

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Department of Biotechnology



**UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA**

DOCTORAL THESIS

**GENOMIC INSTABILITY AS A PREDICTIVE BIOMARKER
FOR THE APPLICATION OF DNA-DAMAGING THERAPIES
IN GYNECOLOGICAL CANCER PATIENTS**

Presented by:

RAQUEL LÓPEZ REIG

Supervised by:

JOSÉ ANTONIO LÓPEZ GUERRERO

ANTONIO FERNÁNDEZ SERRA

VALÈNCIA, JULY 2023

Dr. José Antonio López Guerrero, Head of the Molecular Biology Lab from the Fundación Instituto Valenciano de Oncología.

Dr. Antonio Fernández Serra, Assistant molecular biologist in the Molecular Biology Lab from the Fundación Instituto Valenciano de Oncología.

CERTIFY THAT:

The present doctoral thesis “Genomic instability as a predictive biomarker for the application of DNA-damaging therapies in gynecological cancer patients” has been developed by Raquel López Reig in the Molecular Biology Lab from the Fundación Instituto Valenciano de Oncología under our supervision and meet all the requirements for its deposit and defense.

And in witness thereof, the present document is signed in Valencia, 18th of July of 2023.



Dr. José Antonio López Guerrero



Dr. Antonio Fernández Serra

The present doctoral thesis was partially funded by GVA Grants “Subvencions per a la realització de projectes d’i+d+i desenvolupats per grups d’investigació emergents (GV/2020/158)” and “Ayudas para la contratación de personal investigador en formación de carácter predoctoral” (ACIF/2016/008), “Beca de investigación traslacional Andrés Poveda 2020” from GEICO group and Phase II clinical trial (POLA: NCT02684318, EudraCT 2015-001141-08, 03.10.2015). This study was awarded the Prize “Antonio Llombart Rodriguez-FINCIVO 2020” from the Royal Academy of Medicine of the Valencian Community.

*“Reserve your right to think,
for even to think wrongly is better than not to think at all”*

Hipatia

Acknowledgments

This thesis would not have been possible without the involvement and work of many entities and people that have collaborated and helped during this process, from the initial steps until now. First of all, I would like to thank the Fundación Instituto Valenciano de Oncología, for allowing me to elaborate my thesis in this institution. Being able to develop this kind of project in a reference center in Oncology is an honour.

I would also like to express my deepest gratitude to Andrés Poveda, who, together with Nacho Romero, decided to make me part of their team. This multidisciplinary group of professional, including oncologist, gynaecologist, pathologist, radiotherapists, radiologists, psicologists and physiotherapist, was the perfect breeding ground for great research projects with which to start my scientific career.

I could not have undertaken this journey without the trust, guidance, and expertise of the supervisors of the thesis. José Antonio, your determination and self-demand have made me do my best to try to be up to the task. Antonio, you get on board a little bit later, but you have become essential in keeping the boat afloat. Your knowledge, curiosity, and eternal perseverance have taught me to never stop learning and growing. It is a pleasure to call you a friend. Go BIT!

Additionally, this endeavor would not have been possible without the generous support from each one of the members of the molecular biology lab from the Instituto Valenciano de Oncología. Particularly to our cheerleaders, Reyes, Patri, and Isa, who have made the darker days a little bit brighter.

During all this time, I also had the pleasure of collaborating with different research centers. I would like to acknowledge the invaluable scientific advice and help from Dr. Mar Orzaez and Dr. Mónica Sancho from The targeted therapies and inflammation in cancer lab (Centro de Investigación Príncipe Felipe, CIPF). I am also thankful to Dr. Marta Mendiola and Dr. Victoria Heredia from Cancer molecular pathology and therapeutic targets (IdiPaz), for selfless sharing the cell lines studied in this work. Finally, I would like to express my deepest appreciation to Dr. Delia Mezzanzanica and Dr. Marina Bagnoli, from the Integrated Biology of Rare Tumors at Fondazione IRCCS Istituto Nazionale dei Tumori di Milano. They made my fellowship scientifically and personally worthy.

I would like to extend my sincere thanks to GEICO (Grupo Español de Investigación en Cáncer de Ovario), particularly to the Technical secretary and Executive Committee, who believe in the project and have given us support all along the way.

I cannot forget to mention all the patients that, even in their hardest times, trust in science and decided to participate in this study. In addition, thanks to the Biobank of the Fundación Instituto Valenciano de Oncología for providing the biological samples for the analysis.

In addition, I would like to acknowledge Drs. Belen Picó and Lynne Yenush and F. Javier Martínez for the priceless help mainly during the last steps.

And last but not least, I would like to thank the hard core of people that surround me every day. This journey would not have been half as fun and light without you traveling by my side. Luckily, I treasure friends from every stage of life, so

forgive me for not mentioning you one by one. You know who you are.

To my family, because whoever knows me knows that there is no me without them. From them comes the curiosity to know, to get the answers, and to fight to get them. Each family event has prepared me for a good discussion, extremely important in science. Specially thanks to mamá, papá, and Sergi. As once said, you are my safety net. Taking a step is much easier knowing that, if something does not go as expected, you will be there for me. And of course, I cannot forget the latest additions. Martina, mi bichito favorito, what would I have done without you the last five years? And Marc, the best is yet to come.

Thanks to all of them, you can now call me DOCTOR.

Summary

The natural course of tumors matches the progressive accumulation of genomic alterations, triggering a cascade of events that results in genomic instability (GI). This phenomenon includes copy number alterations and constitutes a genomic hallmark that defines specific outcomes beyond histology and other molecular features of the tumor.

In the context of gynaecologic oncology research, GI has gained strength in the last years allowing the stratification of patients according to prognosis and response to certain DNA-damaging agents, such as platinum-based therapies and PARP inhibitors. Particularly in ovarian and endometrial cancers, it has been described a molecular subgroup characterized by high copy number alterations (CNA) related to good prognosis and better response to chemotherapy. This relationship highlights GI as a predictive and prognostic biomarker. Hence, a GI-based model translated into clinical practice would constitute a tool for optimizing clinical decision-making.

The era of personalised medicine arrived together with the coming of integrative studies, where results of high-throughput techniques are combined to obtain a comprehensive molecular landscape of the diseases, bringing a new paradigm to

characterize the tumors beyond classical anatomic and histological characteristics.

This thesis proposes a global study of the phenomenon of GI as a prognostic and predictive biomarker of treatment response in gynaecological cancers, mainly focused on high-grade ovarian cancer and endometrial cancer. Through the development of an NGS-based strategy with the adaptation of available pipelines of analysis, we obtained GI profiles on formalin-fixed paraffin-embedded samples in a reliable, portable, and cost-effective approach, with the combination of Machine Learning tools to fit prognostic and predictive models based on the integration of omic data.

Based on that premise, we fit and validated, in well-characterized clinical cohorts, three single-source models and an integrative ensemble model (Scarface Score) that proved to be able to predict response to DNA-damaging agents in a clinical scenario of High-Grade Serous Ovarian Cancer. In addition, a mutational-based algorithm (12g algorithm) with prognostic impact was developed and validated for endometrial cancer patients. This algorithm achieved a GI-based stratification of patients. Finally, a panel of ovarian cancer cell lines was characterized at the response, genetic and genomic level, interrogating homologous recombination repair

pathway status and its associated GI profiles, completing the molecular landscape, and establishing the basis and breeding ground of future preclinical and clinical studies.

The results reported in this Doctoral Thesis provide valuable clinical management tools in the accomplishment of a reliable tailored therapy. Additionally, future studies in different tumor types and drugs for implementation of the predictive model can be planned, using as a base the defined one but re-establishing new and specific cut-offs.

Resumen

El curso natural de los tumores va acompañado de la acumulación progresiva de alteraciones genómicas, propiciando una cadena de eventos que resultan en inestabilidad genómica (IG). Éste fenómeno, caracterizado por alteraciones en el número de copias, constituye un *hallmark* genómico con impacto pronóstico más allá de la histología y otras características moleculares del tumor.

En el ámbito de la investigación en oncología ginecológica, la IG ha ganado fuerza en los últimos años, permitiendo la estratificación de pacientes de acuerdo al pronóstico y la respuesta a agentes que dañan el ADN, como las terapias basadas en platinos y los inhibidores de PARP. En el cáncer de ovario, en particular, se ha descrito un subgrupo molecular caracterizado por alta incidencias de alteraciones en el número de copias relacionado con un mejor pronóstico y respuesta a quimioterapia. Esta correlación presenta la IG como un buen marcador predictivo y pronóstico. Así, un modelo basado en la IG trasladable a la práctica clínica constituirá una herramienta útil para la optimización de la toma de decisiones.

La era de la medicina personalizada llegó de la mano de los estudios integrativos, donde las técnicas de alto rendimiento se

aplican de manera combinada para obtener una visión molecular global de los tumores, completando y complementando la caracterización clásica a nivel anatómico e histológico.

Esta tesis propone un estudio global de la IG como biomarcador pronóstico y predictivo de respuesta en cáncer ginecológico, haciendo hincapié en el cáncer de ovario seroso de alto grado y cáncer de endometrio. A través de la aplicación de estrategias basadas en NGS con la adaptación de pipelines de análisis disponibles obtuvimos los perfiles de IG de muestras de tejido fijadas en formol y embebidas en parafina, de una manera fiable, portable y coste efectiva, combinando herramientas de *machine learning* para ajustar modelos predictivos y pronósticos.

Partiendo de esta premisa, ajustamos y validamos, en cohortes clínicas bien caracterizadas, tres modelos a partir de los datos ómicos individuales y un modelo integrativo (*Scarface Score*) que demostró la capacidad de predecir la respuesta a agentes que dañan el ADN en un escenario clínico concreto de pacientes con cáncer de ovario seroso de alto grado.

Paralelamente, desarrollamos y validamos un algoritmo basado en el perfil de mutaciones, con impacto pronóstico, en cáncer

de endometrio. Éste algoritmo consiguió una estratificación que respondía al perfil de IG de los pacientes.

Finalmente, se caracterizó un panel de líneas celulares de cáncer de ovario a nivel de respuesta, genético y genómico. Se interrogó el estatus de la vía de recombinación homóloga y su asociación a patrones de IG, completando el perfil molecular y estableciendo las bases para futuros estudios preclínicos y clínicos.

Los resultados obtenidos en esta tesis doctoral presentan herramientas de gran valor para el manejo clínico en cuanto a la búsqueda de una medicina personalizada. Adicionalmente, diferentes estudios para trasladar el modelo predictivo a otros escenarios clínicos pueden ser planteados, usando como base el planteado pero restableciendo puntos de corte nuevos y específicos.

Resum

El curs natural dels tumors va acompanyat de l'acumulació progressiva d'alteracions genòmiques, propiciant una cadena d'esdeveniments que resulten en inestabilitat genòmica (IG). Aquest fenomen, caracteritzat per la presència de alteracions en el nombre de còpies, constitueix un *hallmark* genòmic amb impacte pronòstic més enllà de la histologia i altres característiques moleculars del tumor.

En l'àmbit de la recerca en oncologia ginecològica, la IG ha guanyat força en els últims anys, permetent l'estratificació de pacients d'acord amb el pronòstic i la resposta d'agents que danyen l'ADN, com les teràpies basades en platins i els inhibidors de PARP. En el càncer d'ovari en particular, s'ha descrit un subgrup molecular caracteritzat per una alta incidència d'alteracions en el nombre de còpies relacionat amb un millor pronòstic i resposta a quimioteràpia. Aquesta correlació presenta la IG com un marcador predictiu i pronòstic adequat. Així, un model basat en la IG traslladable a la pràctica clínica constituirà una eina útil per a l'optimització de la presa de decisions.

L'era de la medicina personalitzada va arribar de la mà dels estudis integratius, on les tècniques d'alt rendiment s'apliquen

de manera combinada per a obtenir una visió molecular global dels tumors, completant i complementant la caracterització clàssica a nivell anatòmic i histològic.

Aquesta tesi proposa un estudi global de la IG com a biomarcador pronòstic i predictiu de resposta en càncer ginecològic, posant l'accent en el càncer d'ovari serós d'alt grau i càncer d'endometri. A través de la aplicació d'estratègies basades en NGS amb l'adaptació de pipelines d'anàlisis disponibles, vam obtenir els perfils de IG de mostres de teixit fixades en formol i embegudes en parafina d'una manera fiable, portable i cost efectiva, combinant eines de *machine learning* per a ajustar models predictius i pronòstics.

Partint d'aquesta premissa, vam ajustar i validar, en cohortes clíniques ben caracteritzades, tres models a partir de les dades omiques individuals i un model integratiu (Scarface Score) que va demostrar la capacitat de predir la resposta a agents que danyen l'ADN en un escenari clínic concret de pacients amb càncer d'ovari serós d'alt grau.

Paral·lelament, desenvoluparem i validarem un algoritme basat en el perfil de mutacions amb impacte pronòstic en càncer d'endometri. Aquest algoritme va aconseguir una estratificació que responia al perfil de IG dels pacients.

Finalment, es va caracteritzar un panell de línies cel·lulars de càncer d'ovari a nivell de resposta, genètic i genòmic. Es varen interrogar l'estatus de la via de recombinació homòloga i la seua associació a patrons de IG, completant el perfil molecular i establint les bases per a futurs estudis preclínic i clínic.

Els resultats obtinguts en aquesta tesi doctoral presenten eines de gran valor per al maneig clínic en quant a la cerca d'una medicina personalitzada. Addicionalment, diferents estudis per a traslladar el model predictiu a altres escenaris clínic poden ser plantejats, usant com a base el propost però restablint punts de tall nous i específic.

Index of contents

0. Glossary

1. Introduction..... 3

1.1. *Gynaecological Cancers*4

1.1.1. Epithelial Ovarian Cancer4

1.1.2. Endometrial Cancer22

1.2. *Genomic Instability*36

1.2.1. Genomic Instability and cancer (leading causes) 36

1.2.2. Leading causes of genomic instability37

1.2.3. Clinical implications of genomic instability41

1.2.4. Methodological approaches to determine deficiencies in DDR pathways44

1.3. *In vitro cellular models*.....49

2. Hypothesis and main objectives 51

2.1. *Hypothesis*.....53

2.2. *Main objectives*.....53

3. Materials and methods 57

3.1. *Patients cohorts*59

3.1.1. COHORT 159

3.1.2. COHORT 262

3.1.3.	COHORT 3	66
3.1.4.	COHORT 4	70
3.2.	<i>Immunohistochemistry analysis</i>	73
3.3.	<i>Genomic approaches</i>	74
3.3.1.	DNA extraction, quantification and quality control 75	
3.3.2.	Truseq Low Input	76
3.3.3.	SureSelect XT HS +Oneseq Backbone 1 mb.....	77
3.3.4.	Microsatellite instability	79
3.3.5.	Sanger Sequencing	80
3.3.6.	Multiplex Ligation-dependent Probe Amplification 81	
3.4.	<i>Transcriptomic approaches</i>	82
3.4.1.	HTG Oncology Biomarker Panel	82
3.5.	<i>Bioinformatics Analysis</i>	85
3.5.1.	Copy-Number Variation pipelines	85
3.5.2.	Data mining in OC cohort	89
3.5.3.	Random Forest Algorithm	94
3.6.	<i>In vitro studies</i>	95
3.6.1.	Cell lines and culturing	96
3.6.2.	Treatment schemes, drug-sensitivity assays and combination treatment.....	97
3.7.	<i>In vitro data processing</i>	106

3.8.	<i>Statistics</i>	106
3.8.1.	Descriptive statistics	106
3.8.2.	Parameters of clinical response	106
3.8.3.	Survival analysis	108
4.	Results	109
4.1.	<i>Characterization of genomic Instability</i>	111
4.1.1.	Establishing the basis of GI: tuning methodological and analytical approaches.....	111
4.1.2.	Multi-gene NGS panel to determine HRD and clinical impact of molecular classification.....	111
4.1.3.	Predictive model fitting and validation to predict response to DNA-damaging agents.....	124
4.1.4.	Another clinical scenario: POLA phase II cohort	133
4.1.5.	OC Cell line Panel	141
4.2.	<i>12g prognostic model in EC cohort</i>	146
4.2.1.	Selection of the multigene-NGS panel and mutational analysis	146
4.2.2.	Impact of 12 genes RF model in the clinical stratification of the disease.....	153
4.2.3.	CNV-based classification: Validation of the 12g model	158
4.2.4.	Expression analysis	161
4.3.	<i>In vitro characterization</i>	165

4.3.1. Drug-sensitivity analysis of ovarian Cancer cell lines: IC50 calculation.....	165
4.3.2. Generation of PM01183-Resistant cell line and functional characterization	170
5. Discussion.....	177
6. Conclusions.....	209
7. References.....	213
Annexes	237
<i>Annexed data</i>	<i>239</i>
<i>Annexed papers</i>	<i>275</i>

Index of figures

Figure 1: Global distribution of deaths caused by Cancer malignancies (GLOBOCAN 2020).....	3
Figure 2: Incidence and mortality stratified by tumor type across global female population	4
Figure 3: Classification of OC tumors	8
Figure 4: PARPi through the years in OC.....	18
Figure 5: Mutational landscape of HGSOC.....	21
Figure 6: Classification of EC tumors.....	24
Figure 7: Available cytoreductive approaches for patients with EC.....	28
Figure 8: Molecular-based classification from EC-TCGA.....	33
Figure 9: Clinically-validated surrogates of the 4-groups prognostic EC-TCGA classification.....	36
Figure 10: Overview of main DNA lesions and their related DDR pathways	38
Figure 11: Schematic representation of HRR pathway and its participants.....	40
Figure 12: Commercial solutions to establish HRDscore	44
Figure 13: Methodological approaches to detect defects in the HR pathway	45
Figure 14: Methodological strategy in cohort 1.....	61
Figure 15: Methodological strategy in cohort 2.....	65

Figure 16: Methodological strategy in cohort 3.....	69
Figure 17: Methodological strategy in cohort 4.....	72
Figure 18: HTG protocol	83
Figure 19: Tested CNV analysis pipelines step by step	88
Figure 20: Different tested algorithm to build the model.	89
Figure 21: Schematic summary of performed in vitro studies..	96
Figure 22: Molecular landscape of Cohort I	113
Figure 23: Log-Rank test and Kaplan-Meier plots	114
Figure 24: Differences in GI parameters according segmentation used in CNVkit pipeline	115
Figure 25: Differences in GI parameters adjusting p-value in CNVkit pipeline.....	116
Figure 26: Differences in GI parameters according tumor burden in CNVkit pipeline	117
Figure 27: GI parameters according pre-filtering step in saasCNV pipeline.....	117
Figure 28: Comparison of GI parameters between implemented pipelines; CNVkit, saasCNV and SureCall	119
Figure 29: Differences in PFI between evaluated pipelines....	121
Figure 30: Comparison of percentage of altered genome between HRR-based groups.....	122
Figure 31: Distribution of different genomic scars and integrative HRDscore between PFI-based stratification..	123

Figure 32: Distribution of different genomic scars and integrative HRDscore between HRR mutations-based stratification	124
Figure 33: Correlation between PFI and fitted models.....	127
Figure 34: ROC curves comparing performance single-source and ensemble model.....	128
Figure 35: Log-rank tests evaluating the performance of different tested molecular classifiers to predict PFS to PARPi	130
Figure 36: Log-Rank tests evaluating the implication of mutational-based classifiers and predictive models with OS	131
Figure 37: Multivariate analysis in terms of PFI	132
Figure 38: Multivariate analysis in terms of OS	133
Figure 39: Molecular landscape of Cohort 2	134
Figure 40: Comparison of GI patterns between cancer types	135
Figure 41: Non-parametric tests (Wilcoxon signed rank test) comparing the GI parameters and HRD status in the whole population	136
Figure 42: Clinical implication of the GI parameters in the OC population (n = 38).....	138
Figure 43: Clinical implication of GI parameters regarding ORR	139
Figure 44: Molecular landscape of OC cell line panel	142

Figure 45: Representation of CNV patterns across OC cell line panel.....	143
Figure 46: Exemplifying GI in HRR defective cell lines over time	145
Figure 47: Molecular landscape of Cohort 3	147
Figure 48: Surrogate of tumor mutational burden (TMB) across four EC prognostic subtypes evaluated by non-parametric test.....	156
Figure 49: Distribution of genetic alteration across the four EC prognostic subtypes	157
Figure 50: Kaplan-Meier plots assessed by log-rank test to evaluate the impact of the 12g model over outcome parameters	158
Figure 51: Unsupervised clustering of EC patients based on CN patterns	159
Figure 52: Distribution of genomic scars among the 12g subpopulation, stratified by 12g Model.....	160
Figure 53: Distribution of GI parameters between CNV-based groups.....	161
Figure 54: Unsupervised analysis of subpopulation of cohort 3 classified by 12g model	162
Figure 55: Clusterization of subpopulation of cohort 3 after DE analysis	162

Figure 56: Log-rank test evaluating the clinical impact of 12g-based classification	163
Figure 57: Clusterization of patients after DE analysis in CNH population	164
Figure 58: Log-rank test evaluating the clinical impact of cluster-based classification	164
Figure 59: Setting-up range of drug-concentration and exposure time to perform cytotoxic assays	166
Figure 60: Cytotoxic assays of PM01183 in monotherapy.....	167
Figure 61: Cytotoxic assays of PARPi in monotherapy.....	168
Figure 62: Cytotoxic assays of PARPi and PM01183 in combination	169
Figure 63: Quantification of changes in the IC50 values after the induction of PM01183 resistance	171
Figure 64: Doubling-time assay	172
Figure 65: Cell cycle analysis	173
Figure 66: Analysis of HRR-related proteins by Western blot following drug treatment.....	174

Index of tables

Table 1: FIGO staging system.	9
Table 2: OC risk factors.....	12
Table 3: FIGO Staging of Uterine Corpus Carcinoma and Carcinosarcoma.....	26
Table 4: Clinicopathological characteristics of cohort 1	60
Table 5: Clinicopathological features of Cohort 2.....	63
Table 6: Clinicopathological characteristics of cohort 3	67
Table 7: Clinicopathological features of Cohort 4.....	71
Table 8: IHC antibodies applied to interrogate MMR and p53 status in EC.....	74
Table 9: Primer sequences for Sanger sequencing of POLE EDM in EC series.	81
Table 10. Tested range of drug-concentrations during cytotoxic assays.....	99
Table 11: Western Blot antibodies applied to interrogate HRR functionality in cell lines.....	104
Table 12: Performance of the single-source and ensemble model.....	126
Table 13: Correlation between mutational status of analysed genes and clinicopathological and outcome parameters.	149

Table 14: Contribution of evaluated parameters to 12g-model	154
Table 15: Performance parameters of the 12g model in the validation series.	155
Table 16: Drug sensitivity data for tested OC cell lines and treatments.....	169

Index of Annexes

Annex 1: Detailed inclusion and exclusion criteria for POLA Phase I/IIB clinical trial.	239
Annex 2: Clinical and molecular characteristics from OC cell line panel.....	241
Annex 3: Profile of STR markers (Identifilier®).....	243
Annex 4: Sequencing metrics cohort 1	244
Annex 5: Variant information cohort 1	247
Annex 6: Germline BRCA mutations cohort 1	249
Annex 7: Comparison of performances between CNV pipelines	250
Annex 8: Correlation between HRD-based classification and different GI parameters	252
Annex 9: CNVkit performance.....	253
Annex 10: Parameter feeding SNP-Model.	254
Annex 11: Parameter feeding HTG-Model.....	255
Annex 12: Sequencing metrics of Cohort 4.....	256
Annex 13: Distribution of variants by annotation in Cohort 4	261
Annex 14: Clinical impact of features included in12g-algorithm	266

Annex 15: Alteration found in cohort 4 sorted by structural annotation among EC prognostic subtypes (Median number of alteration/group).....	268
Annex 16: Distribution of genetic alteration across the four prognostic groups.....	269
Annex 17: Log-Rank tests evaluating the implication of predefined HRD scars parameters from scarHRD package in correlation with PFI and PARPi response	270
Annex 18: Correlation of HRD score obtained on scarHRD package and time-to-event variables.....	271
Annex 19: Contribution of each parameter in the CPP model measured as decreasing of the Gini index	272
Annex 20: Performance of the RF model (CPP) including clinical and pathological parameters (grade, histology and stage)	273
Annexed papers	276
- Prognostic classification of endometrial cancer using a molecular approach based on a twelve-gene NGS panel.	
- Phase 2 Trial (POLA Study) of Lurbinectedin plus Olaparib in Patients with Advanced Solid Tumors: Results of Efficacy, Tolerability, and the Translational Study.	

- The Scarface Score: Deciphering Response to DNA Damage Agents in High-Grade Serous Ovarian Cancer-A GEICO Study

Glossary

AF: Allelic Frequency

AUC: Area under the curve

BER: Base-Excision Repair

CBR: Clinical Benefit Ratio

CBS: Circular Binary Segmentation

CGH: Comparative Genomic Hybridization

CI: Combination Index

CNH: Copy-Number High

CNL: Copy-Number Low

CNV: Copy-Number Variation

CR: Complete Response

DDR: DNA-Damage Response

DE: Differential Expression

DFS: Disease-Free Survival

DSB: Double-Strand Break

EC: Endometrial Cancer

EDM: Exonuclease Domain

EOC: Epithelial Ovarian Cancer

FBS: Fetal Bovine Serum

FFPE: Formalin-Fixed Paraffin-Embedded

FIGO: International Federation of Gynecology and Obstetrics

GAGE: Generally Applicable Gene-set Enrichment

GI: Genomic Instability

GSEA: Gene Set Enrichment Analysis

HGSOC: High-Grade Serous Ovarian Cancer

HRD: Homologous Recombination Deficiency

HRP: Homologous Recombination Proficiency

HRR: Homologous Recombination Repair

HRT: Hormone Replacement Treatment

IC10/30/50: Inhibitory Concentration

IHC: Immunohistochemistry

LOH: Loss of Heterozygosity

LS: Lynch Syndrome

LST: Large-Scale Transition

LTR: Long-term Responder

ML: Machine Learning

MLPA: Multiplex Ligation-dependent Probe Amplification

MMR: Mismatch Repair

MSI: Microsatellite Instability

MTT: 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide

NER: Nucleotide-Excision Repair

NGS: Next-Generation Sequencing

NHEJ: Non-Homologous End Joining

NN: Neural Network

OBP: Oncology Biomarker Panel

OC: Ovarian Cancer

OS: Overall Survival

PARPi: PARP inhibitor

PBS: Phosphate-buffered saline

PCA: Principal Component Analysis

PFI: Platinum-Free Interval

PFS: Progression-Free Survival

PR: Partial Response

RF: Random Forest Algorithm

RT: Radiotherapy

S2N: Signal to Noise

SNP: Single-Nucleotide Polymorphism

SNV: Single-Nucleotide Variant

SSB: Single-Strand Break

STR: Short Tandem Repeat

SVM: Support Vector Machine

TAI: Telomeric Allelic Imbalance

TBS-T: Tris-buffered saline with tween

TCGA: The Cancer Genome Atlas

TFI: Treatment-Free Interval

TNM: Tumor-Node-Metastasis

TMB: Tumor Mutational Burden

WGS: Whole-Genome Sequencing

WES: Whole-Exome Sequencing

INTRODUCTION

1. Introduction

Cancer is one of the leading causes of death worldwide (Figure 1) constituting one of the main health problems that need to be addressed. In 2020 It has been described an incidence of approximately 19.3 million of new cases and a mortality near 10 million. The rapid increase in incidence and mortality mainly reflects aging and growth of population but also modification of prevalence and distribution of risk factors, mostly related to socioeconomic status. In fact, prediction rise up to 28.8 million cases in 2040, meaning a 47% increase compared to 2020, becoming a growing threat to global health¹.

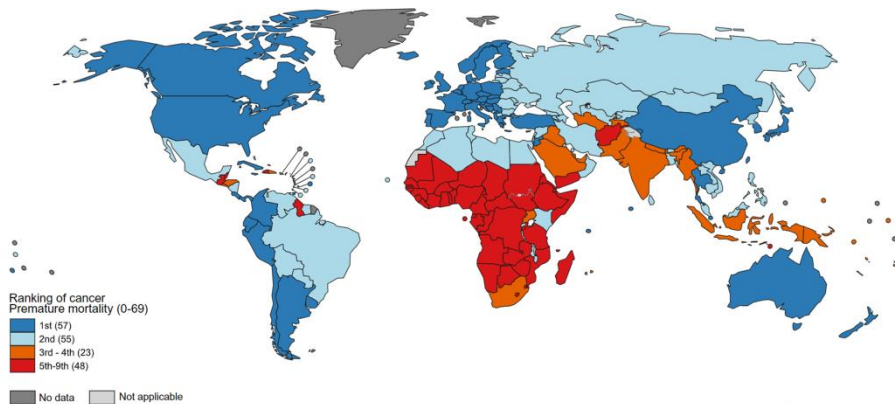


Figure 1: Global distribution of deaths caused by Cancer malignancies (GLOBOCAN 2020)¹. Ranking of Cancer as a Cause of Death at Ages <70 Years in 2019. The legend contains the number of countries included in each level.

1.1. Gynaecological Cancers

1.1.1. Epithelial Ovarian Cancer

1.1.1.1. Epidemiology

Epithelial Ovarian Cancer (EOC) presented 313,959 new cases and causes 207,252 deaths worldwide during 2020². The tendency in Spain is quite similar to those observed in Europe, being the 8th most commonly diagnosed cancer among women and the most lethal gynaecological malignancy with an incidence and mortality of 3513 and 2106 cases respectively (Figure 2) in 2020.

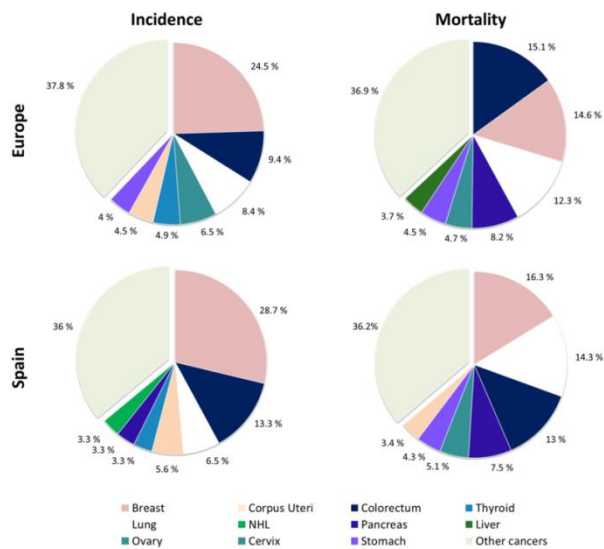


Figure 2: Incidence and mortality stratified by tumor type across global female population (<https://gco.iarc.fr/today/home>, WHO, last access November 2022). Proportion of the total number of cases and deaths for the 8 most common cancers for women in 2020.

1.1.1.2. Diagnosis

Three different cell types exist in the ovary: surface epithelium, stromal cells and germ cells, having the three of them the potential of malignant transformation. Depending on cell origin, the diagnosis, clinical management, and prognosis of arising tumours will be completely different. EOC represents the most common subtype of ovarian cancer (OC), constituting approximately the 90% of all OC cases³.

Clarifying the pathogenesis in early cancer development is fundamental to identify biomarkers of early detection and to define cost-effective strategies for cancer prevention. This task represents an unmet need in those cancers lacking an effective screening strategy like the case of OC⁴.

The diagnosis of EOC constitutes a difficult task for clinicians due to the natural evolution of the disease. This type of malignancy, normally asymptomatic, starts with a series of non-specific symptoms including abdominal bloating, early satiety, nausea, abdominal distension, change in bowel function, urinary symptoms, back pain, fatigue, and weight loss⁵. However, when those symptoms appear, patients already present advanced stage with regional or distant spread in the 75% of cases, dramatically decreasing the 5-year survival rate from 93% in early-stage detection to 29%⁶. Even though there is enough space for huge improvements regarding early

diagnosis, currently there is no established screening test for OC patients, thereby hindering early-stage detection of this disease.

Diagnosis of EOC should combine the evaluation of previous symptoms, risk factors, familiar and personal clinical history of gynaecological or other cancer types, physical exam, laboratory tests⁷ (The most extensively studied tumor marker used in screening for OC is CA-125, a high molecular weight glycoprotein recognized by the murine CA-125 monoclonal antibody as an immunogen⁸) and radiographic imaging. Physical exam and imaging tests of the abdomen and pelvis are usually recommended as an initial step in the diagnosis and will often lead to the identification of a pelvic mass. Testing patients for CA-125 biomarker will also help when there is a suspicion of EOC. Even if CA-125 is not specific of EOC condition (could be elevated in benign conditions or other tumor types), it will support clinical decisions. Detection of serum HE4, a protease inhibitor member of whey acidic four-disulfide core (WFDC) family, overexpressed in OC tissue, is also another potential biomarker for OC⁹. In addition, in 2010 was approved by the Food and Drug Administration (FDA) the ROMA algorithm, a multivariate index which combines both biomarkers, CA-125 and HE4, and menopausal status in a logistic regression model, improving the individual predictive values^{10,11}. In any case, the final diagnosis of ovarian malignancy currently requires pathologic assessment after surgery or biopsy^{12,13}.

1.1.1.2.1. Histopathological classification

Anatomically, EOC is characterized by early intraperitoneal dissemination with seeding of pelvic structures, abdominal organs, and peritoneal surfaces. Molecularly and morphologically, EOC is a heterogeneous group of diseases stratified according to histological subtypes in serous, endometrioid, clear cell, mucinous, malignant Brenner tumours and mixed (Figure 3). These subtypes have different patterns of genomic variation, each carrying different prognostic implications. Serous histology is found as the most common, followed by clear cell and endometrioid. Mucinous, on the contrary, is the less common¹⁴.

Classically, EOC has been grouped as Type I and Type II. Type I tumours include low-grade serous, low-grade endometrioid, clear cell and mucinous carcinomas while Type II comprises high-grade serous, high-grade endometrioid and carcinosarcomas. Clinically, Type I tumors are usually diagnosed at an early stage of the disease while Type II present papillary, glandular and solid patterns. They are highly aggressive and commonly diagnosed at an advanced stage. Type II tumors represent the 75 % of EOC, being High-Grade Serous Ovarian Cancer (HGSOC) the most usual subtype. Regarding the known molecular characteristics, Type I tumors are genetically more stable, occasionally harboring microsatellite instability (MSI) and frequently present mutations in *KRAS*, *NRAS*, *ARID1A*, *CTNBB1* and *PTEN*; whereas Type II tumors are characterized by *TP53* mutation in more

than 80 % of the cases, *CCNE1* amplification and high genomic instability (GI) principally caused by alterations in the Homologous Recombination DNA-Repair machinery (HRR)¹⁵.

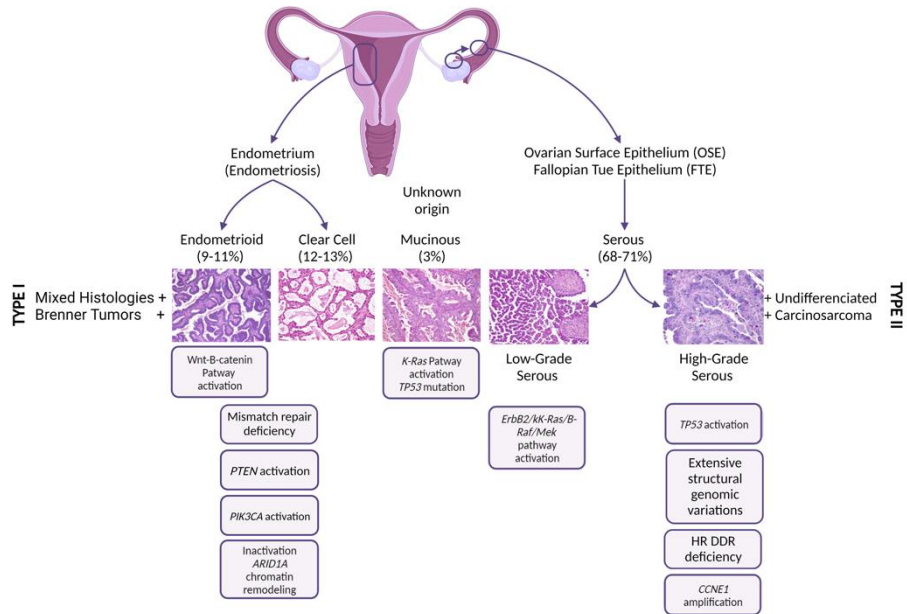


Figure 3: Classification of OC tumors. Clinicopathological-based stratification; differentiation of tumors in Type I and Type II. Information regarding molecular and pathological features specific for each histological subtype are described.

1.1.1.2.2. Staging and grading

While early-stages have similar frequencies independently from the histology, in advanced stages, serous histology is overrepresented. Endometrioid, mucinous and clear cell histologies, however, are normally diagnosed at an early-stage still confined to the ovary¹⁵.

Table 1: FIGO staging system. 2014 FIGO ovarian, fallopian tube, and peritoneal cancer staging system¹⁶.

Stage I	Growth limited to the ovaries
IA	Tumor limited to the ovary, capsule intact, no tumor on surface, negative washing
IB	Tumor involves both ovaries, otherwise like IA
IC	Tumor limited to one or both ovaries
IC1	Surgical spill
IC2	Capsule rupture before surgery or tumor on ovarian surface
IC3	Malignant cells in the ascites or peritoneal washing
Stage II	Tumor involves one or both ovaries with pelvic extension (below the pelvic brim) or primary peritoneal cancer
IIA	Extension and/or implant on uterus and/or fallopian tubes
IIB	Extension to other pelvic intraperitoneal tissues
Stage III	Tumor involves one or both ovaries with cytologically or histologically confirmed spread to the peritoneum outside the pelvis and/or metastasis to the retroperitoneal lymph nodes
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 ± positive retroperitoneal lymph nodes. Includes extension to capsule of liver/spleen
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
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 to the abdominal cavity)

The stratification of tumours according to anatomic location and pathologic characteristics, establishing tumor grade and stage, is an initial key step in oncologic diagnosis. Reproducible histopathologic diagnosis of tumor is essential for successful treatment given that this will determine the clinical course of the disease.

In OC, tumor grade is assessed at two different levels: nuclear and architectural, in a scale from 1 to 3, being the most poorly differentiated part the reference to fix overall tumor grade. Briefly, grade I tumours present mildly enlarged, uniformed nuclei, dispersed chromatin, very small nucleoli and they are well-differentiated with predominant glands content. Grade II tumours show moderately enlarged nuclei, varying in size and shape, small but evident nucleoli and variable mitotic activity. Architecturally, they are moderately differentiated. Lastly, Grade III tumours are characterized by markedly enlarged and pleomorphic nuclei, irregular coarse chromatin, and prominent nucleoli. In addition, mitosis with atypical forms is usually exhibited. These tumours are normally poorly differentiated and they are formed predominantly by solid areas¹⁷.

Regarding staging of tumours, the Tumor-Node-Metastasis (TNM) system is the reference criteria. This classification includes tumor size and local growth (T0 and T4), extent of lymph node metastases (N0 and N3), and occurrence of distant metastases (M0 and M1). Nevertheless, gynaecological cancers have the International Federation of Gynecology and Obstetrics (FIGO) staging system¹⁶, a

staging system, comparable to TNM, which stratifies tumours as detailed in Table 1.

Implementing grading and staging information combined with the remaining clinical data, clinicians can build the final diagnosis, guiding treatment and predicting clinical outcomes^{18,19}.

1.1.1.3. Risk factors

Different risk factors have been described in EOC, however, only a few have proved to have a real impact on the development of the disease (Table 2). Principally, there are four types of risk factors: reproductive factors, the intake of oral contraceptives or/and Hormone Replacement Therapy (HRT) and family and personal cancer history. Reproductive history influences OC risk and is related to exposure to oestrogen and progesterone²⁰. Among factors related to lifetime ovulations, some of them are associated with lower OC risk, such as the use of oral contraceptive, pregnancy and breastfeeding risk^{21,22}, due to the decreased in the number of ovulation cycles. However, those that involve a higher number of menstrual cycles, such as nulliparity, increase the risk²³. In this regard, infertility also represents a negative factor for EOC²⁴. Other risk factors have also been evaluated, failing in demonstrating a real impact on the evolution of the disease. Between them; we can find body complexion (height, weight and body mass), HRT or length of

menstrual life (early age at menarche and late age at menopause)^{25,26}.

Table 2: OC risk factors²⁷. Environmental and personal factors that entail an increase (predisposing), decrease (protective) or controversial risk of developing OC.

Factors		Protective	Predisposing	Controversial
Demographic	Age		X	
Reproductive	Menstrual-related factors		X	
	Age of menarche and menopause			X
	Parity	X		
	Pregnancy characteristics			X
	Higher age of child birth	X		
Gynaecologic	Pelvic Inflammatory Disease			X
	Endometriosis			X
Hormonal	Contraceptive methods		X	
	Hormone Replacement Therapy (HRT)		X	
	Infertility Treatments		X	
Genetic	Family History		X	
	BRCA mutations		X	
	Lynch Syndrome		X	
Lifestyle	Nutrition and Diet			X
	Obesity and physical activity			X
	Alcohol, caffeine and cigarettes			X
Others	Lactation	X		
	Lower socioeconomic status			X

In addition to these, which are commonly found among all histologies, there are others limited to one of them. It is the case of tobacco smoking, associated only with Mucinous Ovarian Cancer²⁸ or endometriosis, associated with 15-20 % of clear-cell and endometrioid histologies²⁹.

Finally, genomic predisposition as well as familiar and personal cancer history represents the most determinant risk factor. OC has been related with different cancer-predisposing hereditary syndromes such as Hereditary Breast and Ovarian Cancer (HBOC) and Lynch Syndrome (LS). In this sense, women with EOC history harbours higher probability of developing this kind of tumours, resulting in a risk three times greater when presenting one affected first-degree relative and even higher if they have been diagnosed before the age of 50^{30,31}. HBOC syndrome is mainly characterized by the presence of mutations in *BRCA* genes, even if they are also present in patients without the syndrome/family history. *BRCA1* has an estimated 40-50% risk of developing OC by age of 70 while *BRCA2* has 10-20%³². Most of these tumours associated with *BRCA* mutations are HGSOE. Other genes also confer a moderately or low risk, predominantly related with HRR pathway, such as *BRIP1*, *RAD51*, *BARD1*, *CHEK2*, *RAD50*, *PALB2* and *ATM*^{33,34}. People with HBOC syndrome may also have an increased risk of other types of cancer, including pancreatic cancer, prostate cancer, and melanoma. LS, however, is characterized by mutations in DNA mismatch repair (MMR) genes, supposing a 3%

to 17% lifetime risk of developing OC, particularly non-serous cancers as clear-cell and endometrioid^{35,36}.

1.1.1.4. Therapeutic approaches

Primary cytoreductive surgery followed by platinum-based chemotherapy has become the standard of care for patients with advanced stage EOC. Achieving no residual tumor (R0) remains as the most important prognostic factor. With this aim, surgical cytoreduction should be done by an experienced gynaecologic oncologist⁵.

Given that EOC is often presented with advanced stage at the time of diagnosis, for optimal cytoreduction, surgical approach should consider including bowel resection and/or appendectomy, diaphragm or peritoneal stripping, splenectomy, partial cystectomy and/or ureteroneocystostomy, partial hepatectomy, partial gastrectomy, cholecystectomy, and/or pancreatectomy. In case optimal cytoreduction is unlikely or it could carry high morbidity and mortality for the patient, neoadjuvant chemotherapy should be considered as a treatment option³⁷.

Fertility conservation surgery could be studied in young woman with early-stage IA or IC disease, unilateral ovarian involvement and favorable histology. In this case, uterus and contralateral ovary will

be left in place. These clinical scenarios should be carefully evaluated, balancing benefits and risk for the patient³⁸.

Regarding systemic treatment, the administration of chemotherapy to early-stage patients has been largely discussed^{39,40}. Hence, only patients that were sub-optimally staged or optimally staged as high-risk (stage IB-C grade 2, any grade 3 or stage IC clear cell histology) demonstrated clinical benefit of receiving systemic treatment⁴¹. In case of EOC staged II-IV, all cases should be treated with chemotherapy after surgery. Standard treatment includes intravenous administration of carboplatin and paclitaxel every 3 weeks, usually for 6 cycles. The combination of paclitaxel and cisplatin is equally effective but carries high toxicity, not being considered a standard of treatment⁴².

Concerning relapse, therapeutically options are chosen based on patient general state, time of recurrence, tumor histology and disease biology. Secondary debulking surgery would be evaluated as a first option⁴³. If there is no surgical option, then, systemic therapy will be used to control the disease and avoid spreading. In the platinum-sensitive recurrence setting, re-challenge with platinum chemotherapy is a standard of care.

Since the arrival of first line therapeutic options that assures better outcome such as the combination of carboplatin plus paclitaxel, radiotherapy (RT) has been reserved to a small number of clinical

scenarios. The first scenario is oligometastatic disease setting, where RT represents a reliable and effective treatment option with a mild toxicity. The second is its usage as palliative treatment. Due to its capacity to significantly decrease abdominal masses, it will properly work to reduce symptoms, predominantly vaginal bleeding and pain relief⁴⁴.

1.1.1.4.1. Targeted therapy and future approaches

In addition to first-line standard treatment, there is a growing range of options that will also play a crucial role in the evolution of the disease. Newly developed targeted therapies aim to improve performance of chemotherapy while decreasing its side effects by combination or maintenance treatment as well as working as an alternative when resistances or toxicities appear.

In EOC, the development of new targeted therapies has been focused on HGSOc subtype, due to its high incidence (75% of all EOC) and specific molecular characteristics. In this regards, two different classes of therapies have been included in the clinical practice. Firstly, the anti-angiogenic drugs, such as Bevacizumab, a humanized monoclonal antibody against VEGF. On the other hand, PARP inhibitors (PARPi) also have strongly entered in the equation. Different clinical trials demonstrated the efficacy of the drug in combination as well as in maintenance in distinct clinical scenarios.

Finally, Bevacizumab is approved in first line in combination with standard of care followed by maintenance until progression. Relapse EOC, both platinum-sensitive and resistant, can also receive Bevacizumab plus chemotherapy followed by maintenance.⁴⁵⁻⁴⁷

On the other hand, PARPi have been successfully implemented in recurrent HGSOC by leveraging inherent defects in HRR pathway present in around 50% of HGSOC⁵. Different drugs have been approved for different clinical scenarios and administered as single agent and maintenance, between them, the most remarkable are Olaparib (first approved PARPi), Rucaparib and Niraparib (Figure 4)⁴⁸⁻⁵⁰.

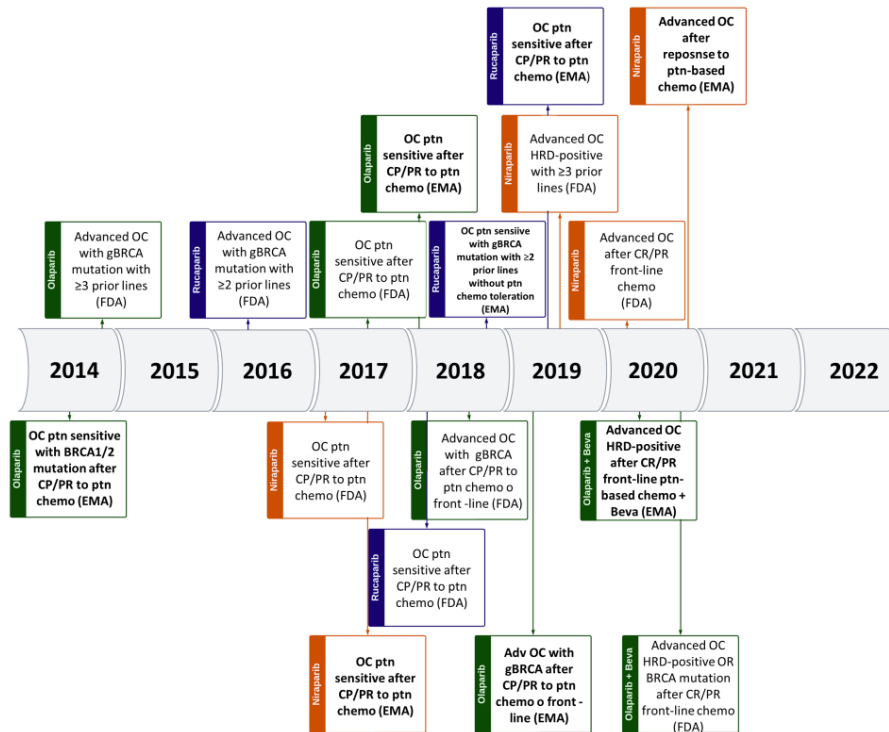


Figure 4: PARPi through the years in OC. Timeline of drug discovery, development, approval and application in clinical routine of different PARPi (Olaparib, Rucaparib and Niraparib) in this particular scenario since 2014. Adapted from Ngoi, N. Y. L., and Tan, D. S. P⁵¹.

1.1.1.5. Molecular biology of EOC: HGSOc

As above mentioned, EOC is a heterogeneous group of diseases that differs from cell origin to histopathological and molecular features, having a direct impact on the prognosis and evolution of the tumor. Together with the advance of science, molecular landscape of EOC is being deciphered and, hence, the term “histotype” is beginning to be

deprecated. Even though the mainstay still depends on conventional histomorphology, complementary molecular tests are being increasingly integrated into clinical decision-making scenarios. Due to HGSOE is the most common subtype of EOC and the main histology studied on this research, from now on the information is going to be focused on this entity.

One of the biggest projects, which work as a starting point on the high-throughput molecular biology studies regarding oncology, is The Cancer Genome Atlas (TCGA). The TCGA established the molecular basis of different tumors, performing thorough studies at different levels and working as a hypothesis generator to following researches.

In 2011, the TCGA published the first integrated analysis of EOC⁵². This study showed a huge mutational heterogeneity and low recurrent gene mutations in this tumor. The highest mutational ratio is presented by *TP53*, the gene that encodes p53 tumor suppressor protein. This gene is the most frequently mutated among tumors and it is altered on approximately the 90% of HGSOE. That shows the essential role of *TP53* in the course of the disease, probably working as a driver mutation⁵³.

In contrast with the relative lack of point mutations across HGSOE patients, it has been described a large and recurrent Somatic Copy Number Aberrations (CNV). Among the 63 regions of punctual recurrent SCNAs, those involving *CCNE1*, *MYC* and *MECOM* were the

most common focal amplifications⁵². Particularly, *CCNE1* has been found to be amplified in HGSOC and it has shown a strong correlation with platinum resistance^{54,55}. The same amplified region of *CCNE1*, 19q12, also implicates the amplification of another anti-apoptotic oncogene, *C19orf2*, that has been related to platinum resistance as well^{56,57}. The incidence of directly follows *CCNE1* and it is mutated or deleted in 10.8 % of the patients⁵².

The other group of genes that were described as characteristically altered in HGSOC was the HRR genes. Until now, *BRCA1* and *BRCA2* have been considered as the directors of the orchestra. However, it has been shown that, additionally to well-known *BRCA* genes, there are other alterations directly implicated in the deficiency of HRR pathway (Figure 5). Alterations in this pathway, due to the inability of cells that carry them to repair DNA double-strand breaks (DSBs), are the main responsible of high rate of platinum sensitivity in HGSOC. This phenomenon, due to common clinical and tumoral effects is known as “*BRCAness*”^{58,59}. Homologous Recombination Repair Deficiency (HRD) can be caused by alterations at both somatic and germline level, including mutations, epigenetic silencing, and big rearrangements. Recent studies have shown a germline mutation rate of *BRCA1/2* around 23% and somatic mutations can also be found in approximately 11% of patients^{60,61}. Additionally, other HRR genes have also been involved, including: *RAD51*, *RAD54*, *DSS1*, *RPA1*, *NBS1*, *ATR*, *ATM*, *CHEK1*, *CHEK2*, *FANCD2*, *FANCA*, and

*FANCC*⁶². Apart from those genes, *PTEN* loss was also related to deficiency in HRR pathway. However, the result of this alteration is still under discussion⁶³⁻⁶⁵. Finally, amplification of *EMSY* was described in 8 % of sporadic HGSOc⁶⁶. Due to its role as an endogenous transcriptional repressor that interacts with the N terminus of *BRCA2*, silencing its activation, *EMSY* has a determinant role in OC. It is exclusively nuclear and relocalizes to sites of DSB following DNA damage⁶⁷. Since overexpression/amplification of *EMSY* and deletion of *BRCA2* have a similar effect, it has also been reported in association with worse survival⁶⁸.

Summarizing, the subset of patients that carry HRD represents near the 50 % of diagnosed HGSOc and that is why this alteration depicts a huge impact biomarker that need to be further studied.

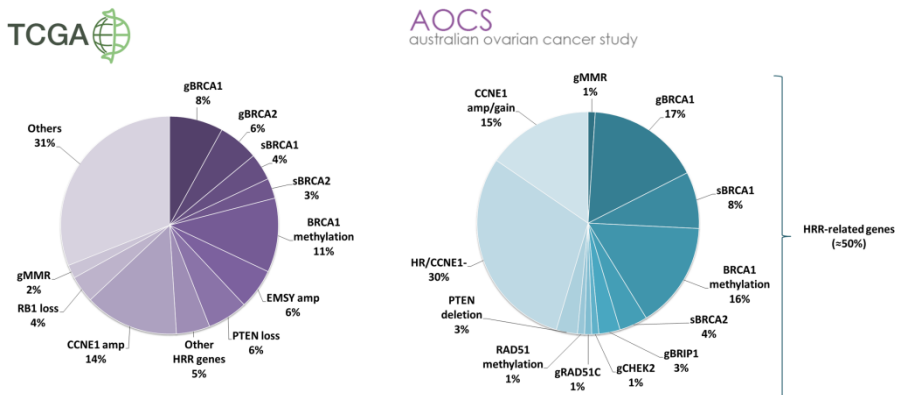


Figure 5: Mutational landscape of HGSOc. Distribution of molecular alterations described in HGSOc in the context of comprehensive studies (TCGA and AOCS project)^{52,69}

1.1.2. Endometrial Cancer

1.1.2.1. Epidemiology

EC represents the most common gynecological malignancy and the fourth cancer regarding incidence in women population, accounting with more than 417,000 cases worldwide. Mortality rate, however, is much lower than previously mentioned OCs, with 97,370 deaths in 2020² (Figure 2).

1.1.2.2. Diagnosis

The diagnosis of EC, due to the appearance of early symptoms, results less challenging than other gynecological malignancies. Hence, the 67 % of women with EC are diagnosed at initial stages, with disease still confined to the uterus and only 21 and 8 % of cases present regional and distant metastasis respectively⁷⁰. However, an increase in mortality rate is occurring, probably as a result of a rise in advanced-stage cancers and high-risk histologies throughout the diagnosis of the disease⁷¹.

The main symptom that prevents women of a possible EC is abnormal uterine bleeding, often found in postmenopausal period. Even though the most suitable diagnostic technique is still under discussion, there is a recommended way to follow when diagnosing EC. Endometrial biopsy is the first approach, normally providing enough information to the process. When the biopsy resulted

negative, in symptomatic patients, fractional dilation and curettage (D&C) is advisable. Additionally, hysteroscopy could be helpful when looking for endometrium lesions. Other imaging tests, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and/or Positron Emission Tomography (PET)/CT will be helpful for detection of distant metastasis and evaluation of tumor spreading^{70,72}.

1.1.2.2.1. Histopathological classification

EC is classically grouped in Type I and II tumors based on pathologic information, including histologic type, tumor grade, stage and lymphovascular and myometrial invasion. Nine different EC histologies have been described, being distributed between both groups (Figure 6). Endometrioid and serous histologies represent most of them. Hence, sporadic endometrial carcinomas, around 70-80%, are mainly designated as type I carcinomas⁷³.

Type I tumors are characterized by hyperestrogenism by association with endometrial hyperplasia, obesity, hormone-receptor positivity, and favorable prognosis. At molecular level, *PTEN* mutations have been described as an early event. MSI is also typical of this subtype, being present in a third part of them. Most tumors show endometrioid differentiation and presented low grade. In addition, mucinous histology is also considered as type I tumors⁷³.

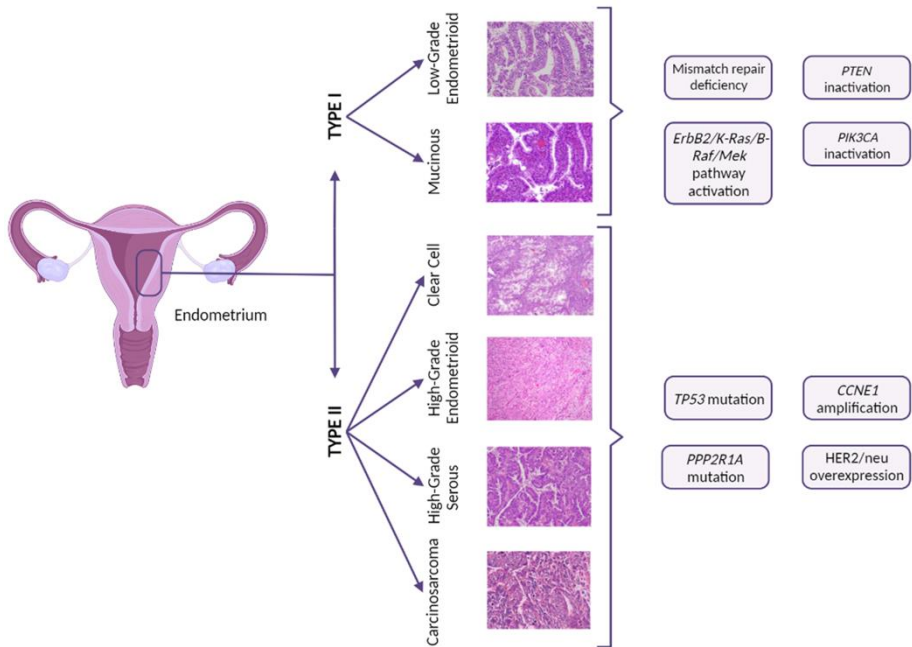


Figure 6: Classification of EC tumors. Clinicopathological-based stratification; differentiation of tumors in Type I and Type II. Information regarding molecular features specific for each histological subtype are described.

Type II tumors, however, are typically diagnosed in non-obese older women and associated with atrophic endometrium. They present infrequent MSI and mutation mainly in *TP53*, followed by inactivation of p16 and e-cadherin together with the amplification of *her2/neu* oncogene. These tumors are normally high-grade carcinomas of non-endometrioid differentiation, principally serous and clear cell, with an aggressive clinical course and poor prognosis. Small cell,

undifferentiated and squamous cell carcinomas are also classified as type II⁷⁴.

1.1.2.2.2. Staging and grading

Cancer staging and grading are fundamental pathologic determination with a huge importance in clinical management of cancer patients, representing a key prognostic and predicting factor.

Regarding stage, in EC it is generally assessed according to the FIGO classification. Even if a new staging system was published on 2009 (Table 3), the information here found is mainly based on the version of the system from 1983⁷⁵. Surgical staging, now implemented instead of classical clinical stage, allows a more accurate assessment of extent of disease. Complete surgical staging should include total hysterectomy with bilateral salpingo-oophorectomy and pelvic and paraaortic lymphadenectomy.

After tumor staging, the most informative prognostic feature is the grade, especially the differentiation between grade 1 and 2 tumors and grade 3 tumor, considered low and high-grade tumor respectively in a binary classification. Currently, FIGO grading system divides tumors between 3 grades, based on degree of glandular differentiation. Grade 1 tumors exhibit $\leq 5\%$ solid non-glandular, non-squamous growth; grade 2 tumors from 6% to 50%; and grade 3 tumors $>50\%$. However, this grading system is only applied to

Endometrioid and mucinous ECs. The remaining histologies, due to its pathologic characteristics, present an intrinsic high-grade^{76,77,78}.

Table 3: FIGO Staging of Uterine Corpus Carcinoma and Carcinosarcoma. Based on staging established by the International Federation of Gynecology and Obstetrics (FIGO) update of 2021⁷⁹

Stage I	Tumor confined to the corpus uteri
IA	No or less than half myometrial invasion
IB	Invasion equal to or more than half myometrium
Stage II	Tumor invades cervical stroma but does not extend beyond the uterus
Stage III	Local and/or regional spread of the tumor
IIIA	Tumor invades serosa of the corpus uteri and/or adnexae
IIIB	Vaginal and/or parametrial involvement
IIIC	Metastasis to pelvic and/or para-aortic lymph nodes
IIIC1	Positive pelvic nodes
IIIC2	Positive para-aortic lymph nodes with or without positive pelvic lymph nodes
Stage IV	Tumor invades bladder and/or bowel mucosa and/or distant metastasis
IVA	Tumor invasion of the bladder and/or bowel mucosa
IVB	Distant metastasis including intra-abdominal and/or inguinal lymph nodes

1.1.2.3. Risk factors

The protective and risk factors related to the development of EC are still under discussion. While some of them present a high confidence level, others present weak significance. Obesity and other metabolic syndromes such as diabetes or polycystic ovary syndrome were found to be risk factors⁸⁰⁻⁸². Other factors also play a role in the

predisposition of women to suffer EC, such as excess of estrogen due to natural causes or HRT. Tamoxifen, classical treatment of breast cancer has also demonstrated to be a risk factor, increasing two times the risk and even four times when administered more than 5 years⁸³⁻⁸⁵. On the contrary, two factors presented protective role to EC development. These are parity and the usage of oral contraceptives, which achieve a reduction of risk by 30 to 40% and the protective effect can increase with longer period of use^{86,87}.

