallopian tube	stroma	Block E5	
	HS15-6498_N_1	right fallopian tube	Normal fallopian tube epithelium	Block E5	
	HS15-6498_T_1	right fallopian tube	p53 signature of right fallopian tube	Block E5	overexpression
	HS15-6498_T_2	right fallopian tube	STIC	Block E5	overexpression
	HS15-6498_T_3	right fallopian tube	primary tumour1	Block E5	overexpression
	HS15-6498_T_4	right fallopian tube	primary tumour2	Block E5	overexpression
	HS15-6498_T_5	right ovary	primary tumor3	Block E9	overexpression
	HS15-6498_T_6	transverse colon+omentum	metastasis tumour1	Block C7	overexpression
100103	HS15-6498_T_7	Omentum	metastasis tumour2	Block A1	overexpression
	100103_N	right fallopian tube	stroma	Block A2	
	100103_T_1	right fallopian tube	p53 signature	Block A2	overexpression
	100103_T_2	right fallopian tube	STIC1	Block A2	overexpression
	100103_T_3	right fallopian tube	STIC2	Block A2	overexpression
	100103_T_4	right fallopian tube	primary tumour1	Block A2	overexpression
HS18-2994	100103_T_5	right fallopian tube	primary tumour2	Block A2	overexpression
	HS18-2994_N	left fallopian tube	stroma	Block B8	
	HS18-2994_T1	left fallopian tube	STIL	Block B8	overexpression
	HS18-2994_T2	left fallopian tube	STIC	Block B8	overexpression
	HS18-2994_T3	left ovary	primary tumour1	Block B5	overexpression
	HS18-2994_T4	Omentum	metastasis tumour1	Block A1	overexpression
	HS18-2994_T5	Right paracolic	metastasis tumour2	Block C1	overexpression
	HS18-2994_T6	Aorto caval lymph node	metastasis tumour3	Block E2	overexpression

Table.5.2 Pathology features of STIC cohort.

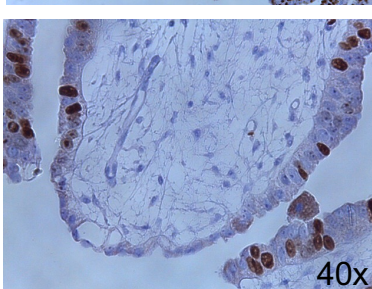
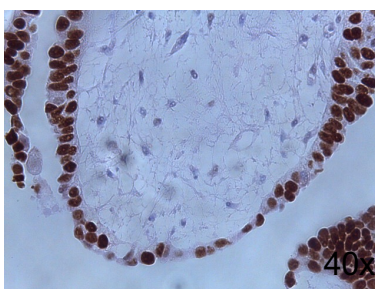
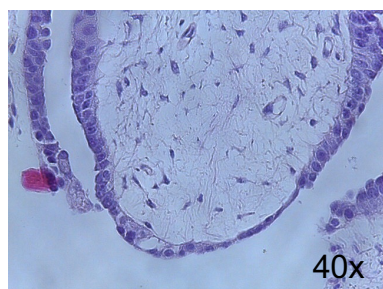
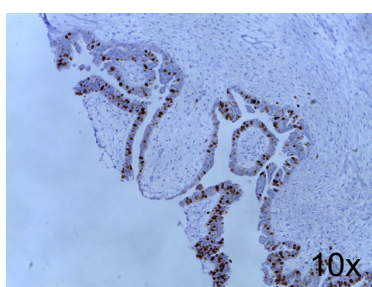
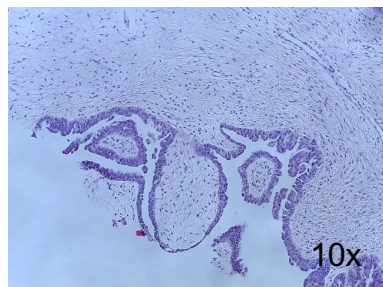


(a) Patient HS15-6498.

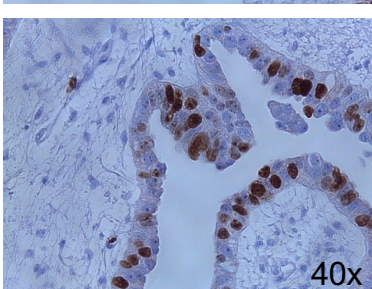
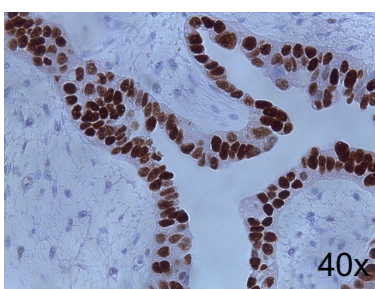
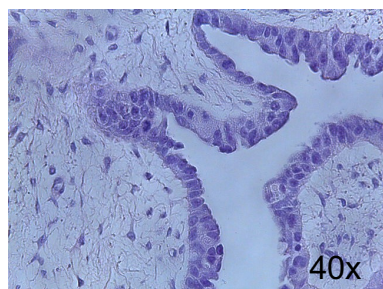
H&E

p53

Ki67



p53 signatures
& STIL



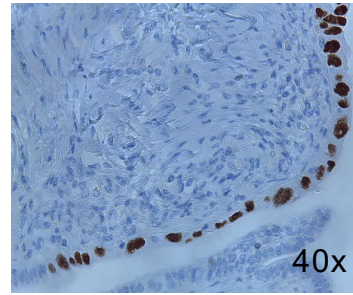
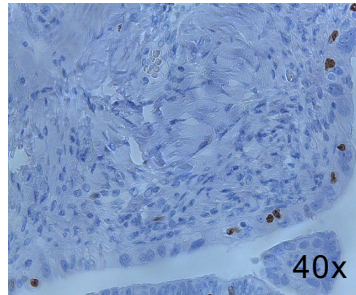
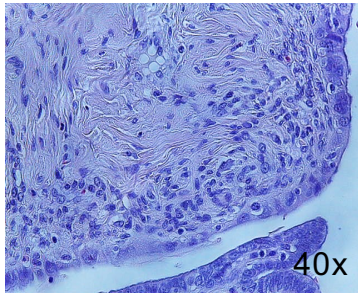
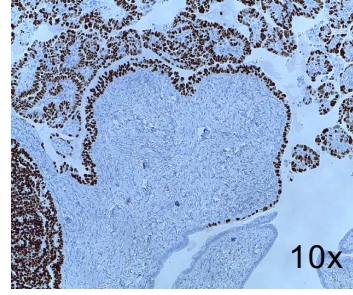
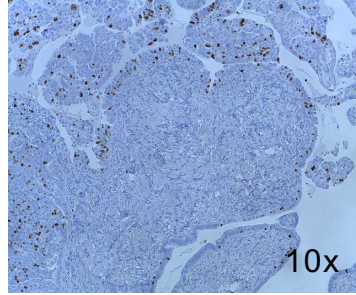
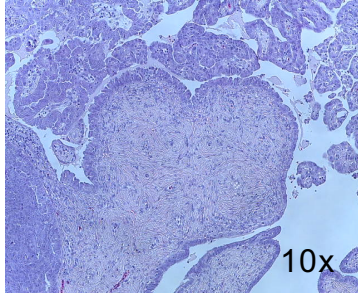
STIC

(b) Patient 100103

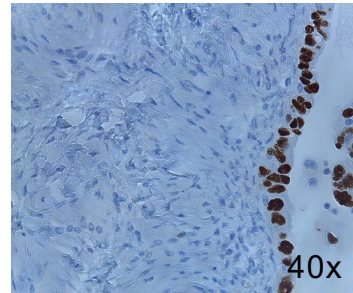
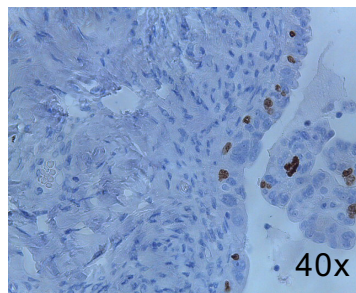
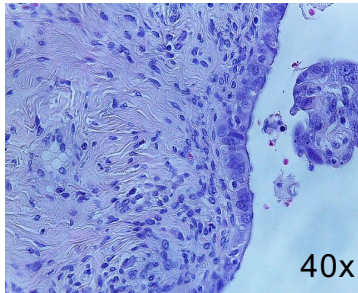
H&E

Ki67

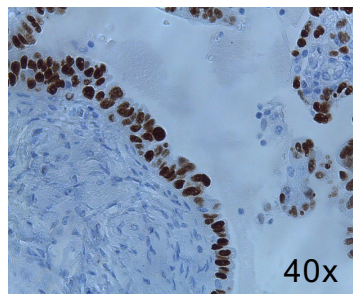
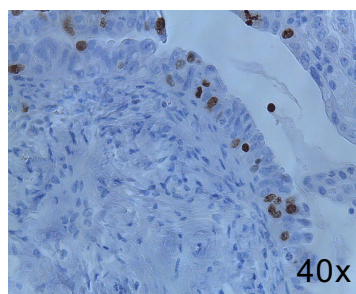
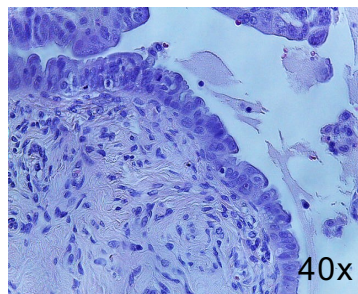
p53



p53
signature

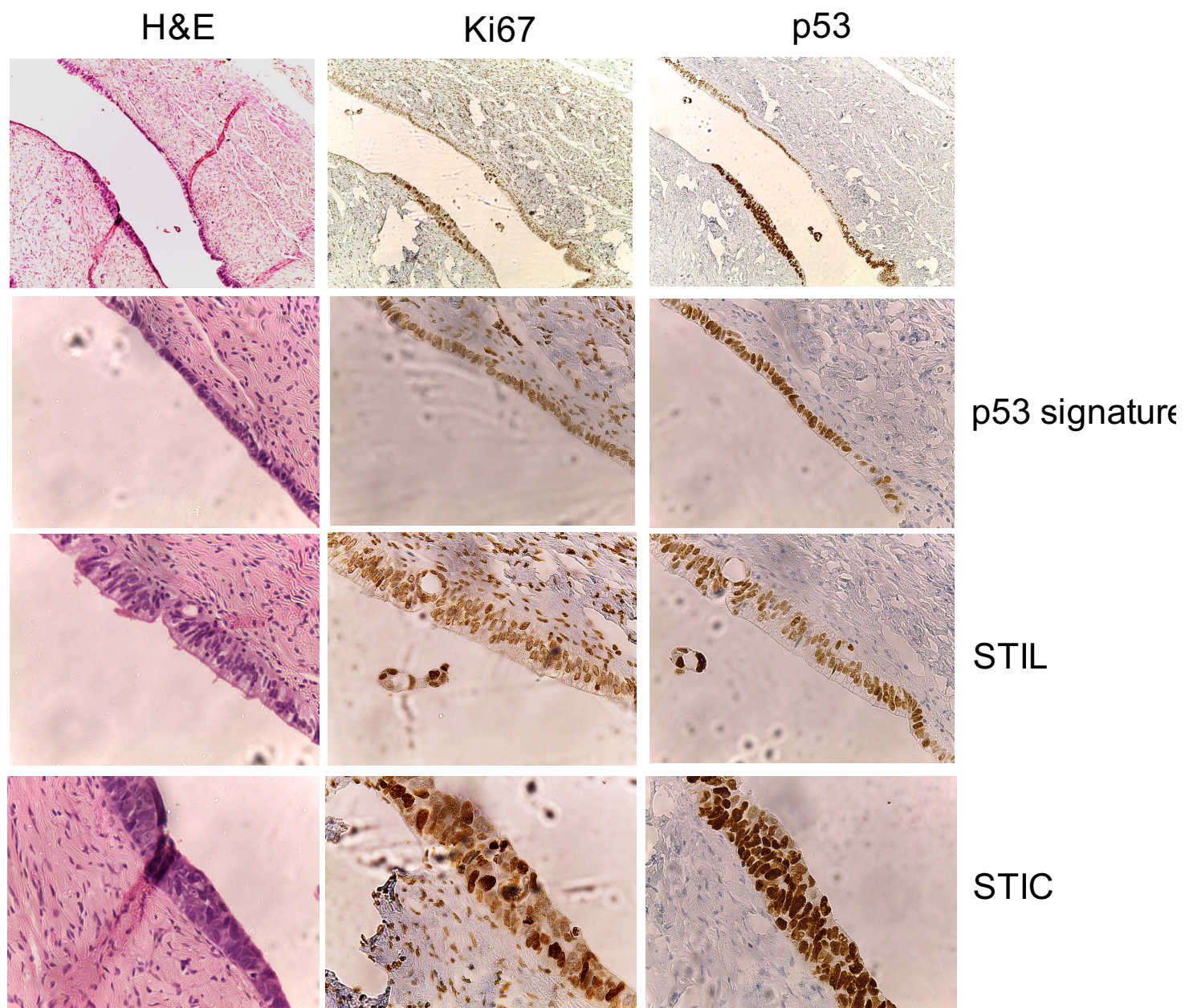


STIL

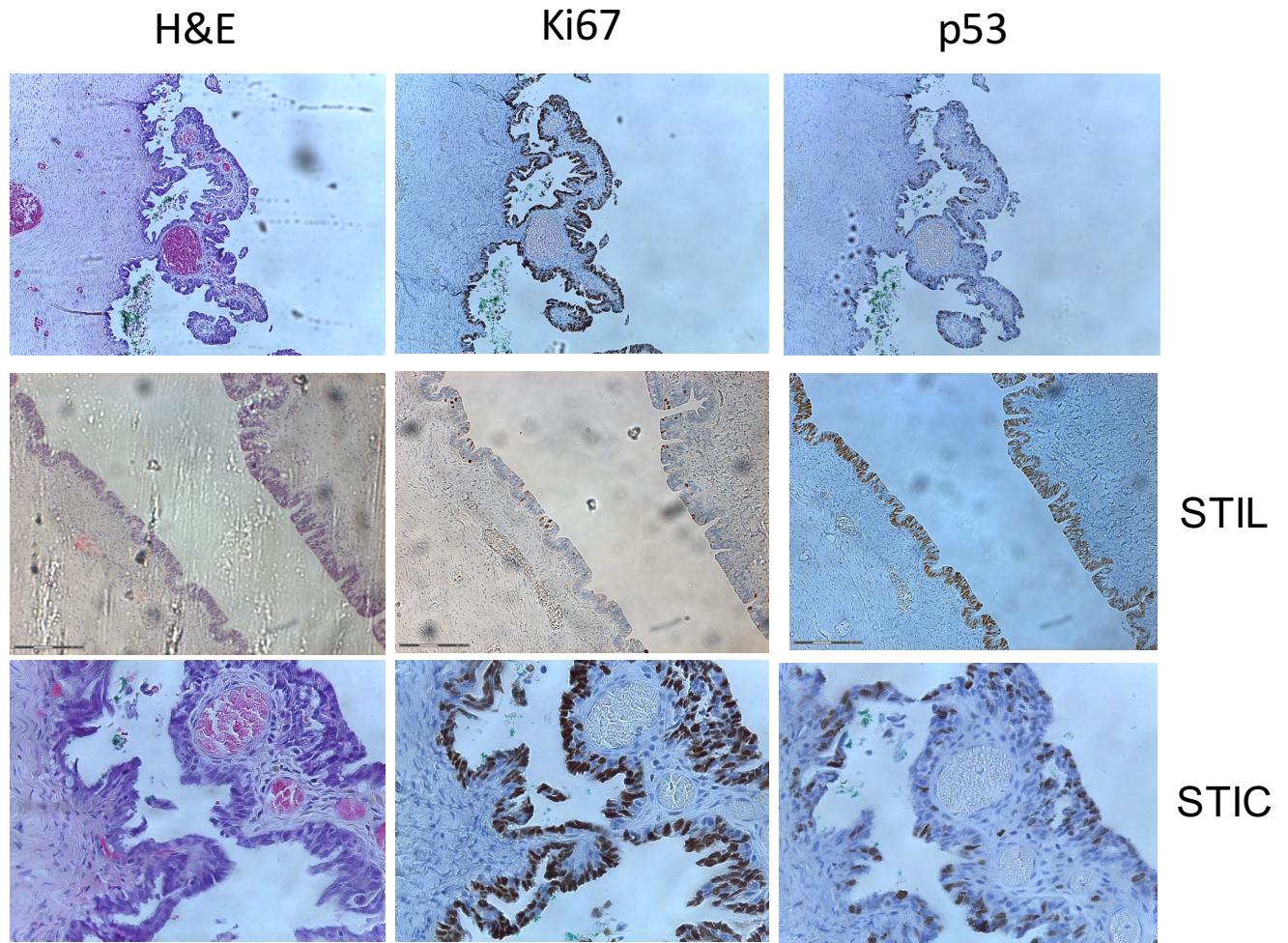


STIC

(c) Patient HS15-533



(d) Patient HS18-8221



(e) Patient HS18-2994

Figure.5.1 (a)-(e). Five representative pathological figures of precursors to high-grade serous carcinoma by staining with p53, Ki67 and H&E. In each patient, the p53 signatures had p53 overexpression, normal morphology, but without high proliferative activity (ie low percentage of Ki67 expression). Serous tubal intraepithelial lesions (STILs) are transitional lesions because they have intermediate morphology findings and proliferative activity between the p53 signature and STIC. Serous tubal intraepithelial carcinoma (STIC) shows cytological atypia and loss of polarity by morphology staining of H&E, p53 overexpression due to p53 mutation, and high proliferative activity with strong Ki67 expression.

5.3.2 Laser capture microdissection of p53 signature and STIC lesions

Laser capture microdissection (LCM) was used to isolate STIC lesions after identification by immunohistochemistry. Tissue was cut onto LCM-compatible polyethylene naphthalate (PEN) membrane slides (MembraneSlide 1.0 PEN, Carl Zeiss, Germany) ^[108]. Laser capture microdissection (LCM) was performed on consecutive p53-stained PEN slides using a Zeiss microscope (Carl Zeiss, Germany) to collect stroma, normal fallopian tube epithelial, p53 signatures, STIC lesions as I described previously in chapter 2. Using H&E, p53 and Ki67 sections as references, I selected the regions for accurate laser capture microdissection. All the p53 signature and STIC lesions from one block of each patient were shown in Figure 5.2 (a)-(e). Some patients, such as HS18-8221, has several STICs, and I collected all the STICs separately for individual analysis. I also collected primary and invasive tumours by macrodissection from the same patient, and pathologist Dr Jackie McDermott scored each primary and invasive tumour for cellularity. The sites of samples from each patient are shown in Figure 5.2. The process for laser capture microdissection is shown in Figure 5.3.

As the yields of DNA extraction from p53 signature and STIC lesions were very low, time was spent optimising the protocol for obtaining enough DNA for downstream analysis. Eric Bonnet et al ^[109] investigated the performance of three DNA extraction kits (QIAamp DNA FFPE Tissue kit, GeneRead DNA FFPE kit from Qiagen, MaxwellTM RSC DNA FFPE Kit from Promega) for whole exome analysis. They checked the DNA quality, coverage, variant analysis and FFPE artefacts between the three kits, and GeneRead kit has the best performance on all of these factors. In addition, the GeneRead kit had lowest median value of

a number of variants difference for each pair of matched FF/FFPE samples for SNVs and INDELs. Also, it had the lowest values for the C > T, G > A substitution rates, most likely due to its artefact correction capabilities specific to FFPE materials. In addition to the Generead kit, I used laser capture microdissection to collect a similar amount of FFPE tissue and extract DNA by QIAamp DNA FFPE Tissue kit. Consistent with Bonnet et al's research, the GeneRead DNA FFPE kit performed better when using low yield low quality FFPE tissue. Another way to increase the yield of DNA is by increasing the number of sections. I utilised 12 slides for each section and kept laser capture microdissection at the same position of stroma, normal fallopian tube epithelial, p53 signature, STIL/STIC, primary tumour. All the lesions were collected individually into an adhesive cap and used as input material. Using multiple sections and GeneRead DNA FFPE kit for DNA extraction, I obtained around 1-5 ng DNA for p53 signature and STIC samples, which was used for shallow whole genome library preparation and sequencing.

In addition to the p53 signature, STIL and STIC lesions, I macrodissected DNA from each patient's FFPE primary ovarian mass and metastasis sites. As described in chapter 3 (page 81-82), we performed Ampliseq sequencing using a custom panel to assess short variants in each gene and allele frequency of any variant identified. The lowest input DNA for Ampliseq sequencing was 10ng for each sample. It was challenging to obtain 10ng of DNA for p53 signature and STIC samples. For primary and metastasis samples obtained from macrodissection, I got much more DNA to perform shallow whole genome and Ampliseq library preparation. The workflow of samples isolation and next-generation sequencing analyses from the five patients is shown in Figure 5.4 and the result of DNA extraction from these samples is shown in Appendix 10.

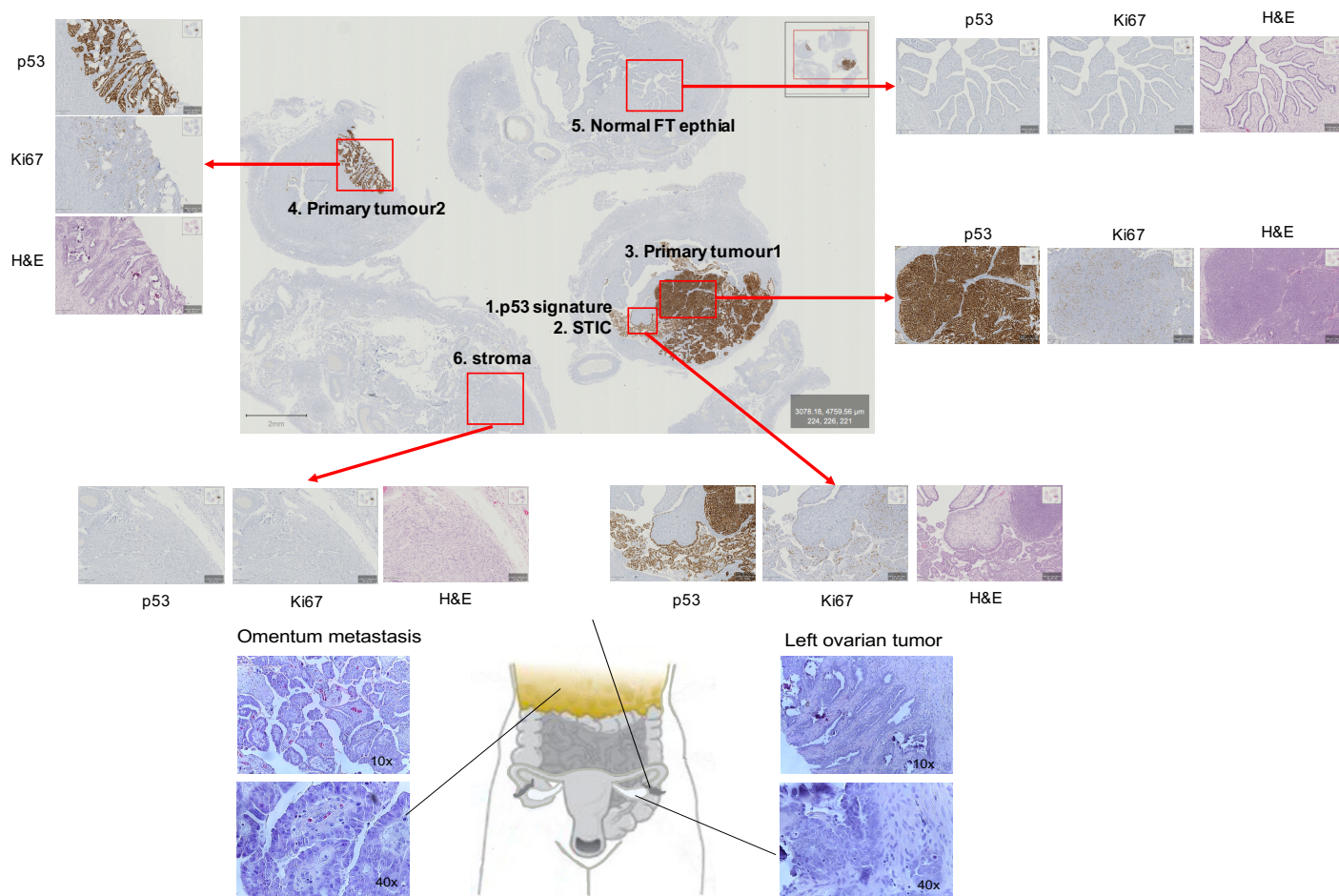
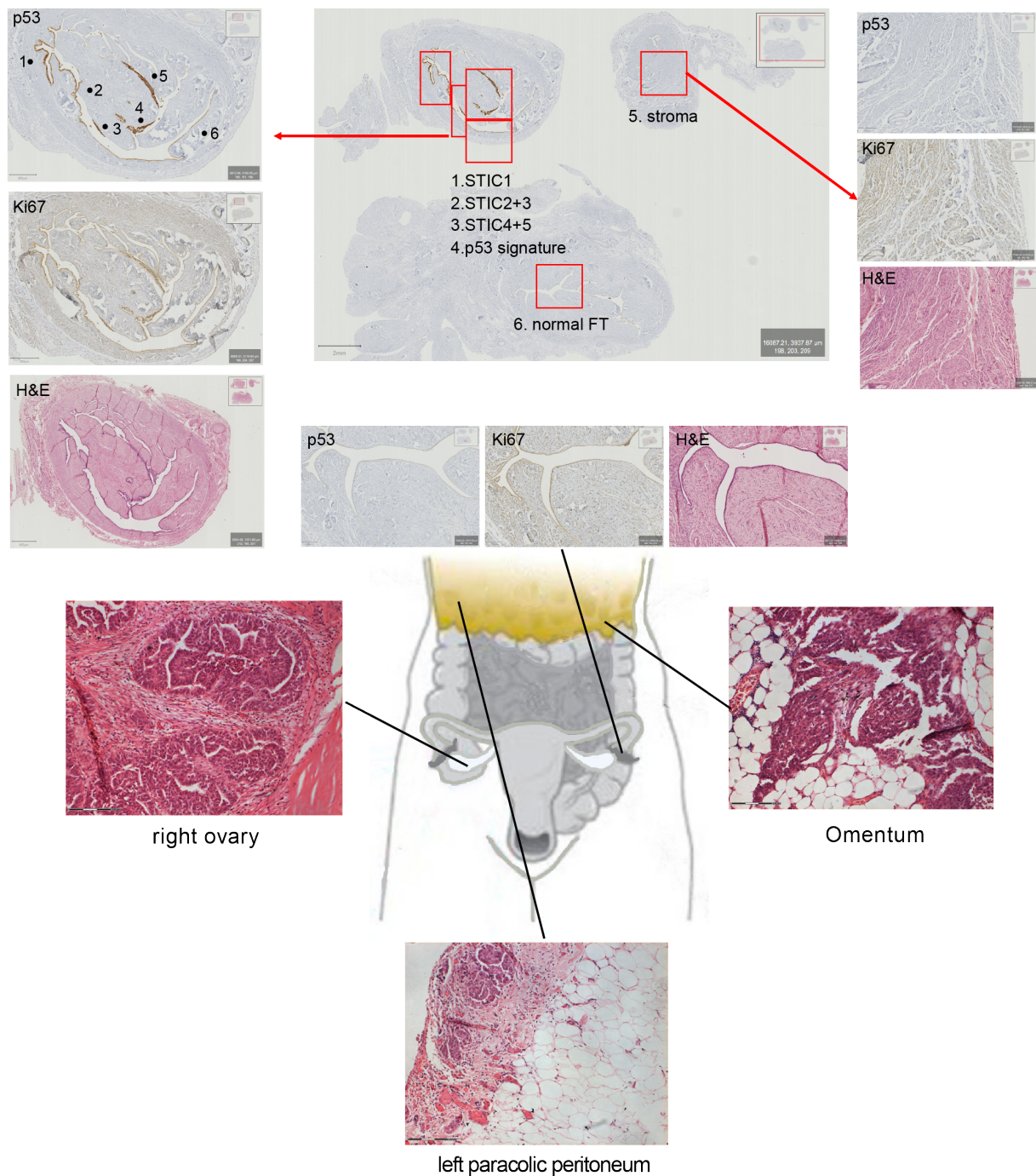
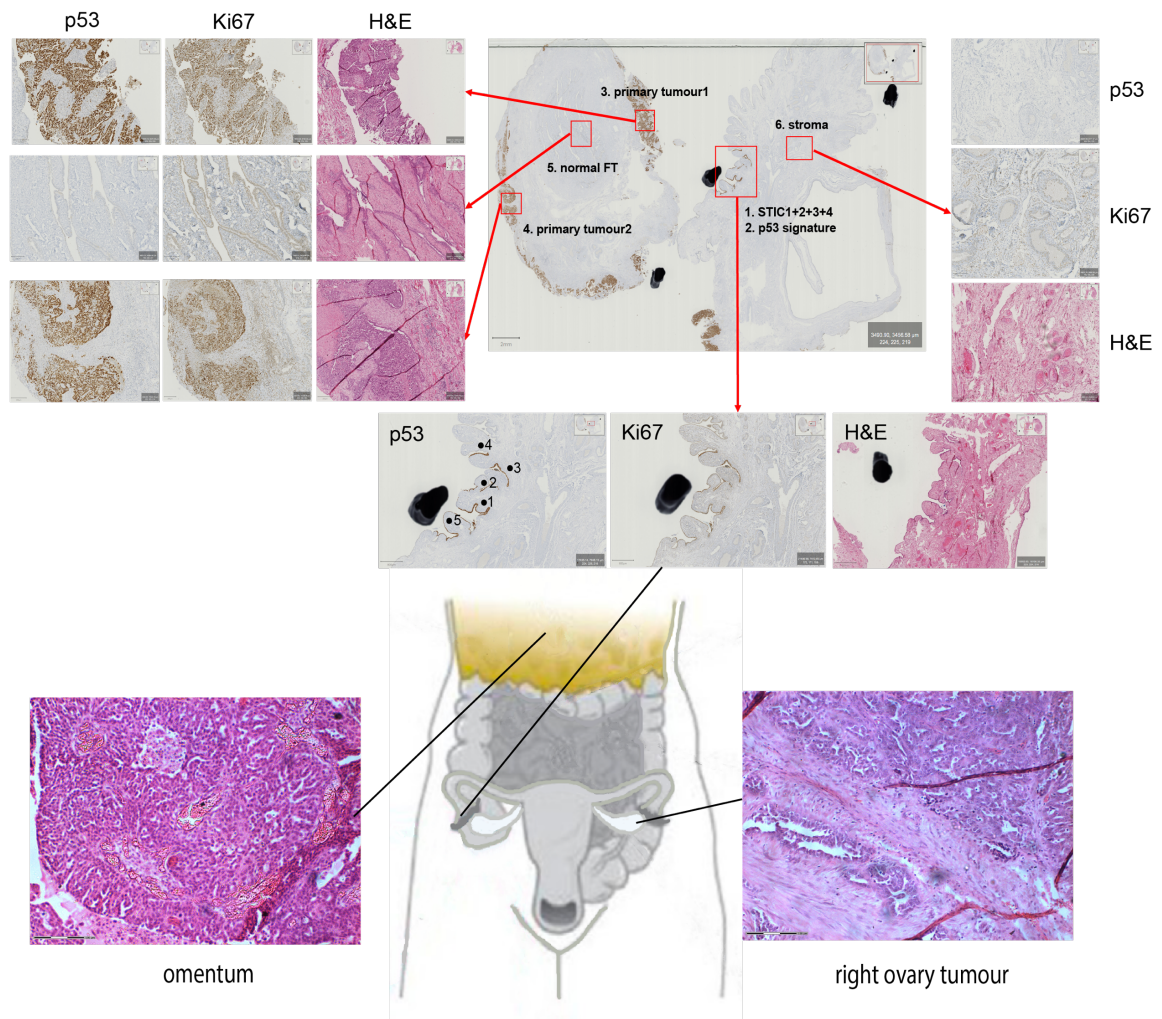


Figure 5.2 (a)-(e) Representative p53, Ki67 and Haematoxylin and Eosin images of five patients with p53 signature, STIC, and primary, metastasis sites.

(a) Sample collection from HS17-533. The anatomical sites in the fallopian tube were selected for laser capture microdissection. P53, Ki67 and H&E staining indicated the position of normal fallopian tube epithelium, stroma, p53 signature, STIC, primary tumour 1 and 2. In addition, primary ovarian tumour was identified in the left ovary, and metastatic carcinoma was found in the omentum. The samples in this patient are: 1. p53 signature; 2. STIC; 3. primary tumour1; 4. primary tumour 2; 5. Normal fallopian tube epithelium; 6. Stroma; 7. Metastasis in omentum.



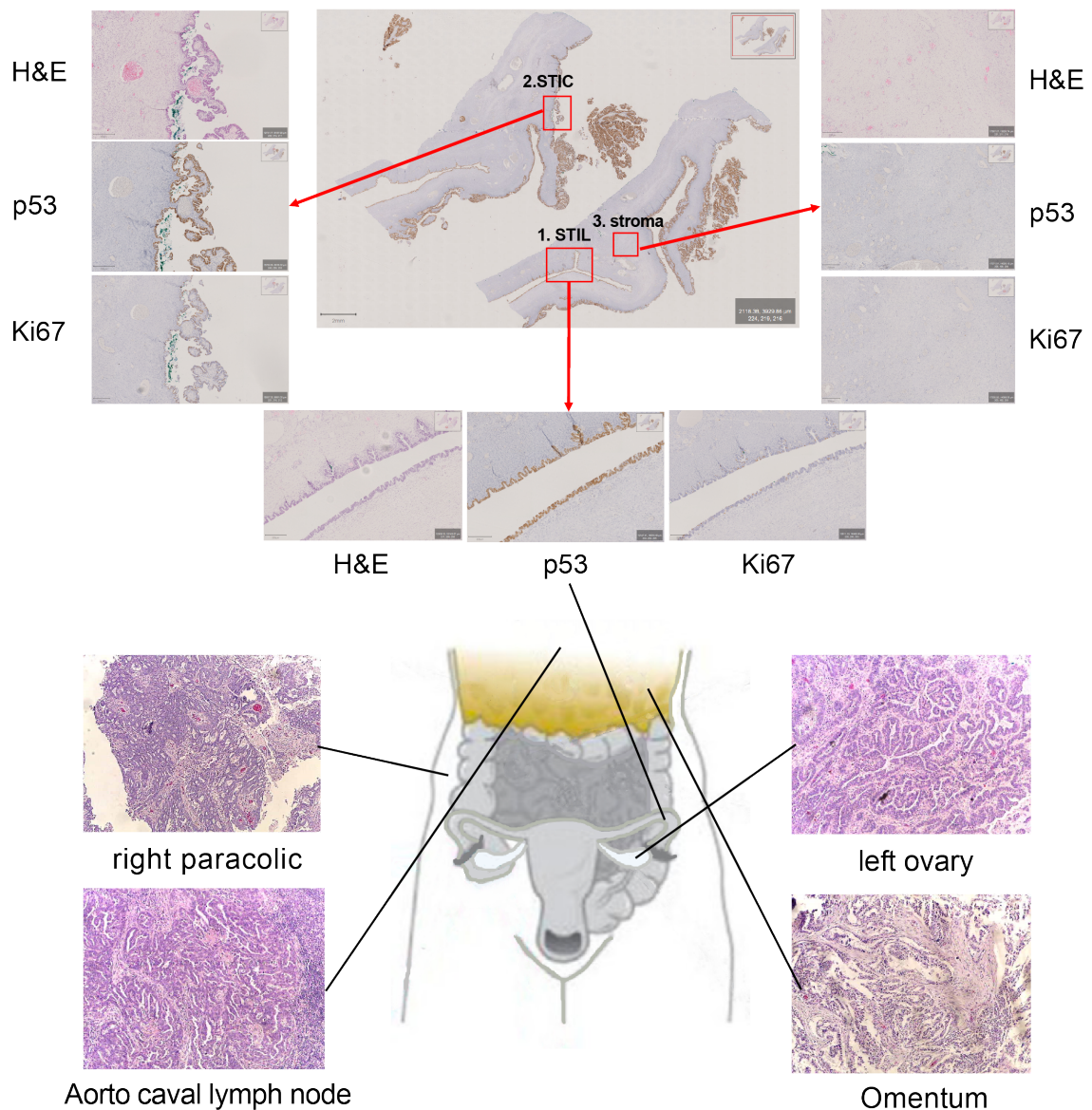
(b) Sample collection from HS15-8221. The anatomical sites in the fallopian tube were selected for laser capture microdissection, and this block contained five STIC lesions. In addition, some STIC samples in the adjacent positions were combined to increase the yield to get enough DNA from STIC lesions. P53, Ki67 and H& staining indicated the position of normal fallopian tube epithelia, stroma, p53 signature, STIC, primary tumour1 and 2. Besides, primary ovarian tumour was identified in the right ovary and metastatic carcinoma was found in the omentum and left paracolic peritoneum. The samples in this patient are: 1. STIC1; 2. STIC2+3; 3. STIC4+5; 4. p53 signature; 5. Stroma; 6. Normal fallopian tube epithelium; 7. Primary tumour in right ovary; 8. Metastatic tumour in omentum; 9. Metastatic tumour in left paracolic peritoneum.



(c) Sample collection from HS15-6498. This block was from the right fallopian tube. Four STIC lesions were combined, and one p53 signature was collected from this block. P53, Ki67 and H&E staining indicated the position of normal fallopian tube epithelia, stroma, p53 signature, STIC (STIC1,2,3,4 together), primary tumour1 and 2. In addition, the right ovary was identified primary ovarian tumour, and metastatic carcinoma was found in the omentum. The samples in this patient are: 1. STIC1+2+3+4; 2. p53 signature; 3. primary tumour 1; 4. primary tumour 2; 5. Normal fallopian tube epithelium; 6. Stroma; 7. Primary tumour in right ovary; 8. Metastatic tumour in omentum.



(d) Sample collection from 100103. This block is from a stage I patient, and there was no metastasis carcinoma in this patient. The anatomical sites in the fallopian tube were selected for laser capture microdissection, and this block contained two p53 signatures and four STIC lesions. Two p53 signatures were combined as a p53 signature sample, and four STIC lesions were combined as two STIC samples based on adjacent locations in the block. P53, Ki67 and H& staining were performed on three slides as an indicator of the position of stroma, p53 signature3+5, STIC1+2, STIC4+6, primary tumour 7 and 8. The samples in this patient are: 1. p53 signature3+5; 2.STIC2+3; 3.STIC4+6; 4.Primary tumour 7; 5.Primary tumour 8; 6.Stroma.



(e) Sample collection from HS15-2994. The anatomical sites in the fallopian tube were selected for laser capture microdissection, and this block contained STIL and STIC lesions. P53, Ki67 and H& staining were performed on three slides as an indicator of the position of stroma, STIL, STIC. The samples in this patient are: 1. STIL; 2. STIC; 3. Stroma; 4. primary tumour in left ovary; 5. metastasis in omentum; 6. metastatic tumour in right paracolic gutter; 7. metastatic tumour in aorto-caval lymph node.

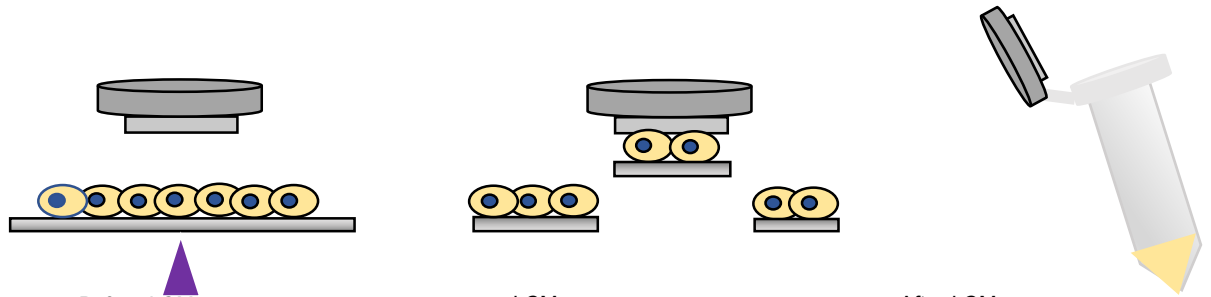
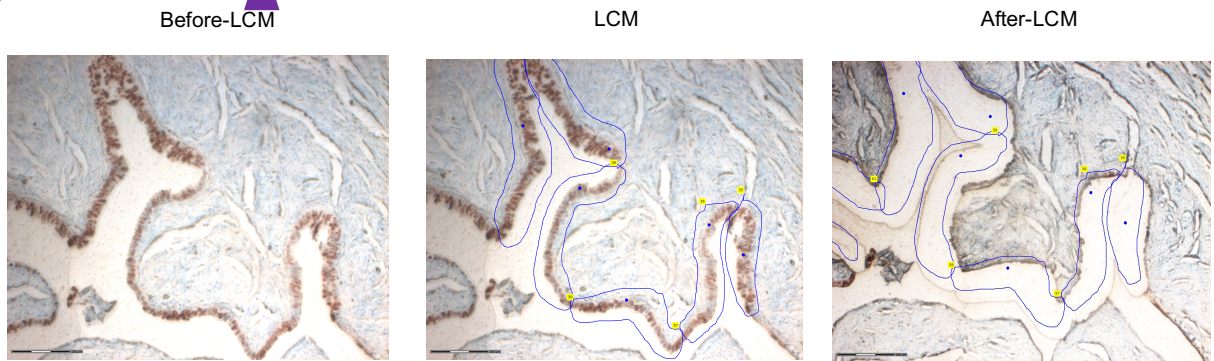
A**B**

Figure 5.3 Laser capture microdissection process for collecting tissue. (A) Schematic figure of laser capture microdissection process by using PALM Zeiss UV laser system. Target cells are marked by PALM software. A UV laser cuts away cells of interest, and the cells and membrane are ejected against gravity into an adhesive cap. The adhesive cap is used to transfer captured cells for downstream DNA extraction. (B) Representative figure of Laser capture microdissection of the STIC lesion before and after the process. The brown tissue represented the STIC lesion that was stained with a p53 antibody.