Even if sporadic cancers represent most of ECs, there is also an approximately 5 % that are caused by a hereditary component/genetic predisposition. These cancers normally appear between 10 and 20 years before sporadic cancers and they are mainly related to Lynch Syndrome (LS) and Cowden syndrome⁸⁸. LS related-tumors are characterized by mutations in MMR genes, which are *MLH1*, *MSH2*, *MSH6*, or *PMS2*. The lifetime risk of developing ECs for patients that carry mutations in those genes is ranged between 32-60⁸⁹. Due to its increased risk, screening should be considered in all patients with EC⁹⁰. On the other hand, there is also a percentage of patients which harbor mutations in *PTEN*, main characteristic of Cowden syndrome⁹¹. In this case, the risk of developing EC is up to 19-28% by age of 70⁹².

Lastly, family history of EC has been associated with a two-to-threefold increased risk of developing the disease⁹³. Recent studies have shown further EC susceptibility candidate genes independently

from LS-associated genes, such as *POLD1*, *POLE*, *NTHL1*, *MUTYH* and *BRCA1*^{78,94,95}.

1.1.2.4. Therapeutic approaches

Surgery plays the main role in the initial treatment of EC. Classically, surgery has been preceded by RT, however, this therapeutic approach has been replaced by initial surgical staging of patients, aiming to avoid unnecessary RT. Hence, initial surgery of EC consists on a total hysterectomy with bilateral salpingo-oophorectomy (Figure 7)⁹⁶. The role of lymphadenectomy, even if lymphatic dissemination is the main route of spread in EC patients, is still under discussion for early stages. Some authors have reported this approach to be associated with improved survival and providing important diagnostic information⁹⁷⁻¹⁰¹.

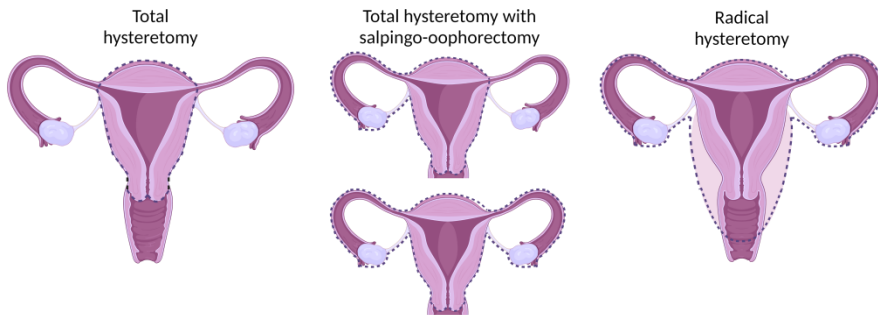


Figure 7: Available cytoreductive approaches for patients with EC. In accordance with clinical recommendations, different types of surgeries, from less to more radical, could be performed depending on tumor nature and personal preferences.

Surgical approach can be done by different methodologies, between them; laparotomy, laparoscopic-assisted vaginal hysterectomy, total laparoscopic hysterectomy or robotic total hysterectomy with pelvic and paraaortic lymphadenectomy¹⁰². Laparoscopic, a minimally invasive procedure has been associated with longer operating time but faster recovery, shorter hospital staying and lower morbidity¹⁰³. Therefore, most gynecologic oncologists recommend this approach when appropriate and feasible.

There are some clinical exceptions to the standard treatment. Only in specific cases of medical comorbidities, vaginal approach will be followed. Primary radiation will also be considered in non-operable tumors¹⁰⁴. For serous and clear-cell subtypes is recommended an omentectomy along with peritoneal and upper abdominal biopsies, similar to OC. Finally, for those patients that have not completed childbearing or that are diagnosed at a younger age, preservative surgery could be considered. Uterus and ovarian preservation will be performed in carefully selected patients to avoid surgical menopause or allow completing a pregnancy. In these cases, tight follow-up is mandatory since recurrences are a common event¹⁰⁵⁻¹⁰⁷.

After initial surgery, approximately the 75% of patients will be classified as early-stage disease. For patients with this presentation of the disease no adjuvant therapy is recommended due to its low rate of recurrence excepting from Intermediate risk scenario, where vaginal brachytherapy has proven its efficacy¹⁰⁸⁻¹¹⁰.

The remaining 25%, classified as advanced stage may need additional chemotherapeutic or radiotherapeutic treatment⁷⁷. Adjuvant chemotherapy, based on cisplatin, doxorubicin, and paclitaxel combination, is now the mainstay of treatment for patients with stage III and IV EC.

Regarding the recurrent disease scenario, EC presented a highly heterogeneous manifestation, from isolated vaginal relapse to widespread disease. Hence, treatment will depend on multiple features such as original stage, location of the recurrence and previous treatments. Surgery, radiation, chemotherapy, and hormonal therapy are all a clinical option. Whole pelvic radiation and vaginal brachytherapy will be the chosen option for patients with relapse at vaginal level, as long as no RT line has been previously received¹¹¹. For systemic disease, hormonal therapy has been presented as an option due to its low toxicity rate. Combine chemotherapy of cisplatin, doxorubicin and paclitaxel can be used as first line treatment for advance disease as well as for recurrence after hormonal therapy. Additionally, radical surgical approaches could be applied to carefully selected patients with locally advanced disease for whom cure would be possible^{112,113}.

1.1.2.4.1. Targeted therapy

There is an unmet need of suitable and effective treatments for EC capable of directly target defects of tumor cells, achieving the development and introduction of personalized therapies in the clinical scenario. In the last few years, a huge number of studies have refined molecular characterization of EC, leading to the development and use of new targeted therapies and, consequently, improving clinical outcome of patients. As a result of these studies, a large number of gene abnormalities and/or pathway deficiencies have been shown as promising targets.

Among the investigated targets and drugs, antiangiogenic and mTOR/PI3K pathway inhibitors have demonstrated clinical activity in different studies and still continue to be under research^{114,115}. Other targeted pathways are showing interesting results, those include glucose metabolism by metformin, EGFR family and Cell-cycle through Cyclin-Dependent Kinases^{116,117}.

However, the most promising and interesting targetable pathway regarding this study are those that involved DRR pathways and immune-related pathways^{118,119}. In the field of immunotherapy, response in ECs has been related to phenotypes characterized by POLE and MSI-H, groups defined as hyper and ultra-mutated in TCGA. In this regards, the main available immunotherapy options are Dostarlimab¹²⁰ and Pembrolizumab¹²¹, checkpoint inhibitor that targets the PD-1/PD-L1 pathway for patients with recurrent or

advanced EC harboring dMMR, MSI high or high tumor mutational burden (TMB). However, there are a wide range of other strategies under discussion that sooner than later will burst into clinical routine.

The other promising target, even if there is no approved option yet is the HR pathway. This pathway presents a key role in the maintenance of GI, working as a nexus between DNA repair and DNA replication. Opposite to HGSOC, high SCNAs or HRD phenotypes have been correlated with adverse outcome in EC, showing the promising role and likely improvement of the use of PARPi in this clinical scenario¹²². In addition, several clinical trials are on-going to evaluate the strategic role of the combination between PARPi and other targeted drugs in EC based on preclinical data¹¹⁹.

1.1.2.5. Molecular biology of EC

Like in the case of EOC, the TCGA project, that globally characterized the EC, marked a milestone in the evolution and knowledge of these tumors. The integrative analyses performed by the consortium based on genomic, transcriptomic and proteomic data using array and sequencing technologies achieve new molecular-based classification that stratified patients in four prognostic groups: *POLE ultramutated*, *MSI hypermutated*, *copy-number low (CNL)* and *copy-number high (CNH)*⁶⁶. The reclassification of EC tumors has an impact on the

treatment scheme of patients, directly affecting the clinical course and management of the disease (Figure 8).

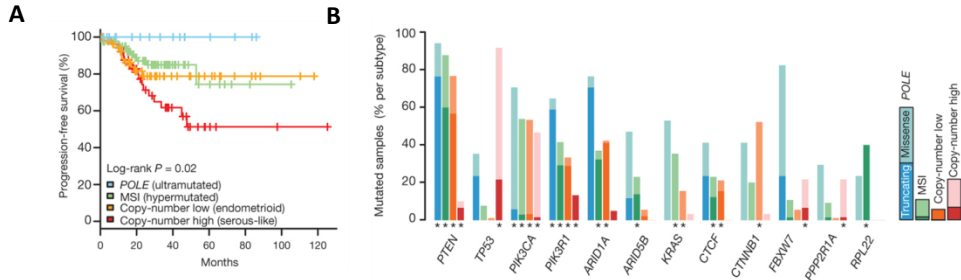


Figure 8: Molecular-based classification from EC-TCGA⁶⁶. A) Integrative Prognostic classification of EC into four groups, from best to worse prognosis: POLE, MSI, CNL and CNH. B) Incidence of 12 most-altered genes across four prognostic groups.

The first group and the most promising one was the POLE ultramutated. The main characteristic of this group is the presence of mutations in *POLE*, gene that encodes the major catalytic subunits of the DNA polymerase epsilon enzyme complex and is involved in replication and repair of the genetic material. Mutations tend to occur in the Exonuclease Domain (EDM), mostly V411L and P286R hotspots, described in 5-8% of EC. These mutations resulted in one of the highest TMB found across human cancers, leading to high neoantigen load and immune-rich microenvironment¹²³. On the contrary, neither CNV nor MSI are commonly found. Other genes frequently mutated in this sub-group are *PTEN* (94%), *PIK3R1* (65%),

PIK3CA (71%) and *KRAS* (53%). Regarding clinical and pathological data, tumors grouped as POLE are younger patients with preferentially endometrial histology, high tumor grade and favorable clinical outcome.

The second group, the MSI hypermutated, presents MSI (mainly as result of *MLH1* promoter hypermethylation) and, like the POLE group, high TMB and low CNV. *KRAS* is frequently altered in this subtype while *TP53* is rarely found mutated. This group is formed by low and high-grade carcinomas.

The third group, in order from better to worse prognosis, is the CNL. Tumors belonging to this subtype are those with no *POLE* mutation or MSI presenting low number of CNV. Normally, these tumors harbor mutations in *PTEN* and *CTNBB1*.

Finally, the CNH group shows the worst prognosis. This subtype is characterized by high number of CNV, low number of mutations and the greatest transcriptional activity through cell cycle deregulation. The gene most altered is *TP53*. Concerning pathological features, tumors classified as CNH are mostly serous and high-grade (25%)^{73,124}. This group is also of particular clinical interest due to its similarities to HGSOc subtype. TCGA found analogous CNV patterns, transcriptomic activity and *TP53* mutation, what could mean new opportunities for overlapping treatment paradigms, such as the administration of iPARPs in serous-like EC.

Due to the clinical implication of this new classification and the methodological and economic difficulties to translate it into the diagnostic routine, different research groups have developed simplified approaches to reproduce the prognosis stratification of patients (Figure 9)¹²⁵. The first one, called the scheme ProMisE (Proactive molecular risk classifier for EC), which consisted on the sequential testing of MMR proteins by immunohistochemistry (IHC), sequencing to detect POLE-EDM mutations and final determination of p53 abnormalities by IHC^{126,127}. Parallely, Stello and colleagues developed another classifier inside PORTEC and transPORTEC clinical trials, with similar aim to ProMisE work, stratifying patients as *POLE* mutated, MSI high, *TP53* mutant and non-specific molecular profile^{128,129}. Even if both approaches fail to totally reproduce TCGA classification and presented some caveats due to applied methodologies, achieve similar prognostic classification of EC¹³⁰.

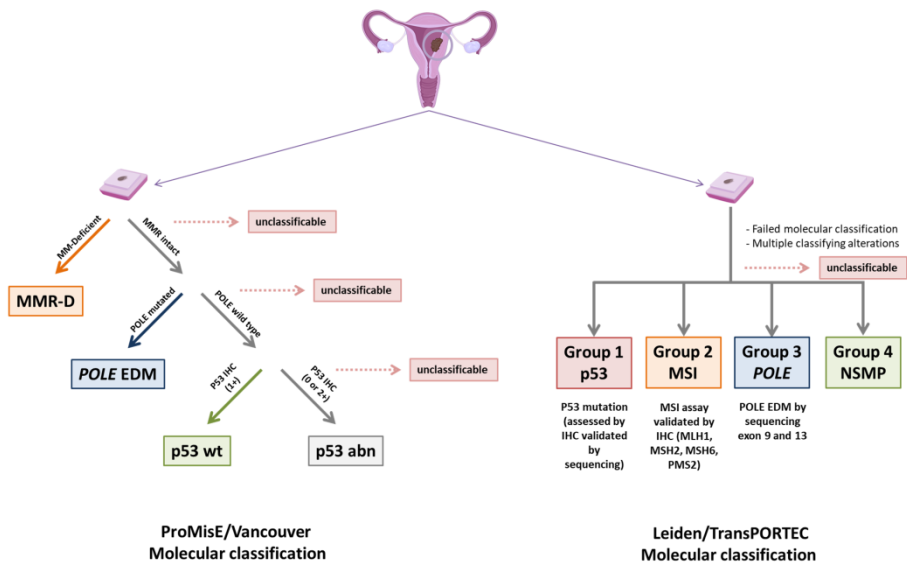


Figure 9: Clinically-validated surrogates of the 4-groups prognostic EC-TCGA classification. Two main simplified approaches based on IHC and molecular data achieve prognostic significance, reproducing TCGA stratification; PROMISE^{126,127} and PORTEC^{128,129} algorithms.

1.2. Genomic Instability

1.2.1. Genomic Instability and cancer (leading causes)

GI is defined as a process of genomic changes or an increased tendency of alterations in the genome during life cycle of cells, working as a driving force of tumorigenesis. Hence, GI, as a hallmark of cancer confers to the tumor cell an advantage above the others and promotes tumor development and progression¹³¹. Even though GI is inherent characteristic of a cancer cell, it varies dramatically between different tumor types. Particularly, some subtypes display

highly instable phenotypes, being characterized by global aneuploidy, amplifications, deletions, loss of heterozygosity (LOH), homozygous deletions, translocations and inversions between others¹³². The presence of a high GI and its underlying cause opens a window of new therapeutic opportunities, taking advantage of an innate genomic feature of tumor cells to combat them.

1.2.2. Leading causes of genomic instability

All human cells have a complex network of interlinked pathways responsible of the maintenance of genome integrity and repair of DNA damage. Throughout the cell cycle, spontaneous and induced DNA damage are going to inevitable occur¹³³.

Among DNA lesions, single-strand breaks (SSBs) represent the vast majority (75%). However, DSBs, less frequent and much more dangerous for cells, could appear by evolution of SSBs¹³⁴. In order to control and repair these type of damage, cells present a wide variety of DNA-damage response (DDR) pathways aiming to recover genome integrity (Figure 10)¹³⁵. The existence of different repair pathways enables the fixation of each type of DNA break with a particular approach. In case of SSBs, generally generated from oxidative damage, at basic sites or from erroneous activity of the DNA TOP1 enzyme^{136,137}, the pathway in charge are Base-Excision Repair (BER), when fixation of modified bases is needed, Nucleotide-Excision Repair (NER) for removal of DNA bulky adducts and MMR,

responsible for mis-incorporated bases during replication. All of them based their repair mechanisms on the excision of the damaged region and insertion of bases to fill the gap. On the other hand, DSBs, typically induced by various chemical and physical DNA damaging agents¹³⁸, are repaired by two distinct pathways: error-free HRR and error-prone Non-homologous end joining (NHEJ) (Figure 11). HR performed the repair of the damage by exchanging equivalent regions of DNA between homologous or sister chromosomes whereas NHEJ did not use any template, commonly leaving insertion and deletion at the breakpoint¹³⁹.

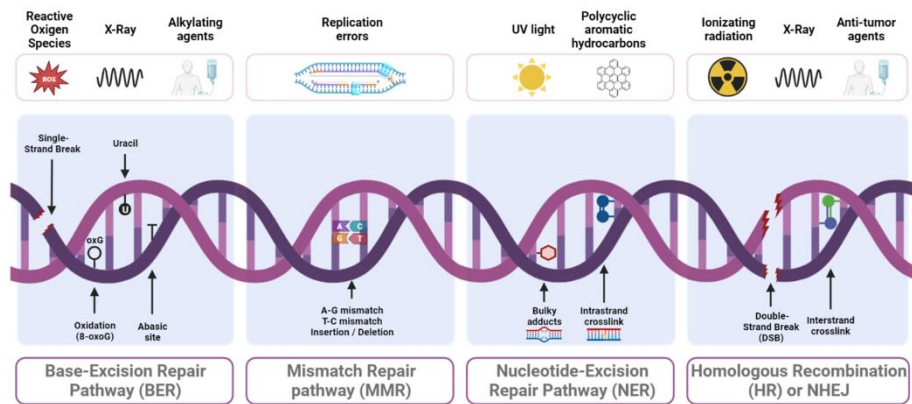


Figure 10: Overview of main DNA lesions and their related DDR pathways. Specific types of DNA damage (SSBs, abasic sites, oxidation, uracil addition, mismatch bases, bulky adducts, intrastrand crosslinks, DSBs, interstrand crosslinks) activate specific DDR pathways (BER, NER, MMR, HR and NHEJ) and signalling cascade.

It is well established that GI mainly arises from dysregulation of these pathways and/or genotoxic stress, originated from cellular processes overwhelming the DNA repair machinery. Alterations in DDR pathways that lead to a different high GI phenotypes in tumors have been extensively studied¹⁴⁰. A good example of a highly instable tumors are *BRCA* associated breast and ovarian cancers¹⁴¹.

Particularly, in these tumors, GI has been mainly associated with defects in the HR pathway, where *BRCA1* and *BRCA2* genes play a central role coordinating the repair of DSBs. However, there are also other genes involved in HRR which alterations can cause a deficiency in the pathway and hence high GI profile¹⁴².

To repair the damage, this pathway requires a homologous sequence that will serve as a template for the repairing process. This allows the recombination machinery to restore any missing genetic information in the vicinity of the break site, using the sister chromatid as the repair template. This restricts recombination to cell cycle stages when the sister chromatid is available, which includes the S and G2 phases. HR is capable of repairing both one- and two-ended DSBs¹⁴³.

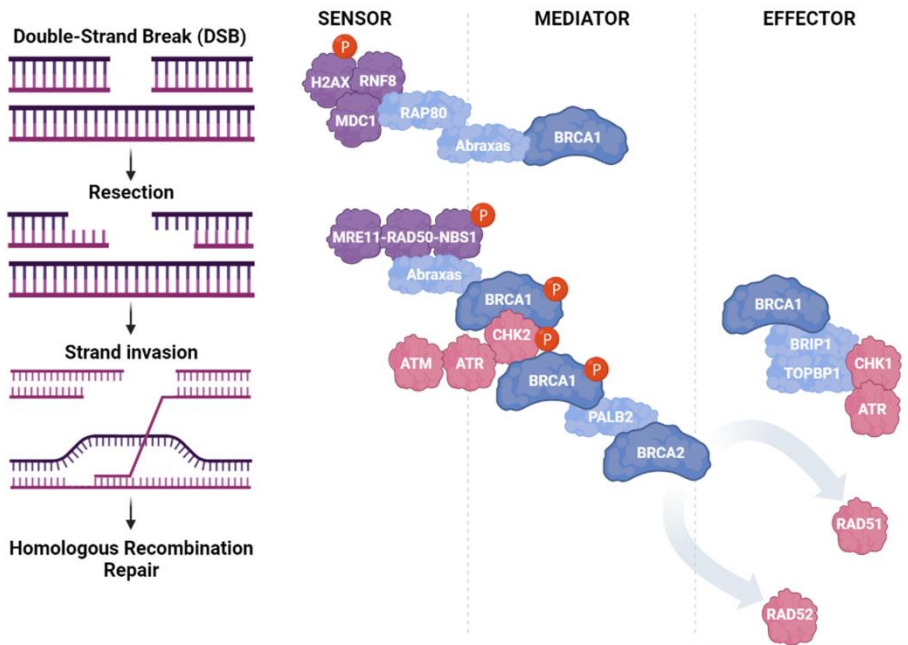


Figure 11: Schematic representation of HRR pathway and its participants. Homologous recombination plays essential roles in the repair of DNA double strand breaks (DSBs). The fundamental reaction in homologous recombination is the exchange of strands between a single-stranded DNA and a homologous double-stranded DNA. In response to DSBs, sensors detect the damage, and signalling mediators recruit or activate effectors that repair the damage. BRCA1-complexes are crucial mediators of the DNA damage response. The CtIP complex associates with the MRN complex, senses DSBs and is responsible for DSB resection. The BRCA1–PALB2–BRCA2 mediates RAD51-dependent HRR. The BRCA1–TOPBP1 complex is associated with DNA repair during replication and may help mediate ATR–CHK1 signalling. DNA damage is also recognized by ATM and ATR kinases, which mediate signalling to form macro-complexes and activate cell cycle checkpoints. Adapted from Macmillan Publishers Ltd: Nature Reviews Cancer¹⁴⁴.

1.2.3. Clinical implications of genomic instability

In the field of oncology there is a growing interest on personalized medicine, aiming to improve treatment outcome by matching specific biological characteristics of a tumor with the most appropriate therapeutic option. For that reason, describing HR defects and being able to use GI as a surrogate of that has generated the appearance of new therapeutic promising approaches with a huge clinical impact. Such a complex biomarker as the GI reflects the advance in the use of Next-Generation Sequencing (NGS). Not only a number of point mutations in HR genes will be identified but also GI based on SCNAs, a pattern of lasting and detectable “genomic scars”¹⁴⁵.

As above mentioned, detecting GI has direct implications in patient prognosis as well as patient management, specifically regarding the choice of therapeutic agents⁵¹. For that reason, different algorithms have been being developed aiming to identify tumors with high GI. Hence, patterns of GI could be applied as predictive and prognosis biomarker, stratifying patients according to predicted sensitivity to DNA damaging agents as platinum-based chemotherapeutics and PARPi. These studies have revealed that the proportion of tumors carrying this phenotype is much higher than previously described¹⁴⁶.

Considering its importance in diagnosis and treatment stratification, many studies have focused on defining clinically relevant surrogate markers of HRD. In this regard, in 2012, three signatures, each measuring a specific type of genomic scars, were published¹⁴⁷. The

first group proposed *Telomeric Allelic Imbalance* (TAI), chromosomal aberrations that involved the presence of sub-telomeric regions with allelic imbalance, which starts beyond the centromere and extend to the telomere. High levels of TAI were associated with platinum-based chemotherapy sensitivity in Triple-Negative Breast Cancer (TNBC) and HGSOC cohorts¹⁴⁸. The following approach was to determine the number of *Large-Scale Transition* events (LST), chromosomal breaks between adjacent regions of at least 10 Mb. This study achieved to predict the status of BRCA1 based on genomic features in basal-like subtype of invasive ductal breast carcinomas¹⁴⁹. Finally, HRD score based on the detection of LOH events, considering regions with LOH larger than 15 Mb but shorter than the whole chromosome, was also presented as a candidate biomarker. In this case, the HRD score was strongly associated with HRD status. The correlation of HRD score and HR deficiency was validated in two independent EOC data sets, as well as breast and pancreatic cancer cell lines¹⁵⁰. All the signatures were based on Single-Nucleotide Polymorphism (SNP) data.

The measurement of all three genomic scar-based signatures exposes an underlying deficiency of DNA repair. Hence, the combination of them will provide a more comprehensive view of genomic landscape of the tumor¹⁴⁷. These findings sooner promoted the appearance of commercial solutions, between them; the two most accepted in clinical use are from Myriad Genetics and Foundation Medicine (Figure 12).

Myriad Genetics solution, myChoice[®]CDx, determines HRD status by detecting *BRCA1* and *BRCA2* (sequencing and large rearrangement) variants and assessing the three genomic scars: LOH, TAI and LST. Foundation Medicine, for its part, developed FoundationOne[®]CDx, a test that detect and quantify LOH, in addition to TMB and mutational status of a wide gene panel. Both tests are suitable for formalin-fixed paraffin-embedded (FFPE) OC tissue samples and ready to use in diagnostic routine, as specified in its corresponding technical information datasheet (myChoice[®]CDx: <https://bit.ly/3eM4Xud> and FoundationOne[®]CDx: <https://bit.ly/3rQpXDU>).

Nevertheless, commercial solutions have two main problems, on one hand, the lack of control about what is being done over the tumor samples and the complete raw data obtain from them. On the other hand, and not less important, the excessive prices that have to be paid for these tests. For that reason, a suitable and simpler academic approach to be perform in the lab setting is something that should be look for.

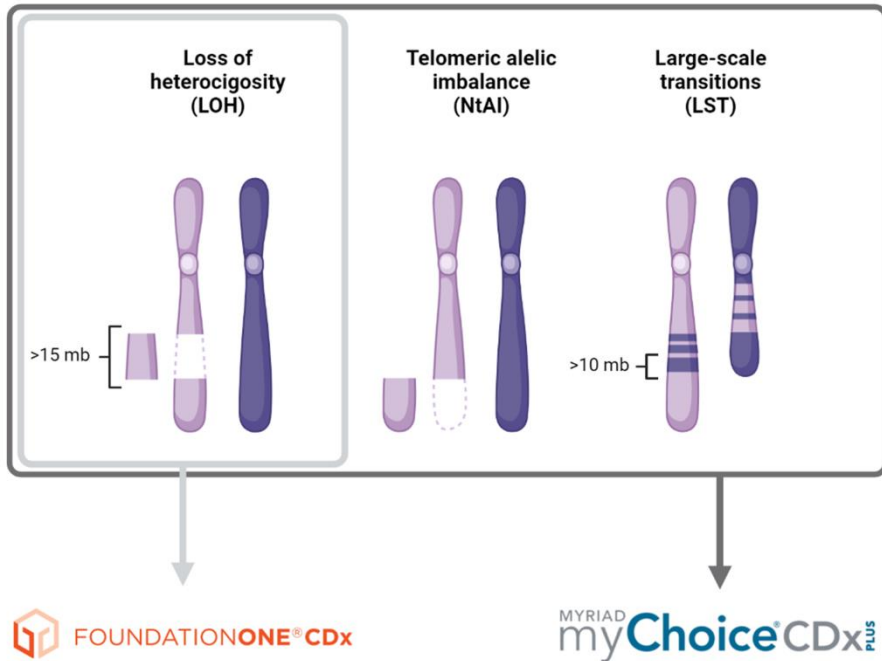


Figure 12: Commercial solutions to establish HRDscore. Clinically-validated approaches from FoundationOne and Myriad based on single or multiple-genomic scar parameters as a surrogate of GI to predict PARPi response.

1.2.4. Methodological approaches to determine deficiencies in DDR pathways

Deciphering GI profiles as a surrogate of HRD requires genome-wide approaches that capture alterations at different levels (Figure 13). In this sense, a huge number of tools have been described so far, not achieving the category of gold standard any of them. From directly sequencing gene-causing disease, such as *BRCA1/2* or multi-genes panel, simpler and less comprehensive approaches, to Whole-

Genome Sequencing (WGS) based on the detection CNV to detect genomic scars, there is a wide range of available options that could be adapted to each particular situation. However, only a few have been clinically validated. While some approaches, such as targeted panels and larger-scale panels are now incorporated in diagnostic routine, a plethora of other more complex methodologies are increasingly gaining force in this field⁵¹.

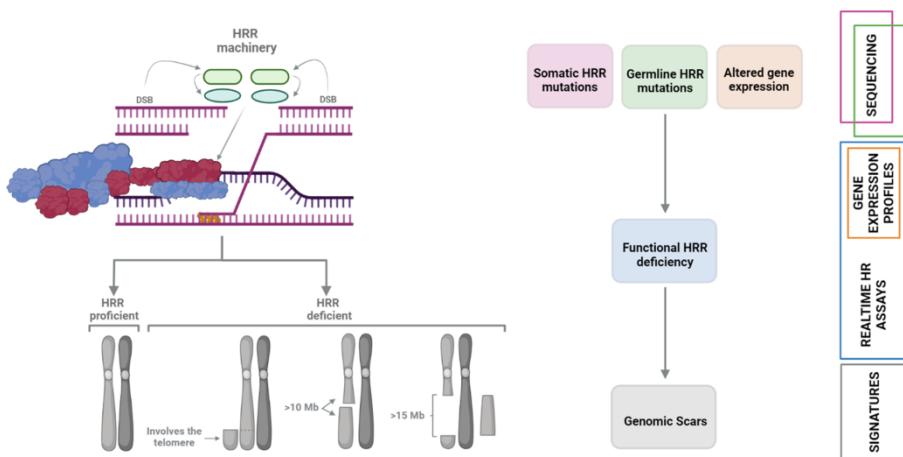


Figure 13: Methodological approaches to detect defects in the HR pathway. Available test examine the different levels of regulation to determine the status of the pathway, interrogating both the causes and the consequences of the deficiency.

Detecting *BRCA1/2* gene mutation is the standard strategy when screening HRD in OC, both at germline and somatic level. Because of *BRCA* alterations over the HR pathway, it is logical to think that tumors with lack of function of other HR components would present

similar response to DNA damaging agents. In fact, there have been described other HR genes as potential biomarkers of HRD, correlating their mutations with clinical sensitivity. Between them *PALB2*, *BARD1*, *BRIP1*, *RAD52B*, *RAD51C*, *ATM*, *CHEK2*, *FANCE* and *FANCM*. However, the low frequency at where the alterations are found together with the unknown functional impact of the majority, cause difficulties when using them as a predictive biomarker¹⁵¹. For that reason, efforts have been focused on the development of most comprehensive and consequence-showing approaches to determine HRD.

Looking for a suitable methodology to detect the consequences of an HRD phenotype independently from the component of the pathway that is affected, technologies able to determine whole-genome CNV where developed. Hence, “genomic scars” that this deficiency leaves the cancer genome could be quantify and used as a biomarker. Currently, “genomic scars” are determine by different high-throughput genomic profiling techniques including array-based comparative genomic hybridization (aCGH), SNP genotyping and NGS¹⁵².

Detection of SCNAs had been performed by cytogenetic techniques such as Fluorescent *In Situ* Hybridization (FISH) or Karyotyping. However, differences in the type of alteration that needed to be detected, the current methodologies did not present enough resolution. Hence, techniques able to detect SCNAs at higher

resolution started to appear in the research context. Firstly, array-based techniques such as the aCGH (100 kb) and later on SNP array (20 kb) were developed^{153,154}. However, these methodologies still had several limitations that needed to be addressed, between them those related to hybridization noise, genome coverage and the inability to detect mutations¹⁵⁵.

The arrival of NGS-based methodologies opened a window of opportunities in the study of SCNAs, radically changing the way to proceed in their detection. These new technologies presented several advantages when compare with array-based approaches. The possibility of simultaneously interrogate point mutation and SCNAs is one of the biggest advances. There are also other not less important such as greater coverage, higher resolution, and accuracy of CN estimation in a faster and cheaper way. Additionally, whole-genome approaches are not limited to genomic regions because of the use of randomly sampled short read in their experimental protocol¹⁵⁶⁻¹⁵⁸. Taking advantage of the better performance of these techniques, several numbers of tools, each one interrogating different features extracted from NGS data, have been developed at both methodological and analytical level¹⁵⁹.

Nevertheless, whole-genome approaches also present some caveats, mainly related to the incorporation to the clinics due to unaffordable costs, generation of huge amount of data and the complexity of the bioinformatics analysis. Attempting to overcome them, clinical

laboratories have opted for the use of targeted panels, designed to interrogate smaller proportion of the genome with greater coverage¹⁶⁰. Point mutations and SCNAs are still able to be detected through these types of techniques, while dealing with more cost-effective and less complex approaches¹⁶¹.

On the other hand, transcriptomic profiling also appears as a suitable technique capturing changes in genomic content through gene expression levels, using it to identify HRD tumors. From expression arrays to RNAseq methodologies, the aim was to establish an expression pattern able to predict deficiency in the HR pathway being possible to use it as a predictive biomarker. Different profiles based on distinct gene selection were tested and some of them showed a favourable result¹⁶²⁻¹⁶⁴. The most representative one regarding transcriptomic studies was the *BRCAness* profile described by Konstantinopoulos et al, distinguishing patients according to platinum sensitivity¹⁶⁵. However, currently none of the tests based on expression profiles have been accepted in clinical diagnostic routine.

The last type of approach, based on the evaluation of functional status of the pathway, will be possible to measure single downstream event that would reflect the proficiency of multiple components of HR. The most studied one has been the expression of RAD51 through quantification of protein loci by immunofluorescence¹⁶⁶. The lack of formation of distinct sub nuclear RAD51 foci after the administration of DNA-damaging agents responds to the inability of tumor cells to

repair DSBs, due to the existence of HRD, making it a feasible biomarker. In the case of functional assays, the limitations respond to other parameters. The quantification of loci in FFPE samples, the existence of a baseline level of expression and the difficulties of post-treatment biopsies are the most crucial ones. For that reason, even the promising applicability, there are several challenges that need to be addressed before incorporating it into clinical routine.

The development of a gold standard biomarker to stratify patient harbouring HRD and, hence, presenting higher sensitivity to DNA-damaging agents will be a key step in the field of HRD-targeted therapies. However, it is unlikely that a single biomarker is sufficient for the establishment of a credible HRD phenotype. Due to the regulation of the pathway at different levels, an HRD scoring that integrates two or more strategies to evaluate the status of the pathway will be highly recommendable^{167,168}.

1.3. In vitro cellular models

Cell line models are essential tool regarding basic science but even more when developing and testing new drugs. Between the multiple applications that a cell line model has, the identification of new targets, testing new therapies, performing drug screening, wide the knowledge about the molecular characteristics of the disease as well as events such as tumor growth, migration, invasion, angiogenesis,

drug delivery and metastasis are some of them. For that reason, the existence of a well-characterized and suitable *in vitro* platform that accurately represents the different histological and molecular subtypes is mandatory in order to correctly establish the basis of every cancer research^{169,170}.

Even though the great efforts done to generate a huge number of cell lines that cover the landscape of existent tumor types and the improvements concerning methodologies, there are still limitations to establish high-quality, permanent cell lines from human primary tumors with high efficiency. Particularly, in the case of OC cell line, the main problems are related to poorly defined origin and hence, lack of information regarding histological and molecular features¹²¹. Additionally, better performances when culturing of high-grade subtypes resulted in a collection of cell lines that do not reproduce the real spectrum of tumors, being some of them underrepresented¹²².

Hence, comprehensive molecular and histological characterization of available cell lines as well as drug sensitivity profiling is highly recommendable to clearly know the type of model you are working with to make the best choice for each specific study and disease¹⁷¹⁻

¹⁷³.

HYPOTHESIS AND MAIN OBJECTIVES

2. Hypothesis and main objectives

2.1. Hypothesis

Gynecological cancers represent a worldwide health problem nowadays. The availability of therapies that directly target molecular features of tumors means a great advantage and advance in the guidance of treatment/clinical routine. Comprehensive molecular characterization of tumors, describing and analyzing known and new molecular hits is imperative to correctly understand and treat tumors.

Particularly, the molecular characterization of the main gynecological tumors at GI level is especially relevant due to the implication of this phenomenon as a prognostic and predictive biomarker. The development of models capable of providing this valuable information based on molecular input could facilitate the stratification and clinical guidance of gynecological patients, candidates to DNA damaging agents.

2.2. Main objectives

Thus, the aim of this PhD project is to develop a comprehensive methodological and analytical approach to determine GI status, overcoming the limitations of current techniques and studying the underlying mechanisms of this phenomenon. With this information, a reliable prognostic and response-predictive classification of

patients relying on generated genomic and transcriptomic data will be determined. Due to the nature of the studied diseases, which differ in many ways, the applied strategies will be different as well.

Specific objectives

Cohort 1

- I. Optimization of technical and analytical workflows to obtain whole-genome CNV profiles from a CNV and SNPs using custom strategy. Optimization and personalization of bioinformatics pipelines to obtain GI (based on CNV data) profiles.

Cohort 2:

- II. Fitting and clinical validation of single-source and integrative platinum-response predictive models based on multi-omic data (SNPs, CNV and transcriptomic data), establishing a predictive biomarker of response. Optimization and validation of *in silico* detection of CN at gene level of *CCNE1*, *EMSY* and *PTEN*.

Cohort 3:

- III. Independent clinical validation of a previously defined in house approach in a clinical trial cohort and different clinical scenario. Correlating the findings in real OC population in prognostic and

response rate levels to PARPi and evaluation of the suitability of the technical approach.

Cohort 4:

- IV. Fitting a prognostic classifier by Random Forest (RF) algorithm based on TCGA-EC genomic data in a real EC cohort.
- V. Validation of mutational-based RF classification in a subpopulation of EC patients obtaining GI profiles.

Pre-clinical studies

- I. Drug-response characterization of OC cell lines to different DNA damaging agents and PARPi (Olaparib, Niraparib y Talazoparib) in monotherapy and in combination.
- II. Generation and molecular characterization of Lurbinectedin-resistant (PM01183) OC cell line.
- III. Evaluation of the inhibitory role of ascites in OC cell lines treated with Lurbinectedin.
- IV. Establishment of GI profiles and genotyping of HRR-genes and correlation with response rate to different therapies.

MATERIALS AND METHODS

3. Materials and methods

3.1. Patients cohorts

3.1.1. COHORT 1

Cohort 1 (Table 4) is composed by 45 FFPE-samples from patients diagnosed with HGSOE and collected from 1999 to 2016. All patients signed an informed consent form approved by our Institutional Review Board in 2014 (ACO-COE-3012-02). Informed consent of patients was obtained in accordance with ethical and legal regulations of the institution. Criteria for inclusion in the study were: age over 18 years; tumours with serous histology; high grade and all stages. Median age at diagnosis being 56.5 years (27.8-81.6 years) and median follow-up of 80.3 months (8.8-249.7 months). The median progression-free survival (PFS) was 18.0 months (2.0-189.0 months) and the median overall survival (OS) 62.6 months (8.5-196.5 months). This cohort will be used to accomplish objective 1 following the approach established in Figure 14.

Table 4: Clinicopathological characteristics of cohort 1

Characteristics	N	%
Stage		
<i>I</i>	1	2.2
<i>II</i>	7	15.55
<i>III</i>	30	66.7
<i>IV</i>	7	15.55
Exitus		
<i>No</i>	23	51.1
<i>Yes</i>	22	48.9
Relapse		
<i>No</i>	7	15.6
<i>Yes</i>	38	84.4
Clinical BRCAness		
<i>No</i>	13	28.9
<i>Yes</i>	20	44.4
<i>NA</i>	12	26.7

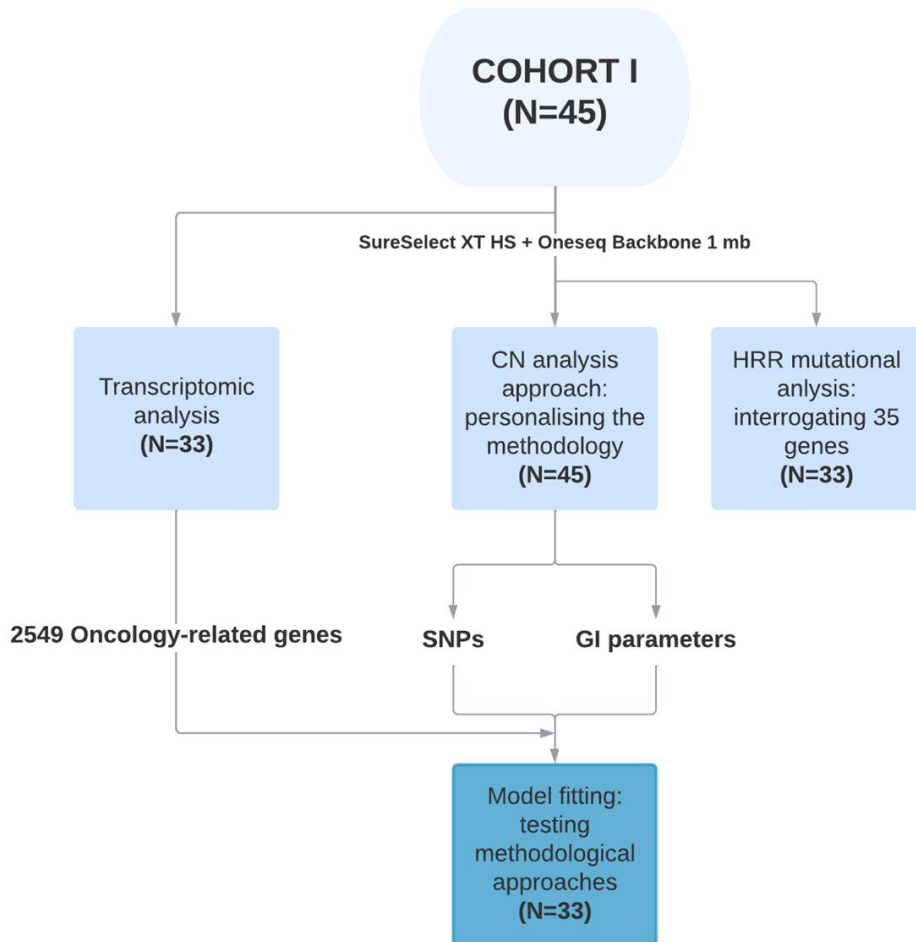


Figure 14: Methodological strategy in cohort 1. Cohort 1 has been used to achieve objective 1, establishing the basis of CNV evaluation and model fitting, setting up methodological and analytical approaches.

3.1.2. COHORT 2

The cohort included 190 HGSOc samples ambispectively collected from 2007 to 2020. All patients signed an informed consent form approved by the required Ethics Committees and the study was approved by the EC of Instituto Valenciano de Oncología (IVO) in 2021 (LBM-02-20, SCARFACE). The informed consent of patients was obtained in accordance with institutional, ethical, and legal regulations. The inclusion criteria in this study were: age ≥ 18 years at inclusion, HGSOc tumors and patients that have received treatment with platinum-based chemotherapy at first-line (Table 5). FFPE tumor blocks of 190 HGSOc patients were analyzed. The median age of the cohort was 59.21 years (27.8-81.6 years) with median follow-up of 31.03 months (5.87-159.27 months). The median platinum-free interval (PFI) after first-line was 16.28 months (0-83.33 months), presenting a recurrence rate of 52.11% (99/190) and a median PFS to PARPi of 11.03 months (1.03-64.63 months). Deaths occurred in 19.80% of patients at the time of data analysis, with a median OS of 31.03 months (5.87-159.29 months). This cohort will be used to accomplish objective 2 and 3 following the approach established in Figure 15.

Table 5: Clinicopathological features of Cohort 2. A) Pathological parameters and
B) Treatment and outcome-related parameters.

A)

Clinical parameter		N	%
Histology	HGSOC	190	100
Stage	IA	7	3.68
	IC1	6	3.16
	IC2	9	4.74
	IIA	4	2.11
	IIB	6	3.16
	IIIA1	8	4.21
	IIIA2	5	2.63
	IIIB	10	5.26
	IIIC	77	40.53
	IVA	12	6.3
	IVB	27	14.21
	NA	19	10
Stage (aggregated)	Localized (I-IIIB)	34	17.89
	Locally Advanced (III-IVA)	120	63.16
	Metastatic (IVB)	36	18.95
Type of biopsy	Excisional	132	69.47
	Incisional	35	18.42
	Tru-Cut	23	12.11
BRCAg	Wildtype	141	71.21
	VUS	13	6.84
	Pathogenic	36	18.95

B)

Clinical parameter		N	%
Surgery	Yes	167	87.89
	No	23	12.11
Primary debulking surgery	Yes	114	68.26
	No	53	31.74
Residual disease after primary debulking surgery	Yes	18	15.79
	No	96	84.21
First-line platinum therapy	All	190	100
Relapse after first-line therapy	Yes	99	52.11
	No	91	47.89
Recived PARP	Yes	59	31.05
	No	131	68.95
Progression with PARPi	Yes	29	49.15
	No	30	50.85
Death	Yes	39	20.53
	No	151	79.47

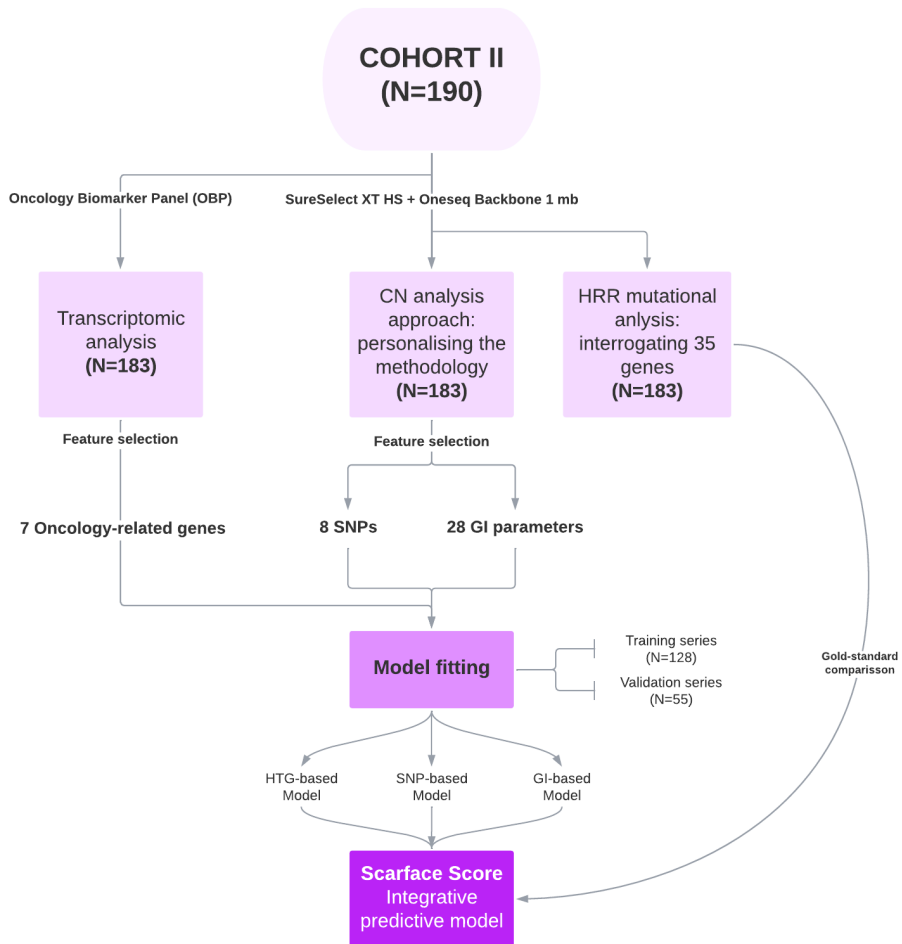


Figure 15: Methodological strategy in cohort 2. Cohort 2 will be used to achieve objective 2 and 3, analysing multi-omic data from 190-FFPE samples to test and validate the response-predictive model to DNA-damaging agents (Scarface Score).

3.1.3. COHORT 3

Cohort 3 (Table 6) belong to a phase II open-label and non-randomized clinical trial (NCT02684318, EudraCT 2015-001141-08, 03.10.2015)^{174,175}, which recruited patients from 6 centres in Spain. The POLA trial is the first phase 2 to assess the efficacy and toxicity of lurbinectedin and olaparib in previously treated gynecological tumors and their correlation with molecular characteristics, specifically GI (Figure 16). The study was approved by the ethics committee of the Hospital Vall d'Hebron and was conducted in accordance with the Declaration of Helsinki, ICH Good Clinical Practice guidelines and the current legislation. Written informed consent was obtained from all patients before study-specific procedures. Detailed inclusion and exclusion criteria are explained in Annex 1. Regarding outcome and response parameters, treatment-free interval (TFI) in first line was 15.86 months (0.23-50.37). During the clinical trial, 42.11% died as consequence of the disease; the median PFS to administered treatment was 3.68 months (0.03-17.69 months) and the median OS 58.67 years (15.4-192.93). This cohort will be used to accomplish objective 4 following the approach established in Figure 16.

Table 6: Clinicopathological characteristics of cohort 3. A) EC series and B) OC series.

A)

Characteristics	N	%
Histology		
Serous	3	15.79
Endometroid	9	47.37
Carcinosarcoma	2	10.53
Others (mixed or non-specified)	5	26.32
Stage		
I	8	42.11
II	0	0
III	2	10.53
IV	7	36.84
NA	2	10.53
Grade		
I	3	15.79
II	4	21.05
III	12	63.16
Exitus		
No	8	42.11
Yes	11	57.89
Relapse		
No	0	0
Yes	19	100

B)

Characteristics	N	%
Origin		
Ovary	33	89.19
Fallopian tube	2	5.41
Peritoneum	2	5.41
Histology		
Serous	34	91.89
Endometroid	3	8.11
Stage		
I	3	8.11
II	2	5.41
III	18	48.65
IV	9	24.32
NA	5	13.51
Grade		
High	31	83.78
NA	6	16.22
Exitus		
No	16	43.24
Yes	21	56.76
Relapse		
No	0	0
Yes	37	100

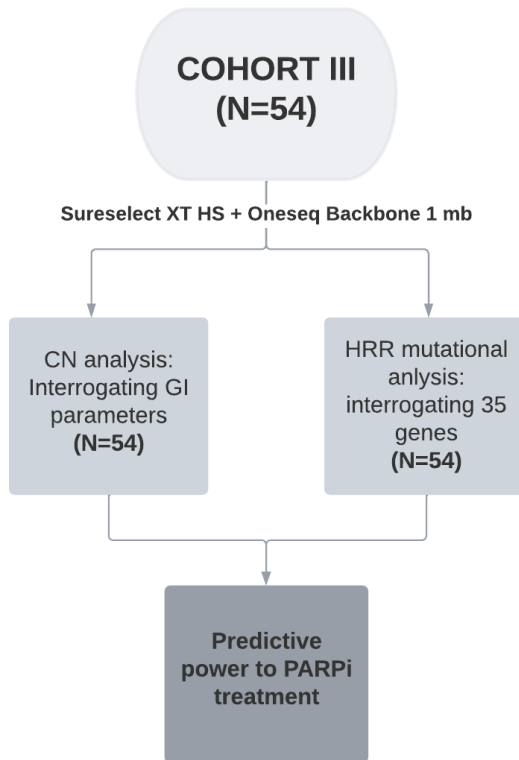


Figure 16: Methodological strategy in cohort 3. Cohort 3 will be used to achieve objective 4, by validating the methodological and analytical approach to interrogate CNV patterns in a clinical-trial scenario with PARPi response as an endpoint.

3.1.4. COHORT 4

This study includes 96 EC patients prospectively collected from 2010 to 2019 within the context of the institutional projects ACOG0901 and ACOG1602 (Table 7). Experimental protocols were approved by Instituto Valenciano de Oncología Institutional Review Board in 2009 and 2016 respectively. At the time of the study, our prospective institutional EC database contained a total of 187 patients. The criteria for inclusion in this study were: age over 18 years; tumors with serous or endometrioid histology; grade I to III and stage I, II and III. A total of 149 fulfilled these criteria, from which 96 were selected according to the best ranked DNA quality and concentration. This cohort will be used to accomplish objective 5 and 6 following the approach established in Figure 17.

Median age at diagnosis being 62 years (range: 36.4-87) and median follow-up of 35.02 months (range: 2.1–91.2 months). During follow-up, 15.2% of the patients recurred, and 10.7% died as consequence of the disease; the median PFS being 33.65 months (range: 2.1-91.2) and the median OS, 35 months (range: 2.1-91.2 months).

Table 7: Clinicopathological features of Cohort 4. A) TCGA and B) Our series.

A)

Stage	Endometrioid			Serous	Total
	Grade 1	Grade 2	Grade 3	All	
I	78 (23)	83 (24)	70 (21)	17 (5)	248 (73)
II	3 (1)	9 (3)	6 (2)	5 (1)	23 (7)
III	7 (2)	12 (4)	26 (8)	25 (7)	70 (21)
Adjuvant therapy					
RT	12 (3)	28 (8)	22 (6)	7 (2)	69 (19)
Chemo	2 (1)	6 (2)	14 (4)	13 (4)	35 (10)
ChemoRT	2 (1)	9 (3)	18 (5)	17 (5)	46 (13)
Unknown	70 (20)	61 (17)	57 (16)	16 (5)	204 (58)
Total	86 (24)	104 (29)	111 (31)	53 (15)	354 (100)

B)

Stage	Endometrioid			Serous	Total
	Grade 1	Grade 2	Grade 3	All	
I	40 (42)	24 (25)	6 (6)	7 (7)	77 (80)
II	0 (0)	1 (1)	0 (0)	1 (1)	2 (2)
III	5 (5)	3 (3)	4 (4)	5 (5)	17 (18)
Adjuvant therapy					
RT	21 (22)	8 (8)	2 (2)	1 (1)	32 (33)
Chemo	2 (2)	1 (1)	2 (2)	7 (7)	12 (13)
ChemoRT	3 (3)	3 (3)	6 (6)	3 (3)	15 (16)
Unknown	2 (2)	1 (1)	0 (0)	2 (2)	5 (5)
Total	28 (29)	13 (13)	10 (10)	13 (13)	64 (67)

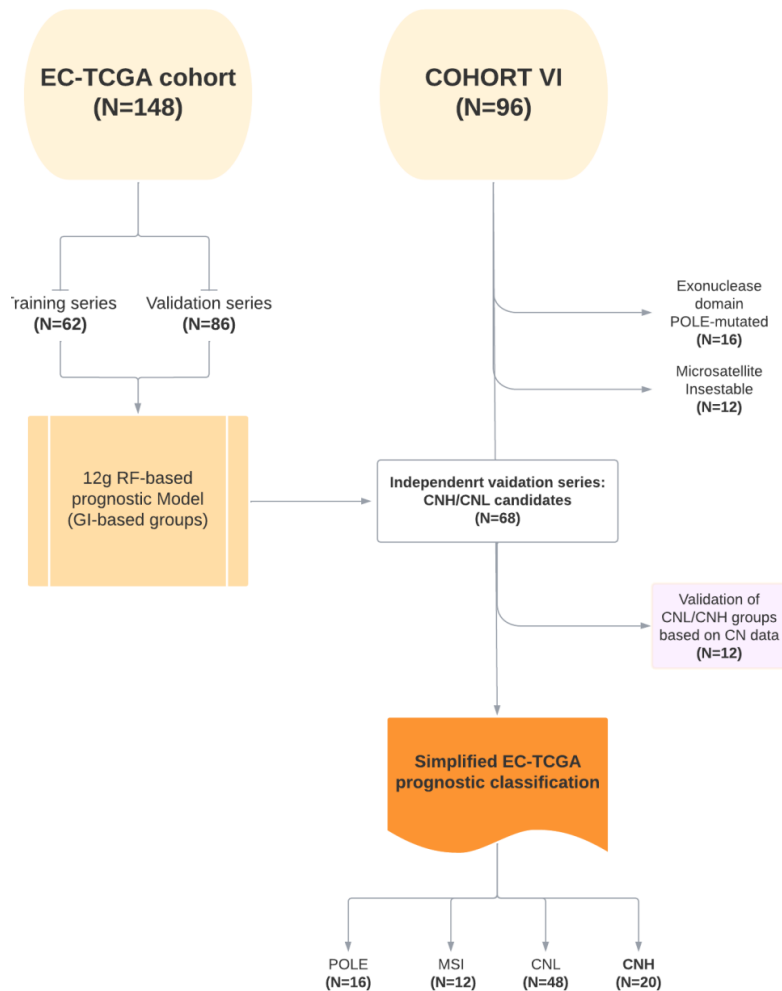


Figure 17: Methodological strategy in cohort 4. Cohort 4 will be used to achieve objective 5 and 6, building an independent model to establish GI profiles in a different clinical scenario of EC patient.

3.2. Immunohistochemistry analysis

IHC has become a clinical routine, being mostly used to guide diagnosis in conjunction with tumor morphology but also helping to answer other questions including characterization of the neoplastic diseases or determination of some pathway alteration, as in the case of MMR proteins. The basis of this methodology is the use of specific antibodies that, conjugated to enzymes, catalyse a reaction when the recognition antibody-antigen is achieved that derivates in the formation of detectable compounds to visualize and localize specific antigens in a tissue sample.

The IHC study was performed in whole 5 μ M-sections from FFPE tissue samples. Microscopic slides (hematoxylin-eosin) from the selected cases were reviewed and confirmed by a gynaecology pathologist.

For IHC analysis, standard protocol was applied. Antigen retrieval was performed by pressure cooker boiling at 1.2 atmospheres for 3 min in 10 μ mol/L citrate buffer (pH 6.0). Labelled Streptavidin–Biotin (LSAB) staining method was done, followed by revelation with 3,30-diaminobenzidine. The panel of markers and specific conditions used in the analysis are presented in Table 8.

In regards to interpretation of IHC results, immunoreactivity was defined as negative in absence of staining while presence was interpreted as positive in all proteins except from p53. For p53,

according to the protocols for Gynecological Cancer Biomarkers from the College of Pathologists (<https://www.cap.org/protocols-and-guidelines/cancer-reporting-tools/cancer-protocol-templates>) the immunostaining was interpreted as normal (between 1-90% of cells express p53) or aberrant (lack of nuclear or cytoplasmic expression, strong, diffuse nuclear expression in greater than 90% of cells or diffuse cytoplasmic staining) .

Table 8: IHC antibodies applied to interrogate MMR and p53 status in EC.

Protein	Dilution	Clone	Manufacturer
MLH1	Prepared to use	IRO79	DAKO
PMS2	Prepared to use	EP51	DAKO
MSH2	Prepared to use	FE11	DAKO
MSH6	Prepared to use	EP49	DAKO
TP53	Prepared to use	DO-7	DAKO

3.3. Genomic approaches

The ability of sequencing the genome of oncologic patients has revolutionized the way of understanding clinical routine and treatment guidance. From simpler and low-throughput technologies based on capillary electrophoresis, such as Sanger sequencing and fragment analysis to high-throughput NGS-based technologies, which allow parallelization of wide number of samples and genes. While

NGS technologies play a main role in oncology, capillary electrophoresis continue being of importance, principally in sequencing validation and targeted sequencing, simplifying and cheapening concrete genomic studies.

3.3.1. DNA extraction, quantification and quality control

NGS-based technologies are high DNA-quality demanding, sometimes limiting the application of these methodologies in routine samples as FFPE. To ensure that samples fulfil needed requirements, in addition to classical fluorimetric and spectrophotometrically quantification methods, more precise approaches, such as qPCR and genomic profiling with a bioanalyzer will be applied.

DNA extraction was performed using the QIAmp DNA FFPE Tissue kit (Qiagen, Valencia, CA) following the manufacturer's instructions. Three FFPE blocks sections of 20 μm -thin were used when tumor content was higher than 30%. In case the sample did not achieve the specified percentage, punch samples were obtain from the block, assuming a 100% of tumoral content in further analysis. The final DNA concentration was measured by two different methods: Spectrophotometrically, by NanoDrop ND1000 (Thermofisher; Waltham, MA, USA) and fluorometrically using PicoGreen™ reagent with a QuantiFluor® dsDNA System (Promega). Additionally, quality and related size of genomic DNA were assessed by the microfluidics-

based platform Agilent 4200 TapeStation with Genomic D1000 Kit (Agilent, Santa Clara, CA) or by qPCR-based approach (QC Illumina kit) (Illumina, San Diego, CA) depending on chosen NGS-protocol. Electropherograms were visualized with the TapeStation Software Analysis A.02.01 SR1 including data collection, peak detection, and interpretation of the different profiles.

3.3.2. Truseq Low Input

The first applied NGS approach was the TruSeq Custom Amplicon Low Input Kit (Illumina, San Diego, CA), an amplicon-based assay for targeted resequencing in FFPE samples. The design of the custom panel, consisting on 13 genes; *POLE*, *PTEN*, *TP53*, *ARID1A*, *ARID5B*, *FBXW7*, *PPP2R1A*, *CTCF*, *CTNNB1*, *RPL22*, *KRAS*, *PIK3CA*, *PIK3R1*, was performed by DesignStudio tool, (Illumina, San Diego, California). Genes were selected based on the results from the EC-TCGA⁶⁶, choosing those that best discriminate between the 4 defined groups, based on relative and absolute frequency of each gene among them. This selection aims to improve the feasibility of the future model.

The recommended initial amount of DNA to construct the library ranged from 10 to 100 ng. In some cases, available DNA did not achieved the recommended quantity, hence, maximum available quantity was assigned to these samples. Finally, used samples presented a median DNA concentration of 49.91 ng/μl (8.77-189.538

ng/μl). Samples were subjected to dual-pool (in order to improve sensitivity) amplicon-based PCR library preparation according to manufacturer's protocol, following the standard steps of: hybridization, amplification and clean up. Subsequent sequencing of pooled libraries was performed in a NextSeq 550 sequencing platform as specified by Illumina, using a 300 cycles High Output FlowCell (Illumina, San Diego, California).

NGS data processing and variant filtering

Data analysis, including alignment to the hg19 human reference genome and variant calling, was done using CASAVA pipeline (Illumina, San Diego, CA). Variants were then annotated by the Illumina VariantStudio v3.0 data analysis software (Illumina, San Diego, CA). Integrative Genomic Viewer (IGV, Broad Institute) was used to visualize the sequence and check the presence of mutations^{176,177}.

3.3.3. SureSelect XT HS +Oneseq Backbone 1 mb

The second NGS approach was a hybrid capture-based target enrichment protocol from Agilent (Santa Clara, CA, USA), which combines SureSelect-XT HS and Oneseq Backbone Kit. This approach allows us to simultaneously interrogate Single-Nucleotide Variant

(SNV) and indels in a custom panel constituted by 35 DNA damage response genes and CN changes, based on the analysis of 147.000 SNPs homogenously distributed along the whole genome. DNA damage response genes included in the custom panel were: *BRCA1*, *BRCA2*, *BARD1*, *BRIP1*, *CHEK1*, *CHEK2*, *FAM175A*, *NBN*, *PALB2*, *ATM*, *MRE11A*, *RAD51B*, *RAD51C*, *RAD51D*, *RAD54L*, *FANCI*, *FANCM*, *FANCA*, *ERCC1*, *ERCC2*, *ERCC6*, *REQL*, *XRCC4*, *HELQ*, *SLX4*, *WRN*, *ATR*, *PTEN*, *CCNE1*, *EMSY*, *TP53*, *MLH1*, *MSH2*, *MSH6* and *PMS2*. Given that the tumor types of the series have been described to present PTEN loss and EMSY and CCNE1 amplification, the panel was reinforced in these regions with more probes.

Methodologically, a starting concentration of extracted DNA ranged between 10-200 ng was enzymatically fragmented to a size between 150 to 200 bp. DNA integrity and fragment size were determined by Genomic DNA assay in the TapeStation 4200 (Agilent, Santa Clara, CA). Each library was then hybridized according to protocol. Pooled library were sequenced in a High-output 300 cycles flow cell (100 bp paired-end) on a NextSeq550 (Illumina, San Diego, CA, USA).

NGS data processing and variant filtering

Data resultant from the sequencing process needs to be processed and analysed. The first step was to convert BCL files into fastq by the software Bcl2fastq. Once fastq files have been obtained, Alignment

and Variant calling steps need to be addressed. BWA-MEM and Haplotypecaller algorithms, from the Broad Institute were used for this aim. Finally, Variant studio 4.0 (Illumina) and SureCall 4.2 (Agilent) were applied for variant annotation (SNVs and Indels). Variants were selected after a filtering process based on the following analytical parameters: coverage > 100x (covered in forward and reverse sense); allelic frequency (AF) above 5 % and classification of variants, including pathogenic (P), likely pathogenic (LP) or Variant of unknown significance (VUS) with the prediction of pathogenicity in ClinVar, frameshift variants as well as variants with both Polyphen and SIFT predictors annotated as damaging or pathogenic.

3.3.4. Microsatellite instability

MSI testing is based on a PCR amplification of DNA regions containing microsatellite repeats and a subsequent capillary gel electrophoresis, detecting abnormal lengths for these sequences.

MSI was performed on 2-3 ng of DNA from paired FFPE and blood samples using the Type-it Mutation Detect PCR Kit (Qiagen) in a Veriti thermocycler (Applied Biosystem, Foster City, CA). Specific primers for the following STR regions: NR27, NR21, NR24, BAT26, BAT25, D5S346, D2S123 and D17S250^{178,179} were used. Primers were labelled with the fluorophores FAM, HEX or NED. PCR conditions were: 5' initial denaturing at 95°C followed by 35 cycles at 95°C of 30'', 1'30''

at 60°C and 30'' at 72°C with a final 10' extension at 68°C. PCR products were denatured with formamide for 5' at 95°C and analysed by capillary electrophoresis in the ABI3130xl Genetic Analyzer (Applied Biosystem, Foster City, CA). In addition, the GeneScan™ 500 LIZ™ Dye Size Standard is used as an internal size standard. For visualization, GeneMapper v4.0 software (Applied Biosystem, Foster City, CA) was used. MSI High was considered when at least 30% of STR regions presented an MSI pattern, while Microsatellite stable (MSS) presented a total lack of instable patterns.

3.3.5. Sanger Sequencing

Aiming to confirm POLE mutations described by NGS, Sanger sequencing was performed. To sequence exons 9–14 of POLE, specific primers were designed (Table 9) and PCR products were amplified with the input of 200 ng FFPE derived DNA. Considered reference sequence was NM_006231.4. Sequencing was performed using BigDye v3.1 terminator cycle sequencing chemistry on the ABI3130XL Genetic Analyzer (Applied Biosystem, Foster City, CA). Variant analysis was performed using Sequencing Analysis Software v5.2 (Applied Biosystems).

Table 9: Primer sequences for Sanger sequencing of POLE EDM in EC series.

Exon	Primer sequence	Product length (bp)	Tm (°C)
Exon 9	GTGTTTCAGGGAGGCCTAATG	248	57.5
	GGGCAGATGCTGCTGTAGTA		
Exon 11	ACTTTGGGAGAGGAATTTGG	250	59.5
	CCCTAAGTCGACATGGGAAGC		
Exon 13	CATCCTGGCTTCTGTTCTCA	298	59.5
	GAGCGGGCTGGCATAACAT		
Exon 14	ACCCTGGGCTCTTGATTTTT	247	57.5
	CACCTCCATTACAGCTCCCAAGT		

3.3.6. Multiplex Ligation-dependent Probe Amplification

To validate the *in silico* assessment of CN amplification and losses at gene level, a Multiplex Ligation-dependent Probe Amplification (MLPA) analysis was performed. MLPA, based on a multiplex PCR, is the gold standard methodology when evaluating CNV at gene-level. For that reason, that was the chosen technique for validation. MLPA reaction can be divided into five main steps: (1) DNA denaturation and probes hybridization; (2) ligation reaction; (3) PCR amplification; (4) separation of amplification products by electrophoresis; (5) data analysis¹⁸⁰. Particularly, CN in *CCNE1*, *PTEN* and *EMSY* genes (previously described in OC population) were evaluated using SALSA® MLPA Probemix p225-E1 and P078-D2 Breast tumor assays. The protocol was performed following the manufacturer’s instructions (MRC Holland). Amplified products were separated using an ABI3130XL Genetic Analyzer (Applied Biosystem, Foster City, CA) and

interpreted with GeneMapper Software v4.0 (Applied Biosystems, Foster City, CA). Quantification of the results of fragment analysis was performed using the Coffalyser software as described by the manufacturer (MRC Holland). Different normal control samples from healthy FFPE tissue were used to normalize the allele dosage.

3.4. Transcriptomic approaches

3.4.1. HTG Oncology Biomarker Panel

Transcriptome analysis was performed with HTG Edgeseq technology (HTG Molecular Diagnostics, Tucson, USA), a very novel technique based on RNA-seq with high sensitivity, requiring low sample input (5µm FFPE section and an area of 15 mm²) and proved robustness. Opposite to other approaches, HTG Panels are specifically designed for each application not analysing the whole transcriptome but focusing on target genes. In this case, the Oncology biomarker Panel (OBP) was selected. This panel is constituted by 2549 mRNAs from 24 tumor-related pathways.

Methodologically, FFPE slides were scraped based on hematoxylin-eosin preparations and lysated. Once samples were primarily processed, pre-hybridization step was performed. This step was done by using specific probes and a quantitative nuclease protection assay (qNPA). Libraries were then amplified by PCR with illumina adapters and quantified using KAPA Library Quant Kit Universal qPCR Mix

(KAPA Biosystems, Wilmington, MA, USA) by ABI 7500 Fast Real-Time PCR System (qPCR). Following final clean up, libraries were ready to be sequenced in a NextSeq 550 System, using a High Output Flowcell (75 cycle) from Illumina. Fastq files obtained after primary analysis were parsed with HTG Parser Software running Bowtie2 in the backend (Figure 18).

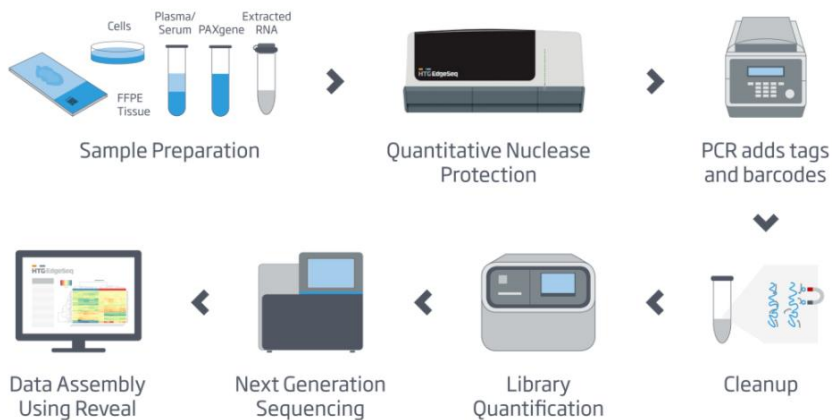


Figure 18: HTG protocol. Step by step methodological protocol to perform targeted transcriptomic analysis, consisting on; sample preparation, Quantitative Nuclease protection, addition of tags and barcodes, clean-up, library quantification, NGS and parsing.

Data processing and DE Analysis

Once sequencing data has been generated from RNAseq protocol, total read counts need to be processed and analysed in order to obtain biologic understandable conclusions. The starting point to do

that will be the Principal Components Analysis (PCA), a useful technique for exploratory data analysis, available as an R function. PCA allows the visualization of the overall variation of the data, identifying common profiles between them.

The Heatmap, similar to PCA, helps to initially visualize the results, grouping samples that present similar gene-expression patterns. To do this, Heatmap2 package from R will be applied over normalized counts. Additionally, this function will also be useful after differential expression analysis, to illustrate resultant gene-selection.

Following to basic/starting visualization tools, DE analysis will be performed. DESeq2¹⁸¹, an R package, will model the raw counts, using normalization factors (size factors) to account for differences in library depth. Then, it will estimate the gene-wise dispersion and shrink these estimates to generate more accurate estimates of dispersion to model the counts. DESeq2 will fit the negative binomial model and perform hypothesis testing using the Wald test or Likelihood Ratio Test. This tool will return differentially expressed genes between studied conditions.

Finally, aiming to extract all the information available in the transcriptomic analysis, Gene Set Enrichment Analysis (GSEA)¹⁸² will be performed. GSEA is a powerful analytical computational method for interpreting gene-expression data. This method drives its power by focusing in gene sets that have common biological function,

determining its statistically significant differences between two biological states. The used GSEA method was Generally Applicable Gene-set Enrichment (GAGE)¹⁸³.

Pathway analysis and functional annotation with Gene Ontology (GO) categories were performed with DESeq2 package, coupled with GAGE and pathview packages from R/Bioconductor. All these analyses were performed with R v4.0.3.

3.5. Bioinformatics Analysis

3.5.1. Copy-Number Variation pipelines

Bioinformatics analysis to obtain CNV, for its part, was independently carried out by the implementation of two in-house pipelines and the available commercial solution from Agilent (Figure 19).

The first approach, based on CNVKit algorithm¹⁸⁴, Different from other pipelines, uses both on-target reads and the non-specifically captured off-target. While the off-target reads alone, due to its low coverage, do not provide enough coverage to call SNVs and other small variants, they can provide useful information on copy number at a larger scale. On- and off-target read depths are combined and normalized to a reference derived from a panel of 10 peritumoral control samples to obtain the values in terms of log₂ copy ratios. For the segmentation step, Circular Binary segmentation (CBS) was used.

In this case, the variant calling step was performed by Mutect2¹⁸⁵. Results can be used for visualization and further analysis.

The second approach based on saasCNV R package¹⁸⁶. This algorithm, based on a joint segmentation algorithm, uses two data dimensions simultaneously, total read depth and Bi-allelic frequency (BAF), incorporating both quantities to improve power and accuracy in identifying and characterizing CNV. SaasCNV follows the steps of SNPs recalibration; alignment, and variant calling, performed according to good practices from Broad Institute (GATK)¹⁸⁷. For the normalization step, a single control sample was used.

Finally, SureCall software was applied with default parameters (Version 4.1.1.5).

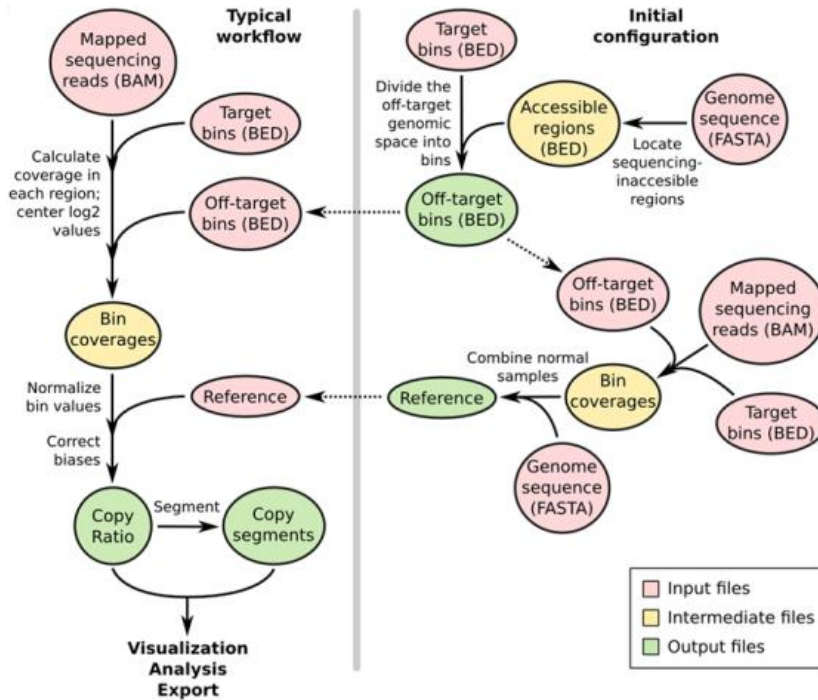
To calibrate the thresholds levels among the numerous CNV derived parameters comparisons at distinct levels were performed. While saasCNV and SureCall are limited to slightly personalization, CNVkit has a greater number of analytical parameters.

Complementary to CN pipelines analysis, scarHRD¹⁸⁸, an R package which determines pre-established genomic scars (TAI, LOH and LST) from NGS data was applied.

In addition to whole-genome CNVs and mutations, we are able to obtain *in silico* CN at gene-level. Thus, due to the importance of *PTEN*, *CCNE1* and *EMSY* in this regard, the custom panel has been

reinforced in probes in these regions. To analyze it, Panel.mops¹⁸⁹ package from R has been used. This pipeline is designed to detect CNVs from targeting sequencing data and it is built upon cn.MOPs. The approach consists of 6 basic steps: Read counting, quality control (QC), control sample selection, normalization, CN detection and segmentation.

A



B

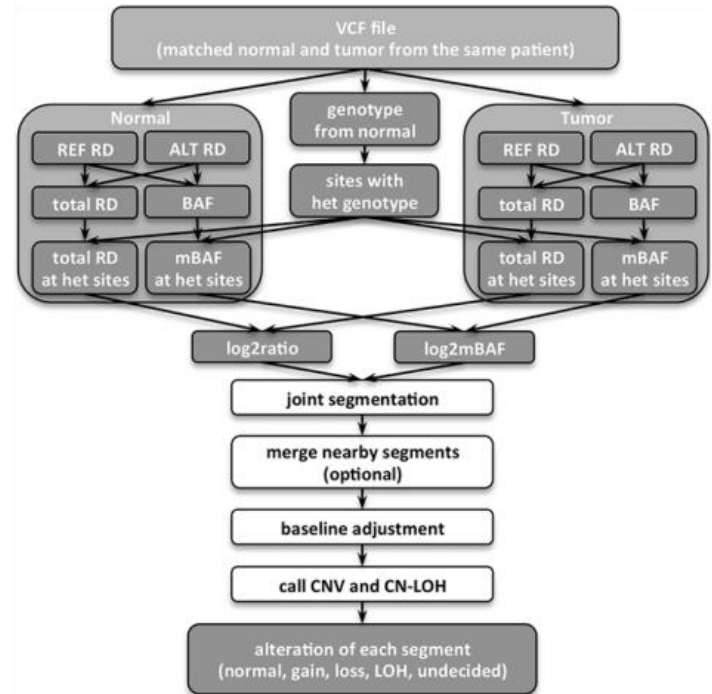


Figure 19: Tested CNV analysis pipelines step by step. A) CNVkit¹⁸⁴ and B) saasCNV¹⁸⁶.

3.5.2. Data mining in OC cohort

Three sources of raw data have been used to fit the model: raw read counts, GI parameters derived from CNVkit algorithm and HTG gene expression panel data (Figure 20).

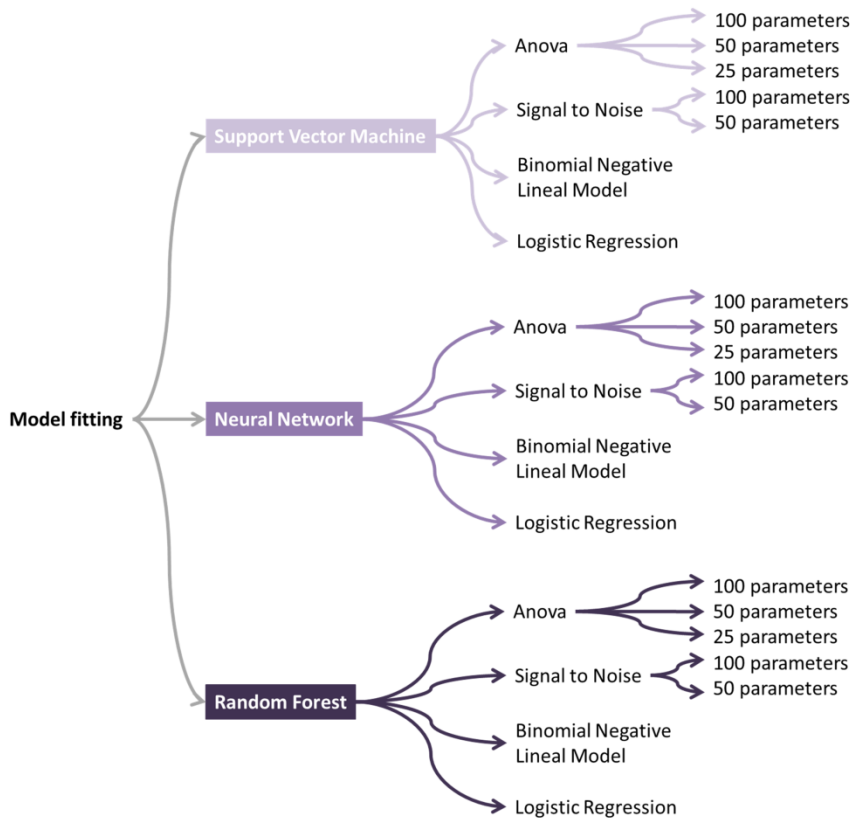


Figure 20: Different tested algorithm to build the model. Each feature selection algorithm and model fitting were subsequently perform to achieve the best performing model.

Due to the complexity and diversity of the data that constituted the levels of the predictive model, different approaches were tested for cleaning of the data to avoid the over-fitting of models, feature selection and model fitting. The first step was cleaning the data and denoising, through which non-informative or redundant predictors are removed. The next step was feature selection. This step is performed to reduce the number of input variables, aiming to decrease the computational cost, to avoid over fitting and, to improve the performance of the model. Thus, features with the strongest relationship with the target variable will be selected measured with not adjusted univariate inference. There are a wide number of methods to do that, however, the suitability of them depends on the data type of both the input and output variables. In this case, the tested methodologies for feature selection were; Anova, Signal to noise (S2N), Binomial negative linear model and logistic regression. Briefly, ANOVA test compares inter and intragroup variance; S2N ratio compares a level of signal power to a level of noise power. It is the quotient between the subtraction from the mean of the groups and the sum of its variances; Binomial negative linear model is a generalization of Poisson regression which loosens the restrictive assumption that the variance is equal to the mean made by the Poisson model and logistic regression is a machine learning (ML) model studies and adjusts to a line the residuals of the data points. Finally, hyper parameter selection and model training step was performed. In this step, the best performer model with

correspondence selected features and hyper parameters were fit. In this step, the tested methods were: Support Vector Machine (SVM), RF and Neural Network (NN). Summarily, SVM algorithms are particular linear classifiers which are based on the margin maximization principle. The SVM accomplishes the classification task by constructing, in a higher dimensional space, the hyper plane that optimally separates the data into two categories. It uses a set of mathematical functions that are defined as the kernel. The kernel functions return the inner product between two points in a suitable feature space. NNs, for its part, are a set of algorithms that are designed to recognize patterns. They are inspired by the biological NN in the brain. The basic computational unit of a NN is a neuron or node. It receives values from other neurons and computes the output. Each neuron is associated with weight (w). This value is given as per the relative importance of that particular node. The activation function determines how the weighted sum of the input is transformed into an output from a node or nodes. Its choice has a large impact on the capability and performance of the NN. RF was aforementioned.

In the first approach, raw read counts from 170080 SNPs obtained from NGS-panel were used to feed the model. These parameters are polymorphisms uniformly distributed along the entire genome. In order to reduce the dimensionality of the data frame, different strategies were applied: SNPs counts were selected sorting the p-

values obtained from an ANOVA test, highest s²n ratios; the significant discriminant SNPs in a binomial negative linear model and in a logistic regression regularized with elastic net.

After feature selection, chosen SNPs were adjusted with SVM applying radial kernel, RF or NN algorithms, performing hyper parameters tuning as follows: a) In SVM the assayed values of the penalty factor for margins violation (cost, C) were 1, 10, 50, 100, 500, 700, 1000, 1500, whereas Gaussian Width, represented by Sigma in the kernel of SVM equation, takes values of 0.0001, 0.001, 0.01; b) In the RF algorithm, the considered number of variables randomly selected at each split (mtry) were 2, 5, 10 and the used split rule was Gini index; c) In the NN tuning, sizes of 5, 10, 15, 20, 40 layers and values of 0.01 and 0.1 in decay in the weights of the loss function were considered.

The second approach applies GI parameters derived from CNVKit algorithm. Due to the lower number of parameters, this approach presents a less stringent feature selection, simplifying the filtering strategy. Parameters derived from a decision tree were curated by an expert in the field with the aim of endowing the dataset with clinical significance. The depurated data frame comprised 28 variables: Number of events, Mb of altered genome, Percentage of altered genome, Number of events excluding copy numbers between 0.5 and 3, Mb of altered genome excluding copy numbers between 0.5 and 3, Percentage of genome altered excluding copy numbers between 0.5

and 3, Total number of gain events, Mb of genome altered by gains, Percentage of genome altered by gains, total number of gain events, Mb of genome altered by gains and percentage of genome altered by gains between 0.5 and 3, total number of loss events, Mb of genome altered by losses, percentage of genome altered by losses, total number of LOH events, mb of genome altered by LOHs, percentage of genome altered by LOHs, total number of LOH>15 Mb, Mb of genome altered by LOHs>15 Mb, percentage of genome altered by LOHs>15 Mb, total number of LOH>10 Mb, Mb of genome altered by LOHs>10 Mb, percentage of genome altered by LOHs>10 Mb and LOH, LST, TAI and scoreHRD from scarHRD package.

The third strategy assesses the expression of 2559 genes related with oncology analysed with HTG OBP. Feature and algorithms selection was analogous to the SNPs scheme. Logistic regression was excluded as a feature selection method considering that only two genes presented discriminant power.

Ensemble of the best performer models comprises: NN (100 SNPs) with greater s2n ratio, SVM with radial kernel with 30 human curated decision trees based on instability parameters and SVM with the 25 HTG genes carrying the lowest ANOVA p-values.

3.5.3. Random Forest Algorithm

The ability to precisely classify observations is extremely valuable in the clinical setting. RF is one of the most-used supervised algorithms due to its simplicity and versatility. It is a ML algorithm that can be used for both classification and regression purposes. RF is an ensemble method based on the construction of multiple decision trees (building blocks of the RF), extraction of each prediction and selection of the best solution. In a RF, trees are not only trained on different sets of data (bagging) but also use different feature selection in each one (feature randomness), avoiding over fitting issues. Hence, a large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituents' models.

Given the nature of the data, this type of algorithm was the chosen one aiming to obtain a surrogate of GI to classify CN-based EC-TCGA subgroups. Thus, the EC dataset from TCGA was used in training and validation steps, while our dataset was used to independently validate the pre-established prognostic model. Dichotomous and categorical variables including mutational status of the studied genes and clinical and pathological parameters such as histology, stage and grade were implemented in the model. Furthermore, a standard bagging approach is applied. To correctly adjust the RF model, the TCGA dataset was randomly split in two cohorts, consisting of 62 patients for the training set and 86 for the validation set.

To train the model, genotyping of 12 genes was included as categorical dichotomous variables. Prior to the adjustment of the RFA model, the number of variables per level on each split was optimized to pre-train the model. The model was validated with 5-fold cross-validation and bagging. The number of trees was empirically estimated to 1000. R v3.4.3 patched was used to build and test all the predictive models.

3.6. In vitro studies

To better understand biology of tumors and its translation in the clinical management scenario it is imperative to obtain representative in vitro models that resemble the different histological and molecular biotypes. In this sense, cell lines constitute an essential tool for basic science and excellent pre-clinical models for developing and testing new drugs. Hence, well-characterized cell lines are desirable to correctly establish the basis of translational cancer research (Figure 21).

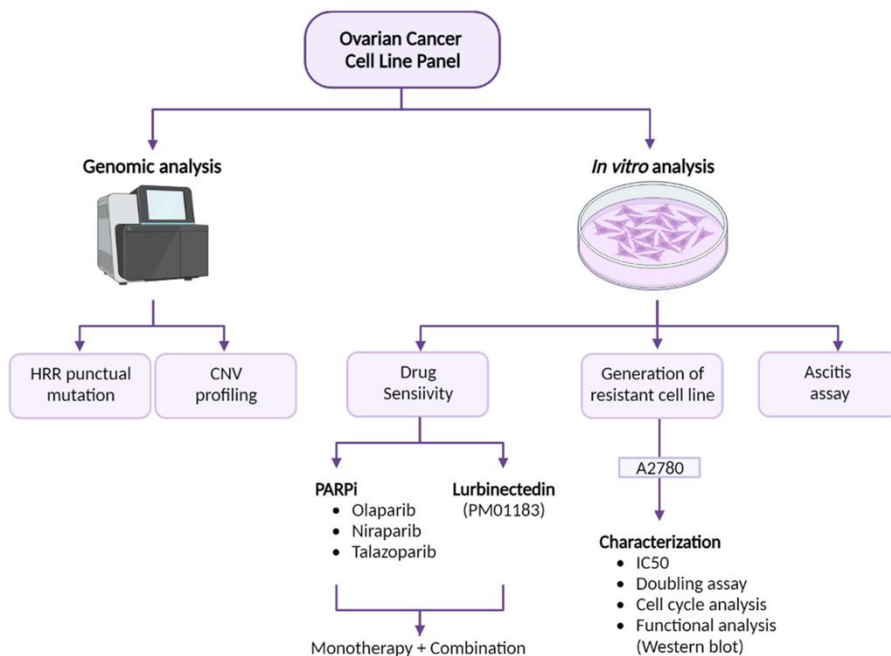


Figure 21: Schematic summary of performed in vitro studies. Cell line panel characterization at genomic and in vitro level.

3.6.1. Cell lines and culturing

A total of 18 OC cell lines, provided by Dr. Marta Mendiola, from La Paz Hospital, were analysed (Annex 2). The complete panel was constituted by: A2780, A2780CIS, PEO1, PEO4, PEO6, TOV112, OVCAR-3, SKOV-3, SCCOHT1, BIN67, PEA1, 59M, TO14, PEO14, PEO23 and PEA2. All cell lines were obtained from the European Collection of Authenticated Cell Cultures (ECACC) and the American Type Culture Collection (ATCC). Short Tandem Repeat (STR) profiles were verified (AmpFISTR® Identifiler® PCR Amplification Kit,

AppliedBiosystem)(Annex 3). Six cell lines were cultured for testing treatment schemes. These cells grew as monolayers and were maintained at 37°C in an incubator with humidified air with 5% CO₂. Disaggregation of the attached cells was performed by trypsinization between passages. A2780, A2780 CIS, OVCAR3 and SKOV3 were cultured in RPMI1640, supplemented with Fetal Bovine Serum (FBS) at 10%, 1% ml of Sodium Pyruvate at 100 mM and 1% Glutamine. OVCAR3 was also supplemented with 1% Hepes 10 mM, 2.5 mg/ml of glucose and 20% FBS. PEO1, PEO4 y PEO6 were cultured in DMEM with FBS at 10%. TOV112D cell were grown in a 1:1 MCDB 105 and Medium 199 with 15% FBS supplemented with 1% Glutamin and 1/1000 Penicilin/streptamicin.

3.6.2. Treatment schemes, drug-sensitivity assays and combination treatment

In vitro tumor models are a necessary tool for assessing the effectiveness of new and available drugs. A broad spectrum of cytotoxicity and cell viability assays are currently in use. Between them, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) is the colorimetric assay most commonly used and the chosen approach for this study. This assay determines principally cell viability through determination of mitochondrial function of cells by measuring activity of mitochondrial enzymes such as succinate dehydrogenase. In this way, reagent will develop a colour in response

to viability, allowing measurement via spectrophotometer in a cheap, reliable and reproducible short-term way. The determination of the cytotoxic activity of several chemotherapeutic agents and identification of variations in susceptibility of different target cells will be performed.

Response curves were generated for Lurbinectedin (PM00183) facilitated from Pharmamar (Madrid, Spain) and PARPi: Olaparib (AZD2281, Ku-0059436), Nirapaib (MK-4827) and Talazoparib (BMN 673), all from Selleck Chemicals GmbH (Munich, Germany). Prior to perform cytotoxic assays, different concentration ranges and exposition times were tested. Cells were continuously exposed to growing concentration of the PARPi in the μM scale during 72, 96 and 120 hours. When referring to PM01183, however, exposition times were 1, 24 and 72 hours and the concentrations used were on the nanomolar range. Finally, the cell lines were cultured in the presence of drugs, both PARPi and PM01183, for 72 hours with concentration ranged as specified in Table 10.

To study possible synergistic effects between PM01183 and different PARPi, cell were simultaneously treated with PM01183 at increasing concentrations specified in Table 10 and fixed dose, consisting on the Inhibitory Concentration (IC) 10 and IC30, of the corresponding PARPi following the same treatment scheme of monotherapy assays.

Table 10. Tested range of drug-concentrations during cytotoxic assays. Different concentrations used during monotherapy and combined treatment assays to obtain dose-response curves for Lurbinectedin, Olaparib, Niraparib and Talazoparib.

Drug	Molarity Range
Olaparib	100-1,5625 μ M (Dilution factor 2)
	25-0,0343 μ M (Dilution factor 3)
Niraparib	25-0,0343 μ M (Dilution factor 3)
Talazoparib	25-0,0343 μ M (Dilution factor 3)
Lurbinectedin (PM01183, PM)	10-0,15625 μ M (Dilution factor 2)
Lurbinectedin (PM01183, PM)	2-0.0625 nM (Dilution factor 2)
In combination	

The MTT colorimetric assay, that measured the metabolic activity of cells by evaluating their capability to reduce the MTT into formazan, its insoluble form, was the methodology applied to determine the chemosensitivity of the cell lines. Response curves were generated for the above-mentioned drug in monotherapy and in combination, looking for synergistic effects. For drug-combination assays, cells were cultured in presence of growing concentrations of PM and fixed PARPi concentration, specific for each cell line and drug (corresponding to IC30 and IC10). Experiments were performed in triplicates. Control cells were treated with appropriated medium.

Absorbance was measured at 570 nM in an Ultra reader (Tecan Group).

3.6.3. Generation of PM01183-Resistant Cell line

The development of resistance is a crucial mechanism for the evasion of tumor cells. However, the mechanisms that result in develop resistance have not been elucidated yet. Preclinical studies provide a key contribution to identify molecular bases of resistance. During this study, Lurbinectedin-resistant cell line was generated and characterized at different level to extend the understanding of the potential mechanisms of resistance. Induction of resistance was done over A2780 cell line.

To generate resistant cell lines, two main approaches could be followed: clinically relevant (one-shot) and high-level laboratory models. The selection of the methodology should rely on the aim of the study and the characteristics of the drug and cell line. Basically, the first approach consists on pulses of treatment follow by recovery in free medium, mimicking the cycles of chemotherapy. This strategy generates a low-level and unstable resistance causing small molecular changing but resembles more reliably the clinical scenario. The second approach, however, consists on continuous exposure escalated doses of the drug. Even if is less relevant to the clinic, this

strategy is useful when studying the mechanisms of resistance, promoting bigger molecular changes easier to identify.

Both treatment strategies were tested to obtain the in vitro resistant model. One-shot strategy consisted of a 1h exposure to highly-concentrated PM01183 (10 nM) and following drug elimination and addition of fresh medium. After 72 hours cells were sub cultured. Concurrently, a second treatment approach was performed. Cells were exposed to concentrations equivalents to IC30 values of PM01183 during 72 hours, when cells were sub cultured with fresh medium and same drug concentration.

3.6.4. Characterization of resistant cell line

The first step to evaluate the evolution of the induced resistance to PM01183 was to determine changes on IC50. As previously mentioned, a drug-response curve was performed over the newly resistant cell line.

3.6.4.1. Cell cycle evaluation

To study the molecular basis of the resistance, the impact of lurbinectedin on cell cycle regulation was studied. As a consequence of the intrinsic ability of lurbinectedin to bind the minor groove of DNA, the drug interferes with the cell cycle, delaying progression

through the S-phase. Lurbinectedin rapidly promoted a cell cycle arrest in the S-phase and the activation of the DNA damage response. To evaluate the type of resistance induced in the A2780 cells, a cell cycle analysis was applied.

The most common method for cell cycle analysis is flow cytometry. This methodology will enable the identification of the cell distribution during the phases of the cell cycle, recognizing the different stages: the G1, S (DNA synthesis phase), G2/M (mitosis). Briefly, the fluorescent dye, which binds to DNA, is incubated with a cell suspension of permeabilized or fixed cells. Since the dye binds to DNA stoichiometrically, the amount of fluorescent signal is directly proportional to the amount of DNA, enabling the distinction of the different phases.

To evaluate the impact of the resistance over cell cycles process, response to PM01183 after 24, 48 and 72 h of exposure to IC30 concentration was assessed on non-resistant and resistant A2780 cell lines. Once cells were seeded and treated during different time-exposures, they were trypsinized, washed and suspended in 1 ml of Phosphate-buffered saline (PBS). Then, 3 ml of 96% iced ethanol was added drop by drop in continuous shaking to fix the cells. Ten minutes centrifugation at 850g, discarding the supernatant and washing on PBS + 5% FBS was done. Finally, 600 μ l of PBS with propidium iodide (50 ng/ml, 50 μ l/ml) plus 1 mg/ml of RNase (Ribonuclease A) were added to cells and incubated overnight in the

dark at 4°C. Following incubation, results were readed in the cytometer. Stained nuclei were analysed with an LSRFortessa system (Becton-Dickinson), data were analyzed and processed with the FlowJo software (Tree Star Inc).

3.6.4.2. Doubling time assay

In addition to cell cycle analysis by flow cytometry, variations in the dynamics of the development and growing of cells in culture were also evaluated by doubling time assay. In this way, A2780, A2780 CIS and A2780 resistant to PM01183 were compared. Cells were seeded at the same starting concentration. Then, they were counted by trypan blue methodology at four different times: 24, 48, 72 and 96 h.

3.6.4.3. Western Blot analysis

Continuing with the characterization of the resistant cell line, a western blot assay was performed. Western blot is used in research to separate and identify proteins and it works as a valuable semi quantitative tool to better understand molecular events, signalling pathways and mechanisms involved in the acquired resistance.

In this technique a mixture of proteins is separated based on molecular weight through gel electrophoresis. Then, the resultant is transferred to a membrane and incubated with labels antibodies

protein-specific. The bound antibodies are then detected by developing the film. The thickness of the band corresponds to the amount of protein present.

Specifically in this study, to assess fluctuations at protein level after treatment with PM01183, protein lysates were analyzed after 72 hours of treatment. Evaluated cell lines were: A2780, A2780 CIS and A2780 PM and interrogated proteins were γ H2AX, RAD51 and PARP (Table 11).

Table 11: Western Blot antibodies applied to interrogate HRR functionality in cell lines.

Antibody	Host	Dilution	Application	Manufacturer and reference
anti-RAD51	rabbit polyclonal antibody	1:1000 (WB)	IF/WB	(PA5-27195, Thermofisher)
Anti-phospho-histone γH2A.X	mouse monoclonal antibody	1:5000 (WB)	IF/WB	(Ser139, clone JLW301, MERCK)
Anti-α-tubulin		1:500	WB	
Anti-PARP	Rabbit monoclonal antibody	1:1000	WB	46d11 9532, Cell Signaling
Anti-RAD51	Rabbit polyclonal antibody	1:100	WB	H-92:SC-8349
γH2AX	Rabbit polyclonal antibody	1:1000 (Biotinilado) En TBS-T/BSA	WB	A300-081A-M Bethyl laboratory

The cell pellet was lysed by adding 50-100 μ l of lysis buffer (25 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8 and 1% SDS supplemented with cocktail of protease and phosphatase inhibitors; leupeptin, pepstatin, and phenylmethylsulfonyl fluoride (PMSF)). Cells were incubated for 10 minutes at 95°C and centrifuged for 10 minutes at 16000g. The entire process was performed maintaining cells on ice. Once the lysate was obtained, protein content was quantified using the commercial BCA Pierce™ kit (ThermoFisher) following the manufacturer's instructions. Starting quantities of 40 μ g of proteins per sample were used. Proteins were separated on a 14% SDS-PAGE gel at 100V. Once finished, proteins were transferred to a nitrocellulose membrane. Transference was carried out at 4°C, 30 V for 3 hours. Subsequently, the membranes were blocked for 1 hour in 5% milk in Tris-Buffered Saline with Tween (TBS-T) buffer. Then, they were incubated overnight with primary antibody in 5% milk at 4°C. After primary antibody incubation, four TBS-T washes were performed for a total of 30 min. Finally, the membrane was incubated at room temperature with secondary antibody for 1 hour and washed four times with TBS-T.

Membrane was revealed using the ECL™ transfer system (Sigma Aldrich) according to the manufacturer's instructions. For normalization of protein concentrations, the membrane was also incubated with the selected housekeeper (control).

3.7. In vitro data processing

Dose-response curves were generated with GraphPad Prism 6 software. Statistical significance was determined applying unidirectional and bidirectional ANOVA test or two-tailed t-student test. Image processing and quantification of western blots was done by Quantity One software.

CompuSyn (CompuSyn, Inc), based on Chou-Talalay method, was employed to establish synergistic or antagonistic combinations between tested drugs and to obtain specific combination indexes (CI).

3.8. Statistics

3.8.1. Descriptive statistics

A Chi-square test and Fisher's exact test were used to compare categorical GI and clinical and pathological variables. For continuous variables, non-parametric Wilcoxon and Kruskal-Wallis tests were applied.

3.8.2. Parameters of clinical response

The lack of an optimal gold standard to determine response to different chemotherapeutic and targeted treatments has promoted the use of a wide number of variables, aiming to establish the most

suitable in each clinical scenario. Due to that, the analysis has been performed considering all the response variables that were available during the study. Evaluation of platinum response has been done based on PFI defined as the time-lapse between end of platinum-based treatment and relapse. Response is categorized in three levels; refractory patients ($PFI < 6$ months), partially sensitive ($PFI \geq 6$ and < 12 months) and sensitive ($PFI \geq 12$ months).

Aiming to evaluate the response to PARPi, two additional concepts were used: Overall response rate (ORR), which is the proportion of patients in a study who presented partial or complete response (CRs) to the treatment within a certain period of time. ORR is generally defined as the sum of CRs, the disappearance of all signs of cancer in the body, no detectable evidence and partial responses (PRs) – decrease in the size of a tumor or in the amount of cancer in the body. On the other hand, Clinical Benefit Ratio (CBR), percentage of patients who achieved a CR, PR, or had stable disease for 6 months or more.

Lastly, two general and commonly used parameters to test the predictive and prognostic value of different biomarkers or approaches are PFS/Disease-Free Survival (DFS), defined as the length of time during and after the treatment of a disease, such as cancer, that a patient lives with the disease but it does not get worse, and OS, which is the period of time from the diagnosis until the exitus date.

3.8.3. Survival analysis

For time-to-event variables (i.e. PFI) Survival analysis was performed using Kaplan Meier estimation and signification was obtained by log-rank test. Statistical signification was considered at $p < 0.05$. All tests were two-tailed. Statistical analysis was performed using R studio (R version 4.0.3)

RESULTS

4. Results

4.1. Characterization of genomic Instability

Due to the clinical relevance of high GI phenotype, working as a suitable prognostic and predictive biomarker, the establishment of GI status has become a major challenge for the selection of patients.

4.1.1. Establishing the basis of GI: tuning methodological and analytical approaches

Despite the advance of high throughput techniques, that have provided insights in cancer characterization and are allowing the unveiling of the genetic landscape of tumors, the analysis of GI still represents a challenge.

In this section, the optimization of methodological, analytical and bioinformatics workflows to achieve whole-genome CN profiles from a CNV and SNVs will be explained (Objective 1). For this purpose a series of 45 cases (Cohort 1) was analysed at three different levels as indicated in Figure 14.

4.1.2. Multi-gene NGS panel to determine HRD and clinical impact of molecular classification

Mutational analysis was performed based on the results of a custom NGS-Panel that interrogates 35 DDR genes. Although HRR genes are

overrepresented in the panel, other genes belonging to the Base Excision Repair (BER), NER, and MMR pathways were also included. The whole series was analyzed in two different batches on which sequencing metrics were comparable. Sequencing metrics are summarized in Annex 4.

A median of 99.5 alterations per case (range: 57-459) was reported (Annex 5). After variant selection, a median of 2 alterations per case (range: 0-4) remained. These variants had a median AF of 53 % (range: 8.2%-94.4%) and a median coverage of 805 reads (range: 114-3406 reads). One of the samples (CO1) was not evaluable for mutational analysis; hence, it was not included for the following statistics. As expected, the most frequently mutated gene was *TP53*, 84.1% (37/44), followed by *BRCA1* and *BRCA2*, with an incidence of 27.3% (12/44) and 18.2 % (8/44) respectively. Other HRR genes were also found to be altered with a lower incidence, including: *PALB2* 9.1 % (4/44), *RAD54L* 4.5 % (2/44), *BRIP1* 4.5 % (2/44), *ATM* 4.5 % (2/44), *ATR* 2.3 % (1/44) and *SLX4* 2.3 % (1/44). Additionally, alterations in other DDR genes were also identified: *PMS2* 4.5 % (2/44), *HELQ* 2.3 % (1/44) and *ERCC6* 2.3 % (1/44) (Figure 22). Germline BRCA mutations for those cases are reported in Annex 6. Mutational data was then used to stratify patients between Homologous Recombination Repair Proficient (HRP), absence of mutation in the HRR-related genes, and HRD, presence of an alteration in any of the HRR-related genes, not referencing in any case to the GI status. Hence, 59.1 % (26/44) of

patients were established to be HRD, harbouring mutations in *BRCA1/2* and other HRR genes in 45.5 % (20/44) and 13.6 % (6/44) of patients respectively (Figure 22). Additionally, *CCNE1* amplification (*CCNE1amp*) was also obtained from sequencing data. Among all cases, 5/44 (11.4 %) presented amplification.

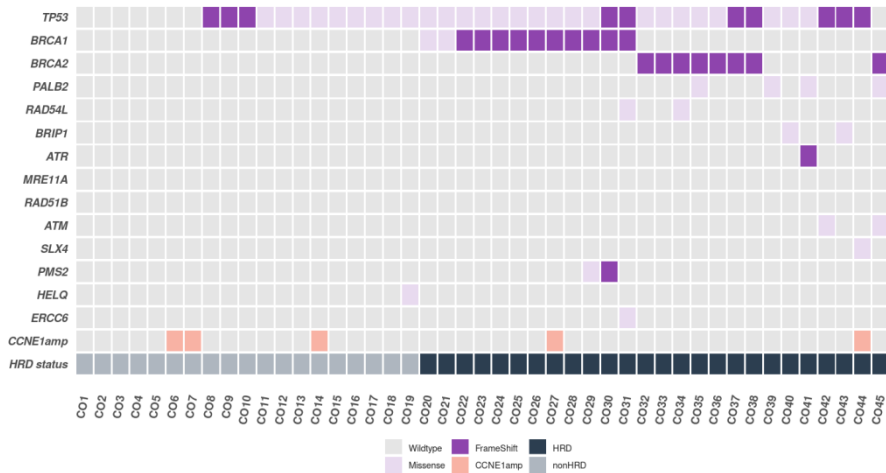


Figure 22: Molecular landscape of Cohort I. Distribution of gene alterations and stratification of patients based on HRR mutational status.

The Log-Rank test was used to evaluate the potential of mutational analysis to predict response to platinum-based drugs. This test revealed differences between HRR groups in PFI when comparing nonHRD and HRD (12.7 and 30.6 months respectively, $p=0.02$) and nonHRD, BRCA HRD and nonBRCA HRD (12.7, 19.5 and 36.8 months respectively, $p=0.045$) demonstrating its prognostic impact (Figure 23).

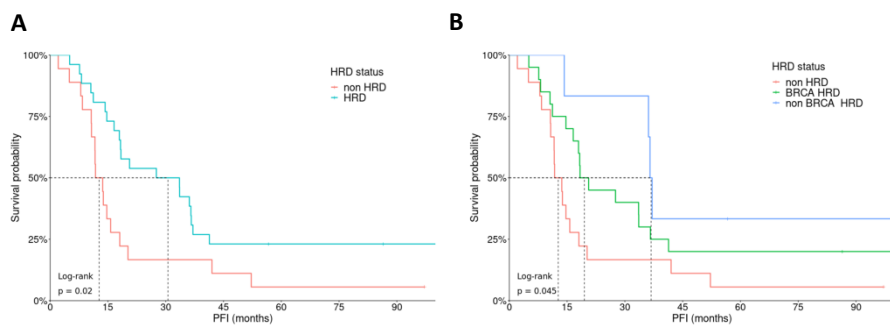


Figure 23: Log-Rank test and Kaplan-Meier plots. A) Two levels corresponding HRD mutated cases (including *BRCA1* and *BRCA2*) and non HRD mutant cases. B) Analysis stratifying population in *BRCA* HRD as cases with mutations only in *BRCA1* and *BRCA2*, non-*BRCA* HRD cases with mutations in HRD excluding *BRCA* genes, and non-HRD cases without mutations in HRR pathway.

4.1.2.1. *Optimization of Genomic Instability detection by tuning selected algorithms*

The detection of CNV, due to its complexity, was performed by using three different pipelines; CNVkit, saasCNV and SureCall. To calibrate the thresholds levels among the numerous CNV derived parameters comparisons at distinct levels were performed. While saasCNV and SureCall are limited to slightly personalization, CNVkit has a greater number of analytical parameters. In this case, segmentation pipeline (CBS or FLASSO), p-value cut-off, personalized or pre-established tumor content of the sample and post analytical filtering were evaluated. Firstly, Segmentation pipelines were evaluated. Differences basically appeared in total number of CN events, higher in CNVkit, while presenting a lower median length. However, the

total percentage of altered genome was not significantly different (Figure 24).

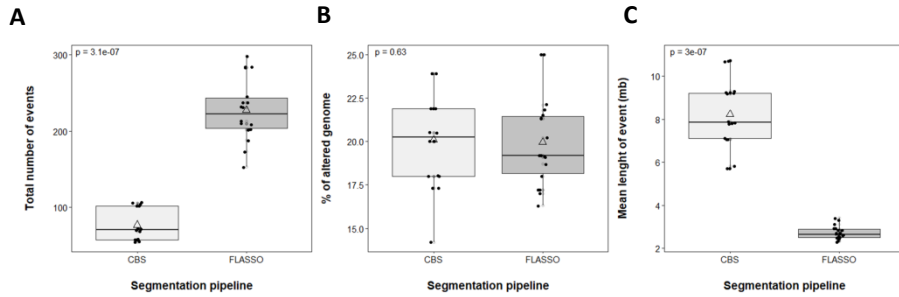


Figure 24: Differences in GI parameters according segmentation used in CNVkit pipeline. Non-parametric tests evaluating performance of CBS and FLASSE regarding A) Total number of events. B) Percentage of altered genome. C) Median length of event

The selected segmentation pipeline, due to higher reliability of events of greater length and the suitability to perform the following analysis to call LOH events, was CBS. A tuning of CNVkit pipeline parameters was performed as follows: A most restrictive of p value in Segmentation algorithm was assayed ($p=0.001$), in this scenario, fragments of greater length are built by segmentation algorithm ($p=0.00035$). The selection of these events caused no differences regarding total number of events while percentage of genome altered was significantly higher ($p=0.00036$) (Figure 25).

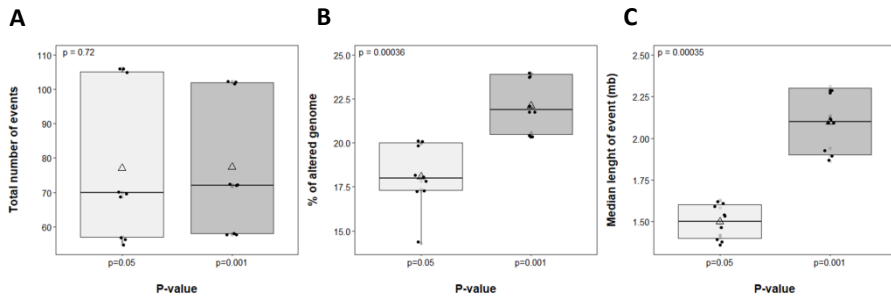


Figure 25: Differences in GI parameters adjusting p-value in CNVkit pipeline. Non-parametric tests evaluating distinct p-value filter regarding A) Total number of events. B) Percentage of altered genome. C) Median length of event

Due to the fact that using a restrictive filter does not mean losing CNV information while obtaining more reliable results, p-value of 0.001 was chosen for the following analysis. Regarding the tumor content, we evaluated if it was decisive in the analysis of CNV. To do that, establishment of values of 50, 80 and 100 % of tumor content for the same sample were applied during the analysis. Results showed significant differences when applying different tumoral content in total number of events ($p=0.00046$), as well as tendency in percentage of tumor altered and CN length (Figure 26).

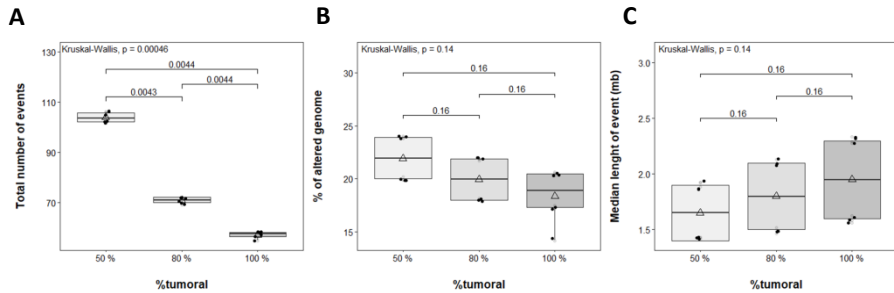


Figure 26: Differences in GI parameters according tumor burden in CNVkit pipeline. Non-parametric tests evaluating the impact of tumor burden regarding A) Total number of events. B) Percentage of altered genome. C) Median length of event

Hence, establishing and applying its own tumor content per each case will be essential for following analysis. Finally, the case of post-analytical filter was slightly different. Filter based on median or quartile CNVkit assigned weight of events were applied after the analysis, followed by another filter by 1 mb in length (Figure 27).

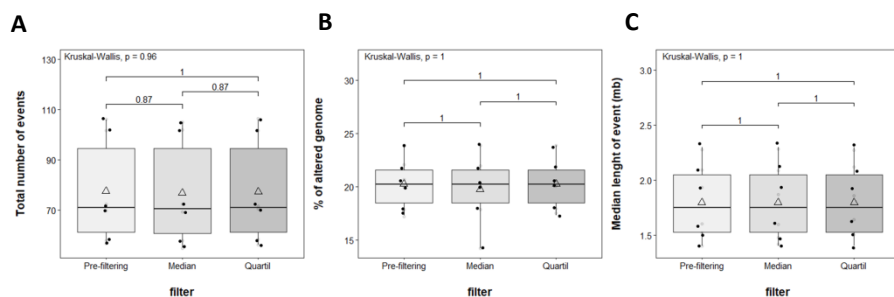


Figure 27: GI parameters according pre-filtering step in saasCNV pipeline. Non-parametric tests evaluating the impact of CNV size filtering regarding A) Total number of events. B) Percentage of altered genome. C) Median length of event

Performing the filtering process before that, resulted on no differences due to the median and 1st quartile value were both below 1 mb. However, considering the high number of total events obtained and trying to simplify following analysis, we decided to apply above median filter after the 1 mb length filter. On the contrary, saasCNV and SureCall were easily adjusted to the nature of the data. For saasCNV, was only evaluated the pre-filtering step during the analytical process. In regards to that, total number of events was considerably lower when not applying pre-filtering steps. Then, the decision was to perform this step, looking for more reliable and handle results. On the other hand, SureCall was applied using the default parameters and chosen healthy control of the performed data.

The final selection of parameters was to apply CBS¹⁹⁰ to incorporate: true tumor content from the sample, instead of its estimation based on sample ploidy; to use a more stringent p-value of 0.001; and to filter results by median weight and fragment size, selecting those shorter than 1 Mb to increase specificity in the CNV calling for CNVkit. SaasCNV adjustment was limited to perform a prefiltering step whereas SureCall was not customized at all.

4.1.2.2. CNVkit: best performing algorithm in GI determination

Once the pipelines were fitted, the ability of each pipeline to assess GI was evaluated. Firstly, values of different GI parameters obtained from the studied pipelines were compared. Significant differences

appeared in all performed comparisons, mainly showing a higher percentage of altered genome in SureCall results, followed by saasCNV and ending with CNVkit, which carries the lowest values (Figure 28).

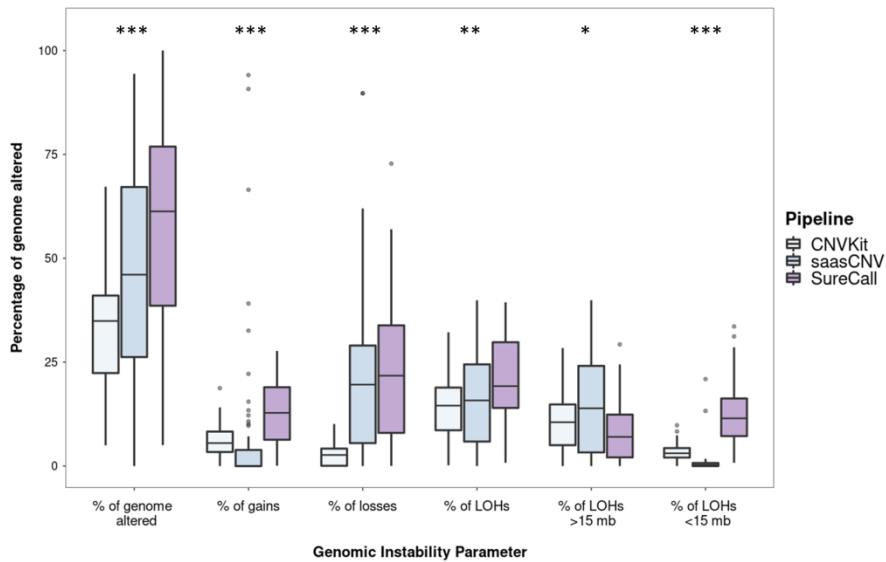


Figure 28: Comparison of GI parameters between implemented pipelines; CNVkit, saasCNV and SureCall. The result of these non-parametric test highlights the weakness and strengths of the pipelines due to differences in the assessment of each feature. (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

Even though all GI parameters followed this hierarchy between pipelines, three differences stood out over the others. CNVkit had worse detection performance of losses in favour of LOHs, probably caused by the analysis methodology to identify this type of alterations, since the baseline to detect heterozygosity loci was

established with a panel of samples. On the other hand, in the case of saasCNV, difficulties appeared in the identification of gain events. Here, there was hardly any gain event detected, showing a weakness of the pipeline. Lastly, Surecall presented a different distribution of LOH events according to size. Whereas saasCNV and CNVkit showed a higher number of LOH larger than 15 mb, SureCall presented more LOH events above 15 mb, having an opposite distribution in size.

Since we lack an objective and robust technical benchmark to compare studied pipelines, the accuracy to determine GI patterns was measured according to clinical outcome, particularly PFI, to endow our findings with clinical significance. Hence, the correlation between each GI parameter from different pipelines and response to platinum was obtained. saasCNV pipeline showed the lowest statistical power in all the evaluated parameters. Among all the established GI parameters, the total number of LOH events presented the highest correlation with clinical outcome in all three pipelines. However, CNVkit continues to lead in this regard. On the contrary, SureCall and CNVkit presented higher statistical significance than saasCNV, each one presenting a correlation with different GI parameters. While CNVkit showed a higher correlation when comparing PFI with global GI (total number of events and percentage of genome altered) as well as LOH events, Surecall presented higher accuracy with parameters related to gains and losses (Annex 7). In all cases, a higher number of CN events, as well as genome altered,

were correlated with longer responses to platinum-based chemotherapy (Figure 29).

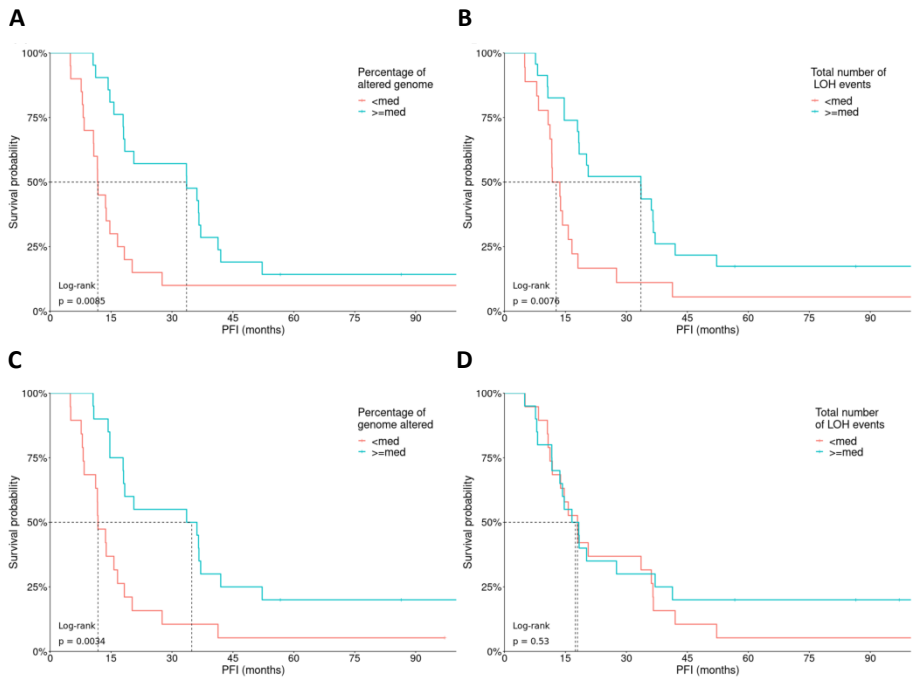


Figure 29: Differences in PFI between evaluated pipelines. Log-rank test interrogating the power to assessed A) Percentage of altered genome and B) Total number of LOH events detected by CNVkit C) Percentage of altered genome and D) Total number of LOH events detected by SureCall.

In addition, mutations in HRR genes were faced with GI parameters (Annex 8). Even though not all GI parameters correlated with the presence of mutations, higher global GI had been shown in HRD patients, particularly in those that harbored mutations in BRCA1/2 (Figure 30).