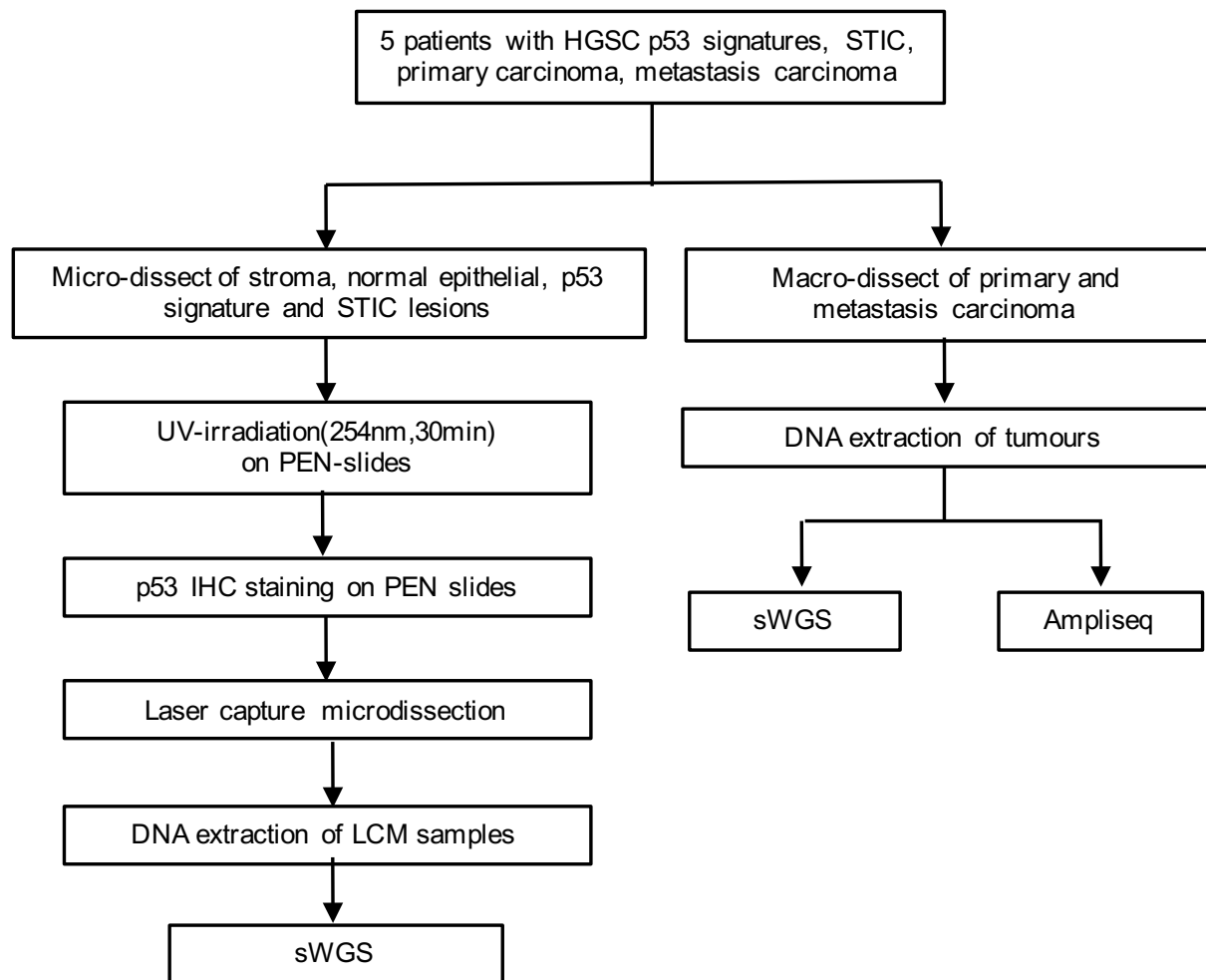


Figure 5.4 Workflow of sample isolation and next-generation sequencing analyses. For p53 signature and STIC samples, slides were stained by immunohistochemical staining of p53. Tumour samples were microdissected for sWGS analysis. Next-generation sequencing analyses were performed for tumour specimens using both sWGS and Ampliseq sequencing for primary and metastasis carcinoma samples.

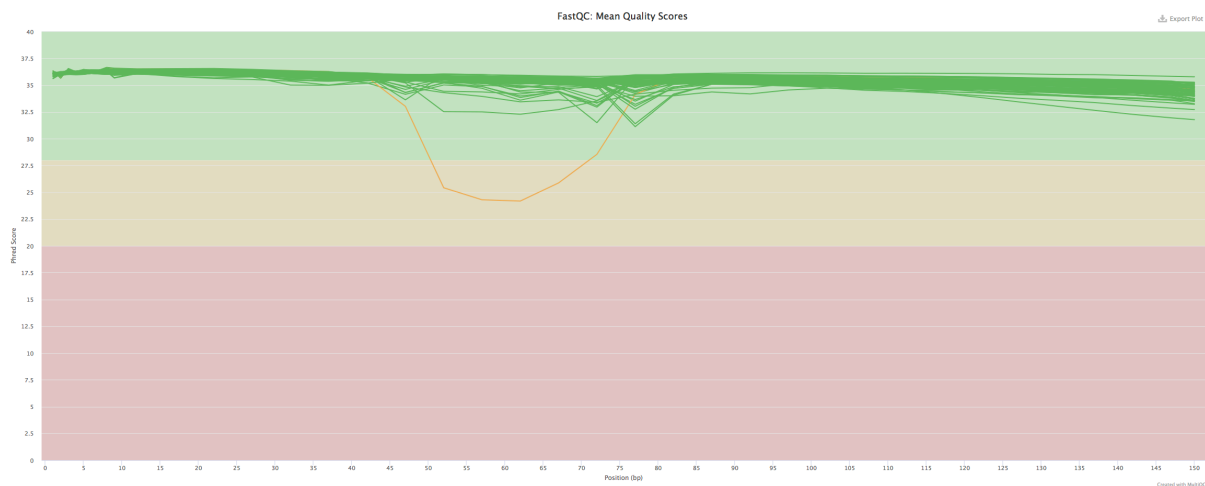
5.3.3 Relative copy number profile of clonal evolution process in HGSC

Shallow WGS (sWGS) data were uploaded into the FastQC package after raw data were transferred from the BRC facility in Imperial College London. First, I generated an HTML web page with different graphs of detailed QC results. Then I used MultiQC software to combine all the HTML web pages into a multiqc summary file. The properties of raw reads (sequence counts, quality scores, mean quality score, adapter content etc.) are shown to assess data quality^[134]; The representative figures are shown in Figure 5.5. The data were of good quality based on the quality control scores and per base sequence quality, which presents the quality score across all bases (Figure 5.5 (a). Acceptable quality control scores were between 30 to 40 quality values, and nearly all the properties were at reasonable variations except one sample, IGF120296. The adapter contamination was seen at a high percentage Figure 5.5(d) and it showed the contamination of Illumina universal adapter in Figure 5.7(A). We used Cutadapt, which is a software package that can search in a single run for multiple adapters and will remove the best matching one. It can optionally search and remove an adapter multiple times, which is useful when library preparation has led to an adapter being appended multiple times^[174]. The sequence for Illumina universal adapter is AGATCGGAAGAG. After the trimming process by Cutadapt, I used MultiQC to test the sequencing quality of all samples again. Figure 5.6 shows the quality control of all samples after trimming. The quality control scores were between 30 to 40, quality values were all acceptable, and no samples were found with any adapter contamination >0.1% in the adapter content (Figure 5.6 (d)). Two presentative figures about the adapter content changes before and after trimming from one sample IGF120285 are shown in Figure 5.7.

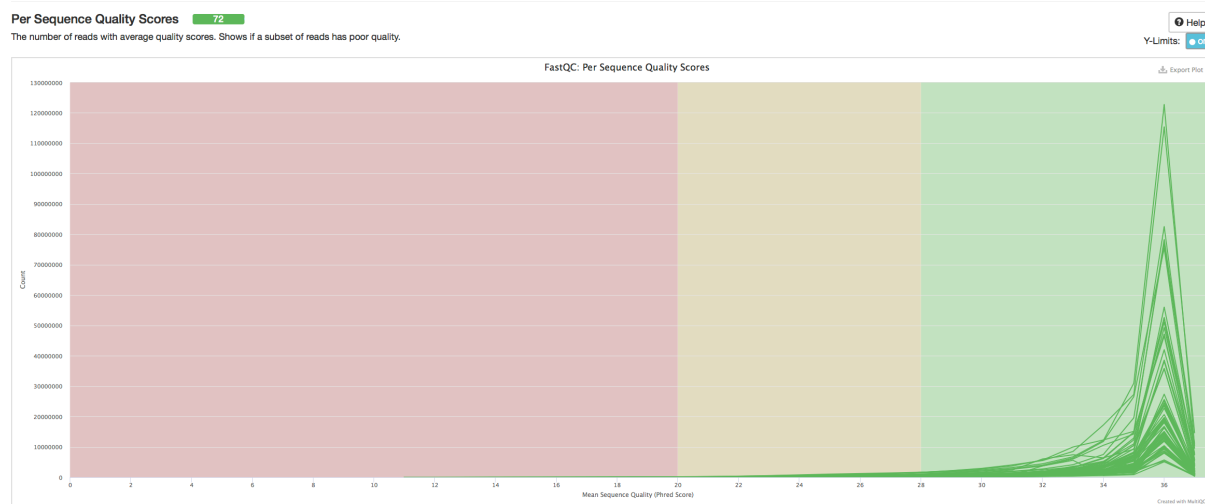
After MultiQC was performed, short sequencing reads were aligned to the hg19 human genome using BWA (Burrows-Wheeler Alignment) pipeline, allowing mismatches and

gaps^[110]. After aligning with the reference genome, SAM (Sequence Alignment Map) format files were obtained in a single mapping file by BWA unified. SAMtools was used to generate binary Alignment/Map formats or (BAM files). SAMtools was also used to sort the reads by chromosomal coordinates and merge the paired and unpaired BAM files to the sorted BAM sample. The package QDNAseq uses sorted BAM files as input and outputs as graphs representing read counts per bin, corrected, filtered, normalised, and optionally log2-transformed in 30kb, 100kb, 500kb and 1000kb bin sizes^[117]. The most critical parameter of QDNAseq is the bin size, which determines the resolution for copy number alterations^[175]. Identifying the optimal bin size can be challenging: the larger the bin size, the more reads they will hold and the less ‘noise’ that will be displayed by the overall output^[176]. On the other hand, the larger bin size will also cause lower resolution. As the input DNA of p53 signature and STIC lesions was very low and noisy, determining the suitable bin size was vital for the downstream analysis. Using different bin sizes, I chose the best fit bin size from the results of the QDNAseq of all samples.

(a) FastQC: Mean Quality Scores



(b) FastQC: Per Sequence Quality Scores



(d) FastQC: Adapter Content before trimming

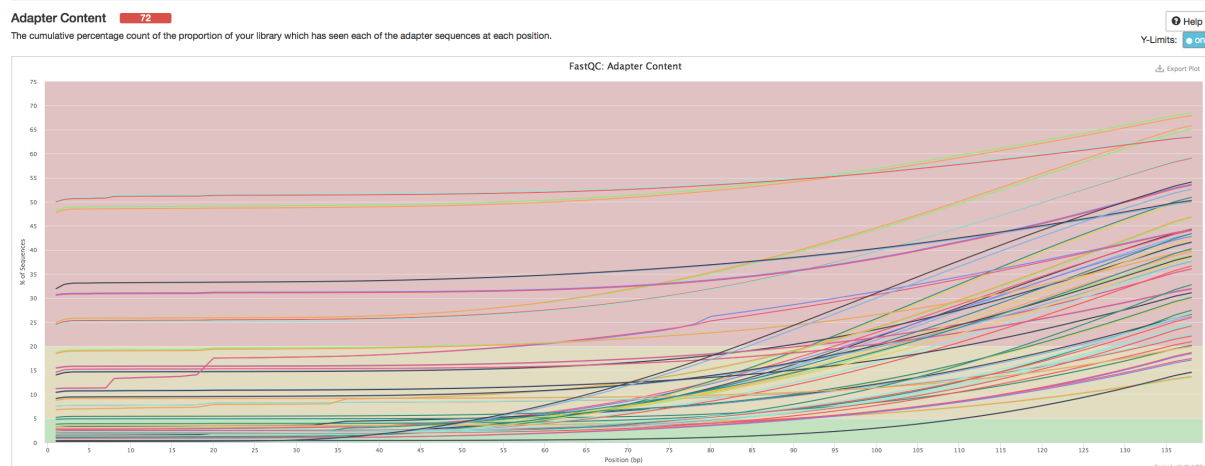
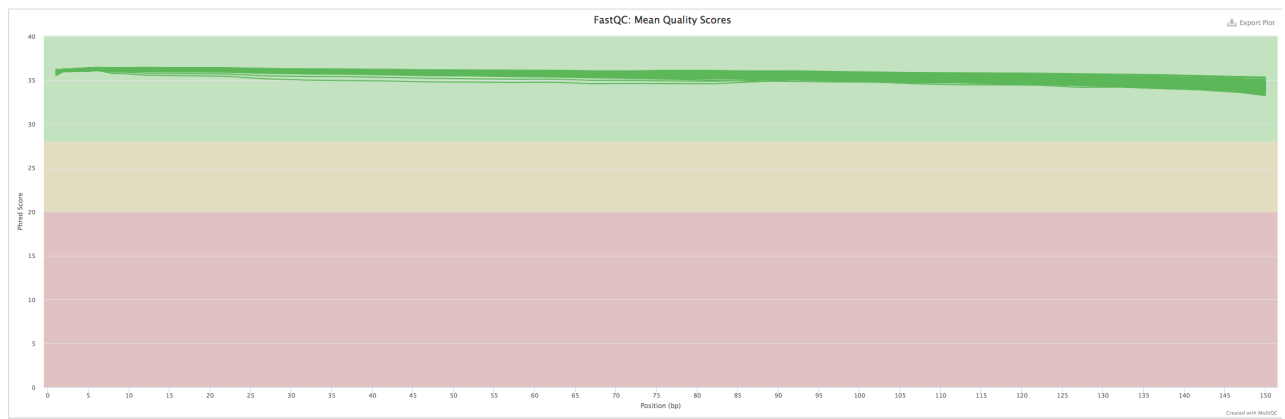
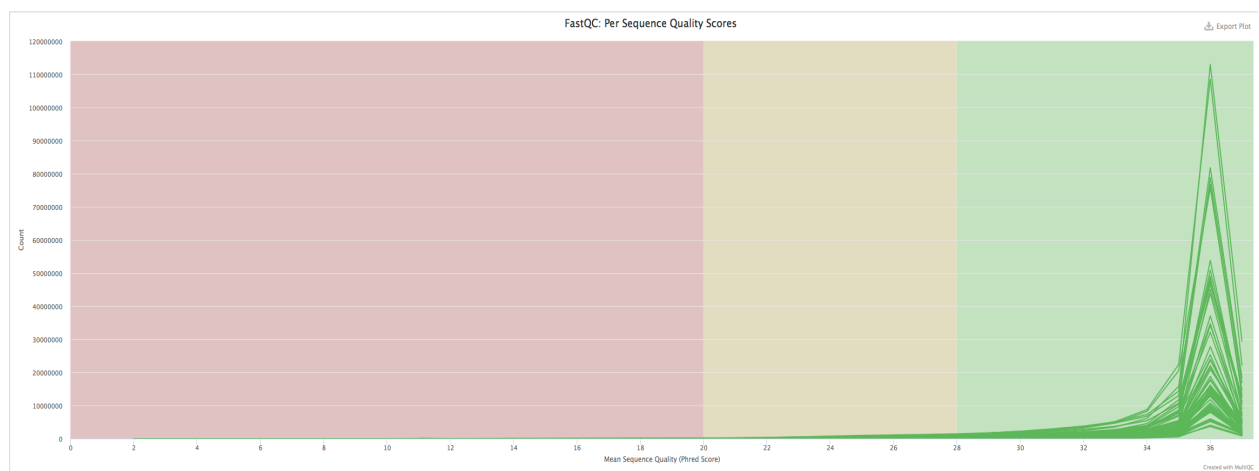


Figure 5.5 (a) Distribution of unique and duplicated reads of sWGS sequencing in STIC cohort samples. (b) Per sequence quality score shows the frequencies of quality scores in all sample. The quality of reads is good quality with the peak shifts to the right side. (c) Per base sequence quality presents the quality score across all bases at each position in the reads. (d) Adapter content shows a high percentage of adapter contamination in these samples.

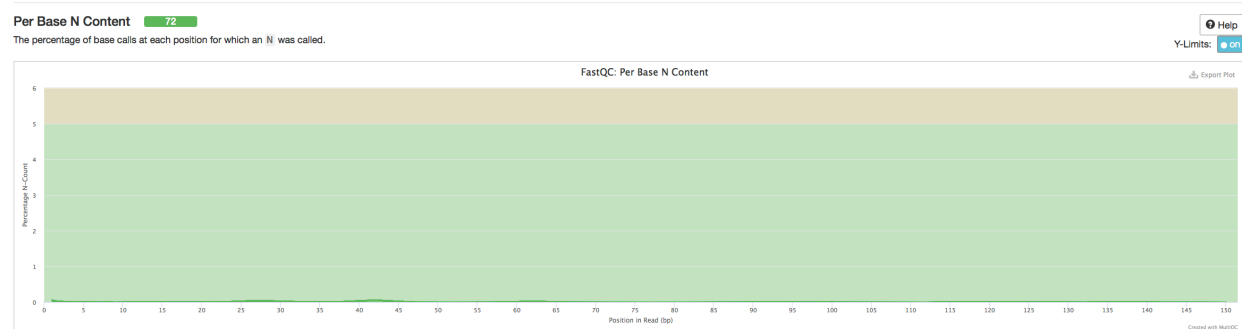
(a) FastQC: Mean Quality Scores



(b) FastQC: Per Sequence Quality Scores



(c)

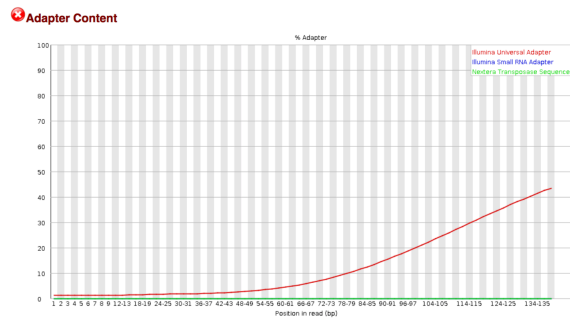


(d)



Figure 5.6 (a) Distribution of unique and duplicated reads of sWGS sequencing in STIC cohort samples after trimming. (b) Per sequence quality score shows the frequencies of quality scores in all samples after trimming. The quality of reads is good quality with the peak shifts to the right side. (c) Per base sequence quality presents the quality score across all bases at each position in the reads. (d) No samples were found with any adapter contamination $>0.1\%$.

(A)



(B)

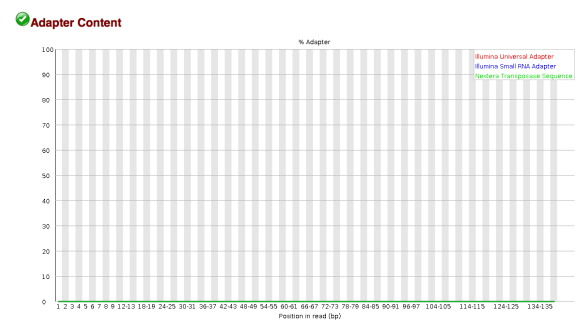


Figure 5.7 Represents adapter content distribution before and after trimming in one sample.
(A) FastQC: Adapter Content before trimming (B) FastQC: Adapter Content after trimming.

5.3.4 Absolute copy number profile of evolution process in HGSC

As described in chapter 3, I used sWGS-absolute and ACE packages to calculate the absolute copy number of the early stage and late stage cohorts from sWGS. The sWGS-absoluteCN (swgs) package generates absolute copy number profiles from shallow whole genome sequencing data and *TP53* allele frequency. However, the input DNA for p53 signature and STIC lesions was very low, and it was difficult to make both sWGS and Ampliseq libraries for these samples. The ACE package is based on model fitting, calculating errors per segment based on the difference between segment value and closest value of an integer copy number, getting the result of a series combination of ploidy and cellularity^[118]. ACE calculated the error of the fit as the root mean square error or mean absolute error of all segments. The fitting process is executed over different ploidies or one fixed ploidy. Relative errors are calculated for each cellularity and the indicated ploidy, whilst the ploidy reflects the number of copies associated with the median segment value. Cellularities (and ploidies when variable) are reported when the error of the fit reaches a minimum. After obtaining cellularities and ploidies from best fit solutions, absolute copy numbers of segments and bins are calculated using the derived cellularity and ploidy.

Based on relative copy number profiles of p53 signatures and STIC lesions, I compared the number of segment sizes, relative errors, purity and ploidy in 30kb, 100kb, 500kb and 1000kb. Within the ACE package, segment size employs the Circular Binary Segmentation (CBS) algorithm to find change points in sequential data and merge similar neighbouring bins^[151]. Calculating the number of segment sizes could provide information about nucleotide-level normalisation of bin size and allow detection of CNVs with high resolution. As shown in Figure 5.8 (a), the number of segments is significantly different between 30kb and 1000kb, and also between 100kb and 1000kb. Too large a bin size will make the absolute copy

number profile have less resolution, while 30kb is too fragmented and brings lots of noise to low input DNA samples. Fig 5.9 shows an example of absolute copy number profile using the different bin sizes. I also made a comparison of the relative errors, purity and ploidy from four bin sizes and found no significant difference between them (Figure 5.8 (B)(C)(D)). I chose 100kb as the standard bin size for downstream analysis based on the previous comparison. The best fit solution of purity and ploidy was picked out from ACE by 100kb bin size. The absolute copy numbers were obtained from ACE by using “getadjustedsegments” function to get information on the actual segments in a handy data frame.

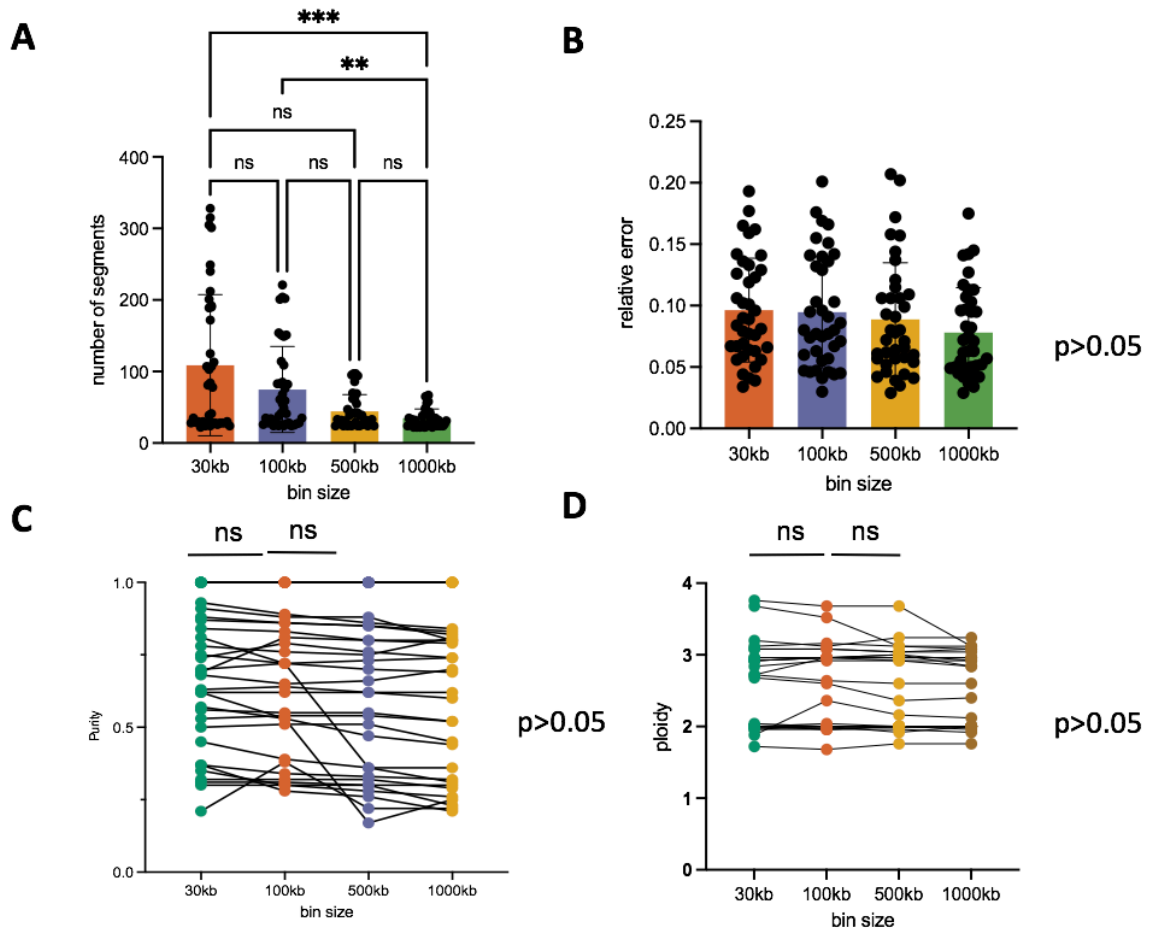


Figure 5.8 Comparison of copy number profiles of STIC project samples using different bin sizes (30kb, 100kb, 500kb, 1000kb) (A) Number of segments in different bin sizes. (B) Relative error distribution between four bin sizes. (C) (D) Purity and ploidy difference between different bin sizes., $p < 0.001$, ***, $p < 0.01$, **; Kruskal-Wallis tests.

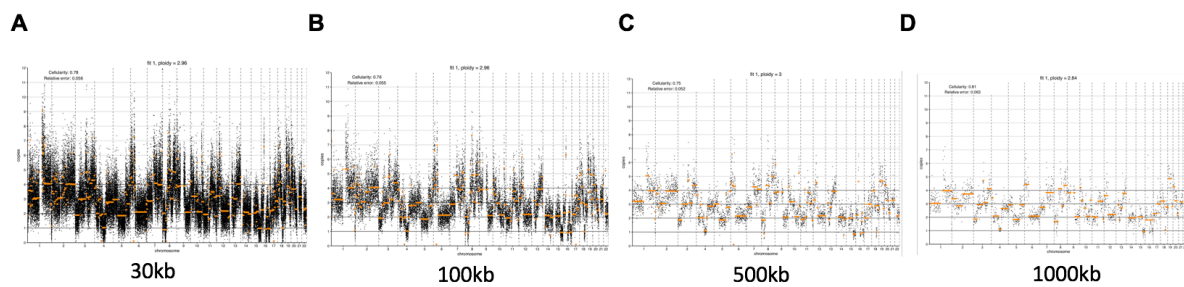


Figure 5.9. Absolute copy number profile of a representative STIC sample. The absolute copy numbers of a STIC sample were calculated by ACE using 30kb, 100kb, 500kb and 1000kb bin sizes.

5.3.5 Amplifications and deletions in the evolution process of HGSC

Technically it was difficult to get SNV information from low yield low quality FFPE STIC samples. However, it was possible to obtain gene amplification and deletions from the sWGS data. To further reveal the somatic copy number alterations (SCNA) in STIC patients, I called all samples SCNA by using GISTIC 2.0.23 on the GenePattern platform in all anatomic sites. The SEG files from ACE were used as an input with the following parameters: focal length cutoff set to 0.50, gene GISTIC set to yes, confidence level set to 0.90, cap value set to the infinite, broad analysis set to on (to call arm-level events), max sample segments set to 10,000, arm peel set to no, and gene collapse method set to extreme. Using these threshold data, the significant chromosomal amplifications and deletions were shown in Figure 5.10. In p53 signatures, I found significant chromosomal loss in peaks at cytobands 6q21.33, and amplification at 6q23 and 22q11.21. In STIC lesions, amplification peaks were found in 6q24, 11q14.1 and 19q13.13 and deletion peaks were found in 4q, 6p21, 8p12, 14q21, 15q25, 17q21. These locations were very similar to the GISTIC analysis of STIC samples from Mark et al, especially some focal amplifications peaked at 11q; included in this wide peak are cyclin D1 (*CCND1*) (11q13.2) and GRB2-associated binding protein 2 (*GAB2*) (Figure 5.10). *GAB2* functions in ErbB signalling and was recently shown to enhance mammary epithelial cell proliferation^[177]. Another commonly amplified gene is *CCNE1*, which is located in 19q12 and is associated with resistance to platinum and paclitaxel used as frontline therapies against HGSC^[178]. *CCNE1* amplification was found in 3/5 patients of our small STIC cohort.

Using GISTIC, I created a heatmap to show the copy number aberrations in all anatomic sites in the five patients (Figure.5.11). Further analysis will focus on common pathways and genes associated with the evolution of HGSC and the evolution tree in the five patients.

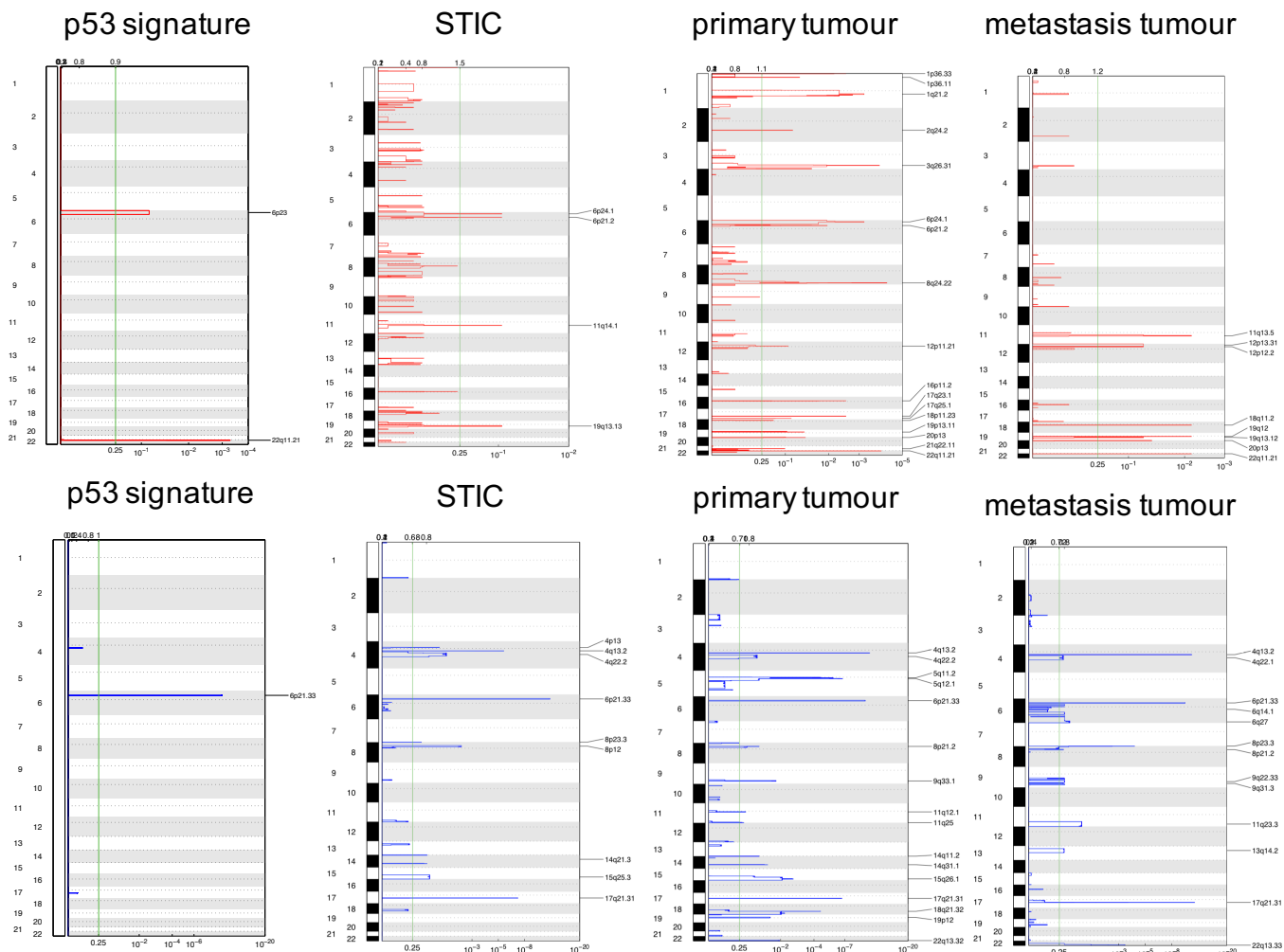


Figure 5.10 Focal copy number peaks specified for STIC project samples. Each peak's gain (red) and loss (blue) are shown. The x-axes represent the normalised amplification signal (top) and the significance of the q-value (bottom). The green line represents the significance cutoff at a q-value of 0.5.

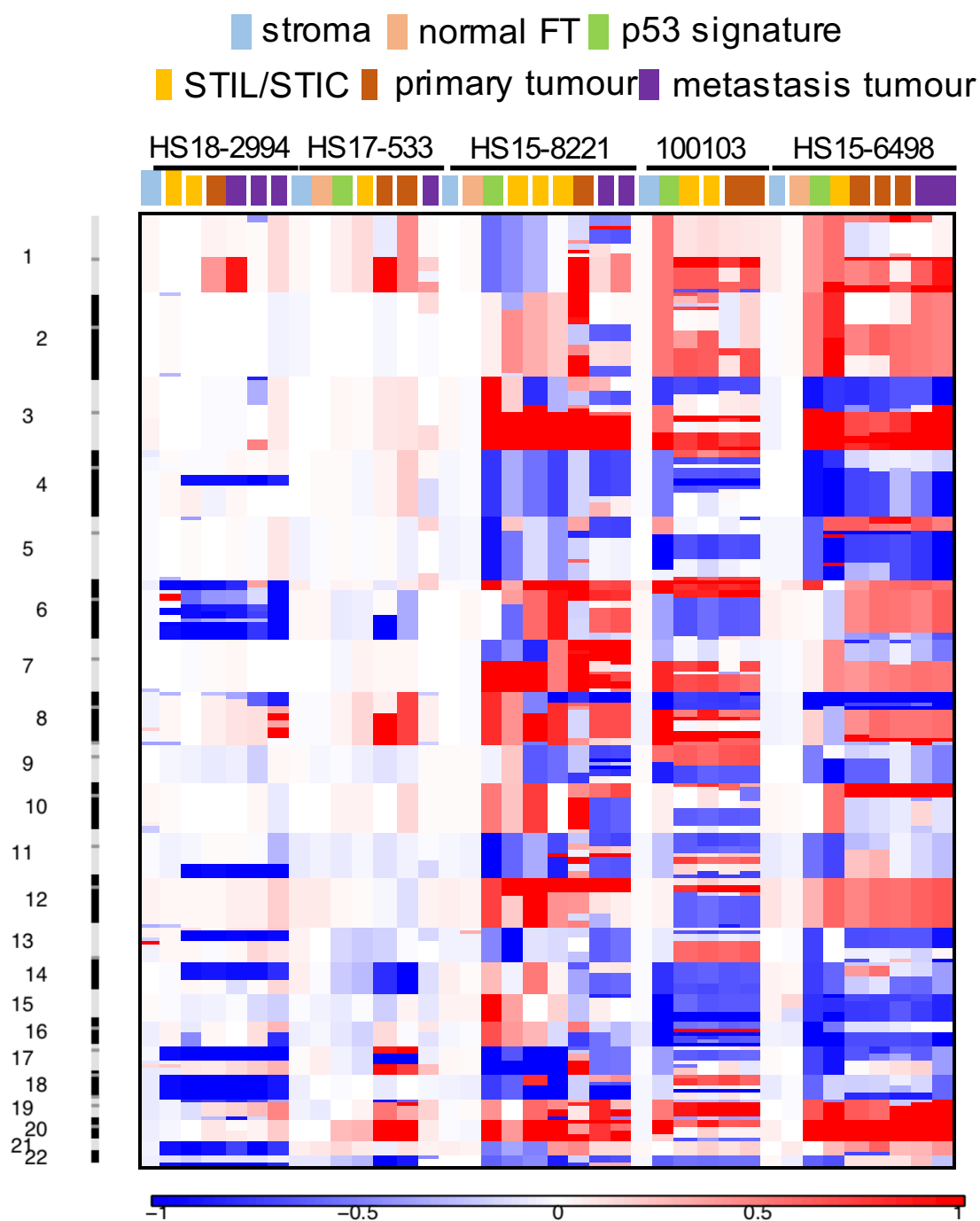


Figure 5.11 Genomic aberration plot of CNA across all anatomic sites and patients in five patients. Images were analysed using GISTIC_2.0 (v7) beta. In each heat map, the samples are arranged from left to right in stroma, normal fallopian tube, p53 signature, STIL/STIC, primary tumour and metastasis tumour. Chromosome numbers are indicated on the margin from top to bottom ordering. Gain is CN red, and CN loss is blue. The copy number aberration number ratio ranges from -1 to 1.

5.3.6 Ploidy change in the clonal evolution process of HGSC

I compared the ploidy between p53 signature, STIC, primary and metastasis tumours. In our sample cohort, the median ploidy was 2.0, 2.5, 2.7 and 2.6 respectively across these different lesions. The distribution of all samples in each lesion site is shown in Figure 5.12 (A).

Although the number of cases is small, the ploidy begins to change at STIC lesions. Figure 5.12 (B) shows that there is no different in ploidy across p53 signature, STIL/STIC, primary and metastasis tumours in patients HS17-533 and HS18-2994, both of whom had advanced disease. However, in the remaining patients, including 100103 who had stage IC disease, p53 signatures were diploid, while ploidy increased to a median of approximately 2.8 in STIC lesions. In the patients with high ploidy in STIC lesions, the same increased ploidy was observed in primary and metastasis tumours. These results suggest that the status of STIC ploidy could be a decisive event in the evolution of HGSC. Interestingly, in patient HS15-8221, I analysed three separate STIC samples. Two samples had ploidy of c.2.6, which was similar to that seen in the primary tumour and metastasis, whilst one was diploid.

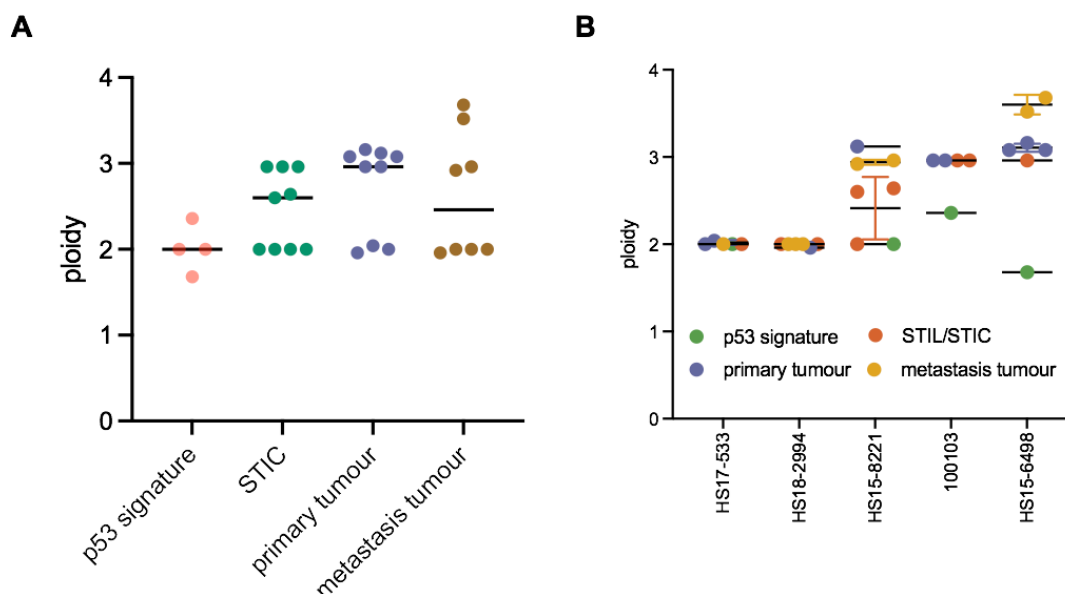


Figure 5.12 Ploidy distribution in STIC project. (A) Frequencies of ploidy by anatomic site. (B) Ploidy comparison of each anatomy site in the five patients.

5.4 Discussion

5.4.1 Whole exome sequencing and shallow whole genome sequencing comparison in the evolution of HGSC

My results provide an evolutionary analysis of multi-site HGSC in five patients. Most previous research on STIC and evolutionary analysis on STIC was performed by whole exome sequencing. My research is the first attempt to apply shallow whole genome sequencing on p53 signature and STIC lesions to explore the genome-wide copy number landscape of early HGSC.

Labidi-Galy et al investigated p53 signatures, STIC, fallopian tube carcinoma, primary and metastases from nine patients by whole exome sequencing. They found that the majority of tumour specific alterations in HGSC were present in STIC. Furthermore, evolutionary analyses revealed that p53 signatures and STICs are early events of ovarian carcinoma. They reported that it may take over seven-years from time of developing a STIC to the initiation of ovarian carcinoma, with metastases following rapidly after^[48]. Eckert et al^[172] also performed whole exome sequencing on specimens from eight HGSC patients, which contained STIC, primary tumours and omental metastases. Similar to Labidi-Galy's result, they also found similar mutational signatures and copy number variation profiles across all anatomic sites, proving genomic instability at early stage HGSC.

However, whole exome sequencing only captures the targets at coding regions of the genome, which only accounts for 1-2% of the whole genome^[179]. In contrast, whole genome sequencing does not require targeted enrichment and provides more information about the samples' ploidy state. As I described in chapter 1, HGSC is a copy number-driven cancer type, and it remains unclear how early the ploidy changes during evolution. By using shallow

whole genome sequencing, I could identify the ploidy dynamics changes and the specific chromosomes that have driven the development of HGSC in the five patients.

5.4.2 Ploidy in the evolution process in HGSC

Consistent with whole exome sequencing of Labidi-Galy et al and Eckert et al^[172], I also found that genomic changes characteristic of HGSC were present in STIC. Interestingly, not all patients have the same ploidy changes, and the evolution trajectories were also different. In cases where the STIC lesions had ploidy of around 2, the primary and metastasis sites also had similar ploidies, while in the cases where STIC samples had high ploidy, the primary and metastasis sites demonstrated ploidy values that were even higher than observed in the STIC. These findings suggested that the ploidy level of STIC could influence the future evolution status in the primary and metastasis tumours. There were no significant ploidy changes in the p53 signature across patients, which indicated the ploidy transformation is likely to happen during the evolution process from p53 signature to STIC. STIL lesions occupy an intermediate status between p53 signature and STIC, and are morphologically p53 mutant but show low proliferation^[180]. I was unable to separate STIL from STIC lesions due to the limited input DNA. However, evaluation of ploidy changes in STIL lesions will be an important future investigation.

One patient, HS15-8221, had five STIC lesions. Due to their size and proximity, these were captured as three samples – STIC1, STIC2+3 and STIC4+5. One sample (STIC4+5) was diploid, whilst the other two had ploidy of c.2.6, which was similar to the primary ovarian mass and the metastatic tumour. This suggests either that STIC4+5 evolved into the others or that these were separate and discrete tumours that developed in parallel. In this patient it

would be most important to identify the *TP53* mutation in all STIC samples to address whether these were, in fact, separate and parallel tumours.