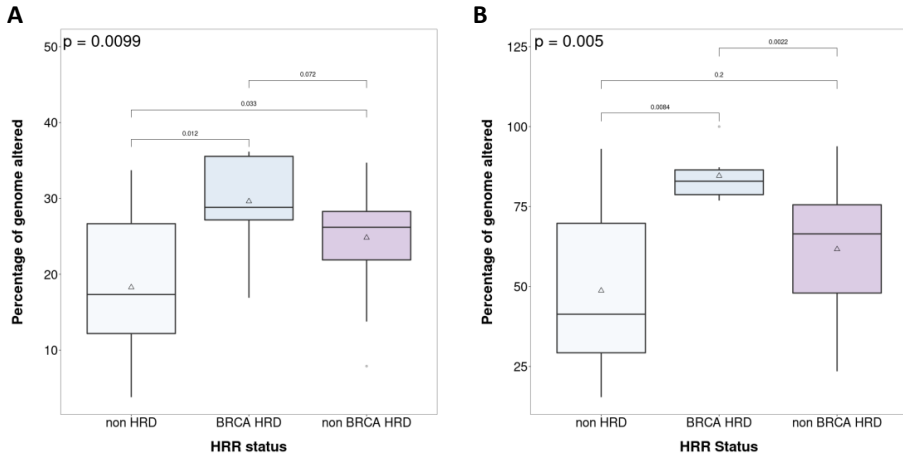


Figure 30: Comparison of percentage of altered genome between HRR-based groups. Differences, assessed by non-parametric tests, were analysed in both pipelines, A) CNVkit B) SureCall.

Finally, CNVkit was chosen according to best performance, higher number of customizable parameters, advantageous use of on-target and off-target read counts, widening the coverage of the panel, and suitability to medium size NGS panels. Hence, following analysis to perform and adjust the predictive model were done using CNVkit pipeline. The relationship between GI parameters and survival is described in Annex 9.

4.1.2.3. *scarHRD pipeline*

Complementary to the establishment of GI by CNVkit, the pre-defined genomic scars (LOH, LST and TAI) were also obtained.

Genomic Scars has been correlated with response to DNA-damaging agents and presence of *BRCA1/2* mutations, used as a surrogate of HRR deficiency. However, annotation data from a clinically validated approach was not available at the time of the study. For that reason, we decided to infer it from our sequencing data by using scarHRD package from R¹⁸⁸. Genomic scars, except from LOH parameter, together with the HRDscore resulted significant when faced to PFI intervals (Figure 31).

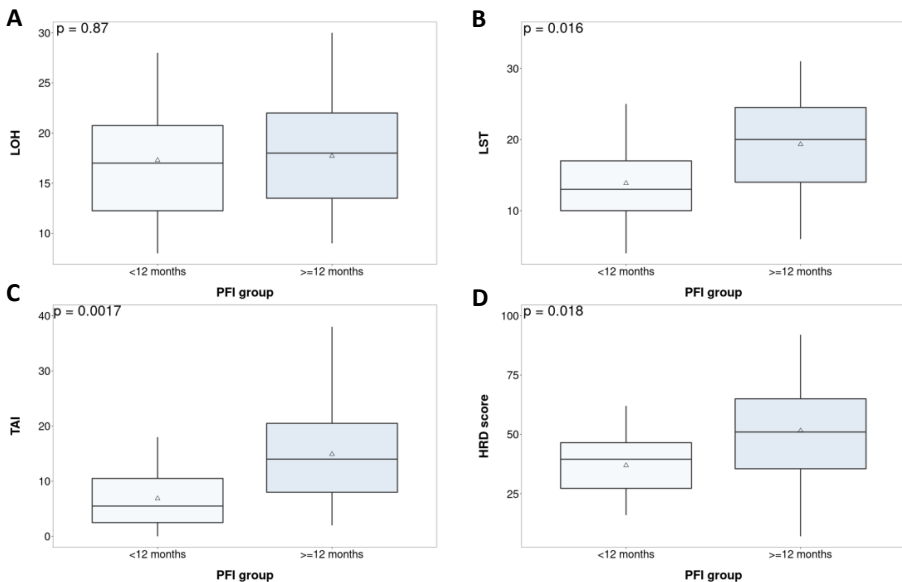


Figure 31: Distribution of different genomic scars and integrative HRDscore between PFI-based stratification. Non-parametric tests evaluating the Incidence of events of A) LOH, B) LST, C) TAI and HRDscore between PFI groups.

Comparisons between Genomic Scars and HRDscore and HRR mutation were also performed. Similar to PFI results, all parameters

except from LOH were significantly correlated with the presence of HRR mutations (Figure 32).

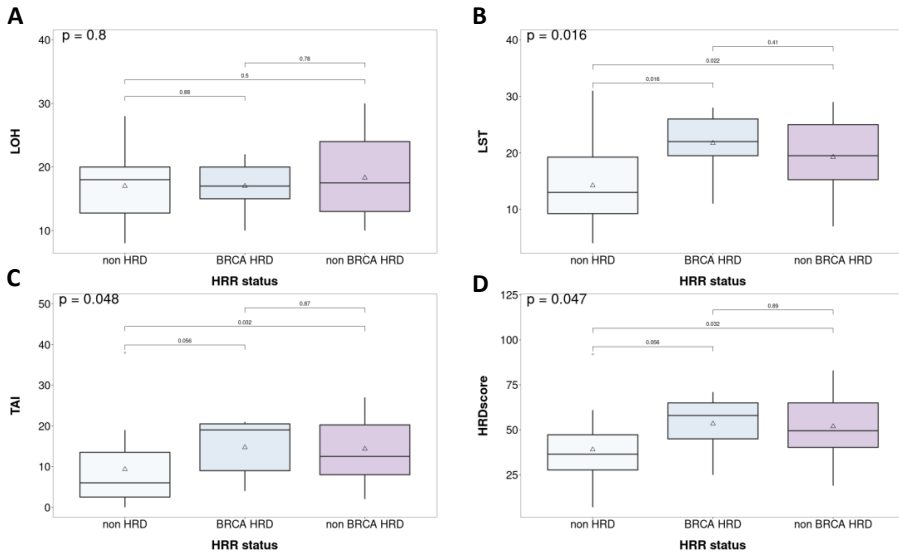


Figure 32: Distribution of different genomic scars and integrative HRDscore between HRR mutations-based stratification. Non-parametric tests evaluating the incidence of events of A) LOH, B) LST, C) TAI and HRDscore between HRR status-based groups.

4.1.3. Predictive model fitting and validation to predict response to DNA-damaging agents

Once optimized the approach to determine GI parameters, we outline a useful academic tool (integrative model) to predict response for DNA-damaging agents in the clinical scenario of HGSOc and, potentially, other HRR deficient tumours in an independent

series of 190 patients as showed in Figure 15 (Objective 2 and 3, Cohort 2).

In order to adjust a ML method to predict response to Platinum-derived therapy, data from three different sources was used to feed the model.

The first one derives from the raw NGS coverage information of 170080 SNPs, while the third model contains gene expression data of 2549 genes obtained from targeted RNAseq experiments. First step in the protocol was denoise the raw data by feature selection as explained in Materials and Methods section. On the other hand, the second model studied the most representative parameters of the GI phenomenon, not being subjected to feature selection due to the reduced number of them. Then, each set of selected parameters was tested coupled with a data mining algorithm. Every possible combination between the algorithm and selected features was checked.

The best-individualized performance models comprise SVM with 8 SNPs (Annex 10), SVM with 28 GI parameters and NN with the expression of 7 genes (Annex 11). Results for each adjusted model are shown in Table 12. Among the three single-source models, the best performance was obtained by the GI model, with an accuracy of 0.8909. Finally, the ensemble model was based on an SVM algorithm, using as an input the 43 previously mentioned selected attributes.

The model was trained with a bootstrapping of 500 iterations, obtaining an accuracy of 0.991 and a kappa index of 0.981, outperforming all the three single-source models.

Table 12: Performance of the single-source and ensemble model. Differences between models regarding capability to predict response to PFI.

	TP/TN/FP/FN	Accuracy (95 % CI)	Sensitivity Specificity	Kappa
SNPs Model	29/13/5/5	0.8077 (0.6747-0.9037)	0.7222 0.8529	0.5752
HTG Model	25/17/1/9	0.8077 (0.6747-0.9037)	0.9444 0.7353	0.6154
GI Model	35/14/2/4	0.8909 (0.7775-0.9589)	0.8750 0.8974	0.7450
Ensemble Model	34/16/2/0	0.9615 (0.8679-0.9953)	0.8889 1.0000	0.9128

*TP: true positive, TN: true negative, FP: false positive, FN: false negative

In addition, the clinical prognostic of each model was tested by using a log-rank test with PFI as a time-to-event variable. All four models, including the ensemble, presented a significant p-value below 0.0001 (Figure 33). While HTG-based model was the one with the lowest predictive power, otherwise, the highest statistical significance was obtained in the ensemble model ($p > 2 \times 10^{-16}$).

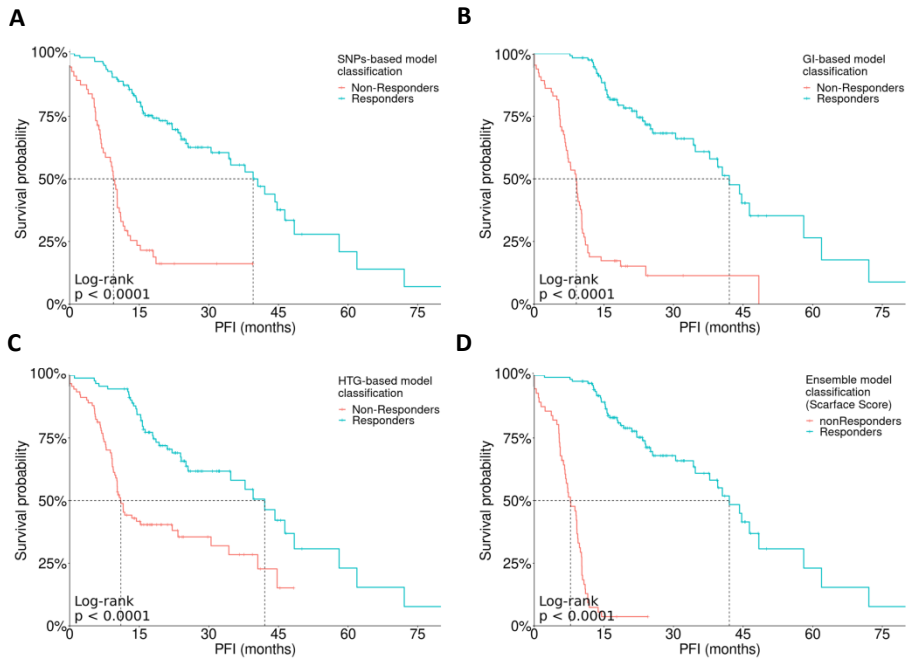


Figure 33: Correlation between PFI and fitted models. Log-rank tests evaluating the performance of: A) SNPs-based model. B) GI-based model. C) HTG-based model. D) Integrative ensemble model.

The goodness-of-fit of the predictive algorithms was also evaluated by ROC curves, showing how well each predictive model discriminates between patients with PFI under or over 12 months. The comparison was carried out in terms of the extracted values of Area under the Curve (AUC). As expected, the highest discriminant power was obtained with the ensemble model, with an AUC of 0.962 and sensitivity and specificity of 0.929 and 0.945 respectively (Figure 34).

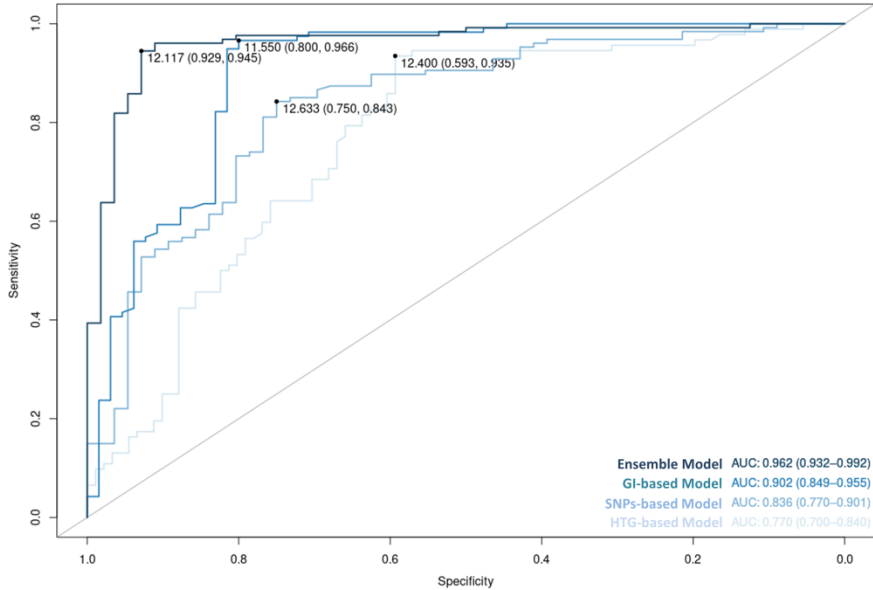


Figure 34: ROC curves comparing performance single-source and ensemble model.

Albeit the algorithms were trained to predict response to platinum-based chemotherapy, the definitive purpose of the study was to fit a model able to establish the candidates for PARPi therapies. Hence, the capability of the models to discriminate the best responders to this type of drug was also tested by log-rank test in a sub-series of 58 patients. Performance of the models was faced with *BRCA* mutation-based stratification, as a gold standard method to select patients to receive PARPi. In this case, the ensemble model outperformed *BRCA*-based classification ($p=0.0048$) with a p -value of 0.00077 (Figure 35), improving the discriminant power of this gold standard.

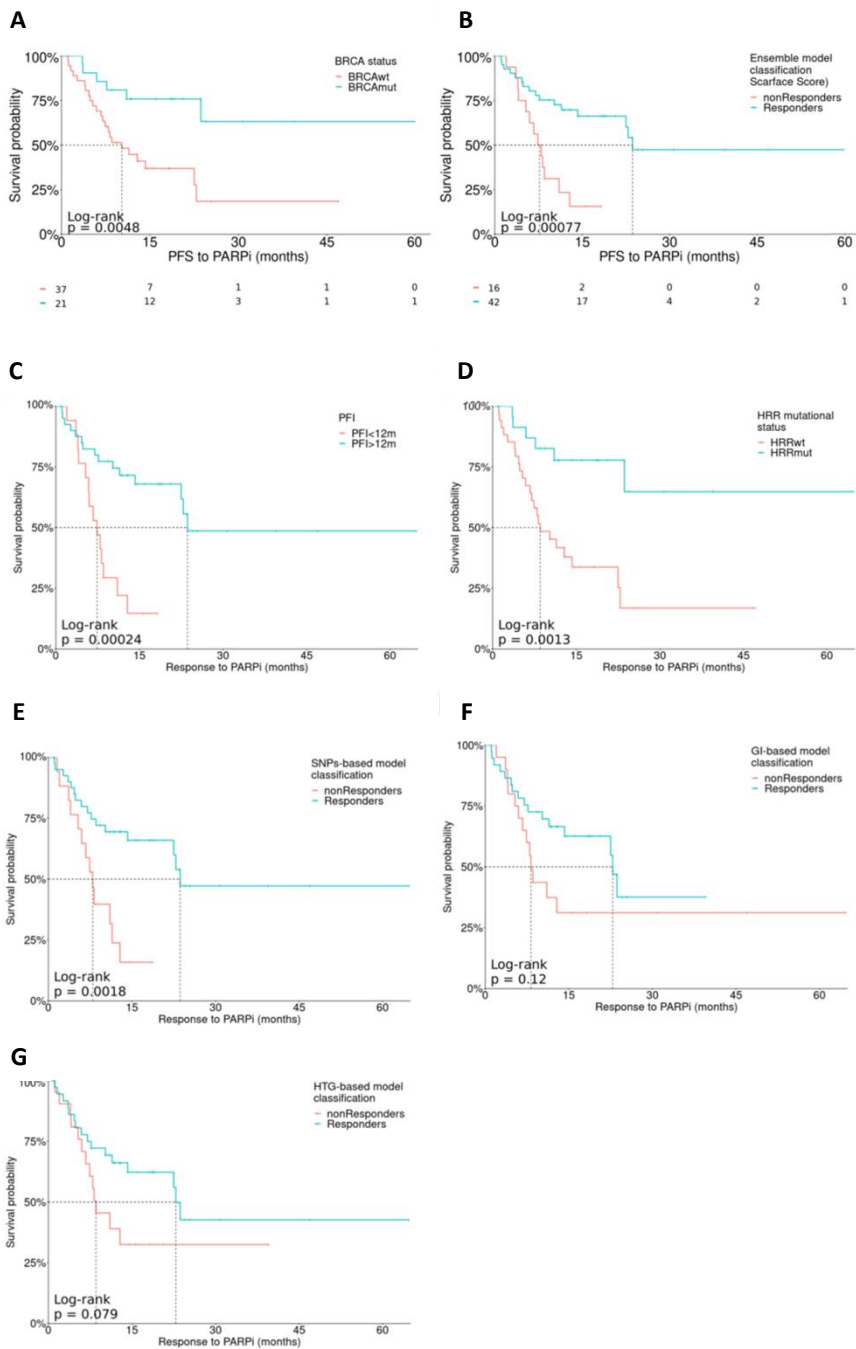


Figure 35: Log-rank tests evaluating the performance of different tested molecular classifiers to predict PFS to PARPi. A) BRCAmutation-based classification, B) Integrative ensemble model C) PFI, D) HRR-based classifier E) SNP-based model, (F) GI-based model and G) HTG-based model regarding PFS to PARPi

The impact of the predictive model over OS was also evaluated. All models reached statistical significance, headed by the ensemble model (Figure 36).

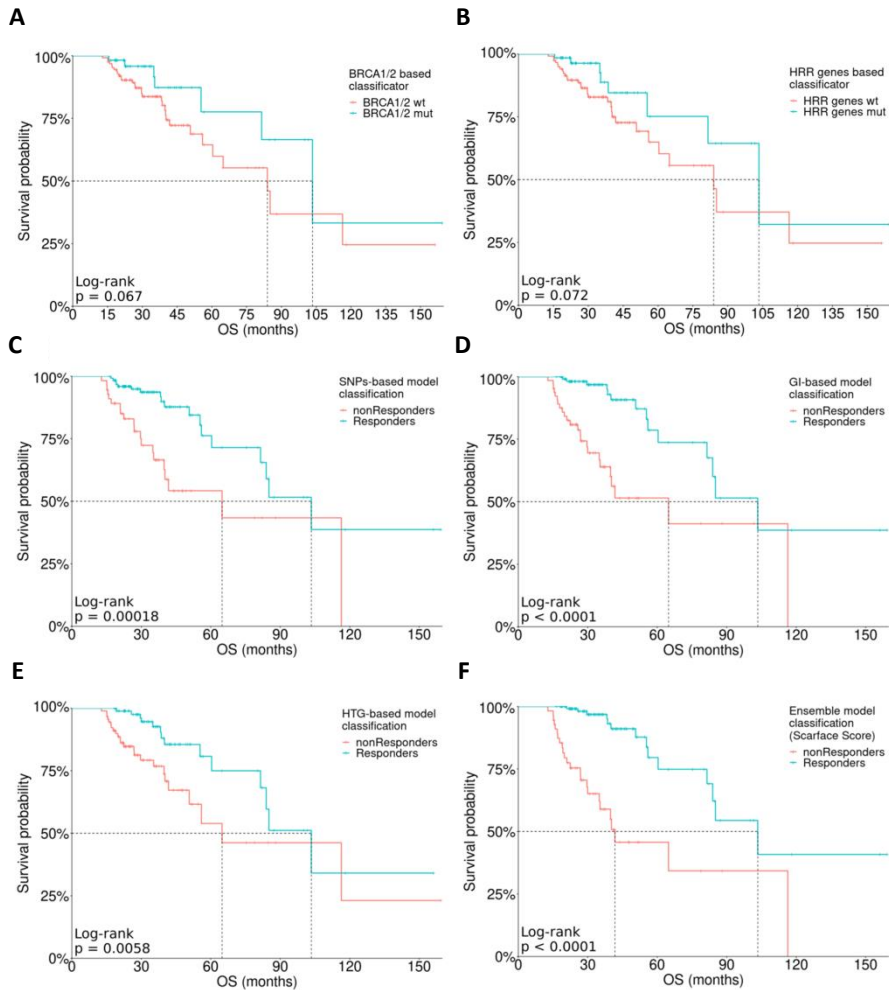


Figure 36: Log-Rank tests evaluating the implication of mutational-based classifiers and predictive models with OS. (A) BRCA1/2-based classifier (B) HRR-based classifier (C) SNP-based model, (D) GI-based model, (E) HTG-based model and (F) Ensemble model

In addition to model performance, a multivariate analysis was performed in order to evaluate the contribution of different clinicopathological and mutational parameters to the stratification of patients. The most discriminant parameter was the ensemble model prediction (HR=0.12). However, other parameters such as tumor extension (Locally advanced; HR=2.18, Metastatic; HR=3.18) and HRR mutation (HR=0.36) also contributed to risk assessment (Figure 37).

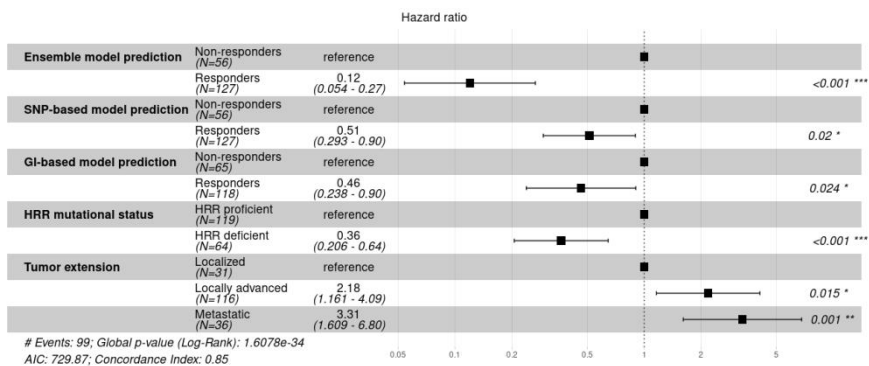


Figure 37: Multivariate analysis in terms of PFI. Analysis performed by Cox regression for clinicopathological parameters, HRR alteration and three-source model performance in addition to ensemble model.

Additional cox analysis was performed, evaluating higher number of variables (Figure 38).

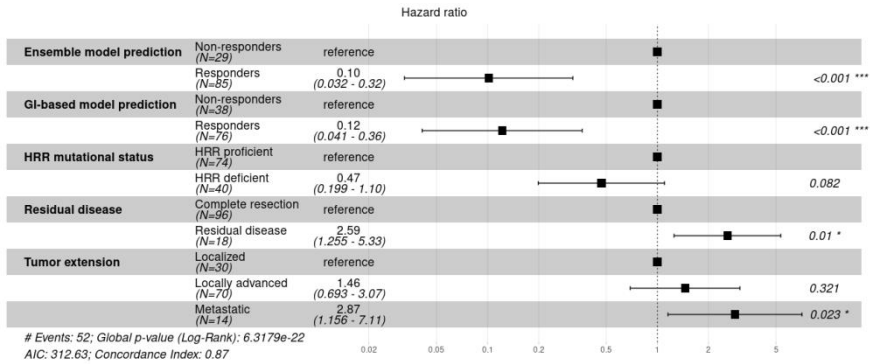


Figure 38: Multivariate analysis in terms of OS. Analysis performed by Cox regression for clinicopathological parameters, HRR alteration and three-source model performance in addition to ensemble model.

4.1.4. Another clinical scenario: POLA phase II cohort

On the other hand, the clinical scenario of PARPi and its correlation with GI was also explored through a series of OC and EC patients enrolled in Phase II clinical trial (POLA: NCT02684318, EudraCT 2015-001141-08, 03.10.2015).

4.1.4.1. Distribution of Genetic Alterations and Clinical Impact

Genetic studies were performed on a total of 57 samples that passed the quality and quantity requirements, corresponding to 19 (33.3%) EC and 38 (66.7%) OC patients. Among all of the mutated genes, considering both cancer types, *TP53* and *PTEN* presented the highest mutational ratios, with 34/57 (59.6%) and 9/57 (15.8%), respectively. *TP53* alterations were mainly present in OC (70.6%), specifically in HGSOC histology, while *PTEN* was preferentially altered in EC (88.9%).

Regarding the HRR pathway, a total of eight genes presented alterations, including *BRCA1* (3, 5.3%), *BRCA2* (1, 1.8%), *ATM* (2, 3.5%), *RAD5L* (1, 1.8%), *ATR* (1, 1.8%), *NBN* (1, 1.8%), *SLX* (1, 1.8%), and *WRN* (1, 1.8%). Overall, HRR gene alterations were reported in 10/57 (17.5%) cases homogeneously distributed between EC and OC, and they were used in the following analysis as an HRD status subrogates. Additionally, mutations in the Fanconi Anemia genes, *FANCM* (1, 1.8%) and *FANCA* (1, 1.8%), were also found. Finally, alterations in the MMR genes were described in two EC cases (Figure 39).

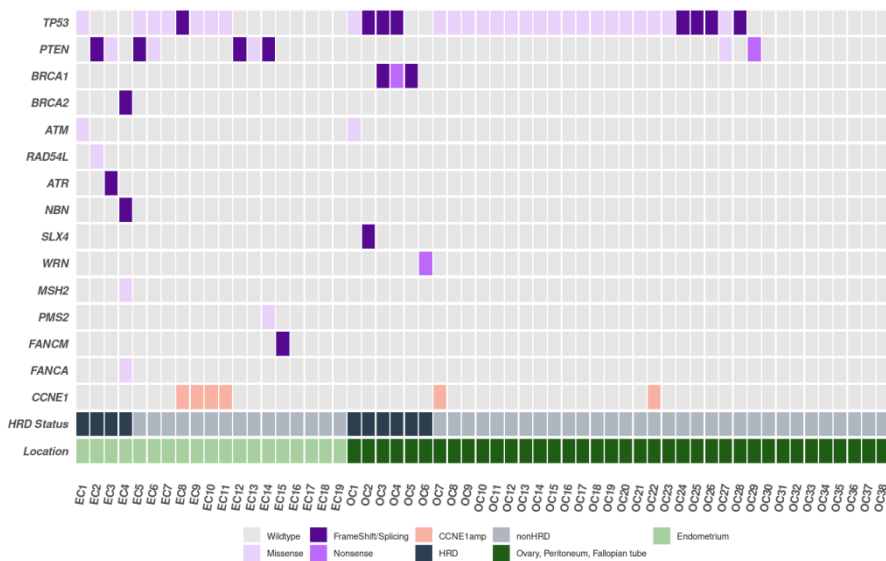


Figure 39: Molecular landscape of Cohort 2. Oncoplot of genetic and genomic alterations across the 35 genes of the custom panel, related to HRR and other DNA repair pathways across the EC (n = 19) and OC (n = 38) cohorts. The Oncoplot shows SNVs and CN at the gene level in CCNE1.

We studied the possible relationship between HRD status and response to treatment. No correlations were found between study treatment ORR or CBR, and HRR mutations. The different GI parameters were compared with the mutation-based stratification. In the whole population, HRD status was associated with losses ($p = 0.0038$) and the percentage of the genomes affected by losses ($p = 0.034$) (Figure 41A and B). Considering that GI caused by HRR gene mutations has been principally described in the OC population, we studied GI patterns according to cancer type (Figure 40).

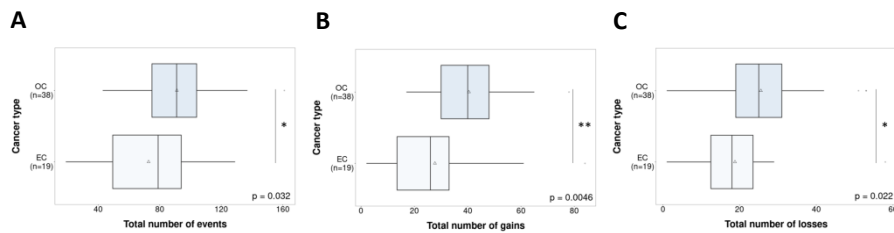


Figure 40: Comparison of GI patterns between cancer types. A) Total number of events, B) Total number of gains, and C) Total number of losses. Non-parametric Wilcoxon Signed Ranks Test was used.

The OC cohort ($n = 38$) showed a significant correlation between HRD status and the total number of events ($p = 0.0059$), loss events ($p = 0.0013$), and percentage of the genome affected by losses ($p = 0.014$)

(Figure 41C, E and F). LOH did not correlate with HRR status ($p = 0.091$). On the other hand, the EC cohort ($n = 19$) did not show any significant results.

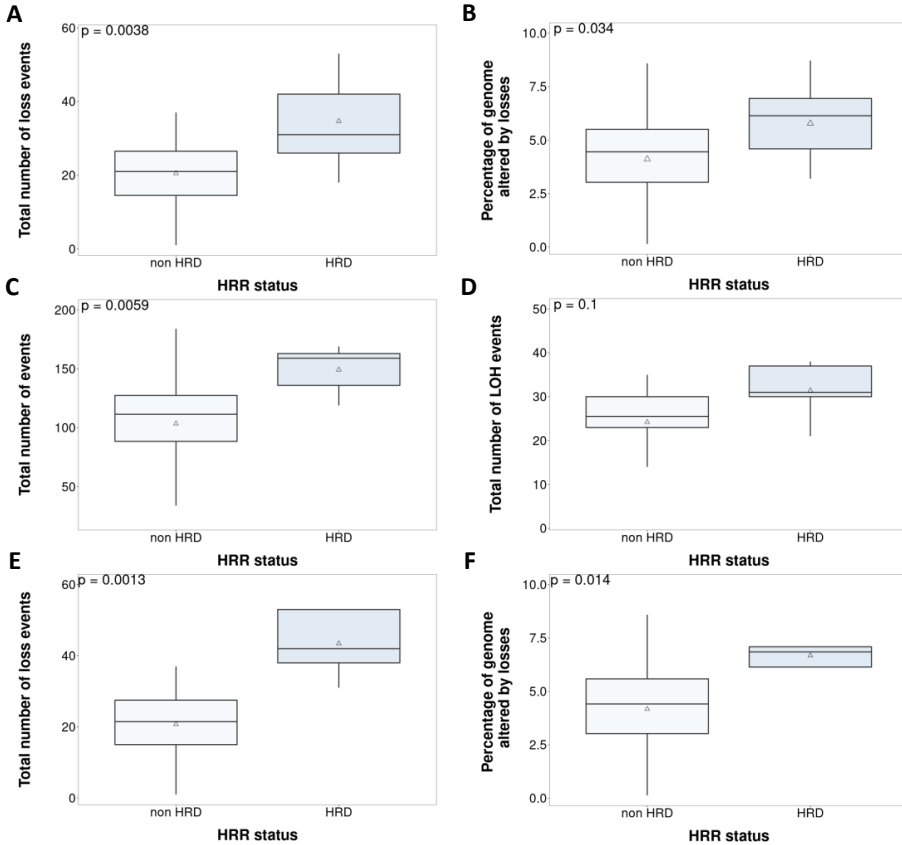


Figure 41: Non-parametric tests (Wilcoxon signed rank test) comparing the GI parameters and HRD status in the whole population. A) loss events and B) Percentage of altered genome by losses and in the OC population ($n = 38$) in C) Total number of events, D) LOH events, E) loss events and F) Percent of altered genome by losses.

4.1.4.2. Characterization of Copy Number Patterns across the Clinical Trial Population: Clinical Impact of Genomic Instability-Based Classification

Finally, the GI parameters were evaluated as a predictive biomarker for the combination of Olaparib and Lurbinectedin. First, in terms of duration of response, Long-Term Responders (LTRs) were assessed. When evaluating the OC population specifically (n = 27), we observed a trend towards an association between LTRs and total LOH events, which did not reach statistical significance with the current sample size (p = 0.055) (Figure 42B). Second, the relationship between GI and ORR was also evaluated. The total number of LOH events was not associated with ORR (p = 0.074) (Figure 43A).

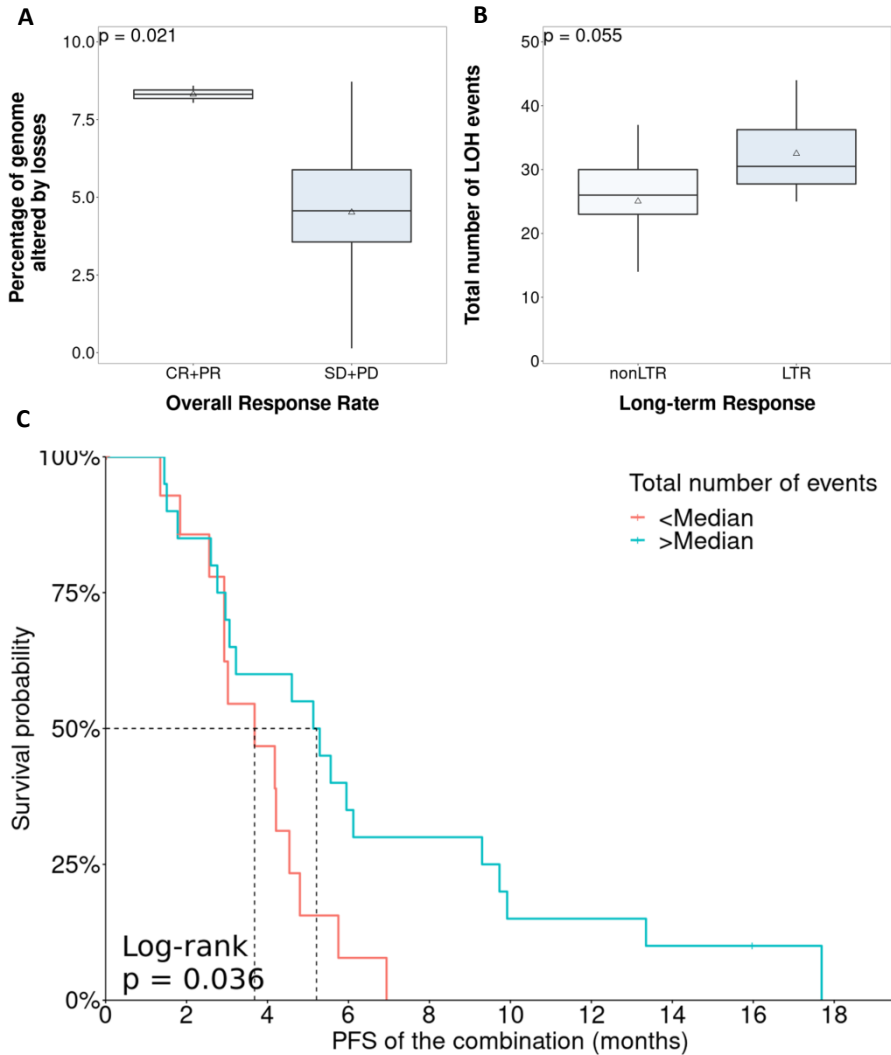


Figure 42: Clinical implication of the GI parameters in the OC population (n = 38). A) Total number of events regarding overall response rate; B) total number of LOHs in LTRs C) survival analysis stratification due to the total number of events. Non-parametric and log-rank tests were used.

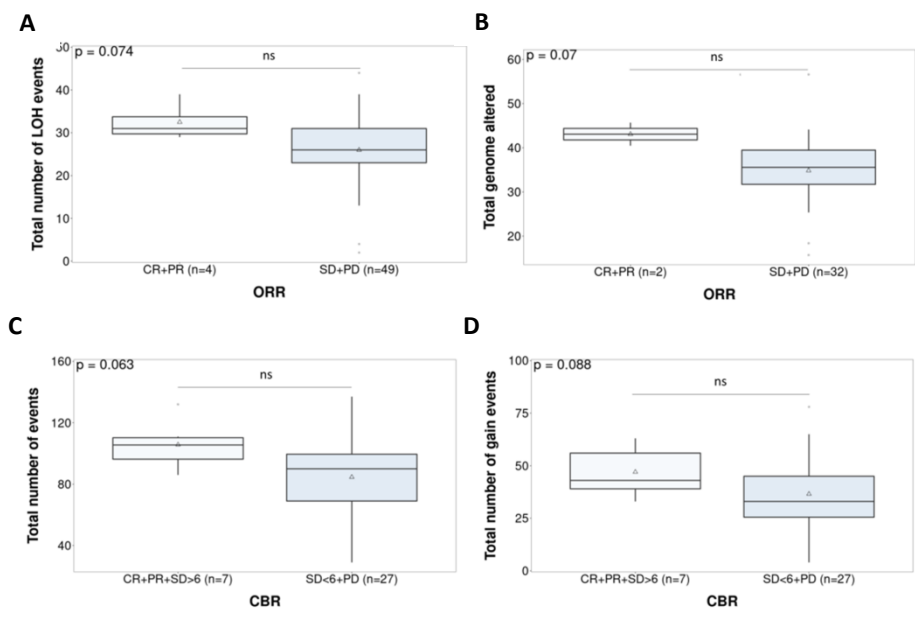


Figure 43: Clinical implication of GI parameters regarding ORR. A) Global population and B) EC population CBR C) Total number of events and D) total number of gains. Non-parametric Wilcoxon Signed Ranks Test was used.

We observed a significant correlation between ORR and the percentage of genome altered by losses ($p = 0.021$), although only two cases qualified as responders with HGS histology (Figure 43A). In the EC population, the percentage of the total genome that was altered was not associated with ORR ($p = 0.07$) (Figure 43B). Finally, the classification of responses as CBR was studied, but did not yield significant associations, for example, with the total number of events ($p = 0.063$) and gains ($p = 0.088$) (Figure 43C and D).

In the HGS population (n = 38), a higher number of events was significantly associated with longer PFS ($p = 0.036$). Although the GI parameters were correlated with the PARPi response in the non-parametric tests, only few parameters showed significance in the univariate survival analysis, and multivariate analysis was not significant. However, the results showed a correlation between higher GI and outcome, which raises the possibility of developing this parameter as a predictive marker. In addition, genomic scars, obtained as mentioned before, were tested in the population. Any of them showed statistical significance with response or HRR mutation.

4.1.4.3. *Assessing CN amplification and losses at gene level*

In addition to the GI profiling, custom panel was also designed to be able to interrogate for CN at gene level. Hence, CN data analyzed by panel mops package were used. A total of 6 and 1 amplification were detected in *CCNE1* and *EMSY* respectively, while 5 patients presented *PTEN* loss among the EC population. Concerning *CCNE1* amplifications, 4 were found in EC and 2 in OC, any of them coinciding with HRR mutation, being mutually exclusive. *PTEN* was also found in the two subpopulations, 2 events in EC and 3 in OC. These alterations were validated by MLPA, the gold standard technique, to assess CN at gene level. While amplification was confirmed in all cases, the validation of *PTEN* loss was not possible. Cases harboring these alterations have not shown correlation with

response to PARPi. Additionally, a sub analysis stratifying OC patients according to the presence of *CCNE1* amplification, HRR mutation or none of them was performed, also lacking significant results.

4.1.5. OC Cell line Panel

Due to the undeniable importance of this biomarker, genomic characterization at CNV and SNV level was performed in a panel of EOC cell lines, looking for better defined in vitro model systems with consistency to valid clinical phenotypes.

4.1.5.1. *Genomic characterization: Punctual mutation and Copy-number alterations*

A panel of 18 OC cell lines has been analyzed by NGS technology. The sequencing panel is specially designed to determine the mutational status of the HRR pathway, highly involved in this type of tumor, particularly regarding the sensitivity to PARPi, previously tested in this study. Additionally, other DNA-damage repair genes were interrogated. The most commonly mutated gene was *TP53*, present in 11 of the 18 (61.1%) evaluated cell lines, ten of them corresponding to HGSOC histology. Alterations in the HRR pathway were detected in 8 cell lines (44.4%): *BRCA2* 4/18 (22.2%), *BARD1* 2/18 (11.1%), *CHEK2* 1/18(%) and *ATR* 1/18 (5.6%). Among them, two correspond to reversions of a pathogenic mutation in *BRCA2* (PE04 and PEO6), recovering the functionality. Other genes were found to

be altered, between them: *MSH6* 4/18 (11.1%), *EMSY* 1/18 (5.6%) and *CCNE1* amplification in OVCAR3 (Figure 44).

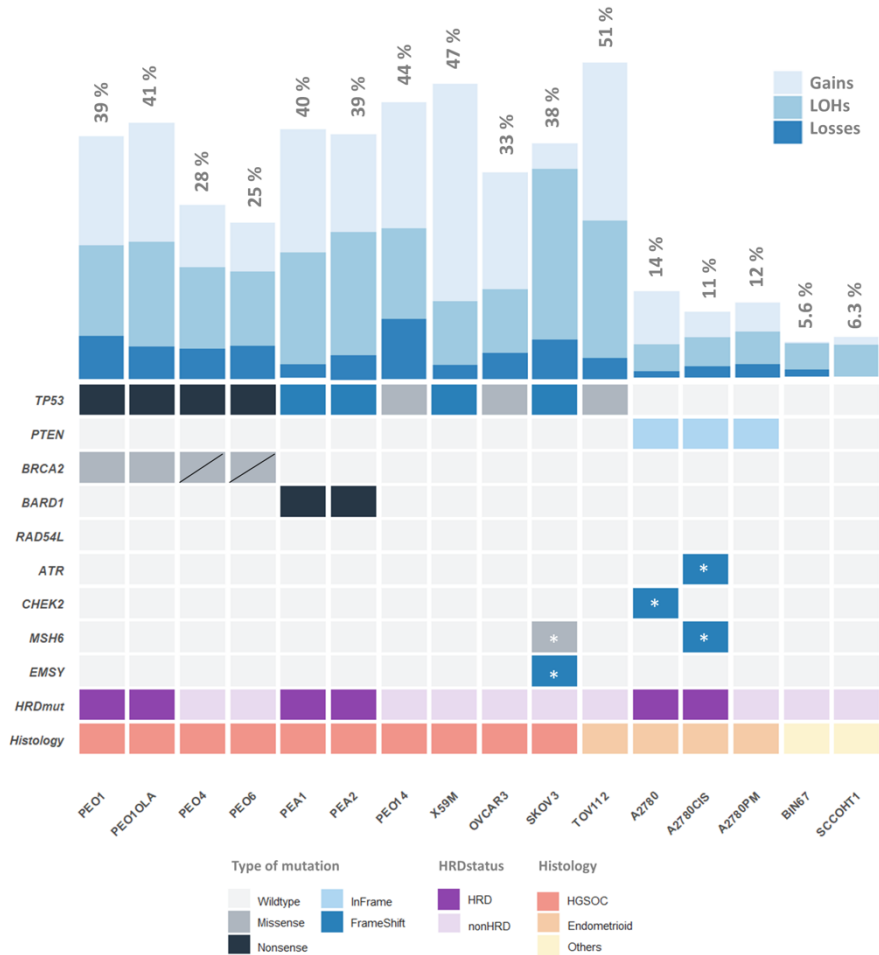


Figure 44: Molecular landscape of OC cell line panel. Distribution of genetic and genomic alterations, showing mutational patterns and incidence of GI through percentage of altered genome parameter. *Alterations not previously described in the bibliography.

Among all the described alterations, 5 of them have not been previously described in the bibliography. Two cell lines were deleted from the study due to possible cross-contamination, TO14 and PEO23.

The analysis of CNV patterns showed two main clusters of cell lines, grouping those with HGSOC and mixed histologies (C2) and endometrioid and small-cell (C1) (Figure 45).

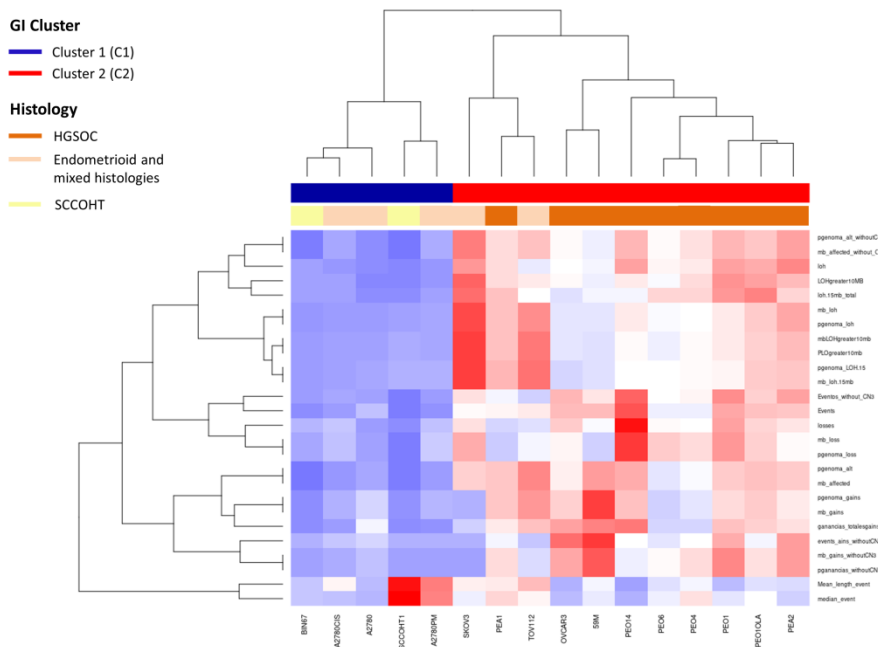


Figure 45: Representation of CNV patterns across OC cell line panel. Heatmap representing distribution and incidence of different GI parameters and its correlation with previously reported histologies.

C1 presented a highly-enriched CNV profile (percentage of genome altered between 25-50%) different from C1 (percentage of genome altered between 5-14%). The lowest GI values, with 6% and 5% of genome altered were obtained from SCCOCHT1 and BIN67 respectively, cell lines from small-cell carcinomas.

Additionally, the presence of GI in the HGSOC cell lines is particularly interesting. The case of PEO1 is a good example to explain correlation between GI and HRD. PEO1, carrier of *BRCA2* mutation presented high GI, while its related cell lines, clinical resistance and terminal stage of the disease from the same patient, with acquired reversion mutation of *BRCA2*, showed lower level of GI, possibly implying loss of sensitivity to certain DNA-damaging drugs. On the other hand, in the case of PEA1 and PEA2 cell lines (both from the same patient at initial and relapse stage) also carriers of HR mutation (*BARD1*) without reversion, maintain high levels of GI over time (Figure 46). However, these data must be carefully considered, since due to the GI inherent to the immortalization and maintenance process of a cell line, this phenomenon can be influenced by many factors.

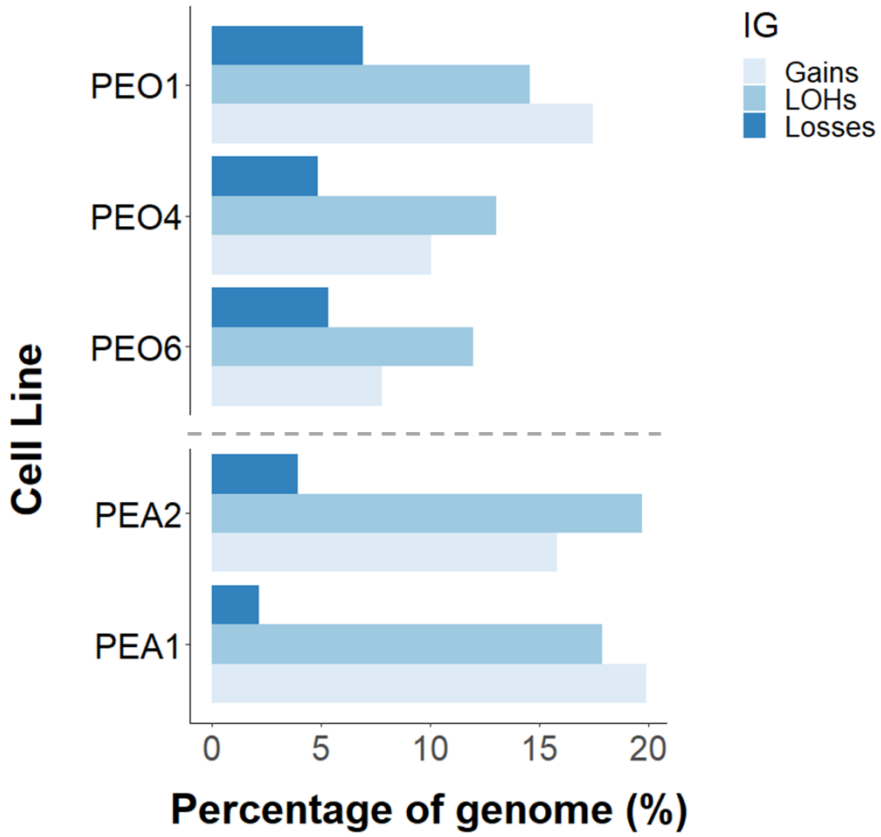


Figure 46: Exemplifying GI in HRR defective cell lines over time. Incidence and distribution of GI between two related cell lines and sub-lines (PEO1/PEO4/PEO6 and PEA2/PEA1).

4.2. 12g prognostic model in EC cohort

The importance of molecularly stratified patients regarding its prognostic or predictive value is evident. Particularly, in the field of gynaecological oncology, the GI has gained force as a classifying feature. Therefore, the development and validation of a reproducible RF model (12g-algorithm) based on mutational data to classify EC according to its GI profile, complementing the four prognostic-group classification, was performed.

4.2.1. Selection of the multigene-NGS panel and mutational analysis

The EC data set from TCGA⁶⁶ defines 48 genes with differential mutation frequencies across the four prognostic groups. A subset of 13 genes, corresponding to those with the highest differences in terms of frequencies between groups, was selected: *POLE*, *PTEN*, *TP53*, *ARID1A*, *KRAS*, *ARID5B*, *FBXW7*, *PPP2R1A*, *CTCF*, *CTNNB1*, *RPL22*, *PIK3CA* and *PIK3R1*. Two separate sequencing runs, containing 48 dual-pool libraries each were performed. The coverage, quality parameters and statistics were comparable between both runs, hence it was possible to merge the data for analysis. Sequencing metrics for analyzed samples are summarized in Annex 12.

A median of 40 genetic alterations per case (range: 13-171) were found (Annex 13). Variants were classified as mutated if they were

already reported in ClinVar or if appeared as predicted pathogenic, likely pathogenic or VUS by PolyPhen and SIFT predictors. Benign and likely benign variants were not considered for the analysis. The presence of mutation was treated as categorical dichotomous variable (presence/absence of mutation).



Figure 47: Molecular landscape of Cohort 3. Frequency of gene mutations in EC patient' series determined by NGS 13 genes panel. *Hotspot *POLE* (p.P286R and p.V411L) 5.2% (5.6% in TCGA population)

The most frequently affected genes in our series were *PTEN* (55.2%), followed by *ARID1A* (49.0%) and *ARID5B* (43.8%), whereas *KRAS*

mutations (9.4%) represent the lowest frequency (Figure 47). MSI was observed in 15 of 96 patients (15.6%): 14 of 15 with endometrioid histology (93.3%), and in just 1 of 13 serous cases (7.7%). The median number of mutations per patient that passed the annotation filter was 9.5 (range: 2-64).

Clinical parameters of stage, grade and histology were tested to gene-mutations by univariate analysis at gene level. The most relevant relationships were *POLE* mutation, which showed a correlation with early stage in EC ($p=0.040$), *PTEN* mutations, enriched in EC with endometrioid histology ($p < 0.001$) and low-grade tumors ($p < 0.001$). In addition, EC with serous histology harbored more *TP53* mutations ($p=0.021$). Finally, *RPL22* mutation showed higher frequency in endometrioid histology ($p=0.005$) and low-grade tumors ($p=0.004$). *KRAS* ($p=0.035$) and *CTCF* ($p=0.05$) mutations were also significantly correlated with low-grade tumors (Table 13A).

Table 13: Correlation between mutational status of analysed genes and clinicopathological and outcome parameters. A) Main clinical and pathological parameters in EC using Chi-square test B) PFS and OS measured by log-rank test

A)

		Histology			Stage			Grade			
		Endometrioid	Serous	p-value	Early-stage	Advanced-stage	p-value	I	II	III	p-value
	mutated	15	1	N.S.	16	0	0.040	10	5	1	N.S.
	non-mutated	68	12		63	17		35	23	22	
PTEN	mutated	53	0	<0.001	44	9	N.S.	30	20	3	<0.001
	non-mutated	30	13		35	8		15	8	20	
TP53	mutated	24	8	0.021	24	8	N.S.	12	8	12	0.050
	non-mutated	59	5		55	9		33	20	11	
KRAS	mutated	8	1	N.S.	9	0	N.S.	8	0	1	0.035
	non-mutated	75	12		70	17		37	28	22	

		Histology			Stage			Grade			
		Endometrioid	Serous	p-value	Early-stage	Advanced-stage	p-value	I	II	III	p-value
CTCF	mutated	25	1	N.S.	22	4	N.S.	16	7	3	0.050
	non-mutated	58	12		57	13		29	21	20	
RPL22	mutated	40	1	0.005	34	7	N.S.	17	19	5	0.004
	non-mutated	43	12		45	10		28	9	18	

B)

		DFS				OS			
		Univariate		Multivariate		Univariate		Multivariate	
Stage	Early	37.40 (2.067-91.2)	0.006	N.S.		42.57 (2.067-91.20)	0.004	N.S.	
	Advanced	15.37 (4.87-91.00)				34.47 (6.30-91.00)			
Grade	I	50,33 (2.07-91.02)	0.003	N.S.		52.83 (2.067-91.20)	<0.001	8.26 (62.50-1.10)	0.04
	II	30.28 (9.70-79.57)				33.10 (9.70-79.57)			
	III	26.38 (4.87-67.60)				32.50 (6.30-67.60)			
Histology	Endome-	43.63 (5.47-	<0.001	8.90 (29.90-	<0.001	44.23 (2.067-	<0.001	N.S.	

	trioid	37.40)		2.71)		91.20)		
	Serous	21.47 (2.067- 91.20)				29.53 (6.30- 38.6)		
TCGA groups	POLE	55.4 (24.27- 77.43)	0.004	N.S.		55.4 (24.27- 77.43)	0.03	N.S.
	MSI	38.33 (11.9- 74.93)				38.6 (11.9- 74.93)		
	CNL	34.43 (2.067- 91.00)				42.57 (2.067- 91.00)		
	CNH	27.7 (4.87- 91.2)				30.53 (6.30- 91.2)		

Regarding the prognostic value of individual gene mutations in our series, mutations in *POLE*, *PTEN*, *PIK3R1*, *ARID5B* and *PPP2R1A* are correlated with better patient outcome as seen in Table 13B. MSI, which was more frequent in early stages: 11/15 (73.3%) stages I-II vs. 4/14 (26.7%) stages III-IV (p=N.S), lacked prognostic value both for PFS and OS (Annex 14).

4.2.2. Impact of 12 genes RF model in the clinical stratification of the disease

RF predictive model for a dichotomous variable (CNL or CNH) was trained using the mutational profile of the 13 selected genes from 148 patients analyzed by the EC TCGA project. To train the model, genotyping of 12 genes was included as categorical dichotomous variables (the so called 12g-model) (Table 14 and Table 15). The *POLE* and MSI groups were directly defined by the presence of *POLE* mutations and MSI respectively.

Our series of 96 EC patients was stratified into the four TCGA prognostic groups based on the genotyping data of the 12-gene NGS panel, MSI status, grade, stage and histology: *POLE*, 16/96 (16.7%); MSI high, 12/96 (12.5%); CNH, 20/96 (20.8%); and CNL, 48/96 (50.0%). As mentioned above, CNH and CNL groups were classified with our RF adjusted model.

Table 14: Contribution of evaluated parameters to 12g-model. The importance of each parameter is measured as the mean decrease of the Gini index of the variables in the models.

Parameter	12g-model
<i>TP53</i>	12.4658
<i>PTEN</i>	6.094
<i>CTNNB1</i>	3.4884
<i>ARID1A</i>	1.8658
<i>PPPR1A</i>	1.5958
<i>CTCF</i>	1.1435
<i>PIK3CA</i>	0.5644
<i>KRAS</i>	0.3994
<i>FBXW7</i>	0.4852
<i>PIK3R1</i>	0.4506
<i>ARID5B</i>	0.2425
<i>RPL22</i>	0

Table 15: Performance parameters of the 12g model in the validation series.

	12g-model RFA
Accuracy (95% CI)	0.9753 (0.9136-0.997)
No Information Rate	0.6049
Kappa	0.9483
McNemar's test p-value	1
Sensitivity	0.9688
Specificity	0.9796
Positive Predictive Value	0.9688
Negative Predictive Value	0.9796
Prevalence	0.3951
Detection Rate	0.3827
Detection prevalence	0.3951
Balanced accuracy	0.9742

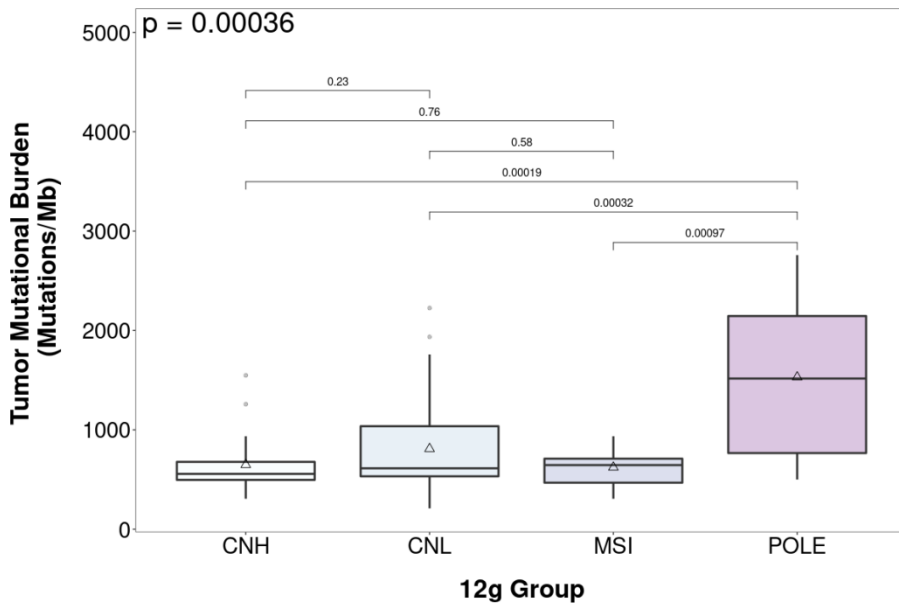


Figure 48: Surrogate of tumor mutational burden (TMB) across four EC prognostic subtypes evaluated by non-parametric test. CNH group shows the lowest mutational rate, whereas POLE mutational rate is the highest.

The POLE group was characterized by a *POLE* EDM mutation in all 16 cases and by the presence of MSI in 3 of the 16 cases (18.7%). This group presented the highest mutational ratio with a median of 94 variants/case (range: 31-171) compared with the other groups ($p < 0.001$). MSI group was characterized by the presence of MSI in 100% of the cases and had no *POLE* mutations. This group presented a lower median of alterations than *POLE* with 40 variants per case (range: 19-93). Among these alterations, the most affected genes were *PTEN* (75.0%), *ARID1A* (58.3%) and *RPL22* (83.3%). CNH presented a median of 32 variants per case (range: 19–96) and was

characterized by mutations in *TP53* (75%), low frequency of *PTEN* mutations (5%) and alterations in *PPP2R1A* (45%). Finally, CNL showed a median of 37 variants per case (range: 13–138) (Figure 48). Gene by gene analysis of these alterations revealed that: *PTEN* (60.4%) and *TP53* (14.6%) presented the highest and the lowest mutation rate respectively, with alterations in other genes as follows: *PIK3R1* (35.4%), *ARID5B* (41.7%), *CTCF* (31.3%) and *RPL22* (39.6%). The distribution of mutations across groups in EC dataset is depicted in Figure 49 (Annex 16).

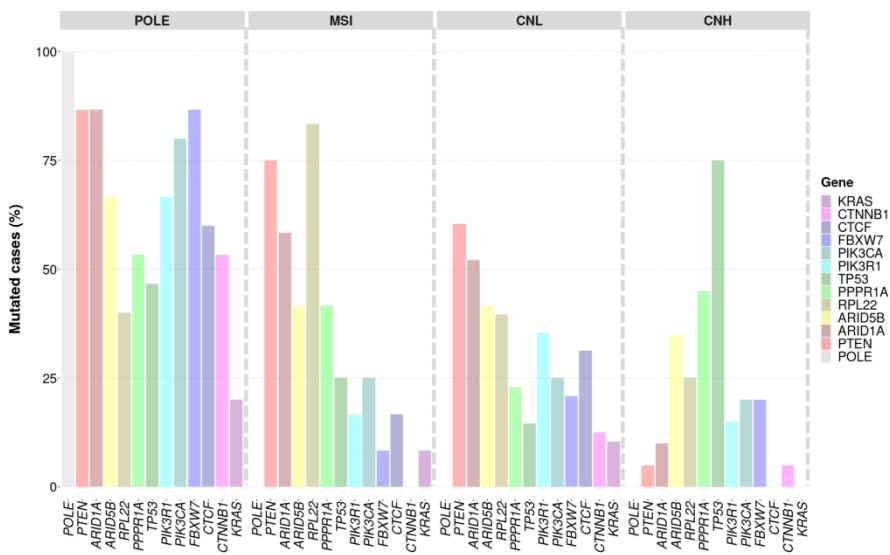


Figure 49: Distribution of genetic alteration across the four EC prognostic subtypes. The four-groups classification presents a particular profile of mutations, being a distinctive characteristic.