5.4.3 High ploidy/whole genome duplication associated with evolution process of HGSC

As we described in chapter 4, whole genome duplication (WGD) is observed in many solid malignancies^[143] and is generally associated with poor prognosis. Across all samples, the median ploidy was approximately 2.5 in STIC lesions, indicating that whole genome duplication may happen as early as the STIC stage. The critical question that I described in chapter 4 is whether the processes that generate high ploidy are the primary drivers of rapid dissemination in HGSC, or whether high ploidy/WGD is simply a time-related marker of evolutionary fitness and is thus more likely to be observed in late-stage than early-stage disease. The presence of high ploidy in STIC lesions and consistent ploidy between STIC and both primary and metastatic tumours suggests strongly that high ploidy/WGD might not be simply time-related but rather an early event in HGSC development.

However, without deep WGS and knowledge of the allele frequency of coding genes, it is challenging to assess both ploidy and extent of LOH^[163], and difficult to determine the origin of different STICs and whether they evolved from separate pure clones or different clone combinations. This is one limitation of my findings. In addition, the results from Ampliseq from primary and metastasis tumours will provide more information about these patients' allele frequency of *TP53* and *BRCA1/2* status. I will also be able to call copy number signatures from primary and metastasis sites to compare the copy number signatures status with my previous early stage and late stage cohorts.

5.4.4 Evolution process at anatomic sites of five HGSC patients

In pan-cancer analyses, the commonest genomic correlate of WGD was a mutation in *TP53*, which usually precedes the duplication, and *CCNE1* amplification and loss of *RBI* ^[163].

Although I do not have the *TP53* allele frequency data in p53 signature and STIC lesions, the immunochemistry staining shows the strong overexpression of p53, in keeping with *TP53* mutation being a very early event in HGSC. *CCNE1* is located on chromosome 19, and Kuhn et al^[80] have found a strong correlation of *CCNE1* copy number between serous tubal intraepithelial carcinoma and high-grade serous carcinoma in the same patient ($p < 0.001$)^[80]. This result is consistent with my analysis for all anatomic sites from STIC, primary and metastasis tumours. Further investigation will focus on the focal genes associated with whole genome duplication in the evolution process of HGSC. In addition, the copy number signature exposures in the primary and metastasis sites could also provide information about the status of the p53 signature and STIC lesions.

5.5 Summary

I collected 38 samples from five HGSC patients with identifiable STIC lesions to identify potential differences in copy number aberrations and ploidies between all anatomic sites. No significant copy number abnormalities were identified in normal fallopian tube and stroma, indicating the normal genome in these samples. p53 signatures have few copy number aberrations change, but some amplifications and deletions were found in chromosomes 6 and 22. The STIC lesions have multiple chromosome changes in my cohort. In addition, the ploidy is markedly higher in STIC than p53 signature, while there were also some differences between each STIC in ploidy. Higher ploidy is consistent with whole genome duplication. The ploidies in primary and metastasis tumours also correlate with those in STIC lesions, suggesting that STIC, primary, and metastasis tumours share fundamental genomic features. The ploidy change may happen as early as the STIC stage, with evidence of whole genome duplication that may provide evolutionary benefit. Overall, the higher ploidy in STIC lesions could provide evidence for the tumour progression in HGSC. However, the biological differences between primary and metastasis tumours may not be time-related markers of evolutionary fitness.

CHAPTER 6. General discussion

6.1 The genomics of evolution of high grade serous carcinoma

6.1.1 The evolution model in HGSC and other cancer types

The work presented in this thesis has investigated the genomic landscape of early stage and precursor lesions of HGSC using shallow whole genome sequencing and targeted panel sequencing. I found no significant differences in our selected panel of somatic mutations or focal copy number alterations between early-stage and late-stage cohorts. Genome-wide copy number abnormality analysis also shows no difference between the two cohorts in chapter 3. Chapter 3 also shows high percentage of *TP53* hotspots in early stage cohort. Mutant p53 protein is classified as “contact” or “structural” function, depending on whether the mutated amino-acid is directly interacts with DNA or causes a global effect on the protein structure. R273H, R248Q and R248W are classified as contact mutations and R175H, G245S, R249S and R282H belongs to conformational mutations^[181]. It will be interesting for further investigation about the relationship between TP53 hotspot types with tumour progression in high grade serous carcinoma.

However, the ploidy was markedly higher in late stage than early stage cohort, and higher signature 4 exposure was found in the late stage cohort, consistent with whole genome duplication. These data show that early and late-stage HGSC share fundamental genomic features. However, late-stage disease appears distinct from early-stage, with evidence of whole genome duplication that may provide evolutionary benefit. It remains an open question whether the ploidy difference reflect biological differences between early and late stage HGSC or simply time-related markers of evolutionary fitness. Chien et al investigated the genetic evidence for early peritoneal spreading in pelvic high grade serous carcinoma and challenged that the HGSC evolution is a stepwise spatial progression from the fallopian primary to ovarian to peritoneal dissemination^[182]. Their research suggested an alternative

progression model where early clones occurs before or in parallel with ovarian metastases. These results also indicates the ploidy changes could happen at early clone and drives the evolution of HGSC.

For further investigation in STIC project, I found no significant copy number abnormalities in normal fallopian tube and stroma, while p53 signatures start to have a few copy number aberrations. The STIC lesions have multiple chromosomal changes, and ploidy is elevated in some STICs from the five patients. The ploidies in STIC lesions are consistent with that seen in primary and metastatic tumours from the same patients, suggesting that STIC, primary, and metastasis tumours share fundamental genomic features. The results indicate the ploidy changes may happen as early as the STIC stage, suggesting that whole genome duplication that may provide very early evolutionary benefit for the tumour progression. These analyses suggest that biological differences between primary and metastasis tumours may not be time-related markers of evolutionary fitness, and the genome pattern of HGSC development may be determined as early as STIC lesions.

These results raise several interesting new questions for further discussions. First of all, high grade serous carcinoma is the most common type and has lethal gynecologic malignancy in epithelial ovarian cancer. Whether the mechanisms of genomic evolution observed in HGSC happens in other cancers or epithelial ovarian cancer types is highly important.

Current theories provide several evolution models that may explain cancer transformation and progression. For example, branching evolution is a model in which clones diverge from a common ancestor and evolve in parallel in the tumour mass, resulting in multiple clonal lineages^[183]. Neutral evolution is an extreme case of branching evolution, which further

hypothesises that there is no selection or fitness changes during most of the lifetime of the tumour^[184]. The linear evolution model posits that only one lineage survives over time. However, it does not imply that there was only ever a single lineage that evolved in a stepwise fashion. The previously described models commonly assume that mutations are acquired sequentially and gradually over time. However, some cancers provide recent evidence that many genomic aberrations may occur at the earliest stages of tumour progression^[185]. This knowledge informs a model for punctuated tumour evolution relevant to cancer development and possibly other malignancies^[186]. The summary of each evolutionary model is shown in Figure 6.1.

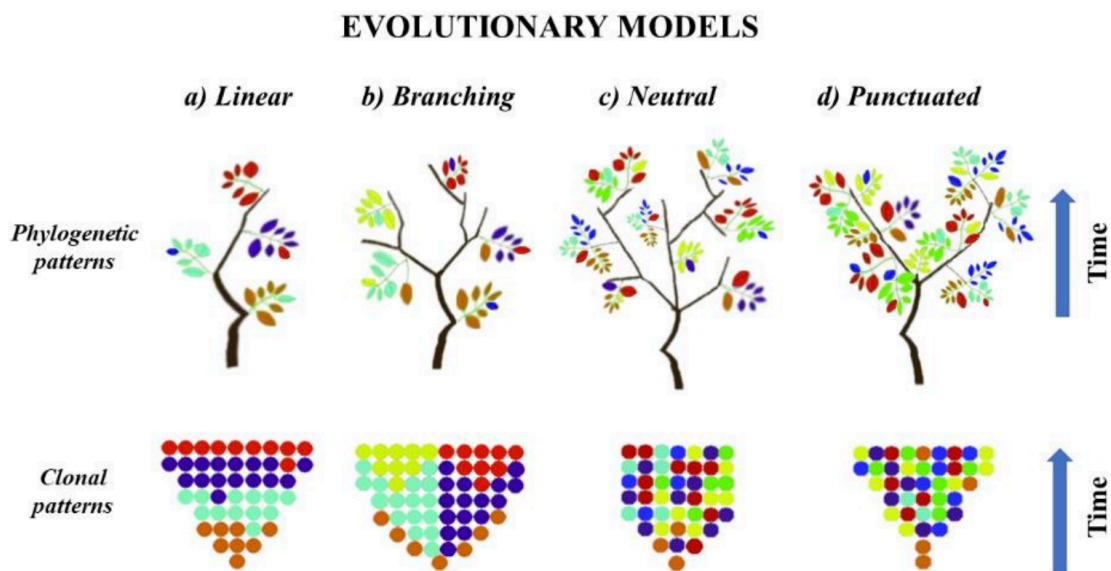


Figure 6.1 Phylogenetic and clonal patterns figure in the tumour evolution process^[187].

Some cancer types have been suggested to follow the punctuated tumour evolution model. Baca et al. sequenced 57 prostate tumour genomes and matched normal tissues to characterise somatic alterations and aimed to study how they accumulate during oncogenesis and progression. They identified abundant DNA translocations and deletions that arose in highly interdependent manner. They termed this phenomenon as “chromoplexy”, which may induce considerable genomic derangements happen over relatively few events, supporting a

model of punctuated cancer evolution in prostate cancer^[186]. Besides prostate cancer, most copy number aberrations are acquired at the earliest stages of tumour evolution in triple-negative breast cancer, with punctuated bursts, followed by stable clonal expansions that form the tumour mass^[188]. My current analysis shows evidence that high grade serous carcinoma might also follow the rule of punctuated cancer evolution, but it needs further analysis for further validation. Overall, these genomic researches indicate that some cancer types such as prostate, triple-negative breast cancer, and high grade serous carcinoma may share similar punctuated cancer evolution paths.

6.1.2. The technical challenges and applicable methods to detect HGSC early

As we have shown before, the key genomic changes in HGSC could occur as early as STIC lesions. According to the analysis of Labidi-Galy et al^[48], the timeline from STIC to invasive carcinoma could be a window of 7 years. Another recent genomic analysis provided striking evidence that the p53 signature of STIC had an ancestral clonal relationship with HGSC. An estimation of the time sequence of their development based on epidemiological, molecular and NGS studies suggests ten years window from the normal tubal epithelium to p53 signature, then another 15 years from p53 signature to STIC and a final 6+ years from STIC to ovarian HGSC^[189].

How to detect HGSC earlier is a challenging question. There is no screening programme for epithelial ovarian cancer anywhere in the world. Current methods to detect epithelial ovarian cancer patients include serum biomarker CA125 and transvaginal sonography (TVS)^[190]. The United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) conducted a large cohort with more than 200,000 postmenopausal women aged 50 – 74 years at average risk. These women were randomised into three groups and followed initially for 14 years^[191].

The three groups were multimodal screening (MMS; longitudinal CA125 and second-line TVS), ultrasound screening (USS; TVS first and second-line test), and a no screening group. Overall, the incidence of stage I or II diseases was higher in the MMS group than in the no screening group ((29.2% invasive epithelial cancers were identified at stage I and II in MMS group compared to 20.4% in the no screening group). Yet, the researchers did not find a significant reduction in ovarian and tubal cancer deaths in the MMS or USS groups compared with the no screening group. The findings suggest that either that this stage shift was insufficient to influence overall survival or that there is a fundamental difference between screen-detected and clinically-detected stage I and II cancers. My data certainly suggest that the latter may be the case, and raises the possibility that earlier diagnosis of invasive epithelial ovarian and tubal cancer in the symptomatic population may not reduce mortality significantly^[191].

Although the extensive population screening using traditional methods did not significantly reduce ovarian and tubal cancer deaths, some new techniques have come out for early detection of HGSC. Noncoding RNA, such as miRNAs in serum, has the potential to correlate with cancer proliferation, migration, invasion, and chemosensitivity and can be used as biomarkers of HGSC^[192]. Cell-free DNA (cfDNA) is another biomarker in blood and cervical secretions, in which cancer-specific mutations can be identified by technologies such as digital PCR (dPCR) and next generation sequencing (NGS)^[193]. The published empirical data correlated tumour size with the expected amount of cfDNA in a 10-mL blood sample and found the relationship between mutant allele fraction in the percentage of circulating tumour DNA compared with the total cfDNA to the likelihood of tumour detection using ctDNA genomic analysis^[194]. A recent study developed a repetitive element aneuploidy sequencing system called RealSeq. This system could detect aneuploidy in

cfDNA and improve sensitivity for early-stage high-grade ovarian cancers with copy-number abnormalities from as little as three pg of DNA in relatively small amounts of plasma^[195]. However, the distribution of p53 signature/STIC in a normal population remains unclear, and the frequency of p53 signature/STIC in the normal fallopian tube in a woman's lifetime is also not known. There is no applicable method to detect the status of p53 signature or STIC lesion on the fallopian tube unless using SEE-FIM techniques for IHC identification – these lesions can only be detected in surgical resection specimens. Applying RealSeq to postmenopausal women to further investigate the sensitivity and accuracy of aneuploidy changes in ctDNA will be interesting. It is challenging to detect the p53 signatures and STIC lesions directly but it will be important to compare the primary tumours in patients who later develop into HGSC to determine how early the aneuploidy changes in the population.

6.1.3. Different immunohistochemistry staining patterns at STIC lesions

Another interesting phenomenon I observed in collecting STIC samples was that the relatively high percentage of STIC lesions have a complete absence/null p53 immunohistochemistry pattern. There are three common p53 immunohistochemistry staining patterns in HGSC; overexpression (OE), complete absence (CA) or wild-type (WT)^[61]. Each immunohistochemistry staining pattern reflects mutations in *TP53*. OE is most commonly associated with nonsynonymous/missense *TP53* mutations, interfering with MDM2-induced ubiquitination and degradation of p53. This process causes the accumulation of p53 in the nucleus. CA is associated with nonsense mutations produced by premature stop codon that triggers nonsense-mediated RNA decay^[196], or indel and splice acceptor mutations that interfere with correct protein translation by introducing frameshifts or aberrant splicing^[61]. The different p53 staining patterns may also reflect distinct biological effects^[133].

Nonsynonymous *TP53* mutations could induce gain of function (GOF) effects and upregulate chromatin regulatory genes, resulting in genome-wide increases of histone methylation and acetylation^[197]. Compared to the OE type, CA is related to *TP53* loss of function mutations, which has been reported to have weaker tumourigenesis ability than GOF mutations in some mouse models^[198]. Given the different immunohistochemistry staining patterns at the start of HGSC evolution, it will be important to research on how the evolution happens according to these staining types. The predominant immunohistochemistry staining pattern in my STIC samples was CA. Therefore, it is worth comparing with OE patients. Utilising sWGS or RNA-sequencing will provide more evidence about how the diversity of HGSC evolution varies according to different immunohistochemistry staining patterns in STIC lesions. This research could also give suggestions for the clinical treatment of patients with CA immunohistochemistry staining patterns in STIC lesions.

6.1.4 Microenvironment of the fallopian tube

The dynamic interactions of cancer cells with their microenvironment consist of stromal cells (cellular part) and extracellular matrix (ECM) components (non-cellular). The microenvironment is essential to stimulate cell heterogeneity and clonal evolution and increase multidrug resistance ending in cancer cell progression and metastasis^[199]. However, little is known about the role the microenvironment in the progression of normal fallopian tube and how the microenvironment supports p53 signatures and STICs or the transition from this state to metastasis^[200]. Hu et al. performed deep single-cell RNA-seq on thousands of nonmalignant epithelial cells from the FTE and classified all the cells into five subtypes. Epithelial-mesenchymal transition (EMT) cells are one of the five subtypes and are strongly linked with poor prognosis^[201]. Separately, Martincorena et al. also found a high burden and pervasive positive selection of somatic mutations in normal human skin. The clonal

expansions induced by driver mutations in normal tissue are critical to understanding the evolution of cancer. The allele frequency of mutation reflects how this mutation correlates accurately with the fraction of cells in a biopsy that carry the mutation and enables the estimate of clone sizes^[202]. However, as I collected stroma and normal fallopian tube epithelial cells from FFPE blocks, the copy number profiles were very normal, without significant amplification and deletions. These results indicate that somatic mutations are likely to be the first step in HGSC development and happen as early as normal fallopian tube epithelial and stromal cells. Chromosome changes are the next step of evolution.

Apart from stroma, the ECM of fallopian tube has been unexplored for STIC development and metastasis. Still, limited studies suggest that the ECM patterns are altered and may influence progression. For example, Kuhn et al. found that laminin γ 1 is highly upregulated in *TP53*-mutated FTE cells relative to normal FTE cells^[203]. Laminin γ 1 is a glycosylated protein enriched in ECM^[204]. These results suggest that the somatic mutations and specific pathways change very early, potentially even before the p53 signature. Therefore, it is worth further research on how normal cells and microenvironment affect the evolution of HGSC.

6.2 Technical challenges for sequencing the early lesions of HGSC

The genomic analysis provides a deeper understanding of copy number variation in disease. Still, most tissues are a mixed population of cells, and most current methods cannot distinguish the unique character of each cell. As epithelial ovarian cancer development is a highly heterogeneous process, it is important to understand how early stage HGSC arises and reconstruct evolutionary history. Single-cell genomic sequencing is an exciting technology that has shed light on biological questions that were previously inaccessible in exploring tumour evolution^[205]. Navin et al. developed a single-cell sequencing technique with flow-sorted nuclei, whole genome amplification and next-generation sequencing to quantify

genomic copy numbers within the individual nucleus. Comparing breast cancer cell line SK-BR-3 single cells to bulk populations, they found that the copy number profile was similar between the two samples^[206]. However, single-cell genomic approaches have been limited mainly to fresh/frozen samples because of chemical alterations from formalin fixation. It remains a significant challenge to detect whole genome copy number profile from the FFPE block and particular location in the block. Mastelotto et al. developed a robust approach to perform single-cell whole-genome copy number (CN) profiling from FFPE tissue in 11 steps. The 11 steps involve tissue microdissection, nuclei preparation and FACS analysis/sorting based on DAPI staining/DNA content, single-nucleus DNA repair to correct FFPE-induced DNA damage, whole-genome amplification (WGA), Illumina library preparation and multiplex sequencing (steps 6 – 10), and bioinformatics analysis. The most important steps are nuclei preparation and FACS analysis/sorting, which isolate cells with different ploidies^[207]. However, the differences between p53 signature and STIC are mixture tissues based on the identification of morphology, p53 and Ki67 expression percentage on immunochemistry, which are difficult to be detected by FACS analysis.

Recently, Trapani et al. developed a DEPArray System, which provided a method for automatic image-based sorting to isolate single cells. DEPArray System comprises three elements: a benchtop instrument, a disposable microfluidic cartridge and CellBrowser software. Initially, FFPE tissue was dissociated into a cell suspension, and immunofluorescence staining by anti-keratin and anti-vimentin antibodies was performed. Non-uniform electric field generated in the silicon chip induced a polarisation of cells, which trapped the cells in stable levitation. Then, the selected cells are moved toward a so-called parking chamber. After a washing and priming phase of the exit chamber, the cells are moved from the parking chamber to the exit chamber for cell recovery.

The FFPE sections that I utilised were stained with p53 or Ki67 to identify the p53 signature and STIC lesions, so I devised a new idea to modify the protocol from the DEPArray system and adapt it to my project more. I have tried using laser capture microdissection to collect p53 signature and STIC lesions. These tiny pieces of tissue were dissociated into several cell suspensions by the enzyme provided by the DEPArray FFPE kit. Then I utilised whole-genome amplification (WGA) to amplify the DNA from each cell suspension and will use the amplified libraries for sequencing. This protocol is still under investigation, and I have tried several pilot experiments from normal tissues. It has some restrictions on the input cell numbers and needs further refinement.

Apart from single-cell DNA sequencing, it will also be important to know how evolution progresses from normal epithelial to p53 signature, STIL and STIC lesions at the transcriptome level. However, major challenges remain with in FFPE tissues; they are incompatible with single-cell transcriptome sequencing because of RNA degradation in storage and RNA damage in extraction. Spatial transcriptome sequencing is a newly emerging tool to perform transcriptome sequencing at a cellular level while corresponding to the original location and morphology in the FFPE block^[208]. Applying spatial transcriptome sequencing to fallopian tube blocks which consist of stroma, normal fallopian tube, p53 signature, STIC and primary tumours will also be a great opportunity to investigate the driver mutations and pathways that play crucial roles in HGSC development process.

6.3 Limitations of early stage and STIC projects

As a retrospective study, the current work may inevitably contain bias during the sample collection. First, all of our samples were derived from FFPE samples and we did not have

match normal sample from patient blood. As I described in the chapter 1, the current method to identify STIC lesions uses the SEE-FIM technique, which can only be used for FFPE blocks. HGSC development is a dynamic process but I could only get genomic information from patients at single specific time points from FFPE samples. In addition, finding STIC depends on how hard the pathologists look for them.

Another shortcoming of my STIC project is lacking allele frequency of *TP53* in p53 signature and STIC lesions. The lowest input DNA for Ampliseq sequencing is 10ng, and it is challenging to provide enough DNA for both sWGS and Ampliseq from such small lesions. Furthermore, without *TP53* allele frequency and mutation type, it is hard to determine whether the STIC lesions come from the same clone or a mixture of separate clones and identify the clonal evolution process in the development of HGSC.

6.4 Future perspectives

A further research objective will include the analysis of the genomic evolution process of p53 signature, STIC, and primary and metastasis tumours in sWGS and Ampliseq data. The result of Ampliseq will give the opportunity of identifying copy number signature exposures in the primary and metastasis tumours. Thus, the exposure of each signature of CN signatures could be inferred in the early lesions of HGSC. Besides, I will also make the phylogenetic tree of the five patients and identify driver players and pathways in this process.

CHAPTER 7. References

1. Macintyre, G., et al., *Copy number signatures and mutational processes in ovarian carcinoma*. Nat Genet, 2018. **50**(9): p. 1262-1270.
2. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012*. Int J Cancer, 2015. **136**(5): p. E359-86.
3. Reid, B.M., J.B. Permuth, and T.A. Sellers, *Epidemiology of ovarian cancer: a review*. Cancer Biol Med, 2017. **14**(1): p. 9-32.
4. Penny, S.M., *Ovarian Cancer: An Overview*. Radiol Technol, 2020. **91**(6): p. 561-575.
5. Vaughan, S., et al., *Rethinking ovarian cancer: recommendations for improving outcomes*. Nat Rev Cancer, 2011. **11**(10): p. 719-25.
6. Oza, A.M., et al., *Standard chemotherapy with or without bevacizumab for women with newly diagnosed ovarian cancer (ICON7): overall survival results of a phase 3 randomised trial*. Lancet Oncol, 2015. **16**(8): p. 928-36.
7. Khazaei, Z., et al., *Worldwide incidence and mortality of ovarian cancer and Human Development Index (HDI): GLOBOCAN sources and methods 2018*. J Prev Med Hyg, 2021. **62**(1): p. E174-E184.
8. Allemani, C., et al., *Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2)*. Lancet, 2015. **385**(9972): p. 977-1010.
9. Sankaranarayanan, R. and J. Ferlay, *Worldwide burden of gynaecological cancer: the size of the problem*. Best Pract Res Clin Obstet Gynaecol, 2006. **20**(2): p. 207-25.
10. Testa, U., et al., *Ovarian Cancers: Genetic Abnormalities, Tumor Heterogeneity and Progression, Clonal Evolution and Cancer Stem Cells*. Medicines (Basel), 2018. **5**(1).
11. Telloni, S.M., *Tumor Staging and Grading: A Primer*. Methods Mol Biol, 2017. **1606**: p. 1-17.
12. Cortez, A.J., et al., *Advances in ovarian cancer therapy*. Cancer Chemother Pharmacol, 2018. **81**(1): p. 17-38.
13. Arora T, M.S., Lekkala MR. , *Ovarian Cancer*. StatPearls 2022.
14. Stewart, C., C. Ralyea, and S. Lockwood, *Ovarian Cancer: An Integrated Review*. Semin Oncol Nurs, 2019. **35**(2): p. 151-156.
15. Bookman, M.A., et al., *Evaluation of new platinum-based treatment regimens in advanced-stage ovarian cancer: a Phase III Trial of the Gynecologic Cancer Intergroup*. J Clin Oncol, 2009. **27**(9): p. 1419-25.
16. Collinson, F., et al., *Optimal treatment of early-stage ovarian cancer*. Ann Oncol, 2014. **25**(6): p. 1165-71.
17. Trimbos, B., et al., *Surgical staging and treatment of early ovarian cancer: long-term analysis from a randomized trial*. J Natl Cancer Inst, 2010. **102**(13): p. 982-7.
18. Trimbos, J.B., et al., *International Collaborative Ovarian Neoplasm trial 1 and Adjuvant ChemoTherapy In Ovarian Neoplasm trial: two parallel randomized phase III trials of adjuvant chemotherapy in patients with early-stage ovarian carcinoma*. J Natl Cancer Inst, 2003. **95**(2): p. 105-12.
19. Lawrie, T.A., et al., *Adjuvant (post-surgery) chemotherapy for early stage epithelial ovarian cancer*. Cochrane Database Syst Rev, 2015(12): p. CD004706.
20. *ICON2: randomised trial of single-agent carboplatin against three-drug combination of CAP (cyclophosphamide, doxorubicin, and cisplatin) in women with ovarian cancer*. ICON Collaborators. International Collaborative Ovarian Neoplasm Study. Lancet, 1998. **352**(9140): p. 1571-6.

21. McGuire, W.P., et al., *Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer*. N Engl J Med, 1996. **334**(1): p. 1-6.
22. Ozols, R.F., et al., *Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally resected stage III ovarian cancer: a Gynecologic Oncology Group study*. J Clin Oncol, 2003. **21**(17): p. 3194-200.
23. Clamp, A.R., et al., *Weekly dose-dense chemotherapy in first-line epithelial ovarian, fallopian tube, or primary peritoneal carcinoma treatment (ICON8): primary progression free survival analysis results from a GCIg phase 3 randomised controlled trial*. Lancet, 2019. **394**(10214): p. 2084-2095.
24. Katsumata, N., et al., *Dose-dense paclitaxel once a week in combination with carboplatin every 3 weeks for advanced ovarian cancer: a phase 3, open-label, randomised controlled trial*. Lancet, 2009. **374**(9698): p. 1331-8.
25. Bois, A.d., et al., *Addition of Epirubicin As a Third Drug to Carboplatin-Paclitaxel in First-Line Treatment of Advanced Ovarian Cancer: A Prospectively Randomized Gynecologic Cancer Intergroup Trial by the Arbeitsgemeinschaft Gynaekologische Onkologie Ovarian Cancer Study Group and the Groupe d'Investigateurs Nationaux pour l'Etude des Cancers Ovariens*. Journal of Clinical Oncology, 2006. **24**(7): p. 1127-1135.
26. Perren, T.J., et al., *A phase 3 trial of bevacizumab in ovarian cancer*. N Engl J Med, 2011. **365**(26): p. 2484-96.
27. Burger, R.A., et al., *Incorporation of bevacizumab in the primary treatment of ovarian cancer*. N Engl J Med, 2011. **365**(26): p. 2473-83.
28. Moore, K., et al., *Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer*. N Engl J Med, 2018. **379**(26): p. 2495-2505.
29. Gonzalez-Martin, A., et al., *Niraparib in Patients with Newly Diagnosed Advanced Ovarian Cancer*. N Engl J Med, 2019. **381**(25): p. 2391-2402.
30. Kalloger, S.E., et al., *Calculator for ovarian carcinoma subtype prediction*. Mod Pathol, 2011. **24**(4): p. 512-21.
31. Pierson, W.E., et al., *An integrated molecular profile of endometrioid ovarian cancer*. Gynecol Oncol, 2020. **157**(1): p. 55-61.
32. Groen, R.S., D.M. Gershenson, and A.N. Fader, *Updates and emerging therapies for rare epithelial ovarian cancers: one size no longer fits all*. Gynecol Oncol, 2015. **136**(2): p. 373-83.
33. Hollis, R.L., et al., *Molecular stratification of endometrioid ovarian carcinoma predicts clinical outcome*. Nat Commun, 2020. **11**(1): p. 4995.
34. Iida, Y., et al., *Clear cell carcinoma of the ovary: a clinical and molecular perspective*. Int J Gynecol Cancer, 2021. **31**(4): p. 605-616.
35. Wiegand, K.C., et al., *ARID1A mutations in endometriosis-associated ovarian carcinomas*. N Engl J Med, 2010. **363**(16): p. 1532-43.
36. Perren, T.J., *Mucinous epithelial ovarian carcinoma*. Ann Oncol, 2016. **27 Suppl 1**: p. i53-i57.
37. Mills, A.M. and E.D. Shanes, *Mucinous Ovarian Tumors*. Surg Pathol Clin, 2019. **12**(2): p. 565-585.
38. Prat, J. and F.C.o.G. Oncology, *FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum: abridged republication*. J Gynecol Oncol, 2015. **26**(2): p. 87-9.

39. Morice, P., S. Gouy, and A. Leary, *Mucinous Ovarian Carcinoma*. N Engl J Med, 2019. **380**(13): p. 1256-1266.
40. Modesitt, S.C., et al., *Ovarian and extraovarian endometriosis-associated cancer*. Obstet Gynecol, 2002. **100**(4): p. 788-95.
41. Matsuo, K., et al., *Trends of low-grade serous ovarian carcinoma in the United States*. J Gynecol Oncol, 2018. **29**(1): p. e15.
42. Vineyard, M.A., et al., *Is low-grade serous ovarian cancer part of the tumor spectrum of Hereditary Breast and Ovarian Cancer?* Gynecologic Oncology, 2011. **120**(2): p. 229-232.
43. Angarita, A.M., D. Cholakian, and A.N. Fader, *Low-grade serous carcinoma: molecular features and contemporary treatment strategies*. Expert Rev Anticancer Ther, 2015. **15**(8): p. 893-9.
44. Malpica, A., et al., *Grading ovarian serous carcinoma using a two-tier system*. Am J Surg Pathol, 2004. **28**(4): p. 496-504.
45. Manu, V., et al., *Serous ovarian tumors: Immunohistochemical profiling as an aid to grading and understanding tumorigenesis*. Med J Armed Forces India, 2020. **76**(1): p. 30-36.
46. Ferguson, D.C., et al., *Comparative analysis of Rb1, P16 and ER as diagnostic, prognostic and potential targets for therapeutic agents in ovarian epithelial tumors: an immunohistochemical study of 130 ovarian carcinomas*. J Ovarian Res, 2015. **8**: p. 34.
47. Xiang, L. and B.H. Kong, *PAX8 is a novel marker for differentiating between various types of tumor, particularly ovarian epithelial carcinomas (Review)*. Oncology Letters, 2013. **5**(3): p. 735-738.
48. Labidi-Galy, S.I., et al., *High grade serous ovarian carcinomas originate in the fallopian tube*. Nat Commun, 2017. **8**(1): p. 1093.
49. Saini, U., et al., *STAT3/PIAS3 Levels Serve as "Early Signature" Genes in the Development of High-Grade Serous Carcinoma from the Fallopian Tube*. Cancer Research, 2018. **78**(7): p. 1739-1750.
50. Cancer Genome Atlas Research, N., *Integrated genomic analyses of ovarian carcinoma*. Nature, 2011. **474**(7353): p. 609-15.
51. Konecny, G.E., et al., *Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer*. J Natl Cancer Inst, 2014. **106**(10).
52. Helleday, T., S. Eshtad, and S. Nik-Zainal, *Mechanisms underlying mutational signatures in human cancers*. Nat Rev Genet, 2014. **15**(9): p. 585-98.
53. Garraway, L.A. and E.S. Lander, *Lessons from the cancer genome*. Cell, 2013. **153**(1): p. 17-37.
54. Monti, P., et al., *Heterogeneity of TP53 Mutations and P53 Protein Residual Function in Cancer: Does It Matter?* Frontiers in Oncology, 2020. **10**.
55. Saleh, A. and R. Perets, *Mutated p53 in HGSC-From a Common Mutation to a Target for Therapy*. Cancers (Basel), 2021. **13**(14).
56. Ogawara, Y., et al., *Akt enhances Mdm2-mediated ubiquitination and degradation of p53*. J Biol Chem, 2002. **277**(24): p. 21843-50.
57. Harris, S.L. and A.J. Levine, *The p53 pathway: positive and negative feedback loops*. Oncogene, 2005. **24**(17): p. 2899-908.

58. Cole, A.J., et al., *Assessing mutant p53 in primary high-grade serous ovarian cancer using immunohistochemistry and massively parallel sequencing*. Sci Rep, 2016. **6**: p. 26191.
59. Tuna, M., et al., *Clinical relevance of TP53 hotspot mutations in high-grade serous ovarian cancers*. Br J Cancer, 2020. **122**(3): p. 405-412.
60. Hashimoto, T., et al., *p53 null mutations undetected by immunohistochemical staining predict a poor outcome with early-stage non-small cell lung carcinomas*. Cancer Res, 1999. **59**(21): p. 5572-7.
61. Kobel, M., et al., *Optimized p53 immunohistochemistry is an accurate predictor of TP53 mutation in ovarian carcinoma*. Journal of Pathology Clinical Research, 2016. **2**(4): p. 247-258.
62. Ahmed, A.A., et al., *Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary*. J Pathol, 2010. **221**(1): p. 49-56.
63. George, S.H. and P. Shaw, *BRCA and Early Events in the Development of Serous Ovarian Cancer*. Front Oncol, 2014. **4**: p. 5.
64. Xu, X., et al., *Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform-deficient cells*. Mol Cell, 1999. **3**(3): p. 389-95.
65. Chang, S., et al., *Tumor suppressor BRCA1 epigenetically controls oncogenic microRNA-155*. Nat Med, 2011. **17**(10): p. 1275-82.
66. Ruscito, I., et al., *BRCA1 gene promoter methylation status in high-grade serous ovarian cancer patients--a study of the tumour Bank ovarian cancer (TOC) and ovarian cancer diagnosis consortium (OVCAD)*. Eur J Cancer, 2014. **50**(12): p. 2090-8.
67. Candido-dos-Reis, F.J., et al., *Germline mutation in BRCA1 or BRCA2 and ten-year survival for women diagnosed with epithelial ovarian cancer*. Clin Cancer Res, 2015. **21**(3): p. 652-7.
68. Turner, N., A. Tutt, and A. Ashworth, *Hallmarks of 'BRCAness' in sporadic cancers*. Nat Rev Cancer, 2004. **4**(10): p. 814-9.
69. Telli, M.L., et al., *Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer*. Clin Cancer Res, 2016. **22**(15): p. 3764-73.
70. Li, L., et al., *Androgen receptor inhibitor-induced "BRCAness" and PARP inhibition are synthetically lethal for castration-resistant prostate cancer*. Sci Signal, 2017. **10**(480).
71. Patch, A.M., et al., *Whole-genome characterization of chemoresistant ovarian cancer*. Nature, 2015. **521**(7553): p. 489-94.
72. Wang, Y.K., et al., *Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes*. Nat Genet, 2017. **49**(6): p. 856-865.
73. Bronder, D., et al., *TP53 loss initiates chromosomal instability in fallopian tube epithelial cells*. Disease Models & Mechanisms, 2021. **14**(11): p. 1754-8403.
74. Di Fiore, R., et al., *RB1 in cancer: different mechanisms of RB1 inactivation and alterations of pRb pathway in tumorigenesis*. J Cell Physiol, 2013. **228**(8): p. 1676-87.
75. Luo, G., J. Kim, and K. Song, *The C-terminal domains of human neurofibromin and its budding yeast homologs Ira1 and Ira2 regulate the metaphase to anaphase transition*. Cell Cycle, 2014. **13**(17): p. 2780-9.
76. Xu, G.F., et al., *The catalytic domain of the neurofibromatosis type 1 gene product stimulates ras GTPase and complements ira mutants of S. cerevisiae*. Cell, 1990. **63**(4): p. 835-41.

77. Qiao, G.Y., et al., *Neurofibromin 1 expression is negatively correlated with malignancy and prognosis of epithelial ovarian cancer*. International Journal of Clinical and Experimental Pathology, 2019. **12**(5): p. 1702-1712.
78. Lee, Y.R., M. Chen, and P.P. Pandolfi, *The functions and regulation of the PTEN tumour suppressor: new modes and prospects*. Nat Rev Mol Cell Biol, 2018. **19**(9): p. 547-562.
79. Roh, M.H., et al., *High-grade fimbrial-ovarian carcinomas are unified by altered p53, PTEN and PAX2 expression*. Mod Pathol, 2010. **23**(10): p. 1316-24.
80. Kuhn, E., et al., *CCNE1 amplification and centrosome number abnormality in serous tubal intraepithelial carcinoma: further evidence supporting its role as a precursor of ovarian high-grade serous carcinoma*. Mod Pathol, 2016. **29**(10): p. 1254-61.
81. Cotton, A.M., et al., *Spread of X-chromosome inactivation into autosomal sequences: role for DNA elements, chromatin features and chromosomal domains*. Human Molecular Genetics, 2014. **23**(5): p. 1211-1223.
82. Amr, S.S., et al., *Allele-Specific Droplet Digital PCR Combined with a Next-Generation Sequencing-Based Algorithm for Diagnostic Copy Number Analysis in Genes with High Homology: Proof of Concept Using Stereocilin*. Clin Chem, 2018. **64**(4): p. 705-714.
83. Ciriello, G., et al., *Emerging landscape of oncogenic signatures across human cancers*. Nat Genet, 2013. **45**(10): p. 1127-33.
84. Huang, R.Y., et al., *Histotype-specific copy-number alterations in ovarian cancer*. BMC Med Genomics, 2012. **5**: p. 47.
85. Ramakrishna, M., et al., *Identification of candidate growth promoting genes in ovarian cancer through integrated copy number and expression analysis*. PLoS One, 2010. **5**(4): p. e9983.
86. Meyer, N. and L.Z. Penn, *Reflecting on 25 years with MYC*. Nat Rev Cancer, 2008. **8**(12): p. 976-90.
87. Nanjundan, M., et al., *Amplification of MDS1/EVI1 and EVI1, located in the 3q26.2 amplicon, is associated with favorable patient prognosis in ovarian cancer*. Cancer Res, 2007. **67**(7): p. 3074-84.
88. Au-Yeung, G., et al., *Selective Targeting of Cyclin E1-Amplified High-Grade Serous Ovarian Cancer by Cyclin-Dependent Kinase 2 and AKT Inhibition*. Clin Cancer Res, 2017. **23**(7): p. 1862-1874.
89. Shi, J., et al., *Highly frequent PIK3CA amplification is associated with poor prognosis in gastric cancer*. BMC Cancer, 2012. **12**.
90. Lyu, S., et al., *High expression of eIF4A2 is associated with a poor prognosis in esophageal squamous cell carcinoma*. Oncol Lett, 2020. **20**(5): p. 177.
91. Sangha, N., et al., *Neurofibromin 1 (NF1) defects are common in human ovarian serous carcinomas and co-occur with TP53 mutations*. Neoplasia, 2008. **10**(12): p. 1362-72, following 1372.
92. Zhu, Z., et al., *Whole-exome sequencing identifies prognostic mutational signatures in gastric cancer*. Ann Transl Med, 2020. **8**(22): p. 1484.
93. Goranova, T., et al., *Safety and utility of image-guided research biopsies in relapsed high-grade serous ovarian carcinoma-experience of the BriTROC consortium*. Br J Cancer, 2017. **116**(10): p. 1294-1301.

94. Popova, T., et al., *Ovarian Cancers Harboring Inactivating Mutations in CDK12 Display a Distinct Genomic Instability Pattern Characterized by Large Tandem Duplications*. *Cancer Res*, 2016. **76**(7): p. 1882-91.
95. Nik-Zainal, S., et al., *Landscape of somatic mutations in 560 breast cancer whole-genome sequences*. *Nature*, 2016. **534**(7605): p. 47-54.
96. Chien, J., et al., *TP53 mutations, tetraploidy and homologous recombination repair defects in early stage high-grade serous ovarian cancer*. *Nucleic Acids Res*, 2015. **43**(14): p. 6945-58.
97. Alexandrov, L.B. and M.R. Stratton, *Mutational signatures: the patterns of somatic mutations hidden in cancer genomes*. *Curr Opin Genet Dev*, 2014. **24**: p. 52-60.
98. Shih, I.M., Y. Wang, and T.L. Wang, *The Origin of Ovarian Cancer Species and Precancerous Landscape*. *Am J Pathol*, 2021. **191**(1): p. 26-39.
99. Lee, Y., et al., *A candidate precursor to serous carcinoma that originates in the distal fallopian tube*. *J Pathol*, 2007. **211**(1): p. 26-35.
100. Visvanathan, K., et al., *Fallopian Tube Lesions in Women at High Risk for Ovarian Cancer: A Multicenter Study*. *Cancer Prevention Research*, 2018. **11**(11): p. 697-705.
101. van der Steen, S., et al., *Changes in the Extracellular Matrix Are Associated With the Development of Serous Tubal Intraepithelial Carcinoma Into High-Grade Serous Carcinoma*. *Int J Gynecol Cancer*, 2017. **27**(6): p. 1072-1081.
102. Arora, S., B.R. Yelikar, and M.H. Karigoudar, *Evaluation of SEE-FIM (Sectioning and Extensively Examining the FIMbriated End) Protocol in Identifying Fallopian Tube Precursor Lesions in Women with Ovarian Tumors*. *J Obstet Gynaecol India*, 2019. **69**(Suppl 2): p. 153-159.
103. Kuhn, E., et al., *Ki-67 labeling index as an adjunct in the diagnosis of serous tubal intraepithelial carcinoma*. *Int J Gynecol Pathol*, 2012. **31**(5): p. 416-22.
104. Kito, M., et al., *Expression of Cell Competition Markers at the Interface between p53 Signature and Normal Epithelium in the Human Fallopian Tube*. *PLoS One*, 2016. **11**(6): p. e0156069.
105. Visvanathan, K., et al., *Diagnosis of serous tubal intraepithelial carcinoma based on morphologic and immunohistochemical features: a reproducibility study*. *Am J Surg Pathol*, 2011. **35**(12): p. 1766-75.
106. Ando, K., et al., *Discrimination of p53 immunohistochemistry-positive tumors by its staining pattern in gastric cancer*. *Cancer Med*, 2015. **4**(1): p. 75-83.
107. Vang, R., et al., *Validation of an algorithm for the diagnosis of serous tubal intraepithelial carcinoma*. *Int J Gynecol Pathol*, 2012. **31**(3): p. 243-53.
108. Nik, N.N., et al., *Origin and pathogenesis of pelvic (ovarian, tubal, and primary peritoneal) serous carcinoma*. *Annu Rev Pathol*, 2014. **9**: p. 27-45.
109. Bonnet, E., et al., *Performance comparison of three DNA extraction kits on human whole-exome data from formalin-fixed paraffin-embedded normal and tumor samples*. *PLoS One*, 2018. **13**(4): p. e0195471.
110. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. *Bioinformatics*, 2009. **25**(14): p. 1754-60.
111. Sandmann, S., et al., *Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data*. *Sci Rep*, 2017. **7**: p. 43169.
112. Benjamin, D., et al., *Calling Somatic SNVs and Indels with Mutect2*. *bioRxiv*, 2019: p. 861054.

113. Koboldt, D.C., et al., *VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing*. Genome Res, 2012. **22**(3): p. 568-76.
114. Saunders, C.T., et al., *Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs*. Bioinformatics, 2012. **28**(14): p. 1811-7.
115. Poplin, R., et al., *Scaling accurate genetic variant discovery to tens of thousands of samples*. bioRxiv, 2018: p. 201178.
116. McLaren, W., et al., *The Ensembl Variant Effect Predictor*. Genome Biol, 2016. **17**(1): p. 122.
117. Scheinin, I., et al., *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, 2014. **24**(12): p. 2022-32.
118. Poell, J.B., et al., *ACE: absolute copy number estimation from low-coverage whole-genome sequencing data*. Bioinformatics, 2019. **35**(16): p. 2847-2849.
119. Kristensen, K., et al., *TMB: Automatic Differentiation and Laplace Approximation*. Journal of Statistical Software; Vol 1, Issue 5 (2016), 2016.
120. Charrad, M., et al., *NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set*. Journal of Statistical Software, 2014. **61**(6): p. 1-36.
121. Therneau, T. *A Package for Survival Analysis in R*. . 2020; Available from: <https://cran.r-project.org/package=survival>.
122. González-Martín, A., et al., *Niraparib in Patients with Newly Diagnosed Advanced Ovarian Cancer*. N Engl J Med, 2019. **381**(25): p. 2391-402.
123. Ray-Coquard, I., et al., *Olaparib plus Bevacizumab as First-Line Maintenance in Ovarian Cancer*. N Engl J Med, 2019. **381**(25): p. 2416-2428.
124. TCGA, *Integrated genomic analyses of ovarian carcinoma*. Nature, 2011. **474**(7353): p. 609-615.
125. Forshew, T., et al., *Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA*. Sci Transl Med, 2012. **4**(136): p. 136ra68.
126. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
127. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples*. Nat Biotechnol, 2013. **31**(3): p. 213-9.
128. Van der Auwera, G.A., et al., *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline*. Curr Protoc Bioinformatics, 2013. **43**: p. 11 10 1-11 10 33.
129. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data*. Bioinformatics, 2014. **30**(15): p. 2114-20.
130. Mafficini, A., et al., *BRCA somatic and germline mutation detection in paraffin embedded ovarian cancers by next-generation sequencing*. Oncotarget, 2016. **7**(2): p. 1076-83.
131. Baugh, E.H., et al., *Why are there hotspot mutations in the TP53 gene in human cancers?* Cell Death and Differentiation, 2018. **25**(1): p. 154-160.
132. Hainaut, P. and G.P. Pfeifer, *Somatic TP53 Mutations in the Era of Genome Sequencing*. Cold Spring Harb Perspect Med, 2016. **6**(11).
133. Muller, P.A. and K.H. Vousden, *Mutant p53 in cancer: new functions and therapeutic opportunities*. Cancer Cell, 2014. **25**(3): p. 304-17.
134. Andrews, S.; Available from: <https://tinyurl.com/y4gacv2u>.

135. Mao, Y.H., et al., *Low tumor purity is associated with poor prognosis, heavy mutation burden, and intense immune phenotype in colon cancer*. Cancer Management and Research, 2018. **10**: p. 3569-3577.
136. Etemadmoghadam, D., et al., *Amplicon-dependent CCNE1 expression is critical for clonogenic survival after cisplatin treatment and is correlated with 20q11 gain in ovarian cancer*. PLoS One, 2010. **5**(11): p. e15498.
137. Cully, M., et al., *Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis*. Nature Reviews Cancer, 2006. **6**(3): p. 184-192.
138. Martins, F.C., et al., *Combined image and genomic analysis of high-grade serous ovarian cancer reveals PTEN loss as a common driver event and prognostic classifier*. Genome Biol, 2014. **15**(12): p. 526.
139. Rust, K., et al., *Routine germline BRCA1 and BRCA2 testing in patients with ovarian carcinoma: analysis of the Scottish real-life experience*. BJOG, 2018. **125**(11): p. 1451-1458.
140. Kobel, M., et al., *Ovarian carcinoma subtypes are different diseases: implications for biomarker studies*. PLoS Med, 2008. **5**(12): p. e232.
141. Tang, Y.C. and A. Amon, *Gene copy-number alterations: a cost-benefit analysis*. Cell, 2013. **152**(3): p. 394-405.
142. Newcomb, R., et al., *Context-dependent effects of whole-genome duplication during mammary tumor recurrence*. Sci Rep, 2021. **11**(1): p. 14932.
143. Bielski, C.M., et al., *Genome doubling shapes the evolution and prognosis of advanced cancers*. Nat Genet, 2018. **50**(8): p. 1189-1195.
144. Storchova, Z. and D. Pellman, *From polyploidy to aneuploidy, genome instability and cancer*. Nat Rev Mol Cell Biol, 2004. **5**(1): p. 45-54.
145. Dewhurst, S.M., et al., *Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution*. Cancer Discov, 2014. **4**(2): p. 175-185.
146. Lopez, S., et al., *Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution*. Nature Genetics, 2020. **52**(3): p. 283-+.
147. Morden, C.R., et al., *Chromosome instability is prevalent and dynamic in high-grade serous ovarian cancer patient samples*. Gynecologic Oncology, 2021. **161**(3): p. 769-778.
148. Lemaitre, C., et al., *Precise detection of rearrangement breakpoints in mammalian chromosomes*. BMC Bioinformatics, 2008. **9**: p. 286.
149. Korbel, J.O. and P.J. Campbell, *Criteria for inference of chromothripsis in cancer genomes*. Cell, 2013. **152**(6): p. 1226-36.
150. Nadeau, J.H. and B.A. Taylor, *Lengths of chromosomal segments conserved since divergence of man and mouse*. Proc Natl Acad Sci U S A, 1984. **81**(3): p. 814-8.
151. Venkatraman, E.S. and A.B. Olshen, *A faster circular binary segmentation algorithm for the analysis of array CGH data*. Bioinformatics, 2007. **23**(6): p. 657-63.
152. Park, B.J. and D. Lord, *Application of finite mixture models for vehicle crash data analysis*. Accid Anal Prev, 2009. **41**(4): p. 683-91.
153. Yu, J.W.Z.W.C.Y.N.W.X., *Optimization of the number of components in the mixed model using multi-criteria decision-making*. Applied Mathematical Modelling, 2012. **36**(9): p. 4227-4240.
154. Leisch, F., *FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R*. Journal of Statistical Software, 2004. **11**(8): p. 1 - 18.

155. Lin, X. and P.C. Boutros, *Optimization and expansion of non-negative matrix factorization*. BMC Bioinformatics, 2020. **21**(1): p. 7.
156. Hubschmann, D., et al., *Analysis of mutational signatures with yet another package for signature analysis*. Genes Chromosomes & Cancer, 2021. **60**(5): p. 314-331.
157. Gardiner, J.C., Z. Luo, and L.A. Roman, *Fixed effects, random effects and GEE: What are the differences?* Statistics in medicine, 2009. **28**(2): p. 221-239.
158. Lee, S., *Detecting Differential Item Functioning Using the Logistic Regression Procedure in Small Samples*. Appl Psychol Meas, 2017. **41**(1): p. 30-43.
159. Egozcue, J.J., et al., *Isometric logratio transformations for compositional data analysis*. Mathematical Geology, 2003. **35**(3): p. 279-300.
160. Liu, X.C., et al., *A knowledge-driven way to interpret the isometric log-ratio transformation and mixture distributions of geochemical data*. Journal of Geochemical Exploration, 2020. **210**.
161. Au, N.H., et al., *Evaluation of immunohistochemical markers in non-small cell lung cancer by unsupervised hierarchical clustering analysis: a tissue microarray study of 284 cases and 18 markers*. J Pathol, 2004. **204**(1): p. 101-9.
162. Rulli, E., et al., *Assessment of proportional hazard assumption in aggregate data: a systematic review on statistical methodology in clinical trials using time-to-event endpoint*. Br J Cancer, 2018. **119**(12): p. 1456-1463.
163. Dentre, S.C., et al., *Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes*. Cell, 2021. **184**(8): p. 2239-2254 e39.
164. Walton, J.B., et al., *CRISPR/Cas9-derived models of ovarian high grade serous carcinoma targeting Brca1, Pten and Nf1, and correlation with platinum sensitivity*. Sci Rep, 2017. **7**(1): p. 16827.
165. Trope, C., et al., *Randomized study on adjuvant chemotherapy in stage I high-risk ovarian cancer with evaluation of DNA-ploidy as prognostic instrument*. Ann Oncol, 2000. **11**(3): p. 281-8.
166. Nelson, L., et al., *A living biobank of ovarian cancer ex vivo models reveals profound mitotic heterogeneity*. Nat Commun, 2020. **11**(1): p. 822.
167. Kurman, R.J. and M. Shih Ie, *The Dualistic Model of Ovarian Carcinogenesis: Revisited, Revised, and Expanded*. Am J Pathol, 2016. **186**(4): p. 733-47.
168. Lengyel, E., *Ovarian cancer development and metastasis*. Am J Pathol, 2010. **177**(3): p. 1053-64.
169. Medeiros, F., et al., *The tubal fimbria is a preferred site for early adenocarcinoma in women with familial ovarian cancer syndrome*. American Journal of Surgical Pathology, 2006. **30**(2): p. 230-236.
170. Mehra, K., et al., *STICS, SCOUTs and p53 signatures; a new language for pelvic serous carcinogenesis*. Front Biosci (Elite Ed), 2011. **3**(2): p. 625-34.
171. Ducie, J., et al., *Molecular analysis of high-grade serous ovarian carcinoma with and without associated serous tubal intra-epithelial carcinoma*. Nature Communications, 2017. **8**.
172. Eckert, M.A., et al., *Genomics of Ovarian Cancer Progression Reveals Diverse Metastatic Trajectories Including Intraepithelial Metastasis to the Fallopian Tube*. Cancer Discov, 2016. **6**(12): p. 1342-1351.
173. Vang, R., M. Shih Ie, and R.J. Kurman, *Fallopian tube precursors of ovarian low- and high-grade serous neoplasms*. Histopathology, 2013. **62**(1): p. 44-58.

174. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads*. EMBnet.journal, 2011. **17**(1): p. 2226-6089.
175. Van Roy, N., et al., *Shallow Whole Genome Sequencing on Circulating Cell-Free DNA Allows Reliable Noninvasive Copy-Number Profiling in Neuroblastoma Patients*. Clin Cancer Res, 2017. **23**(20): p. 6305-6314.
176. Raman, L., et al., *WisecondorX: improved copy number detection for routine shallow whole-genome sequencing*. Nucleic Acids Res, 2019. **47**(4): p. 1605-1614.
177. Bocanegra, M., et al., *Focal amplification and oncogene dependency of GAB2 in breast cancer*. Oncogene, 2010. **29**(5): p. 774-9.
178. Raab, M., et al., *Boosting the apoptotic response of high-grade serous ovarian cancers with CCNE1 amplification to paclitaxel in vitro by targeting APC/C and the pro-survival protein MCL-1*. Int J Cancer, 2020. **146**(4): p. 1086-1098.
179. Dunn, P., et al., *Next Generation Sequencing Methods for Diagnosis of Epilepsy Syndromes*. Front Genet, 2018. **9**: p. 20.
180. Kito, M., et al., *Expression of Cell Competition Markers at the Interface between p53 Signature and Normal Epithelium in the Human Fallopian Tube (vol 11, e0156069, 2016)*. Plos One, 2016. **11**(7).
181. Kim, M.P. and G. Lozano, *Mutant p53 partners in crime*. Cell Death Differ, 2018. **25**(1): p. 161-168.
182. Chien, J., et al., *Genetic Evidence for Early Peritoneal Spreading in Pelvic High-Grade Serous Cancer*. Front Oncol, 2018. **8**: p. 58.
183. Davis, A., R. Gao, and N. Navin, *Tumor evolution: Linear, branching, neutral or punctuated?* Biochim Biophys Acta Rev Cancer, 2017. **1867**(2): p. 151-161.
184. Kimura, M., *Rare variant alleles in the light of the neutral theory*. Mol Biol Evol, 1983. **1**(1): p. 84-93.
185. Gould, S.J. and N. Eldredge, *Punctuated equilibrium comes of age*. Nature, 1993. **366**(6452): p. 223-7.
186. Baca, S.C., et al., *Punctuated evolution of prostate cancer genomes*. Cell, 2013. **153**(3): p. 666-77.
187. Nicos, M. and P. Krawczyk, *Genetic Clonality as the Hallmark Driving Evolution of Non-Small Cell Lung Cancer*. Cancers, 2022. **14**(7): p. 1813.
188. Gao, R., et al., *Punctuated copy number evolution and clonal stasis in triple-negative breast cancer*. Nat Genet, 2016. **48**(10): p. 1119-30.
189. Wu, N.Y., et al., *Natural history of ovarian high-grade serous carcinoma from time effects of ovulation inhibition and progesterone clearance of p53-defective lesions*. Mod Pathol, 2020. **33**(1): p. 29-37.
190. Bast, R.C., Jr., et al., *Biomarkers and Strategies for Early Detection of Ovarian Cancer*. Cancer Epidemiol Biomarkers Prev, 2020. **29**(12): p. 2504-2512.
191. Menon, U., et al., *Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial*. Lancet, 2021. **397**(10290): p. 2182-2193.
192. Staicu, C.E., et al., *Role of microRNAs as Clinical Cancer Biomarkers for Ovarian Cancer: A Short Overview*. Cells, 2020. **9**(1).
193. Asante, D.B., et al., *Liquid biopsy in ovarian cancer using circulating tumor DNA and cells: Ready for prime time?* Cancer Letters, 2019. **468**: p. 59-71.
194. Fiala, C. and E.P. Diamandis, *Utility of circulating tumor DNA in cancer diagnostics with emphasis on early detection*. BMC Med, 2018. **16**(1): p. 166.

195. Douville, C., et al., *Assessing aneuploidy with repetitive element sequencing*. Proc Natl Acad Sci U S A, 2020. **117**(9): p. 4858-4863.
196. Li, Z., et al., *USP9X controls translation efficiency via deubiquitination of eukaryotic translation initiation factor 4A1*. Nucleic Acids Res, 2018. **46**(2): p. 823-839.
197. Zhu, J., et al., *Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth*. Nature, 2015. **525**(7568): p. 206-211.
198. Olive, K.P., et al., *Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome*. Cell, 2004. **119**(6): p. 847-60.
199. Baghban, R., et al., *Tumor microenvironment complexity and therapeutic implications at a glance*. Cell Commun Signal, 2020. **18**(1): p. 59.
200. Micek, H.M., et al., *The Many Microenvironments of Ovarian Cancer*. Adv Exp Med Biol, 2020. **1296**: p. 199-213.
201. Hu, Z., et al., *The Repertoire of Serous Ovarian Cancer Non-genetic Heterogeneity Revealed by Single-Cell Sequencing of Normal Fallopian Tube Epithelial Cells*. Cancer Cell, 2020. **37**(2): p. 226-242 e7.
202. Martincorena, I., et al., *High burden and pervasive positive selection of somatic mutations in normal human skin*. Science, 2015. **348**(6237): p. 880-886.
203. Kuhn, E., et al., *The diagnostic and biological implications of laminin expression in serous tubal intraepithelial carcinoma*. Am J Surg Pathol, 2012. **36**(12): p. 1826-34.
204. Chen, Z.L. and S. Strickland, *Laminin gamma 1 is critical for Schwann cell differentiation, axon myelination, and regeneration in the peripheral nerve*. Journal of Cell Biology, 2003. **163**(4): p. 889-899.
205. Evrony, G.D., A.G. Hinch, and C. Luo, *Applications of Single-Cell DNA Sequencing*. Annu Rev Genomics Hum Genet, 2021. **22**: p. 171-197.
206. Navin, N., et al., *Tumour evolution inferred by single-cell sequencing*. Nature, 2011. **472**(7341): p. 90-U119.
207. Martelotto, L.G., et al., *Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples*. Nat Med, 2017. **23**(3): p. 376-385.
208. Liu, Y., et al., *Spatial transcriptome sequencing of FFPE tissues at cellular level*. 2020.

CHAPTER 8.Appendix

Appendix 1. Clinical information of early stage cohort

Patient_ID	Sample_ID	Diagnosis_date	Stage	Diag_age	OS (months)	Event	Months_since_diagnosis	Surgery
IRITROC-046	IM_092	21/04/2008	2a	47.2	122.4	0	157	TAH, BSO, omentectomy, R pelvic LN
IRITROC-187	IM_277	10/06/2013	1a	74.4	46.6	1	96	TAH, BSO, omentectomy
IRITROC-260	IM_426	02/02/2012	2a	67.6	77.3	1	112	TAH, BSO, omentectomy, pelvic LN, para-aortic LN, para-colic gutter, peritoneal biopsies
S_001	IMPCO-001	15/03/2017	1a	63.9	10.1	0	51	TAH, BSO, omentectomy, pelvic LN, paracolic gutter, diaphragm
S_003	IMPCO-005	17/12/2015	1c	59.3	NA	NA	66	TAH, BSO, omentectomy
S_007	IMPCO-015	03/03/2009	1c	84.0	109.3	0	147	TAH, BSO, omentectomy
S_009	IMPCO-018	04/12/2012	1c	70.0	69.3	0	102	BSO, omentectomy
S_010	IMPCO-020	21/09/2010	1c	51.0	89.4	0	128	TAH, BSO, omentectomy
S_011	IMPCO-022	09/03/2010	1a	60.9	NA	1	135	TAH, BSO, omentectomy
S_012	IMPCO-024	24/07/2007	1c	60.4	125.9	0	166	BSO, omentectomy, previous TAH
S_013	IMPCO-026	03/11/2011	1c	81.6	84.1	0	115	TAH, BSO, omentectomy
S_014	IMPCO-028	02/08/2011	1b	67.9	87.1	0	118	TAH, BSO, omentectomy
S_015	IMPCO-030	14/12/2011	1c	76.6	85.3	0	114	TAH, BSO, omentectomy, para-aortic LN
S_017	IMPCO-060	15/07/2011	1c	42.6	86.0	0	119	TAH, BSO, omentectomy, pelvic LN, para-aortic LN
S_018	IMPCO-061	12/09/2006	1c	71.3	118.1	0	177	TAH, BSO, omentectomy
S_019	IMPCO-063	17/11/2010	1c	84.6	NA	1	127	TAH, BSO, omentectomy
S_020	IMPCO-065	27/07/2010	1c	48.3	93.5	0	130	TAH, BSO, omentectomy, para-aortic LN
S_021	IMPCO-067	16/11/2011	1c	47.4	81.0	0	115	TAH, BSO, omentectomy
S_022	IMPCO-071	01/12/2015	1c	49.8	NA	1	66	TAH, BSO, omentectomy
S_023	IMPCO-072	08/12/2015	1c	49.5	NA	0	66	TAH, BSO, omentectomy
S_024	IMPCO-073	22/04/2016	1c	66.9	NA	0	61	TAH, BSO, omentectomy, appendicectomy
S_027	IMPCO-074	23/09/2016	2a	58.9	52.9	0	56	BSO, omentectomy, pelvic LN, prior TAH
S_028	IMPCO-076	26/05/2017	2a	40.3	43.2	0	48	TAH, BSO, omentectomy, R iliac LN
S_029	IMPCO-077	15/11/2017	1b	57.4	37.8	0	43	TAH, BSO, omentectomy
S_030	IMPCO-078	27/11/2017	1c	42.9	NA	NA	42	TAH, BSO, omentectomy, para-aortic LN, pelvic peritoneum
S_031	IMPCO-079	29/11/2016	2a	61.4	31.5	0	54	TAH, BSO, omentectomy, pelvic LN
S_032	IMPCO-080	29/09/2015	1b	42.9	17.8	0	68	LSO, omentectomy
S_033	IMPCO-083	22/08/2014	2a	69.6	NA	NA	81	TAH, BSO, omentectomy
S_034	IMPCO-084	22/12/2016	2a	61.3	15.0	0	53	TAH, BSO, omentectomy
S_035	IMPCO-085	23/03/2016	1c	67.7	14.6	0	62	LSO, omentectomy
S_036	IMPCO-086	06/06/2019	2a	52.3	27.0	0	24	TAH, BSO, omentectomy, para-aortic LN, falciform ligament
S_037	IMPCO-087	07/03/2017	2a	67	42.7	0	51	TAH, BSO, omentectomy
S_038	IMPCO-088	16/05/2017	2a	82	56.6	0	49	TAH, BSO, omentectomy
S_039	IMPCO-089	27/05/2015	1c	48	28.2	0	72	TAH, BSO, omentectomy, appendicectomy
S_040	IMPCO-091	06/02/2014	2a	60	40.0	0	88	TAH, BSO, omentectomy, pelvic peritonectomy
S_041	IMPCO-093	21/09/2016	1c	73.9	0.9	0	56	BSO
S_042	IMPCO-094	08/02/2018	1c	81.2	31.7	0	40	TAH, BSO, omentectomy
S_043	IMPCO-096	21/11/2018	2a	59.2	17.5	0	30	TAH, BSO, omentectomy, pelvic LN
S_044	IMPCO-098	29/08/2014	1c	55.7	6.6	0	81	TAH, BSO, omentectomy, common iliac LN
S_047	IMPCO-003	23/04/2015	2a	81	27.6	1	73	TAH, BSO, omentectomy, common iliac LN
S_048	IMPCO-104	09/03/2018	1a	73.6	12.1	0	39	BSO, omentectomy
S_049	IMPCO-106	15/08/2016	1c	78.0	32.9	0	58	TAH, BSO, omental biopsy, R pelvic LN
S_050	IMPCO-107	23/03/2018	2a	53.0	8.0	0	38	BSO, omentectomy

Appendix 2. Clinical information of late stage cohort

patient_ID	Sample_ID	Diagnosis_date	Stage	Diag_age	OS (months)
BRITROC-002	IM_005	11/03/2012	3c	73.0	27.2
BRITROC-003	IM_007	06/06/2011	3c	71.0	27.4
BRITROC-005	IM_018	24/08/2009	3c	49.0	60.3
BRITROC-009	IM_023	21/06/2010	3c	47.0	74.6
BRITROC-011	IM_025	22/12/2010	3c	74.0	73.1
BRITROC-013	IM_033	28/01/2008	3c	69.0	80.6
BRITROC-023	IM_056	27/02/2009	3c	42.3	111.6
BRITROC-030	IM_068	05/11/2012	4	50.0	25.0
BRITROC-034	IM_323	16/11/2012	3c	75.0	23.6
BRITROC-037	IM_081	18/04/2012	3c	50.0	44.5
BRITROC-061	IM_108	15/01/2010	3c	70.0	101.9
BRITROC-065	IM_113	19/06/2011	3c	70.0	83.6
BRITROC-067	IM_115	16/07/2008	3c	62.3	91.3
BRITROC-068	IM_116	12/06/2012	3c	51.0	48.7
BRITROC-069	IM_117	07/08/2012	3c	76.0	38.2
BRITROC-074	IM_124	28/12/2007	3c	64.0	126.5
BRITROC-092	IM_143	01/07/2013	3c	65.0	30.3
BRITROC-099	IM_149	21/01/2013	4	63.0	65.5
BRITROC-101	IM_202	01/07/2013	3c	68.0	24.8
BRITROC-103	IM_282	22/03/2011	4	76.0	64.0
BRITROC-105	IM_211	21/07/2008	3c	61.8	91.9
BRITROC-106	IM_326	07/04/2011	3c	66.0	80.6
BRITROC-107	IM_153	29/01/2013	3c	53.0	34.7
BRITROC-113	IM_155	21/02/2011	4	58.0	65.5
BRITROC-114	IM_186	20/11/2012	3c	65.0	40.2
BRITROC-115	IM_180	24/03/2011	3c	65.4	88.2
BRITROC-124	IM_214	19/12/2012	3c	63.0	66.7
BRITROC-135	IM_193	30/04/2014	3c	65.0	17.0
BRITROC-139	IM_196	01/03/2013	3c	66.0	47.0
BRITROC-141	IM_169	04/04/2012	3c	45.0	43.9
BRITROC-146	IM_255	08/05/2013	3c	57.0	48.8
BRITROC-147	IM_198	08/06/2014	4	74.0	21.8
BRITROC-156	IM_220	05/09/2010	3c	65.1	70.0
BRITROC-161	IM_352	30/06/2014	3c	49.8	31.1
BRITROC-170	IM_243	09/06/2014	3c	51.6	48.9
BRITROC-172	IM_269	09/07/2013	3c	51.7	61.2
BRITROC-185	IM_357	30/06/2006	3c	44.4	143.5
BRITROC-192	IM_360	13/12/2013	3c	45.8	55.0
BRITROC-194	IM_361	06/10/2014	4	71.6	45.2
BRITROC-199	IM_319	28/08/2012	3c	57.4	70.7
BRITROC-206	IM_298	03/06/2015	3c	64.0	23.0
BRITROC-209	IM_295	05/08/2014	3c	40.2	47.4
BRITROC-218	IM_332	11/09/2014	4	34.0	34.3
BRITROC-224	IM_348	02/12/2015	3c	68.1	31.0
BRITROC-234	IM_399	18/05/2016	4	62.3	10.5
BRITROC-241	IM_423	28/12/2014	4	50.8	41.8
BRITROC-248	IM_403	20/04/2015	4	54.0	27.4

BRITROC-250	IM_372	09/10/2014	3c	73.5	40.2
BRITROC-252	IM_388	10/09/2014	3c	40.4	32.8
BRITROC-255	IM_385	22/07/2015	3c	58.2	24.9
BRITROC-256	IM_381	04/07/2013	3c	59.1	51.5
BRITROC-268	IM_411	03/08/2015	3c	58.0	31.8

Appendix.3 SNV mutations in early stage by samples by Mutect2, HaplotypeCaller, Strelka and Varscan2 pipelines.

s_Name	Patient_ID	Location	Gene	REF	ALT	Mutation_Type	Protein_position	Amino_acids	Mutect2.Alt.Freq	HaplotypeCaller.Alt.Freq	Strelka.Alt.Freq	VarScan2.Freq	Clinical_significance
1	br55	13:32944692	BRCA2	C	T	stop_gained,splice_region_variant	2829	Q/*	0.136	NA	0.4123	NA	pathogenic
1	br55	17:7577071	TP53	GGA	G	frameshift_variant	289	L/X	NA	0.5341	0.5341	NA	NA
7	br187	17:7577541	TP53	T	TTC A	inframe_insertion	p.M246dup		0.263	NA	NA	NA	NA
6	br260	17:7577129	TP53	A	G	missense_variant	270	F/S	0.457	0.472	0.5042	NA	pathogenic,likely_pathogenic
6	br260	10:89692800	PTEN	C	T	missense_variant	95	P/L	NA	NA	0.2397	NA	pathogenic
6	br260	10:89653807	PTEN	G	A	missense_variant	35	M/I	NA	NA	0.2157	NA	Uncertain significance
6	br46	17:7577121	TP53	G	T	missense_variant	273	R/S	0.58	0.584	0.9515	NA	uncertain_significance,likely_pathogenic,pathogenic
	br46	17:56780641	RAD51C	T	C	missense_variant	219	L/S	0.354	0.4065	NA	NA	Conflicting_interpretations_of_pathogenicity
↳0001	ES_0001	17:7578397	TP53	T	G	missense_variant	178	H/P	0.787	0.790660225	0.775	0.7862	NA
↳0006	ES_0003	17:7578416	TP53	C	A	missense_variant	172	V/F	0.55	0.533575318	0.558	0.5582	likely_pathogenic
↳0015	ES_0007	17:7578521	TP53	G	C	missense_variant	137	L/V	0.444	0.547457627	0.517	0.5175	NA
↳0015	ES_0007	17:7578442	TP53	T	C	missense_variant	163	Y/C	0.194	0.283478261	0.519	0.3427	likely_pathogenic,pathogenic
↳0018	ES_0009	17:7577539	TP53	G	A	missense_variant	248	R/W	0.823	0.865384615	0.867	0.8638	pathogenic,likely_pathogenic
↳0020	ES_0010	17:7577120	TP53	C	A	missense_variant	273	R/L	0.697	0.713554987	0.717	0.7184	pathogenic,likely_pathogenic
↳0022	ES_0011	17:7578416	TP53	C	A	missense_variant	172	V/F	0.456	0.449526814	0.476	0.4771	likely_pathogenic
↳0025	ES_0012	17:7577114	TP53	C	T	missense_variant	275	C/Y	0.319	0.320920785	0.616	0.6112	likely_pathogenic
↳0026	ES_0013	17:7577124	TP53	C	A	missense_variant	272	V/L	0.436	0.413533835	0.424	0.4331	pathogenic,likely_pathogenic
↳0028	ES_0014	17:7577121	TP53	G	A	missense_variant	273	R/C	0.491	0.527836505	0.545	0.5416	not_provided,uncertain_significance,pathogenic,pathogenic/likely_pathogenic,likely_patho
↳0030	ES_0015	WT	TP53						NA	NA	NA	NA	
↳0060	ES_0017	17:7578428	TP53	GC	G	frameshift_variant	167	Q/X	0.398	NA	0.393	NA	pathogenic
↳0062	ES_0018	17:7578190	TP53	T	C	missense_variant	220	Y/C	0.337	0.381294964	0.345	0.3417	likely_pathogenic,pathogenic
↳0064	ES_0019	17:7577559	TP53	G	C	missense_variant	241	S/C	0.869	0.899470899	0.901	NA	likely_pathogenic
↳0065	ES_0020	17:7578502	TP53	A	G	missense_variant	143	V/A	0.755	0.696226415	0.698	0.6982	NA
↳0067	ES_0021	17:7578190	TP53	T	C	missense_variant	220	Y/C	0.471	0.456582633	0.475	0.4783	likely_pathogenic,pathogenic
↳0071	ES_0022	17:41256153	BRCA1	C	A	stop_gained	143	E/*	0.905	0.903508772	0.9	0.9108	pathogenic
↳0071	ES_0022	17:7578552	TP53	G	C	stop_gained,splice_region_variant	126	Y/*	0.761	NA	0.755	0.4783	likely_pathogenic
↳0072	ES_0023	17:7577559	TP53	G	C	missense_variant	241	S/C	0.47	0.529610829	0.51	0.5076	likely_pathogenic
↳0073	ES_0024	17:7578406	TP53	C	T	missense_variant	175	R/H	NA	0.760171306	0.776	0.7791	pathogenic,likely_pathogenic,not_provided
↳0076	ES_0028	17:41276047	BRCA1	C	CT	frameshift_variant	22-23	-/X	0.855	0.856589147	0.862	NA	pathogenic
↳0076	ES_0028	17:7577121	TP53	G	A	missense_variant	273	R/C	0.714	0.715634218	0.717	0.715634218	not_provided,uncertain_significance,pathogenic,pathogenic/likely_pathogenic,likely_patho
↳0077	ES_0029	17:41226539	BRCA1	C	T	splice_acceptor_variant	-	-	0.39	0.297560976	0.795	0.7946	pathogenic,likely_pathogenic
↳0077	ES_0029	17:7577121	TP53	G	A	missense_variant	273	R/C	0.665	0.633303809	0.614	0.6171	not_provided,uncertain_significance,pathogenic,pathogenic/likely_pathogenic,likely_patho
↳0078	ES_0030	17:7577539	TP53	G	A	missense_variant	248	R/W	0.592	0.575141243	0.574	0.5875	pathogenic,likely_pathogenic
↳0079	ES_0031	WT	TP53						NA	NA	NA	NA	
↳0080	ES_0032	17:7578406	TP53	C	T	missense_variant	175	R/H	0.629	0.640926641	0.645	0.6485	pathogenic,likely_pathogenic,not_provided
↳0080	ES_0032	17:41243482	BRCA1	G	A	stop_gained	1060	Q/*	0.761	0.79175705	NA	0.7884	pathogenic
↳0081	ES_0032	17:7578406	TP53	C	T	missense_variant	175	R/H	0.743	0.730050934	0.753	0.7522	pathogenic,likely_pathogenic,not_provided
↳0081	ES_0032	17:41243482	BRCA1	G	A	stop_gained	1060	Q/*	0.892	0.889848812	NA	0.8599	pathogenic
↳0083	ES_0033	17:7578212	TP53	G	A	stop_gained	213	R/*	0.73	0.74695122	0.713	0.7141	pathogenic
↳0084	ES_0034	17:7577538	TP53	C	T	missense_variant	248	R/Q	0.736	NA	0.716	NA	pathogenic,likely_pathogenic,pathogenic/likely_pathogenic
↳0085	ES_0035	WT	TP53						NA	NA	NA	NA	
↳0085	ES_0035	17:29508478	NF1	C	T	missense_variant			NA	0.797918474	NA	0.8329	pathogenic
↳0086	ES_0036	17:7577539	TP53	G	A	missense_variant	248	R/W	0.712	0.638865721	0.643	NA	likely_pathogenic,pathogenic
↳0087	ES_0037	17:7578190	TP53	T	C	missense_variant	220	Y/C	0.801	0.741477273	0.799	0.8001	likely_pathogenic,pathogenic
↳0088	ES_0038	17:7576896	TP53	TG	T	frameshift_variant	317	Q/X	0.43	NA	0.423	NA	pathogenic
↳0091	ES_0040	17:7578526	TP53	C	A	missense_variant	135	C/F	0.85	0.88179669	0.847	0.8151	likely_pathogenic
↳0092	ES_0040	17:7578526	TP53	C	A	missense_variant	135	C/F	0.883	0.874806801	0.858	0.8585	likely_pathogenic
↳0094	ES_0042	17:7578221	TP53	TTC	T	frameshift_variant	209	R/X	0.535	NA	0.555	NA	pathogenic
↳0095	ES_0042	17:7578221	TP53	TTC	T	frameshift_variant	209	R/X	0.376	NA	0.339	NA	pathogenic
↳0103	ES_0047	17:7577569	TP53	A	G	missense_variant	238	C/R	0.685	0.689282203	0.692	0.6912	likely_pathogenic,pathogenic
↳0106	ES_0049	17:7578406	TP53	C	T	missense_variant	175	R/H	0.311	NA	0.274	0.2686	pathogenic,likely_pathogenic,not_provided
↳0107	ES_0050	WT	TP53						NA	NA	NA	NA	

Appendix 4. SNV mutations in late stage by samples by Mutect2, HaplotypeCaller,

Strelka pipelines.

Sample_ID	Patient_ID	Location	Gene	REF	ALT	Mutation_Type	Protein_position	Amino_acids	Mutect2.Alt.Freq.GATK.4.0	Strelka.Alt.Freq	HaplotypeCaller_Freq	CLNSIG
IM_202	br101	17:7578552	TP53	G	T	stop_gained,splice_region_variant	126	Y/*	0.608	0.6239	NA	NA
IM_282	br103	17:7578461	TP53	C	A	missense_variant	157	V/F	0.603	0.5656		0.701 Conflicting_interpretations_of_pathogenicity
IM_282	br103	2:215633966	BRN1	C	T	stop_gained	462	W/*	0.4115	NA	NA	Pathogenic
IM_211	br105	17:7579414	TP53	C	T	stop_gained	91	W/*	0.212	NA	NA	0.78 Pathogenic
IM_326	br106	17:7578222	TP53	TTC	T	frameshift_variant	209	R/X	NA	0.6275		0.766 NA
IM_153	br107	17:41246432	BRCA1	C	T	stop_gained	372	W/*	0.579	NA	NA	Pathogenic
IM_153	br107	17:7577106	TP53	G	C	missense_variant	278	P/A	0.997	0.4374		0.8305 Conflicting_interpretations_of_pathogenicity
IM_25	br11	17:7578265	TP53	A	G	missense_variant	195	I/T	0.276	0.4479		0.5815 Pathogenic/Likely_pathogenic
IM_155	br113	17:7577120	TP53	C	T	missense_variant	273	R/H	0.578	0.5763		0.7195 Pathogenic
IM_186	br114	17:7578508	TP53	C	T	missense_variant	141	C/Y	0.75	0.766		0.716 Pathogenic
IM_187	br114	17:7578508	TP53	C	T	missense_variant	141	C/Y	0.75	0.766		0.762 Pathogenic
IM_214	br124	17:7578413	TP53	C	T	missense_variant	173	V/M	0.808	0.2958		0.81 Pathogenic
IM_28	br13	17:7579414	TP53	C	T	stop_gained	91	W/*	0.37	NA		0.694 Pathogenic
IM_29	br13	17:7579414	TP53	C	T	stop_gained	91	W/*	0.37	NA		0.671 Pathogenic
IM_30	br13	17:7579414	TP53	C	T	stop_gained	91	W/*	0.37	NA		0.745 Pathogenic
IM_31	br13	17:7579414	TP53	C	T	stop_gained	91	W/*	0.37	NA		0.778 Pathogenic
IM_32	br13	17:7579414	TP53	C	T	stop_gained	91	W/*	0.37	NA		0.799 Pathogenic
IM_33	br13	17:7579414	TP53	C	T	stop_gained	91	W/*	0.37	NA		0.6695 Pathogenic
IM_34	br13	17:7579414	TP53	C	T	stop_gained	91	W/*	0.37	NA		0.635 Pathogenic
IM_193	br135	17:7577575	TP53	A	G	missense_variant	236	Y/H	0.525	0.517		0.6735 Conflicting_interpretations_of_pathogenicity
IM_194	br135	17:7577575	TP53	A	G	missense_variant	236	Y/H	0.525	0.517		0.6735 Conflicting_interpretations_of_pathogenicity
IM_196	br139	17:7577550	TP53	C	A	missense_variant	244	G/V	NA	0.6931		0.6205 Likely_pathogenic
IM_167	br141	17:7577124	TP53	C	T	missense_variant	272	V/M	0.685	0.7244		0.809 Uncertain_significance
IM_168	br141	17:7577124	TP53	C	T	missense_variant	272	V/M	0.685	0.7244		0.819 Uncertain_significance
IM_169	br141	17:7577124	TP53	C	T	missense_variant	272	V/M	0.685	0.7244		0.759 Uncertain_significance
IM_170	br141	17:7577124	TP53	C	T	missense_variant	272	V/M	0.685	0.7244		0.759 Uncertain_significance
IM_171	br141	17:7577124	TP53	C	T	missense_variant	272	V/M	0.685	0.7244		0.759 Uncertain_significance
IM_231	br141	17:7577124	TP53	C	T	missense_variant	272	V/M	0.685	0.7244		0.759 Uncertain_significance
IM_255	br146	17:7578442	TP53	T	C	missense_variant	163	Y/C	0.794	0.7707		0.7335 Pathogenic
IM_256	br146	17:7578442	TP53	T	C	missense_variant	163	Y/C	0.794	0.7707		0.7985 Pathogenic
IM_257	br146	17:7578442	TP53	T	C	missense_variant	163	Y/C	0.794	0.7707		0.7565 Pathogenic
IM_258	br146	17:7578442	TP53	T	C	missense_variant	163	Y/C	0.794	0.7707		0.7695 Pathogenic
IM_259	br146	17:7578442	TP53	T	C	missense_variant	163	Y/C	0.794	0.7707		0.885 Pathogenic
IM_197	br147	17:7579733	TP53	G	T	frameshift	c.314dupG	p.S106Qfs*43	1	NA		0.1095 NA
IM_198	br147	17:7579733	TP53	G	T	frameshift	c.314dupG	p.S106Qfs*43	1	NA		0.413 NA
IM_199	br147	17:7579733	TP53	G	T	frameshift	c.314dupG	p.S106Qfs*43	1	NA		0.5445 NA
IM_232	br147	17:7579733	TP53	G	T	frameshift	c.314dupG	p.S106Qfs*43	1	NA		0.1125 NA
IM_220	br156	17:41215926	BRCA1	C	T	missense_variant	523	G/E	0.227	NA	NA	Pathogenic
IM_220	br156	17:7577096	TP53	T	A	missense_variant	281	D/V	0.712	0.3225		0.5405 Pathogenic/Likely_pathogenic
IM_352	br161	17:7577580	TP53	T	C	missense_variant	234	Y/C	0.502	0.4886		0.6955 Pathogenic/Likely_pathogenic
IM_354	br161	17:7577580	TP53	T	C	missense_variant	234	Y/C	0.502	0.4886		0.787 Pathogenic/Likely_pathogenic
IM_355	br161	17:7577580	TP53	T	C	missense_variant	234	Y/C	0.502	0.4886		0.787 Pathogenic/Likely_pathogenic
IM_243	br170	17:7578535	TP53	T	C	missense_variant	132	K/R	NA	NA		0.278 Uncertain_significance
IM_357	br170	17:7578535	TP53	T	C	missense_variant	132	K/R	NA	NA		0.278 Uncertain_significance
IM_245	br170	17:7578535	TP53	T	C	missense_variant	132	K/R	NA	NA		0.278 Uncertain_significance
IM_238	br172	17:41245354	BRCA1	C	A	stop_gained	732	E/*	NA	NA		0.837 Pathogenic
IM_238	br172	17:7577141	TP53	C	A	missense_variant	266	G/V	NA	0.4932		0.5195 Uncertain_significance
IM_239	br172	17:41245354	BRCA1	C	A	stop_gained	732	E/*	NA	NA		0.837 Pathogenic
IM_239	br172	17:7577141	TP53	C	A	missense_variant	266	G/V	NA	0.4932		0.5195 Uncertain_significance
IM_240	br172	17:41245354	BRCA1	C	A	stop_gained	732	E/*	NA	NA		0.837 Pathogenic
IM_240	br172	17:7577141	TP53	C	A	missense_variant	266	G/V	NA	0.4932		0.5195 Uncertain_significance
IM_269	br172	17:41245354	BRCA1	C	A	stop_gained	732	E/*	NA	NA		0.837 Pathogenic
IM_269	br172	17:7577141	TP53	C	A	missense_variant	266	G/V	NA	0.4932		0.5195 Uncertain_significance
IM_357	br185	17:7577551	TP53	C	T	missense_variant	244	G/S	0.597	0.6487		0.7755 Pathogenic
IM_358	br185	17:7577551	TP53	C	T	missense_variant	244	G/S	0.597	0.6487		0.6475 Pathogenic
IM_360	br192	17:41242963	BRCA1	G	A	stop_gained,splice_region_variant	1099	Q/*	NA	0.2017	NA	Pathogenic
IM_360	br192	17:7578190	TP53	T	C	missense_variant	220	Y/C	NA	0.4788		0.6115 Pathogenic
IM_361	br194	17:7574000	TP53	C	A	stop_gained	343	E/*	0.572	0.4444		0.6135 NA
IM_319	br199	17:7577570	TP53	C	A	missense_variant	237	M/I	0.731	0.4163		0.774 Likely_pathogenic
IM_5	br2	17:7577120	TP53	C	T	missense_variant	273	R/H	0.992	0.9892		0.9755 Pathogenic
IM_6	br2	17:7577120	TP53	C	T	missense_variant	273	R/H	0.992	0.9892		0.952 Pathogenic
IM_298	br206	17:7576927	TP53	C	T	splice_acceptor_variant	-	-	0.585	0.6067		0.5345 Pathogenic/Likely_pathogenic
IM_295	br209	17:7576208	TP53	AT	A	frameshift_variant	214	H/X	NA	0.4408		0.372 NA
IM_332	br218	17:7574002	TP53	C	T	missense_variant	342	R/R	0.539	0.6877		0.5375 Pathogenic/Likely_pathogenic
IM_348	br224	17:7577082	TP53	C	T	missense_variant	286	E/K	0.651	0.4994		0.6875 Pathogenic
IM_56	br23	17:7577561	TP53	A	C	missense_variant	240	S/R	0.346	0.3525		0.4265 Uncertain_significance
IM_399	br234	17:7578413	TP53	C	A	missense_variant	173	V/L	0.66	0.5836		0.4675 Pathogenic
IM_400	br234	17:7578413	TP53	C	A	missense_variant	173	V/L	0.66	0.5836		0.771 Pathogenic
IM_401	br234	17:7578413	TP53	C	A	missense_variant	173	V/L	0.66	0.5836		0.741 Pathogenic
IM_402	br234	17:7578413	TP53	C	A	missense_variant	173	V/L	0.66	0.5836		0.72 Pathogenic
IM_423	br241	17:7577581	TP53	A	T	missense_variant	234	Y/N	0.575	0.5992		0.549 Conflicting_interpretations_of_pathogenicity
IM_403	br248	17:7578555	TP53	C	G	splice_acceptor_variant	-	-	0.172	NA		0.6255 NA
IM_372	br250	17:7577124	TP53	C	T	missense_variant	272	V/M	0.281	0.3254		0.575 Uncertain_significance
IM_388	br252	17:7578190	TP53	T	C	missense_variant	220	Y/C	0.545	0.5545		0.5425 Pathogenic
IM_389	br252	17:7578190	TP53	T	C	missense_variant	220	Y/C	0.545	0.5555		0.4665 Pathogenic
IM_385	br255	17:7577538	TP53	C	T	missense_variant	248	R/Q	0.921	0.9217		0.737 Pathogenic
IM_386	br255	17:7577538	TP53	C	T	missense_variant	248	R/Q	0.921	0.9217		0.737 Pathogenic
IM_380	br256	17:7578478	TP53	G	T	missense_variant	151	P/H	0.203	0.1289		0.3 Uncertain_significance
IM_381	br256	17:7578478	TP53	G	T	missense_variant	151	P/H	0.203	0.1289		0.3 Uncertain_significance
IM_411	br268	17:7578416	TP53	C	A	missense_variant	172	V/F	0.451	0.3106		0.528 Likely_pathogenic
IM_412	br268	17:7578416	TP53	C	A	missense_variant	172	V/F	0.451	0.3106		0.503 Likely_pathogenic
IM_10	br3	17:7578535	TP53	T	A	missense_variant	132	K/M	0.335	NA		0.313 Uncertain_significance
IM_11	br3	17:7578535	TP53	T	A	missense_variant	132	K/M	0.335	NA		0.314 Uncertain_significance
IM_12	br3	17:7578535	TP53	T	A	missense_variant	132	K/M	0.335	NA		0.171 Uncertain_significance
IM_13	br3	17:7578535	TP53	T	A	missense_variant	132	K/M	0.335	NA		0.1585 Uncertain_significance
IM_7	br3	17:7578535	TP53	T	A	missense_variant	132	K/M	0.335	NA		0.327 Uncertain_significance
IM_8	br3	17:7578535	TP53	T	A	missense_variant	132	K/M	0.335	NA		0.361 Uncertain_significance
IM_9	br3	17:7578535	TP53	T	A	missense_variant	132	K/M	0.335	NA		0.3535 Uncertain_significance
IM_68	br30	17:7576852	TP53	C	A	splice_donor_variant	-	-	0.884	0.8692		0.9165 Pathogenic
IM_323	br34	17:7578536	TP53	T	C	missense_variant	132	K/E	0.768	NA		0.676 Conflicting_interpretations_of_pathogenicity
IM_81	br37	17:7577529	TP53	A	T	missense_variant	251	I/N	0.213	NA		0.6505 Uncertain_significance
IM_81	br37	17:41226546	BRCA1	G	A	intron_variant	-	-	0.392	0.3538	NA	Conflicting_interpretations_of_pathogenicity
IM_18	br5	17:7577094	TP53	G	A	missense_variant	282	R/W	0.859	0.8611		0.853 Pathogenic/Likely_pathogenic
IM_19	br5	17:7577094	TP53	G	A	missense_variant	282	R/W	0.859	0.8611		0.7785 Pathogenic/Likely_pathogenic
IM_108	br61	17:7578394	TP53	T	C	missense_variant	179	H/R	0.84	NA		0.6435 Conflicting_interpretations_of_pathogenicity
IM_113	br65	17:7574003	TP53	G	A	stop_gained	342	R/*	NA	NA		0.541 Pathogenic
IM_115	br67	12:25398222	KRAS	C	A	missense_variant	33	D/Y	NA	NA		0.095 NA
IM_115	br67	16:23619180	PALB2	C	T	intron_variant	-	-	0.916	NA	NA	Conflicting_interpretations_of_pathogenicity
IM_115	br67	17:7578190	TP53	T	C	missense_variant	220	Y/C	0.071	NA		0.4985 Pathogenic
IM_115	br67	10:89717623	PTEN	TG	T	frameshift_variant	216	V/X	0.3424	NA	NA	NA
IM_116	br68	17:7578190	TP53	T	C	missense_variant	220	Y/C	NA	NA		0.6475 Pathogenic
IM_117	br69	17:7577539	TP53	G	A	missense_variant	248	R/W	0.541	0.5873		0.5885 Pathogenic
IM_117	br69	3:178952039-178952040	PIK3CA	GA	TT	missense_variant	1032	E/L	0.223	NA	NA	NA
IM_118	br69	17:7577539	TP53	G	A	missense_variant	248	R/W	0.541	0.5873		0.709 Pathogenic
IM_118	br69	3:178952039-178952040	PIK3CA	GA	TT	missense_variant	1032	E/L	0.223	NA	NA	NA
IM_119	br69	17:7577539	TP53	G	A	missense_variant	248	R/W	0.541	0.5873		0.5305 Pathogenic
IM_119	br69	3:178952039-178952040	PIK3CA	GA	TT	missense_variant	1032	E/L				