The Log-Rank test was used to evaluate the prognostic capacity of our molecular classification. This test confirmed that the molecular stratification of our patients revealed differences in both PFS ($p=0.0037$) and OS ($p=0.030$), suggesting that the POLE and CNH biotypes constituted the best and the worst prognostic groups respectively, mirroring the groups defined by the TCGA (Figure 50). In addition, a multivariate analysis was performed, being statistically significant only for histology (Table 13B).

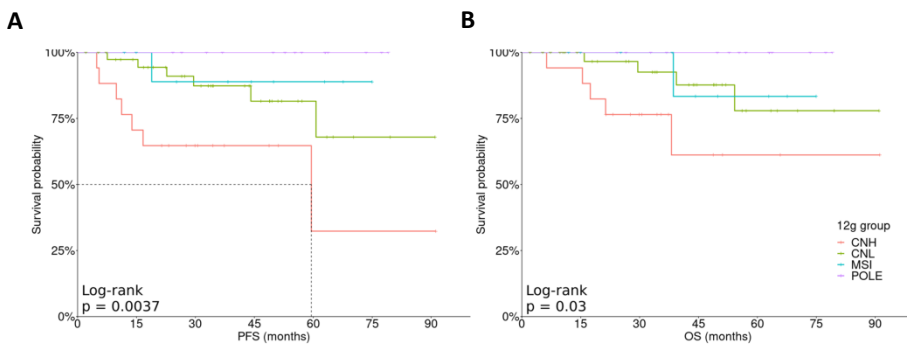


Figure 50: Kaplan-Meier plots assessed by log-rank test to evaluate the impact of the 12g model over outcome parameters. A) DFS and B) OS according to 12g stratification. Both parameters reach the statistical significance.

4.2.3. CNV-based classification: Validation of the 12g model

The 12g Model classify the CNL and CNH group based on the mutational profile. Due to the differential approach between the establishment of groups in the TCGA study and this study, a

validation step to confirm the CN profile concordance was performed in a 14-samples subpopulation. All parameters of GI and genomic scars were combined to obtain the CN profile of each sample (Figure 51). As seen in Figure 51, clustering of samples reproduces the CNH and CNL classification from 12g algorithm.

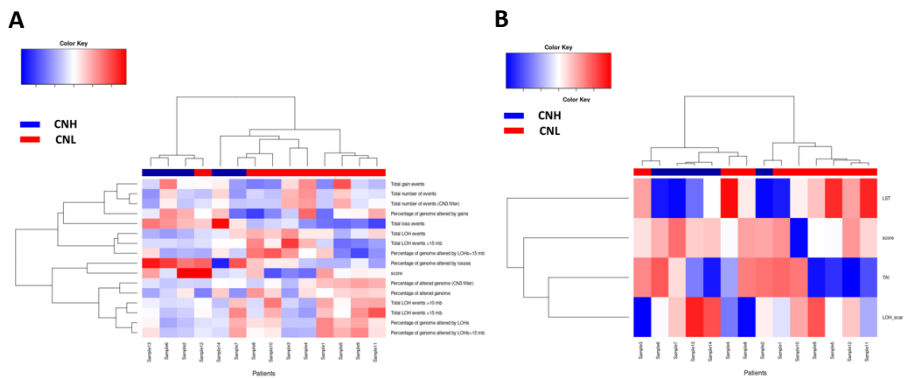


Figure 51: Unsupervised clustering of EC patients based on CN patterns. A) GI parameters derived from CNVkit pipeline and B) Genomic Scars from scarHRD package.

In addition to unsupervised clusterization based on CN data, the presence of each parameter per case was quantified. Genomic Scars, as shown in Figure 52, presented higher incidence in samples classified as CNH, specially enriched in LST events.

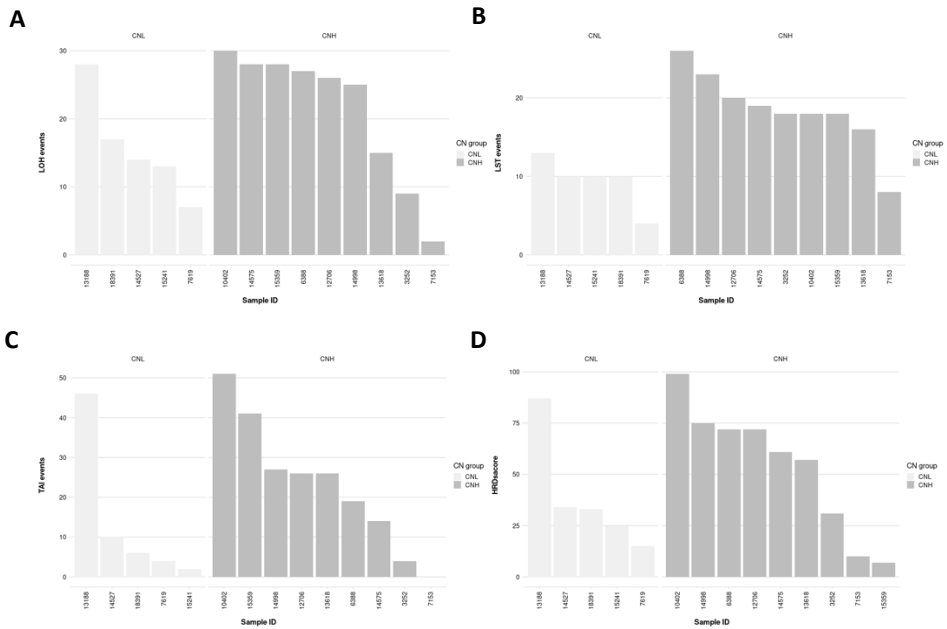


Figure 52: Distribution of genomic scars among the 12g subpopulation, stratified by 12g Model. Incidence of A) LOH events, B) LST events, C) TAI events and D) HRD score.

Regarding the distribution of GI parameters, obtain from CNVkit, between CN groups established by 12g algorithm, CNH group harbours higher presence of GI events, being specifically pronounced the distribution of LOH at event level ($p=0.0071$) and percentage of genome altered ($p=0.007$) represented in Figure 53.

However, even if 12g-Model classification seems to coincide with CN-based classification, GI parameters as well as Genomic scars, should be used as a whole to define GI pattern and not independently.

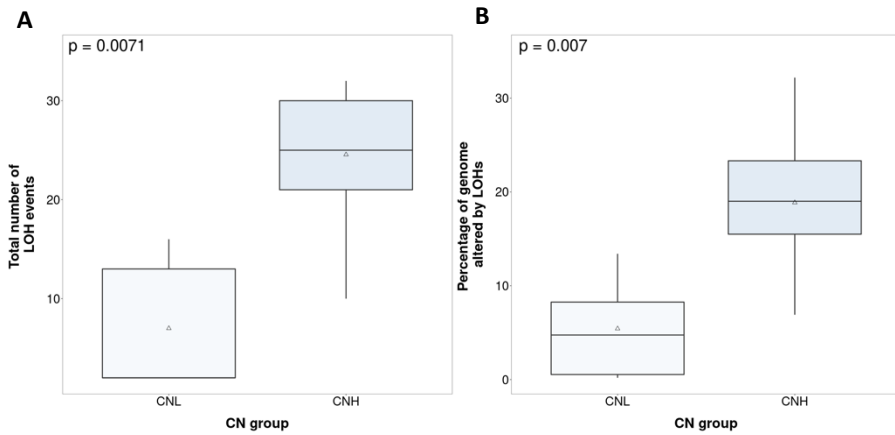


Figure 53: Distribution of GI parameters between CNV-based groups. Non-parametric tests evaluating A) LOH events and B) Percentage of genome altered by LOHs among the 12g subpopulation, stratified by 12g Model.

4.2.4. Expression analysis

Considering all the previous information generated from 12g cohort, a transcriptional approach was also performed, in order to complete the analysis. Due to the differential clinical implication of GI between OC and EC, the characterization of the CNH group was especially interesting, since could shed light to this difference. However, aiming to sub classify this group, CNH cases were faced to CNL. Unsupervised analysis correctly grouped samples according to 12g classification (Figure 54), showing a differential expression pattern.

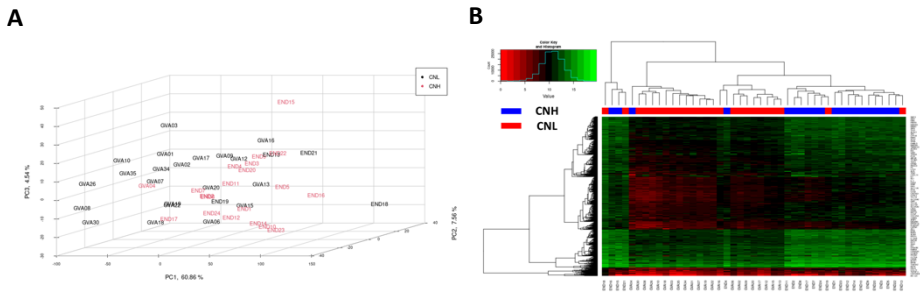


Figure 54: Unsupervised analysis of subpopulation of cohort 3 classified by 12g model. Results from A)PCA plot and B)Heatmap.

When DE analysis was performed between 12g-groups, samples seems to group in three different clusters; C1 mainly enriched in CNH cases, C2 enriched in CNL cases and C3 with mixed presence of cases. DE resulted in four main genes differentially expressed between groups (Figure 55B).

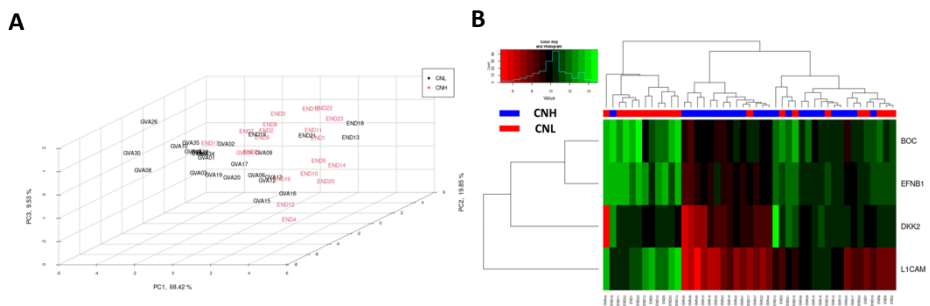


Figure 55: Clusterization of subpopulation of cohort 3 after DE analysis. Results from A)PCA plot and B)Heatmap.

The clinical implication of the 3-cluster classification was evaluated. Survival analysis facing Cluster 1 and Cluster 2 and 3 was significant in terms of PFS ($p=0.039$) and OS ($p=0.0016$), presenting the worst prognosis the C1 cluster enriched in CNH patients (Figure 56).

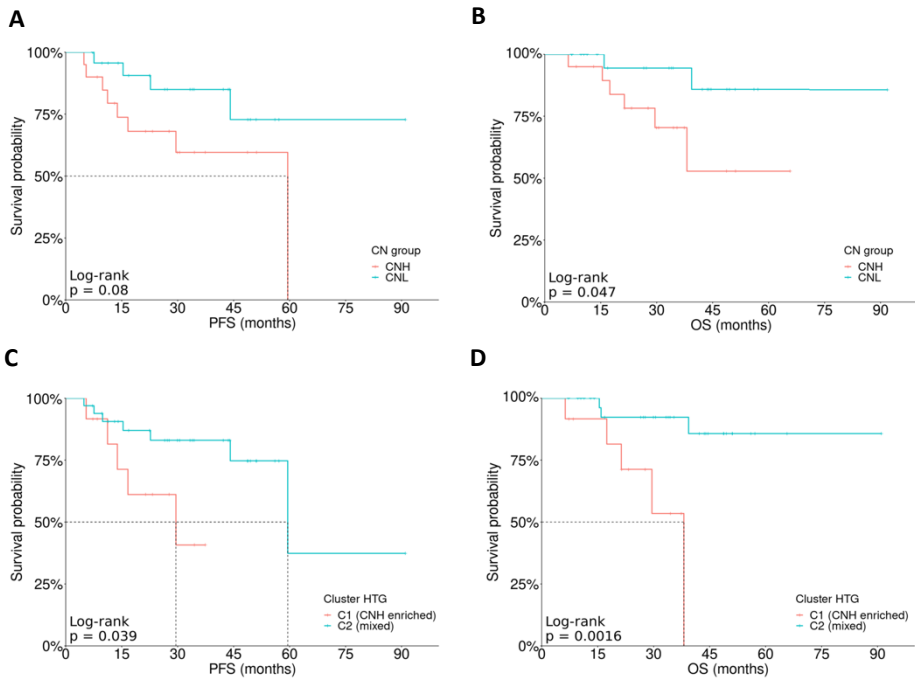


Figure 56: Log-rank test evaluating the clinical impact of 12g-based classification. Analysis regarding A) PFS and B) OS and DE-cluster based classification for C) PFS and D) OS.

In addition, DE analysis was also performed exclusively in the CNH population, classified by clusters. CNH samples were perfectly

clusterized when assessed DE. Five DE genes were obtained from this comparison (Figure 57)

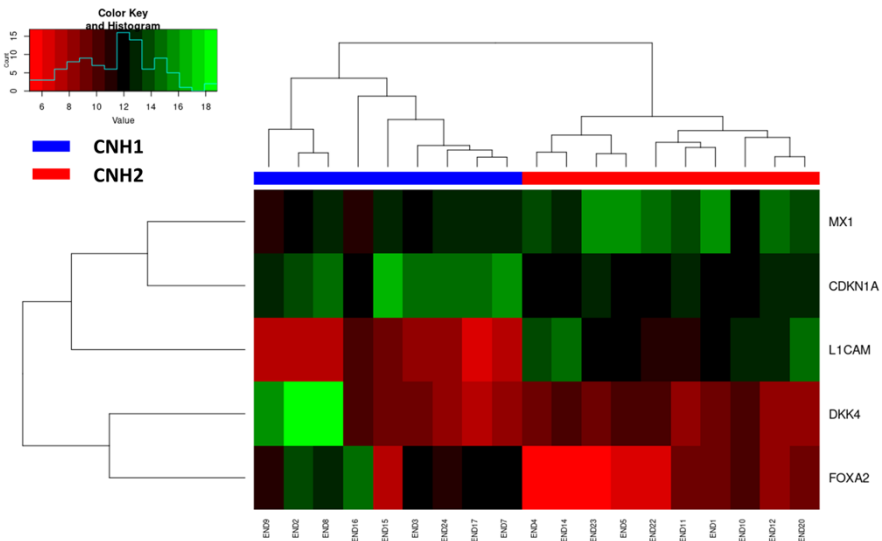


Figure 57: Clusterization of patients after DE analysis in CNH population. Results from A) PCA plot and B) Heatmap.

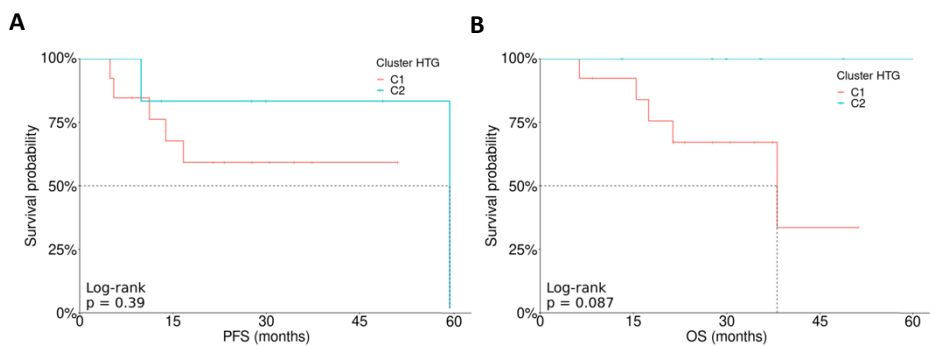


Figure 58: Log-rank test evaluating the clinical impact of cluster-based classification. Analysis regarding A) PFS and B) OS in CNH subpopulation.

This clusterization also maintains the prognostic impact of the classification in terms of OS (Figure 58).

4.3. In vitro characterization

In cancer research, cell culture systems represent the starting point of any study. For that reason, the existence of appropriate models that predict clinical efficacy and mimic the evolution of the disease is a need. Hence, characterizing and understanding the molecular biology of tumours allow the generation of suitable in vitro models.

4.3.1. Drug-sensitivity analysis of ovarian Cancer cell lines: IC50 calculation

Pharmacological characterization of the different cell lines was performed, obtaining corresponding dose-response curves for PM01183 (Lurbinectedin), Olaparib, Niraparib and Talazoparib in each cell line. Thus, IC50 values were established in the OC panel.

First, culture conditions were set up for each cell line. Regarding PM01183, exposure times of 1, 24 and 72 h were tested. While 1 h of treatment was discarded due to not reaching IC50 inside the evaluated range of concentrations, 24 and 72 h gave comparable results (Figure 59A). PARPi, evaluated at 72, 96 and 120 h after

treatment, showed a proportional decrease in IC50 value with increase of exposure time (Figure 59B). Based on these results, and considering that tested drugs will be administered also in combination, 72 h was chosen as optimal treatment condition for both type of drugs. In this way, cells were seeded. After 24 h, drugs were added with fresh medium and maintained during 72 h. Finally, cell survival was evaluated by MTT assay.

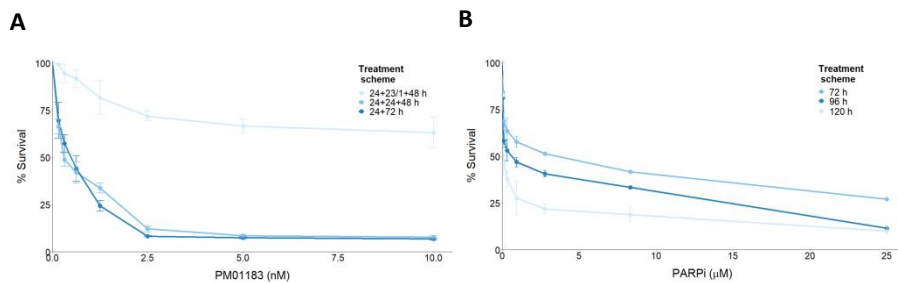


Figure 59: Setting-up range of drug-concentration and exposure time to perform cytotoxic assays. Tested conditions to establish best treatment scheme for A) Lurbinectedin and B) iPARP in monotherapy through three different time lapses.

Once established optimal time exposure and concentration ranges, cytotoxic assays were performed. All evaluated cell lines presented high sensitivity to PM01183, being the IC50 lower than 1 nM, except from OVCAR-3 (Figure 60). In the case of OVCAR-3, was not possible to apply the same treatment scheme.

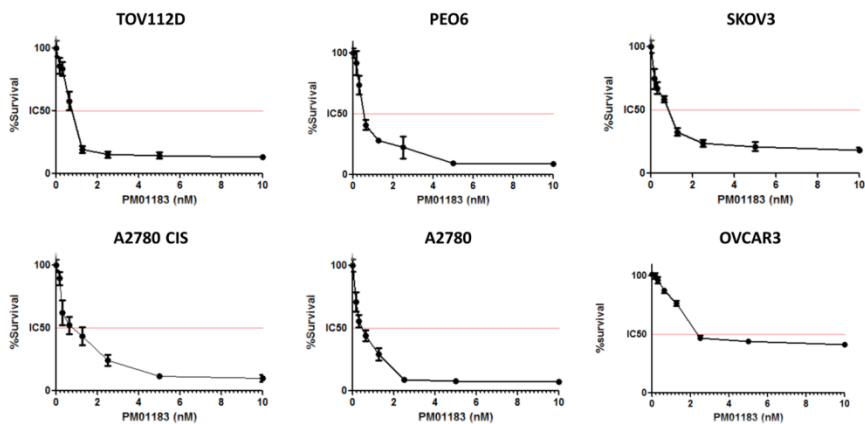


Figure 60: Cytotoxic assays of PM01183 in monotherapy. Drug-response curves of OC panel cell lines to PM01183.

Concerning the PARPi, even if cell lines were also sensitive, the active range of the drugs was in the μM scale (Figure 61). The most sensitive cell line to PM01183 (0.46 nM) and to different PARPi (ranged between 0.01-10 μM) was A2780, while SKOV3 presented the lowest sensitivity, excluding OVCA3. Talazoparib was the PARPi that presented the lowest IC50 values in comparison with Olaparib and Niraparib, being the most effective (Regarding combination treatment, as mentioned above, due to high sensitivity of cell lines to single drugs, the obtention of accurate results to prove synergistic effect was not feasible with this methodology).

Table 16).

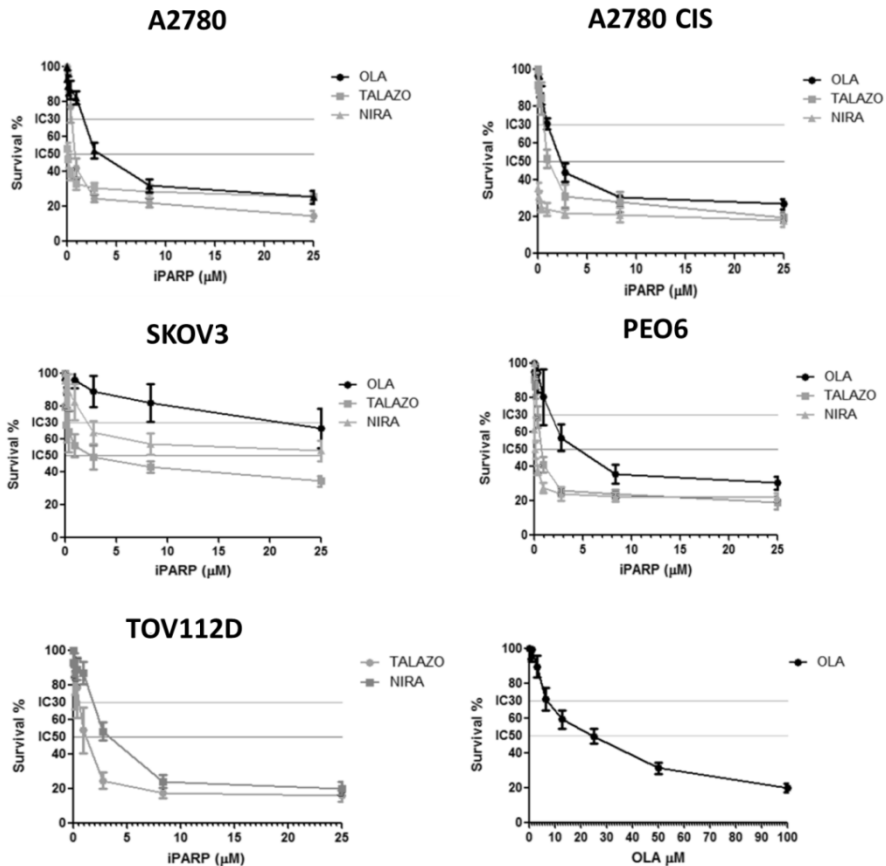


Figure 61: Cytotoxic assays of PARPi in monotherapy. Drug-response curves of OC panel cell lines to Olaparib, Niraparib and Talazoparib.

Regarding combination treatment, as mentioned above, due to high sensitivity of cell lines to single drugs, the obtention of accurate results to prove synergistic effect was not feasible with this methodology.

Table 16: Drug sensitivity data for tested OC cell lines and treatments. IC50 values for PM01183 and different PARPi in monotherapy.

Cell line	PM01183 (nM)	Olaparib (mM)	Niraparib (mM)	Talazparib (mM)
A2780	0.8764	3.9	0.98	0.06
A2780 CIS	0.4315	2.9	1.5	0.003
PEO6	0.6047	5.1	0.87	0.091
TOV112D	0.7191	21.7	3.5	1.16
SKOV3	0.7153	105.7	21.65	25.31
OVCAR-3	4.128	-	-	-

Only TOV112D, which presented lower single-drug sensitivity, showed a significant synergistic effect ($CI > 1$) for the combination (Figure 62).

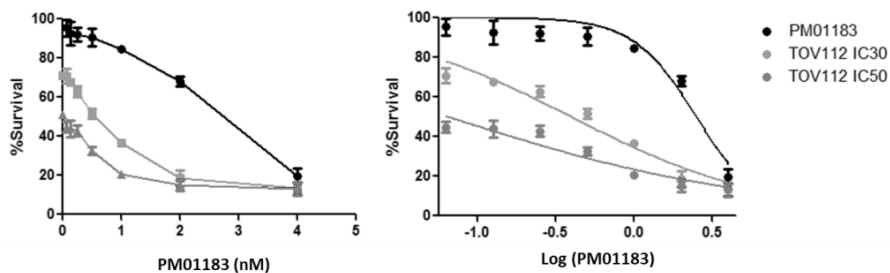


Figure 62: Cytotoxic assays of PARPi and PM01183 in combination. Dose-response curve for combination treatment of PM01183 and Olaparib in TOV112D cell line.

4.3.2. Generation of PM01183-Resistant cell line and functional characterization

As a final step, we decided to study the problem of resistance, complication that would lead to treatment failure and therefore, there is an urgent need of understanding its molecular mechanisms in order to overcome them.

Treatment of A2780 cells with one shot strategy, expose to 10 nM of PM01183, resulted in significant growth delay, slow recovery periods and extreme morphological changes, characterized by stellated morphology, long prolongations, and significant increase in cellular volume. These cells were unstable, not being possible to maintain them in culture. After the first pass, cells stopped growing and finally died.

Due to that, continuous exposure was the selected approach. Even though changes were less prominent, the possibility of culturing them, allowed the generation of the resistance over time. In this way, cells continuously treated with IC30 concentrations of PM01183 were maintained in culture until pass 8th. At this point, IC50 values were re-evaluated. The IC50 concentration of PM01183 resistant cells was 1.5 nM compared to 0.4 nM in the original parental cell line, a 3.75-fold increase (Figure 63). Cell line at the eighth passage of treatment was considered partially resistant (A2780-PM) and changes at functional and genomic level were evaluated.

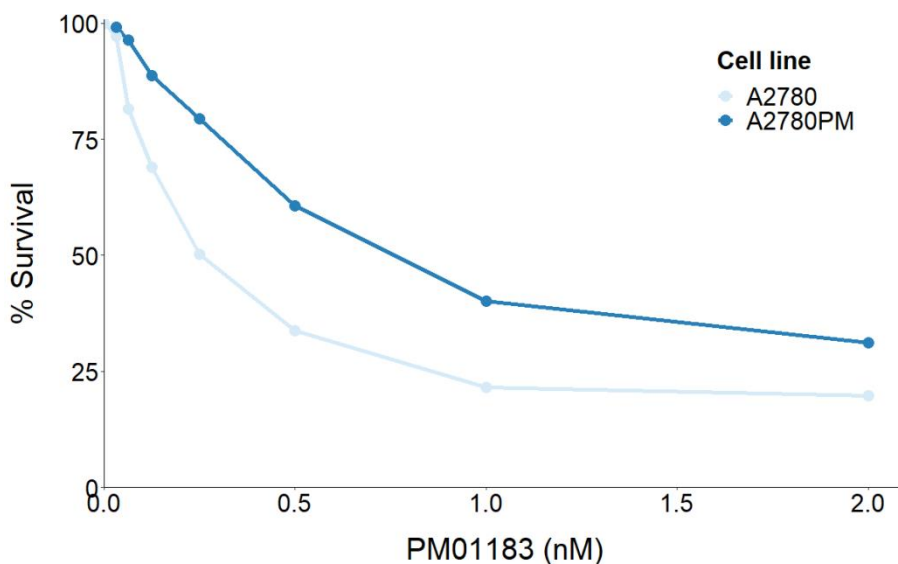


Figure 63: Quantification of changes in the IC50 values after the induction of PM01183 resistance. Modified IC50 after generation of PM01183 resistant cell line. Comparison between parental and partially resistant cell lines.

To evaluate alterations in the doubling time, cell growing ability was analysed over time. Different cell lines (A2780, 2780-cis and A2780-PM) were seeded at equal concentration and then cells were counted at five different time; 0, 24, 48, 72 and 96 h. While there were no significant differences between the A2780 cell line and the A2780-cis, A2780-PM exhibits a much lower growth rate (Figure 64), with a delay in cell cycle progression that was also confirmed by the analysis of cell cycle. Significant difference in the proliferation capacity/ability between parental cell line and their corresponding PM01183 resistant sub-line indicates the emergence of a resistant phenotype.

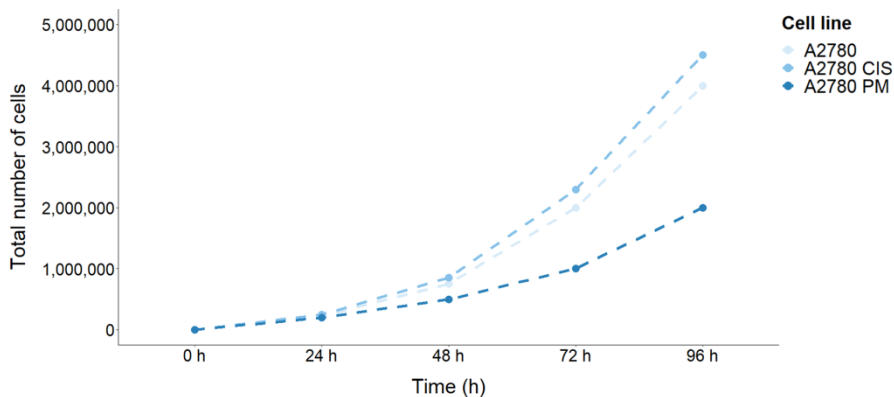


Figure 64: Doubling-time assay. Evaluation of growth capability changes induced by PM and CIS resistance in A2780, A2780-cis and A2780-PM.

Cell lines characterization also included cell cycle analysis in the parental and partially resistant cell lines. Observed variations, although subtle, showed a delay in the advancement of the cell cycle in partially resistant cell line to PM01183 compared to parental line. Treatment with PM00183, promoted an accumulation of cells in the G0/G1 phase (Figure 65). Treatment induced an accumulation of cells in the G0/G1 phase concomitantly with a decrease in the S phase of the cell cycle. This blockade in the advance of the cell cycle, consistent with the doubling time studies, may play an important role in the PM01183 resistant phenotype.

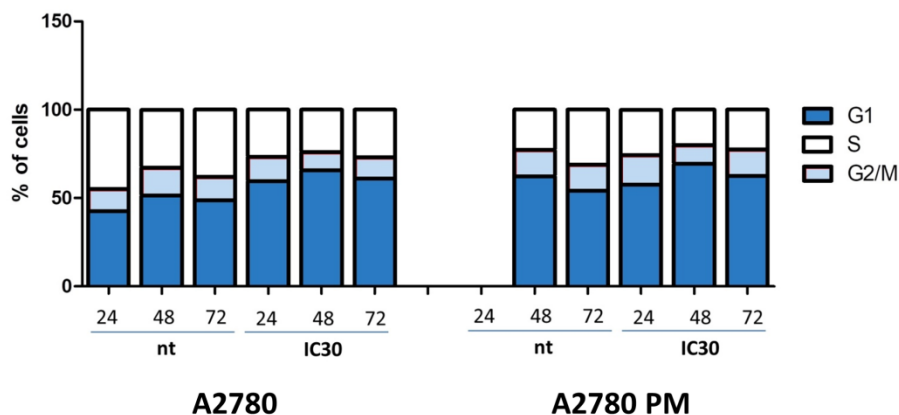


Figure 65: Cell cycle analysis. Quantification of fluctuations in cell cycle components before and after PM treatment in primary and resistant cell lines, A2780 and A2780PM.

To investigate DNA-DSB repair pathway status in cell lines and characterize the signalling pathway activation following drug treatment, expression of γ H2AX, RAD51 and PARP were measured by western blot in parental and resistant cell lines (Figure 66).

In general terms, results showed a trend to increase the expression of γ H2AX during treatment in parental and PM-resistant lines, with high basal levels in the last one, while CIS-resistant remains unaffected. Regarding RAD51 expression, this seems to decrease in parental and CIS-resistant, while PM-resistant maintains elevated but constant values (Figure 66). No significant decrease in the intact form of PARP was evident in any of the cell lines.

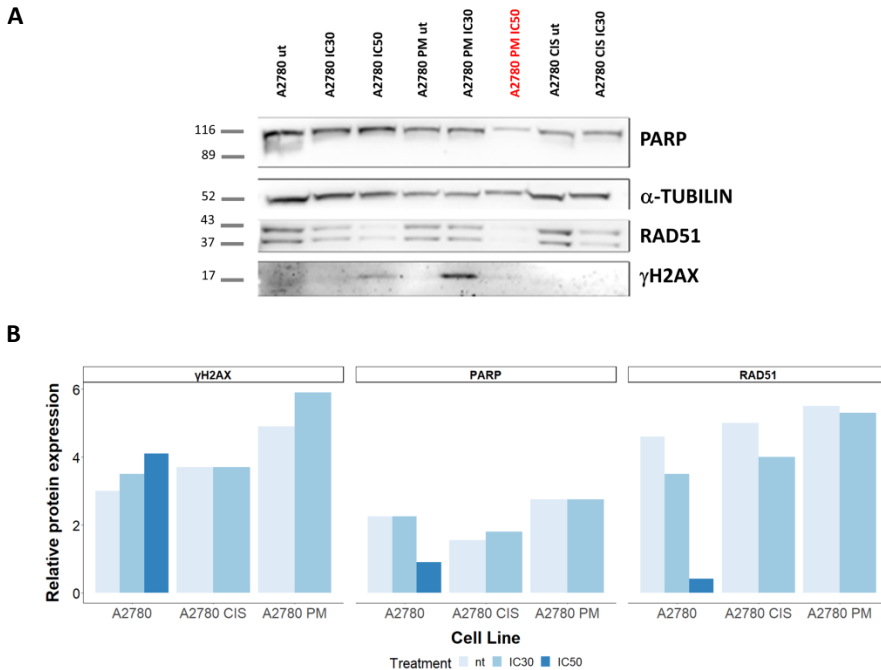


Figure 66: Analysis of HRR-related proteins by Western blot following drug treatment. Differential expression between parental and PM01183 and cisplatin resistant cell lines illustrated in A) Western blot gel and B) Protein normalized quantification from western blot raw data. Levels of γ H2AX, RAD51 and PARP were evaluated using α -tubulin as loading control.

Although it is necessary to extend these studies to obtain valid biological conclusions, the most relevant data are the high levels of phosphorylated histone observed in the A2780-PM following PM01183 treatment, without a change of RAD51 thus suggesting the ability of this cell line to sense and therefore repair DNA damage. The low levels of PARP observed in A2780-cis can possibly explain the

high sensitivity of this line to the different PARPi as well as to PM01183, although it is resistant to cisplatin.

DISCUSSION

5. Discussion

Over the past decades, there have been rapid advances in high-throughput technologies, which enable a range of genomic analyses to quantify the different layers of biological elements that contribute to the emergence and development of cancer. These variations in cellular machinery are driven by molecular aberration in several omics layers such as genome, epigenome, transcriptome, proteome, and metabolome within cancer cells¹⁹¹.

The advent of high-throughput technologies has permitted the growing of system biology and multi-omic approaches¹⁹². In this particular context, the use of integrative analysis is mandatory to depict cancer at multiple molecular levels and move forward towards the precision medicine paradigm¹⁹³

The integration and analysis of these multi-omics datasets is a crucial and critical step to gain actionable knowledge in a precision medicine framework. Integrated approaches allow comprehensive views of cancer phenomena. This integration provides a platform to connect the genomic or epigenomic alterations to transcriptome, proteome, and metabolome networks¹⁹¹. Focusing on the analysis of the generated data, bioinformatics has always had a central role and, hence, powerful and sophisticated computational tools have been developed to identify the interconnection between them¹⁹⁴. Thus, a comprehensive recognition of molecular networks based on multi-omics data has an important scientific role to understand the

molecular mechanisms of cancer.¹⁹⁵. The computational approaches play a central role in improving our current cancer management capabilities and are mainly based on ML techniques and other techniques such as graph analysis¹⁹⁶⁻¹⁹⁸.

The main goal of the precision oncology research driven by multi-omics data is to develop predictive/prognostic models that are refined by experimental validations. Hence, selecting patients based on personalized multi-omics data and stratifying them for prognosis or outcome^{194,199}.

In this context, several efforts have been made to generate comprehensive multi-omics profiles of cancer patients. TCGA, (<https://portal.gdc.cancer.gov/>), which provides detailed clinical, genomics, transcriptomics, and proteomics data for a variety of cancer types is a good example of this kind of approaches¹⁹⁸. These studies have been undeniably useful in the field of oncology, helping to decipher the molecular landscape of multiple tumors and working as a hypothesis-generator of a wide number of research works²⁰⁰. However, the deepness and hugeness of this type of projects is a limitation when translating them into the clinics. For that reason, there is a need to reproduce these classifiers in a simplified and more suitable approach, enabling their use in clinical routine.

Based on all the previously exposed data arises the hypothesis and objectives of this research, to develop an integrative prognostic and

response-predictive model relying on generated multi-omic data, facilitating the stratification and clinical guidance of gynecological patients.

a. OC approach

The picture of OC shifted after the publication of the OC-TCGA in 2011⁵². From now on, the research started to focus on the study of alterations in the DNA HRR pathway and its clinical implications, together with the development and approval of targeted drugs against that such as the PARPi²⁰¹. The TCGA project showed that defects in HRR genes were presented in almost 50% of cases²⁰². Hence, half of tumors will have to lay on more error-prone systems to repair DNA-damage, such as NHEJ, leading to an increased GI, also called “Genomic Scars”⁶⁵. These alterations serve as an indirect measure of HRD as they represent a permanent historical footprint of genomic changes induced by DNA repair deficiency, irrespective of the underlying etiology²⁰³. The main clinical impact arrives when the assessment of GI through omic-based algorithms showed association with better prognosis and response to platinum-based chemotherapy and PARPi, even improving the predictive power of the until-now gold standard, *BRCA* mutations^{203,204}. At this point, the key consequences of HRR defects in the development of tumours reached enough attention to focus research in that direction. However, interpreting the consequences of alterations in HR-related genes remains challenging since only some present the predictive

value, depending on its specific function within the repair process and the type of mutation involved¹⁶⁷. Low frequency of the individual mutations also complicates the assessment of their relevance²⁰⁵. In addition, different studies have shown that nearly the 20% of patient presented GI in absence of HRR mutation, also benefiting from Platinum-based chemotherapy and PARPi²⁰⁶. Therefore, limiting the evaluation of HRD phenotype to HRR-gene panel would restrict the number of patients that would benefit from these therapies. Due to that, the assessment of GI status as a surrogate of HRD started to be present in different clinical trials including PAOLA²⁰⁷, PRIMA²⁰⁸, and VELIA²⁰⁹, confirming its clinical relevance as a predictive and prognostic biomarker. Consequently, the identification of GI has risen as an essential tool in OC and become a challenge for the selection of patients that will benefit from these therapeutic approaches. However, even though GI-based (or equivalently HR-based) stratification would ideally be useful to understand individualized treatment options for patients. Otherwise, the lack of clinically validated academic tool together with the limited access to testing in many countries and associated socioeconomic challenges of the available tools limit its widespread use in clinical routine⁵¹.

- ***Technical and analytical methodology: CNV***

The lack of an academic tool in conjunction with the lack of a gold standard technology to decipher SCNAs patterns mainly respond to the complexity of this genomic feature²¹⁰. At technical level, there

are several available approaches to determine CNV, from Array-based technologies (aCGH and SNP array), affordable and relatively high-resolution assays²¹¹ to NGS-based solutions, such as WGS and whole-exome sequencing (WES). Sequence-based technologies have rapidly emerged as a viable option with higher resolution and accuracy, becoming primary strategies in this field²¹²⁻²¹⁷. However, there are still caveats that need to be overcome in every strategy. While aCGH and SNP array faces limitations regarding time-consumption, limited resolution and poorer sensitivity and precision, especially in FFPE samples, NGS protocols tend to require big technical and bioinformatics efforts, not only related to preparation but also related to computing power, storage, and costs^{218,219}.

With the aim of outperforming all these limitations, in this project a tuning of both methodological and analytical approaches to identify and characterize GI was achieved. This analysis was performed by using a NGS-derived method to determine SCNAs by combining the interrogation of 170080 SNPs along the genome plus a DNA HRR-genes custom panel. Thus, we are able to evaluate SCNAs at genomic level together with the interrogation of alteration in the HRR pathway without losing resolution, cost-effectiveness and avoiding the need of huge computational work.

Along with the methodological approach, the challenge was to develop a bioinformatic tool able to accurately detect SCNAs¹⁵⁹. At analytical/bioinformatics level, many algorithms have been

developed to facilitate CNV detection, but these algorithms are hindered by several factors²²⁰. However, despite improvements in CNV detection methods, the identification of CNV is still a challenging approach^{160,221}.

The first challenge was to identify pipelines that best suits our type of data and further analysis. In the absence of a standard analytical method for FFPE samples, we tested different available tools and selected three to be evaluated: *Surecall*, the commercial solution from Agilent; and two in-house pipelines based on *CNVkit*¹⁸⁴ and *saasCNV*¹⁸⁶ packages coded in python and R respectively. Reasons for this choice were in turn: feasibility to analyze FFPE samples using NGS strategy, adaptation to the computer resources available and possibility to use in targeted panels^{222,223}.

In general terms, these algorithms are constituted by four basic steps: mapping of reads and calculation of coverage; normalization; segmentation; and estimation of copy number^{160,224}. Depending on the nature of the pipeline, the different steps are more open to personalization. Due to the aforementioned complexity of the determination, being able to adapt them is a considerable advantage. Hence, adaptation and parameter tuning was performed for each tested approach. This process enable us to identify the best parameter setting to obtain reliable results in the sense of being

endowed with the most predictive information, maximizing the sensitivity required for a clinical context.

Due to the customizable nature of CNVkit pipeline, three different parameters were tested due to its relevance in CN calling. On the one hand, the addition of tumor burden, obtained by a specialized pathologist, as an input feature, was evaluated. Tumor samples are mostly contaminated by non-tumor cells. As a result, read count values do not completely reflect copy number of tumor cells. This introduces difficulties in calling copy number segments²¹⁷. Values of 50, 80 and 100 % of tumor content for the same sample were applied during the analysis (Figure 26). Results showed significant differences when applying different tumoral content in total number of CNV events ($p=0.00046$), as well as tendency in percentage of tumor altered and CN length. On the other hand, applied segmentation pipeline was also shown to be decisive when calling CNAs (Figure 24). The selected segmentation pipeline, due to higher reliability of events of greater length and the suitability to perform the following analysis to call LOH events, was CBS. Lastly, different p-values were tested to select and filter CN events (Figure 25). The proven fact was that using a restrictive filter does not mean losing CNV information while obtaining more reliable results. Thus p-value of 0.001 was chosen for the subsequent analysis.

Whereas *CNVkit* offered a wide range of adjustable parameters, *saasCNV* and *SureCall* were quite limited in this regard, applying

default parameters and pre-filtering steps respectively. Additionally, selection and use of controls were diverse between pipelines. While *saasCNV* and *SureCall* approaches only allow the addition of one control sample, when running *CNVkit*, a panel of controls was applied. In both cases, 10 FFPE corresponding to peritumoral samples of the ovary were used. However, these divergences will mark the difference in later analysis, mainly in the step of CN calling.

Following the adjustment, pipelines were benchmarked according to established GI parameters. Regarding calling of GI parameters between pipelines, significant differences appeared. These differences seemed to be related to the controls that had been used, normalization method, segmentation pipeline and default or personalized features applied on each pipeline (Figure 28). To evaluate the performance of these CN callers in terms of GI assessment, we used response to Platinum-based drugs as a surrogate of outcome. Hence, *saasCNV* presented the worst performance rate except for LOH calling, while *CNVkit* and *Surecall* were significantly higher (Figure 29). Similar results were obtained when comparing them to HRR gene mutations (Figure 30).

Taking all this information into account, when comparing performances of these tools, *CNVkit* appeared as the most suitable to our data due to its customizable fashion as well as its good correlation with clinical outcome, being selected for further analysis.

- **Mutational analysis**

Classically, the determination of HRD status has relied on *BRCA1* and *BRCA2* genotyping. However, other non-*BRCA* HRR genes have been also described to be implicated in the HRD phenotype, in a frequency between 4-9%, depending on the NGS panel used²²⁵. Additionally, preclinical data suggested that patients, whose tumors show HRD caused by other mutations in the HR pathway, may benefit from therapy with PARPi²²⁶. Due to its maintained clinical relevance, alterations in genes involved in the HRR pathways were also interrogated.

In our series, 59.1% of cases presented HRR mutations, representing *BRCA1* and *BRCA2* the 27.4% and 13.6% of cases respectively (Figure 22). The high frequency of the *BRCA* mutations in our series in comparison to other previously published data²²⁷ is due to the enrichment in *BRCA* mutated cases to optimize our GI status approach. We also identified mutations in other non-*BRCA* HRR genes, including *PALB2* 9.1 % (4/44), *RAD54L* 4.5 % (2/44), *BRIP1* 4.5 % (2/44), *ATM* 4.5 % (2/44), *ATR* 2.3 % (1/44) and *SLX4* 2.3 % (1/44) (Figure 22).

Whereas the detection of these mutations seems to be technically manageable, there are another caveats. First, several types of proteins are involved in protect genome integrity²⁰⁶. On the other hand, there are issues with the interpretation of the consequences of alterations in a clinical context such us variant selection, which includes *in silico* prediction of pathogenesis, might have an impact on

sensitivity prediction. Some of those mutations may not have a real loss of function and, hence, may not present the HRD phenotype. Besides, non-BRCA genes present low frequency of the individual mutations complicating the assessment of their relevance. This point still remains challenging¹⁶⁷. Additionally, it has been describe that nearly 20% of the study population was HRD positive without having a *BRCA* mutation²⁰⁶, drawing attention to the fact that GI not always overlap HRR-gene mutation and suggesting higher accuracy of the GI score over an HRR gene panel to define an HRD phenotype. Due to its implication in the clinical scenario, especially regarding response prediction, approaches at different levels such as genomic scars, identifying HRD without knowing the underlying mechanisms are gaining relevance^{205,228,229}.

Finally, to expose the clinical advantage that a more comprehensive approach means over the mutational analysis, as previously underlined based on bibliography, we compare the ability of GI parameter and HRR mutation to predict response. Here we showed that GI parameter presented a better correlation with response than mutation. These differences respond to the nature of the phenomenon that we are analysing. While the interrogation of gene-mutation brings us to the possible cause of GI, the evaluation of SCNAs bring us to consequence. Even if sometimes the specific cause of a genomic event is the data that we are looking for, in the case of HRD, the presence of GI will give us real information about the

current status and functionality of the pathway, directly impacting over clinical behaviour of the tumor²²⁸.

- ***Model fitting***

Despite some GI parameters individually present better performance than mutations in the HRR pathway, there is yet space for improvement. For that reason, the possibility of combining genomic and transcriptomic data was interesting to explore. Besides representation of a complex phenomenon like GI, involves several biological and regulatory mechanisms. However, the combination of different biological layers remains a challenging process. Hence, the need to integrate multi-omics data has led to the development of new algorithms and methods that are able to extract biologically significant information of clinical relevance²³⁰. The identification of GI as a surrogate of HRD has risen as a prognostic and predictive tool in HGSOC²⁰⁴. Even though HRR-based stratification based on any alteration or effect in the genome is widely recognized as essential, many efforts have been made to obtain and clinically validate academic tools based on different approaches²³¹⁻²³³.

ML plays a leading role in the design and development of these studies¹⁹¹. This discipline has been proved to be capable of solving many biomedical problems, providing insights into the molecular mechanisms, revealing interactions between different omics, and have been used in prognostic tools^{193,230,234}.

In this study, we established three single-source models based on SNPs, GI and RNA expression analysis and an integrative algorithm (Scarface model) built to predict response to DNA-damaging agents, particularly platinum-based chemotherapy, and PARPi.

Proposed models are built on the basis of three layers: SNP deep NGS sequencing, CNV profile using in-silico algorithms and targeted RNAseq using HTG EdgeSeq technology. Each layer has its strengths and limitations, but ultimately, each one underpins the others. This design responds to the different mechanisms in which HRD is produced, trying to mimic the complex biological context (genomic, transcriptomic, epigenetic, etc...). These different levels of biological information could be better represented by a multi-omic approach²³⁵.

Due to the large amounts of heterogeneous data generated by the different approaches, the removal of non-informative characteristics reveals to be a crucial step during the fitting process, simplifying the model and increasing its performance. Hence, feature selection selects the variables that contribute most to the prediction, removing the irrelevant or less important features that can negatively contribute to the performance of the model^{191,193}.

Feature selection was carried out by different strategies using ANOVA test, s2n ratio, significant parameters from logistic regression, recursive feature extraction²³⁶ and Boruta algorithm²³⁷.

Then, each set of selected parameters was tested coupled with a data mining algorithm. Several ML models (SVM, RF, NN, Decision Tree, and Naïve Bayes) were adjusted with different parameters and hyperparameters. Every possible combination between the algorithm and selected features was tested to ensure best performing. The Scarface model, constituted by the best single-source models, comprises 8 SNPs (Annex 10), 28 GI parameters and the expression of 7 genes (Annex 11), and achieves an accuracy of 0.9615 and a kappa of 0.9128.

However, even the great performance of the Scarface Score, single-source models could also work as a suitable and efficient tool in the real clinical setting, helping to guide the clinical management of patients simplifying the approach and decreasing costs (Table 12).

The lack of a gold standard for the definition and assessment of GI has motivated a wide number of studies to find an accurate approach. However, only two of them have been commercially approved: Myriad MyChoice[®] and FoundationOne CDx from Foundation Medicine^{®51}.

Those solutions, based on the identification of genomic scars, HRR gene mutations and LOH have proven their clinical utility in the context of clinical trial^{150,205}. Notwithstanding, they all succeeded in the prediction of *BRCA1/2* status, aim they were trained for. These tests, however, do not cover other molecular mechanisms such as

epigenetic or gene expression, which causes a loss of information concerning other HRD-causing mechanisms independent from *BRCA* gene status occurs^{205,238}.

Due to the lack of possibility of comparing the Scarface model with gold standard tests, such as those above mentioned, we compare classification based on HRR genes mutations (only *BRCA1/2* and including all HRR genes) and score from scarHRD pipeline¹⁸⁸, facing them to our integrative algorithm.

Stratifications based on *BRCA1/2* mutation (pPFI<0.0001, pPARPi= 0.0048) and whole HRR genes (pPFI < 0.0001, pPARPi = 0.0013) resulted significant regarding PFI and PFS to PARPi (Figure 23). In this particular case, the addition of other HRR genes together with *BRCA1/2* when classifying patients improved the statistical power, adding a prognostic and predictive value to the survival analysis. However, as recently reported, they do not always overlap GI, suggesting higher accuracy of the GI score over an HRR gene panel to define an HRD phenotype^{239,240}.

On the other side, scarHRD pipeline was applied to compare the performances of the classifiers. This pipeline has been trained to obtain the genomic scars evaluated by the validated commercial solutions, LOH, LST, TAI and HRDscore. However, the results were not as good as expected. Differences regarding methodological and analytical procedures caused a loss of statistical significance when

analyzing our series; this could be due to several causes. Firstly, in this approach, GI data was derived from NGS data covering a backbone and a medium-size panel whereas myChoice kit was validated and calibrated by using aCGH. Secondly, in this approach CNVkit method is used with the parameters specifically tuned in our clinical scenario, including pre-analytical factors such as Tumor burden in the sample. The best results were obtained when the series was stratified based on the median number of LOH events (pPFI = 0.0071, pPARPi = 0.07) and median HRDscore (pPFI = 0.031, pPARPi = 0.28), only being significant for PFI and not PFS to PARPi (Annex 17, Annex 18).

Nevertheless, the Scarface Model achieve the highest statistical significance for both platinum-based chemotherapy ($p < 2 \times 10^{-16}$) and PFS to PARPi ($p = 0.00077$), improving the predictive performance of until now-used classifiers (Figure 33).

As aforementioned, the predictive algorithm was trained and validated in an ambispective, multi-centric and real cohort of HGSOc using PFI as an endpoint. Thus, information regarding PFS to PARPi was not as accurate as expected due to this real-life design. PFS to PARPi data was collected for different lines, different treatment combinations and schemes and different PARPi drugs, not being a suitable parameter to use for training and validating the model. Due to that, and admitting the presented algorithm also showed predictive value of HRR mutations in this clinical scenario, the model

should be evaluated in a cohort with homogeneous and available PARPi response to validate its clinical benefit in this specific scenario.

With the Scarface model, we integrate GI parameters, equivalent to an HRD status that will more accurately differentiate patients according to PFI rather than the HRR mutations. In addition, information about gene expression is also provided, supporting the GI and contributing to the responder-phenotype processes. This approach has the advantage of studying the GI phenomenon as a whole at the genomic, chromosomal, and transcriptomic levels.

In conclusion, the Scarface Score could constitute a useful academic tool to predict response for DNA-damaging agents in the clinical scenario of HGSOC and, potentially, other HRR deficient tumours. This algorithm covers the needs not accomplished by available and validated commercial solutions, addressing GI and the molecular biology of the tumor from a more comprehensive point of view.

- **Other clinical scenarios**

Besides, due to the fact that this model addresses GI from different points of view, it seems a plausible strategy to calibrate the model to predict response with different cut-offs in other tumors, in which GI may play an important role in response to therapy, such as advanced prostate cancer with *BRCA* mutations, pancreatic cancer or TNBC.

Analogously, model could be adjusted to other similar drugs, such as PARPi⁵¹.

The translational study associated with the POLA clinical trial (NCT02684318, EudraCT 2015-001141-08, 03.10.2015) was designed to describe the correlation between different GI parameters and the benefit of the treatment, establishing GI as a predictive biomarker in this clinical scenario. In total, 57 cases were evaluated at the gene and genomic levels, defining those that presented HRD based on the mutational status of HRR genes (10 patients, 17.5%). *BRCA1/2* tumor mutations were present in 10% of cases, which is below the 20% of germline and somatic cases reported in HGSOC (Figure 39)²⁴¹. Regarding the correlation between HRD classifications based on mutational status and response, no significant association was found. The lack of predictive power of HRR gene mutations could be explained by differences regarding the characteristics of the population. Due to the nature of the population, with patients who have been previously treated with PARPis and, therefore, potentially mutated and responsive patients being excluded, the incidence of cases with *BRCA* mutations suffered an evident decrease. Clinical and methodological issues might also have an impact on the results. For instance, the fact that the genetic and genomic analysis was performed on the primary tumor and not at the moment of relapse, previously to study entry, could affect the concordance between HRD status and treatment response. The mutational/LOH patterns are not

reverted when a tumor recovers HR function, so they may not be accurate in predicting PARPi sensitivity in patients who have previously received and progressed on DNA damaging chemotherapy, such as platinum.

A significant correlation between response and different GI parameters, such as loss events, mainly present in the OC cohort and HRD population was found. Besides, a correlation between different GI parameters and better response to the studied combination was established. The results concerning the OC population were particularly interesting, where a higher percentage of losses ($p = 0.021$) appeared to be correlated with ORR (Figure 42A). At the same time, without reaching statistical significance, a trend was observed between the number of LOH events ($p = 0.055$) with LTRs (Figure 42B). In addition, the total number of events was also significant in the log-rank test (Figure 42C). However, all these results should be considered as preliminar because of the limited sample size.

Despite the promising results of GI parameters and its ability to predict response, the adjustment and application of the Scarface Model as validation was not possible due to the low number of cases, heterogeneous characteristics in the population and limited clinical data. However, it will be desirable to get well characterized and homogeneous populations to obtain a predictive model in the PARPi

context combining these pre-defined parameters and using response as the endpoint.

On the other hand, due to the undeniable importance of this GI phenomenon as a biomarker, in this study, we have performed genomic characterization at CNV and SNV level of a panel of EOC cell lines, representing different histotypes. Somatic mutations, taking into account only evaluated panel of genes, highly correlate with those previously published¹⁷¹, except from SK-OV-3 and A2780 related cell lines. SK-OV-3, presented alterations in MSH6 and EMSY while A2780 lack ATM and BRCA mutations but acquired alterations in MSH6, CHEK2 and ATR (Figure 44). Regarding CNV profiles, the evaluated cell lines showed a wide range of GI, from 5 to 50 % of altered genome, demonstrating the high dynamic range of this parameter, conferring a potential discriminant power (Figure 44). As expected, the most unstable genomes corresponded to HGSOC subtype. This characterization presents huge clinical relevance due to the indication of PARPi treatment beyond BRCA genes mutation and help to stratify tumors²⁴². Small cell and endometrioid histologies showed lower values. Interestingly, the mixed histologies (TOV112D and SKOV3), classically used as HGSOC models, were also grouped in C2, providing insights to a better knowledge and classification of this kind of tumors (Figure 45). However, we should be careful when evaluating GI and somatic mutations in cell lines. Their unlimited proliferation capability together with the passages previously

performed could imprint this type of profiles, being responsible of differences with bibliography²⁴³.

The characterization of an OC cell line panel at genetic and genomic level, interrogating HRR pathway status and GI profiles, completes the molecular landscape and establishes the basis and breeding ground of future preclinical and clinical studies.

b. EC approach

Analogously to the case of OC, the evolution of the molecular landscape of EC was marked by the data generated in the EC-TCGA⁶⁶. The stratification of EC patients has changed among the years, from the first subtype classification in 1983, which distinguished between Type I and Type II based on clinical and hormonal features⁷⁵, until before-TCGA classification, predicated on tumor morphology and tumor grade (based on glandular architecture and nuclear grade), in addition to the genetic profiling of different histologic subtypes, that has led to the identification of early driver mutations²⁴⁴.

However, techniques to determine histotype and grade in EC still have some caveat that need to be solved²⁴⁵⁻²⁴⁸. One of the main problems in the management of EC patients is inter-observer agreement when assigning histology and tumor grade by microscopic techniques^{246,249-252}. In addition, the overlap between histologic

subtypes and grade determination, as the difficulties differencing grade 3 endometrioid carcinomas and serous carcinomas or the poor correlation of histology and grade between diagnostic and final tissue samples complicates final diagnosis²⁵³⁻²⁵⁵.

Stratifications based on these data are associated with different natural histories, treatment scheme and patient outcomes, all of which will influence clinical decision making. Thus, accurate pathological assessment of histology and grade is essential in prognosis assessment and patient management²⁵². However, this scenario is frequently idealistic.

Those limitations present in our current system of EC classification and risk stratification have promoted researches aiming to identify new biologically informative tools²⁵⁶⁻²⁶⁰.

The incorporation of molecular classification into the standard histologic classification would more precisely define the subtypes of EC and guide development and use of molecular diagnostic techniques and targeted therapies, resulting in improved diagnostic accuracy and clinical outcomes²⁶¹.

In the post-genomic era, multi-omic information is redefining tumor classification. In this context, the arrival of the TCGA project completely changed current classification of EC and now constitutes an excellent source of data to mine prognostic models⁶⁶.

In 2013, the TCGA reported a large scale, comprehensive, and integrated genomic analysis of 373 endometrial carcinomas. The analyses identified four categories of endometrial carcinomas with distinct clinical, pathologic, and molecular features. However, even the information generated is extremely valuable and acts as an hypothesis generator, the comprehensive TCGA based approach is unaffordable in clinical scenarios, which require methodologies to be feasible, reproducible, and cost effective¹³⁰.

Aiming to overcome those limitations, we designed a small NGS gene panel with data from the EC TCGA dataset consisting of 13 of the most discriminant genes. Those selected were the genes with highest absolute and differential mutational frequency among the TCGA groups. Our objective was to develop a method based on the genotyping of only 12 genes with the definition and implementation of a reproducible model to classify EC into the four prognostic groups (POLE, MSI, CNL and CNH).

The most challenging task was to define a surrogate to classify CNH and CNL groups, which currently requires sophisticated technology as well as technical and analytical training. To achieve this, a RF model (12g-algorithm) was built by using the EC TCGA dataset (Table 14). This 12 g model accurately defines CNH and CNL groups (97%) and considers the contribution of each gene to discriminate between groups (Table 15). Finally, we validated the model with our prospective and independent EC patients series resulting in a total of

20 cases classified as CNH (21%) and 48 as CNL (50%). These frequencies were similar to those reported by the TCGA (26% and 39% respectively)⁶⁶. Due to the different approach between the establishment of groups in the TCGA study, based on CNV determination and this study, based on mutational approach, a validation step analysing CNV patterns was performed. Clustering of samples correctly reproduce the CNH and CNL classification from 12g algorithm as shown in Figure 51.

As a whole, the classificatory algorithm showed a good correlation with the TCGA groups and matched its prognostic value, with differences in both PFS ($p=0.0037$) and OS ($p=0.030$) (Figure 50). In addition, 12g prognostic model classified the patients independently of IHC, thus avoiding the intrinsic inter-observer subjectivity.