Appendix 5. Amplicons for Ampliseq sequencing for Early stage cohort

Target ID	ULSO Sequence	DLSO Sequence
AMPL529821	TCCTTCGAGCTGGTTTATAGACTGT	TTGTGAGAGCCCAGTCATTGTACTG
AMPL529822	CTGCACACAGTATATCATCTATGAA	TGAGTTGCTTCCTCTTGACAGCTGA
AMPL529823	GGGGGCCAGATCCTCAGTAGAAAGC	TTATCACCCAGAGAGGGTTCCGGCAG
AMPL529824	TCTCATTATTTTCAGTTGCCAAAGCT	AGTGACGTGACTCAGACCATCAATA
AMPL529825	TTCAGGTAAGGGATATTCAACTTGT	TGATGGATGCTACTTCTATTGTGG
AMPL529832	CAGGAAGAAAAGTATGAAATTCCTG	TTTCAACAGCTACAGGTGACTGATT
AMPL529833	GCTCACTTGATACTTAGTTTGCTTT	AGAAGCAGGCTCAACAGAGAACAGC
AMPL529826	TGCAGTTCAAAGTACCTTGAAAGTGT	ATGCCTGTGATTAAGGAAC TTGATG
AMPL571014	TATCTTTCTTTCTCTTACTGCTTTT	TCTCAATGGATGCTGGATTCTAAAA
AMPL529803	CAGTGGGCTGTCTTCAGAACACAG	AGATGATCCTTTTGAGAACTACCT
AMPL623349	AATGTTTTTAACACTGGTATCTCCT	TGAGCTTGCAGTAATTCTTAAGGCT
AMPL529816	ACCAGAGAAGAATGAATCATCCTCA	CCAAACAATTATTGAGCACTTCCTG
AMPL623351	GACTTACTTAAACACATCTGCACTT	CTCAGTAGTAAGTATGGATTTAGCT
AMPL529835	CCACTTCACGATGCAGCCAAGAATG	TCAAGGTGTGTGATAAAGCAGACTT
AMPL529836	GTAGTGGAATTATTGCTCCAGCATA	ATAGTCAAGCTGTTACTTTTCTATG
AMPL623353	TTTCTTATCACTTCTTTCACTTCTT	AACACCACGGGTATCAAAATGACT
AMPL529801	CCAAATGTTAAAGACCATGCTGGAT	TCTGCCACATATACTCTTTTGTAGTC
AMPL529802	TTTTCTTTCTTTCTTAAATGCTTT	GTAGTTGTCTGGTTTTTATTCTCAT
AMPL529805	ATAGAGGAGAGACTTTGCTCCATAT	CAGACATGATACTTCTCTGAACCA
AMPL529806	CCTTCTACATTAAGTAGTTCAAGTT	AGGTAGGATGCTTACTCTGAAATAC
AMPL529807	ACCATTGCCTGAATGTTCTTCACCA	ATGTCTAGTCCCTCAGCAATGAAGC
AMPL529808	TCACAATAGACTTTCCAGTCCCAT	ACGTAAAGTTGGTGGTACATCAGGG
AMPL529809	GTCTCCAGACACTAAGAGCAGGAAT	TAGAACCAGCATCTGAGCACCAGT
AMPL529810	CAGAATCTGAATGTTTTGGAAGTTT	TCCTGAGAAGGTCTGCAAAAATTAT
AMPL529811	GAGGAATCTAAGCAAAAGCTGGTAT	CTTTACCATTGGCTGAGCAATAGA
AMPL529812	AAGGCTTCTGCAAGATCTGAAAAA	CAACCATCTGTTATCTCCAGTCTC
AMPL529813	GTGCAAACCCAGCCTGCAATAAAAA	AACTTTAGCTGAAATCAACCAAAA
AMPL529814	TTAATGATGCAGGAAACAAGAAGAA	TCAGCAAGACTCATATGAATTTGTT
AMPL623350	AGCAATAAATTTTCATGTAACAGATT	TGTGGTTTAGCCCTCGAAGTAAGAA
AMPL529828	GTAAGCTTCGAAATTTGCTACATGA	TGGAGTTCTGGAATACGTATTTAG
AMPL529829	GGATACAAGACTTGAAGATAAATAG	CAGGTAAGAACTATCCCTATCTCTC
AMPL529831	GTTTATCTAACAATGACTCAAATCT	ACGGTTTGATGTATATAGTACAAC
AMPL529819	CCCGCCATGGAACCGGATGGTCGCG	GAATCTTTTCAAATCTTCCGTTTCC
AMPL529842	GCCGGATAATCGGCAGCCGAGGAAC	GTAAGACGGAGCTTCTTGGGGGTG
AMPL529843	GCAGCTTCCCTGTGGTTTCCCGAGG	TCCGGGAACGAGCCTCGTTCCGCGC
AMPL32540	ACTTACCTGTGACTCCATAGAAAAAT	TGTCTCTAGCTAGTCTGTTACTCTG
AMPL32541	AGATCAGCCAAATTCAGTTATTTTT	TCAGTGATTTTCAGAGAGAGATCTC
AMPL1893424	TGCATGCTGTTAATTGTGTGGAAG	TCAAAAGATTGTAGTTCTGGCATT
AMPL1893425	TGGATTGTGCAATTCCTATGCAATC	TTTTGTTGTCCAGCCACCATGATGT

AMPL193785	CTCTCCGGCGTTCCCAGCCCTGGAA	CACGGCTTCCACCTTCCCTTTCAGG
AMPL230050	GGCGCACGGGAGGTTTAAACCGGC	ACAGGCGACTCAGACCCCCCTCG
AMPL535652	GGAGAAGACGAATAATCCTCCGAAC	GTGTTGGAGGCAGTAGAAGGGGAGA
AMPL193788	GGCGGGGCTTCTTCTGCAGGATGGA	TCCTCAGCAGCCAGAGGCCTGGCAG
AMPL156298	CTTTTGTCTCTGCTAACGATCTCTT	ACTTGGCGGTAGCTGATGCCCTCG
AMPL202862	GTGACAGAAAGGTAAAGAGGAGCAG	GTGATGTCTGGGAGCCTGTGGCTGA
AMPL3181	ACTACATCATCAATATTGTTCTCTGT	GGAAAGAAATATCTGAAATATGCAG
AMPL3182	AAATCTTTTCTAAATGAAAACACAA	GCAGGAAATCCCATAGCAATAATGT
AMPL2576	TCTAACAAGCAGATAAATTCTACTT	ACAAAAAGCCACCATTAAACAAAAT
AMPL3269	TACATTTCATACCTACCTCTGCAATT	AACATTGCCTGAATCTTTATAATGT
AMPL3270	GATAATCTGGATGACTCATTATTGT	TGTCATAATGTCTTTCAGCACAACT
AMPL3191	GTTATGGTCTTCAAAAGGATATTGT	CACTGGATTCCAGAGAAAAATAATG
AMPL3192	CATTACACCAGTTCGTCCCTTTCCA	CAACTGTGGTAAAAAGATAACCTCA
AMPL3193	GCCTCTTGTGCCTTTAAAAATTTGC	CATGATTGTCATCTTCACTTAGCCA
AMPL3238	ATAATCCAGATGATTCTTTAACAGG	TGGTAACTGGGTCGTAGCCATTTAA
AMPL3239	AATTTCAGCACTTACTGCAAGTTC	TGACTGGGAATAGTTACTCCCTGGT
AMPL96778	GGTAACGGCTGAGGGAACCAAGT	CCTTTAACTGTCAAAGTGTAAAAA
AMPL219645	ATTTTGGATATTTCTCCCAATGAAA	CATGAAGTTGTCTTCCCGTCGTGTG
AMPL156290	AGGTTTCCTCTGGTCTGGTATGAA	CCTATGTTAAACATTTGCAAAAAGT
AMPL156291	TGCAAATGCTATCGATTTCTTGATC	CCCCAAAAGTGAAACATTTTGTCTTA
AMPL156293	ATTAAACACACATCACATACATACA	AGTACTAGATATTCCTTGTATTAT
AMPL219759	CAGAATATCTATAATGATCAGGTTT	AATGAAAACCCACAAATATGACTCA
AMPL219760	TTTATGGTGTTTTATCCCTCTTGAT	CATCTGGTGTACAGAAGTTGAACT
AMPL456704	AATATGATCCAACAATAGAGGTAAA	CTGAAATGTACCTTGGGTTTCAAGT
AMPL456705	GGCTGCTGAAATGACTGAATATA	TATGCATATTACTGGTGCAGGACCA
AMPL3237	CTGGTGGAGTATTTGATAGTGATT	CAAGAGTGCCTTGACGATACAGCTA
AMPL1104	CAGCGTGTCTTAAAAATTTCAAAAA	TGTGACTGAGGTAAAACCTCAGAAGC
AMPL1105	TACTGGGTTTTTTAGCAAGCATTTTT	TCTTTGGATCCAATAGGCATTTTTTA
AMPL993	TTTCAAGTCTTCTGCAGGTTTCAAG	TTGTGACCCAGAACAGTTAAAGATC
AMPL994	CAGAGTCAGCCCTTGCTCTTTGAAT	AGCTTCTGAAGAAAGTTCTTCAAAC
AMPL995	CCTAAATTCCTAGTTTGTAGTTCTC	TGAAGCCAGCTGATTATAAGATGGT
AMPL1024	CTGTGCGAAGACTTTTATGTCTACT	TGGGAATGAGAATGTATAGGGAGTT
AMPL1025	ATCATACCTTTCTACTAAGACAAGAA	TCCTTCCTGAAACAGTACAATAATT
AMPL1026	TCTTAGCCAAAATATTAGCATAAAA	ACAGGAAACATCATCTGCTTGATCC
AMPL1015	AAACAACTCCACATACCACTGGG	TGTAGAACAACAGGACTAAAAATAA
AMPL1100	CTTCACAACTTTGGTGTATGAAAC	TTTGTGGTGTACATGTGTACATTG
AMPL1011	CAGGATCCACCTCAGCTCCTAGACT	TTGCCTGTATGAGGCAGATGCTAG
AMPL555246	TCTGCTCTTTCTTGTAATACACAT	ATGTTTTGGTGTCTGACGACCCTGG
AMPL1020	TACTTACAGCAGTAGTATCATGAGG	GCACATCTACTATTTGATTACATGA
AMPL1021	CTCAAAGGCTTAGATAAATTACAGA	AAATACAGTTTCAGATGCTTCTTCA
AMPL156253	CAGCTTCTCTTTGATTTGTGTTTTT	CTCTCAATGCACATATAGTAGTAGT
AMPL156254	AAACCGAGATCACGGGTGACAGAGC	CAGAAGCGATAAATCTATCATTTTTT
AMPL981	TACAACTGTTTCATATACTTCATCT	AAAACAGAAGCACATTAATATTTTA

AMPL982	CATCAGCGTTTGCTTCATGGAAAAT	TGGCATTGACTTTCCAATGTGGTCT
AMPL983	CCACTTCAGATACAAATGAGTATTT	GTCTTGCTAGTTCTTACTTTTGTA
AMPL984	GACCATTACAGGCCAAAGACGGTA	GGTTTTTAGATTTTTCACATTCATC
AMPL985	CTGTGTCTAATAGGTCTTTTCTGA	GAGATTTTGTCACTTCCACTCTCAA
AMPL986	ACTGTTTCCTCATTTAATGGCTTCT	TGGTCACATGAAGAAATATGCAATA
AMPL987	CTTTACTGCAAGAATGCAGTCTGTA	AGGCTAGAAATACGTGGCAAAGAAT
AMPL988	GAAAGTCTCTTTAGGTGATCTCTT	ATGCTGCTCTTCATCTCTCTATTT
AMPL989	CCTCCTTCTGTGAGCAAACAGTATG	AGACTTTTTGATACCTGAAATGAA
AMPL990	GTGGATATTAAACCTGCATTCTTCA	CACTTTCAGAGGCTTCAGTTCTTT
AMPL991	TTTGCTTCAAACCTGGGCTGAACAGT	ACAGAATTCTGTGTGGTGGTGGCTG
AMPL992	CACAGAAGGAATCGTCATCTATAAA	GTTCTGATTTTGGTCTTTCGGTAT
AMPL1032	CAAAGTTGGTTCTTCAGAATCATTC	AAAGGTAGTGTTTGGGCACAGTGGC
AMPL1033	TCTTTATAATCAAGATCCTGAGAGA	TCTTTTCACAGAAGAATGCAATAAA
AMPL1034	GGATCATTTTCACACTGTCTTCCT	CATTTCTAGAACATTTCTCAGAAT
AMPL1035	GATTGAAAGTCAGTATCACTGTATT	TGACAGAGAATCAGCTTCTGGGGTA
AMPL1036	CTAGGTTTGACAGAACATCCTTGGA	TGTTGTACTGGGTGACATGCTGCAG
AMPL1037	CATCTTGATTCTTTTCCATGGGAAT	AGGAGTTAAATAAGAGTGTGGCA
AMPL1038	AATTGTACCTTTCTTGAAGGTGATG	TTCATAATTGTTACCTTTGAGCTTG
AMPL1039	GTCTGGATTGACAGTTATTTTGA	TATTTTTCAGGTGGCAACAGCTCAA
AMPL1041	GAGTTCTTGAAAATGGGTTCTGTTTA	TCTGAGAAAAGTTCTTCAGAGTCTG
AMPL1042	TGCAAGAACATAAACCAAATCTTTT	TCTGTTTCATGAAGTTCCTTAGTAT
AMPL1043	TCCGATTTTAAATCTTGACCTAGAG	TTGGGTTGCTTGTTTATCACCTGTG
AMPL1044	TCTGAAGCTACCTCCAAAACGTGTA	TGCTGCTTTACACTATTTTGTCT
AMPL1045	TTCAACACAAGCTAAACTAGTAGGA	ACCTAAGAGTCCTGCCCATTTGTTT
AMPL1046	TACTCTGTAAATGTGCAGATACAGT	TTGCTCTTCTTAATGTTATGTTTCT
AMPL1047	CTTTTGGCTAGGTGTTAAATTATGG	TTGCTCAGTTTCTTTTGATTATCTA
AMPL1048	AAATTCAAACCTGACTTCCTGATTCT	ATCCTGCTTGAAAAATAACATCTGA
AMPL1050	GTTTAATTTCAACTGTACCTTCAAA	TCAAATGTACTCTTCTGCAATATGT
AMPL1051	AGTTTTGTGCCATGAGCAGAATAAA	TGTCTACCTGACCAATCGATGGGGC
AMPL1052	ATAAACTTATTGGATGTACCTCTGC	CCCCTTCATTTTCATCTGTAAAT
AMPL1053	GCCAGTAGTCATTTCAATATTATTT	ACCTCTGCAGAAGTTTCTCACTAA
AMPL1055	TCATTTTACTTGAATCACTGCCAT	AACAAAAGTGCCAGTAGTCATTTCA
AMPL1056	GATAATTTAAGACATATGTTGTGCT	ATGAGAATTTCTACTGGCAGCAGTA
AMPL1057	CTTGAGCTTTCGCAACTTCCAAAAA	AGCAAGTCCGTTTCATCTTTATGAA
AMPL305	GCTCCGTTTTAGTAGCAGTTAACTG	CAAATCTTCTTTAATCTGAGTGT
AMPL1058	ATTTTCCCCTTGACAGTCTGAAAA	AAGTATTACCATGACATGCTTCTTG
AMPL1060	TTTCTCATAACTTAGAATGTCCAT	AATGACTCTTGGCGACACTAATAT
AMPL1061	TGGGACACTTCTTTCAGTATTTTG	AAGGAAAAGTTATGCAATCTTCTG
AMPL1062	CTAGTACCTTGCTCTTTTTCATCAA	TGGAAGGTCAGTAGTTGATTTCCAG
AMPL1063	ATGGTCTCACATGCTAATTCAGGT	ACTTGTCCAAAGATTCCTTTGCAA
AMPL1064	AATTATCACTTAAGAGCTTAGGTGG	TCTCTGTACTTTAGGGTCTTTGCCC
AMPL1065	CTTTTCTACATTTTCATGTACTTT	ACAGTCTCAATAGAAACAAGGTTTT
AMPL1067	TCAAATATTCCTTCTCTAAGCCATT	GTAACAAGTGCAGGACTTTTGTCT

AMPL1068	GATGATTTTTGTCATTTTCAGCTAT	AGTAATGAAGTCTGACTCACAGAAG
AMPL1069	AGCTGTTAGACATGCTACTGTTACT	AATCTGCAGTATTTATTCTTTCTGG
AMPL1071	GCATTTGCATCTTTTACATTGGATA	ATATACCTCATCAGAATGGTAGGAA
AMPL1072	ATTTTTCAGGGTGAAGAGCTAGTC	TTTTGATCTTCAACATTCTTCAATA
AMPL1073	CATGTGAAACACAAACGATTTTACC	AACGCAAATATCTTCATTTACAGTT
AMPL1074	TCATCCAATGCCTCGTAACAACCTG	GCTATCCTAAATGCAGGTGGCCCTA
AMPL330	GTCAGCAAAAACCTTATGTGAATGC	TCGTTTGGCAAATTTTGTATTATT
AMPL331	TCCAGACATATTTTGGTTATGTTGT	ATGAAGAATATCCTCTGAATCATCC
AMPL1076	GATACCTGGACAGATTTTCCACTTG	TGAAGTTTCCAACTAACATCACAA
AMPL1077	TGAGCTGGTCTGAATGTTCTGTTACT	ATCCCAACAAGTATTGCAGATGAGA
AMPL1078	TTGCTGTACTAAATCCAGAGAAAGC	CTTTGGAAGAACTTGCTTGGTACT
AMPL1079	GTGAAGACTATGCTCAGTTCTGATT	CCACATTATATGAAAAGCCTTTTTG
AMPL1080	TCCATTTCTGAGTTTACACAGTGCT	CTCTAACACTCCCTTAACTTTGTGT
AMPL1081	GTTGAAATTGAGAGAGATATGGAGA	GGTTTCTCTTATCAACACGAGGAAG
AMPL1082	AAAACATGAATGTTCTCAACAAGTG	AGTGATTATTTTCTGAAGAACCACC
AMPL1084	AGTTCATCATCTTCCATAAAAGCTT	CATTTTACGTTTTTAGGTGAAGCC
AMPL1085	TTTCGGGACATGTAAAAGAGAATG	ACTGCTTCTGTTTCAAAGTAGTTTT
AMPL156246	CTTTTGATTGAGGGTTCTCCTAAAA	ACTTACAACAAATAGTGACCTAATG
AMPL89070	TATAGAGGGAGAACAGATATAAATA	ATTGAGGGTTCTCCTAAAAAGAAAT
AMPL1027	GACTTTTCTCATACTGTATTAGAAT	ATCTTTTATTGTGCCTAGGAAACAA
AMPL1087	GGTGCGGTAAAATTTGGATTCTGTA	TGCTACATGGTGATACTATTCTGGT
AMPL1088	GCTTGAAGATTTTCCAAAGTCAGA	TCCTTAGTTGTGCTGCAATGGGGAG
AMPL1089	TGTAGTAATCAAGTGCTCATTTTTT	ATGAGATTTAGACAGAAATCTTGA
AMPL1090	TTTTGCTTTTGTCTGTTTTCTCCCA	AGCAGAACTTGATAAAATGGATGT
AMPL1091	TTCACACTTTGTGAAAGTTACAGCT	ATTCCTAACACACTGTTCAACTCTG
AMPL1092	AACCATCAGGACATTATTTAACAAC	GCTTGATTGGAGTTGTTTTGTAA
AMPL979	GCTGCTTTCAGAGAGATTCGAGGCA	TGGCATTCTGAAGACTTGTAATTAA
AMPL173	CATAAAAGCCATCAGTATTGTAGAC	TTTGCAAGATACAGACTGCCTGGCT
AMPL1102	GGAGGGTATGAGCCATCCACCATCA	AGCTACACAAAATAAACACACACAA
AMPL1103	AGGGAATACATAAAAGTTAACACAC	TCCTTTTCCAGTCCATAAACTTTCC
AMPL1093	TTCCATATGATCCATCTATAGTGAT	ACCACATAGGATGATACTGAATTCA
AMPL1094	TTTCTGGGCTTAGGCATCTATTAGC	ATTCTAGAAATAAGCTTTGGATCCA
AMPL1095	CTGCCGTATATGATTACGTAATGTA	GAAAGGCACATTCCATAGCTGCCAG
AMPL1003	CTCAATGAAATTATGTGAGAAACAC	AGGAAGTGTGACTCTAGAATTCCAC
AMPL1004	TACCACCCATCTGTAAGTTCAATAA	GTTTTTGCAGCTGTGTCATCCCTTT
AMPL1005	TCCATGAAGAATAATCTTCTGACCA	TGGGTATCTGCACTACTAGTTTTAT
AMPL555208	TGACTGATTTTTACCAAGAGTGCAA	GCCATTCTTTAAGACAGCTAAGAGG
AMPL191427	AAAGAATCCAAGTTTGGTATACCAG	AGAGTTTTCATTCTCTAGAACTGC
AMPL191428	CCTGTATAGGGTATGCTCTTTGAAT	AGTACTGTTAGCAGAAATCTGGACA
AMPL191429	GGGCAAGAGACCGAACTCCATCTC	CACAACCAACATTTCTCCATCACT
AMPL1096	TGGGCCTCCACATATTTTGTGCTT	GTAAATTGTATCAGGCCAGGCACAG
AMPL1097	TATTAGAAATAACAATGTGTACCAT	TCTCTTTCATTGCGAAATATGTATA
AMPL1016	CTTCATAAAGCTCTGCACCATCTTG	GAACCAAGCATAAAACACCCAAAGA

AMPL352520	TTGAATAATCATCAAGCCTCATTAT	CTTGCTGTCTTGTTAGTGCACGTGA
AMPL1028	CATTTCCTGTGATTATTCAAGGCT	AGAGCTCATCTTAAGGGTGTGGTTA
AMPL1029	TTCCTTTTGTTCAGCAGATTCCATG	TTCCTGAAATAACCCTGTGACCAT
AMPL1030	ACATACTTACCTGAATCTTTTCTT	CTGAGCTTGTTTCTTATCATTTCAAC
AMPL1031	TTGCTTCTCTGATATAAACTAACTT	CTTACAATACGCAACTTCCACACGG
AMPL1002	ACTGTATGTTAGCTCTTTCAGATTT	TGAGAAAGATGCAATGGAAGAAGTG
AMPL200	TACCGGTAGTTGTTGATACTGAGTT	TCTGTATCTCTTTCCTTCTGTTAAC
AMPL1018	AAATCCTATTAGGTCCACCTCAGAA	TGGAGATTCCATAAACTAACAAGCA
AMPL1019	AGCTCCAACTAATCATAAGAGATTT	ATGGCTGAAAGTCTGGATCTAAAAA
AMPL1008	AATAATGTCTTCATTAAGGTCTATC	AGATGAATTTTAAGATGCAAGAAAG
AMPL1009	GCAGAAAACACAGAAAAATCTCCAG	TGCCAGTAAATTGTAACATTCGTCT
AMPL554964	CCAAAATGTGTGGTGATGCTGAAAA	AGAAGGCCTGATTTGGATTCTGGTC
AMPL1022	GGCCTGAAGTACAGTCTTTAGTTG	TGCAAACCCACATATTTATGTTTCC
AMPL1023	AACAGAATATACGATGGCCTCCATA	TTGGGATCATTTGCATGCAGTATAT
AMPL996	GGCCATACAAAGTGATAAAGGACTT	CGCAAATATTACACACAGTCTCCCT
AMPL230	TTACAAGACTTTGAAGTCATCTGGG	CTCACAATTAGGAGAAGACATCTGA
AMPL997	AAATGGGACTAACAGGTGGAGGTAA	GTGGAAACAGACTTCTTTTGGCCA
AMPL156239	CATCTGAGGAGAATTCAGTTCTTTT	TCAAGAAATCCAAGGCTCTTCTCTT
AMPL156240	TGTTTTTCTCCTGTGTAACCAGACA	TGTTTCGTATTTGGTGCCACAACCTC
AMPL156241	GGCCTGGGAACCTCCTGTCTTTTG	GGGTATTTTCAATGCAAGTTCTTC
AMPL156242	TAAACTGGAAAGGTTAAGCGTCAAT	AGTACAACGTCGTTTCAGTCTGAGA
AMPL193816	CCTGCTCTGGGTCCTCCTCAGGAGG	AGCGAGCTGTGGAGGAGCGCCGGGG
AMPL193817	CGACGGCGGCTCTGCTCGCTCACCT	CATGACGCCTTCCGCGGCGGGAGC
AMPL193818	CTGTCACCATTCTGCAGACGCTCCG	TCGCTCACCTGACGAGAGGCAGGTC
AMPL156387	ACATAATGCAGTAAATCAGGTTCT	TGCACATAACACTACTGTGAAAAAG
AMPL156388	ACGTTTTAAGAAAAATCCTTACCAAT	TCTTCAAACCAAGCCTACCAATG
AMPL156389	AAATTTCTCTGGGTAATGGAATTA	TCCATCCACAGATGAACTTTCTCC
AMPL156380	AGCTCAGTAAAAGTGAACGACATCT	TCACACACTGGATACTATGTTAAAA
AMPL156381	AGTCTATTGAGAGGAAAAATCCAGAA	ACTGCTGCAATAAAGATACAGATTC
AMPL96988	TGGTACTGGTATCAATTTCTTTTAG	AAAGATTATCAGCTCTACATCAAAT
AMPL96989	TTACTTTACCTTTCCAATTTGCTGA	TTTATGGACACTACAAAGGAAAGAA
AMPL96990	ATTAAAAGGGACAAGTCTAAGAATT	GCTGAAGAGTGCAACAATACATCA
AMPL156391	ATCAAGATGTTTGAGATTATTCCAT	GTCTGAAAAGAAAAACATTATGAC
AMPL97176	CTTTAGCTAATAAAAAATGTGATCCA	TCTTTTGGGTGTTTCTTTTGTGCA
AMPL156375	ATGCCAATTTAATTATTACATTTGA	GCACCAATGCAGAATTTATTTCAGT
AMPL156394	AACATGGGAGGTGAGAGTTAATAA	TGAGAGAAAATCGCAGGGTAGAGTA
AMPL219754	ACTGCTGAATGAGAAAGTAAAAGGG	ACACATAGCATTAAGTGAATGAAA
AMPL110239	TTTACAGAGAACTTCAATAATTCTT	ACAGCTGTTTCTGTAAAAGTGGTAA
AMPL156407	TTACATCTAAATCTACTTTAACTGT	TAGTTGTTTTGCTATCCGTGCACTC
AMPL156398	TTAGATGTTACAAGTCCAAGAGAAT	AGGGTAAGTTAGAGGTGTCAATCCC
AMPL219756	CCACAGTCTCAAAACATTAATATTT	TTCACCTGAAAAGATGAAGTTACA
AMPL156412	TACCTGTCTATAGAATCAGTCTGAA	CCTTTACATGTCATTACATCTCTCA
AMPL156413	TCTAAAGGTCACCTAAGCTAAAGACT	TGATCCAAAATAATCTTGATCTCA

AMPL96788	GAGGAATTACATTCACCTCTTCATC	AGCAGTATAATGCTTCTGTGTCTC
AMPL96789	AAATCTGAAACACTATAAAGCCATG	TTCGTGGTGTCTCTGTGTTCAAA
AMPL219751	TGGCTTACGTTAAAAATAGGAAATCA	TGTGTTAAGCAATGAAGGGAGAAAA
AMPL110241	AGAAAAGATTATGGATACTACATG	AGGAAATCAGATTTTCTGAAGGTTG
AMPL156417	TCTTTTCAGTATACTTTCTTTTGGA	AAATGTGGAACACTTCCATAATAA
AMPL156418	CGAATTACAATGAATTCAAGTTACC	GCAGTTCTTTAGGAGGTAAAAACAG
AMPL113479	GGGATATAGTATCTGACAGTATTTT	AATTTGACACAACCCGTGCCACA
AMPL156419	CTTACTGATTTAAGCATGGATTCCA	AGAGCCTGCTATTTTAGAAAAACAC
AMPL156420	TCTTGATGCCTTGACCTCCTGATCT	TACAAGCGAACTCCAAGTTTGTATC
AMPL193819	GAATGGATAATCGTTCTTCTCTCTG	TTTTTAGATGGTGTCTGTCTGT
AMPL156405	ATTGAGAGTTTGCTAAAGGAAAAA	GGATAATCGTTCTTCTCTCGAAAC
AMPL156393	AAGAAACACACCACATTTTAACCTT	GTTGTCATTGAGAGTTTGTCTAAAG
AMPL156414	AGCACATTCAGATCCATGGGAAAG	AAGAAAACAGTAGAAAGGTTAAACC
AMPL156415	GATGTTACATCGTTCTAAATGTTT	CCAGAATCAAGATTCTGAGATGTAC
AMPL219755	ATGCACATGAATGAATTTCTTCAAT	TTGTCAAGTTGCCTTCTGCTTTGAT
AMPL219752	TGGAGAGGAAGATTAAGAGGACAAG	CCACATTAGTAAGCATAATTTTCCC
AMPL219753	TGCAGTTTGAATGGTCAACATAACA	TGATCAGTTGGTCTTCTCGGTCCT
AMPL219749	AGAGAGGTAGATTTCAATGGCTTCT	GGAAAAGTCTTCATTTCAAGTTGGC
AMPL219750	CCTAGTTTCAGAGTCCATGCTCTTG	TGAGGTTGCTTGTGCTCTGCATTT
AMPL96981	AAGGCGTTCACAAAGTGATTTAGC	GCTGTTAAGTAGCATTTTAACAGAA
AMPL96982	TCTCTCATGAGTTCATACTCATCT	TAGCCGATACACTGTGGGAATAAGA
AMPL96983	AACAAGTAAGTAGGAGGAGAGAAG	TGGAAAAGGGTCCAGATGATATGTT
AMPL96770	TGCTGTTACAATGATTTTGAATTTA	AATACCATTTTCTTCTTTTAGCCT
AMPL96771	TGGAAAATTACCTACCTCCTGAACA	ATGCCATACATGGAACACATCATAA
AMPL97135	ACTCCTCTTCTTTGATCAAAACACG	ATACTTTTTGACTGGTAAGAAGCAC
AMPL156378	TGGTGGAAGCATACTGCAAAATATT	TGAATGTCTGAGGAAGAACAGTAAA
AMPL156379	TTGGTGGACCCATTACATTAGATTT	TCTGCATGAAGACCGAGTTATAGAA
AMPL219757	AATATAGATGTTCCCTCCAGGAATC	TTTGGTGGACCCATTACATTAGATT
AMPL219758	CATTCTTTACTACTTCCCTAAAGAG	ACTAGGAACTTGTAAAGGGCTTCGA
AMPL156384	CCCGAATGATTCACCAATTGATACT	TCTACCCTAGAACCACTGACAAACT
AMPL156385	GGAAATTCAAATTATAGAAAGCTTT	AATTGATACTAAGATTCTGTCAAGT
AMPL156408	TTTCAGTGGTTTAGGAGGGTGCTT	ATGCCAGTGTTTCATAGTTAGCAAC
AMPL156409	GGAGCAGTGCTGAGACTCTGGATTC	CACTTCTTTTGAGCACACGGTCGCT
AMPL156396	GTACTTACTCATTTCTGCCAGTTTC	TGATGCTTTTCGATGACTTACTTTAT
AMPL156397	CCAACTGAAGAAATTATACATTCTT	TTGGACTCTCCTGGGAGATGTTTAC
AMPL156386	AGACAATGAATCCAGAGGTGTACAC	CAGCATTGATGGTACTGGGTAAGTT
AMPL573598	CCAGATGATCGGGAGATACTTGAGC	AAAACCTCTGGAACGAGAAGCGGGA
AMPL520748	GCCGCGACAAGCAACATCATCGT	AAAAGCGTTCTTTGCCGTCCGCTCA
AMPL520749	ACCCGCCATTCTCTAGACACAACCTG	TCAGTTCCGGAGCTGCAACCCGGAG
AMPL520751	ACCAGCGGTAGAAATTGTACATGAC	CAATTGGTAGGGTAAATCCACAGGG
AMPL520752	CGGATACCCATCACCTGGTAGCAA	TAAAGGTCTTTCCAGTCCGGTAGG
AMPL576854	AGCTCTTACTTCAAAATTAGTCCAG	CATGAAGACCACCTTTCTGAAGGG
AMPL520773	CTCTTACTGCACCATATTTCTCTCC	GCAAACATGCTTTGGTAACATTAAG

AMPL520774	CCGAGAGCTTTATGAGCTTCATCAA	TGTGTAGACCCTGAAAATAAAAGTG
AMPL573573	AATTCTCTTTAACAGTACAGTTACA	TTTATTTTCAGCAGCGGGACAAGCTC
AMPL576857	AAGCTTTATTTTCTTAACGTGAATA	AAAAAGTTCTTTAGTTTCATCTGTT
AMPL520797	AATCTTCAGAACGAAGCTCTATCTG	GCAGTAAAAATTTTGAAAGAAGCACC
AMPL520798	TACCTGGATATAGGTCTTTTGATG	GGTTAGTAATAACTTGTTCACAGC
AMPL520799	TCCAAATCCTGTTAGCTGAAATAAT	TTCACCAAGCGGAACAATAAGCTTT
AMPL520697	CTGATCTCTTGCCAGAATTATCTGA	TGCTCTAAAGTCCTTTTTCAGTAC
AMPL520698	TCAGTGAAAATTTAATCTGTCATCA	ATTTGGGATATCCCTTCTCATCAAA
AMPL520768	TGCTGCAATAATTCATAACCATGAT	TTGTCAGGCCATAGTCAGTGTAACA
AMPL520769	TTAGTTCATCCATAATCCACAAA	CAAACCTCTCCCTCGATTATGCCTTG
AMPL520741	GAAGTCTTCATTTTCGGCCAAGTTCA	TTCAAGTGTAACATGAAAACCAGA
AMPL520742	TTGGATAGCAGAAATACCATTGCT	TGTCATCCCTTCAGAAATTAAGAAA
AMPL520743	GCATCAAAGGAGCAAAATTTCAAAA	TGTACGTGCAACATACACTCTAGA
AMPL573564	TACCTCTTGCAATAACTTTGTCTTG	CTCCTGGAACAAAAACAAAATATCA
AMPL326097	CTATCTCGAAATGAAGAGAAGATCA	ACCTGAAAGTGTTAATGCTAACGAA
AMPL573574	TGGCCGACAAAAGTCATTACTCTAA	TCATCACGTTTCTTTTCAGTAGTGT
AMPL573575	AACAAGTGCAAAGCTAATCAATCTT	TGCTGTGAAAGCATTTCTGCAATTT
AMPL520722	CCTTCTTCACCCACACAGGTAGAAA	GAGAACTGTAGATCAAAGCATATTT
AMPL573571	ACGTTTACGGCCAGTTCTACCCATT	TAACCACCGTCACGAAACTGTTTCA
AMPL576852	TCAAGTGTCAAATTGTGTTTCTGCA	ACGAATTGGGCTCTCTGGGAATCA
AMPL520800	TTCTTTTGTGTTGACTGACTCTGATT	GATCCACTTAAATACTTAGAAAATG
AMPL573566	GGCTTCTCTGGTTCATAGACCCAT	TTTTGTTGACTGACTCTGATTATA
AMPL573567	ATACGACCTGCTAAGTAAGAACAGG	ATTTTGTGTAATTTTGGGTGATTTC
AMPL520732	CCCTTAATCTATAAAGTCTGTTCCA	CCACAATATCCATTCAAACCTCTGTC
AMPL520733	ACTTACTGGTTTGTCTTCTCATTT	AAGAACCAATCTTTCTTTAGGCTAC
AMPL520734	TTCCCAAAGCTATTGCTTGATCATT	AAACTGAACTTGAGGCAATGTTATC
AMPL573565	GATCAACTTGATGTGTAGGCAAAGG	TGAAAACATACTGAGATTTACAAA
AMPL520727	GCAAGAGAAAGTTACAATAAAACCC	ACAGTCTCCATTGAGAGAGAGAG
AMPL520728	TTCTTTACTGAGATCACAAAGAACT	GTCTCATTCCTCTATCATTTGCAT
AMPL328830	AAATGTTGAGGTAACATCTTCCATT	GCCAAATCTAAGAGAGGAGTTTAC
AMPL520702	GACGATTTCTTGTGAGTGATAAAGG	TTCAACTTCCAATTCATAGCTGCAT
AMPL326077	AGCAAGTGAAGTAGTGATTATTTT	TGTCCTTTTATCATTTGTTAGGTCT
AMPL520704	CTTCTTCGAATGGAAGAAATAAGTT	CTCTGTTAATAACACACACGGTTCT
AMPL520706	CGGAGGAGAATAAGATAAAATCTC	AAAGCTCTTCTTCGAATGGAAGAAA
AMPL520707	AAATTCTCAAGTGCAGTACCCTTAG	TAGTACCTCTTTTGTCAATGAGTGA
AMPL520708	AATTCTGTTGAGAATTCACAGCTGA	TCCAAGTCTGAGAGTCCACTGAGAG
AMPL520709	TATGTTTATGTACATCACAGTCACA	GTGACAGCAACAGGTGCATTTATC
AMPL520710	TCACTGTTATCAACATCATGATTGT	ACTTGGCTCATCAGAAATATTATCA
AMPL520711	AAAGATTGCTTTTTACTGTTCAAGG	AAATTCTGTGCAGGGCTTCTGTGTA
AMPL520712	TTCAGAATTATTGTCCAACAAAAGT	ACATCGTCGAACCTGTATTAACAT
AMPL520713	GAGCAATCAAAAATATCCTCATGAT	TAAGAAGTGAAGAGACCAGAGGCGTT
AMPL520714	CTATTGCTATCTGATGTATGTTCCA	TCTTCCTGTACACCTCTCTGATCAC
AMPL520715	ACATAATTTGCATCACTTATCTCCT	TCAGAATCTGGACTACAGAATCCTA

AMPL520717	AGAAAAC TGAGAATAACCAGGAGAT	TATGATCTCTTGGTATTAGTGCTTG
AMPL520718	GTTTGATTAGAGAGTGGTGACTC	ACTCAGTGGCAAATGTACATAATTT
AMPL520720	ATAAAGATGGAATCTTACCTTCAAC	GTTTACTTTTTCTTACTAGAAATCT
AMPL520744	CATTCCTTTTGCTCTTTTAACTTT	GGTTAATAGATTTAGGTTTACCCAA
AMPL573561	GGTATAAAAGGAAAATAGCCTTTGA	TCTCGTCATCTTCACTTTCGTTACT
AMPL573597	GGAATTTGGATTACTATACATTACC	AAAAGTCTCAGCAGAAGCCCAAGGT
AMPL520700	TGAAGTCTTCTAATGTGAACATGG	GCTCTTACAAC TCAGGCCATGTTA
AMPL520701	GATTACCTTTAAGTGACTCTGTCTC	TCTCACTCTCATCTAGATGATAA
AMPL520735	TATTCTGCATCTTCTTCAGAAAGTT	ATTCTGGAAAAC TGATTCTTATCA
AMPL520736	TACCATTTATAGCCTGTGAAAGTTG	TCTAAAACTTCCTAGCTACATGCT
AMPL520737	ATTTAATTACCTTTAAGGTCTAGTT	TGAAAGTTGAGTTTCATCATTTAAA
AMPL520761	TTCATCATTGGACTACGCAAAGATT	TCAAATTTACTTGATTCAAACCTTG
AMPL576855	TCATCATCCAGAAATTCTCCATTAT	GCTCTCATTT CAGAATCTGTGTAAA
AMPL520776	TTCTTCACTTGATTGGCCTTTGCAA	ACTTCTAGGCACATGAAATCTGCAT
AMPL520777	ATGTGCACAATTTTGTTCCATCATT	TCCTCTTCATCAACACAAAACTAT
AMPL520778	TAACCGCAATATTAGATTCTCTTTT	TACTGCACGTCGAGTTT TACTTTT
AMPL520779	CAGAAGGCACTGAATTTAAACAATG	TGTTCTCCTCCTCACTTGAATCATC
AMPL520780	ATCCTTTAAATT CAGCGATGTCTGT	TGTTCTTCTTAACAGTGCTTGGGTT
AMPL520781	AAATTTTCTACAGTCTTTCTGTGAA	TGCTAATGGATTAACTCTTGGAGTA
AMPL520782	TATTTAATACAAAACACACACACAA	GAAGGGTGGTGTGGTAGACTGGACT
AMPL520756	TACAAGAATACAGGTTCTTTTTCCT	ACTGGTCACATT CAGGTTAAATGCA
AMPL520757	ACTCACGATGTAATCAGGCCATTA	TCTAAGAGAAGTATGTGTCCAGCT
AMPL520758	CATATTCTTTCAAACATACTCTGCA	TTCTACTTGCAACCATGAATTGCT
AMPL520759	AACCGTAAAGTTCTTTTACAGTCTT	TGCTCAATGAAC TTGTTCTTATTGA
AMPL576856	GCAGGAAC TGAAAAGAATTCGGATT	ATTTAGGGAAGGCAAAC TTCAGTA
AMPL520765	TGCCTCACTTTTATTACTATTACC	TAAGGTAGTCAGCAGGCTGTCATAG
AMPL520766	ACAACAGACATACCTGTTAGCCATC	ATGAATACCAACATCTTTCTTTGT
AMPL520767	CAGGCATGATGCCGCTTAAATTTTT	TGAAAAC TGGTGACACATATTTAAT
AMPL520783	TCTCCTCAGCCTTCTGATGAGTTAC	GGAAATCATCTCAAGCATTATCCGC
AMPL573595	CCCATCTTGAGCAGCTTGATTATAT	TGGAGATTTCTTGAAGTGAGCTGAA
AMPL440950	TTTACCTGACAGGTAAGGATCTGAT	ACAGAATATAACATAAGTGATAGTG
AMPL440951	TATTTACAATTAGCCTATGAAAAAG	ACCTGACAGGTAAGGATCTGATGTC
AMPL440942	CATAGAAGTTCATGGACACCTCGAT	AGGAAACCTTCAATTGATACTGCTT
AMPL440943	TCTGCATGTGCCTCATCCCTAAAAA	TCAGACCAGTCACCTTCATAAGCTC
AMPL795670	AGATAAGAATGCTGGTGAGAAATCA	AACCCTAATCCATCCTTTTCCCTTC
AMPL1895375	GTCAATGGAAGAAGGTACGAAAAAT	TGCTTTTATCCCATAGCCTAAACA
AMPL440947	AAATGCTCATCATTATACAAAAC TG	ACTTTGCCATATATGCAGTCCAAC T
AMPL440948	TGCAGACTCTGTGTCAATGTACACC	AACCTGGTGGACCTGTAATCTAAAA
AMPL440949	CCTATGTGTAAATGTTTCATGCTGT	TAATCCTCCCATGTTGGTGGGTAAT
AMPL440937	GTTAAAATATCTGGGAAAACGGGAT	CAATAGAGCAGGTATAAGCAAAATG
AMPL440938	ATCTAAAGCAGCATACCTTTGTAGA	TCAACCAGTCTGTGGGTGAAAAAAA
AMPL440939	GAACACTGAAAAAGTTACTCCACAT	ACAGGTGAGTTCCCGATAAAGATGA
AMPL440940	TGTGCATCAAACCTCTTTCTGACCA	AAGGTATTGCAAAGGCTAAGCAACT

AMPL440941	AACTCCTCAGCCAAATACTTCAAGG	ACAGAGTCAAGAATCACAAAGTTTAA
AMPL440945	CTCCACTCAGATGGGTGTGAATCTG	GAAGACTGATAAATACACAGGTAGC
AMPL1192853	CTATGTCAGTGTCAAGTTTCAGCCT	TAACCTGTGAACAGACAGGGAGAAG
AMPL1895376	GGCTATCACACAGCTGGATCCAGAA	CACAGAAGTTGAGGAATCAGTGAAG
AMPL1895377	CCTTCATAGTCAGAGAATACTGTCA	GAAATGGTAAAAGCACGCTTTCAGA
AMPL1192854	GCAGCACTGAGCATAAGCTTAAAGA	AGGGTTAGTCAAAGGATCAGGGATG
AMPL440936	CTTCCAAGCAGTACGGTCTGTTTT	ACTTCCCACCAGCAACCTAAATGGA
AMPL440933	CCAATATTTTGTCTGTGGCAGAA	TGCATAAAGCCTACTGAGGCTAAAA
AMPL1232455	GCACATCTGTAGTCTAGCTGCTCA	CCTGAGGCTTCTGAAGAGGATGGT
AMPL1232456	TGAGGTGGGTGGATCTCTTGAAGCC	GAGGATTTCTGGAGCCAGTTCAGT
AMPL202873	AACATTGGTCTTTTGTGAAATGGTC	CAATTTTCTGGATATATTGGGCCTC
AMPL188911	CTGTGCAGCAGCAATCTTGACTTCT	CTCATTTGCTGGCTGGACAAAAGA
AMPL188912	TTGGGAACATGGTTTTGACCTTTTT	CATTGGGACTTACTTCTCGGTCAG
AMPL188889	CCTGTGCCAAAGAGAGTGAGTCGTT	CATTCTGAAAGGCACTGTGCAACA
AMPL202872	TACTCTTGACAGTCTATTTGGGATA	TGTTTCAGCTCATTGTGATTAACCC
AMPL188887	AAGCTTCAGTCTGTCAAAAGCCTA	TGGATAGCAATGTTTGTCTTTCTT
AMPL202871	TTGGAAGAATGTGATCAGCTTATTT	TAAGTAATGACTGGCTGGGACCACT
AMPL576813	AGAAGCTCTGCTTGGTACTACTATT	ACATGATTGTGGTTTTGTGTTGCTG
AMPL576814	CAAAGTACTGACTTTTCATACTGTT	TGTTATTTGGTAAGCTTTCCTCTA
AMPL188904	CTGATCAACAAGTAGAAGTCATGAC	TAGGATCTCGTTTTCTGTGCTGGG
AMPL188905	AGTCTGGAATATAAAAGCTGTGCT	ATGGAGGGTAAGAAAAGCATTGATT
AMPL188906	TAAAGGTTACTCCTCACATCACCC	AGAGGAGGCTAGTTAGTAGCAGTGG
AMPL202876	AAACAACCAAGCATAATTTTGGCT	GAGTGATTTGTTTTCTCTCATCTT
AMPL202874	ATGTATCATAACTGCTTGCGAAGAT	TAGATGGGCAGCTTACCAAAATTGG
AMPL576812	GAGTGCCACTTTAACAGAACTGTTG	TTGGAAAGCTCTGGATGCTTGGCAG
AMPL188901	AGACATTATACATAAAGTGTAGACT	TTCCAGTGTCTCGTGTCTTACATAT
AMPL188891	ATCAGCTCTGTAGAAACACATGCCA	CGTTTAAGAATGGTTTGACGTGTTT
AMPL188892	CAAGTGTCAGGCAGGCAAGGACAAC	AACATTCGGTCGAACAGGTACAATC
AMPL188893	CTACAGAAGTTGCTGGACGAACCTG	TACCTGTGACTGTGACTCTGTCCCG
AMPL188894	CGGTTGCGCCTGATGATAATGACAG	AACTTGCTCATTTGAAAGACTCAGT
AMPL188895	CTACCAGGAAAATCACATCCCAAAA	ACATGTGTTACCTGCTTTCCCAT
AMPL188896	TTTTCCAGAGGAAGTGAAGCTTAAA	CAAAGCCAGCATACAAGACGGGCC
AMPL188897	GACTTTGGACCTCTTAAGCTTGAAA	AGAAATGGAGGACTTAGAAGAGGAC
AMPL188898	TTTATCCTTGGATGATGATGCTTTC	TGCTCAGAAAAACCAGTGGAGCCCT
AMPL188899	TCTGTTTTGTGGGTTTTGTACTA	TCATAGGGATGGAATGCTGAGTTTA
AMPL188913	TTTGCCAACTTCTAGCCTGTGATT	TATTTGCCTGGCACTTCTTTTCTT
AMPL188914	CTGAAGATAATGACTTGTCTAGGAA	CAAGGAAGAAGTCACCTCACACAAA
AMPL188915	AGTGAAATTAGGATGTCTGGCACAT	AAGCACCTGGTAGAAGATACACAGG
AMPL188916	AGAGTCATTTGGATGTCAAGAAAAA	CCAAGTTCAAGAACCTCTCAGAAAC
AMPL109758	TGTCCAATTGCCAGAGGAAAGTAGC	AATTTAAACCTTTCCAATGAGGAAA
AMPL188918	AAATCTTTAAAATCTCCAGTGACA	TGAAGGCCTTCTGTTTCTGCAGAA
AMPL188919	AAGTTCTAATTTAGAGGCAAATATT	CAGGAATGAAAATCTTCAGGAAAGT
AMPL109764	CTGTCTCTACAGATAACCTCCTTGT	AAATGAAGTACCTACAATAACTTA

AMPL109765	GAACTTACTACTCACGACCTAAAAA	TAAGTAAAAGTGGCCAACTGCCAC
AMPL188920	GACTACAGTTCCTTTACAGACTCTA	ACTTCACCTGTAAGTTTGGAGGCAC
AMPL188921	TTACAGAAATTAATGAAGACAGTGT	TAGTAGTCAGCACCTTGAACACATT
AMPL188922	ACAGGAAGAAATCAGTAGCAAAAT	AACTGCCCCAACAGAAAAAGGTGTT
AMPL188923	GCAGCAGAAGAGGACATTTATTTCA	ACCAGTAACTGAAATAAGAACTCAC
AMPL188924	AGATGGCCCAGGAGGATTACCTATA	CTGTGTCTTTGGCACTGATTCACTC
AMPL188925	TTATGACAAGTTACACATCAAAACC	TGACACCCAAGAACATTTTCCCCAC
AMPL188902	GAGCTGAAAAGATTAAGCATTCTAT	TTCCCAGGTATATTTTGGCTATTGT
AMPL188903	CAATTTAAAAACAGTTCTTGACTCT	TAGAAGAACAAGATTGTTTGTCTCA
AMPL202868	ACTCCACCTTTCCACTTGCCAGTA	GTACAAACAATTTAAAAACAGTTCT
AMPL202867	GCGACGGCTGCTCTTTTCGTTCTGT	TTTCCCAGGGTTTTAGGCCTGGCTT
AMPL2514	AAGTCCAAAAAGGGTCAGTCTACCT	TGAACCCCTTGCTTGCAATAGGTGTG
AMPL43	TTCAAAGCATTGGTCAGGGAAGG	AAACTCATGTTCAAGACAGAAGGGC
AMPL8	GTGGGCGTGAGCGCTTCGAGATGTT	TGGAAAGCCATAGGATTCCATTCTC
AMPL9	AACCATCTTTTAACTCAGGTACTGT	AATGAGGCCTTGGAACCAAGGATG
AMPL17	AGAGCATGAAATGGTTCTATGACT	TCTAAGGAACATCATAGCTTAGCCT
AMPL2572	AGACAATGGCTCCTGGTTGTAGCTA	CTTGACTTACGATGGTGTTACTTCC
AMPL16	GTTATGCCTCAGATTCACTTTATC	TCAAGTGGAAAGTTTCCAGTCTAAC
AMPL26	AGGAAGAGAATCTCCGCAAGAAAGG	TTTCCTTGCCCTCTTTCCTAGCACTG
AMPL27	ATCCTGAGTAGTGGTAATCTACTGG	AGCCTAAGCGAGGTAAGCAAGCAG
AMPL28	TGGGAGTAGATGGAGCCTGGTTTTT	TGTTTGTGCCTGTCTGGGAGAGAC
AMPL44	AGGCCATCCTCACCATCATCACAC	ATCCACCTCTCATCACATCCCCGGC
AMPL189161	CTCCCCAAGGCGCACTGGCCTCATC	AAGACTCCAGGTGAGGAGCCACTTG
AMPL14	TCCTCAGCATCTTATCCGAGTGGAA	GTTAAGGGTGGTTGTCAGTGGCCCT
AMPL15	AGCGCTGCTCAGATAGCGATGGTGA	TGTGGAGTATTTGGATGACAGAAAC
AMPL18	CGCGTCCGCGCCATGGCCATCTACA	ACTGATTGCTCTTAGGTCTGGCCCC
AMPL19	TGACTTTCAACTCTGTCTCCTTCCT	AGTCACAGCACATGACGGAGGTTGT
AMPL20	TGGCCCCTGTCATCTTCTGTCCCTT	TCAATGCCTGGCCGTATCCCCCTGC
AMPL117516	ATATTGAACAATGGTTCACTGAAGA	CAGAAAACCTACCAGGGCAGCTACG
AMPL22	CTGCTCTTTTCACCCATCTACAGTC	CAGATGAAGCTCCCAGAATGCCAGA
AMPL23	CAACGTTCTGGTAAGGACAAGGGTT	CAAGCAATGGATGATTTGATGCTGT
AMPL6	CTCTGAGTCAGGAAACATTTTCAGA	TTTCAGACTTCCTGAAAACAACGTT
AMPL7	GGGTTGGAAGTGTCTCATGCTGGAT	TGTGAGTGGATCCATTGGAAGGGCA
AMPL173198	CACGGGCCGGTTACCTGCTCGTCGA	AGGGAAGAGGCAAGGAGAGGGTCTG
AMPL173199	CCTTCCTTTCCCTCCAGAGCCTGAGG	ACGGCCTGGACCCATTCCACCGGCC
AMPL29784	TCCTTGTGTGCTCAGTACTGACTT	TTAGGAAAACCTGCATCAACCACAGA
AMPL29785	CAAAACACAGTAACCCAAATACTCA	TTCTGCTGTCTGTTTTTATTGGAA
AMPL29786	GAAAATCACTTTCCATACTTCAAAA	ACATTCTTTAAAATAGTAGTGAGGC
AMPL28653	TCACAAAGCCTGCCTACTTAAGATC	AAAAGTTAGTTTCACTTTCTGCTAC
AMPL30445	AGCAACTGTTTGACCAGCATCGTTT	GAACACCTTACATACACCACCTA
AMPL30446	CCAGAGGCAGAATTCCGAAGTTCAG	ATTGTGTCCTTTGGTTGCTAAAATA
AMPL267900	TACTTCAGTAGTCCCATGTGGATTT	TTTCCTTCACGACAGGTGTGAAGAA
AMPL173477	CACATTGATATACTGTAACAATTCT	ACAGGTGTGGTATCTCACACCTGTA

AMPL173478	TGATACTAGTTTTTGACCCAGTGAT	ATCAACATTGTCTTCTGAACAAACA
AMPL25764	TTGTAACCTACCTTTTCCAGGCTAT	AACTCAGAGCAAAATAACTTTCCAT
AMPL25765	TTTTGATTCAAGGATGCTAACAAACAG	AACTGCGCAACCTTCTTTAGGGCTT
AMPL93965	GATAATTTTCTACCCAGTTCCAAAA	TCCTGAAAAATGAATGTGAACTATC
AMPL93966	ACAAGTAGGCATTTAAAAGACAAAA	AAAGAACCACATGTCTTTACAGTGA
AMPL26194	ATTTCTGGACACAAGATAAGGAGAA	AGATAGGAACAAAATCCCTGGCATT
AMPL26195	GAGTAAAAATAAAGTGTTCACATCT	CAAAGTGTGCTTTACGTTTGGTGC
AMPL23956	CACCTTCTGTGAGCTGCCTACTTCC	ATAGCAAATTAACAGCCTCTAAAAC
AMPL31923	GAACAAGTAGGAAAAATGACAGAGTT	CAAGAGCTTTTCGTAGACTGTCCAG
AMPL31924	TCTTTGACATAATACTTATGCTAGA	TGATGTAAGTACTTGTCTTACACAG
AMPL24485	GCAGTCAATCATTAGATCCACATCT	ACACATTACCCACACAAATGGCAGT
AMPL24486	TGTTTATAAAGGATAACAGCATCAG	GCCTCTTGAGAATGGCTTACTTGGGA
AMPL26720	CGATGGAGTGAATTTACCAGCACAT	CCAAAAAGAACAGCAAATAGAAAGT
AMPL26721	TTGAAAACCAAGAGTGCATTTCTTA	GAAGGTGAATTCTGAGCCAGGCAGA
AMPL267901	CAACCGAGTGACAATACACAGCATC	GGAAGTCTTAAGACCATGCAAACA
AMPL267902	AGGAGGTGAGATTCAAATTCGTGA	ACCAATCCAATGCGGACTAGAGAAA
AMPL193773	TGAATTAGTTTCACCATGGACAAGA	AAGACTCAGGATAATAGCTATCCAA
AMPL193774	CTAAACCCAATTAAGTGTGTGAT	TAGCTTCTTGTCTCCAGGTCTGTAG
AMPL93990	TTGCTTCCTGAGCAATCTCTGGCAT	AAATCGAGAGAGATTAGAAGGAGGA
AMPL93991	AGGAACATCATGAATGTACATAGTT	AGTTGGACGAGCCCTGTAATTAATT
AMPL21260	TAATCTCCCAAAATGTTTCTACAGG	TGGTAAACATAAACTGGAGTACTT
AMPL21261	AGTGTAAGTTTTACAATGTACTAAG	GGATTCCACAAATCAATGTCTATCTA
AMPL31933	AATATTTCCCGCAACCACTTGAGAA	TACTAGCATCATCAATAACCTAATG
AMPL31934	CCATTCCAAATATTCTTCCATTTTT	CTAAGCATTTGATGACTAGTTAATT
AMPL31445	AGAGAGGCTCCAGGAGTACGTAGTA	TAAATAAGCTTTACTGAAGCACTGA
AMPL31446	CATTGAGAAAACAAACAGAGCACAT	TCCATGGACATTTGACTGGTATTTTC
AMPL25754	TTCAGTGTGAGGGTTCCACAGAAAC	GAAGACATATATTTGCAACTTGGGT
AMPL25755	TATAGTTGGGCAAGAGGTTATGCAC	CACCTCTAGTTTGGTCTGGGCTTGT
AMPL25756	TGCTTTGAGGCAGACTGAGTAAAAA	CACACCGGATATCTGCTTCTCACA
AMPL27810	CTCAGCAGTGCCATCACTCTTTTCT	AAGAGCAGAGCAAATGCTTCAAATG
AMPL27811	AACAAGAGGCTTCCAAATTGAGGGT	CTTCCTAAGCAAATACAACAGCAAA
AMPL25757	GATCAGTTAACAGACAAAAGTCAAC	ACTTGAAAGCAACTTAGTCTAGAGC
AMPL26168	CTATGGATCCTCCTCCACTCACATG	GCTTTATCCAACATTTCTTCCATGA
AMPL26169	GGCCAGAATTGCTTCTCTGCTGGAG	TCTTAACAATGGTCTTGTGAAGGCT
AMPL26171	AGCCGATCCATAAATTTGCTGACAG	TATAGGTTGCCAGGCCAGAATTGCT
AMPL26172	ATACAGAGCAGGACTCAATTCTAGA	CTGAAGACATCACTGAAATCATAGA
AMPL26173	CTATTCATAGAGAAAGGTGAAAAAT	CTTAACATTGGTCCGTATTGGAAGT
AMPL19472	GTTTCAATGCTAGCTTGCCCTAGAT	CATTAATGATTTGCCATCAATGCCC
AMPL267904	GTGTAAGAGGTTTTTCTTATCTCTT	CTAGCAAGTCTTTCATTATAGCTAT
AMPL29755	GAGGTCATCTCTCCTTGCCATCATT	AGAGAAGGCATTCTAAATACTTTAG
AMPL29756	GATTTACAAGACCTACATTGCTCT	TTGACACAGTTTCGTTTTTATTTGA
AMPL23990	AGTAAAACCCACTAATACTTGAAGG	AGAGAAACAGCAAACAATGTTTTTG
AMPL24831	GATTGGCTTCCATCAATTCACAC	TGAATCTTTCAAGTCTTTCACTTC

AMPL24832	TATTTGCTTTTGGCTTTATGTTTTT	CAGAGGGAGACCAGCTAGAAAGTGAA
AMPL20234	AGTTTGACATTGCAAGGACCGTACA	AGACTGACCTTTACATAGTGAAGGC
AMPL20235	GTGCATGAGACCACTGTCTACGTTG	TGTTTGCGCACTTTTCATCTTCAACT
AMPL26687	GAACCTCCATAAATGTAGCTCTTGT	CTCTTACATGCCAGTTCTCTAGGTT
AMPL26688	GTTCTCCTTGATCACCCATCATTGT	TGTGGTAACCTAAGCCTAATGTGGA
AMPL26689	TTCTGAAGGATTTGCTATGTGCCAG	CCAATCTCTCAAACCGATCAGCCAA
AMPL28911	TGCCAATTCTACTTCTTTAGAAAAC	TTCTTAGGTAGGAAGCTAAAGAGTA
AMPL28912	ACATACACACAAAATGAATGATACA	AGTTGGTAGAGTAAATGCCGAGAAT
AMPL28913	TAGGAGTTTATTATACCTGCTCCAT	AGAATGTCATTATTTTACTGGCCAA
AMPL620150	ACAAACCTGGTAGGATCCACTTCAA	CACGAATAGGAGTTTATTATACCTG
AMPL93974	GCAGTGTTAGTAAAACAAACAAACA	TGCCAATCAGAGGATGTGATCACAA
AMPL20668	AAGAACTTTTCAGTCATCTGAAGGA	ACATTAAGCCAAGTCAACGTGTAG
AMPL20669	TATAAACAGTGGCACACACTTCGAA	TTTTCTCAAGGCTCTCTGATGGTT
AMPL20670	GATCCTTCTTTGATAAAGCATTCTT	AATTCTGAGGAGGAAGTGTGATGG
AMPL23893	AGATTTTGTTTTGGCTTTGAAAGAA	AGTCACAGCACCGACTACATACAGC
AMPL93967	CCTGCTTCATACGGTGAGACAATGG	AGACATGAAAACAAAGAGTCAAAG
AMPL93968	ATCAAAATAATTCACCTTTGACATT	AATCTGAGGAACATGGCACTTCTTA
AMPL93969	CTGAAAACCATAGGGTATTTCAAAT	CCTTTCGATTCTAGGTGGTGGCTTT
AMPL20313	TTTTGTGAAGAGAACATGATTGGCA	TCTGCTAGTACTTAAAAATCAAACA
AMPL267898	AAACAGAGCAACTGAGTAAGTGGCA	CTCTGAAGTATCTACACAATAAAAC
AMPL22549	TACTGCCCAATTTTCTCCTGATTGT	GCAAAGTCATTAGAATTTTCAACAC
AMPL22550	GTAAAAAGCACTATTCATGACCAAT	CGATGTAAAGCAAGCACATTGCCGT
AMPL25541	CTGGAGGACCCAGGTATGCAAGAAG	GCTTTTCTTTTACTGTGCTAGTGAT
AMPL25542	ATAAATTCCTCAAACCTTTGAAGTGG	CTTATCAAAAGGTCGTCTTCCAACA
AMPL25543	CTCTCCTTCTCAACCAAAAAGAAAG	CAGTGTGTATCTGCCACAGGTTTGT
AMPL93978	AGCTTTGGAAGTCCCAGCTTGGTAG	GGCTTTATGTTGTCTACTGAGTATG
AMPL93979	GCCATCTCTCTATATTTGCTATAAT	CGTTTTCAAAGCCTTGAATCTTCT
AMPL20609	CAGCAAATCACCATTGATTTGACCA	AAACAATTCTTCAAAGTCTGGATT
AMPL20610	AAGTCTGTTTTAAAGCGATTGCTAG	AGTTTTGAACCTAAGGAAAATGACA
AMPL20611	TAGATATAGACTGCGGAGACGTTGT	ACAATTTTCATATGGCTTTGCATAAT
AMPL20612	TCAGCCAGTTTCCCAGGACAGTCTA	CAGGAAAAACAACAACCACTTAGA
AMPL20613	ATACCTTCAGGTCCTCTTCTAAAGC	CTTTTGCTACCTTTGAGGCCAGTCA
AMPL20614	TTAGTCAAGAAAAGCAATGAATCGT	CAGGTAGTTTCTGTTGTTTCATGCTC
AMPL25353	TTTTGTTGCTCTGCTGAAGTTACT	CCATTTAAGGAAGAGTCCCACCAT
AMPL25354	AGGCAGATTCTTCAATTTCCGAAG	AGAACCAACCTGGAAGTGGTGGTG
AMPL25355	CGGGTCCGGATATGAATGATAGACT	TCATTTAGAAAGACTGATTGCCCTA
AMPL25356	AACTCGGGTCAGAACTGCCTAAATT	TCACACTCCTGGTGCATGAAGGTGA
AMPL25357	GTTTTAAAATTCATTAAATGGTCTC	TATTGAGCAGTGTCCAGGACATC
AMPL30159	TGTTGTTGGCAGGGATACATAAACC	CAAATGAAGTTGTTTTAAACGTAT
AMPL30160	AATCCAGAAATACACTCTTCCAAA	GTAAGTGGCCCTCGATTTTAAATT
AMPL30161	AAATCTGAAATGAAAGGTTTCT	CTCATTTGGCTGCCAGTGTCTTACTA
AMPL31831	GGTCATTGTTATCAGCTTGTCAGA	ACACTATTAGCATGTGATGGTCAAT
AMPL31832	TAAATATGTGCACAAAGGAGATACT	AACTCTTTGTCGTTTGGCATCATCA

AMPL25924	GCACTGGTTTTGATGAACTGTCTA	TGAAAATTCTATTTTACCTGAGGCC
AMPL25925	GAAGCCAAAGCTACAGCAGTATCTG	AGCAGATCTGTAATCTAGTTGAAGA
AMPL25926	ATGCTATAGTACAGAAGGCAAAAGA	GCTTTTATTGATCCCAAGCCACCTG
AMPL24875	TAAATAGCAATATCATCCCACATA	ATGGGACTCAAAGTTTTCAGACAAA
AMPL24876	GAGCGGACCTGTGGCTACTAAGAAA	TAAAGTAGGAGTTGGAGATAAGCAT
AMPL268113	AACTGGTAAACACAAATTCCTTTCC	GAAGAGGTAGGGAAGATGAGCTGCC
AMPL21440	TGTCAGACTGAGTCTCAAACTTGC	ACATAATCATGTTCCCTAAATCAAA
AMPL21441	CCCGGTAACGGAACGGAAGGCAAT	TCTGTAAGAGAATAAAGGAACAAGT
AMPL21442	CATTTTGGCTTCTATACCTCCATGA	ACTTGACTTTGCTAATGCCAAACAG
AMPL21443	ATGCAAGTGTTTTATTTAAGCATT	AGCTTCTGTGACTGTTTCCAAGGAT
AMPL19623	TACCTTTGAGCTAGTCTGTCCACT	TGACCTGTTACCAATAACACAGTCC
AMPL19624	TTTCAACAATATGCATATTTTACA	TTGGAATATCTCTCATGCATGCCTA
AMPL29022	CCCATGAGACACTCGTTTGCTAATA	TCCACAGAACTTATAAAAACCTTTG
AMPL29023	GATATGCTTTACAACCTTGAGAACCA	AACAAGAGCTCTTGGTTGCAGGGAT
AMPL26350	TGACTGTTGTAAGTGTGAGGTCCCTT	TGCTAGAATGATTTCTCATGCTATG
AMPL267903	CCAAAATGAATGCACTCATTTTCTA	TCAAGTGCCTTTTAGGAAAACAGAA
AMPL30207	GTATGCAGGTTTTGTTCAAGAAGTG	AGGCTAGTAAGATGAATCATGACAT
AMPL267899	AGTAATCTAGGAACCTCAAGGCAAA	GAATACAAGTTGACCTCATCAAGCT
AMPL27692	TACCTTTTAAAAGGTGTCCAACCAA	GAGATGGTTGATGAGGTTACACAG
AMPL27693	ATAAGGGCTAACTTACTTCAATTTA	GTTAAAGTTAGAGTTGAAATTGAGT
AMPL23237	GTAAAATTCTGACTGTTCTTGCAAC	TCTGCAAGGAAAGGAAATATAACTC
AMPL267905	CGCTCTGTGTATTCACTTCAAATTT	AAGGATGCCTGTACCTTACAAAAAA
AMPL267906	TAGGTATTTGTACTTTTTGGGTGAT	TGTGTTTGTAAACCAGAGTTAGTAG
AMPL28883	TGCAGCGACTTCGAACTTCTTCAGA	CTGTTAAATGTGCACATATTTTGGGA
AMPL28884	GAGTAAATCCACTTACCTATAGGAA	AAGCTAAAAGGAAAACGGTGGCAGG
AMPL28885	TACCTTCTAGATTACTCTGCTGCTT	TGATGAATGGGATATGTATCCATAG
AMPL22047	CTGCAAGGTATCCTTCAGAACCTTT	TGTCCTATTTTCCTAACACAGCAAT
AMPL22048	CTTCTTAGTGTTGGCCTGAGAAGGT	ATGGCTGAGTCTCCTTTAGTGTCTT
AMPL22049	AAATTTTGATAATCCTGAAATTTCA	CAGGCTCATGGATTTCTGCTCGG
AMPL25520	GATTCCATTTCTTGCCCTTTTAGGAG	TTGGACATAAAGTGTTTAAGCTTTC
AMPL25521	ACCTTATTTTCAATGATCTAGATCA	ATCAAGTGATCAAACTTTTCCTTG
AMPL25096	ATCCAAGAGAACATTTGATTACAGAC	ACCTGGATAATCACTTCACCTAACT
AMPL25097	CCTAAAAGTAGACTGGAATAAAAAAT	ACGTAAATGTGGGTGCTGTTGTGAT
AMPL30402	TGCTAAGTATTCTATAAAGAATTCGT	TTGGGTGAGAAGTAGAAGACTGTAT
AMPL229900	ACGAAGGTGAATTTAAATCAAAAGG	ATCTGTGGTATATTTTACCAGTGTA
AMPL31647	TTTGGATCTTGGCACAATGATAACA	CCAGAATAATGTGCCAAAATTAAAG
AMPL31648	TTGCAGGTAAGATGTTTGGTATTGT	TGTTGATCTTAGAGTCCAACAAATT
AMPL31649	CTGTAGTCAGTGCATTCTACAACAG	TTCATGGTACACCACACTCTGCACA
AMPL23325	TACCTTTGAAAACGGTCTTGCAAAC	ACCCTATGTACATGTTGAAAACATG
AMPL23326	ACTAATTTGAACAGAAACCTTCATT	AAAACCAAAACCTGCAACAAATCAA
AMPL19504	TTTCTTCATCAATTCAGGCAGGTA	CTGAAGCCAAAAATATTAATTTTGT
AMPL19505	GAGGTCATGGAATTAGAAAGGTTAA	AGGCATCAAGAACTTAACAATAAG
AMPL19506	GTCTTATATTGTTGCTCAAAGTCAT	ATACTAAGCTGGCTCTGCAGTGCAG

AMPL23486	AATGCTATTACGTTTGAACTGCCA	AGCCAGTTACTTCTTCTTTTCCTTC
AMPL23487	CTAGTGAAGAGCCCATGTTAAGTTG	CTTCTGCACTTGGCTTGCGGATCCG
AMPL623321	GTTGCTTAACTCCACTGACTCCTTA	TTGGGAAACAGGGAAGCATTGGGGA
AMPL222539	CACCATCGAGGGAGCAGGAGCATCA	TAGAGGCTGAAACCTTGCAACTAAA
AMPL222540	GTGACCAACCACATAACTCGAGACA	CAAATCTTCCCGACAGGTGAGCCAG
AMPL794189	CCTCTATACTAGCATTATGGATCTG	GATTCTCCTGGACACCATCGAGGGA
AMPL1015515	TCTGGAGTCTGGTCTCTGGCCAGCC	AGGCTTGGCCAATGAACCTGATTTTC
AMPL222537	AGGCTGACAGGTTTCATGAGTGCTTC	AGTGGAATGTGACCTCGCCTCCCAG
AMPL222521	TTCTTTTCCACCGCTTCAGGCAGAA	GTGCTATCTCTGTGGGCAGGTGTCC
AMPL222522	TCTTCCTTCTCAGCCTTACCTTGAC	TGCATGCATTTGACATCTTCCAGAT
AMPL623320	TGTCTCTGTATGGCAGCAAATGTGG	TCCATTGAGATTCTGACACAGGAAA
AMPL222530	AGCCCATTTGTGTTGCTGGAATCTG	CAGCAAAACGTCCTATATGTAGATT
AMPL222517	TGCTTGATGCTGGTCTCTATACTGG	CGTACTAATGGGTTTCAGAGCCCAAC
AMPL222518	TACTTTTCTTCTTGGTACCTTTTCC	AAATGTAGAGGCCCAGGTAGCGG
AMPL115313	GGGGTCAGGAATGCTCTGAAGAAGG	GTTACTGTTTAAATCATCATTCTAG
AMPL1015514	TGGCCTTGGGCTTAGGGAGTTTATA	ATGGAGGAACAGATGGACTTTCGCT
AMPL222523	GATCTCTACGAGGAAGTGAAGACCT	ATGCTTGAGCTAGACGGGAGGGGT
AMPL222524	CGGGCAATTTGTTGCTGAGCAGTCT	CATTGGCAGGTTTGTGTGTACTTT
AMPL623322	CAGTCTCTCTGCCAATCCGCTATA	AAAAAGTCCATCCCAAGCCAGCTGG
AMPL222519	CGAGGAGATGATCCAGCTTCTCAGG	CCTCCATTCTAGAGGGCCGCGCTT
AMPL222520	GGACCCCCCGCAATGCCACGTGA	CAAGACAGGTGACCGGGTGTGCGCG
AMPL213612	GGTGACGCTCTCTGCTGTTAGAGCT	AGCACTCCAAAAAGCAACCCACC
AMPL213613	CTCCACCAAAGGTTTGATAACTGTG	CTGCAAAAGTCCAGAAGCTCCTCCA
AMPL213614	CGGTCTAGTTTGAAGGCCATGTCAT	CCATGTCTTTGGAGTGTGTTGACTT
AMPL213615	GGTCTGTTTAGCTTTTAGTAAGTCC	TCGGAATCAGAGCTGATATCATCAT
AMPL213617	GCCTTCCCATAGTCATCAGTCTCCT	TCTGTTTAGCTTTTAGTAAGTCCCG
AMPL213618	TCCGGTCTTTGTGCCCAGACTTCAG	TTTGAACCTCCCGATATCCGGTCCT
AMPL213619	TTGTTTGGAGCTGTCAGACCCTTC	GTCTTCTCCTTGTGGAGCTTGGATG
AMPL213620	TTTTCTTGTAGGAGTCATAATTGCT	GATCTCCGTTTGGGCTGTCCACTG
AMPL213621	GACAGATCTCTGTCGCCTACTGTAG	ATCGTGAGGGACTAAGGTCATAATC
AMPL213623	GGACTGTTAGACAAATAGCTCACCT	TCCTGCTATAGGGACTGACAGATCT
AMPL793541	CTAAATGAGGAAGGACTGTTAGACA	CTTCTTTCGTAGCTGGACGACCGTC
AMPL213656	ACTGGATCTCTTCTTTTACTATGA	TGGACTACAGTAGTATATGGAGGAA
AMPL213657	CTCCGAGCTGGAATTAAGTGGAAG	TCTTGAATATGCAGGACTTCTACTT
AMPL213658	TTTTCTAGGCAAAAATACAGGTGAA	AGATACTGGAATGACGACTGCGTGA
AMPL213659	TGTTACATTCACCAGTTCAGTATCT	CTTCCATCCATCTTTGCTGCAGCA
AMPL213660	TCAAATCTTTAACAAGATGCTTCTC	TTCCAATTTTACACTTCTGGGTAAC
AMPL213661	GTTTCTGTCTCCTTTGGAGTAACAA	CCAACCTTTACTTTCCCTGTATCTGA
AMPL213662	GGTATTGGAGGCAAAGGTGGCAAAG	AGTGCTATGGGTTTAGAGTCTCTTG
AMPL213663	GGGGGCAAAGTTGAAGTACTGGAAG	GAGAAGCAATTGTGGGAAGAGGTGG
AMPL213664	TACTTGAGTCTTCACAGAAACCTGT	CTAAATGCTGGCTGAGAAGGAGGCA
AMPL213665	TGTCATCATCTCCAGGTAATAAGGG	AGAATTTGCCTGAGAGGACACAGCA
AMPL213666	AGTGTTGACAAAGATTAACACAGAA	AAGGAGGCAACGTTGAAGTTTTCAG