However, even though clinical impact was significant and genomic and genetic features were comparable to those of the TCGA groups, some discrepancies appeared in the incidence and distribution of alterations between them. Firstly, it should be highlighted that our cohort was significantly high in *POLE* mutations (16.7%) compared with the TCGA dataset (7%)⁶⁶. This mismatch is presumably due to different NGS technical approaches implemented in the two projects. WES, used in the analysis of EC TCGA samples, achieved a lower coverage (20X) than the targeted panel which was implemented in our work (600X). Therefore, mutations with lower variant AF could not be detected in WES. However, the percentage of mutations

found in POLE hotspots (p.P286R and p.V411L) defined by the EC TCGA was concordant between both datasets; 5.3% (our series) and 5.2% (TCGA series). In addition, filtering of alteration regarding annotation process could also impact in the number of variants finally selected and, hence, causing differences on incidence. This is particularly challenging in cases with rare variants, of which the effect on proofreading function is unclear.

Additionally, the incidence of MSI was considerably different. While, TCGA series presented around 28-20% of instable cases, our cohort only identify 16.6 %. Even if differences would be expected when distinct methodologies are applied, in this case, both fragment analysis and IHC were performed, obtaining a 96% of concordance between them, reinforcing the main role of the size of the analysed series in frequency of MSI unbalance.

Due to the relevance of the field, several research groups have also made great efforts to overcome these limitations and to identify the best combination of molecular markers to be used. With this aim, different approaches have been applied to reproduce prognostic classification simplifying the methodology.

Two main studies, Leiden/Trans PORTEC and Vancouver/ProMisE, have evaluated approaches to generate practical and feasible molecular classification systems. Although the subgroups of these two studies are not identical to the TCGA study, they integrate similar

molecular alterations^{127-129,262}. Similarly, the ProMisE/Vancouver group provided a molecular classification based on IHC of p53 as a surrogate of the CNH/CNL TCGA groups^{127,130}. However, the aforementioned inter-observer variability implicit in IHC makes standardization difficult between labs. This is underlined by the discrete (70%) concordance found in our global series, which is significantly improved in the CNH subgroup (84%), between the determination of *TP53* mutational status using IHC and NGS approach.

Since the known relevance of clinicopathological parameters, besides sequencing and adjusting the 12 g-RF model, we trained another model including these features (histology, grade and stage) to study the influence of these parameters, the so-called CPP model. Although there was a slight improvement in the performance parameters of the RFA (Annex 19, Annex 20), it is important to consider that our series came from a monographic oncology hospital. Additionally, pathological assessment was performed by a single gynecological pathologist highly trained in the diagnosis of EC, possibly masking the subjective effect, which should be considered with this kind of determination.

Taking all in account, the proposed approach overcomes subjectivity and technical difficulties related to the definition of CNH and CNL groups. The assessment of the mutational status by NGS technology

constitutes a highly objective methodology, drastically simplifying TCGA approach. Furthermore, the common availability of NGS and trained staff in clinical labs will facilitate the implementation of the proposed workflow in the diagnostic routine.

c. Complementary transcriptomic assays

Complementary transcriptomic analysis was mainly interesting to assess differences and similarities between CNH and HRD/high GI

subgroups from OC and EC, due to their differential clinical implication regarding survival. The three different clusters obtained from DE analysis give some clues about the distinct causing events of GI (Figure 51). The fact that C1 mainly enriched in CNH cases, and presenting worse prognosis while C3 with mixed presence of cases and better prognosis could be stratifying CNH patients with different prognosis, suggesting a differential pathways implicated in the high GI pattern (Figure 58). Interestingly, a strong similarity between CNH and HGSOC has been suggested. Consequently, it is likely that a subgroup of CNH-EC presented HRR deficiency¹¹⁹, similar to HGSOC, probably corresponding to CNH from the C3 cluster, with better prognosis. These results could provide a rationale for the use of PARPi in this subset of EC patients.

d. In vitro analysis

The improved understanding of the molecular landscape of OC has also drawn attention to the unmet need of developing more accurate in vitro models that reflect the clinical behavior and genomic phenotype^{263,264}. Hence, well characterized models will be useful to correctly guide patient and treatment selection, leading to higher response rates¹⁷². As aforementioned, HRD stratification appeared to be the most relevant molecular discovery in OC⁵¹. However, multiple studies revealed that alteration in this pathway not only respond to mutation in HRR-related genes but also by genomic scars or GI. Widen the concept of HRD by including GI allowed to capture clinically relevant information, and especially when obtaining correlation with drug response²⁶⁵.

Thus, in addition to genomic profiling, response to Lurbinectedin and different PARPi, treatment scheme from POLA clinical trial, was evaluated. Analysis of dose-response curves was focused on 6 of the total panel of OC cell lines, all showing high sensitivity to PM01183, with IC50s lower than 1 nM, except from OVCAR-3 (Regarding combination treatment, as mentioned above, due to high sensitivity of cell lines to single drugs, the obtention of accurate results to prove synergistic effect was not feasible with this methodology.

Table 16, Figure 60). In the case of OVCAR-3 was not possible to apply the same treatment scheme due to longer doubling times, not being comparable. Different range should be applied in order to achieve accurate data. Regarding, response to PARPi, Talazoparib was the PARPi that presented the lowest IC50 values (Table 16 and Figure 61), as previously reported²⁶⁶.

The obtention of reliable results from the combination was not possible following the established methodology and treatment scheme due to the extremely high sensitivity of cell lines to PM01183. This limited the correct determination of combination index and synergies. Although it is beyond the scope of this study a colony assay could shed light in this aspect²⁶⁷. Additionally, it will be interesting to discard possible toxicities of the drug beyond tumor cells.

As a final step, we decided to front the problem of resistance, complication that would lead to treatment failure and therefore, there is an urgent need of understanding its molecular mechanisms and how to overcome them. We generated a PM01183 partially-resistant model and characterized it in terms of proliferative potential, cell cycle distribution and HR-functionality. Two strategies for resistance induction were tested: One-shot and chronic exposure. In spite of one-shot strategy was more clinically relevant, its unstable resistance makes it impossible to work with²⁶⁸. Finally, PM01183 resistant cell line was established over time through chronic in vitro

exposure to IC30 concentrations of the drug. Differences were observed between parent and resistant cells when evaluating proliferative and cell cycle assays (Figure 64, Figure 65), as well as IC50 values (Figure 63), evidencing the starting of a resistance phenotype. However, this value did not achieve significant increase in order to consider cell line as resistant.

Therefore this study presents considerable research value to those looking for better defined model systems with consistency to valid clinical phenotypes, completing the molecular landscape and establishing the basis and breeding ground of future preclinical and clinical studies. However, it is needed to wide and complete data regarding chemo sensitivity as well as proving significance correlation between genomic studies and in vitro assays, aiming to work as a basis to new clinical trials. Additionally, generation of PM partially-resistant cell line will understand and overcome this phenomenon, being able to anticipate treatment failures.

CONCLUSIONS

6. Conclusions

1. A Laboratory Developed Test (LDT) to obtain genomic instability status, based on combined NGS data from a HRR-genes custom panel and SNPs sequencing, has been developed and validated in a retrospective High-Grade serous Ovarian Cancer cohort.
2. An independent validation of the developed LDT demonstrated a significant correlation with different clinically-validated parameters of response to PARP inhibitors in the ovarian cancer cohort of the POLA clinical trial.
3. Three independent models and an integrative ensemble one (The Scarface Score) have been validated to predict response to platinum-based chemotherapy in patients diagnosed with High-Grade Serous Ovarian Cancer.
4. All models showed better performance than currently used gold standard biomarkers, showing clinical impact at Platinum-response level and other clinical outcomes like PFS to PARPi and overall survival.
5. The *12g Model* stratifies endometrial cancer patients according to genomic instability profiles. The model combined with *POLE* mutations and Microsatellite Instability data

reproduces endometrial cancer TCGA prognostic classification with 98% of accuracy.

6. The *12g model* define group of patients with different genomic instability patterns in endometrial cancer which concordance was verified.
7. Genomic landscapes and cytotoxicity profiles of an ovarian cancer cell line panel have been assessed. This data was used to start the generation of a Lurbinectedin partially-resistant cell line. The generated information will constitute a hypothesis generator for subsequent analysis.

REFERENCES

7. References

- 1 Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, doi:10.3322/caac.21660 (2021).
- 2 Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a cancer journal for clinicians* 71, 209-249, doi:10.3322/caac.21660 (2021).
- 3 Bhoola, S. & Hoskins, W. J. Diagnosis and management of epithelial ovarian cancer. *Obstetrics and gynecology* 107, 1399-1410, doi:10.1097/01.AOG.0000220516.34053.48 (2006).
- 4 Shih, I. M., Wang, Y. & Wang, T. L. The Origin of Ovarian Cancer Species and Precancerous Landscape. *The American journal of pathology* 191, 26-39, doi:10.1016/j.ajpath.2020.09.006 (2021).
- 5 Lheureux, S., Gourley, C., Vergote, I. & Oza, A. M. Epithelial ovarian cancer. *Lancet* 393, 1240-1253, doi:10.1016/S0140-6736(18)32552-2 (2019).
- 6 Torre, L. A. *et al.* Ovarian cancer statistics, 2018. *CA: a cancer journal for clinicians* 68, 284-296, doi:10.3322/caac.21456 (2018).
- 7 Zhang, R., Siu, M. K. Y., Ngan, H. Y. S. & Chan, K. K. L. Molecular Biomarkers for the Early Detection of Ovarian Cancer. *International journal of molecular sciences* 23, doi:10.3390/ijms231912041 (2022).
- 8 Partridge, E. E. & Barnes, M. N. Epithelial ovarian cancer: prevention, diagnosis, and treatment. *CA: a cancer journal for clinicians* 49, 297-320, doi:10.3322/canjclin.49.5.297 (1999).
- 9 Rowswell-Turner, R. B. *et al.* HE4 Overexpression by Ovarian Cancer Promotes a Suppressive Tumor Immune Microenvironment and Enhanced Tumor and Macrophage PD-L1 Expression. *Journal of immunology* 206, 2478-2488, doi:10.4049/jimmunol.2000281 (2021).
- 10 Moore, R. G. *et al.* A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecologic oncology* 112, 40-46, doi:10.1016/j.ygyno.2008.08.031 (2009).
- 11 Wang, H., Liu, P., Xu, H. & Dai, H. Early diagnosis of ovarian cancer: serum HE4, CA125 and ROMA model. *American journal of translational research* 13, 14141-14148 (2021).
- 12 Colombo, N. *et al.* Newly diagnosed and relapsed epithelial ovarian carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology* 21 Suppl 5, v23-30, doi:10.1093/annonc/mdq244 (2010).
- 13 American College of, O. & Gynecologists' Committee on Practice, B.-G. Practice Bulletin No. 174: Evaluation and Management of Adnexal Masses. *Obstetrics and gynecology* 128, e210-e226, doi:10.1097/AOG.0000000000001768 (2016).

- 14 Rojas, V., Hirshfield, K. M., Ganesan, S. & Rodriguez-Rodriguez, L. Molecular Characterization of Epithelial Ovarian Cancer: Implications for Diagnosis and Treatment. *International journal of molecular sciences* 17, doi:10.3390/ijms17122113 (2016).
- 15 McCluggage, W. G. Morphological subtypes of ovarian carcinoma: a review with emphasis on new developments and pathogenesis. *Pathology* 43, 420-432, doi:10.1097/PAT.0b013e328348a6e7 (2011).
- 16 Prat, J. & Oncology, F. C. o. G. Staging classification for cancer of the ovary, fallopian tube, and peritoneum. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics* 124, 1-5, doi:10.1016/j.ijgo.2013.10.001 (2014).
- 17 Telloni, S. M. Tumor Staging and Grading: A Primer. *Methods in molecular biology* 1606, 1-17, doi:10.1007/978-1-4939-6990-6_1 (2017).
- 18 Colombo, N. *et al.* ESMO-ESGO consensus conference recommendations on ovarian cancer: pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society*, doi:10.1136/ijgc-2019-000308 (2019).
- 19 Redondo, A. *et al.* SEOM clinical guideline in ovarian cancer (2020). *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* 23, 961-968, doi:10.1007/s12094-020-02545-x (2021).
- 20 Sueblinvong, T. & Carney, M. E. Current understanding of risk factors for ovarian cancer. *Current treatment options in oncology* 10, 67-81, doi:10.1007/s11864-009-0108-2 (2009).
- 21 Rooth, C. Ovarian cancer: risk factors, treatment and management. *British journal of nursing* 22, S23-30, doi:10.12968/bjon.2013.22.Sup17.S23 (2013).
- 22 Brekelmans, C. T. Risk factors and risk reduction of breast and ovarian cancer. *Current opinion in obstetrics & gynecology* 15, 63-68, doi:10.1097/00001703-200302000-00010 (2003).
- 23 La Vecchia, C. Ovarian cancer: epidemiology and risk factors. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation* 26, 55-62, doi:10.1097/CEJ.0000000000000217 (2017).
- 24 Ness, R. B. *et al.* Infertility, fertility drugs, and ovarian cancer: a pooled analysis of case-control studies. *American journal of epidemiology* 155, 217-224, doi:10.1093/aje/155.3.217 (2002).
- 25 Schildkraut, J. M. *et al.* Age at natural menopause and the risk of epithelial ovarian cancer. *Obstetrics and gynecology* 98, 85-90, doi:10.1016/s0029-7844(01)01388-6 (2001).

- 26 Olsen, C. M. *et al.* Obesity and the risk of epithelial ovarian cancer: a systematic review and meta-analysis. *European journal of cancer* 43, 690-709, doi:10.1016/j.ejca.2006.11.010 (2007).
- 27 Momenimovahed, Z., Tiznobaik, A., Taheri, S. & Salehiniya, H. Ovarian cancer in the world: epidemiology and risk factors. *International journal of women's health* 11, 287-299, doi:10.2147/IJWH.S197604 (2019).
- 28 Collaborative Group on Epidemiological Studies of Ovarian, C. *et al.* Ovarian cancer and smoking: individual participant meta-analysis including 28,114 women with ovarian cancer from 51 epidemiological studies. *The Lancet. Oncology* 13, 946-956, doi:10.1016/S1470-2045(12)70322-4 (2012).
- 29 Pavone, M. E. & Lyttle, B. M. Endometriosis and ovarian cancer: links, risks, and challenges faced. *International journal of women's health* 7, 663-672, doi:10.2147/IJWH.S66824 (2015).
- 30 Jervis, S. *et al.* Ovarian cancer familial relative risks by tumour subtypes and by known ovarian cancer genetic susceptibility variants. *Journal of medical genetics* 51, 108-113, doi:10.1136/jmedgenet-2013-102015 (2014).
- 31 Stratton, J. F., Pharoah, P., Smith, S. K., Easton, D. & Ponder, B. A. A systematic review and meta-analysis of family history and risk of ovarian cancer. *British journal of obstetrics and gynaecology* 105, 493-499, doi:10.1111/j.1471-0528.1998.tb10148.x (1998).
- 32 Boyd, J. Specific keynote: hereditary ovarian cancer: what we know. *Gynecologic oncology* 88, S8-10; discussion S11-13, doi:10.1006/gyno.2002.6674 (2003).
- 33 Song, H. *et al.* Contribution of Germline Mutations in the RAD51B, RAD51C, and RAD51D Genes to Ovarian Cancer in the Population. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 33, 2901-2907, doi:10.1200/JCO.2015.61.2408 (2015).
- 34 Ramus, S. J. *et al.* Germline Mutations in the BRIP1, BARD1, PALB2, and NBN Genes in Women With Ovarian Cancer. *Journal of the National Cancer Institute* 107, doi:10.1093/jnci/djv214 (2015).
- 35 Ketabi, Z. *et al.* Ovarian cancer linked to Lynch syndrome typically presents as early-onset, non-serous epithelial tumors. *Gynecologic oncology* 121, 462-465, doi:10.1016/j.ygyno.2011.02.010 (2011).
- 36 Lynch, H. T., Casey, M. J., Lynch, J., White, T. E. & Godwin, A. K. Genetics and ovarian carcinoma. *Seminars in oncology* 25, 265-280 (1998).
- 37 Kuroki, L. & Guntupalli, S. R. Treatment of epithelial ovarian cancer. *Bmj* 371, m3773, doi:10.1136/bmj.m3773 (2020).
- 38 Bentivegna, E. *et al.* Fertility-sparing surgery in epithelial ovarian cancer: a systematic review of oncological issues. *Annals of oncology : official journal of the European Society for Medical Oncology* 27, 1994-2004, doi:10.1093/annonc/mdw311 (2016).

- 39 Collinson, F. *et al.* Optimal treatment of early-stage ovarian cancer. *Annals of oncology : official journal of the European Society for Medical Oncology* 25, 1165-1171, doi:10.1093/annonc/mdu116 (2014).
- 40 Lawrie, T. A., Winter-Roach, B. A., Heus, P. & Kitchener, H. C. Adjuvant (post-surgery) chemotherapy for early stage epithelial ovarian cancer. *The Cochrane database of systematic reviews* 2015, CD004706, doi:10.1002/14651858.CD004706.pub5 (2015).
- 41 Oseledchyk, A. *et al.* Adjuvant chemotherapy in patients with stage I endometrioid or clear cell ovarian cancer in the platinum era: a Surveillance, Epidemiology, and End Results Cohort Study, 2000-2013. *Annals of oncology : official journal of the European Society for Medical Oncology* 28, 2985-2993, doi:10.1093/annonc/mdx525 (2017).
- 42 Karam, A. *et al.* Fifth Ovarian Cancer Consensus Conference of the Gynecologic Cancer InterGroup: first-line interventions. *Annals of oncology : official journal of the European Society for Medical Oncology* 28, 711-717, doi:10.1093/annonc/mdx011 (2017).
- 43 Bristow, R. E., Puri, I. & Chi, D. S. Cytoreductive surgery for recurrent ovarian cancer: a meta-analysis. *Gynecologic oncology* 112, 265-274, doi:10.1016/j.ygyno.2008.08.033 (2009).
- 44 Iorio, G. C., Martini, S., Arcadipane, F., Ricardi, U. & Franco, P. The role of radiotherapy in epithelial ovarian cancer: a literature overview. *Medical oncology* 36, 64, doi:10.1007/s12032-019-1287-8 (2019).
- 45 Aghajanian, C. *et al.* OCEANS: a randomized, double-blind, placebo-controlled phase III trial of chemotherapy with or without bevacizumab in patients with platinum-sensitive recurrent epithelial ovarian, primary peritoneal, or fallopian tube cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 30, 2039-2045, doi:10.1200/JCO.2012.42.0505 (2012).
- 46 Coleman, R. L. *et al.* Bevacizumab and paclitaxel-carboplatin chemotherapy and secondary cytoreduction in recurrent, platinum-sensitive ovarian cancer (NRG Oncology/Gynecologic Oncology Group study GOG-0213): a multicentre, open-label, randomised, phase 3 trial. *The Lancet. Oncology* 18, 779-791, doi:10.1016/S1470-2045(17)30279-6 (2017).
- 47 Ledermann, J. A. *et al.* Cediranib in patients with relapsed platinum-sensitive ovarian cancer (ICON6): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet* 387, 1066-1074, doi:10.1016/S0140-6736(15)01167-8 (2016).
- 48 Grimm, C., Cropet, C. & Ray-Coquard, I. Maintenance olaparib plus bevacizumab (bev) after platinum-based chemotherapy plus bev in patients (pts) with newly diagnosed advanced high-grade ovarian cancer (HGOC): Efficacy by timing of surgery and residual tumor status in the Phase III PAOLA-1 trial. *Gynecologic oncology* 159, 19 (2020).
- 49 González-Martín, A. *et al.* (BMJ Specialist Journals, 2019).

- 50 Coleman, R. L. *et al.* VELIA/GOG-3005: Integration of veliparib with front-line chemotherapy and maintenance in women with high-grade serous carcinoma of ovarian, fallopian tube, or primary peritoneal origin. *Annals of Oncology* 30, v851-v934 (2019).
- 51 Ngoi, N. Y. L. & Tan, D. S. P. The role of homologous recombination deficiency testing in ovarian cancer and its clinical implications: do we need it? *ESMO open* 6, 100144, doi:10.1016/j.esmoop.2021.100144 (2021).
- 52 Cancer Genome Atlas Research, N. Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609-615, doi:10.1038/nature10166 (2011).
- 53 Nigro, J. M. *et al.* Mutations in the p53 gene occur in diverse human tumour types. *Nature* 342, 705-708, doi:10.1038/342705a0 (1989).
- 54 Etemadmoghadam, D. *et al.* Integrated genome-wide DNA copy number and expression analysis identifies distinct mechanisms of primary chemoresistance in ovarian carcinomas. *Clinical cancer research : an official journal of the American Association for Cancer Research* 15, 1417-1427, doi:10.1158/1078-0432.CCR-08-1564 (2009).
- 55 Farley, J. *et al.* Cyclin E expression is a significant predictor of survival in advanced, suboptimally debulked ovarian epithelial cancers: a Gynecologic Oncology Group study. *Cancer research* 63, 1235-1241 (2003).
- 56 Kudoh, K. *et al.* Gains of 1q21-q22 and 13q12-q14 are potential indicators for resistance to cisplatin-based chemotherapy in ovarian cancer patients. *Clinical cancer research : an official journal of the American Association for Cancer Research* 5, 2526-2531 (1999).
- 57 Li, Y. *et al.* Establish of an Initial Platinum-Resistance Predictor in High-Grade Serous Ovarian Cancer Patients Regardless of Homologous Recombination Deficiency Status. *Frontiers in oncology* 12, 847085, doi:10.3389/fonc.2022.847085 (2022).
- 58 Turner, N., Tutt, A. & Ashworth, A. Hallmarks of 'BRCAness' in sporadic cancers. *Nature reviews. Cancer* 4, 814-819, doi:10.1038/nrc1457 (2004).
- 59 Bast, R. C., Jr. & Mills, G. B. Personalizing therapy for ovarian cancer: BRCAness and beyond. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 28, 3545-3548, doi:10.1200/JCO.2010.28.5791 (2010).
- 60 Hennessy, B. T. *et al.* Somatic mutations in BRCA1 and BRCA2 could expand the number of patients that benefit from poly (ADP ribose) polymerase inhibitors in ovarian cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 28, 3570-3576, doi:10.1200/JCO.2009.27.2997 (2010).
- 61 Alsop, K. *et al.* BRCA mutation frequency and patterns of treatment response in BRCA mutation-positive women with ovarian cancer: a report from the Australian Ovarian Cancer Study Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 30, 2654-2663, doi:10.1200/JCO.2011.39.8545 (2012).

- 62 McCabe, N. *et al.* Deficiency in the repair of DNA damage by homologous recombination and sensitivity to poly(ADP-ribose) polymerase inhibition. *Cancer research* 66, 8109-8115, doi:10.1158/0008-5472.CAN-06-0140 (2006).
- 63 Mendes-Pereira, A. M. *et al.* Synthetic lethal targeting of PTEN mutant cells with PARP inhibitors. *EMBO molecular medicine* 1, 315-322, doi:10.1002/emmm.200900041 (2009).
- 64 Shen, W. H. *et al.* Essential role for nuclear PTEN in maintaining chromosomal integrity. *Cell* 128, 157-170, doi:10.1016/j.cell.2006.11.042 (2007).
- 65 Konstantinopoulos, P. A., Ceccaldi, R., Shapiro, G. I. & D'Andrea, A. D. Homologous Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian Cancer. *Cancer discovery* 5, 1137-1154, doi:10.1158/2159-8290.CD-15-0714 (2015).
- 66 Cancer Genome Atlas Research, N. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* 497, 67-73, doi:10.1038/nature12113 (2013).
- 67 Hughes-Davies, L. *et al.* EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. *Cell* 115, 523-535, doi:10.1016/s0092-8674(03)00930-9 (2003).
- 68 Brown, L. A. *et al.* Amplification of 11q13 in ovarian carcinoma. *Genes, chromosomes & cancer* 47, 481-489, doi:10.1002/gcc.20549 (2008).
- 69 Patch, A. M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* 521, 489-494, doi:10.1038/nature14410 (2015).
- 70 Koh, W. J. *et al.* Uterine Neoplasms, Version 1.2018, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network : JNCCN* 16, 170-199, doi:10.6004/jnccn.2018.0006 (2018).
- 71 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA: a cancer journal for clinicians* 68, 7-30, doi:10.3322/caac.21442 (2018).
- 72 Trojano, G. *et al.* Conservative treatment in early stage endometrial cancer: a review. *Acta bio-medica : Atenei Parmensis* 90, 405-410, doi:10.23750/abm.v90i4.7800 (2019).
- 73 Abdulfatah, E. *et al.* Gynecologic Cancers: Molecular Updates 2018. *Clinics in laboratory medicine* 38, 421-438, doi:10.1016/j.cl.2018.02.007 (2018).
- 74 Lax, S. F. Molecular genetic pathways in various types of endometrial carcinoma: from a phenotypical to a molecular-based classification. *Virchows Archiv : an international journal of pathology* 444, 213-223, doi:10.1007/s00428-003-0947-3 (2004).
- 75 Bokhman, J. V. Two pathogenetic types of endometrial carcinoma. *Gynecologic oncology* 15, 10-17, doi:10.1016/0090-8258(83)90111-7 (1983).
- 76 Soslow, R. A. *et al.* Endometrial Carcinoma Diagnosis: Use of FIGO Grading and Genomic Subcategories in Clinical Practice: Recommendations of the International Society of Gynecological Pathologists. *International journal*

- of gynecological pathology : official journal of the International Society of Gynecological Pathologists* 38 Suppl 1, S64-S74, doi:10.1097/PGP.0000000000000518 (2019).
- 77 Sorosky, J. I. Endometrial cancer. *Obstetrics and gynecology* 120, 383-397, doi:10.1097/AOG.0b013e3182605bf1 (2012).
- 78 Dork, T., Hillemanns, P., Tempfer, C., Breu, J. & Fleisch, M. C. Genetic Susceptibility to Endometrial Cancer: Risk Factors and Clinical Management. *Cancers* 12, doi:10.3390/cancers12092407 (2020).
- 79 Koskas, M., Amant, F., Mirza, M. R. & Creutzberg, C. L. Cancer of the corpus uteri: 2021 update. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics* 155 Suppl 1, 45-60, doi:10.1002/ijgo.13866 (2021).
- 80 Lauby-Secretan, B. *et al.* Body Fatness and Cancer--Viewpoint of the IARC Working Group. *The New England journal of medicine* 375, 794-798, doi:10.1056/NEJMSr1606602 (2016).
- 81 Barry, J. A., Azizia, M. M. & Hardiman, P. J. Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Human reproduction update* 20, 748-758, doi:10.1093/humupd/dmu012 (2014).
- 82 Saed, L. *et al.* The effect of diabetes on the risk of endometrial Cancer: an updated a systematic review and meta-analysis. *BMC cancer* 19, 527, doi:10.1186/s12885-019-5748-4 (2019).
- 83 Sjogren, L. L., Morch, L. S. & Lokkegaard, E. Hormone replacement therapy and the risk of endometrial cancer: A systematic review. *Maturitas* 91, 25-35, doi:10.1016/j.maturitas.2016.05.013 (2016).
- 84 Bernstein, L. *et al.* Tamoxifen therapy for breast cancer and endometrial cancer risk. *Journal of the National Cancer Institute* 91, 1654-1662, doi:10.1093/jnci/91.19.1654 (1999).
- 85 Bland, A. E. *et al.* Relationship between tamoxifen use and high risk endometrial cancer histologic types. *Gynecologic oncology* 112, 150-154, doi:10.1016/j.ygyno.2008.08.035 (2009).
- 86 Collaborative Group on Epidemiological Studies on Endometrial, C. Endometrial cancer and oral contraceptives: an individual participant meta-analysis of 27 276 women with endometrial cancer from 36 epidemiological studies. *The Lancet. Oncology* 16, 1061-1070, doi:10.1016/S1470-2045(15)00212-0 (2015).
- 87 Brinton, L. A. *et al.* Reproductive, menstrual, and medical risk factors for endometrial cancer: results from a case-control study. *American journal of obstetrics and gynecology* 167, 1317-1325, doi:10.1016/s0002-9378(11)91709-8 (1992).
- 88 Passarello, K., Kurian, S. & Villanueva, V. Endometrial Cancer: An Overview of Pathophysiology, Management, and Care. *Seminars in oncology nursing* 35, 157-165, doi:10.1016/j.soncn.2019.02.002 (2019).

- 89 Aaltonen, M. H., Staff, S., Mecklin, J. P., Pylvanainen, K. & Maenpaa, J. U. Comparison of lifestyle, hormonal and medical factors in women with sporadic and Lynch syndrome-associated endometrial cancer: A retrospective case-case study. *Molecular and clinical oncology* 6, 758-764, doi:10.3892/mco.2017.1211 (2017).
- 90 Gupta, S. *et al.* NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 3.2017. *Journal of the National Comprehensive Cancer Network : JNCCN* 15, 1465-1475, doi:10.6004/jnccn.2017.0176 (2017).
- 91 Mester, J. & Eng, C. Cowden syndrome: recognizing and managing a not-so-rare hereditary cancer syndrome. *Journal of surgical oncology* 111, 125-130, doi:10.1002/jso.23735 (2015).
- 92 Staff, S. *et al.* Endometrial cancer risk factors among Lynch syndrome women: a retrospective cohort study. *Br J Cancer* 115, 375-381, doi:10.1038/bjc.2016.193 (2016).
- 93 Win, A. K., Reece, J. C. & Ryan, S. Family history and risk of endometrial cancer: a systematic review and meta-analysis. *Obstetrics and gynecology* 125, 89-98, doi:10.1097/AOG.0000000000000563 (2015).
- 94 Johnatty, S. E. *et al.* Family history of cancer predicts endometrial cancer risk independently of Lynch Syndrome: Implications for genetic counselling. *Gynecol Oncol* 147, 381-387, doi:10.1016/j.ygyno.2017.08.011 (2017).
- 95 O'Mara, T. A., Glubb, D. M., Kho, P. F., Thompson, D. J. & Spurdle, A. B. Genome-Wide Association Studies of Endometrial Cancer: Latest Developments and Future Directions. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 28, 1095-1102, doi:10.1158/1055-9965.EPI-18-1031 (2019).
- 96 Wright, J. D., Barrera Medel, N. I., Sehouli, J., Fujiwara, K. & Herzog, T. J. Contemporary management of endometrial cancer. *Lancet* 379, 1352-1360, doi:10.1016/S0140-6736(12)60442-5 (2012).
- 97 Kilgore, L. C. *et al.* Adenocarcinoma of the endometrium: survival comparisons of patients with and without pelvic node sampling. *Gynecol Oncol* 56, 29-33, doi:10.1006/gyno.1995.1005 (1995).
- 98 Cragun, J. M. *et al.* Retrospective analysis of selective lymphadenectomy in apparent early-stage endometrial cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 23, 3668-3675, doi:10.1200/JCO.2005.04.144 (2005).
- 99 Lutman, C. V. *et al.* Pelvic lymph node count is an important prognostic variable for FIGO stage I and II endometrial carcinoma with high-risk histology. *Gynecol Oncol* 102, 92-97, doi:10.1016/j.ygyno.2005.11.032 (2006).

- 100 Chan, J. K. *et al.* Therapeutic role of lymph node resection in endometrioid corpus cancer: a study of 12,333 patients. *Cancer* 107, 1823-1830, doi:10.1002/cncr.22185 (2006).
- 101 Goudge, C., Bernhard, S., Cloven, N. G. & Morris, P. The impact of complete surgical staging on adjuvant treatment decisions in endometrial cancer. *Gynecol Oncol* 93, 536-539, doi:10.1016/j.ygyno.2004.02.020 (2004).
- 102 Homesley, H. D., Boike, G. & Spiegel, G. W. Feasibility of laparoscopic management of presumed stage I endometrial carcinoma and assessment of accuracy of myoinvasion estimates by frozen section: a gynecologic oncology group study. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society* 14, 341-347, doi:10.1111/j.1048-891x.2004.014219.x (2004).
- 103 Ghezzi, F. *et al.* Laparoscopic staging of early ovarian cancer: results of a multi-institutional cohort study. *Annals of surgical oncology* 19, 1589-1594, doi:10.1245/s10434-011-2138-9 (2012).
- 104 Podzielinski, I. *et al.* Primary radiation therapy for medically inoperable patients with clinical stage I and II endometrial carcinoma. *Gynecol Oncol* 124, 36-41, doi:10.1016/j.ygyno.2011.09.022 (2012).
- 105 Lee, T. S. *et al.* Ovarian preservation during the surgical treatment of early stage endometrial cancer: a nation-wide study conducted by the Korean Gynecologic Oncology Group. *Gynecol Oncol* 115, 26-31, doi:10.1016/j.ygyno.2009.06.041 (2009).
- 106 Richter, C. E. *et al.* Ovarian preservation and staging in reproductive-age endometrial cancer patients. *Gynecol Oncol* 114, 99-104, doi:10.1016/j.ygyno.2009.03.032 (2009).
- 107 Wright, J. D. *et al.* Safety of ovarian preservation in premenopausal women with endometrial cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 27, 1214-1219, doi:10.1200/JCO.2008.19.8150 (2009).
- 108 Sorbe, B. G. *et al.* External pelvic and vaginal irradiation versus vaginal irradiation alone as postoperative therapy in medium-risk endometrial carcinoma: a prospective, randomized study—quality-of-life analysis. *International Journal of Gynecologic Cancer* 22 (2012).
- 109 Nout, R. *et al.* Vaginal brachytherapy versus pelvic external beam radiotherapy for patients with endometrial cancer of high-intermediate risk (PORTEC-2): an open-label, non-inferiority, randomised trial. *The Lancet* 375, 816-823 (2010).
- 110 Wortman, B. G. *et al.* Ten-year results of the PORTEC-2 trial for high-intermediate risk endometrial carcinoma: improving patient selection for adjuvant therapy. *British journal of cancer* 119, 1067-1074 (2018).
- 111 Huh, W. K. *et al.* Salvage of isolated vaginal recurrences in women with surgical stage I endometrial cancer: a multiinstitutional experience. *International journal of gynecological cancer : official journal of the*

- International Gynecological Cancer Society* 17, 886-889, doi:10.1111/j.1525-1438.2007.00858.x (2007).
- 112 Bristow, R. E. *et al.* Salvage cytoreductive surgery for recurrent endometrial cancer. *Gynecol Oncol* 103, 281-287, doi:10.1016/j.ygyno.2006.03.011 (2006).
- 113 Oaknin, A. *et al.* Endometrial cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology* 33, 860-877, doi:10.1016/j.annonc.2022.05.009 (2022).
- 114 Yetkin-Arik, B. *et al.* Angiogenesis in gynecological cancers and the options for anti-angiogenesis therapy. *Biochimica et biophysica acta. Reviews on cancer* 1875, 188446, doi:10.1016/j.bbcan.2020.188446 (2021).
- 115 Roncolato, F., Lindemann, K., Willson, M. L., Martyn, J. & Mileschkin, L. PI3K/AKT/mTOR inhibitors for advanced or recurrent endometrial cancer. *The Cochrane database of systematic reviews* 10, CD012160, doi:10.1002/14651858.CD012160.pub2 (2019).
- 116 Lee, Y. C., Lheureux, S. & Oza, A. M. Treatment strategies for endometrial cancer: current practice and perspective. *Current opinion in obstetrics & gynecology* 29, 47-58, doi:10.1097/GCO.0000000000000338 (2017).
- 117 Lee, Y. C., Lheureux, S., Mirza, M. R. & Oza, A. M. in *Management of Endometrial Cancer* 249-276 (Springer, 2020).
- 118 Marin-Jimenez, J. A., Garcia-Mulero, S., Matias-Guiu, X. & Piulats, J. M. Facts and Hopes in Immunotherapy of Endometrial Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 28, 4849-4860, doi:10.1158/1078-0432.CCR-21-1564 (2022).
- 119 Musacchio, L. *et al.* PARP Inhibitors in Endometrial Cancer: Current Status and Perspectives. *Cancer management and research* 12, 6123-6135, doi:10.2147/CMAR.S221001 (2020).
- 120 Oaknin, A. *et al.* Clinical Activity and Safety of the Anti-Programmed Death 1 Monoclonal Antibody Dostarlimab for Patients With Recurrent or Advanced Mismatch Repair-Deficient Endometrial Cancer: A Nonrandomized Phase 1 Clinical Trial. *JAMA oncology* 6, 1766-1772, doi:10.1001/jamaoncol.2020.4515 (2020).
- 121 Ott, P. A. *et al.* Safety and antitumor activity of pembrolizumab in advanced programmed death ligand 1-positive endometrial cancer: results from the KEYNOTE-028 study. *Obstetrical & Gynecological Survey* 73, 26-27 (2018).
- 122 Siedel, J. H. *et al.* Clinical significance of homologous recombination deficiency score testing in endometrial Cancer. *Gynecologic oncology* 160, 777-785 (2021).
- 123 Rayner, E. *et al.* A panoply of errors: polymerase proofreading domain mutations in cancer. *Nature reviews. Cancer* 16, 71-81, doi:10.1038/nrc.2015.12 (2016).

- 124 McDonald, M. E. & Bender, D. P. Endometrial Cancer: Obesity, Genetics, and Targeted Agents. *Obstetrics and gynecology clinics of North America* 46, 89-105, doi:10.1016/j.ogc.2018.09.006 (2019).
- 125 Bell, D. W. & Ellenson, L. H. Molecular Genetics of Endometrial Carcinoma. *Annual review of pathology* 14, 339-367, doi:10.1146/annurev-pathol-020117-043609 (2019).
- 126 Talhouk, A. *et al.* A clinically applicable molecular-based classification for endometrial cancers. *British journal of cancer* 113, 299-310, doi:10.1038/bjc.2015.190 (2015).
- 127 Talhouk, A. *et al.* Confirmation of ProMisE: A simple, genomics-based clinical classifier for endometrial cancer. *Cancer* 123, 802-813, doi:10.1002/cncr.30496 (2017).
- 128 Stelloo, E. *et al.* Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a TransPORTEC initiative. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 28, 836-844, doi:10.1038/modpathol.2015.43 (2015).
- 129 Stelloo, E. *et al.* Improved Risk Assessment by Integrating Molecular and Clinicopathological Factors in Early-stage Endometrial Cancer-Combined Analysis of the PORTEC Cohorts. *Clinical cancer research : an official journal of the American Association for Cancer Research* 22, 4215-4224, doi:10.1158/1078-0432.CCR-15-2878 (2016).
- 130 Talhouk, A. & McAlpine, J. N. New classification of endometrial cancers: the development and potential applications of genomic-based classification in research and clinical care. *Gynecologic oncology research and practice* 3, 14, doi:10.1186/s40661-016-0035-4 (2016).
- 131 Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer discovery* 12, 31-46, doi:10.1158/2159-8290.CD-21-1059 (2022).
- 132 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* 144, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
- 133 Shen, Z. Genomic instability and cancer: an introduction. *Journal of molecular cell biology* 3, 1-3, doi:10.1093/jmcb/mjq057 (2011).
- 134 Trenner, A. & Sartori, A. A. Harnessing DNA Double-Strand Break Repair for Cancer Treatment. *Frontiers in oncology* 9, 1388, doi:10.3389/fonc.2019.01388 (2019).
- 135 Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* 168, 644-656, doi:10.1016/j.cell.2017.01.002 (2017).
- 136 Wang, J. C. Cellular roles of DNA topoisomerases: a molecular perspective. *Nature reviews. Molecular cell biology* 3, 430-440, doi:10.1038/nrm831 (2002).
- 137 Hegde, M. L., Hazra, T. K. & Mitra, S. Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell research* 18, 27-47, doi:10.1038/cr.2008.8 (2008).

- 138 Pfeiffer, P., Goedecke, W. & Obe, G. Mechanisms of DNA double-strand
break repair and their potential to induce chromosomal aberrations.
Mutagenesis 15, 289-302, doi:10.1093/mutage/15.4.289 (2000).
- 139 Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair, and
mutagenesis. *Environmental and molecular mutagenesis* 58, 235-263,
doi:10.1002/em.22087 (2017).
- 140 Mehrotra, S. & Mitra, I. Origin of Genome Instability and Determinants of
Mutational Landscape in Cancer Cells. *Genes* 11,
doi:10.3390/genes11091101 (2020).
- 141 Pikor, L., Thu, K., Vucic, E. & Lam, W. The detection and implication of
genome instability in cancer. *Cancer metastasis reviews* 32, 341-352,
doi:10.1007/s10555-013-9429-5 (2013).
- 142 Pilie, P. G., Gay, C. M., Byers, L. A., O'Connor, M. J. & Yap, T. A. PARP
Inhibitors: Extending Benefit Beyond BRCA-Mutant Cancers. *Clinical cancer
research : an official journal of the American Association for Cancer
Research* 25, 3759-3771, doi:10.1158/1078-0432.CCR-18-0968 (2019).
- 143 Ranjha, L., Howard, S. M. & Cejka, P. Main steps in DNA double-strand
break repair: an introduction to homologous recombination and related
processes. *Chromosoma* 127, 187-214, doi:10.1007/s00412-017-0658-1
(2018).
- 144 Roy, R., Chun, J. & Powell, S. N. BRCA1 and BRCA2: different roles in a
common pathway of genome protection. *Nature reviews. Cancer* 12, 68-
78, doi:10.1038/nrc3181 (2011).
- 145 Zhao, E. Y., Jones, M. & Jones, S. J. M. Whole-Genome Sequencing in
Cancer. *Cold Spring Harbor perspectives in medicine* 9,
doi:10.1101/cshperspect.a034579 (2019).
- 146 Nik-Zainal, S. From genome integrity to cancer. *Genome medicine* 11, 4,
doi:10.1186/s13073-019-0617-y (2019).
- 147 Marquard, A. M. *et al.* Pan-cancer analysis of genomic scar signatures
associated with homologous recombination deficiency suggests novel
indications for existing cancer drugs. *Biomarker research* 3, 9,
doi:10.1186/s40364-015-0033-4 (2015).
- 148 Birkbak, N. J. *et al.* Telomeric allelic imbalance indicates defective DNA
repair and sensitivity to DNA-damaging agents. *Cancer discovery* 2, 366-
375, doi:10.1158/2159-8290.CD-11-0206 (2012).
- 149 Popova, T. *et al.* Ploidy and large-scale genomic instability consistently
identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer
research* 72, 5454-5462, doi:10.1158/0008-5472.CAN-12-1470 (2012).
- 150 Abkevich, V. *et al.* Patterns of genomic loss of heterozygosity predict
homologous recombination repair defects in epithelial ovarian cancer.
British journal of cancer 107, 1776-1782, doi:10.1038/bjc.2012.451 (2012).
- 151 Riaz, N. *et al.* Pan-cancer analysis of bi-allelic alterations in homologous
recombination DNA repair genes. *Nat Commun* 8, 857,
doi:10.1038/s41467-017-00921-w (2017).

- 152 Hasty, P. & Montagna, C. Chromosomal Rearrangements in Cancer: Detection and potential causal mechanisms. *Molecular & cellular oncology* 1, doi:10.4161/mco.29904 (2014).
- 153 Lucito, R. *et al.* Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res* 13, 2291-2305, doi:10.1101/gr.1349003 (2003).
- 154 Carter, N. P. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nature genetics* 39, S16-S21 (2007).
- 155 Snijders, A. M. *et al.* Assembly of microarrays for genome-wide measurement of DNA copy number. *Nature genetics* 29, 263-264 (2001).
- 156 Goldstein, D. B. *et al.* Sequencing studies in human genetics: design and interpretation. *Nature reviews. Genetics* 14, 460-470, doi:10.1038/nrg3455 (2013).
- 157 Baudhuin, L. M. A new era of genetic testing and its impact on research and clinical care. *Clinical chemistry* 58, 1070-1071, doi:10.1373/clinchem.2012.185314 (2012).
- 158 Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nature Reviews Genetics* 11, 685-696 (2010).
- 159 Pirooznia, M., Goes, F. S. & Zandi, P. P. Whole-genome CNV analysis: advances in computational approaches. *Frontiers in genetics* 6, 138, doi:10.3389/fgene.2015.00138 (2015).
- 160 Zhao, M., Wang, Q., Wang, Q., Jia, P. & Zhao, Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC bioinformatics* 14 Suppl 11, S1, doi:10.1186/1471-2105-14-S11-S1 (2013).
- 161 Roca, I., González-Castro, L., Fernández, H., Couce, M. L. & Fernández-Marmiesse, A. Free-access copy-number variant detection tools for targeted next-generation sequencing data. *Mutation Research/Reviews in Mutation Research* 779, 114-125 (2019).
- 162 Zhuang, S. *et al.* A transcriptional signature detects homologous recombination deficiency in pancreatic cancer at the individual level. *Molecular therapy. Nucleic acids* 26, 1014-1026, doi:10.1016/j.omtn.2021.10.014 (2021).
- 163 Kang, J., Lee, J., Lee, A. & Lee, Y. S. Prediction of homologous recombination deficiency from cancer gene expression data. *The Journal of international medical research* 50, 3000605221133655, doi:10.1177/03000605221133655 (2022).
- 164 Peng, G. *et al.* Genome-wide transcriptome profiling of homologous recombination DNA repair. *Nature communications* 5, 3361, doi:10.1038/ncomms4361 (2014).
- 165 Konstantinopoulos, P. A. *et al.* Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in

- patients with epithelial ovarian cancer. *Journal of clinical oncology* 28, 3555 (2010).
- 166 Castroviejo-Bermejo, M. *et al.* A RAD51 assay feasible in routine tumor samples calls PARP inhibitor response beyond BRCA mutation. *EMBO molecular medicine* 10, doi:10.15252/emmm.201809172 (2018).
- 167 Hoppe, M. M., Sundar, R., Tan, D. S. P. & Jeyasekharan, A. D. Biomarkers for Homologous Recombination Deficiency in Cancer. *Journal of the National Cancer Institute* 110, 704-713, doi:10.1093/jnci/djy085 (2018).
- 168 Ladan, M. M., van Gent, D. C. & Jager, A. Homologous Recombination Deficiency Testing for BRCA-Like Tumors: The Road to Clinical Validation. *Cancers (Basel)* 13, doi:10.3390/cancers13051004 (2021).
- 169 Beaufort, C. M. *et al.* Ovarian cancer cell line panel (OCCP): clinical importance of in vitro morphological subtypes. *PLoS one* 9, e103988 (2014).
- 170 Lopez-Reig, R. & Lopez-Guerrero, J. A. The hallmarks of ovarian cancer: proliferation and cell growth. *EJC supplements : EJC : official journal of EORTC, European Organization for Research and Treatment of Cancer ... [et al.]* 15, 27-37, doi:10.1016/j.ejcsup.2019.12.001 (2020).
- 171 Beaufort, C. M. *et al.* Ovarian cancer cell line panel (OCCP): clinical importance of in vitro morphological subtypes. *PLoS one* 9, e103988, doi:10.1371/journal.pone.0103988 (2014).
- 172 Domcke, S., Sinha, R., Levine, D. A., Sander, C. & Schultz, N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nature communications* 4, 2126, doi:10.1038/ncomms3126 (2013).
- 173 Mirabelli, P., Coppola, L. & Salvatore, M. Cancer Cell Lines Are Useful Model Systems for Medical Research. *Cancers* 11, doi:10.3390/cancers11081098 (2019).
- 174 Poveda, A. *et al.* A phase I dose-finding, pharmacokinetics and genotyping study of olaparib and lurbinectedin in patients with advanced solid tumors. *Scientific reports* 11, 4433, doi:10.1038/s41598-021-82671-w (2021).
- 175 Poveda, A. *et al.* Phase 2 Trial (POLA Study) of Lurbinectedin plus Olaparib in Patients with Advanced Solid Tumors: Results of Efficacy, Tolerability, and the Translational Study. *Cancers* 14, doi:10.3390/cancers14040915 (2022).
- 176 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* 14, 178-192, doi:10.1093/bib/bbs017 (2013).
- 177 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature biotechnology* 29, 24-26, doi:10.1038/nbt.1754 (2011).
- 178 Park, J., Shin, S., Yoo, H. M., Lee, S. W. & Kim, J. G. Evaluation of the Three Customized MSI Panels to Improve the Detection of Microsatellite

- Instability in Gastric Cancer. *Clinical laboratory* 63, 705-716, doi:10.7754/Clin.Lab.2016.161029 (2017).
- 179 Umar, A. *et al.* Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *Journal of the National Cancer Institute* 96, 261-268, doi:10.1093/jnci/djh034 (2004).
- 180 Stuppia, L., Antonucci, I., Palka, G. & Gatta, V. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *International journal of molecular sciences* 13, 3245-3276, doi:10.3390/ijms13033245 (2012).
- 181 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 182 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
- 183 Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC bioinformatics* 10, 161, doi:10.1186/1471-2105-10-161 (2009).
- 184 Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS computational biology* 12, e1004873, doi:10.1371/journal.pcbi.1004873 (2016).
- 185 Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology* 31, 213-219, doi:10.1038/nbt.2514 (2013).
- 186 Zhang, Z. & Hao, K. SAAS-CNV: A Joint Segmentation Approach on Aggregated and Allele Specific Signals for the Identification of Somatic Copy Number Alterations with Next-Generation Sequencing Data. *PLoS computational biology* 11, e1004618, doi:10.1371/journal.pcbi.1004618 (2015).
- 187 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20, 1297-1303, doi:10.1101/gr.107524.110 (2010).
- 188 Sztupinszki, Z. *et al.* Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *NPJ breast cancer* 4, 16, doi:10.1038/s41523-018-0066-6 (2018).
- 189 Povysil, G. *et al.* panelcn.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics. *Human mutation* 38, 889-897, doi:10.1002/humu.23237 (2017).
- 190 Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data.

- Biostatistics* (Oxford, England) 5, 557-572, doi:10.1093/biostatistics/kxh008 (2004).
- 191 Pettini, F., Visibelli, A., Cicaloni, V., Iovinelli, D. & Spiga, O. Multi-Omics Model Applied to Cancer Genetics. *International journal of molecular sciences* 22, doi:10.3390/ijms22115751 (2021).
- 192 Du, W. & Elemento, O. Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. *Oncogene* 34, 3215-3225, doi:10.1038/onc.2014.291 (2015).
- 193 de Anda-Jauregui, G. & Hernandez-Lemus, E. Computational Oncology in the Multi-Omics Era: State of the Art. *Frontiers in oncology* 10, 423, doi:10.3389/fonc.2020.00423 (2020).
- 194 Chakraborty, S., Hosen, M. I., Ahmed, M. & Shekhar, H. U. Onco-Multi-OMICS Approach: A New Frontier in Cancer Research. *BioMed research international* 2018, 9836256, doi:10.1155/2018/9836256 (2018).
- 195 Long, Y., Lu, M., Cheng, T., Zhan, X. & Zhan, X. Multiomics-Based Signaling Pathway Network Alterations in Human Non-functional Pituitary Adenomas. *Frontiers in endocrinology* 10, 835, doi:10.3389/fendo.2019.00835 (2019).
- 196 Ma'ayan, A. Network integration and graph analysis in mammalian molecular systems biology. *IET systems biology* 2, 206-221, doi:10.1049/iet-syb:20070075 (2008).
- 197 Parkinson, D. R., Johnson, B. E. & Sledge, G. W. Making personalized cancer medicine a reality: challenges and opportunities in the development of biomarkers and companion diagnostics. *Clinical cancer research : an official journal of the American Association for Cancer Research* 18, 619-624, doi:10.1158/1078-0432.CCR-11-2017 (2012).
- 198 Nicora, G., Vitali, F., Dagiati, A., Geifman, N. & Bellazzi, R. Integrated Multi-Omics Analyses in Oncology: A Review of Machine Learning Methods and Tools. *Frontiers in oncology* 10, 1030, doi:10.3389/fonc.2020.01030 (2020).
- 199 Guhathakurta, D., Sheikh, N. A., Meagher, T. C., Letarte, S. & Trager, J. B. Applications of systems biology in cancer immunotherapy: from target discovery to biomarkers of clinical outcome. *Expert review of clinical pharmacology* 6, 387-401, doi:10.1586/17512433.2013.811814 (2013).
- 200 Knijnenburg, T. A. *et al.* Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell reports* 23, 239-254 e236, doi:10.1016/j.celrep.2018.03.076 (2018).
- 201 Javle, M. & Curtin, N. J. The role of PARP in DNA repair and its therapeutic exploitation. *British journal of cancer* 105, 1114-1122, doi:10.1038/bjc.2011.382 (2011).
- 202 Damia, G. & Broggin, M. Platinum Resistance in Ovarian Cancer: Role of DNA Repair. *Cancers* 11, doi:10.3390/cancers11010119 (2019).

- 203 Ledermann, J. A., Drew, Y. & Kristeleit, R. S. Homologous recombination
deficiency and ovarian cancer. *European journal of cancer* 60, 49-58,
doi:10.1016/j.ejca.2016.03.005 (2016).
- 204 da Cunha Colombo Bonadio, R. R., Fogace, R. N., Miranda, V. C. & Diz, M.
Homologous recombination deficiency in ovarian cancer: a review of its
epidemiology and management. *Clinics* 73, e450s,
doi:10.6061/clinics/2018/e450s (2018).
- 205 Wagener-Ryczek, S., Merkelbach-Bruse, S. & Siemanowski, J. Biomarkers
for Homologous Recombination Deficiency in Cancer. *Journal of
personalized medicine* 11, doi:10.3390/jpm11070612 (2021).
- 206 Ray-Coquard, I. *et al.* Olaparib plus Bevacizumab as First-Line Maintenance
in Ovarian Cancer. *The New England journal of medicine* 381, 2416-2428,
doi:10.1056/NEJMoa1911361 (2019).
- 207 Lwin, Z. *et al.* LBA41 LEAP-005: Phase II study of lenvatinib (len) plus
pembrolizumab (pembro) in patients (pts) with previously treated
advanced solid tumours. *Annals of Oncology* 31, S1170 (2020).
- 208 Gonzalez-Martin, A. *et al.* Niraparib in Patients with Newly Diagnosed
Advanced Ovarian Cancer. *The New England journal of medicine* 381,
2391-2402, doi:10.1056/NEJMoa1910962 (2019).
- 209 Kaseb, A. *et al.* Open-label, perioperative phase 2 study evaluating
nivolumab alone versus nivolumab plus ipilimumab in patient with
resectable HCC. *Ann. Oncol* 30, v851-v934 (2019).
- 210 Qin, M. *et al.* SCNVSim: somatic copy number variation and structure
variation simulator. *BMC bioinformatics* 16, 66, doi:10.1186/s12859-015-
0502-7 (2015).
- 211 Carter, N. P. Methods and strategies for analyzing copy number variation
using DNA microarrays. *Nature genetics* 39, S16-21, doi:10.1038/ng2028
(2007).
- 212 Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer
genomes through second-generation sequencing. *Nature reviews.
Genetics* 11, 685-696, doi:10.1038/nrg2841 (2010).
- 213 Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid
leukaemia genome. *Nature* 456, 66-72, doi:10.1038/nature07485 (2008).
- 214 Ku, C. S., Loy, E. Y., Salim, A., Pawitan, Y. & Chia, K. S. The discovery of
human genetic variations and their use as disease markers: past, present
and future. *Journal of human genetics* 55, 403-415,
doi:10.1038/jhg.2010.55 (2010).
- 215 Moreno-Cabrera, J. M. *et al.* Evaluation of CNV detection tools for NGS
panel data in genetic diagnostics. *European journal of human genetics :
EJHG* 28, 1645-1655, doi:10.1038/s41431-020-0675-z (2020).
- 216 Kerkhof, J. *et al.* Clinical Validation of Copy Number Variant Detection
from Targeted Next-Generation Sequencing Panels. *The Journal of
molecular diagnostics : JMD* 19, 905-920,
doi:10.1016/j.jmoldx.2017.07.004 (2017).

- 217 Zare, F., Dow, M., Monteleone, N., Hosny, A. & Nabavi, S. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. *BMC bioinformatics* 18, 286, doi:10.1186/s12859-017-1705-x (2017).
- 218 Li, W. & Olivier, M. Current analysis platforms and methods for detecting copy number variation. *Physiological genomics* 45, 1-16, doi:10.1152/physiolgenomics.00082.2012 (2013).
- 219 Torri, F. *et al.* Next generation sequence analysis and computational genomics using graphical pipeline workflows. *Genes* 3, 545-575, doi:10.3390/genes3030545 (2012).
- 220 Pös, O. *et al.* Copy Number Variation: Methods and Clinical Applications. *Applied Sciences* 11, 819 (2021).
- 221 Liu, B. *et al.* Computational methods for detecting copy number variations in cancer genome using next generation sequencing: principles and challenges. *Oncotarget* 4, 1868-1881, doi:10.18632/oncotarget.1537 (2013).
- 222 Zhao, L., Liu, H., Yuan, X., Gao, K. & Duan, J. Comparative study of whole exome sequencing-based copy number variation detection tools. *BMC bioinformatics* 21, 97, doi:10.1186/s12859-020-3421-1 (2020).
- 223 Kim, H. Y., Choi, J. W., Lee, J. Y. & Kong, G. Gene-based comparative analysis of tools for estimating copy number alterations using whole-exome sequencing data. *Oncotarget* 8, 27277-27285, doi:10.18632/oncotarget.15932 (2017).
- 224 Pereira, R., Oliveira, J. & Sousa, M. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *Journal of clinical medicine* 9, doi:10.3390/jcm9010132 (2020).
- 225 Faraoni, I. & Graziani, G. Role of BRCA Mutations in Cancer Treatment with Poly(ADP-ribose) Polymerase (PARP) Inhibitors. *Cancers* 10, doi:10.3390/cancers10120487 (2018).
- 226 Hodgson, D. R. *et al.* Candidate biomarkers of PARP inhibitor sensitivity in ovarian cancer beyond the BRCA genes. *British journal of cancer* 119, 1401-1409, doi:10.1038/s41416-018-0274-8 (2018).
- 227 Li, W. *et al.* Germline and somatic mutations of multi-gene panel in Chinese patients with epithelial ovarian cancer: a prospective cohort study. *Journal of ovarian research* 12, 80, doi:10.1186/s13048-019-0560-y (2019).
- 228 Pujade-Lauraine, E. *et al.* Homologous recombination repair mutation gene panels (excluding BRCA) are not predictive of maintenance olaparib plus bevacizumab efficacy in the first-line PAOLA-1/ENGOT-ov25 trial. *Gynecologic oncology* 162, S26-S27 (2021).
- 229 Pellegrino, B., Mateo, J., Serra, V. & Balmaña, J. Controversies in oncology: are genomic tests quantifying homologous recombination repair deficiency (HRD) useful for treatment decision making? *ESMO open* 4 (2019).

- 230 Yoo, B. C., Kim, K. H., Woo, S. M. & Myung, J. K. Clinical multi-omics strategies for the effective cancer management. *Journal of proteomics* 188, 97-106, doi:10.1016/j.jprot.2017.08.010 (2018).
- 231 Despierre, E. *et al.* Somatic copy number alterations predict response to platinum therapy in epithelial ovarian cancer. *Gynecologic oncology* 135, 415-422, doi:10.1016/j.ygyno.2014.09.014 (2014).
- 232 Bogush, T. A. *et al.* The expression and clinical significance of ERbeta/ERalpha in ovarian cancer: can we predict the effectiveness of platinum plus taxane therapy? *Irish journal of medical science* 191, 2047-2053, doi:10.1007/s11845-021-02842-6 (2022).
- 233 Staropoli, N. *et al.* A Prognostic and Carboplatin Response Predictive Model in Ovarian Cancer: A Mono-Institutional Retrospective Study Based on Clinics and Pharmacogenomics. *Biomedicines* 10, doi:10.3390/biomedicines10051210 (2022).
- 234 Syed-Abdul, S., Iqbal, U. & Jack Li, Y. C. Predictive Analytics through Machine Learning in the clinical settings. *Computer methods and programs in biomedicine* 144, A1-A2, doi:10.1016/S0169-2607(17)30552-7 (2017).
- 235 Stevens, L. M., Mortazavi, B. J., Deo, R. C., Curtis, L. & Kao, D. P. Recommendations for Reporting Machine Learning Analyses in Clinical Research. *Circulation. Cardiovascular quality and outcomes* 13, e006556, doi:10.1161/CIRCOUTCOMES.120.006556 (2020).
- 236 Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* 46, 389-422, doi:10.1023/A:1012487302797 (2002).
- 237 Kurşa, M. B. & Rudnicki, W. R. Feature Selection with the Boruta Package. *Journal of Statistical Software* 36, 1 - 13, doi:10.18637/jss.v036.i11 (2010).
- 238 Miller, R. E. *et al.* ESMO recommendations on predictive biomarker testing for homologous recombination deficiency and PARP inhibitor benefit in ovarian cancer. *Annals of oncology : official journal of the European Society for Medical Oncology* 31, 1606-1622, doi:10.1016/j.annonc.2020.08.2102 (2020).
- 239 Pujade-Lauraine, E. *et al.* Homologous recombination repair mutation gene panels (excluding *BRCA*) are not predictive of maintenance olaparib plus bevacizumab efficacy in the first-line PAOLA-1/ENGOT-ov25 trial. *Gynecologic oncology* 162, S26-S27, doi:10.1016/S0090-8258(21)00695-8 (2021).
- 240 Pellegrino, B., Mateo, J., Serra, V. & Balmana, J. Controversies in oncology: are genomic tests quantifying homologous recombination repair deficiency (HRD) useful for treatment decision making? *ESMO open* 4, e000480, doi:10.1136/esmoopen-2018-000480 (2019).
- 241 Kaufman, B. *et al.* Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 33, 244-250, doi:10.1200/JCO.2014.56.2728 (2015).