AMPL213642	GTGAGTAAGTGACGTGTCCTCTGTT	ATGCCTGGTTACGAAAAGGTTAGAA
AMPL213643	TTTGGTCTCTTTTATATGGTTGCT	TTCACAGGTTTTGAAGGAAGAGTTT
AMPL213644	GGTTATAGAAGAAACATGAGAGATT	ATTGCCTTTGGTTCTGGAGAGTCTG
AMPL213654	TTTATATACTTGCCATAGGTTCCCT	ACAAAATAGTTATTTAAGAAAATGA
AMPL213655	AGAATTTACTTTTGCTAAATCCACT	CCCAATAATGTCAAACCTGTCCACA
AMPL813556	AACTGACGAAGGATTTTGATTTCAC	TGTGAACCAGCAATATATGCCCTTG
AMPL813557	GTCTACCCGTGTAATAATGTGATTCT	ATTGGGAAGCCCTCTTCTCATTGT
AMPL213628	CCATTAAGTCATGGTCCATATACTC	AGTTGACCACAGGTCTCTAGAAAA
AMPL833220	TCCATTAGCTGTTTCATGAACGACT	AAAAGGCACCTACATAAAGACGAAA
AMPL833221	AGGAAAGTGCTTAATCTTATTGGT	TCAGAAAAGTGCACCAACCAGATT
AMPL213667	CAAAACAAACCTTATTTCCACATG	TAGCAGTGGCCTCCAAAATACACCT
AMPL213607	TTCTCTCCTAGCAGTAGTTCTGGA	TAGGTGCCAAATGAAGATGAGACTG
AMPL213608	GAAACAAAGTACAAGTCAACAGTA	CAAAGTAATGACTTTGTTGTGTAA
AMPL232064	GCTGTACCTGATCAGTTCTAGCTGA	AGGTGAGAAGAGAAATCTTAGAGC
AMPL232065	TTAGAACTGGAGGATGAGAATACC	ATTGGCTTGAAAAATAGGCTTCTTT
AMPL213668	CTTCTTCGGTTTCATGGTGTGAAG	AGAAGCAGAAATTATTACCTCCTGA
AMPL232071	CACAGAAACGAGATGTTTATTTTCC	GATAACATCAGGCCACACAGCTGGA
AMPL833218	TTACTAGGATCTAGTGTGAGCATGT	TTTCAAAGGGTCTCACACCTGTAAT
AMPL833219	AAGATTTAGGGTCAACATGGCTA	TCAAGTGCTGCAGAAGGAATGCTAT
AMPL213609	TCTTCGACTACAACACCACCTTTGTC	GGATTAAAAATGGGAAGGCAAAGACT
AMPL213610	GTTCTTCACAGGCTCAGTACTTGTC	CGTTTCTTACTCCACAACCTCATGGC
AMPL820591	AATACGCTTAGCAAAAGCACAAGG	TTCTTTTCGAGAAGTTTGGATGGA
AMPL213635	GGTTTAATAACACTGCCAATTCACT	AACTGGCAACAGCAACTTACAATAT
AMPL213636	TCAGGGCACTGATGGATTGGTTCAG	GTTGTGTGATGTCAGCAAGGCCTGT
AMPL213637	GGGTGCTTCCTTCAAAGACTCCTCT	CATCTCTGGGTTGGAGTGGATGTTA
AMPL794060	TGCTACTGTTGTTTTCTCCAGACT	TCGTCTGTTCTGCTGAAGGCAGGAT
AMPL794061	CCTGAGCTCAATGAGATTCATGACC	CTTGGGTTTTTCATCAGCTGACTCAA
AMPL213645	GCTGGAAGATCGCCTTCAACCAGA	GTGATTATAAACACAAACACACACA
AMPL213646	CAGGCTGGGATAAAAGTTGCAGCAA	CTCTTCTCTGGTGAAGAATGTGAG
AMPL213648	TGTCAATGGCACTGAACCTGTTTC	TCTGCTTCAGGCTGGGATAAAAGTT
AMPL213649	CCTGAGAAGGTCCGTGTTCTTCACCA	TGTTTCCATAAGTCTGTTTGGACG
AMPL213650	CCCACCACCGGGTGTAAACGCTAAG	AAGGATTCTGTCAAGGCTGGACCAG
AMPL213651	CCCCATAGTTTTTGCAATTTGGTCTC	CAAAACGGAGGTCTGGTCCCCTGG
AMPL213652	CCCCTCTTGGTGGGACTCTGTAGG	TACCACAGAATTGCTGCTTCCCTCA
AMPL798605	TAGGAAAGGAATCTTTCAGGACACT	GTTTTCCATAAGCAGAAGACTGAGT
AMPL913	TCCAGCAATTGGGCAGATGTGTGAG	ACAGGTACAGAGCCACAGGACCCCA
AMPL914	AGCCTAGTCCAGGAGAATGAATTGA	TGTTGGACAGTGTAGCACTTACCA
AMPL934	GTTCCAGTAGTCTACTTTGACACT	GTCCTTTTGCATGGGTTTGGTTTAT
AMPL920	GTATGCAGATTACTGCAGTGATTTT	TCAGAGAGGGAGGACACAATATTCT
AMPL916	TCTCTCTTCTCTCTTCTTCCAGAT	ATAAGCAAGATCCCACGATGGGGGT
AMPL917	TCTTCTTTTTTGAAAGTCTGTTTTT	GCTAGAAATCTGTTGCTATGGGCCC
AMPL127631	AGGTCCAAAGCGAGCAAGAGAATCC	GTGCTAGATACTTTCACACAGGTTC
AMPL555150	TCTTATCCTGATGGGTTGTGTTTGG	GTAAAGCTCCCTCCCTCAAGTTGAC

AMPL973	AAAGAGCACGTTCTTCTGCTGTATG	TGTATTTGCTTCCTTCCATCAATGC
AMPL127628	GTGAACGGACACTGAAATATTTTCT	TGCTGATGCTGAGTCTGAGTTACCA
AMPL32187	TCCTGATTTTGTCTTCAACTTCTAA	GAGGAAAATGGGTAGTTAGCTATTT
AMPL925	CTAGAGGTAACCTCATGATAATGGAA	CAGAATACCTTGCATCTGCTGCATA
AMPL909	AAAGTGTGAGCAGGGAGAAGCCAGA	CATTCTGGAAGAGTTTTATGTAGG
AMPL910	ACCATCTTCAACCTCTGCATTGAAA	CAACAGAAAGGGTCAACAAAAGAAT
AMPL911	TCTGGAATCAGCCTCTTCTCTGATG	GAAAGTTGCAGAATCTGCCCAGAGT
AMPL912	TGTAATTCAACATTCATCGTTGTGT	GATCCTTCTGAAGACAGAGCCCCAG
AMPL932	ATAGAACTACCCATCTCAAGAGGA	TGCTGTATTGGAACAAACACTTTGA
AMPL555255	TGGCGATGGTTTTCTCCTTCCATTT	TTGTTGATGTGGAGGAGCAACAGCT
AMPL923	TAAGCCAGAATCCAGAAGGCCTTTC	AAGATGCTGTGGTATCTGACATCTT
AMPL924	ATTATCACTATCAGAACAAAGCAGT	TTGAGGTGTCTGCAGATAGTTCTAC
AMPL189162	ACCCGTTCCCTTGATGTCTACAATT	AAAGATGGTGGTTCTGGCCAAAAGG
AMPL930	CTAGAAGCTGTGTTAGAACAGCATG	TGATATCTTAAGCAAAATTCTTTCC
AMPL349810	GTATTTCATTTTCTTGGTGCCATTT	TCTAACAGCTACCCCTTCCATCATAA
AMPL935	GCTCAGGGCTATCCTCTCAGAGTGA	TGCTGACATCATCTCTTTGAATTAA
AMPL555015	CTTTTACATCTGAACCTCTGTTTTT	CTCAGGTAAAAAGCGTGTGTGTGTG
AMPL127624	AGTTGGTCTGAGTGACAAGGAATTG	AGTGAGCTAAATGTTTATGCTTTTG
AMPL127625	CAGTGCAGTGAATTGGAAGACTTGA	TGAAGAAAGAGGAACGGGCTTGAA
AMPL127626	GAAGAATAGCTTAAATGACTGCAGT	AACCCCAGGATCCTTTCTTGATTG
AMPL127627	CAACACTTGTTATTTGGTAAAGTAA	TATTGGCAAAGGCATCTCAGGAACA
AMPL941	CCCATACACATTTGGCTCAGGGTTA	TCTCAGTCTACTAGGCATAGCACCG
AMPL942	GATACTAGTTTTGCTGAAAATGACA	CCAAGAAATTAGAGTCCTCAGAAGA
AMPL943	TAACTTAGAACAGCCTATGGGAAGT	TCTGCTGTTTTTAGCAAAAGCGTCC
AMPL944	CTTCCTGGAAGTAATTGTAAGCATC	TCAGGTTTGTCTGAGACACCTGAT
AMPL945	GTTCCAGTGATGAAAACATTCAAGC	AGCAAGAATATGAAGAAGTAGTTCA
AMPL946	AATTAGCCGTAATAACATTAGAGAA	AGAAACAGAGGGCCAAAATTGAATG
AMPL947	CATTCAATGTCACCTGAAAGAGAAA	AGCAATATTAATGAAGTAGTTCCA
AMPL948	TAAACATGGACTTTTACAAAACCCA	AACATTCCAAGTACAGTGAGCACAA
AMPL949	GAAAGATAAGCCAGTTGATAATGCC	ACCACTTTTTCCCATCAAGTCATTT
AMPL950	GTCACTTTTGAATGTGAACAAAAGG	CAAAGGAGGCTCTAGGTTTTGTCTA
AMPL951	AAGGTTTCAAAGCGCCAGTCATTTG	GGAAAGAATGAGTCTAATATCAAGC
AMPL952	CACAGAAGGCTTTAAGTATCCATTG	AATCCAGGAAATGCAGAAGAGGAAT
AMPL953	GGAAGGCAAAAACAGAACCAATAA	TAACCACAGTCGGGAAACAAGCATA
AMPL954	AAGATCTGTAGAGAGTAGCAGTATT	AGTGTGCAGCATTTGAAAACCCCAA
AMPL955	CTTCCAAGAGAAGAAAAAGAAGAGA	TGGTACTGATTATGGCACTCAGGAA
AMPL956	TGGTCTTTTACTAAGTGTCAAAT	GTAAAGTGTCTAATAATGCTGAAG
AMPL958	CAGAAACCTACAACCTCATGGAAGGT	TTTCCCAGAGCTGAAGTTAACAAT
AMPL959	TGATAGTTGTTCTAGCAGTGAAGAG	AACTGGAGCCAAGAAGAGTAACAAG
AMPL960	GAGGAAGTCTTCTACCAGGCATATT	AAAGTACAACCAATGCCAGTCAGG
AMPL961	AGAATCTGCTTTCAAACGAAAGCT	ACTAGTAGTCAGTAGAAATCTAAGC
AMPL962	ACGGAGCAGAATGGTCAAGTGATGA	CAGCAGTATAAGCAATATGGAACCTC
AMPL963	TCAGGCCTTCATCTGAGGATTTTA	AGTGGTCATGAGATAAAACAAAAG

AMPL964	AACTTAAGCCATGTAACTGAAAATC	GATTGGCAGTTCAAAGACTCCTG
AMPL965	TCACTCCAAATCAGTAGAGAGTAAT	GCATTTGTTACTGAGCCACAGATAA
AMPL966	TAGATGAATATTCTGGTTCTTCAGA	AATATTTGGGAAAACCTATCGGAAG
AMPL967	AGTTAATGAGTGGTTTTCCAGAAGT	TACTGGCCAGTGATCCTCATGAGGC
AMPL968	TGTGAGAGAAAAGAATGGAATAAGC	AGGTTCTGATGACTCACATGATGGG
AMPL969	CCAACATAACAGATGGGCTGGAAGT	TGCTCAGAGAATCCTAGAGATACTG
AMPL970	AGACAGAATGAATGTAGAAAAGGCT	TAATGATAGGCGGACTCCCAGCACA
AMPL971	GTTCTGTTTCAAACCTGTCATGTGGA	TAAAAGCAAACAGCCTGGCTTAGCA
AMPL972	TAACAAATACTGAACATCATCAACC	CAAATACTCATGCCAGCTCATTACA
AMPL919	CCAAGGTGTATGAAGTATGATTTTT	ATTTGAACACCACTGAGAAGCGTGC
AMPL929	TGCATACATGTAAGTAGTGTCTT	CCAACCTAACAGGCACTGAAAAGAG
AMPL127643	TACCGTTAATAAGGCAACTTATTGC	AGGGATGTATGTGCAGTTTGTATA
AMPL974	ATCTGTTTTTCAACAAGTACATTTT	AGAGAACCTTTGTCTATGAAGCTGG
AMPL127629	CTGTGAGAACTCTGAGACAAAGCA	ATTTGTGGTAGGCAGCTTTGGGAAG
AMPL127630	CTTCAGGAGGAAAAGCACAGAACTG	CTCAAAAGACGTCTGTCTACATTGA
AMPL554992	CAAAGTATGGGCTACAGAAACCGTG	AGGCTAGAAGCAGTCTCTCGCCTT
AMPL554993	ACTGCAAACATAATGTTTTCCCTTG	CTACAGAGTGAACCGAAAATCCTT
AMPL928	TAATCACTTGCTGAGTGTGTTCTC	AGAATGACACTCAAGTGCTGTCCAT
AMPL906	ATTTTGCATGCTGAAACTTCTCAAC	AGCAACCACAGTAGGAAAAAGTAGA
AMPL907	TTGCCTTTTGAGTATTCTTTCTACA	AAGAAAGGGCCTTCACAGTGCCTT
AMPL921	CTGGAGTTGATCAAGGAACCTGTCT	AACCCAGGAAACACCTAACCTTATA
AMPL922	CATAACAGCTCAAAGTTGAACTTAT	ACCACATATTTTGCAAGTAAGTTTG
AMPL202863	TCTAATGTGTTAAAGTTCAATGGAA	ATTTGGGATTCCATGATTATCTCC
AMPL222544	CCGCTGGAGACAGCGGAAACTCAC	TCAGACGTAAAGCGGAAGGGGCCGG
AMPL222545	CAGGAGTCGTTACCTTTGCTAAGCT	CATTTCAAAGCGGAACGTCTTCCCG
AMPL193550	GAGGAAGCAGAATCTAACGGAGACT	ACCTCTAGGAGTTCTCTCAGAGTCT
AMPL222550	AGACATTCTCTTCTGATAATTGCA	CAACAGTGATGCTAGGAGTGGAGAA
AMPL222551	AGTGCTGAACAGAAGGTGATTATGA	AAGGCTTCTGCTTTAGATATCCCAA
AMPL222552	ACCTTAAAAGGAGAACACTTTATT	ATGCTCCTGCTCAAGAAGTTCCAGT
AMPL222553	TACACACTTAAATTAAGAGCACAT	GTGTTTTTCCAACACCTGGTGCACC
AMPL623288	AGGCAGTAGCAAGGTCTACCACTCT	ACCAATCATGATGATGATCTAAGG
AMPL222562	GGCACAAATGCAAATTACTTTTATT	AACCTCCCTCTGTATCAATAAAAAC
AMPL222542	AATCTGGAAGAAGATAAACTTGTGC	ACCTGTTTGGATGTATTGGCAATTG
AMPL222543	TTGATTTTATGCTACTGTACTGGTT	TGTGTAGTCACGACAGCGAAAATAA
AMPL115338	CATCTAGGTCATGACGAAATGGAAA	ATGCTCTCTCTCTCCAAGAGAGCAG
AMPL222554	TATTTGCAAGGCTGATCATTTGCTG	CCACTATCACTAGTCGAACCTAAAT
AMPL222555	CGTATTTTGACATTTCTAGACATT	TTAATAACCGAGTACGAAGAGACAG
AMPL222558	GATTAATTACCCACCTAATGCAGGA	ATCAAGTAGACCAGTCTGGCCAAGA
AMPL222559	AGCGTATAGTACTCAATGAGAATCA	AAGGCCTGATTTCTATCAATCTTTG
AMPL222548	TTCTGTACTGACCTTTGCTTTCGGT	TCTCAGAACGCCTCTGATCAAATAT
AMPL222549	AAGAAACCTGTAAGTGTTCATTGAG	ATTAGCCGTATTGTAGCAGCATGTC
AMPL222556	CAGTACTGTGCATTCTTCTGGCTG	AAGTTCAGATGACCAAACTCATTAT
AMPL222557	GAAGCAAGTTGTAAATTCTCTCTTT	CAATGTTGCCAACCTGCTTAAAGAA

AMPL222546	CCTTCTGTTTGCAATGAACATGCAG	TACTTTATTCTAGGGCCAGGCGCAG
AMPL222547	GTACACATTGATTTCACAACATCAA	GTATCTCTAAATCCCTGAGGCTGAA
AMPL623307	TTGATGGTAATGTAAAAACAACCTG	CTCTCAAGCTAAGTAAAAATATGTC
AMPL222724	AAGTAGATTTCAGCCAGAGAAGTGAA	TGGAAGTGGGAAAAACTCATGAAAT
AMPL222725	ACCTAGCTGAAACTGATAGAGGAAA	GCATAGATACAAAGTTGAATGGAAT
AMPL222726	TGAAACAGAAGCAGAAGATGAATCT	ACAATTCAGATTGCATTTTAGCTAA
AMPL518797	ACATCGTTTGGATCATGCCCTCAAT	AGTGAAGAAGATAAACAGTCCACTT
AMPL518798	ATCAGAGAATAGTGCCCTCTAGTCCT	ATTTTCATCATTAAGATTGATGCCA
AMPL518799	AGCCCAACTTTCAACAAACAAACAA	ACAGAGAAGATGGAAAGTAAACTG
AMPL518800	CTAGGAGCAGAAAGTTAAATCTCTT	TGGTCAAGCTTTAATCTTTGGGAC
AMPL232665	ATATGTGTCCAGGAACACAGTGTC	TCCCATTTACCTTTGATAGAGAAGT
AMPL623304	AGTGACCTCTTTAAAGTACAGTACC	ACCAAAAATTACCTCTACCAAGTA
AMPL623305	AATTTTCCAAAAAGCATCAAAAGT	ACTGGAAGCAGCAAGTCATCTATCA
AMPL1015529	AACCAAATTTCCCTGGGTTACCT	CCATAAAGGACAGAACCAATATACA
AMPL222703	TTTAGGAATAACCCAAGTCGCTATA	ACATGAGCTCTACTATCTTATTAAG
AMPL222704	GAAAACTGGATTTATTTCCGATGTT	ATCTCAGCTGGGCTTAGAGTAAAAA
AMPL222720	TGAAATTCAAGCATACAGGCCTTA	ACTGGAACCATAGGAAGTGGTATGT
AMPL222721	GAACTAAAACGACAATACAATGACC	CCTTGGTAGGTAAGTGGAGTGAAAT
AMPL623292	CCGTGCTGTCATAACAATAGGAATT	TTGCGCTTTGCTTTTATAGTGCTTT
AMPL623293	TTTCTGAGAATTTTGGATATGCTTT	TGTGAAAGATCTACAGGTAGGCCAT
AMPL222705	GACGCAATCAAATACAAAGGAGAGA	TCCTGAGTGAAAAGAAAATAAATTG
AMPL222706	TTATGGCATAATCTGGAGTTGGTGA	TGGAATATTATTTTCATGGTCGGTGAA
AMPL222707	TTAAGTTTGTTTAGCTAGACACACA	GTAGAACCACAGGGAGGAGAAAAAA
AMPL222694	TTGTTATCTGTGTGCCAGACTGTGA	TTCTAGGGTAGAGGCAACAAGGTA
AMPL222695	ATCTCTGTGCTACCTTCCAGAATAC	TTGTGTTTCTTGCCATCTTACAAGG
AMPL623297	TTTCAGGCATAACTCTAATAATGAC	AGTTCCAAGATGAAGTGGGAGCACT
AMPL222680	TTCTGTCAGAACTTGGTGTTACATTT	TGAAGTGCCAGCAAGTAAATCTTGG
AMPL222681	GTTTTCTTTTAGGCCTTTTTCAGATA	GGAGGCTAATCATATCATTAATAAAT
AMPL222682	CAGACTCCTAGACTCAAGCGATCTT	GTTTCAGACCATTGTTTTGACATCTG
AMPL1015530	TTACTCCTGGACAAATCAGATTGAT	TGTAAATGGAAATTTTGTCTTCTTG
AMPL1015531	TAGGTTTTAAAAGTTTTGCCACTTA	AAATGGGTTGTTGGTTCTACCAAAA
AMPL222683	AACTCAAATAATGCTTAAAGGACTT	TGGATGCTAGCATGTGTTTAATACA
AMPL222684	TGCTGTTCTTCAAAAAGAGGAAAAA	TGACTATCTTTTTAGGCAAAATAGC
AMPL222685	TTTTGTTGGGTTTCATTGTATTTTT	TTATGGTAAAGAGGAGGCAAGAGAA
AMPL222691	GGAAATGAAATGCTCTTAACCTTTAC	TCCAGATTAAATCTAGTGAGTAACC
AMPL623296	TGATACTGGTTGACACAATTGTTTT	ATCACCCTGCTACTTTTCCCATTT
AMPL222677	TCAGCAAGTTACAGTGTAACAGAAG	TGTCATTAGTGTATTTGTGCTGCC
AMPL222678	TTAAATCTGAAAGAACAGGTGTCA	TTTGCTCGGGATGAACTAGATAGTA
AMPL222679	TGTTGTAAATCTATTTTGCCAAAT	GCTCATAACATCGAGGACTGTGCTC
AMPL623300	AAGATGCTGACATCATATTTTGTC	TCCTACTCATGGAGATGTAAATGTT
AMPL623301	TACAGACTTTCCAAGGATGTGCAA	TTCTAGATGCACAAATAAGGGAAAG
AMPL623302	TATTGGTTCACATTTTCAGTTATGCA	TAGAAGAACTTGTGAGCCTGGGGAA
AMPL222708	TCCTGAGGTAGTCGGTAACTTCAAC	TTTTAGTGATAATCAACTTCGAAA

AMPL222709	ATATTTTGGGACACGCACACACAAG	GAAGTGCATGGAATTGCTAGATGGG
AMPL222710	AAACAAGGAAACAGTCAAGAGTCAT	GATTACTAGAGAGCTCCGGAGGACG
AMPL222711	CCATGTGAGGTTTGATAACGTTTTT	AAGAAGGATCATAACAGGGAAATCCA
AMPL222699	AGTTGATTCAGGAAAGACTGTAAAA	ACCAATTTTGGACAAAAGAGGAACA
AMPL222700	CTAGCCAGTTAACTTCTAGTTCAC	ACTGGAAGATAAACTCCTTTTCG
AMPL222697	CTCTGGCTGCAAAGTTATCTGCTAA	ACAGCTGAACATGGTAGTCAAGTAT
AMPL222698	TAGAAAAATTGCCTACCTGTAAGTT	CCATATACAGAGATGAAAATGATGA
AMPL222687	TACAATTGTCATGTTGTTGTGCATG	TTTGCCTGGGTCAGTTGATATCTGT
AMPL222688	TGTAGAGAATGATTGATGTTTATTT	ATTTTACAAACAATGACATGAACCA
AMPL222689	CCACAGGAAGTGGAAGCTTAGC	TATCCAGACTGGTTCAGCCATGCTG
AMPL222690	AGATGGAGAGAGACCTTTATAAAGT	CTGCTTTAGCATGGCAACAATCTCT
AMPL222701	GAAGATTTACTTTCCTTATAAAGCT	TGCATTCATATTCTCAGTAAATGGA
AMPL222702	TCTCAATATTTGACATTTTCATTC	TCACAGCTTGCTATGATGAATTCTG

Note: *ULSO* stands for Upstream Locus-Specific Oligo and *DLSO* stands for Downstream Locus-Specific Oligo

Appendix 6. Indexes for pooling guide of Ampliseq sequencing

I7 index

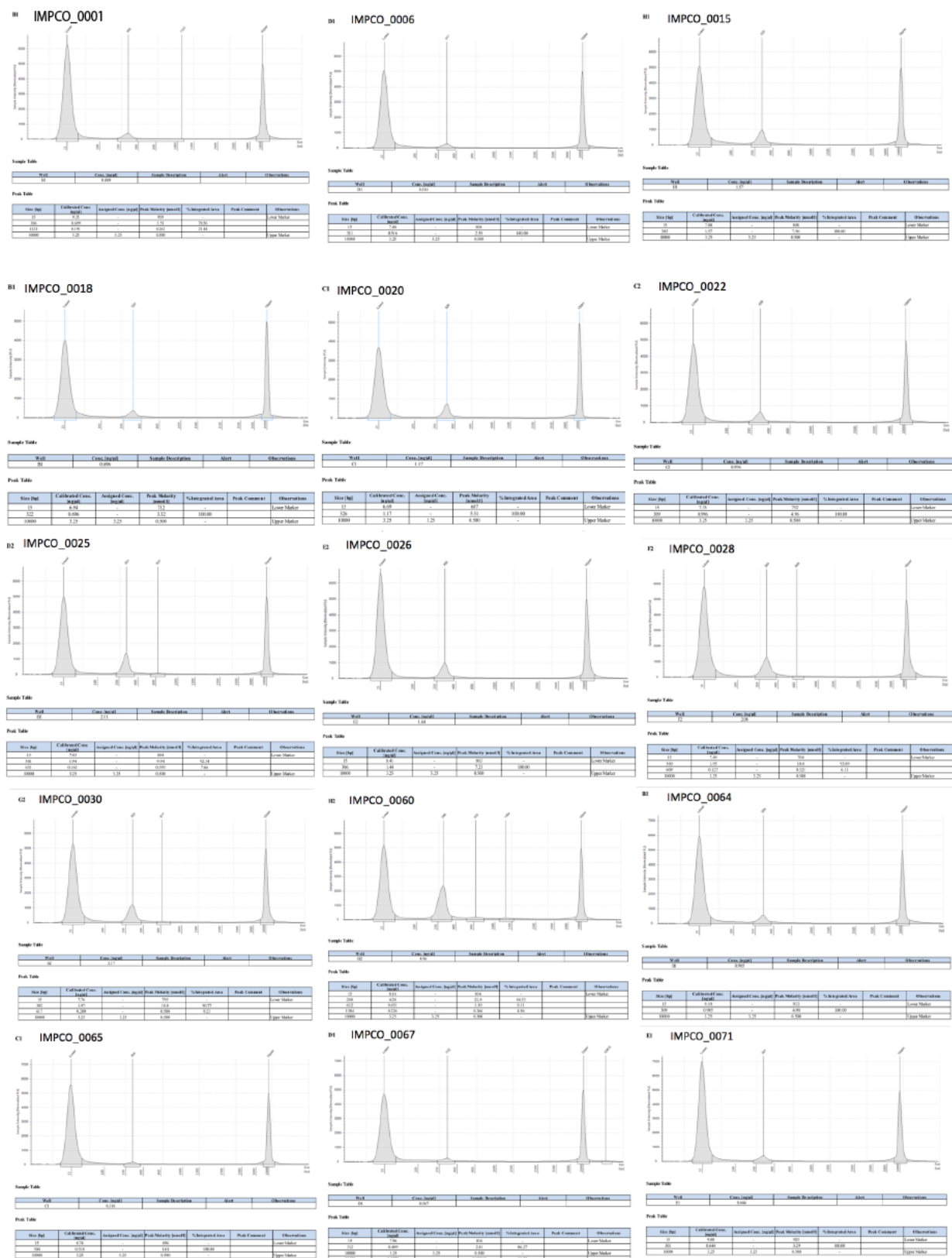
i7 Index Name	i7 Bases for Sample Sheet
Q7005	GTGAATAT
Q7006	ACAGGCGC
Q7007	CATAGAGT
Q7008	TGCGAGAC
Q7015	TCTCTACT
Q7016	CTCTCGTC
Q7017	CCAAGTCT
Q7018	TTGGACTC
Q7023	GCAGAATT
Q7024	ATGAGGCC
Q7025	ACTAAGAT
Q7026	GTCGGAGC
Q7027	AGCCTCAT
Q7028	GATTCTGC
Q7029	TCGTAGTG
Q7030	CTACGACA
Q7035	ATGGCATG
Q7036	GCAATGCA
Q7039	CTTATCGG
Q7040	TCCGCTAA
Q7041	GATCTATC
Q7042	AGCTCGCT
Q7047	ACACTAAG
Q7048	GTGTCGGA

I5 index

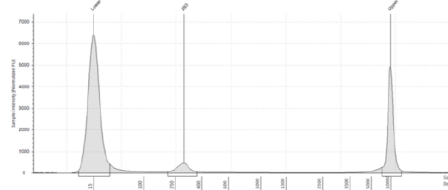
I5 Index Name	I5 Bases for Sample Sheet NovaSeq, MiSeq, HiSeq 2000/2500	I5 Bases for Sample Sheet iSeq, HiSeq 3000/4000, MiniSeq, NextSeq
Q5001	AGCGCTAG	CTAGCGCT
Q5002	GATATCGA	TCGATATC
Q5003	CGCAGACG	CGTCTGCG
Q5004	TATGAGTA	TACTCATA
Q5007	ACATAGCG	CGCTATGT
Q5008	GTGCGATA	TATCGCAC
Q5009	CCAACAGA	TCTGTTGG
Q5010	TTGGTGAG	CTCACCAA
Q5013	AACCGCGG	CCGCGGTT
Q5014	GGTTATAA	TTATAACC
Q5017	CTAGCTTG	CAAGCTAG
Q5018	TCGATCCA	TGGATCGA

Q5025	ATACCAAG	CTTGGTAT
Q5026	GCGTTGGA	TCCAACGC
Q5027	CTTCACGG	CCGTGAAG
Q5028	TCCTGTAA	TTACAGGA
Q5029	CCTCGGTA	TACCGAGG
Q5030	TTCTAACG	CGTTAGAA
Q5031	CGCTCGTG	CACGAGCG
Q5032	TATCTACA	TGTAGATA
Q5035	CATTGTTG	CAACAATG
Q5036	TGCCACCA	TGGTGGCA
Q5039	ACGCCGCA	TGCGGCGT
Q5040	GTATTATG	CATAATAC

Appendix 7. Tapestation results of Ampliseq libraries preparation from early stage samples



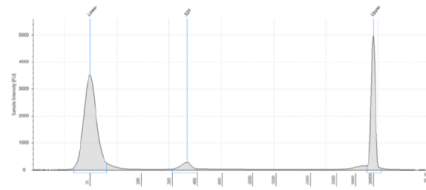
F1 IMPCO_0072



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.03	-	720	-	-	Lower Marker
220	0.786	-	0.13	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

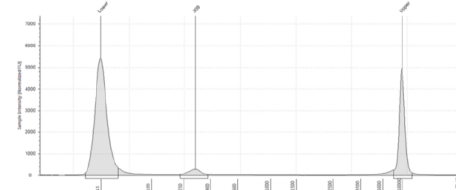
D1 IMPCO_0073



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	-	-	-	Lower Marker
120	0.663	-	2.33	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

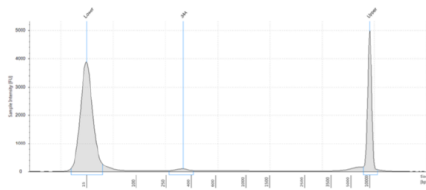
A2 IMPCO_0076



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	720	-	-	Lower Marker
100	0.514	-	2.30	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

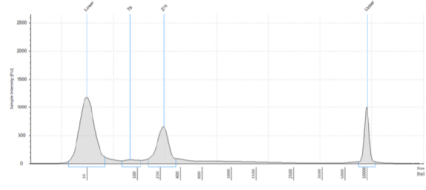
D1 IMPCO_0077



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	720	-	-	Lower Marker
344	0.317	-	0.072	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

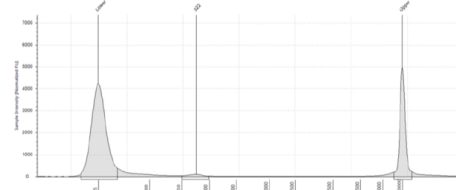
B1 IMPCO_0078



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	7200	-	-	Lower Marker
270	0.315	-	1800	89.12	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

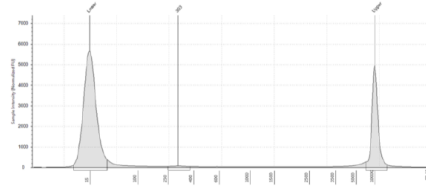
D2 IMPCO_0079



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	720	-	-	Lower Marker
221	0.214	-	1.42	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

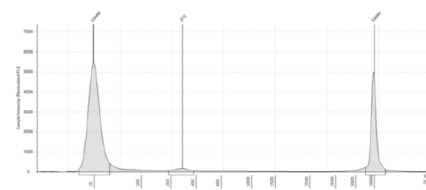
E2 IMPCO_0080



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	816	-	-	Lower Marker
310	0.313	-	0.533	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

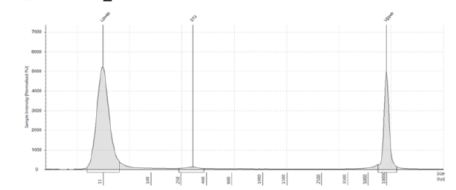
F2 IMPCO_0081



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	7200	-	-	Lower Marker
312	0.314	-	0.50	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

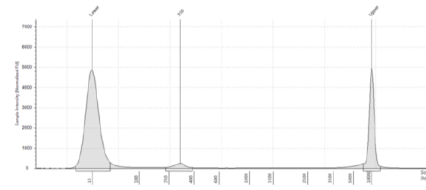
G2 IMPCO_0083



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	816	-	-	Lower Marker
313	0.273	-	1.34	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

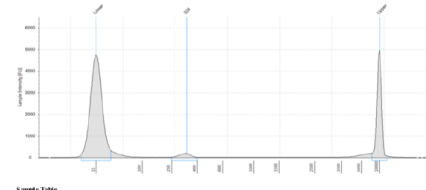
H2 IMPCO_0084



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	804	-	-	Lower Marker
314	0.080	-	1.06	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

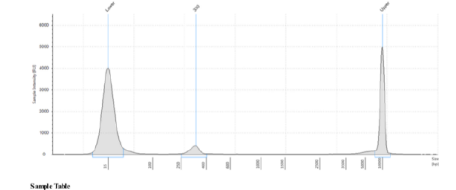
H2 IMPCO_0085



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	720	-	-	Lower Marker
320	0.373	-	1.74	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

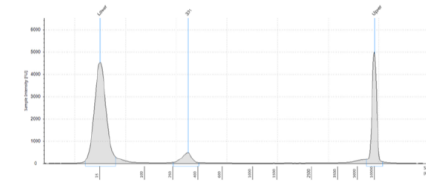
C2 IMPCO_0086



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	804	-	-	Lower Marker
330	0.057	-	2.93	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

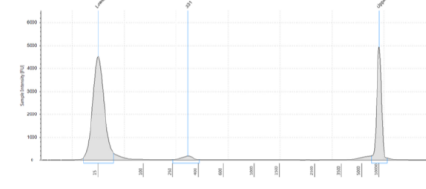
D2 IMPCO_0087



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	720	-	-	Lower Marker
235	0.778	-	3.40	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

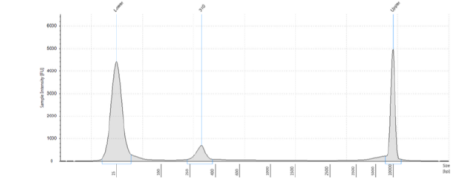
E2 IMPCO_0088



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	724	-	-	Lower Marker
331	0.344	-	1.60	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

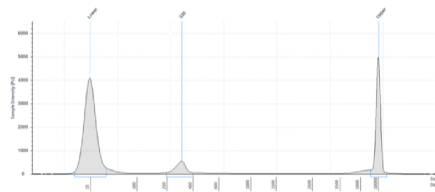
F2 IMPCO_0091



Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Molarity (mM)	% Integrated Area	Peak Comment	Observations
15	0.00	-	724	-	-	Lower Marker
340	0.09	-	5.43	100.00	-	-
10000	3.25	3.25	0.500	-	-	Upper Marker

H2 IMPCO_0092



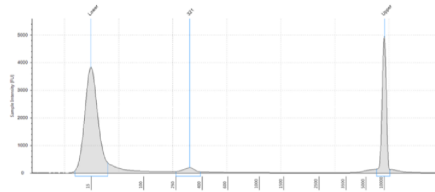
Sample Table

Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Peak Table

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Mobility (mmol/L)	% Integrated Area	Peak Comment	Observations
15	6.50	-	7.01	100.00	-	Lower Marker
320	0.921	-	4.34	-	-	-
10000	3.25	3.25	0.900	-	-	Upper Marker

G1 IMPCO_0094



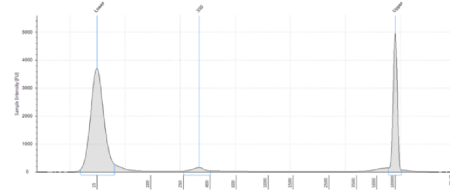
Sample Table

Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Peak Table

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Mobility (mmol/L)	% Integrated Area	Peak Comment	Observations
15	7.23	-	7.52	-	-	Lower Marker
320	0.418	-	2.00	100.00	-	-
10000	3.25	3.25	0.900	-	-	Upper Marker

H1 IMPCO_0095



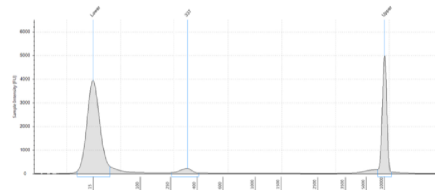
Sample Table

Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Peak Table

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Mobility (mmol/L)	% Integrated Area	Peak Comment	Observations
15	7.50	-	7.54	-	-	Lower Marker
320	0.752	-	1.04	100.00	-	-
10000	3.25	3.25	0.900	-	-	Upper Marker

G2 IMPCO_0103



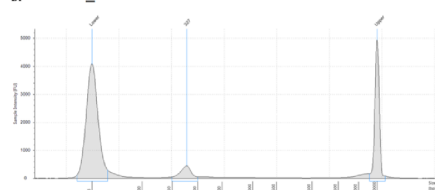
Sample Table

Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Peak Table

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Mobility (mmol/L)	% Integrated Area	Peak Comment	Observations
15	6.51	-	6.98	-	-	Lower Marker
327	0.420	-	1.92	100.00	-	-
10000	3.25	3.25	0.900	-	-	Upper Marker

B1 IMPCO_0062



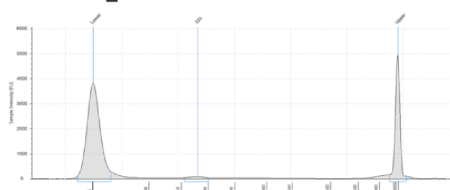
Sample Table

Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Peak Table

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Mobility (mmol/L)	% Integrated Area	Peak Comment	Observations
15	6.51	-	7.08	-	-	Lower Marker
327	0.762	-	3.53	100.00	-	-
10000	3.25	3.25	0.900	-	-	Upper Marker

E1 IMPCO_0106



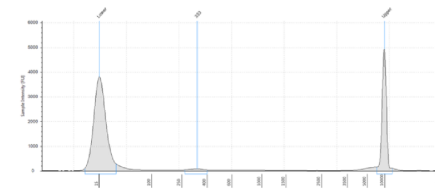
Sample Table

Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Peak Table

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Mobility (mmol/L)	% Integrated Area	Peak Comment	Observations
15	6.51	-	7.01	-	-	Lower Marker
327	0.206	-	0.954	100.00	-	-
10000	3.25	3.25	0.900	-	-	Upper Marker

E1 IMPCO_0106



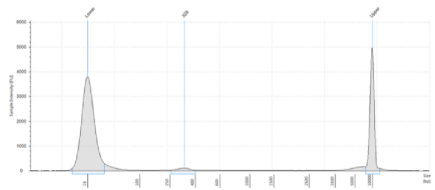
Sample Table

Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

Peak Table

Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Mobility (mmol/L)	% Integrated Area	Peak Comment	Observations
15	6.51	-	7.01	-	-	Lower Marker
327	0.206	-	0.954	100.00	-	-
10000	3.25	3.25	0.900	-	-	Upper Marker

F1 IMPCO_0107



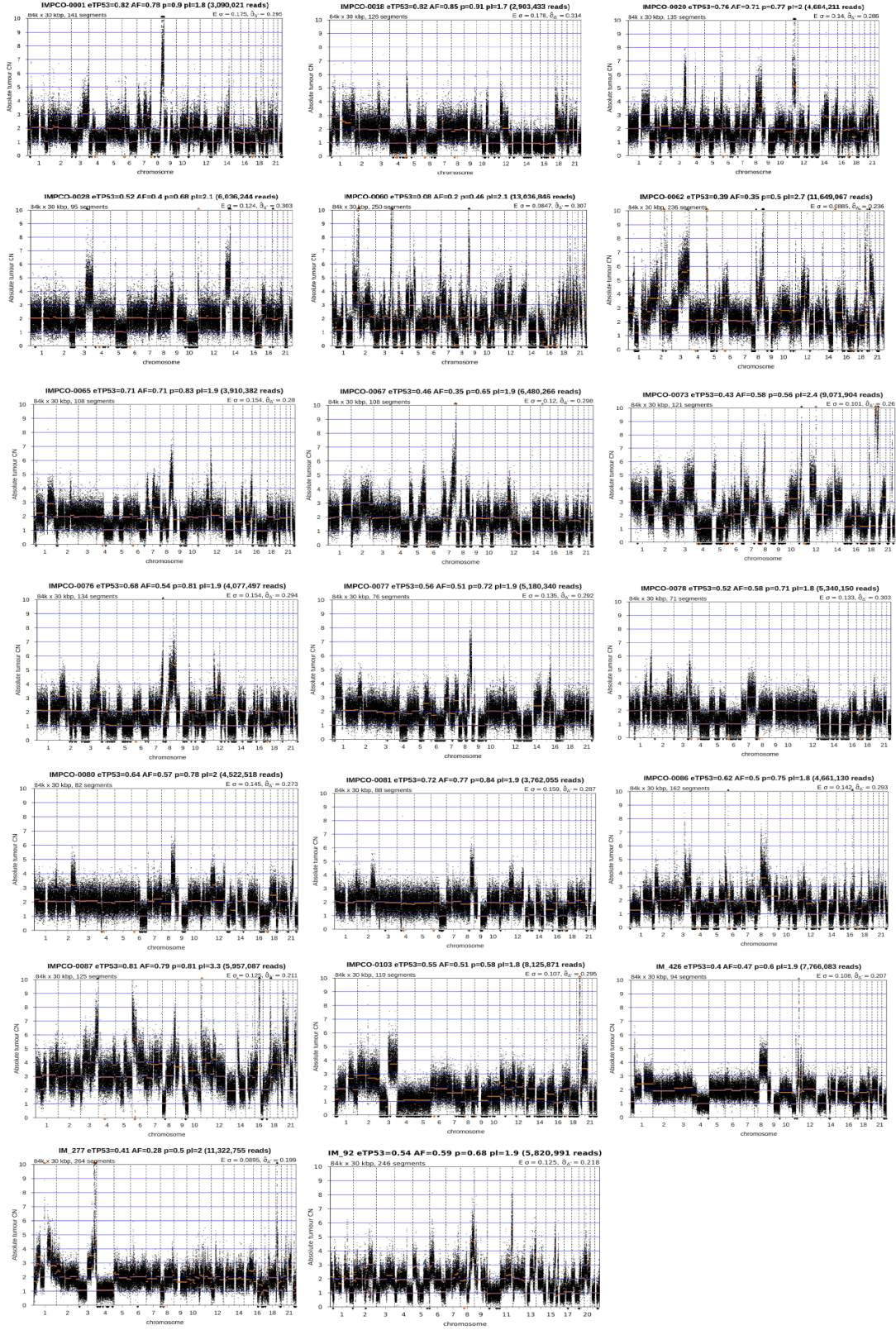
Sample Table

Well	Conc. (ng/mL)	Sample Description	Alert	Observations
------	---------------	--------------------	-------	--------------

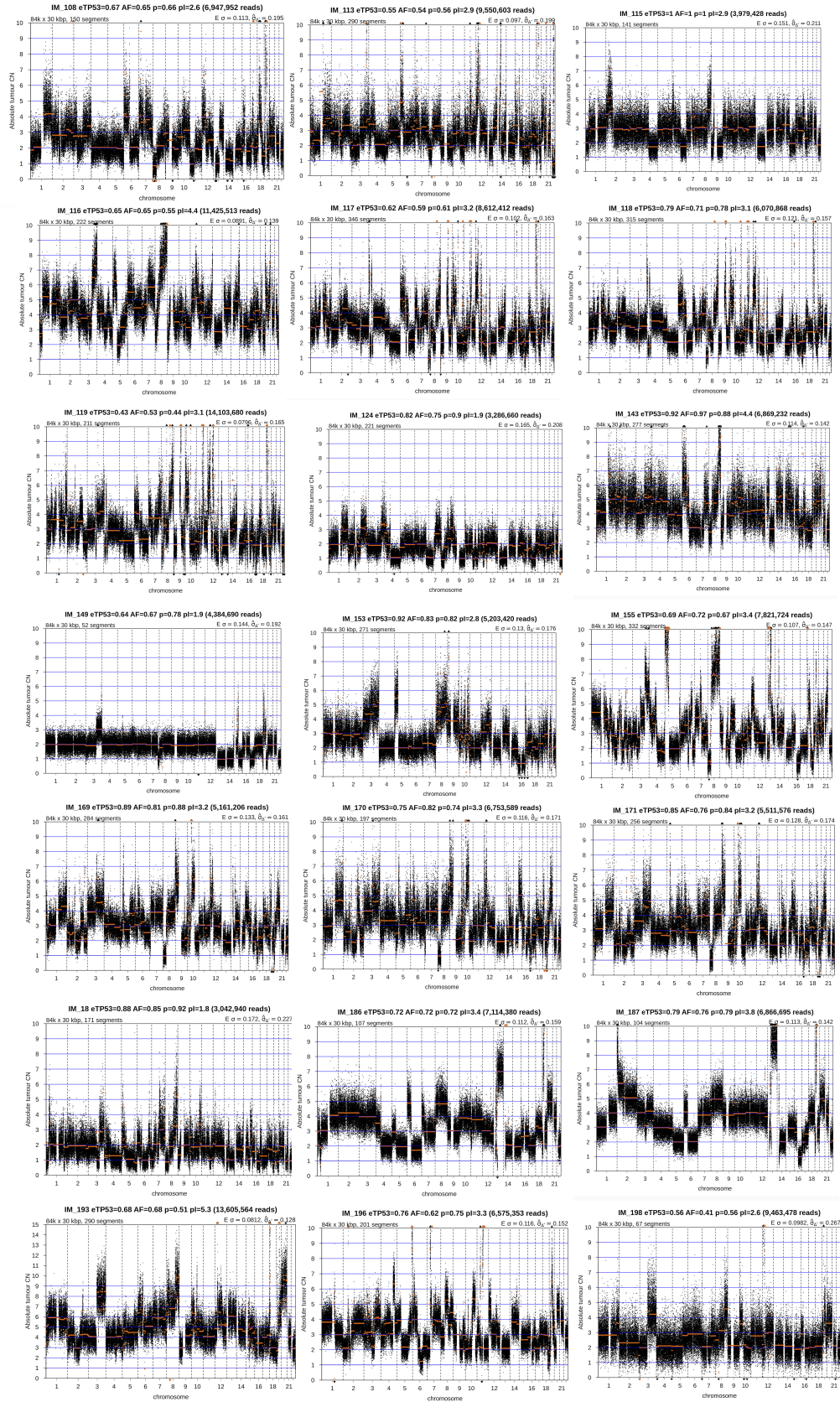
Peak Table

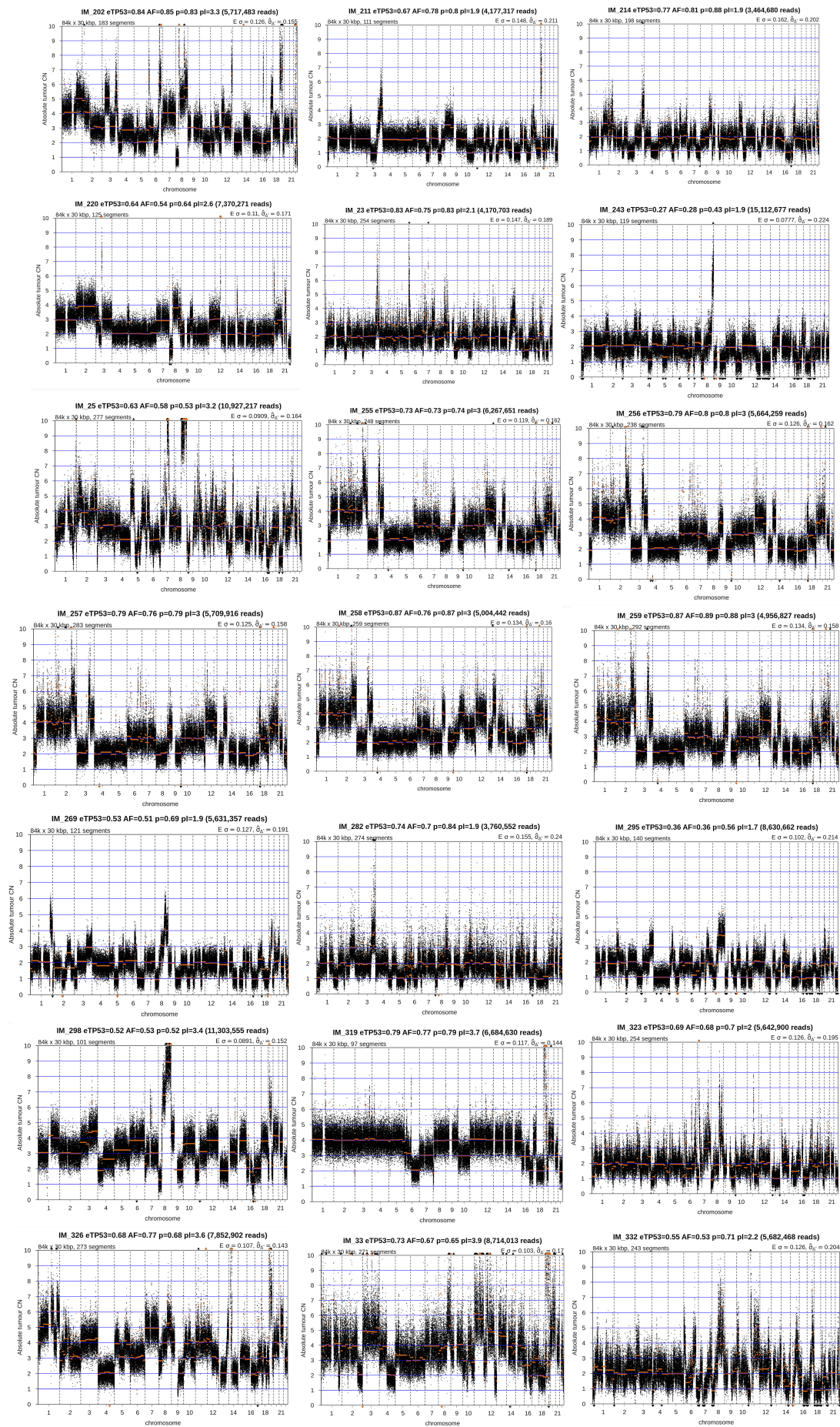
Size (bp)	Calibrated Conc. (ng/mL)	Assigned Conc. (ng/mL)	Peak Mobility (mmol/L)	% Integrated Area	Peak Comment	Observations
15	7.07	-	7.23	-	-	Lower Marker
320	0.252	-	1.10	100.00	-	-
10000	3.25	3.25	0.900	-	-	Upper Marker

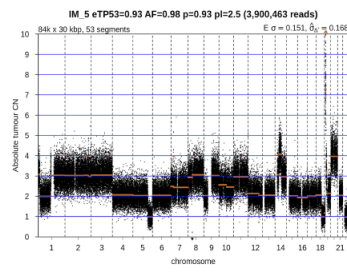
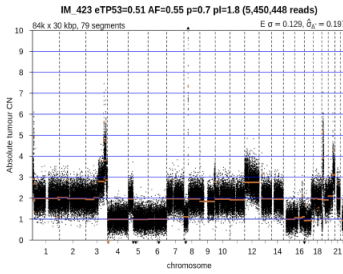
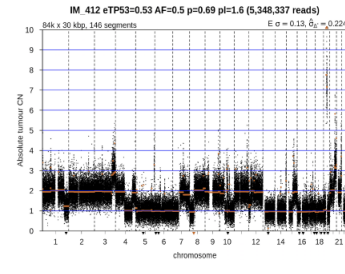
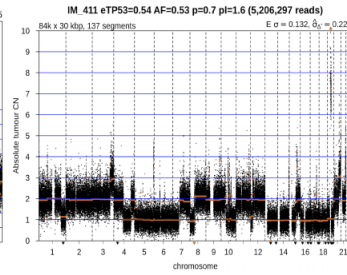
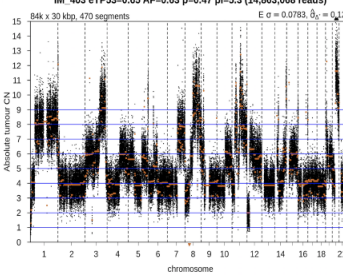
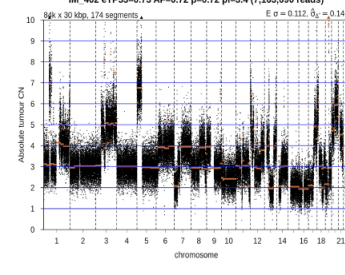
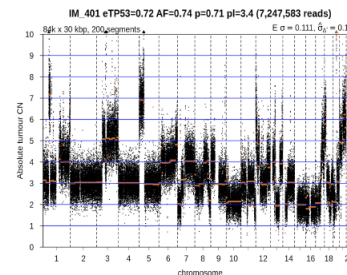
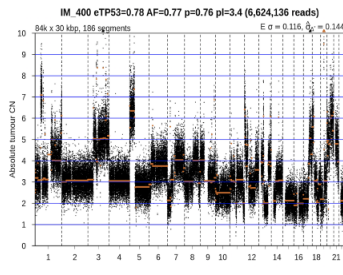
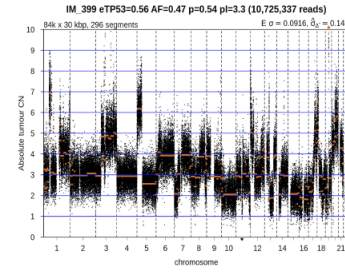
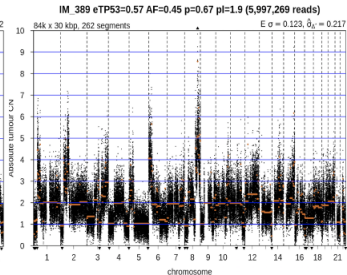
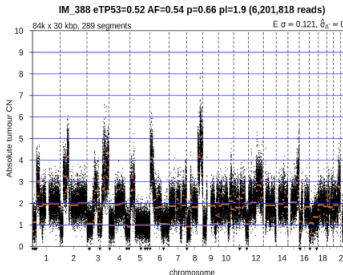
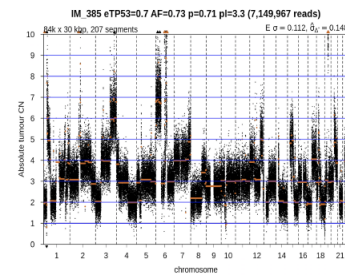
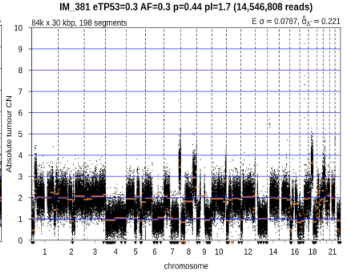
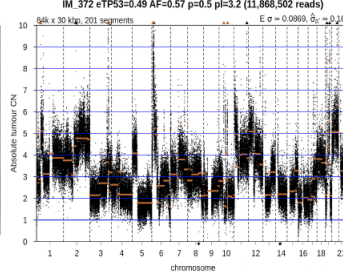
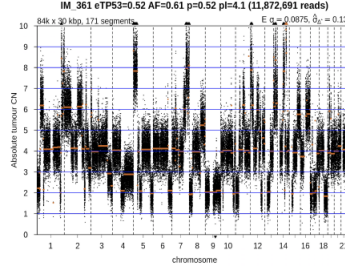
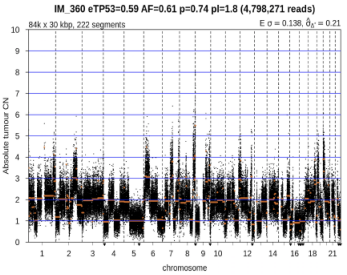
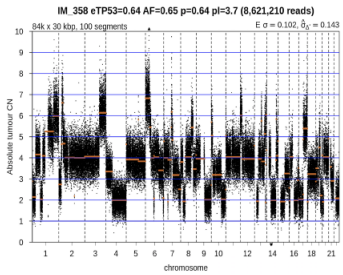
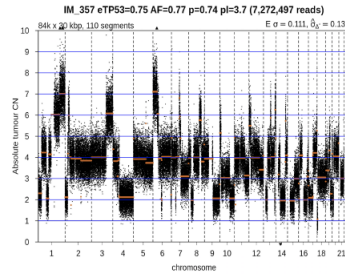
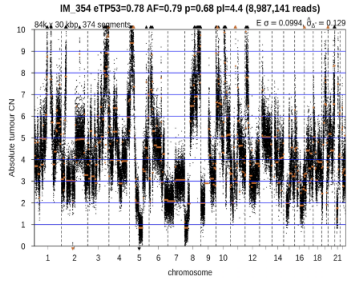
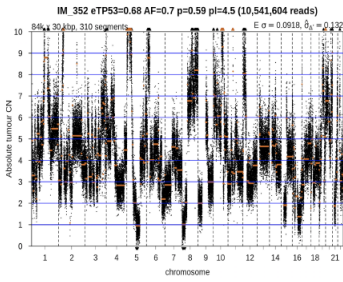
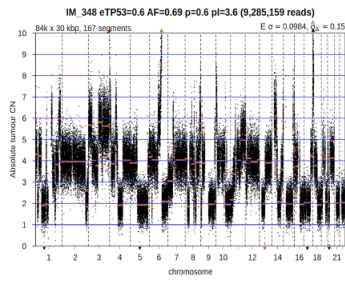
Appendix 8. Absolute copy number profiles from swgs-absn pipeline in early stage cohort

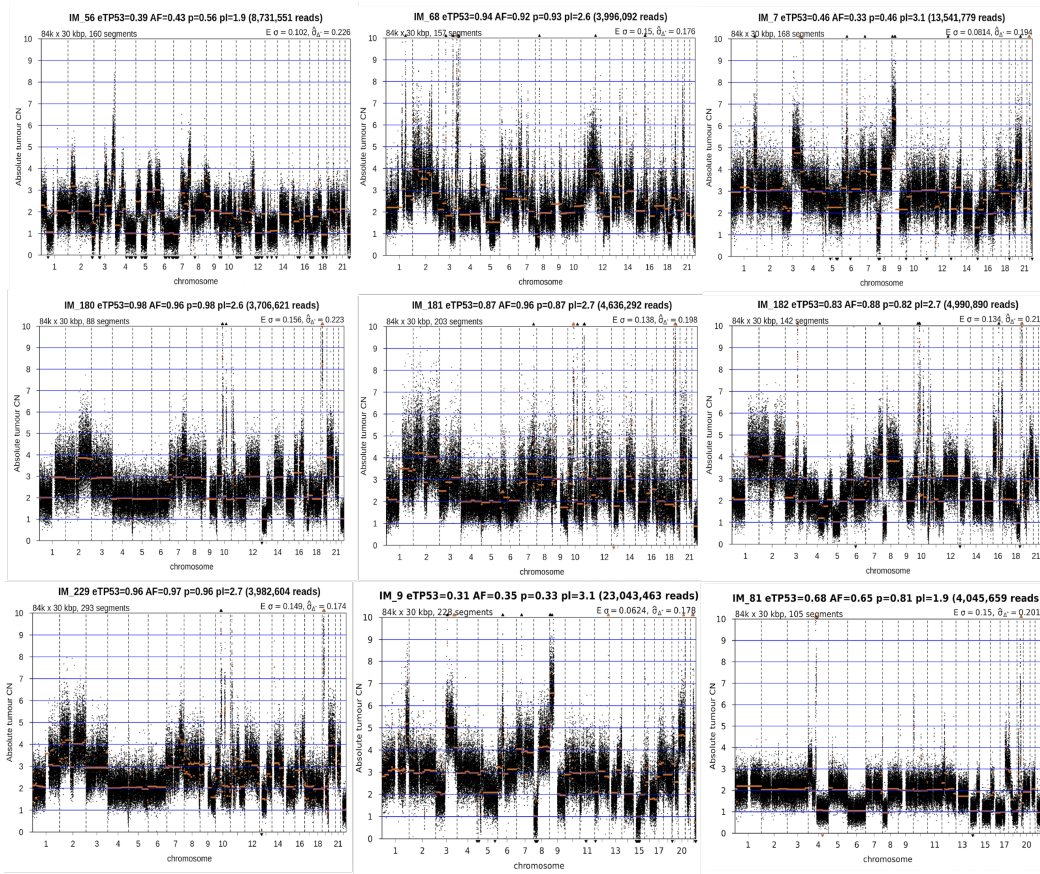


Appendix 9. Absolute copy number profiles from swgs-absn pipeline in late stage cohorts







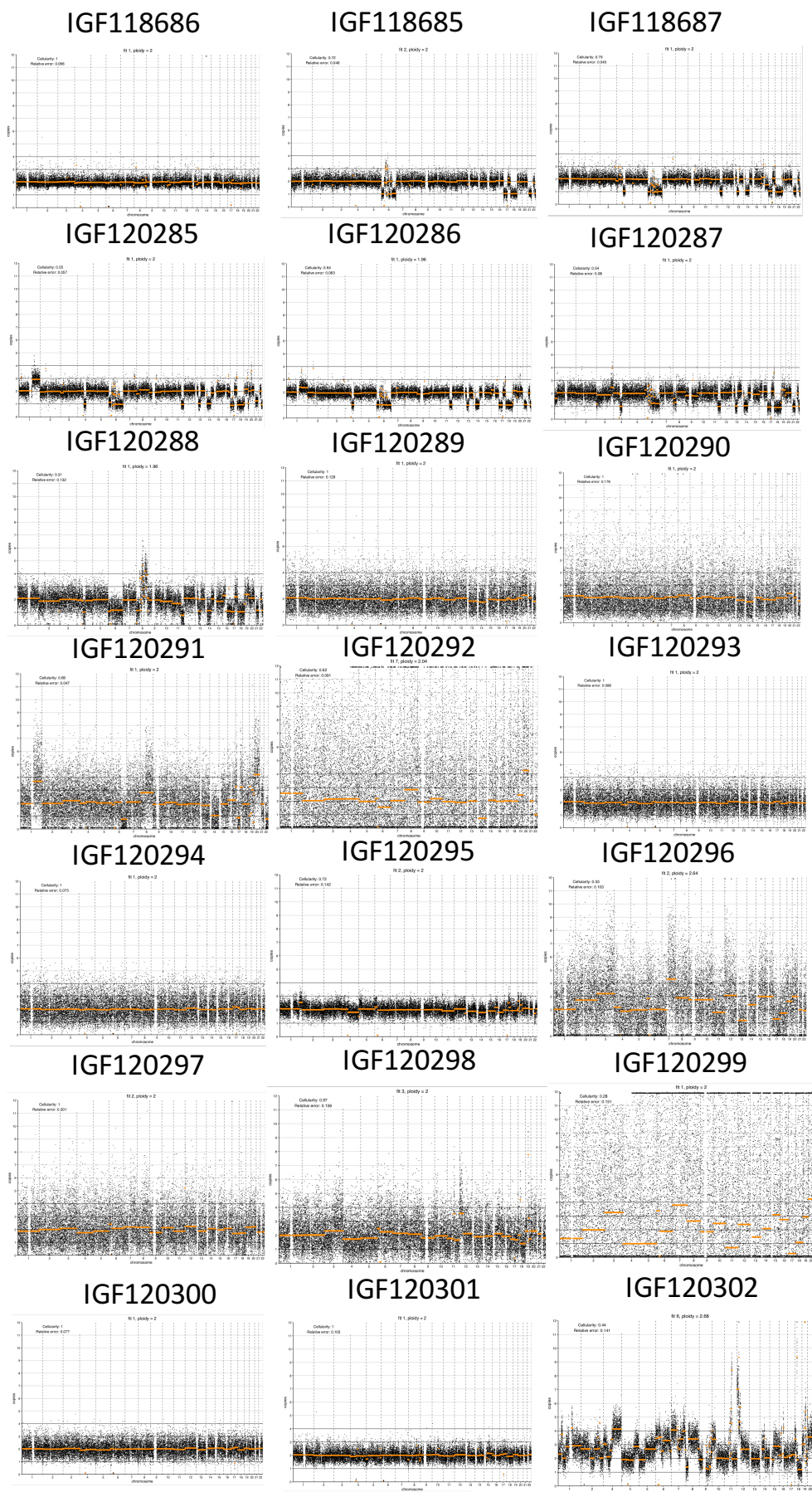


Appendix 10. DNA extraction result from all the samples in STIC project

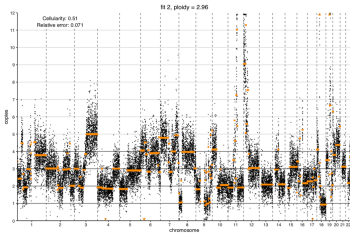
Sample name	Concentration(ng/ul)	Total yield(ng)
HS18-2994-B7_stroma	0.204	4.08
HS18-2994-B7_STIL	0.448	8.96
HS18-2994-B7_STIC	0.171	3.42
HS18-2994-A1	530	10600
HS18-2994-B5	570	11400
HS18-2994-C1	510	10200
HS18-2994-E2	700	14000
HS15-8221-C12-STIC1	0.088	1.76
HS15-8221-C12-STIC2+3	0.053	1.06
HS15-8221-C12-STIC4+5	0.075	1.5
HS15-8221-C12-p53_signature	<0.05	
HS15-8221-C12-normal_epithelial	0.248	4.96
HS15-8221-C12-stroma	0.152	3.04
HS15-8221-E1	28.9	578
HS15-8221-B3	53	1060
HS15-8221-C19	91.4	1828
100103_A2_tumour1	0.176	3.52
100103_A2_tumour2	0.114	2.28
100103_A2_stroma	<0.05	
100103_A2_STIC1+2	0.204	4.08
100103_A2_STIC3+4	<0.05	
100103_A2_p53_signature	<0.05	
HS15-6498-E5-p53_signature	0.08	1.6
HS15-6498-E5-STIC-1+2+3+4	0.382	7.64
HS15-6498-E5-primary_tumour1	0.354	7.08
HS15-6498-E5-primary_tumour2	0.218	4.36
HS15-6498-E5-normal_epithelial	0.28	5.6
HS15-6498-E5-stroma	0.086	1.72
HS15-6498-E9	31.6	632
HS15-6498-C7	33	660
HS15-6498-A1	4.88	97.6
HS17-533-A7-primary_tumour1	0.059	1.18

HS17-533-A7- primary_tumour2	<0.05	
HS17-533-A7-stroma	0.15	3
HS17-533-A7-STIC	<0.05	
HS17-533-A7-p53_signature	0.062	1.24
HS17-533-A7- normal_epithelial	0.123	2.46
HS17-533-B2	125	2500

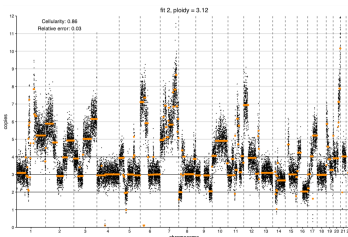
Appendix 11. Absolute copy number profiles in all anatomic sites in five patients



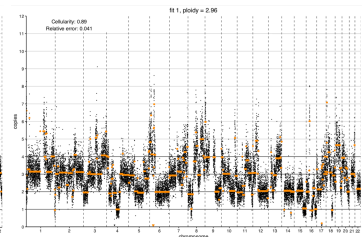
IGF120303



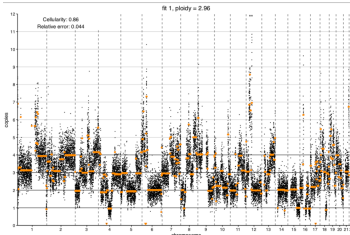
IGF120304



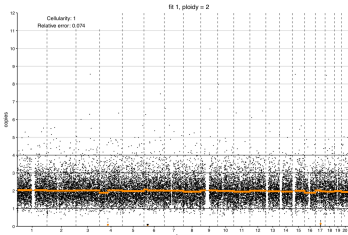
IGF120305



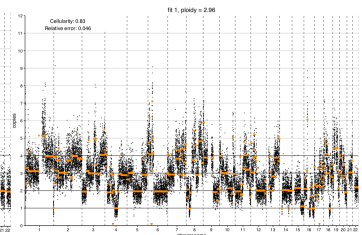
IGF120306



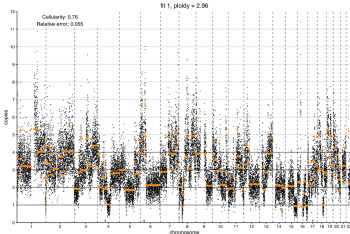
IGF120307



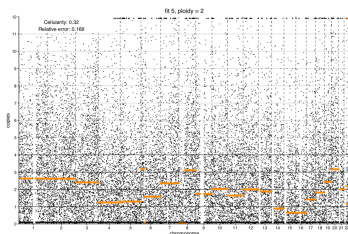
IGF120308



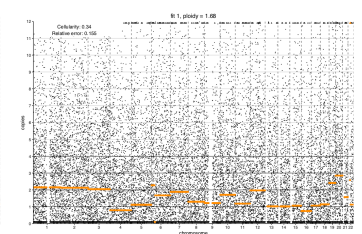
IGF120309



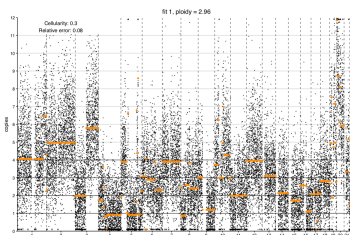
IGF120310



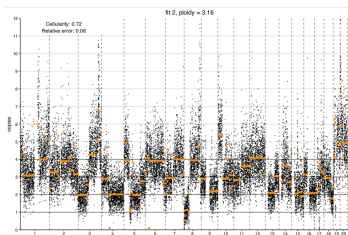
IGF120311



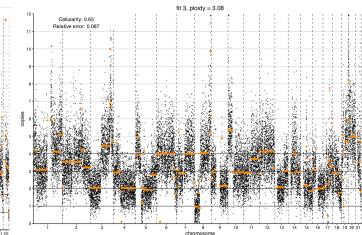
IGF120312



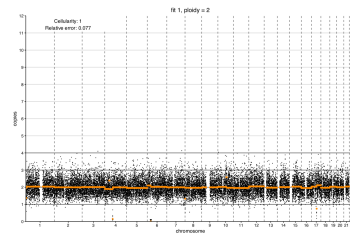
IGF120313



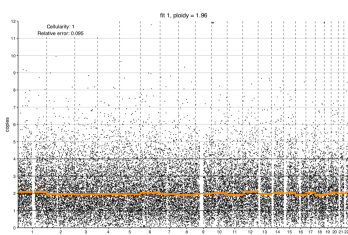
IGF120314



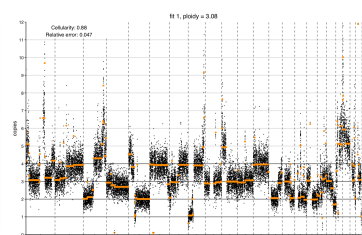
IGF120315



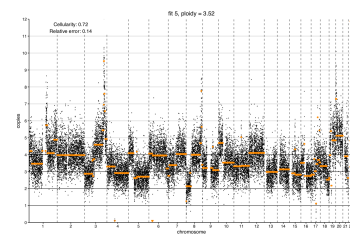
IGF120316



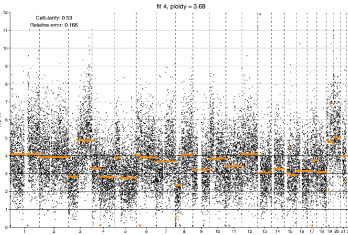
IGF120317



IGF120318



IGF120319



Copyright permission for each figure used in this thesis

1. Figure 1.1

MDPI Open Access Information and Policy

All articles published by MDPI are made immediately available worldwide under an open access license. This means:

- everyone has free and unlimited access to the full-text of *all* articles published in MDPI journals;
- everyone is free to re-use the published material if proper accreditation/citation of the original publication is given;
- open access publication is supported by the authors' institutes or research funding agencies by payment of a comparatively low **Article Processing Charge (APC)** for accepted articles.

Permissions

No special permission is required to reuse all or part of article published by MDPI, including figures and tables. For articles published under an open access Creative Common CC BY license, any part of the article may be reused without permission provided that the original article is clearly cited. Reuse of an article does not imply endorsement by the authors or MDPI.

2. Figure 1.2

MDPI Open Access Information and Policy

All articles published by MDPI are made immediately available worldwide under an open access license. This means:

- everyone has free and unlimited access to the full-text of *all* articles published in MDPI journals;
- everyone is free to re-use the published material if proper accreditation/citation of the original publication is given;
- open access publication is supported by the authors' institutes or research funding agencies by payment of a comparatively low **Article Processing Charge (APC)** for accepted articles.

Permissions

No special permission is required to reuse all or part of article published by MDPI, including figures and tables. For articles published under an open access Creative Common CC BY license, any part of the article may be reused without permission provided that the original article is clearly cited. Reuse of an article does not imply endorsement by the authors or MDPI.

3.Figure 1.3

03/04/2022, 18:04

RightsLink Printable License

OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Apr 03, 2022

This Agreement between Miss. Zhao Cheng ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License
Number 5281450532494

License
date Apr 03, 2022

Licensed
content
publisher Oxford University Press

Licensed
content
publication Clinical Chemistry

Licensed
content title Allele-Specific Droplet Digital PCR Combined with a Next-Generation
Sequencing-Based Algorithm for Diagnostic Copy Number Analysis in Genes
with High Homology: Proof of Concept Using Stereocilin

Licensed
content
author Amr, Sami S; Murphy, Elissa

Licensed
content date Apr 1, 2018

Type of Use Thesis/Dissertation

Institution
name

Title of
your work Tracking the evolution of ovarian high grade serous carcinoma

Publisher of Imperial College London
your work

Expected
publication date May 2022

Permissions
cost 0.00 GBP

Value added
tax 0.00 GBP

Total 0.00 GBP

Title Tracking the evolution of ovarian high grade serous carcinoma

Institution
name Imperial College London

Expected
presentation date May 2022

Portions Figure.2

Requestor
Location Miss. Zhao Cheng
Hammersmith Hospital,
Du Cane Road
London, Please Select W12 0NN
United Kingdom
Attn: Miss. Zhao Cheng

Publisher
Tax ID GB125506730

Total 0.00 GBP

Terms and Conditions

4. Figure 1.4

Emerging landscape of oncogenic signatures across human cancers

SPRINGER NATURE

Author: Giovanni Ciriello et al

Publication: Nature Genetics

Publisher: Springer Nature

Date: Sep 26, 2013

Copyright © 2013, The Author(s)

Creative Commons

The request you have made is considered to be non-commercial/educational. As the article you have requested has been distributed under a Creative Commons license (Attribution-Noncommercial), you may reuse this material for non-commercial/educational purposes without obtaining additional permission from Springer Nature, providing that the author and the original source of publication are fully acknowledged (please see the article itself for the license version number). You may reuse this material without obtaining permission from Springer Nature, providing that the author and the original source of publication are fully acknowledged, as per the terms of the license. For license terms, please see <http://creativecommons.org/>

BACK

CLOSE WINDOW

5. Figure 1.5



[BMC Med Genomics](#). 2012; 5: 47.

Published online 2012 Oct 18. doi: [10.1186/1755-8794-5-47](https://doi.org/10.1186/1755-8794-5-47)

PMCID: PMC3567940

PMID: [23078675](https://pubmed.ncbi.nlm.nih.gov/23078675/)

Histotype-specific copy-number alterations in ovarian cancer

[Ruby YunJu Huang](#),^{#1,2} [Geng Bo Chen](#),^{#3} [Noriomi Matsumura](#),⁴ [Hung-Cheng Lai](#),⁵ [Seiichi Mori](#),² [Jingjing Li](#),³ [Meng Kang Wong](#),² [Ikuo Konishi](#),⁴ [Jean-Paul Thiery](#),^{2,6} and [Liang Goh](#)^{#3,7,8}

► [Author information](#) ► [Article notes](#) ▼ [Copyright and License information](#) [Disclaimer](#)

[Copyright](#) ©2012 Huang et al.; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License

(<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Attribution 2.0 Generic (CC BY 2.0)

This is a human-readable summary of (and not a substitute for) the [license](#). [Disclaimer](#).

You are free to:

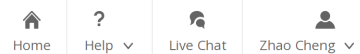
Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.



6. Figure 1.6





Copy number signatures and mutational processes in ovarian carcinoma
Author: Geoff Macintyre et al
Publication: Nature Genetics
Publisher: Springer Nature
Date: Aug 13, 2018
Copyright © 2018, The Author(s)

Order Completed

Thank you for your order.

This Agreement between Miss. Zhao Cheng ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	5296481328270	Printable Details
License date	Apr 26, 2022	

✔ Licensed Content		📋 Order Details	
Licensed Content Publisher	Springer Nature	Type of Use	Thesis/Dissertation
Licensed Content Publication	Nature Genetics	Requestor type	non-commercial (non-profit)
Licensed Content Title	Copy number signatures and mutational processes in ovarian carcinoma	Format	electronic
Licensed Content Author	Geoff Macintyre et al	Portion	figures/tables/illustrations
Licensed Content Date	Aug 13, 2018	Number of figures/tables/illustrations	1
		High-res required	no
		Will you be translating?	no
		Circulation/distribution	1 - 29
		Author of this Springer Nature content	no

📄 About Your Work		📁 Additional Data	
Title	Tracking the evolution of ovarian high grade serous carcinoma	Portions	Figure.3
Institution name	Imperial College London		
Expected presentation date	May 2022		

📍 Requestor Location		📄 Tax Details	
	Miss. Zhao Cheng 70 mellitus St, Shepherd's bush,London		
Requestor Location	London, Please Select W12 0AS United Kingdom Attn: Miss. Zhao Cheng		

\$ Price	
Total	0.00 GBP

Total: 0.00 GBP

7.Figure 1.7



Changes in the Extracellular Matrix Are Associated With the Development of Serous Tubal Intraepithelial Carcinoma Into High-Grade Serous Carcinoma

Author: Sophieke C.H.A. van der Steen,Johan Bulten,Koen K. Van de Vijver,Toin H. van Kuppevelt,Leon F.A.G. Massuger

Publication: International Journal of Gynecological Cancer

Publisher: BMJ Publishing Group Ltd.

Date: Jul 1, 2017

Copyright © 2017, BMJ Publishing Group Ltd. & International Gynecologic Cancer Society & European Society of Gynaecological Oncology

Order Completed

Thank you for your order.

This Agreement between Miss. Zhao Cheng ("You") and BMJ Publishing Group Ltd. ("BMJ Publishing Group Ltd.") consists of your license details and the terms and conditions provided by BMJ Publishing Group Ltd. and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number 5296500641990

[Printable Details](#)

License date Apr 26, 2022

Licensed Content

Licensed Content Publisher	BMJ Publishing Group Ltd.
Licensed Content Publication	International Journal of Gynecological Cancer
Licensed Content Title	Changes in the Extracellular Matrix Are Associated With the Development of Serous Tubal Intraepithelial Carcinoma Into High-Grade Serous Carcinoma
Licensed Content Author	Sophieke C.H.A. van der Steen,Johan Bulten,Koen K. Van de Vijver,Toin H. van Kuppevelt,Leon F.A.G. Massuger
Licensed Content Date	Jul 1, 2017
Licensed Content Volume	27
Licensed Content Issue	6

About Your Work

Title	Tracking the evolution of ovarian high grade serous carcinoma
Institution name	Imperial College London
Expected presentation date	May 2022

Requestor Location

Requestor Location	Miss. Zhao Cheng Hammersmith Hospital, Du Cane Road London, Please Select W12 0NN United Kingdom Attn: Miss. Zhao Cheng
--------------------	---

\$ Price

Total	0.00 GBP
-------	----------

Order Details

Type of Use	Dissertation/Thesis
Requestor Type	Individual
Format	Electronic
Portion Used	Figure/table/extract
Number of figure/table/extracts	1
Will you be translating?	No

Additional Data

Portions	Figure.3
Circulation	1

Tax Details

Publisher Tax ID	GB674738491
------------------	-------------

Total: 0.00 GBP

RESEARCH ARTICLE

Expression of Cell Competition Markers at the Interface between p53 Signature and Normal Epithelium in the Human Fallopian Tube

Masahiko Kito^{1,2}, Daichi Maeda^{1*}, Yukitsugu Kudo-Asabe¹, Naoki Sato², le-Ming Shih³, Tian-Li Wang³, Masamitsu Tanaka⁴, Yukihiro Terada², Akiteru Goto¹

1 Department of Cellular and Organ Pathology, Graduate School of Medicine, Akita University, Akita, Japan, **2** Department of Obstetrics and Gynecology, Graduate School of Medicine, Akita University, Akita, Japan, **3** Department of Pathology, Oncology and Gynecology and Obstetrics, Johns Hopkins Medical Institutions, Baltimore, MD, United States of America, **4** Department of Molecular Medicine and Biochemistry, Graduate School of Medicine, Akita University, Akita, Japan

* maeda-ky@umin.ac.jp



OPEN ACCESS

Citation: Kito M, Maeda D, Kudo-Asabe Y, Sato N, Shih I-M, Wang T-L, et al. (2016) Expression of Cell Competition Markers at the Interface between p53 Signature and Normal Epithelium in the Human Fallopian Tube. PLoS ONE 11(6): e0156069. doi:10.1371/journal.pone.0156069

Editor: Hiroyoshi Ariga, Hokkaido University, JAPAN

Received: February 5, 2016

Accepted: May 9, 2016

Published: June 3, 2016

Copyright: © 2016 Kito et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: This work was supported by a Grant-in-Aid for Scientific Research (KAKENHI) from the Japan Society for the Promotion of Science. Grant number: 26870129.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

There is a growing body of evidence regarding cell competition between normal and mutant mammalian cells, which suggest that it may play a defensive role in the early phase of carcinogenesis. *In vitro* study in the past has shown that overexpression of vimentin in normal epithelial cells at the contact surface with transformed cells is essential for the cell competition involved in epithelial defense against cancer. In this study, we attempted to examine cell competition in human tissue *in vivo* by investigating surgically resected human fallopian tubes that contain p53 signatures and serous tubal intraepithelial lesions (STILs), a linear expansion of p53-immunopositive/*TP53* mutant tubal epithelial cells that are considered as precursors of pelvic high grade serous carcinoma. Immunofluorescence double staining for p53 and the cell competition marker vimentin was performed in 21 sections of human fallopian tube tissue containing 17 p53 signatures and 4 STILs. The intensities of vimentin expression at the interface between p53-positive cells at the end of the p53 signature/STIL and adjacent p53-negative normal tubal epithelial cells were compared with the background tubal epithelium. As a result, the average vimentin intensity at the interfaces relative to the background intensity was 1.076 (95% CI, 0.9412 – 1.211 for p53 signature and 0.9790 (95% CI, 0.7206 – 1.237) for STIL. Thus, it can be concluded that overexpression of the cell competition marker vimentin are not observed in human tissue with *TP53* alterations.

Introduction

Normal and transformed epithelial cells often compete with each other for cell survival, a phenomenon called cell competition.[1,2] Cell competition was initially confirmed in *Drosophila* in 1975 and has been recently shown in mammalian cells.[3–7] Cell competition is now considered a unique type of short-range cell-to-cell communication that may take place during the early stages of carcinogenesis.[7,8] Genes such as *myc*, *scribble*, *sparc* and *mahjong* have been shown to play an important role in this process.[9–13] Based on the results of *in vitro* studies

Review

Genetic Clonality as the Hallmark Driving Evolution of Non-Small Cell Lung Cancer

Marcin Nicoś  and Paweł Krawczyk

Department of Pneumology, Oncology and Allergology, Medical University of Lublin, 20-954 Lublin, Poland; krapa@poczta.onet.pl

* Correspondence: marcin.nicos@umlub.pl

Simple Summary: Limited knowledge about NSCLC evolution has affected therapeutic strategies for many decades. The application of NGS-based techniques to studies on ITH has provided genetic insight into the contribution of clonality primary seeding, as well as to distant dissemination. To date, multiregional ITH affects accurate diagnosis and treatment decisions and is considered the main hallmark of anticancer therapy failure. Understanding the evolutionary trajectories that drive the metastatic process is critical for improving treatment strategies for this deadly condition. In this review, we discuss how the clonality of genetic alterations influence the seeding of primary and metastatic lesions of NSCLC, highlighting that wide genetic analyses may reveal the phylogenetic lineages of NSCLC evolution.

Abstract: Data indicate that many driver alterations from the primary tumor of non-small cell lung cancer (NSCLC) are predominantly shared across all metastases; however, disseminating cells may also acquire a new genetic landscape across their journey. By comparing the constituent subclonal mutations between pairs of primary and metastatic samples, it is possible to derive the ancestral relationships between tumor clones, rather than between tumor samples. Current treatment strategies mostly rely on the theory that metastases are genetically similar to the primary lesions from which they arise. However, intratumor heterogeneity (ITH) affects accurate diagnosis and treatment decisions and it is considered the main hallmark of anticancer therapy failure. Understanding the genetic changes that drive the metastatic process is critical for improving the treatment strategies of this deadly condition. Application of next generation sequencing (NGS) techniques has already created knowledge about tumorigenesis and cancer evolution; however, further NGS implementation may also allow to reconstruct phylogenetic clonal lineages and clonal expansion. In this review, we discuss how the clonality of genetic alterations influence the seeding of primary and metastatic lesions of NSCLC. We highlight that wide genetic analyses may reveal the phylogenetic trajectories of NSCLC evolution, and may pave the way to better management of follow-up and treatment.

Keywords: NSCLC; tumor evolution; driver mutations; clonal/subclonal alterations; metastases



Citation: Nicoś, M.; Krawczyk, P. Genetic Clonality as the Hallmark Driving Evolution of Non-Small Cell Lung Cancer. *Cancers* **2022**, *14*, 1813. <https://doi.org/10.3390/cancers14071813>

Academic Editor: Matthias Wirth

Received: 20 February 2022

Accepted: 29 March 2022

Published: 2 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Darwinian theory of variation, heredity and selection had provided a basic evolutionary framework that has been further adopted to develop models of tumor evolution [1,2]. Based on this, today we postulate that cancer may arise from normal cells as a consequence of somatic or germline genetic alterations that represent a key hallmark related to the tumorigenesis [3]. The genetic disorders may affect the cell's survival and proliferations as "drivers" of tumorigenesis, or stay biologically neutral as "passengers" of the evolutionary processes [4]. Moreover, the alterations may be limited only to a subset of cancer cells, as clonal changes that are private within a single taxon, or subclonally distributed in all cancer cells sharing abnormalities between the trunks of the phylogenetic tree [5–7]. However, even a homogenous tumor with the clonal origin may accumulate