- 242 Parkes, E. E. & Kennedy, R. D. Clinical Application of Poly(ADP-Ribose)
Polymerase Inhibitors in High-Grade Serous Ovarian Cancer. *The*
oncologist 21, 586-593, doi:10.1634/theoncologist.2015-0438 (2016).
- 243 Frattini, A. *et al.* High variability of genomic instability and gene expression
profiling in different HeLa clones. *Scientific reports* 5, 15377,
doi:10.1038/srep15377 (2015).
- 244 Kurman, R. J., Visvanathan, K. & Shih Ie, M. Bokhman's dualistic model of
endometrial carcinoma. Revisited. *Gynecologic oncology* 129, 271-272,
doi:10.1016/j.ygyno.2013.03.029 (2013).
- 245 Hoang, L. N. *et al.* Histotype-genotype correlation in 36 high-grade
endometrial carcinomas. *The American journal of surgical pathology* 37,
1421-1432, doi:10.1097/PAS.0b013e31828c63ed (2013).
- 246 Gilks, C. B., Oliva, E. & Soslow, R. A. Poor interobserver reproducibility in
the diagnosis of high-grade endometrial carcinoma. *The American journal*
of surgical pathology 37, 874-881, doi:10.1097/PAS.0b013e31827f576a
(2013).
- 247 Han, G. *et al.* Reproducibility of histological cell type in high-grade
endometrial carcinoma. *Modern pathology : an official journal of the*
United States and Canadian Academy of Pathology, Inc 26, 1594-1604,
doi:10.1038/modpathol.2013.102 (2013).
- 248 Guan, H. *et al.* Prognosis and reproducibility of new and existing binary
grading systems for endometrial carcinoma compared to FIGO grading in
hysterectomy specimens. *International journal of gynecological cancer :
official journal of the International Gynecological Cancer Society* 21, 654-
660, doi:10.1097/IGC.0b013e31821454f1 (2011).
- 249 Ritterhouse, L. L. & Howitt, B. E. Molecular Pathology: Predictive,
Prognostic, and Diagnostic Markers in Uterine Tumors. *Surgical pathology*
clinics 9, 405-426, doi:10.1016/j.path.2016.04.006 (2016).
- 250 Soslow, R. A. Endometrial carcinomas with ambiguous features. *Seminars*
in diagnostic pathology 27, 261-273, doi:10.1053/j.semdp.2010.09.003
(2010).
- 251 Fadare, O. *et al.* The diagnosis of endometrial carcinomas with clear cells
by gynecologic pathologists: an assessment of interobserver variability and
associated morphologic features. *The American journal of surgical*
pathology 36, 1107-1118, doi:10.1097/PAS.0b013e31825dd4b3 (2012).
- 252 Clarke, B. A. & Gilks, C. B. Endometrial carcinoma: controversies in
histopathological assessment of grade and tumour cell type. *Journal of*
clinical pathology 63, 410-415, doi:10.1136/jcp.2009.071225 (2010).
- 253 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2017. *CA: a cancer*
journal for clinicians 67, 7-30, doi:10.3322/caac.21387 (2017).
- 254 Phelippeau, J. *et al.* Preoperative diagnosis of tumor grade and type in
endometrial cancer by pipelle sampling and hysteroscopy: Results of a
French study. *Surgical oncology* 25, 370-377,
doi:10.1016/j.suronc.2016.08.004 (2016).

- 255 Fader, A. N., Santin, A. D. & Gehrig, P. A. Early stage uterine serous carcinoma: management updates and genomic advances. *Gynecologic oncology* 129, 244-250, doi:10.1016/j.ygyno.2013.01.004 (2013).
- 256 Murali, R., Soslow, R. A. & Weigelt, B. Classification of endometrial carcinoma: more than two types. *The Lancet. Oncology* 15, e268-278, doi:10.1016/S1470-2045(13)70591-6 (2014).
- 257 Bendifallah, S. *et al.* Just how accurate are the major risk stratification systems for early-stage endometrial cancer? *British journal of cancer* 112, 793-801, doi:10.1038/bjc.2015.35 (2015).
- 258 Bendifallah, S., Darai, E. & Ballester, M. Predictive Modeling: A New Paradigm for Managing Endometrial Cancer. *Annals of surgical oncology* 23, 975-988, doi:10.1245/s10434-015-4924-2 (2016).
- 259 Morice, P., Leary, A., Creutzberg, C., Abu-Rustum, N. & Darai, E. Endometrial cancer. *Lancet* 387, 1094-1108, doi:10.1016/S0140-6736(15)00130-0 (2016).
- 260 Talhouk, A. *et al.* Molecular classification of endometrial carcinoma on diagnostic specimens is highly concordant with final hysterectomy: Earlier prognostic information to guide treatment. *Gynecologic oncology* 143, 46-53, doi:10.1016/j.ygyno.2016.07.090 (2016).
- 261 Yen, T. T., Wang, T. L., Fader, A. N., Shih, I. M. & Gaillard, S. Molecular Classification and Emerging Targeted Therapy in Endometrial Cancer. *International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists* 39, 26-35, doi:10.1097/PGP.0000000000000585 (2020).
- 262 Lopez-Reig, R. *et al.* Prognostic classification of endometrial cancer using a molecular approach based on a twelve-gene NGS panel. *Scientific reports* 9, 18093, doi:10.1038/s41598-019-54624-x (2019).
- 263 Berns, E. M. & Bowtell, D. D. The changing view of high-grade serous ovarian cancer. *Cancer research* 72, 2701-2704, doi:10.1158/0008-5472.CAN-11-3911 (2012).
- 264 Vaughan, S. *et al.* Rethinking ovarian cancer: recommendations for improving outcomes. *Nature reviews. Cancer* 11, 719-725, doi:10.1038/nrc3144 (2011).
- 265 How, J. A. *et al.* Modification of Homologous Recombination Deficiency Score Threshold and Association with Long-Term Survival in Epithelial Ovarian Cancer. *Cancers* 13, doi:10.3390/cancers13050946 (2021).
- 266 Valabrega, G., Scotto, G., Tuninetti, V., Pani, A. & Scaglione, F. Differences in PARP Inhibitors for the Treatment of Ovarian Cancer: Mechanisms of Action, Pharmacology, Safety, and Efficacy. *International journal of molecular sciences* 22, doi:10.3390/ijms22084203 (2021).
- 267 Sumantran, V. N. Cellular chemosensitivity assays: an overview. *Methods in molecular biology* 731, 219-236, doi:10.1007/978-1-61779-080-5_19 (2011).

268 McDermott, M. *et al.* In vitro Development of Chemotherapy and Targeted Therapy Drug-Resistant Cancer Cell Lines: A Practical Guide with Case Studies. *Frontiers in oncology* 4, 40, doi:10.3389/fonc.2014.00040 (2014).

ANNEXES

Annexes

Annexed data

Annex 1: Detailed inclusion and exclusion criteria for POLA Phase I/IIB clinical trial.

Patients aged ≥ 18 years were eligible if they had histologically confirmed advanced or metastatic high-grade serous or endometrioid (no mucinous nor clear cell) platinum resistant not refractory (not primary nor secondary refractory) ovarian, fallopian or primary peritoneal cancer, EC (any grade, not platinum refractory) or triple negative breast cancer; Eastern Cooperative Group (ECOG) performance status ≤ 2 ; life expectancy of ≥ 3 months; measurable disease per Response Evaluation Criteria in Solid Tumors (RECIST) criteria version 1.1; have receive at least one line of standard therapy for locally advanced or metastatic disease and developed progression disease afterwards; hemoglobin ≥ 10 g/dL; absolute neutrophil count $\geq 1,500/\mu\text{L}$; platelets $\geq 100,000/\mu$; total bilirubin ≤ 1.5 times the institutional upper limit of normal (ULN); aspartate aminotransferase and alanine aminotransferase ≤ 2.5 times ULN; albumin ≥ 3 g/dL; creatinine ≤ 1.5 times the ULN or a creatinine clearance ≥ 30 mL/min. Patients were illegible if they had received previous treatment with a PARP inhibitor or lurbinectedin; received chemotherapy or radiotherapy within 2 weeks of the study entry; required treatment with inhibitors or inducers of CYP3A4; another malignancy within the past 5 years; know active severe disease or

immunosuppression; persistent grade ≥ 2 toxicities (Common Terminology Criteria for Adverse Events version XX NCI-CTCAE v. 4.03) caused by previous treatment (excluding alopecia); resting ECG with QTc > 470 msec on 2 or more time point within a 24-hour period; existing gastrointestinal disorders that could interfere with the absorption of the study medication or had condition that preclude to swallow oral medication; baseline features suggestive of acute myelogenous leukemia or myelodysplastic syndrome.

Annex 2: Clinical and molecular characteristics from OC cell line panel. Available information of clinicopathological, treatment and molecular features.

Cell line	Tissue	Site	Status	Hormone Receptor	Previous Treatment	Diagnosis	Reported alterations
A2780	T	PT		-	Untreated	E	ATM p.Pro604Ser PTEN p.Lys128_Arg130del
A2780CIS	T	PT	GR	-	Chronic exposure to cisplatin	E	ATM p.Pro604Ser PTEN p.Lys128_Arg130del
A2780PM	T	PT	GR	-	Chronic exposure to PM	E	-
PEO1	A	MS		ER+	5-fluorouracil and chlorambucil	C	TP53 p.Gly244Asp BRCA2 p.Tyr1655Ter
PEO1OLA	A	MS	GR	-	-	C	-
PEO4	A	MS	CR	ER+	5-fluorouracil and chlorambucil	C	TP53 p.Gly244Asp BRCA2 p.Tyr1655Tyr
PEO6	A	MS	A, TS	ER+	Cisplatin, 5-fluorouracil and chlorambucil	C	TP53 p.Gly244Asp BRCA2 p.Tyr1655Tyr
TOV112	-	-		-	-	E	TP53 p.Arg175His
OVCAR-3	-	-		-	-	A	TP53 p.Arg248Gln
SKOV-3	A	-	PS	-	-	A	APC p.Thr1556fs*9 FBXW7 p.Arg505Leu PIK3CA p.His1047Arg TP53 p.Ser90fs*33
SCCOHT1	-	-		-		SCCOT1	-

Cell line	Tissue	Site	Status	Hormone Receptor	Previous Treatment	Diagnosis	Reported alterations
BIN67	P	MS		-	-	SCCHT1	SMARCA4 c.2438+1G>A
PEA1	PE	MS		ER+	Untreated	C	BARD1 p.Arg658Cys TP53 p.Cys242Alafs*5.
59M	A	MS		-	-	HGSOC	-
TO14	O	MS		ER-		C	TP53 p.Cys277Phe
PEO14	A	MS	RL	ER-	Cisplatin and chlorambucil	C	TP53 p.Cys277Phe
PEO23	A	MS		ER-	Untreated	C	TP53 p.Cys277Phe
PEA2	PE	MS	RL	-	cisplatin and prednimustine	C	BARD1 p.Arg658Cys TP53 p.Cys242Alafs*5

* All cell lines presented Ovarian Origin from Female patients, epithelial morphology and adherent growing. Site: PT, Primary tumor r MS, metastatic site. Tissue: T, Tumor, A, ascitis, P, pelvis, O, omentum or PE, Pleural effusion. Status: GR, Generated resistance, CR, Clinical resistance, A, Advanced stage, TS, Terminal stage, PS, Primary tumor and RL, relapse. Hormone receptor: ER+, Oestrogen positive and ER-, Oestrogen negative. Diagnosis: E, endometrioid, C, cystadenocarcinoma, SCCHT1, small-cell carcinoma hypercalcemic Type 1, A, adenocarcinoma and HGS, High-grade serous.. All this information has been obtain from different databases and bibliograpy: ATCC, ECCAC, Cellosaurus, 10.1371/journal.pone.0122284 and 10.1038/ncomms8419.

Annex 3: Profile of STR markers (Identiflier®). Characterization of OC cell line panel at STR-markers level to ensure the fidelity regarding the public profiles.

Marker	CSF1PO	D2S1338	D3S1358	D5S818	D7S820	D8S1179	D13S317	D16S539	D18S51	D19S433	D21S11	FGA	TH01	TPOX	vWA	% of coincidence with public data bases
OVCAR3	11,12	17,21	17,18	11,12	10	10,15	12	12	13	16,2	29,31.2	21	9,9.3	8	17	100
PEO7	10,12	20,21	16	11,12	10	13,14	10	9	16,17	13,15	30,32.2	20	9.3	9,11	15,16	98
A2780CIS	10,11	21,22	13,16	11	10	15,18	13	11,13	16,17	12	28	19,24	6	8,10	15,16	100
PEO1	10,12	20,21	16	11,12	10	13,14	10	9	16,17	13,15	32.2	20	9.3	9,11	15,16	100
PEO4	10,12	20,21	16	11,12	10	13,14	10	9	16,17	13,15	30,32.2	20	9.3	9,11	15,16	100
SKOV4	11	18,23	14	11	13,14	14,15	8,11	12	16,17	14,14.2	30,31.2	23,24,25	9,9.3	8,11	18	100
TOV112D	12	19,24	15	10	9,10	9,13	8	11,12	17	14	31	20	6	8,11	18	100
A2780	10,11	21,22	14,16	11,12	10	15,17	12,13	11,12,13	16,18	12	28	19,24	6	8,10	15,16	100
PEA1	14	19,25	16,17	11	10	14	8	12,13	14,17	12,13	30	22,25	9	8,12	18	100
PEA2	11,14	19	16	11,13	10,12	14	8	12,13	17	12,13	30	25	9	8,12	17,18	100
PEO14	11	19	16	12	9,10	13,14	11,12	11	15	14,15	27,30	20	6	8,11	15,16	97
PEO23	10,11	20,23	13,16	11	10	15,18	13	11,13	16,17	12	28	18,24	6	8	15,16	90
TO14	10,11	20,23	13,16	11	10	15,18	13	11,13	16,17	12	28	19,24	6	8	15,16	94
59M	13	19,20	15,17	11	9,10	10,14	11	9	12	14,16	31.35.2	23,30	7	8	17,19	89
BIN67	10,12	17,25	15,19	11,12	9	14	10,12	12	13,19	14,15.2	30	22,2	6,9.3	11	14,18	94
SCCOHT-1	10,11	17	15,17	11,12	10,11	13,14	8,12	9,13	13	15.2,16	28,31.2	25,26	7,8	8,9	14,16	Not previously described
A2780PM	10,17	21,22	14,16	11,12	10	15,17	12,13	11,13	16,18	12	28	19,25	6	8,10	15,16	93
PEO1OLA	10,12	20,21	16	11,12	10	13,14	10	9	16,17	13,15	32.2	20	9.3	9,11	15,16	100

*All cell lines were X for Amelogenin marker.

Annex 4: Sequencing metrics cohort 1

Sample	Total aligned Reads	Total aligned Reads (%)	Targeted Aligned reads	Read Enrichment (%)	Median Region Coverage Depth	Target Coverage 100X (%)
12482	22277084	7.34	3803089	68.51	257	74
13742	98968514	32.60	4444242	21.61	333	83
15092	18214140	6.00	5176898	73.48	232	74
15296	24190126	7.97	4305008	65.15	251	75
15403	25076454	8.26	4701465	73.31	337	81
16732	11853728	3.90	2121826	62.80	127	59
16733	13741696	4.53	5545198	61.94	220	72
16734	19793354	6.52	2208859	62.61	209	70
16735	10246434	3.38	3766684	66.33	88	45
17561	13741696	4.53	5883440	61.94	127	58
18389	23076694	7.60	5410662	64.82	205	69
18720	22388788	7.38	1712684	46.26	145	60
13726	23701142	2.91	924006	62.76	236	77
13727	16690956	2.05	2254052	44.32	104	51
13728	23425194	2.88	4195091	46.49	173	69
13729	19406064	2.39	5211536	55.82	172	68
13730	22081854	2.72	5140839	60.65	226	77
13732	20212616	2.49	4385546	56.35	181	70
13734	22565478	2.77	1977853	44.36	164	68

Sample	Total aligned Reads	Total aligned Reads (%)	Targeted Aligned reads	Read Enrichment (%)	Median Region Coverage Depth	Target Coverage 100X (%)
13735	16817662	2.07	5577291	65.64	187	73
13736	27440478	3.37	3690801	61.29	249	77
13738	23831292	2.93	3792185	64.49	225	74
13740	17606562	2.16	3308214	45.55	108	53
15948	21679520	2.67	4407397	61.93	212	74
15949	22644328	2.78	318424	61.51	234	77
15950	20476602	2.52	4738696	61.94	199	74
15952	74933690	9.21	3779027	31.24	413	84
15953	9771584	1.20	700381	35.09	49	22
15954	10723472	1.32	805311	53.99	98	49
15955	19892626	2.45	2808450	62.05	208	74
15956	20942190	2.57	4739107	54.16	180	69
15957	28262764	3.48	496183	59.39	301	82
15959	70256830	8.64	433802	64.56	775	86
15961	24399080	3.00	3848533	57.75	236	77
15962	22245538	2.74	4330348	66.31	235	78
15963	54956126	6.76	669822	64.75	598	86
15964	10915424	1.34	5049329	55.98	100	50
15965	33281256	4.09	4050099	57.72	332	83
15966	16207554	1.99	4723055	63.91	179	75
15967	37139140	4.57	5603188	70.46	448	85

Sample	Total aligned Reads	Total aligned Reads (%)	Targeted Aligned reads	Read Enrichment (%)	Median Region Coverage Depth	Target Coverage 100X (%)
15969	20633066	2.54	5492957	59.06	210	74
15970	8652210	1.06	748620	61.01	101	50
15974	32708974	4.02	2594118	67.06	384	84
15977	18796972	2.31	1776993	48.90	151	65

Annex 5: Variant information cohort 1

Sample	Total aligned Reads	Total aligned Reads (%)	Targeted Aligned reads	Read Enrichment (%)	Median Region Coverage Depth	Target Coverage 100X (%)
12482	22277084	7.34	3803089	68.51	257	74
13742	98968514	32.60	4444242	21.61	333	83
15092	18214140	6.00	5176898	73.48	232	74
15296	24190126	7.97	4305008	65.15	251	75
15403	25076454	8.26	4701465	73.31	337	81
16732	11853728	3.90	2121826	62.80	127	59
16733	13741696	4.53	5545198	61.94	220	72
16734	19793354	6.52	2208859	62.61	209	70
16735	10246434	3.38	3766684	66.33	88	45
17561	13741696	4.53	5883440	61.94	127	58
18389	23076694	7.60	5410662	64.82	205	69
18720	22388788	7.38	1712684	46.26	145	60
13726	23701142	2.91	924006	62.76	236	77
13727	16690956	2.05	2254052	44.32	104	51
13728	23425194	2.88	4195091	46.49	173	69
13729	19406064	2.39	5211536	55.82	172	68
13730	22081854	2.72	5140839	60.65	226	77
13732	20212616	2.49	4385546	56.35	181	70
13734	22565478	2.77	1977853	44.36	164	68
13735	16817662	2.07	5577291	65.64	187	73
13736	27440478	3.37	3690801	61.29	249	77
13738	23831292	2.93	3792185	64.49	225	74

Sample	Total aligned Reads	Total aligned Reads (%)	Targeted Aligned reads	Read Enrichment (%)	Median Region Coverage Depth	Target Coverage 100X (%)
13740	17606562	2.16	3308214	45.55	108	53
15948	21679520	2.67	4407397	61.93	212	74
15949	22644328	2.78	318424	61.51	234	77
15950	20476602	2.52	4738696	61.94	199	74
15952	74933690	9.21	3779027	31.24	413	84
15953	9771584	1.20	700381	35.09	49	22
15954	10723472	1.32	805311	53.99	98	49
15955	19892626	2.45	2808450	62.05	208	74
15956	20942190	2.57	4739107	54.16	180	69
15957	28262764	3.48	496183	59.39	301	82
15959	70256830	8.64	433802	64.56	775	86
15961	24399080	3.00	3848533	57.75	236	77
15962	22245538	2.74	4330348	66.31	235	78
15963	54956126	6.76	669822	64.75	598	86
15964	10915424	1.34	5049329	55.98	100	50
15965	33281256	4.09	4050099	57.72	332	83
15966	16207554	1.99	4723055	63.91	179	75
15967	37139140	4.57	5603188	70.46	448	85
15969	20633066	2.54	5492957	59.06	210	74
15970	8652210	1.06	748620	61.01	101	50
15974	32708974	4.02	2594118	67.06	384	84
15977	18796972	2.31	1776993	48.90	151	65

Annex 6: Germline BRCA mutations cohort 1. Details of alterations at germline level regarding *BRCA1/2* gene.

Sample	Gene	Alteration
13726	BRCA1	c.3264dupT p.Gln1089Serfs*10
13727	BRCA1	c.213-12A>G
13728	BRCA1	c.3481_3491del11 p.Glu1161Phefs* and c.3264depT p.Gln1089Serfs*10
13729	BRCA1	Deletion exons8-13
13730	BRCA1	c.4307_4308delCT p.Ser1436Phefs*4
13732	BRCA1	c.4287CA> p.Tyr1429*
13734	BRCA1	c.1961delA p.Lys654Serfs*47
13735	wild type	No informative
13736	wild type	No informative
13738	BRCA1	c.5095C>T (p.Arg1699Trp)
13740	wild type	No informative
13742	BRCA1	c.3770_3771delAG (p.Glu1257Glyfs)
15948	BRCA1	c.68_69delAG p.Glu23Valfs*17
15949	wild type	No informative
15950	wild type	No informative
15952	wild type	No informative
15953	BRCA2	No informative
15954	wild type	No informative
15955	wild type	No informative
15956	wild type	No informative
15957	wild type	No informative
15959	wild type	No informative
15961	wild type	No informative
15962	BRCA2	c.2246_2261delGTGATACTGACTTTCA p.Ser749Asnfs*18
15963	wild type	No informative
15964	wild type	No informative
15965	wild type	No informative
15966	wild type	No informative
15967	wild type	No informative
15969	BRCA2	c.6024dupG p.Gln2009Alafs*9
15970	BRCA2	c.7435+1G>A 14%
15974	BRCA2	c.2808_2811delACAA p.Ala938Profs*21
15977	wild type	No informative

Annex 7: Comparison of performances between CNV pipelines. Non-parametric tests to decipher strengths and weakness of applied pipelines.

Variable	Group	CNVKit			saasCNV			SureCall		
		N	Median [range]	p	N	Median [range]	p	N	Median [range]	p
Total number of events	<median	20	38.5 [3-168]	0.002	21	20 [4-49]	0.032	19	147 [68-300]	0.011
	>median	21	64 [32-102]		21	32 [0-53]		20	204 [146-240]	
Percentage of altered genome	<median	21	19.26 [3.8-28.54]	<0.001	21	51.52 [3.44-89.71]	1.000	19	41.33 [15.39-86.65]	0.001
	>median	20	28.61 [0.98-36.16]		21	46.3 [0-90.75]		20	76.355 [38.71-100]	
Total number of gain events	<median	21	8.5 [0-57]	0.044	21	0 [0-21]	0.900	19	25 [1-80]	0.007
	>median	20	15 [2-49]		21	0 [0-35]		20	51 [9-84]	
Percentage of genome altered by gains	<median	21	3.445 [0-11.56]	0.025	21	0 [0-66.5]	0.750	19	6.48 [0.093-27.42]	0.008
	>median	20	5.92 [0.77-13.12]		21	0 [0-90.75]		20	17.28 [3.44-27.7]	
Total number of loss events	<median	21	11 [0-87]	0.087	21	9 [0-49]	0.190	19	26 [1-118]	0.022
	>median	20	16 [4-42]		21	14 [0-28]		20	66.5 [17-104]	

Variable	Group	CNVKit			saasCNV			SureCall		
		N	Median [range]	p	N	Median [range]	p	N	Median [range]	p
Percentage of genome altered by losses	<median	21	2.93 [0-10.1]	0.083	21	16.59 [0-89.71]	0.950	19	12.14 [0.034-56.98]	0.009
	>median	20	4.05 [1.39-7.95]		21	21.72 [0-45.89]		20	30.05 [4.53-72.76]	
Total number of LOH events	<median	21	18 [2-49]	<0.001	21	7 [0-19]	0.008	19	82 [10-193]	0.899
	>median	20	26 [12-42]		21	13 [0-29]		20	80 [39-198]	
Percentage of genome altered by LOHs	<median	21	12.265 [2.67-24.01]	0.009	21	13.29 [0-34.9]	0.240	19	20.68 [6.87-38.63]	0.180
	>median	20	16.94 [0.77-25.1]		21	22.03 [0-39.62]		20	28.845 [13.91-39.4]	
Total number of LOH>15 mb events	<median	21	7 [0-13]	0.009	21	5 [0-16]	0.060	19	6 [0-46]	0.061
	>median	20	11 [3-20]		21	11 [0-23]		20	12.5 [0-25]	
Percentage of genome altered by LOHs>15 mb	<median	21	7.255 [0-24.01]	0.042	21	13 [0-34.8]	0.540	19	6.73 [0-23.71]	0.026
	>median	20	13.57 [4.53-22.08]		21	20.91 [0-39.23]		20	9.98 [0-29.29]	

Annex 8: Correlation between HRD-based classification and different GI parameters. Non-parametric tests to evaluate the grade of concordance between the presence of HRR-genes mutations and GI.

Variable	Group	N	Median [range]	<i>p-value</i>
Total number of events	nonHRD	15	39 [3-136]	0.068
	HRD noNBRCAs	7	68 [32-91]	
	HRD BRCA	19	55 [17-168]	
Percentage of altered genome	nonHRD	15	17.09 [3.8-32.25]	0.006
	HRD noNBRCAs	7	28.82 [16.9-36.16]	
	HRD BRCA	19	27.14 [13.77-34.71]	
Total number of gain events	nonHRD	15	8 [0-36]	0.120
	HRD noNBRCAs	7	25 [2-49]	
	HRD BRCA	19	13 [3-57]	
Percentage of genome altered by gains	nonHRD	15	1.86 [0-10.59]	0.013
	HRD noNBRCAs	7	8.52 [0.77-12.55]	
	HRD BRCA	19	5.06 [1.2-13.12]	
Total number of loss events	nonHRD	15	13 [0-52]	0.870
	HRD noNBRCAs	7	16 [4-34]	
	HRD BRCA	19	15 [0-87]	
Percentage of genome altered by losses	nonHRD	15	2.95 [0-7.62]	0.570
	HRD noNBRCAs	7	4.2 [1.39-7.14]	
	HRD BRCA	19	3.69 [0-10.11]	
Total number of LOH events	nonHRD	15	18 [2-49]	0.260
	HRD noNBRCAs	7	21 [12-42]	
	HRD BRCA	19	25 [12-30]	
Percentage of genome altered by LOHs	nonHRD	15	10.11 [2.67-24.01]	0.050
	HRD noNBRCAs	7	14.92 [9.48-25.1]	
	HRD BRCA	19	16.94 [3.86-23.61]	
Total number of LOH>15 mb events	nonHRD	15	6 [0-14]	0.035
	HRD noNBRCAs	7	8 [5-20]	
	HRD BRCA	19	11 [1-16]	
Percentage of genome altered by LOHs>15 mb	nonHRD	15	6.35 [0-24.01]	0.037
	HRD noNBRCAs	7	11.22 [4.92-22.08]	
	HRD BRCA	19	13.86 [0.49-21.72]	

Annex 9: CNVkit performance. Ability of different pre-defined GI parameters to predict clinical outcome.

Variable	Group	N	Median (95% CI)	<i>p</i>value
Total number of events	<median	19	11.67 (±0.411)	0.015
	>median	22	20.60 (±10.42)	
Percentage of altered genome	<median	20	11.667 (±0.149)	0.008
	>median	21	33.6 (±11.850)	
Total number of gain events	<median	19	13.63 (±1.57)	0.088
	>median	22	20.60 (±9.17)	
Percentage of genome altered by gains	<median	20	11.77 (±2.2)	0.028
	>median	21	33.6 (±13.554)	
Total number of loss events	<median	20	13.63 (±2.42)	0.620
	>median	21	18.37 (±1.70)	
Percentage of genome altered by losses	<median	20	11.77 (±2.236)	0.210
	>median	21	18.37 (±2.009)	
Total number of LOH events	<median	18	11.77 (±2.09)	0.008
	>median	23	33.57 (±10.70)	
Percentage of genome altered by LOHs	<median	20	13.63 (±2.273)	0.057
	>median	21	33.567 (±11.621)	
Total number of LOH>15 mb events	<median	20	14.267 (±0.671)	0.280
	>median	21	20.6 (±11.901)	
Percentage of genome altered by LOHs>15 mb	<median	20	13.80 (±2.19)	0.110
	>median	21	20.60 (±11.83)	

Annex 10: Parameter feeding SNP-Model.

SNP	Chr	Position (GRCh38)	Alleles	Consequence	Clinical significance
rs876261	8	95230391	C>G	None	Not Reported in ClinVar
rs142099227	X	102701669	C>A	ARMCX5-GPRASP2 : Intron Variant	Not Reported in ClinVar
rs13135475	4	39498624	T>C T>G	UGDH : 500B Downstream Variant	Not Reported in ClinVar
rs56747986	13	97118261	T>A	None	Not Reported in ClinVar
rs540649069	4	97223716	T>G	STPG2 : Intron Variant	Not Reported in ClinVar
rs761256207	X	70504827	G>A	DLG3 : 3 Prime UTR Variant	Not Reported in ClinVar
rs13401599	2	834796	A>G A>T	LINC01115 : Intron Variant	Not Reported in ClinVar
rs562439697	3	14679706	G>A G>T	C3orf20 : Intron Variant	Not Reported in ClinVar

Annex 11: Parameter feeding HTG-Model.

Gene	Location	Cytogenetic band	Size (bp)	Function
RUVBL1	chr3:128,064,611-128,153,914	3q21.3	89,304	Activity ATPase associated with diverse cellular activities.
ADORA2A	chr22:24,417,879-24,442,357	22q11.23	24,479	Guanine nucleotide-binding protein (G protein)-coupled receptor (GPCR).
ABCD4	chr14:74,285,269-74,303,062	14q24.3	17,794	ATP-binding cassette (ABC) transporters. ABC proteins transport various molecules across extra- and intra-cellular membranes.
PVR	chr19:44,643,798-44,666,162	19q13.31	22,365	Transmembrane glycoprotein belonging to the immunoglobulin superfamily.
PFDN2	chr1:161,100,556-161,118,055	1q23.3	17,500	The encoded protein is one of six subunits of prefoldin, which is a chaperone protein that binds and stabilizes newly synthesized polypeptides.
SIL1	chr5:138,946,724-139,293,557	5q31.2	346,834	This gene encodes a N-linked glycoprotein with an N-terminal ER targeting sequence, 2 putative N-glycosylation sites, and a C-terminal ER retention signal.

Annex 12: Sequencing metrics of Cohort 4.

Sample	Total aligned Reads	Total aligned Reads (%)	Targeted Aligned reads	Read Enrichment (%)	Median Region Coverage Depth	Target Coverage 100X	Target Coverage 700X	SNVs Reported
10031	4598467	2.08	3803089	82.70	11543	100	99.17	40
10293	1168220	0.53	924006	79.10	2963	99.17	92.95	21
11218	3388612	1.53	2254052	66.52	6815	100	98.76	37
11282	5124886	2.31	4195091	81.86	13150	99.59	99.17	44
11525	6169307	2.79	5211536	84.48	15373	100	99.17	45
12395	6105083	2.76	5140839	84.21	16774	99.59	98.76	35
12447	5392192	2.43	4385546	81.33	12699	100	99.17	34
12514	4677751	2.11	1977853	42.28	5491	99.59	95.02	41
13033	6932913	3.13	5577291	80.45	16556	100	99.17	37
13324	4712742	2.13	3690801	78.32	10862	100	99.17	40
13618	4891463	2.21	3792185	77.53	11203	99.59	99.17	36
13652	4133615	1.87	3308214	80.03	10252	98.76	98.34	130
13841	5667250	2.56	4444242	78.42	13186	100	99.59	39
14032	6078507	2.74	5176898	85.17	15698	100	99.59	34
14039	5246469	2.37	4305008	82.06	13275	100	99.17	36
14061	5688289	2.57	4701465	82.65	13829	100	99.59	36
14078	5276367	2.38	4407397	83.53	13611	100	99.17	39
14202	382024	0.17	318424	83.35	928	92.12	62.66	13
14206	5433163	2.45	4738696	87.22	13188	100	99.17	26
14257	4573663	2.07	3779027	82.63	11068	100	99.17	77

Sample	Total aligned Reads	Total aligned Reads (%)	Targeted Aligned reads	Read Enrichment (%)	Median Region Coverage Depth	Target Coverage 100X	Target Coverage 700X	SNVs Reported
14317	831817	0.38	700381	84.20	2183	98.76	89.21	18
14537	938562	0.42	805311	85.80	2230	98.76	89.21	25
14575	4629694	2.09	2808450	60.66	8468	99.59	97.51	78
14623	5870553	2.65	4739107	80.73	14109	100	99.17	35
14890	939669	0.42	496183	52.80	1530	99.17	83.82	19
15447	671034	0.30	433802	64.65	1336	98.34	71.78	29
15636	5503367	2.49	3848533	69.93	12241	99.59	99.17	39
15815	5175822	2.34	4330348	83.66	12784	100	99.17	40
16265	827609	0.37	669822	80.93	2038	99.17	88.38	19
17842	5998116	2.71	5049329	84.18	12945	100	99.17	22
18217	4566355	2.06	4050099	88.69	12266	100	98.76	31
18644	5971319	2.70	4723055	79.10	14436	100	99.17	49
3252	7365762	2.24	5603188	76.07	750	98.14	25.65	13
3075	6531400	2.95	5492957	84.10	17188	100	99.17	49
3474	1024933	0.46	748620	73.04	2358	99.17	92.12	19
4072	3284082	1.48	2594118	78.99	7495	100	98.76	32
4273	2228365	1.01	1776993	79.74	5428	100	98.76	33
4712	3497351	1.58	2121826	60.67	6253	100	98.76	33
5244	6765709	3.06	5545198	81.96	16523	100	99.59	47
5653	3137473	1.42	2208859	70.40	6543	100	98.76	43
6202	4646968	2.10	3766684	81.06	11843	99.17	97.93	35
6272	7777797	3.51	5883440	75.64	18029	100	99.17	42

Sample	Total aligned Reads	Total aligned Reads (%)	Targeted Aligned reads	Read Enrichment (%)	Median Region Coverage Depth	Target Coverage 100X	Target Coverage 700X	SNVs Reported
6691	6674023	3.01	5410662	81.07	16098	100	99.17	47
7824	2065368	0.93	1712684	82.92	5001	100	98.34	25
7926	3038922	1.37	2406975	79.20	7101	100	98.76	30
8774	4920697	2.22	3498165	71.09	10606	99.59	99.17	102
9352	5042059	2.28	4183288	82.97	13182	100	99.17	46
9692	6042409	2.73	4990079	82.58	15550	100	99.17	35
9831	2851563	1.29	2187688	76.72	6549	100	98.76	32
10402	4304993	1.55	3601452	83.66	9771	99.58	99.17	24
11535	5688265	2.05	2001592	35.19	5703	99.58	98.33	29
12670	5384295	1.94	4329967	80.42	13226	99.58	99.17	89
13169	4837981	1.74	3977094	82.21	11635	100	98.75	94
13539	8732122	3.46	6457114	73.95	864	97.76	41.26	41
13188	4788291	1.72	3955243	82.60	12224	98.75	97.92	34
13672	7338031	2.64	6218334	84.74	19146	100	99.58	52
14208	2375408	0.86	1737958	73.16	5170	96.25	89.17	27
14262	5956424	2.15	3548475	59.57	8972	99.58	98.33	30
14406	5313507	1.91	4329544	81.48	13367	99.58	99.17	40
14953	3126311	1.13	2474465	79.15	7717	99.17	98.75	41
14998	2832890	1.02	2298279	81.13	6895	97.92	90.83	19
15213	2681321	0.97	2349804	87.64	7326	96.25	91.25	33
16300	4449066	1.60	3463807	77.85	10441	100	99.58	29
17305	5408446	1.95	4650477	85.99	14243	99.17	99.17	86

Sample	Total aligned Reads	Total aligned Reads (%)	Targeted Aligned reads	Read Enrichment (%)	Median Region Coverage Depth	Target Coverage 100X	Target Coverage 700X	SNVs Reported
17364	5443701	1.96	3822513	70.22	11902	100	99.17	31
17654	5714914	2.06	4455107	77.96	13457	100	99.58	35
17836	5480066	1.97	4318424	78.80	13011	100	98.75	33
18715	5702422	2.05	4608851	80.82	14444	99.58	99.17	33
19166	3683450	1.33	2906697	78.91	8796	99.58	98.75	29
19173	4852416	1.75	3843651	79.21	11718	99.58	98.75	37
19191	6420569	2.31	5541841	86.31	16681	99.58	99.17	41
2797	7262801	2.62	5313960	73.17	16603	100	99.17	138
2924	7772194	2.80	6422778	82.64	19198	100	99.17	79
2932	7867688	2.83	5105399	64.89	15641	100	99.17	97
3104	3853896	1.39	2827460	73.37	8147	99.17	97.50	146
4373	3802540	1.37	1998333	52.55	6116	99.58	97.50	120
4483	6960219	2.51	5707522	82.00	17258	99.58	99.17	88
4574	3605445	1.30	2562902	71.08	7486	99.58	98.33	134
4598	6383926	2.30	4203209	65.84	12242	99.58	98.75	171
5491	8453975	3.05	6558101	77.57	19316	100	99.58	100
5799	6233467	2.25	5275053	84.62	15725	100	99.17	56
5991	4044883	1.46	2527164	62.48	7544	99.58	97.08	132
6001	8098650	2.92	6158951	76.05	19473	100	99.58	96
6020	7409096	2.67	4090395	55.21	11747	100	99.17	138
6388	11357066	3.46	8406022	74.02	1125	99.63	52.04	20
6186	5718245	2.06	4912155	85.90	10738	100	95.83	89

Sample	Total aligned Reads	Total aligned Reads (%)	Targeted Aligned reads	Read Enrichment (%)	Median Region Coverage Depth	Target Coverage 100X	Target Coverage 700X	SNVs Reported
6667	7159257	2.58	5919789	82.69	16353	99.58	99.17	57
7004	7274183	2.62	5537072	76.12	16717	99.58	99.58	109
7153	9582534	2.92	9332364	97.39	1249	99.63	68.7	27
7443	6187109	2.23	4525018	73.14	12386	99.17	97.92	93
7619	8845944	3.19	7188904	81.27	22438	100	99.17	43
8081	7637004	2.75	5163648	67.61	15902	99.58	99.17	85
8721	6094668	2.20	3829490	62.83	11474	99.58	99.17	65
9417	2107803	0.76	1725302	81.85	4594	98.33	87.50	31
9558	9020276	3.25	7193906	79.75	21643	100	99.17	64
9935	6572137	2.37	5621340	85.53	17290	99.58	98.75	43

Annex 13: Distribution of variants by annotation in Cohort 4.

Sample	Pathogenic	Likely Pathogenic	VUS	Likely Benign	Benign	Other	Total filtered Variants	Total Unfiltered Variants	% Variants Passing Filters
10031	7	2	3	4	0	24	12	40	30.00
10293	6	1	0	0	0	14	7	21	33.33
11218	2	3	3	6	0	23	8	37	21.62
11282	6	0	6	4	0	28	12	44	27.27
11525	1	1	7	5	1	31	9	45	20.00
12395	1	0	4	0	2	30	5	35	14.29
12447	2	1	7	3	0	21	10	34	29.41
12514	1	3	2	6	1	29	6	41	14.63
13033	2	1	4	4	0	26	7	37	18.92
13324	6	6	5	4	2	19	17	40	42.50
13618	6	2	4	4	1	20	12	36	33.33
13652	15	16	12	17	1	70	43	130	33.08
13841	2	0	7	6	1	24	9	39	23.08
14032	1	2	7	4	0	20	10	34	29.41
14039	0	1	5	6	0	24	6	36	16.67
14061	4	0	8	4	0	20	12	36	33.33
14078	1	3	4	4	1	27	8	39	20.51
14202	2	1	0	0	0	10	3	13	23.08
14206	0	3	2	5	1	16	5	26	19.23
14257	8	8	1	0	0	60	17	77	22.08

Sample	Pathogenic	Likely Pathogenic	VUS	Likely Benign	Benign	Other	Total filtered Variants	Total Unfiltered Variants	% Variants Passing Filters
14317	2	0	0	1	0	15	2	18	11.11
14537	1	0	1	2	1	21	2	25	8.00
14575	9	18	9	10	15	32	36	78	46.15
14623	3	1	3	5	1	23	7	35	20.00
14890	0	1	2	4	2	12	3	19	15.79
15447	4	3	3	9	0	10	10	29	34.48
15636	5	2	4	2	1	26	11	39	28.21
15815	2	1	4	5	1	28	7	40	17.50
16265	2	1	0	3	1	13	3	19	15.79
17842	2	1	5	3	1	13	8	24	33.33
18217	1	0	2	4	0	24	3	31	9.68
18644	5	4	7	4	1	29	16	49	32.65
3252	1	1	5	1	1	5	7	13	53.85
3075	3	6	8	4	0	28	17	49	34.69
3474	3	0	0	0	0	16	3	19	15.79
4072	0	3	3	4	0	22	6	32	18.75
4273	1	2	3	3	0	24	6	33	18.18
4712	2	1	3	4	0	23	6	33	18.18
5244	3	4	7	6	0	27	14	47	29.79
5653	3	3	5	7	0	25	11	43	25.58
6202	3	2	3	2	0	25	8	35	22.86
6272	1	0	7	4	0	30	8	42	19.05
6691	6	2	4	4	0	31	12	47	25.53

Sample	Pathogenic	Likely Pathogenic	VUS	Likely Benign	Benign	Other	Total filtered Variants	Total Unfiltered Variants	% Variants Passing Filters
7824	0	0	3	3	0	19	3	25	12.00
7926	1	2	3	4	0	20	6	30	20.00
8774	4	3	6	7	0	82	13	102	12.75
9352	3	10	3	7	0	23	16	46	34.78
9692	2	1	7	4	0	21	10	35	28.57
9831	1	1	3	5	0	22	5	32	15.63
10402	1	3	1	2	0	17	5	24	20.83
11535	1	1	2	3	0	22	4	29	13.79
12670	13	11	16	0	0	49	40	89	44.94
13539	2	3	6	4	5	26	11	41	26.83
13169	6	13	13	12	0	50	32	94	34.04
13188	2	2	5	3	0	22	9	34	26.47
13672	8	2	7	5	0	30	17	52	32.69
14208	1	3	1	3	0	19	5	27	18.52
14262	1	0	4	4	0	21	5	30	16.67
14406	1	0	8	4	0	27	9	40	22.50
14953	2	7	5	4	0	23	14	41	34.15
14998	0	1	1	5	0	12	2	19	10.53
15213	0	4	4	3	0	22	8	33	24.24
16300	1	0	5	3	0	20	6	29	20.69
17305	0	10	19	14	0	43	29	86	33.72
17364	1	0	4	5	0	21	5	31	16.13
17654	1	0	6	5	0	23	7	35	20.00

Sample	Pathogenic	Likely Pathogenic	VUS	Likely Benign	Benign	Other	Total filtered Variants	Total Unfiltered Variants	% Variants Passing Filters
17836	0	1	4	7	0	21	5	33	15.15
18715	0	0	5	4	0	24	5	33	15.15
19166	0	1	3	3	0	22	4	29	13.79
19173	0	1	2	0	0	34	3	37	8.11
19191	1	1	4	5	0	30	6	41	14.63
2797	23	18	20	16	1	61	61	138	44.20
2924	6	5	9	10	0	49	20	79	25.32
2932	18	12	9	4	0	54	39	97	40.21
3104	20	20	16	14	0	76	56	146	38.36
4373	3	21	11	25	0	60	35	120	29.17
4483	6	6	13	7	0	56	25	88	28.41
4574	8	19	25	22	1	60	52	134	38.81
4598	19	28	17	28	0	79	64	171	37.43
5491	12	15	13	18	0	42	40	100	40.00
5799	2	8	1	7	1	38	11	56	19.64
5991	15	33	11	17	0	56	59	132	44.70
6001	4	0	4	41	0	47	8	96	8.33
6020	21	15	9	20	1	73	45	138	32.61
6388	1	3	7	6	3	33	11	50	22.00
6186	10	11	9	12	0	47	30	89	33.71
6667	4	6	7	6	0	34	17	57	29.82
7153	2	3	5	7	10	41	10	58	17.24
7004	12	3	14	9	0	71	29	109	26.61

Sample	Pathogenic	Likely Pathogenic	VUS	Likely Benign	Benign	Other	Total filtered Variants	Total Unfiltered Variants	% Variants Passing Filters
7443	5	13	10	15	0	50	28	93	30.11
7619	2	3	5	3	0	30	10	43	23.26
8081	6	1	8	9	0	61	15	85	17.65
8721	10	8	9	6	0	32	27	65	41.54
9417	2	2	2	5	0	20	6	31	19.35
9558	5	2	7	7	0	43	14	64	21.88
9935	3	8	6	5	0	21	17	43	39.53

Annex 14: Clinical impact of features included in12g-algorithm. Implication of gene-alteration in the outcome of patients; OS and DFS.

		n	Events	%	DFS (pvalue)	n	Events	%	OS (pvalue)
MSI	MSI low	68	13	68.3	N.S.	64	9	77.9	N.S.
	MSI high	14	1	91.7		14	1	85.7	
POLE	Non mutated	68	14	64.0	0.038	64	10	74.4	N.S.
	Mutated	14	0	100		14	0	93.3	
PTEN	Non mutated	34	10	35.5	0.002	31	9	48.0	<0.001
	Mutated	48	4	85.4		47	1	94.5	
ARID1A	Non mutated	41	9	53.7	N.S.	38	7	66.0	N.S.
	Mutated	41	5	85.5		40	3	89.2	
ARID5B	Non mutated	42	9	64.4	N.S.	42	8	64.0	0.016
	Mutated	40	5	80.0		36	2	92.0	
RPL22	Non mutated	46	9	60.7	N.S.	42	8	74.3	N.S.
	Mutated	36	5	79.1		36	2	87.8	
PPPR1A	Non mutated	50	10	62.6	N.S.	46	9	62.4	0.005
	Mutated	32	4	82.0		32	1	92.9	
PIK3R1	Non mutated	53	11	62.6	N.S.	49	9	66.8	0.028

		n	Events	%	DFS (pvalue)	n	Events	%	OS (pvalue)
	Mutated	29	3	83.5		29	1	93.2	
TP53	Non mutated	55	6	79.8	N.S.	51	4	82.5	N.S.
	Mutated	27	8	60.6		27	6	73.0	
PIK3CA	Non mutated	55	10	64.7	N.S.	51	7	75.8	N.S.
	Mutated	27	4	82.1		27	3	84.8	
FBXW7	Non mutated	57	10	68.1	N.S.	53	7	74.6	N.S.
	Mutated	25	4	78.8		25	3	85.1	
CTCF	Non mutated	58	10	70.0	N.S.	54	8	72.3	N.S.
	Mutated	24	4	77.6		24	2	90.5	
CTNNB1	Non mutated	69	13	64.3	N.S.	66	9	76.5	N.S.
	Mutated	13	1	92.3		12	1	88.9	
KRAS	Non mutated	74	13	70.5	N.S.	70	10	75.7	N.S.
	Mutated	8	1	85.7		8	0	100	

* n corresponds to total number of cases used in the statistical analysis

Annex 15: Alteration found in cohort 4 sorted by structural annotation among EC prognostic subtypes (Median number of alteration/group).

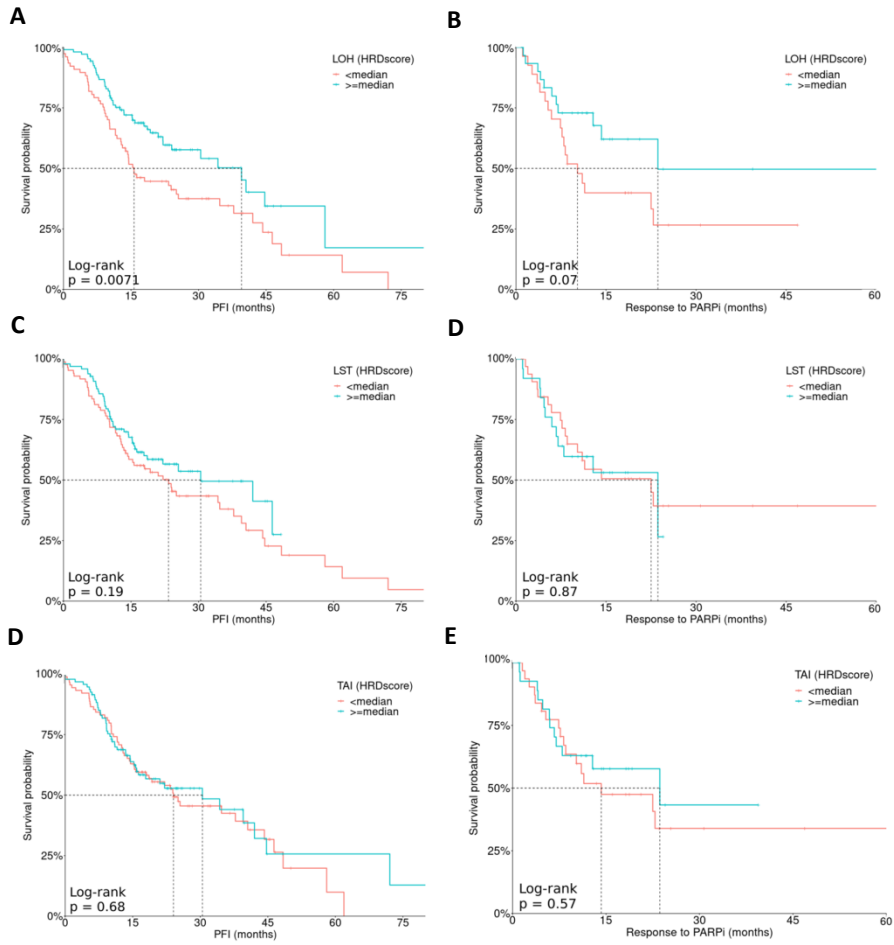
	Molecular group (variants/group)			
	POLE	MSI	CNL	CNH
Regulator	11.0	5.7	6.9	4.8
Frameshift	6.4	3.3	2.5	2.0
In frame	3.3	2.9	2.8	2.0
Splicing events	9.3	5.1	4.6	3.8
Synonymous	24.5	13.1	14.5	9.5
Intron Variant	13.5	8.7	10.0	6.9
Nonsense	3.9	0.8	0.8	0.5
Missense	35.2	11.4	10.5	9.0

Annex 16: Distribution of genetic alteration across the four prognostic groups.

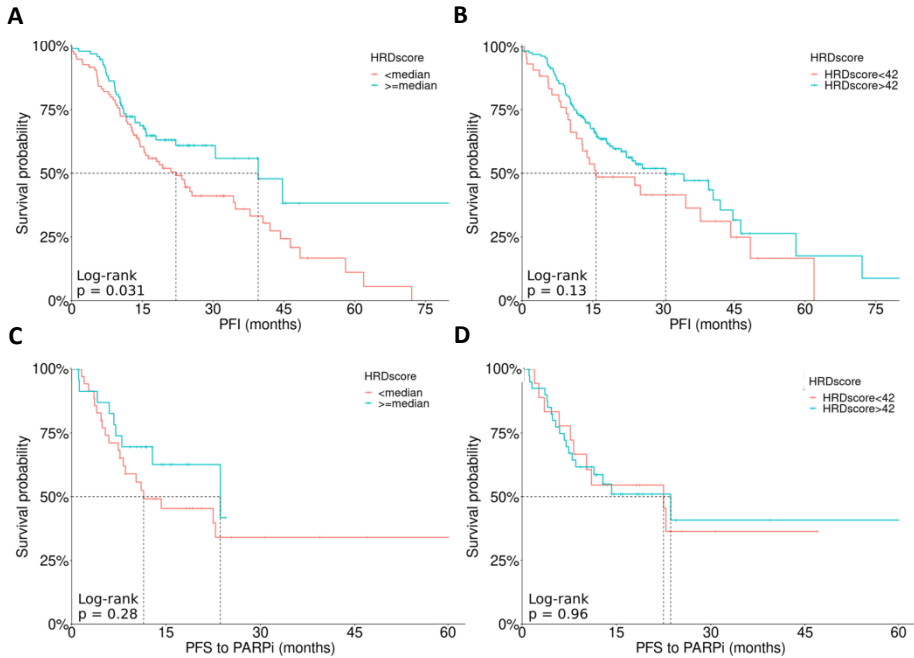
Parameter	Group			
	POLE	MSI	CNL	CNH
MSI (%)	3 (18.7)	12 (100)	0 (0)	0 (0)
POLE (%)	16 (100)	0 (0)	0 (0)	0 (0)
PTEN (%)	14 (87.5)	9 (75)	29 (60.4)	1 (5)
TP53 (%)	7 (43.7)	3 (25)	7 (14.6)	15 (75)
PI3K (%)	12 (75.0)	3 (25)	12 (25.0)	4 (20)
PIK3R1 (%)	11 (68.7)	2 (16.7)	17 (35.4)	3 (15)
ARID1A (%)	13 (81.2)	7 (58.3)	25 (52.1)	2 (10)
ARID5B (%)	10 (62.5)	5 (41.7)	20 (41.7)	7 (35)
KRAS (%)	3 (18.7)	1 (2.0)	5 (10.4)	0 (0)
CTCF (%)	9 (56.2)	2 (16.7)	15 (31.3)	0 (0)
CTNNB1 (%)	8 (50)	0 (0)	6 (12.5)	1 (5)
FBXW7 (%)	13 (81.2)	1 (2.0)	10 (20.8)	4 (20)
PPP2R1A (%)	9 (56.2)	5 (41.7)	11 (22.9)	9 (45)
RPL22 (%)	7 (43.7)	10 (83.3)	19 (39.6)	5 (25)

Annex 17: Log-Rank tests evaluating the implication of predefined HRD scars parameters from scarHRD package in correlation with PFI and PARPi response.

Log-rank tests evaluating performance of: (A) LOH, (C) LST and (E) TAI versus PFI and (B) LOH, (D) LST and (F) TAI performance versus PARPi response.



Annex 18: Correlation of HRD score obtained on scarHRD package and time-to-event variables. Log-rank tests evaluating: (A) PFI prediction regarding HRD score median-based stratification. (B) Pre-established Myriad-based cut-off stratification (42). (C) PFS to PARPi prediction regarding HRD score median-based stratification. (D) Pre-established Myriad-based cut-off stratification (42).



Annex 19: Contribution of each parameter in the CPP model measured as decreasing of the Gini index.

Parameter	CPP-model
TP53	8.765
Grade	7.3384
Histology	4.7434
PTEN	4.2490
CTNNB1	1.6679
ARID1A	1.4816
Stage	0.7857
PPPR1A	0.6392
CTCF	0.5489
PIK3CA	0.4308
KRAS	0.4176
FBXW7	0.3973
PIK3R1	0.3395
ARID5B	0.2071
RPL22	0

Annex 20: Performance of the RF model (CPP) including clinical and pathological parameters (grade, histology and stage).

	CPP-model RFA
Accuracy (95% CI)	0.9808 (0.8974-0.9995)
No Information Rate	0.6923
Kappa	0.9541
McNemar's test p-value	1
Sensitivity	0.9375
Specificity	1
Positive Predictive Value	1
Negative Predictive Value	0.9730
Prevalence	0.3077
Detection Rate	0.2855
Detection prevalence	0.2855
Balanced accuracy	0.9688

Annexed papers

OPEN

Prognostic classification of endometrial cancer using a molecular approach based on a twelve-gene NGS panel

Raquel López-Reig^{1,6}, Antonio Fernández-Serra^{1,6}, Ignacio Romero², Cristina Zorrero³, Carmen Illueca⁴, Zaida García-Casado¹, Andrés Poveda⁵ & José Antonio López-Guerrero^{1*}

Endometrial Cancer (EC) is one of the most common malignancies in women in developed countries. Molecular characterization of different biotypes may improve clinical management of EC. The Cancer Genome Atlas (TCGA) project has revealed four prognostic EC subgroups: POLE, MSI; Copy Number Low (CNL) and Copy Number High (CNH). The goal of this study was to develop a method to classify tumors in any of the four EC prognostic groups using affordable molecular techniques. Ninety-six Formalin-Fixed Paraffin-embedded (FFPE) samples were sequenced following a NGS TruSeq Custom Amplicon low input (Illumina) protocol interrogating a multi-gene panel. MSI analysis was performed by fragment analysis using eight specific microsatellite markers. A Random Forest classification algorithm (RFA), considering NGS results, was developed to stratify EC patients into different prognostic groups. Our approach correctly classifies the EC patients into the four TCGA prognostic biotypes. The RFA assigned the samples to the CNH and CNL groups with an accuracy of 0.9753 ($p < 0.001$). The prognostic value of these groups was prospectively reproduced on our series both for Disease-Free Survival ($p = 0.004$) and Overall Survival ($p = 0.030$). Hence, with the molecular approach herein described, a precise and suitable tool that mimics the prognostic EC subtypes has been solved and validated. Procedure that might be introduced into routine diagnostic practices.

Endometrial Cancer (EC) is the most common gynecological neoplasm and the fourth most frequent cancer in women in developed countries, with 280000 cases per year worldwide¹. This cancer principally affects post-menopausal women, with the peak incidence between 55 and 65 years¹. Clinically, the presence of metrorrhagia in 80% of patients allows both early diagnosis and treatment, resulting in an improved five-year survival². Among newly-diagnosed women, 68% will present localized disease in the uterine cavity, 20% will show disease in pelvic organs and lymph nodes, and about 8% will suffer distant metastasis at diagnosis³. Prognosis varies dramatically according to the stage of the disease. Stage I has an 80–90% five-year survival rate, whereas for Stage IV this rate decreases up to 20%^{4,5}.

Considering the biology and clinical parameters, EC is classified into two groups: type I carcinomas comprise 80% of newly-diagnosed EC and are characterized by alterations in *PTEN*, *KRAS*, and *CTNNB1* and by microsatellite instability (MSI). These tumors are associated with better prognosis^{6,7}. Type II tumors are defined by *TP53* mutations, high Ki-67 score, p16 inactivation and *CDH1* and *HER2* amplification^{8,9}.

Integration of clinicopathological information and genetic data provides more accurate classification of EC into different prognostic groups, facilitating the use of specific therapeutic interventions. The integrated genomic characterization of EC performed by the Cancer Genome Atlas (TCGA) consortium¹⁰ defined four prognostic EC subgroups, with a prognosis from the best to the worst as follows: POLE group, comprising tumors with *POLE* exonuclease domain mutations; MSI group, composed of EC with MSI; Copy Number Low (CNL) and Copy

¹Laboratory of Molecular Biology, Services of Fundación Instituto Valenciano de Oncología, Valencia, Spain.

²Medical Oncology, Fundación Instituto Valenciano de Oncología, Valencia, Spain. ³Gynecology, Fundación Instituto Valenciano de Oncología, Valencia, Spain. ⁴Pathology, Fundación Instituto Valenciano de Oncología, Valencia, Spain.

⁵Department of Oncology, INITIA ONCOLOGY, Hospital Quirón Salud, Valencia, Spain. ⁶These authors contributed equally: Raquel López-Reig and Antonio Fernández-Serra. *email: jalopez@fivo.org

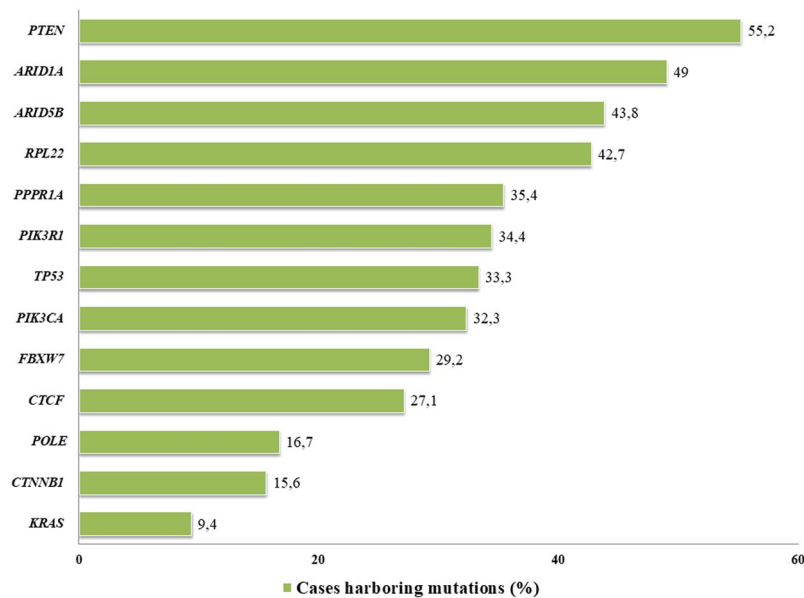


Figure 1. Frequency of gene mutations in EC patient series determined by NGS 13 genes panel. *Hotspot *POLE* (p.P286R and p.V411L) 5.2% (5.6% in TCGA population).

Number High (CNH) groups. CN groups are defined by a differential profile of CN alterations (CNA), CNH group particularly presenting an elevated incidence of *TP53* alterations¹⁰.

The aim of this study was to develop a molecular prognostic classifier for EC that mimics the four TCGA prognostic groups, by using only a small multi-gene NGS panel and MSI determination.

Results

Selection of the multigene-NGS panel and mutational analysis. The EC data set from TCGA¹⁰ defines 48 genes with differential mutation frequencies across the four prognostic groups. A subset of 13 genes, corresponding to those with the highest differences in terms of frequencies between groups, was selected: *POLE*, *PTEN*, *TP53*, *ARID1A*, *KRAS*, *ARID5B*, *FBXW7*, *PPP2R1A*, *CTCF*, *CTNNB1*, *RPL22*, *PIK3CA*, *PIK3R1*. Two separate sequencing runs, containing 48 dual-pool libraries each were performed. The coverage, quality parameters and statistics were comparable between both runs, hence it was possible to merge the data for analysis. Sequencing metrics for analyzed samples are summarized in Supplementary File 1.

A median of 40 genetic alterations per case (range: 13–171) were found (Supplementary File 2). Variants were classified as mutated if they were already reported in ClinVar or if appeared as predicted pathogenic, likely pathogenic or VUS by PolyPhen and SIFT predictors. Benign and likely benign variants were not considered for the analysis. The presence of mutation was treated as categorical dichotomous variable (presence/absence of mutation).

The most frequently affected genes in our series was *PTEN* (55.2%), followed by *ARID1A* (49.0%) and *ARID5B* (43.8%), whereas *KRAS* mutations (9.4%) represent the lowest frequency (Fig. 1). The median number of mutations per patient was 9.5 (range: 2–64). Univariate analysis at gene level showed a correlation between *POLE* mutation and early stage EC ($p = 0.040$), *PTEN* mutations were enriched in EC with endometrioid histology ($p < 0.001$) and low-grade tumors ($p < 0.001$). EC with serous histology harbored more *TP53* mutations ($p = 0.021$). Finally, *RPL22* mutation showed higher frequency in endometrioid histology ($p = 0.005$) and low-grade tumors ($p = 0.004$). *KRAS* ($p = 0.035$) and *CTCF* ($p = 0.05$) mutations were also related with low-grade tumors (Table 1A).

Regarding the prognostic value of individual gene mutations in our series, mutations in *POLE*, *PTEN*, *PIK3R1*, *ARID5B* and *PPP2R1A* are correlated with better patient outcome as seen in Supplementary File 3.

Distribution of microsatellite instability in paired blood and FFPE samples. MSI was observed in 15 of 96 patients (15.6%): 14 of 15 with endometrioid histology (93.3%), and in just 1 of 13 serous cases (7.7%) ($p = \text{N.S.}$). MSI was more frequent in early stages: 11/15 (73.3%) stages I–II vs. 4/14 (26.7%) stages III–IV ($p = \text{N.S.}$). This parameter lacked prognostic value both for PFS and OS (Supplementary File 3). The status of Mismatch repair (MMR) proteins was also evaluated by immunohistochemistry (IHC, Supplementary Information) obtaining a concordance with MSI results of 96%.

Building a predictive multi gene model using a Random Forest approach. A random forest (RF) predictive model for a dichotomous variable (CNL or CNH) was trained using the mutational profile of the 13 selected genes from 148 patients analyzed by the EC TCGA project¹⁰. To correctly adjust the RF model, the TCGA dataset was randomly split in two cohorts (training and validation), based on the distribution of the dichotomous response variable; hence, the groups consisted of 62 patients for the training set and 86 for the validation set.

		Histology			Stage			Grade			
		Endometrioid	Serous	p-value	Early-stage	Advanced-stage	p-value	I	II	III	p-value
(A)											
POLE	mutated	15	1	N.S.	16	0	0.040	10	5	1	N.S.
	non-mutated	68	12		63	17		35	23	22	
PTEN	mutated	53	0	<0.001	44	9	N.S.	30	20	3	<0.001
	non-mutated	30	13		35	8		15	8	20	
TP53	mutated	24	8	0.021	24	8	N.S.	12	8	12	0.050
	non-mutated	59	5		55	9		33	20	11	
KRAS	mutated	8	1	N.S.	9	0	N.S.	8	0	1	0.035
	non-mutated	75	12		70	17		37	28	22	
CTCF	mutated	25	1	N.S.	22	4	N.S.	16	7	3	0.050
	non-mutated	58	12		57	13		29	21	20	
RPL22	mutated	40	1	0.005	34	7	N.S.	17	19	5	0.004
	non-mutated	43	12		45	10		28	9	18	
		DFS			OS						
		Univariate		Multivariate	Univariate		Multivariate				
(B)											
Stage	Early	37.40 (2.067–91.2)		0.006	N.S.	42.57 (2.067–91.20)		0.004	N.S.		
	Advanced	15.37 (4.87–91.00)				34.47 (6.30–91.00)					
Grade	I	50.33 (2.07–91.02)		0.003	N.S.	52.83 (2.067–91.20)		<0.001	8.26 (62.50–1.109)	0.040	
	II	30.28 (9.70–79.57)				33.10 (9.70–79.57)					
	III	26.38 (4.87–67.60)				32.50 (6.30–67.60)					
Histology	Endometrioid	43.63 (5.47–37.40)		<0.001	8.90 (29.90–2.71)	<0.001		44.23 (2.067–91.20)	<0.001	N.S.	
	Serous	21.47 (2.067–91.20)				29.53 (6.30–38.6)					
TCGA groups	POLE	55.40 (24.27–77.43)		0.004	N.S.	55.40 (24.27–77.43)		0.030	N.S.		
	MSI	38.33 (11.9–74.93)				38.6 (11.9–74.93)					
	CNL	34.43 (2.067–91.00)				42.57 (2.067–91.00)					
	CNH	27.70 (4.87–91.2)				30.53 (6.30–91.2)					

Table 1. Correlation between mutational status of analyzed genes and (A) Main clinical and pathological parameters in EC using Chi-square test (B) PFS and OS measured by log-rank test.

To train the model, genotyping of 12 genes was included as categorical dichotomous variables (the so called 12g-model) (Tables 2 and 3). Prior to the adjustment of the RFA model, the number of variables per level on each split was optimized to pre-train the model. The model was validated with 5-fold cross-validation and bagging¹¹.

The POLE and MSI groups were directly defined by the presence of *POLE* mutations and MSI respectively.

Impact of 12 genes RF model in the clinical stratification of the disease. Our series of 96 EC patients was stratified into the four TCGA prognostic groups based on the genotyping data of the 12-gene NGS panel, MSI status, grade, stage and histology: PO LE, 16/96 (16.7%); MSI-H, 12/96 (12.5%); CNH, 20/96 (20.8%); and CNL, 48/96 (50.0%). As mentioned above, CNH and CNL groups were classified with our RF adjusted model.

The POLE group was characterized by a *POLE* exonuclease domain mutation in all 16 cases and by the presence of MSI in 3 of the 16 cases (18.7%). This group presented the highest mutational ratio with a median of 94 variants/case (range: 31–171) compared with the other groups ($p < 0.001$) (Fig. 2). MSI group was characterized by the presence of microsatellite instability in 100% of the cases and had no *POLE* mutations. This group presented a lower median of alterations than POLE with 40 variants per case (range: 19–93). Among these alterations, the most affected genes were *PTEN* (75.0%), *ARID1A* (58.3%) and *RPL22* (83.3%). CNH presented a median of 32 variants per case (range: 19–96) and was characterized by mutations in *TP53* (75%), low frequency of *PTEN* mutations (5%) and alterations in *PPP2R1A* (45%). Finally, CNL showed a median of 37 variants per case (range: 13–138) (Table 4 and Supplementary Fig. S1). Gene by gene analysis of these alterations revealed that: *PTEN* (60.4%) and *TP53* (14.6%) presented the highest and the lowest mutation rate respectively, with alterations in other genes as follows: *PIK3R1* (35.4%), *ARID5B* (41.7%), *CTCF* (31.3%) and *RPL22* (39.6%). The distribution of mutations across groups in EC dataset is depicted in Fig. 3 (Supplementary Table S1).

The Log-Rank test was used to evaluate the prognostic capacity of our molecular classification. This test confirmed that the molecular stratification of our patients revealed differences in both PFS ($p = 0.004$) and OS ($p = 0.030$), suggesting that the POLE and CNH biotypes constituted the best and the worst prognostic groups respectively, mirroring the groups defined by the TCGA (Fig. 4). In addition, a multivariate analysis was performed, being statistically significant only for histology (Table 1B).

Parameter	12 g-model
TP53	12.4658
PTEN	6.094
CTNNB1	3.4884
ARID1A	1.8658
PPPR1A	1.5958
CTCF	1.1435
PIK3CA	0.5644
KRAS	0.3994
FBXW7	0.4852
PIK3R1	0.4506
ARID5B	0.2425
RPL22	0

Table 2. Contribution of evaluated parameters to 12-g-model measured as mean decrease of Gini index of the variables in the models.

	12 g-model RFA
Accuracy (95% CI)	0.9753 (0.9136–0.997)
No Information Rate	0.6049
Kappa	0.9483
McNemar's test p-value	1
Sensitivity	0.9688
Specificity	0.9796
Positive Predictive Value	0.9688
Negative Predictive Value	0.9796
Prevalence	0.3951
Detection Rate	0.3827
Detection prevalence	0.3951
Balanced accuracy	0.9742

Table 3. Performance parameters of 12 g model.

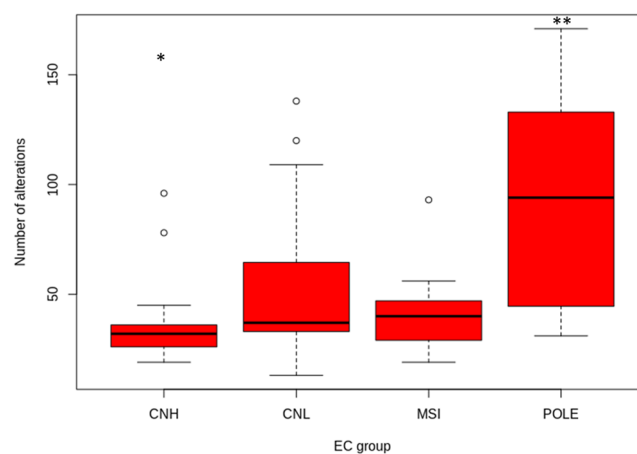


Figure 2. Mutational load across four EC prognostic subtypes. CNH group shows the lowest mutational rate (* $p < 0.05$), whereas POLE mutational rate is the highest (** $p < 0.001$).

Discussion

One of the main problems in the management of EC patients is inter-observer agreement when assigning histology and tumor grade by microscopic techniques. These classifications are associated with different natural histories, treatment scheme and patient outcomes, all of which will influence clinical decision making. Thus, accurate pathological assessment of histology and grade is essential in prognosis assessment and patient management¹². However, this scenario is frequently idealistic. For instance, a misclassification in grade assignment, especially in high-grade EC tumors, has already been reported¹². In addition, there is a poor correlation of histology and grade

	Molecular group (variants/group)			
	POLE	MSI	CNL	CNH
Regulator	11.0	5.7	6.9	4.8
Frameshift	6.4	3.3	2.5	2.0
In frame	3.3	2.9	2.8	2.0
Splicing events	9.3	5.1	4.6	3.8
Synonymous	24.5	13.1	14.5	9.5
Intron Variant	13.5	8.7	10.0	6.9
Nonsense	3.9	0.8	0.8	0.5
Missense	35.2	11.4	10.5	9.0

Table 4. Occurrence of mutations sorted by functional annotation among EC prognostic subtypes (Median number of alteration/group).

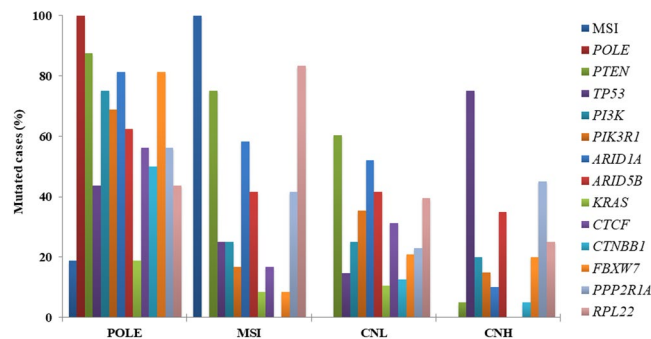


Figure 3. Distribution of genetic alteration across the four EC prognostic subtypes.

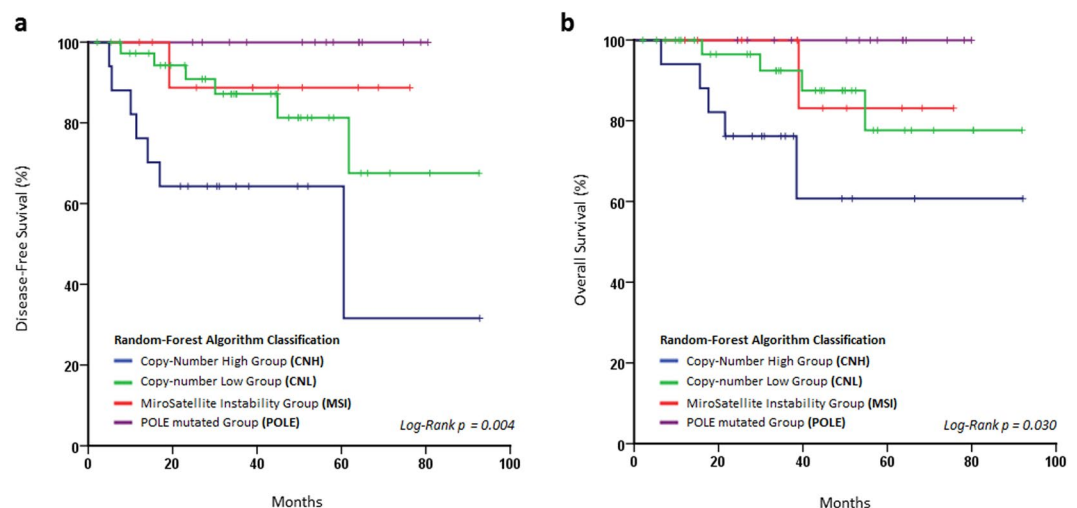


Figure 4. Kaplan-Meier plots assessed by log-rank test to evaluate. (a) Disease free survival and (b) Overall Survival according to 12g stratification. Both parameters reach the statistical signification.

between diagnostic and final tissue samples^{13,14}. Moreover, inter-observer grade agreement has also shown only moderate consistency having a kappa index of 0.41–0.68¹⁵.

In the post-genomic era, multiomic information is redefining tumor classification. In this context, the EC TCGA project was developed and now constitutes an excellent source of data to mine prognostic models¹⁰. EC TCGA described four prognostic groups based on multiomics data. However, this approach is unaffordable in clinical scenarios due to the lack of availability of omic techniques in standard clinical laboratories. Several research groups have made great efforts to overcome these limitations. With this aim, different approaches have been applied to reproduce prognostic classification simplifying the methodology. Stello *et al.* used IHC for p53 and MMR protein assessment and Sanger sequencing for *POLE* hotspots genotyping as surrogate of the EC TCGA subgroups^{16–18}. Similarly, the ProMisE/Vancouver group provided a molecular classification based on p53 IHC as a surrogate of the CNH/CNL TCGA groups^{19,20}. However, the aforementioned inter-observer variability

implicit in IHC makes standardization difficult between labs (Supplementary Fig. S2). This is underlined by the discrete (70%) concordance found in our global series, which is significantly improved in the CNH subgroup (84%), between the determination of *TP53* mutational status using IHC and NGS approach. For these reasons our objective was to develop a method based on the genotyping of only 12 genes with the definition and implementation of a reproducible RF model (12g-algorithm) to classify EC into the four prognostic groups.

We designed a small NGS gene panel with data from the EC TCGA dataset consisting of 13 of the most discriminant genes which presented the highest absolute and differential mutational frequency among the groups. The *POLE* ultramutated group was defined by mutations in the exonuclease domain of this gene. This group presents the highest mutational load and the best prognosis, as previously described by EC TCGA¹⁰. It should be highlighted that our cohort was significantly high in *POLE* mutations (16.7%) compared with the TCGA dataset (7%)¹⁰. This mismatch is presumably due to different NGS technical approaches implemented in the two projects. Whole Exome Sequencing (WES), used in the analysis of EC TCGA samples, achieved a lower coverage (20X) than the targeted panel which was implemented in our work (600X). Therefore, mutations with lower variant allelic frequency could not be detected in WES. However, the percentage of mutations found in *POLE* hotspots (p.P286R and p.V411L) defined by the EC TCGA¹⁰ was concordant between both datasets; 5.3% (our series) and 5.2% (TCGA series). Additionally, the MSI Group was obtained by the determination of MSI status using the eight microsatellite markers (NR27, NR21, NR24, BAT26, BAT25, D5S346, D2S123 and D17S250) by fragment analysis. The IHC for MMR proteins was also performed, obtaining a 96% of concordance between IHC and MSI results. This group was characterized by high mutational ratio, although lower than the *POLE* group (94 vs 40 median of mutations per case respectively).

The most challenging task was to define a surrogate to classify CNH and CNL groups, which currently requires sophisticated technology as well as technical and analytical training. To achieve this, we adjusted a RF model (12g-algorithm) by using the EC TCGA dataset. This 12g model accurately defines CNH and CNL groups (97%) and considers the contribution of each gene to discriminate between groups. Finally, we validated the model with our prospective and independent EC patients series resulting in a total of 20 cases classified as CNH (21%) and 48 as CNL (50%). These frequencies were similar to those reported by the TCGA (26% and 39% respectively)¹⁰. As expected, these groups had lower mutational load than the *POLE* group and were characterized by mutations in *TP53* and *PTEN*. As a whole, this approach showed a good correlation with the TCGA groups and matched its prognostic value. In addition, our prognostic model classified the patients independently of IHC, thus avoiding the intrinsic inter-observer subjectivity.

Besides sequencing and adjusting the 12g-RF model, we trained another model including clinicopathological features (histology, grade and stage) to study the influence of these parameters, the so-called CPP model.

Although there was a slight improvement in the performance parameters of the RFA (Supplementary Tables S2 and S3), it is important to take into account that our series came from a monographic oncology hospital. Additionally, pathological assessment was performed by a single gynecological pathologist highly trained in the diagnosis of EC, possibly masking the subjective effect.

Our approach overcomes subjectivity and technical difficulties related to the definition of CNH and CNL groups. The assessment of the mutational status by NGS technology constitutes a highly objective methodology, drastically simplifying the approach. Furthermore, the common availability of NGS and trained staff in clinical labs will facilitate the implementation of the proposed workflow in the diagnostic routine.

In conclusion, we have defined a prognostic model to classify EC prognostic biotypes based on the analysis of a multi-gene NGS panel; which could be easily implemented as a molecular diagnostic tool.

Material and Methods

Patients. This study includes 96 EC patients prospectively collected from 2010 to 2019 within the context of the institutional projects ACOG0901 and ACOG1602. Experimental protocols were approved by Instituto Valenciano de Oncología (IVO) Institutional Review Board in 2009 and 2016 respectively. All methods used during the study were performed in accordance with the relevant guidelines and regulations.

At the time of the study, our prospective institutional EC database contained a total of 187 patients. Criteria for inclusion in this study was: age over 18 years; tumors with serous or endometrioid histology; grade I to III and stage I, II and III. A total of 149 fulfilled these criteria, from which 96 were selected according to the best ranked DNA quality and concentration.

All analyzed samples were formalin-fixed paraffin-embedded (FFPE) tumor tissue retrieved from the IVO Biobank. Informed consent of patients was obtained in accordance with our institution's ethical and legal regulations.

Clinical and pathological information for the whole series was integrated into a prospective database, median age at diagnosis being 62 years (range: 36.4–87) and median follow-up of 35.02 months (range: 2.1–91.2 months). During follow-up, 15.2% of the patients recurred, and 10.7% died as consequence of the disease; the median progression-free survival (PFS) being 33.65 months (range: 2.1–91.2) and the median overall survival (OS) 35 months (range: 2.1–91.2 months) (Table 5).

Multi-gene next generation sequencing. DNA extraction was performed using the QIAamp DNA FFPE Tissue kit (Qiagen, Valencia, CA) following the manufacturer's instructions. Three FFPE blocks sections of 20 μm -thin with tumor content higher than 50% were used. The final DNA concentration was measured fluorometrically using PicoGreen™ reagent in a Quantifluor instrument (Promega, Fitchburg, Wisconsin). DNA sample quality for NGS selection was estimated using a qPCR-based approach (QC illumina kit) (Illumina, San Diego, CA). In addition, quality and related size of genomic DNA were assessed by the microfluidics-based platform Agilent 4200 TapeStation with Genomic D1000 Kit (Agilent, Santa Clara, CA). Electropherograms were visualized

Stage	Endometrioid			Serous	
	Grade 1	Grade 2	Grade 3	All	Total
(A)					
I	78 (23)	83 (24)	70 (21)	17 (5)	248 (73)
II	3 (1)	9 (3)	6 (2)	5 (1)	23 (7)
III	7 (2)	12 (4)	26 (8)	25 (7)	70 (21)
Adjuvant therapy					
RT	12 (3)	28 (8)	22 (6)	7 (2)	69 (19)
Chemo	2 (1)	6 (2)	14 (4)	13 (4)	35 (10)
ChemoRT	2 (1)	9 (3)	18 (5)	17 (5)	46 (13)
Unknown	70 (20)	61 (17)	57 (16)	16 (5)	204 (58)
Total	86 (24)	104 (29)	111 (31)	53 (15)	354 (100)
(B)					
I	40 (42)	24 (25)	6 (6)	7 (7)	77 (80)
II	0 (0)	1 (1)	0 (0)	1 (1)	2 (2)
III	5 (5)	3 (3)	4 (4)	5 (5)	17 (18)
Adjuvant therapy					
RT	21 (22)	8 (8)	2 (2)	1 (1)	32 (33)
Chemo	2 (2)	1 (1)	2 (2)	7 (7)	12 (13)
ChemoRT	3 (3)	3 (3)	6 (6)	3 (3)	15 (16)
Unknown	2 (2)	1 (1)	0 (0)	2 (2)	5 (5)
Total	28 (29)	13 (13)	10 (10)	13 (13)	64 (67)
32 (33) patients did not receive any treatment					
Median follow-up (months)	34.45 (1.8–91.2)				
Median PFS (months)	33.1 (1.87–91.2)				
Median OS (months)	34.45 (1.87–91.2)				
Relapse (%)	14.6				
Exitus (%)	11.4				

Table 5. Distribution of patients based on most relevant clinical and pathological parameters in (A) TCGA series, (B) Our series.

with the TapeStation Software Analysis A.02.01 SR1 including data collection, peak detection, and interpretation of the different profiles.

For NGS, the median starting DNA concentration was 49.91 ng/μl (8.77–189.538 ng/μl). According to the manufacturer's protocol, the initial amount of DNA required to construct the library is between 10 and 100 ng. In some cases, recommended DNA quantity was not achieved, so maximum available quantity was assigned to these samples. Library preparation was conducted using TruSeq Custom Amplicon Low Input Kit (Illumina, San Diego, CA) in combination with a custom-designed panel (DesignStudio, Illumina, San Diego, California), interrogating the whole coding regions of the following 13 genes: *POLE*, *PTEN*, *TP53*, *ARID1A*, *ARID5B*, *FBXW7*, *PPP2R1A*, *CTCF*, *CTNBN1*, *RPL22*, *KRAS*, *PIK3CA*, *PIK3R1*. These genes were selected based on the sequencing results of the TCGA. By selecting the 13 genes that best discriminate between the 4 groups, based on relative and absolute frequency of each gene among the groups, it is possible to improve the feasibility of the model. Samples were subjected to dual-pool amplicon-based PCR library preparation according to the manufacturer's instructions. Subsequent sequencing of pooled libraries was performed in a NextSeq 550 sequencing platform (Illumina, San Diego, California).

Data analysis, including alignment to the hg19 human reference genome and variant calling, was done using CASAVA pipeline (Illumina, San Diego, CA). These variants were then annotated using the Illumina VariantStudio v3.0 data analysis software (Illumina, San Diego, CA). Integrative Genomic Viewer (Broad Institute) was used to visualize the sequence and check for the presence of mutations^{21,22}. Variants were selected based on a minimum coverage of 600X, minimum frequency of mutated allele of 5% and previously describe or *in silico* as pathogenic, likely pathogenic or variant of unknown significance (VUS).

Microsatellite instability. MSI was performed on 2–3 ng of DNA from paired FFPE and blood samples using the Type-it Mutation Detect PCR Kit (Qiagen) in a Veriti thermocycler (Applied Biosystem, Foster City, CA) and specific primers labelled with the fluorophores FAM, HEX or NED for the following STR regions: NR27, NR21, NR24, BAT26, BAT25, D5S346, D2S123 and D17S250²³. PCR conditions were: 5' initial denaturing at 95 °C followed by 35 cycles at 95 °C of 30", 1'30" at 60 °C and 30" at 72 °C with a final 10' extension at 68 °C. PCR products were denatured with formamide for 5' at 95 °C and visualized, after capillary electrophoresis in the ABI3130xl Genetic Analyzer (Applied Biosystem, Foster City, CA), using the GeneMapper v4.0 software (Applied Biosystem, Foster City, CA). MSI-High (MSI-H) was considered when at least 30% of STR regions presented an MSI pattern.

Random forest algorithm (RFA). The EC dataset from TCGA¹⁰ was used to train a Random Forest algorithm (RFA) to define a prognostic model. Dichotomous and categorical variables including mutational status of the studied genes and clinical and pathological parameters such as histology, stage and grade were implemented in the model. Furthermore, a standard bagging approach is applied. Briefly, the dataset is internally split in three sets in order to internally cross-validate the predictor's performance. The number of trees was empirically estimated to 1000. R v3.4.3 patched was used in all the predictive models built and tested.

Survival analysis. Statistical analysis was performed to define the correlations between clinicopathological and molecular parameters for time-to-event variables [i.e., PFS and OS]. Log-rank test with Kaplan–Meier estimations were performed to compare groups. SPSS v20.0 software was used for statistics.

For categorical variables frequency inference a chi-square test was employed. For median comparison between continuous variables non-parametric tests (Kruskal–Wallis and Wilcoxon) were used.

For RFA classification validation, survival analysis of the four established groups was performed using log-rank test.

Data availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

Received: 17 June 2019; Accepted: 18 November 2019;

Published online: 02 December 2019

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2017. *CA Cancer J Clin* **67**, 7–30, <https://doi.org/10.3322/caac.21387> (2017).
2. Cohn, D. E. *et al.* Should the presence of lymphovascular space involvement be used to assign patients to adjuvant therapy following hysterectomy for unstaged endometrial cancer? *Gynecol Oncol* **87**, 243–246 (2002).
3. Cramer, D. W. The epidemiology of endometrial and ovarian cancer. *Hematol Oncol Clin North Am* **26**, 1–12, <https://doi.org/10.1016/j.hoc.2011.10.009> (2012).
4. Creasman, W. T. *et al.* Carcinoma of the Corpus Uteri. *Int J Gynaecol Obstet* **95**(Suppl 1), S105–S143, [https://doi.org/10.1016/S0020-7292\(06\)60031-3](https://doi.org/10.1016/S0020-7292(06)60031-3) (2006).
5. Mutch, D. FIGO Update: Vancouver, Canada, October 2015. *Gynecol Oncol* **140**, 6–7, <https://doi.org/10.1016/j.ygyno.2015.12.002> (2016).
6. Lax, S. F. Molecular genetic pathways in various types of endometrial carcinoma: from a phenotypical to a molecular-based classification. *Virchows Arch* **444**, 213–223, <https://doi.org/10.1007/s00428-003-0947-3> (2004).
7. Lax, S. F., Kendall, B., Tashiro, H., Slebos, R. J. & Hedrick, L. The frequency of p53, K-ras mutations, and microsatellite instability differs in uterine endometrioid and serous carcinoma: evidence of distinct molecular genetic pathways. *Cancer* **88**, 814–824 (2000).
8. Fadare, O. & Zheng, W. Insights into endometrial serous carcinogenesis and progression. *Int J Clin Exp Pathol* **2**, 411–432 (2009).
9. Okuda, T. *et al.* Genetics of endometrial cancers. *Obstet Gynecol Int* **2010**, 984013, <https://doi.org/10.1155/2010/984013> (2010).
10. Cancer Genome Atlas Research, N. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73, <https://doi.org/10.1038/nature12113> (2013).
11. Chen, X. & Ishwaran, H. Random forests for genomic data analysis. *Genomics* **99** (2012).
12. Clarke, B. A. & Gilks, C. B. Endometrial carcinoma: controversies in histopathological assessment of grade and tumour cell type. *J Clin Pathol* **63**, 410–415, <https://doi.org/10.1136/jcp.2009.071225> (2010).
13. Phelippeau, J. *et al.* Preoperative diagnosis of tumor grade and type in endometrial cancer by pipelle sampling and hysteroscopy: Results of a French study. *Surg Oncol* **25**, 370–377, <https://doi.org/10.1016/j.suronc.2016.08.004> (2016).
14. Talhouk, A. *et al.* Molecular classification of endometrial carcinoma on diagnostic specimens is highly concordant with final hysterectomy: Earlier prognostic information to guide treatment. *Gynecol Oncol* **143**, 46–53, <https://doi.org/10.1016/j.ygyno.2016.07.090> (2016).
15. Gilks, C. B., Oliva, E. & Soslow, R. A. Poor interobserver reproducibility in the diagnosis of high-grade endometrial carcinoma. *Am J Surg Pathol* **37**, 874–881, <https://doi.org/10.1097/PAS.0b013e31827f576a> (2013).
16. Billingsley, C. C. *et al.* Polymerase varepsilon (POLE) mutations in endometrial cancer: clinical outcomes and implications for Lynch syndrome testing. *Cancer* **121**, 386–394, <https://doi.org/10.1002/ncr.29046> (2015).
17. Stelloo, E. *et al.* Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a TransPORTEC initiative. *Mod Pathol* **28**, 836–844, <https://doi.org/10.1038/modpathol.2015.43> (2015).
18. Stelloo, E. *et al.* Improved Risk Assessment by Integrating Molecular and Clinicopathological Factors in Early-stage Endometrial Cancer-Combined Analysis of the PORTEC Cohorts. *Clin Cancer Res* **22**, 4215–4224, <https://doi.org/10.1158/1078-0432.CCR-15-2878> (2016).
19. Talhouk, A. *et al.* Confirmation of ProMisE: A simple, genomics-based clinical classifier for endometrial cancer. *Cancer* **123**, 802–813, <https://doi.org/10.1002/ncr.30496> (2017).
20. Talhouk, A. & McAlpine, J. N. New classification of endometrial cancers: the development and potential applications of genomic-based classification in research and clinical care. *Gynecol Oncol Res Pract* **3**, 14, <https://doi.org/10.1186/s40661-016-0035-4> (2016).
21. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26, <https://doi.org/10.1038/nbt.1754> (2011).
22. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178–192, <https://doi.org/10.1093/bib/bbs017> (2013).
23. Park, J., Shin, S., Yoo, H. M., Lee, S. W. & Kim, J. G. Evaluation of the Three Customized MSI Panels to Improve the Detection of Microsatellite Instability in Gastric Cancer. *Clin Lab* **63**, 705–716, <https://doi.org/10.7754/Clin.Lab.2016.161029> (2017).

Acknowledgements

This work has been performed thanks to the ACIF/2016/008 grant from the Generalitat Valenciana. Authors also thank the Biobank of the Fundación Instituto Valenciano de Oncología for providing the biologic samples for the analysis.

Author contributions

R.L.R., A.F.S., I.R., A.P. and J.A.L.G. contributed to the study design. I.R., Z.G., C.Z. and C.I. contributed to the data acquisition. R.L.R. and A.F.S. contributed to data analysis and statistics. R.L.R., A.F.S., I.R. and J.A.L.G. contributed to manuscript preparation, editing and reviewing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-54624-x>.

Correspondence and requests for materials should be addressed to J.A.L.-G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.






Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Article

Phase 2 Trial (POLA Study) of Lurbinectedin plus Olaparib in Patients with Advanced Solid Tumors: Results of Efficacy, Tolerability, and the Translational Study

Andres Poveda ^{1,*}, Raquel Lopez-Reig ^{2,3} , Ana Oaknin ⁴, Andres Redondo ⁵, Maria Jesus Rubio ⁶ , Eva Guerra ⁷, Lorena Fariñas-Madrid ⁴, Alejandro Gallego ⁵, Victor Rodriguez-Freixinos ^{4,8}, Antonio Fernandez-Serra ^{2,3}, Oscar Juan ⁹, Ignacio Romero ¹⁰ and Jose A. Lopez-Guerrero ^{2,3,11} 

- ¹ Oncogynecologic Department, Initia Oncology, Hospital Quironsalud, Avda Blasco Ibañez, 14, 46010 Valencia, Spain
 - ² Laboratory of Molecular Biology, Fundación Instituto Valenciano de Oncología, 46009 Valencia, Spain; rlopez@fivo.org (R.L.-R.); afernandez@fivo.org (A.F.-S.); jalopez@fivo.org (J.A.L.-G.)
 - ³ IVO-CIPF Joint Research Unit of Cancer, Príncipe Felipe Research Center (CIPF), 46012 Valencia, Spain
 - ⁴ Medical Oncology Department, Vall d'Hebron University Hospital, Vall d'Hebron Institute of Oncology (VHIO), 08035 Barcelona, Spain; aoaknin@vhio.net (A.O.); lfarinas@vhio.net (L.F.-M.); victor.rodriguezfreixinos@sunnybrook.ca (V.R.-F.)
 - ⁵ Medical Oncology Department, Hospital Universitario La Paz-IdiPAZ, Universidad Autónoma de Madrid (UAM), 28049 Madrid, Spain; andres.redondos@uam.es (A.R.); alejandro.gallego@salud.madrid.org (A.G.)
 - ⁶ Medical Oncology Department, University Hospital Reina Sofia, 14004 Cordoba, Spain; mjesusrubio63@gmail.com
 - ⁷ Medical Oncology, Hospital Universitario Ramón y Cajal, 28034 Madrid, Spain; eva_m_guerra@hotmail.com
 - ⁸ Department of Medical Oncology and Hematology, Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto, ON M4N 3M5, Canada
 - ⁹ Medical Oncology, Pivotal SLU, 28023 Madrid, Spain; oscar.juan@pivotalcr.com
 - ¹⁰ Medical Oncology, Fundación Instituto Valenciano de Oncología, 46009 Valencia, Spain; iromero@fivo.org
 - ¹¹ Department of Pathology, School of Medicine, Catholic University of Valencia 'San Vicente Mártir', 46001 Valencia, Spain
- * Correspondence: apoveda@initiaoncologia.com



Citation: Poveda, A.; Lopez-Reig, R.; Oaknin, A.; Redondo, A.; Rubio, M.J.; Guerra, E.; Fariñas-Madrid, L.; Gallego, A.; Rodriguez-Freixinos, V.; Fernandez-Serra, A.; et al. Phase 2 Trial (POLA Study) of Lurbinectedin plus Olaparib in Patients with Advanced Solid Tumors: Results of Efficacy, Tolerability, and the Translational Study. *Cancers* **2022**, *14*, 915. <https://doi.org/10.3390/cancers14040915>

Academic Editors: Christina M. Annunziata and Adam R. Karpf

Received: 21 December 2021

Accepted: 11 February 2022

Published: 12 February 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/>).

Simple Summary: Genomic instability (GI) is a transversal phenomenon in oncology, constituting a hallmark of cancer. In gynecological malignancies, the predictive value of GI has been described and is mainly caused by alterations in the homologous recombination repair (HRR) genes, such as *BRCA1/2*. The POLA clinical trial constitutes an ideal substrate used to study the correlation between GI and response to combined therapy of lurbinectedin plus olaparib in solid tumors. In this context, we developed an approach based on next-generation sequencing, capable of shedding information about Copy Number Variations (CNV) as a surrogate of GI and genotyping of homologous recombination repair genes. Additionally, some algorithms used to extract GI parameters were tested and benchmarked, selecting the most informative mutational and GI features as potential predictive biomarkers for the drug combination explored in the POLA trial.

Abstract: We hypothesized that the combination of olaparib and lurbinectedin maximizes DNA damage, thus increasing its efficacy. The POLA phase 1 trial established the recommended phase 2 dose of lurbinectedin as being 1.5 mg (day 1) and that of olaparib as being 250 mg/12 h (days 1–5) for a 21-day cycle. In phase 2, we explore the efficacy of the combination in terms of clinical response and its correlation with mutations in the HRR genes and the genomic instability (GI) parameters. Results: A total of 73 patients with high-grade ovarian ($n = 46$), endometrial ($n = 26$), and triple-negative breast cancer ($n = 1$) were treated with lurbinectedin and olaparib. Most patients (62%) received ≥ 3 lines of prior therapy. The overall response rate (ORR) and disease control rate (DCR) were 9.6% and 72.6%, respectively. The median progression-free survival (PFS) was 4.54 months (95% CI 3.0–5.2). Twelve (16.4%) patients were considered long-term responders (LTR), with a median PFS of 13.3 months. No clinical benefit was observed for cases with HRR gene mutation. In ovarian LTRs, although a direct

association with GI and a total loss of heterozygosity (LOH) events was observed, the association did not reach statistical significance ($p = 0.055$). Globally, the total number of LOHs might be associated with the ORR ($p = 0.074$). The most common grade 3–4 toxicities were anemia and thrombocytopenia, in 6 (8.2%) and 3 (4.1%) patients, respectively. Conclusion: The POLA study provides evidence that the administration of lurbinectedin and olaparib is feasible and tolerable, with a DCR of 72.6%. Different GI parameters showed associations with better responses.

Keywords: ovarian cancer; endometrial cancer; lurbinectedin; olaparib; genomic instability

1. Introduction

The treatment of ovarian cancer has seen increasing improvement over recent years. Today, the most critical advance has been the use of poly (ADP-Ribose) polymerase inhibitors (PARPis). In 2009, a phase I study on olaparib presented the first clinical evidence of PARPi having an effect in patients with *BRCA1/2* mutations, with the benefits being of a magnitude never observed before [1]. However, the clinical benefit of PARPi is not limited to patients with *BRCA1/2* mutations; the entire population of high-grade serous (HGS) ovarian cancer or triple-negative breast cancer has observed its benefits [2]. In patients with the absence of *BRCA* alterations, the efficacy of PARPi is more pronounced in those with homologous recombinant deficiency (HRD). Several phase II and III trials have demonstrated the efficacy of PARPi in patients with ovarian cancer [3] and have led to the approval of three PARPis—olaparib, niraparib, and rucaparib—as maintenance therapy for platinum-sensitive recurrent ovarian cancer [4–8].

Regarding maintenance, PARPis have been administered as monotherapy in a continuous oral dosing schedule. The combination of PARPis with chemotherapy or other agents is an interesting approach to increasing their efficacy, especially in previously treated patients and those with unknown DNA repair deficits [9]. However, combination trials using continuous olaparib with chemotherapy, such as cisplatin or carboplatin alone, or combined with gemcitabine, had to be stopped prematurely due to high hematological toxicity. An intermittent dose of olaparib, especially with a short course, has shown better tolerability when combined with chemotherapy [10–13]. Myelosuppression is the main effect caused by toxicity of PARPi and is considered a “class toxicity”. However, only the PARPi veliparib has been associated with lower myelosuppression in monotherapy studies, and the continuous administration of veliparib has been successfully combined with chemotherapy [14,15].

Trabectedin is an anticancer drug structurally related to ecteinascidins and approved in many countries to treat patients with relapsed platinum-sensitive ovarian cancer. Lurbinectedin (PM01183) is a novel synthetic alkaloid structurally similar to trabectedin. Both contain a pentacyclic skeleton composed of two fused tetrahydroisoquinoline rings (subunits A and B) responsible primarily for DNA recognition and binding. However, the additional module (ring C) in lurbinectedin is a tetrahydro β -carboline rather than the additional tetrahydroisoquinoline present in trabectedin. This structural difference may confer pharmacokinetic benefits and intrinsic activity [16]. Lurbinectedin joins covalently to the DNA, inducing DNA double-strand breaks that initiate apoptosis [17] and reducing tumor-associated macrophages and the inflammatory microenvironment by inhibiting inflammatory factors [17]. Lurbinectedin has been recently approved by the U.S. Food and Drug Administration (FDA) based on a phase 2 single-arm study in 105 platinum-sensitive and platinum-resistant adult patients with metastatic small cell lung cancer and disease progression on or after platinum-based chemotherapy [18]. In a small randomized phase 2 study, lurbinectedin showed high activity in patients with platinum-resistant ovarian cancer [19]. However, a recently published phase 3 study contradicted these results, with lurbinectedin showing a similar antitumor efficacy to topotecan or liposomal doxorubicin but having a better toxicity profile [20].

Combining a PARPi (olaparib) with a DNA damaging agent (trabectedin or lurbinectedin) is an exciting approach to maximizing the effect of DNA damage. In preclinical models, the combination of both agents was synergistic and led to biologically significant deregulation of the DNA damage repair machinery that elicited relevant antitumor activity [21,22]. However, overlapping hematological toxicity may represent a limitation of the combination. The lurbinectedin dose adjusted to the body surface area showed lower hematological toxicity (57%) than flat dose [19,23]. Similarly, an intermittent schedule of olaparib is feasible and has a lower rate of hematologic adverse events than a continuous course when combined with chemotherapy [10,11].

Recently, we reported the results of a phase I dose-finding study with a short course of olaparib and lurbinectedin in patients with ovarian and endometrial cancer. The dose-limiting toxicity was grade 4 neutropenia, and the recommended phase 2 dose (RP2D) was 1.5 mg/m² of lurbinectedin administered intravenously on day 1 and 250 mg of olaparib administered as oral capsules twice a day (BID) on days 1–5 of a 21-day cycle [24]. Most adverse events were mild, and the treatment was well-tolerated. Moreover, we obtained a disease control rate (DCR) of 60% (but no responses). Overall, the favorable safety profile and preliminary efficacy results deserved further investigation.

The POLA is the first phase 2 trial to assess the efficacy and toxicity of lurbinectedin and olaparib in previously treated gynecological tumors and their correlation with molecular characteristics.

2. Materials and Methods

2.1. Study Population

This is a phase 2, open-label, non-randomized study that recruited patients from five centers in Spain. Patients aged ≥ 18 years were eligible if they had histologically confirmed advanced or metastatic HGS or endometrioid (no mucinous and no clear cells) platinum-resistant—Not refractory (neither primary nor secondary)—Ovarian cancer, fallopian cancer, primary peritoneal cancer, endometrial cancer (any grade, not platinum-refractory), or triple-negative breast cancer; had an Eastern Cooperative Group (ECOG) performance status (PS) ≤ 2 ; had a life expectancy of ≥ 3 months; had a measurable disease according to the Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1; received at least one line of standard therapy for locally advanced or metastatic disease and developed progression disease afterwards (no limit was placed on the number of prior therapies); had hemoglobin ≥ 10 g/dL; had an absolute neutrophil count $\geq 1500/\mu\text{L}$; had platelets $\geq 100,000/\mu$; had total bilirubin ≤ 1.5 times the institutional upper limit of normal (ULN); had aspartate aminotransferase and alanine aminotransferase ≤ 2.5 times ULN; had albumin ≥ 3 g/dL; and had creatinine ≤ 1.5 times the ULN or a creatinine clearance ≥ 30 mL/min. Patients were ineligible if they had received previous treatment with a PARPi or lurbinectedin.

The study (NCT02684318, EudraCT 2015-001141-08, 03.10.2015) was approved by a centralized ethics committee and was conducted following the Declaration of Helsinki, ICH Good Clinical Practice guidelines, and the current legislation. Written informed consent was obtained from all patients before conducting study-specific procedures.

2.2. Study Treatment

The patients received 1.5 mg/m² of lurbinectedin intravenously on day 1 in combination with oral administration of 250 mg of olaparib/12 h on days 1–5 BID of a cycle of 21 days according to the RP2D determined in the phase I trial. The study treatments were given until objective disease progression according to the RECIST 1.1, unacceptable toxicity, or patient withdrawal of consent. At screening, patients underwent a history and physical examination, baseline hematological and chemistry assessments and urinalysis, blood sampling for pharmacogenomics (PG) analysis, ECG, and tumor assessment. The patients were seen on day 1 and day 15 of cycles 1 and 2 and every 3 weeks for the rest of the cycles for history and physical examination, hematological and chemistry assessment,

and PG sampling (only in cycle 1). Tumor response was assessed by the investigators using the same method used during screening, which was in line with the RECIST v1.1 every 2 cycles (6 weeks) until disease progression or death. All toxic effects were graded using the National Cancer Institute-Common Terminology Criteria for Adverse Events (NCI-CTCAE) version 4.0.3 (https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm#ctc_40 (accessed on 22 January 2022)).

2.3. Outcomes

The primary endpoint was overall response rate (ORR), defined as a complete response (CR) or partial response (PR) according to the RECIST v1.1. The secondary endpoints were progression-free survival (PFS), overall survival (OS), safety, and translational studies. The exploratory objectives included ORR and PFS by tumor type and by the number of previous treatment lines, duration of response, and long-term responders (LTR).

2.4. Translational Studies

2.4.1. DNA Extraction

DNA extraction was performed using $3 \times 20 \mu\text{m}$ sections of formalin-fixed paraffin-embedded (FFPE) archived tumors and the QIAmp DNA FFPE Tissue kit (Qiagen Iberica S.L., Spain). DNA integrity, concentration, and fragment size were determined using a Genomic DNA ScreenTape assay (TapeStation 4200, Agilent, Santa Clara, CA, USA).

2.4.2. Next-Generation Sequencing (NGS) Panel

The libraries were prepared using the Agilent (Santa Clara, CA, USA) SureSelect-XT HS Target Enrichment Kit combined with OneSeq backbone 1 Mb. Briefly, 200 ng of extracted DNA were enzymatically fragmented to a size between 150 and 200 bp. Each library was then hybridized to a SureSelect custom panel (Agilent) according to the manufacturer's protocol. The custom panel, designed to evaluate the HRD status, includes 35 genes involved in different DNA repair pathways: *BRCA1*, *BRCA2*, *BARD1*, *BRIPI1*, *CHEK1*, *CHEK2*, *FAM175A*, *NBN*, *PALB2*, *ATM*, *MRE11A*, *RAD51B*, *RAD51C*, *RAD51D*, *RAD54L*, *FANCI*, *FANCM*, *FANCA*, *ERCC1*, *ERCC2*, *ERCC6*, *REQL*, *XRCC4*, *HELQ*, *SLX4*, *WRN*, *ATR*, *PTEN*, *CCNE1*, *EMSY*, *TP53*, *MLH1*, *MSH2*, *MSH6*, and *PMS2*, and 147,000 SNPs distributed homogeneously along the genome that served to obtain Copy Number (CN) profiles. The pooled library was sequenced (2×100 cycles) on a NextSeq550 using a high output flow cell (Illumina, San Diego, CA, USA). A secondary analysis was performed with Haplotypcaller (Broad Institute) for variant calling and Variant studio 4.0 for annotation (Illumina, San Diego, CA, USA). The variants were considered when classified as pathogenic (P), likely pathogenic (LP), or variant of unknown significance (VUS) with pathogenic prediction or variants with both in silico predictors, SIFT and Polyphen, predicted as pathogenic. The variants were filtered based on the coverage and functional annotation. The minimum coverage for a variant was established at $100\times$. Mutations were accepted with a frequency higher than 5%. For CN calling at the gene level, the PanelMops package [25] from R was applied. The genomic instability was established using NGS OneSeq kit (Agilent) data. Briefly, the cnvkit algorithm [26] was used with bam alignment files as the input. Filtering by a *p*-value of 0.001 was applied, and the copy number events were adjusted to the tumor burden of every sample; this tuning was applied during cns file creation. The genome's LOH regions were established by comparing the heterozygote regions of a panel of five controls. Lastly, a post-analytical filter removing alterations shorter than 1 Mb, which were assigned as probable technical artifacts, was applied previously to the data analysis.

The studied parameters were the number of copy number variation (CNV) events, the average length per event, the length of the genome altered by these events, percentage of the genome altered, the same four parameters removing borderline events with a biallelic frequency (BAF) between 0.3 and 0.7, the number of gains, the length of the genome affected by gains, the percentage of genomes affected by gains, the three gain events removing a BAF between 0.3 and 0.7, the total number of losses, the length of genomes

affected by copy number losses, the length of the genome suffering LOH events [27], the percentage of genomes altered by LOH, the number of events, and the length of genomes and the percentage of genomes altered with LOH spanning more than both 15 and 10 Mb. Continuous variables were categorized according to their median and quartiles.

The parameter settings and codes used for GI determination with *cnvkit* software and script to extract analytical features are available at https://github.com/afernandezse/Pola_Phase2_GI_traslational (accessed on 22 January 2022).

2.4.3. Multiplex Ligation-Dependent Probe Amplification (MLPA) Analysis

To validate the *in silico* assessment of CN amplification and losses at the gene level in *CCNE1*, *PTEN*, and *EMSY* (previously described in the OC population), an MLPA analysis was performed. SALSA[®] MLPA Probemix p225-E1 and P078-D2 Breast tumor assays were used, and the protocol was performed following the manufacturer's instructions (MRC Holland, Amsterdam, The Netherlands). Amplified products were separated using an ABI3130XL Genetic Analyzer (Applied Biosystem, Foster City, CA, USA) and interpreted with GeneMapper Software v4.0 (Applied Biosystems, Foster City, CA, USA). Quantification of the fragment analysis results was performed using the Coffalyser software as described by the manufacturer (MRC Holland). Different normal control samples from healthy FFPE tissue were used to normalize the allele dosage.

The subrogates of the deficiency of the homologous recombination repair (HRR) pathway were HRD status, defined as single nucleotide variants (SNVs) and indels in HRR-genes, and different pre-established GI parameters (See the Supplementary Materials). To assess their predictive power, these parameters faced response-based rates. ORR was the result of grouping CR and PR versus SD and PD, whereas Clinical Benefit Rate (CBR) grouped CR, PR, and SD of more than 6 months versus SD of fewer than 6 months and PD. Their associations with LTR were also explored.

2.5. Statistical Analysis

According to the Fleming method for phase 2 trials, 73 patients had to have more than a 90% chance of their ORRs being significantly different (0.05), considered the minimum (historical control 24%) and optimal ORRs for the proposed experimental schedule (estimated at 40%). For both safety and efficacy analyses, all patients who received at least one dose of the study treatment were included. The time-to-event analysis (PFS, OS, and duration of response) was analyzed using the Kaplan–Meier method. The Clopper–Pearson method was used to present the number and percentage of patients achieving a response with a two-sided 95% confidence interval (CI).

A statistical analysis was performed to define the correlations between clinicopathological and molecular parameters for time-to-event variables (i.e., PFS and OS). Differences between a Kaplan–Meier curves tests were determined using a log-rank test. R version 4.03 was used for statistics. A side effect was estimated using Cohen's D test.

Further statistical analyses of all endpoints were performed following the Statistical Analysis Plan.

3. Results

3.1. Efficacy

A total of 84 patients were screened, and 73 patients received at least one dose of the study drugs (Supplementary Figure S1). When the study database was locked (January 2019), all patients had discontinued treatment. The median treatment duration was 15 weeks (the minimum duration was 7 weeks, and the maximum was 25 weeks). The principal reason for treatment discontinuation was radiological progression disease in 59 patients (73%). Other reasons included patient decision (four patients), adverse event (two patients), death (one patient), protocol violation such as an overdose or skipped dose of olaparib (two patients), clinical progression (two patients), or clinical deterioration (three patients).

The clinical characteristics of the patients are summarized in Table 1. Most patients had HGS or high-grade endometrioid ovarian, fallopian tube, or primary peritoneum cancer ($n = 46$ patients, 63%), with endometrial cancer being the second most common tumor type ($n = 26$ patients, 35.6%). There was only one patient with triple-negative breast cancer. Twenty-seven (37%) patients presented visceral metastases at the time of the inclusion, with the most common site being lymph nodes in 16 patients (59.3%), followed by the lungs in 10 patients (37%). The median time from the diagnosis to the inclusion in the trial was 43.4 months (range: 6.3–171.8 months), and most patients ($n = 45$, 61.6%) had received three or more prior lines of therapy.

Table 1. Patient characteristics.

Characteristic	Lurbinectedin plus Olaparib, $n = 73$ Patients
Age, median (range), years	65 (22–80)
Gender	
Females, n (%)	73 (100)
ECOG PS, n (%)	
0	40 (54.8)
1	33 (45.2)
Primary tumor type, n (%)	
Ovarian carcinoma	46 (63)
High-grade serous	44 (60.3)
High-grade endometrioid	2 (2.7)
Endometrial carcinoma	23 (31.5)
Endometrial carcinosarcoma	3 (4.1)
Triple negative breast cancer	1 (1.4)
Metastasis at baseline, n (%)	27 (37)
Lung	10 (37)
Liver	5 (18.5)
Lymph nodes	16 (59.3)
Bone	1 (3.7)
Others	12 (44.4)
Number of previous treatment regimens	
<3 treatments, n (%)	28 (38.4)
≥ 3 treatments, n (%)	45 (61.6)

ORR evaluated per RECIST in the intention-to-treat population (73 patients) was 9.6%: one (1.4%) and six (8.2%) patients achieved CR and PR, respectively (Supplementary Table S1). However, the disease control rate (DCR = CR + PR + SD) was 72.6%. The best percentage of change from baseline in target lesions is shown in Figure 1. Five (6.8%) patients were unevaluable.

In the subgroup analysis, patients with endometrial cancer had higher ORR than patients with ovarian cancer (15.4% vs. 6.6%, respectively; $p = 0.057$). The number of previous lines of therapy influenced ORR; in patients with less than three previous lines ($n = 28$), the ORR was 21.4%, whilst in patients with three or more previous lines ($n = 45$), the ORR was 2.2% ($p = 0.02$). The sole patient with CR had been diagnosed with endometrial cancer and had received less than three lines of previous therapies.

The median PFS was 4.54 months (95% CI 3.0, 5.2) (Supplementary Figure S2A). No significant statistical differences were found in terms of PFS according to the primary site of the tumor. The median PFS was 4.5 (95% CI 3.0, 5.1) months for ovarian cancer and 4.8 (95% CI 1.9, 6.8) months for endometrial cancer (Supplementary Figure S2B,C). For the whole population, the PFS rate at 6 months was 28.56% (95% CI 18.34, 39.62) (Supplementary Figure S2A).

An exploratory analysis was performed to characterize the subset of patients deriving long-term benefit from the combination of lurbinectedin and olaparib. Long-term responders (LTRs) were defined as patients whose PFS was equal to or greater than the double estimated median PFS (4.54 months). In total, 12 (16.4%) of the 73 patients were considered

LTR, with a median PFS of 13.3 months (Supplementary Table S2). The median OS for the entire population was 15.19 (95% CI 12.13, 17.69) (Supplementary Figure S3A). No differences were found in OS according to the tumor type (Supplementary Figure S3B).

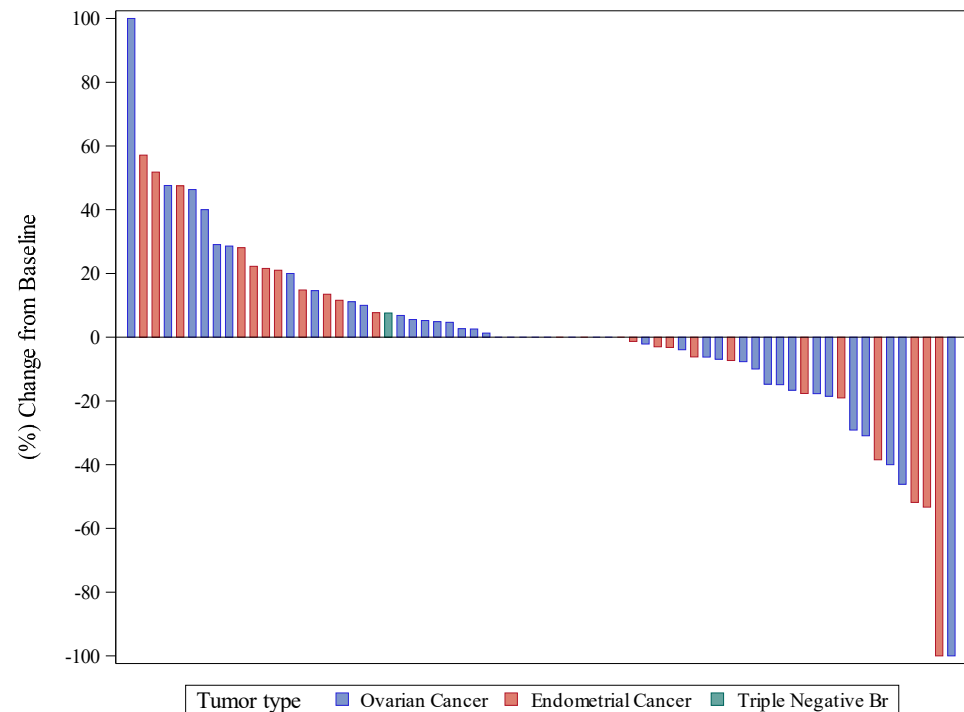


Figure 1. Waterfall plot of best response, as a percentage of change in target lesions.

3.2. Safety and Tolerability

All 73 patients had at least one treatment-emergent adverse event (TEAE) (Table 2). Overall, treatment with lurbinectedin and olaparib was well-tolerated, with most TEAEs being grade 1 or 2. A total of 26 patients (35.5%) experienced grade 1–2 TEAEs, most commonly asthenia, nausea, vomiting, constipation, diarrhea, abdominal pain, dysgeusia, and anemia. The most common grade > 3 TEAE was hematological toxicity, predominantly neutropenia, which was reported in 28 patients (38.3%). Grade 3–4 anemia and thrombocytopenia were observed in 6 (8.2%) and 3 (4.1%) patients, respectively. The most common grade 3–4 non-hematologic toxicity was asthenia, reported in 6 patients (8.2%). Serious TEAEs were observed in 22 patients (30.1%), which were related to study drugs in 3 of the patients: 1 patient had grade 3 diarrhea, 1 patient had grade 3 constipation, and 1 patient had grade 3 cardiac disorders. No deaths were related to adverse events.

Over the course of the treatment, 42 patients (57.5%) required a dose reduction in at least one drug due to adverse events. Six patients (8.2%) discontinued treatment due to toxicity as the main reason.

3.3. Translational Studies

3.3.1. Distribution of Genetic Alterations and Clinical Impact

Genetic studies were performed on a total of 57 samples that passed the quality and quantity requirements, corresponding to 19 (33.3%) endometrial cancer and 38 (66.7%) ovarian cancer patients. Among all of the mutated genes, considering both cancer types, *TP53* and *PTEN* presented the highest mutational ratios, with 34/57 (59.6%) and 9/57 (15.8%), respectively, excluding CNVs. *TP53* alterations were mainly present in ovarian cancer (70.6%), specifically in HGS histology, while *PTEN* was preferentially altered in endometrial cancer (88.9%). Regarding the HRR pathway, a total of eight genes presented alterations, including *BRCA1* (3, 5.3%), *BRCA2* (1, 1.8%), *ATM* (2, 3.5%), *RAD5L* (1, 1.8%), *ATR* (1, 1.8%), *NBN* (1, 1.8%), *SLX* (1, 1.8%), and *WRN* (1, 1.8%). Overall, HRR gene

alterations were reported in 10/57 (17.5%) cases homogeneously distributed between endometrial cancer and ovarian cancer, and they were used in the following analysis as an HRD status subrogates. Additionally, mutations in the Fanconi Anemia genes, *FANCM* (1, 1.8%) and *FANCA* (1, 1.8%), were also found. Finally, alterations in the *MMR* genes were described in two endometrial cancer cases (Figure 2).

Table 2. Treatment-emergent adverse event ($\geq 5\%$) by maximum grade per patient.

TEAE	Grade 1		Grade 2		Grade 3		Grade 4		Total	
	<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)
Anemia	8	(10.9)	7	(9.5)	6	(8.2)	0	(0)	21	(27.3)
Leukopenia	3	(4.1)	5	(6.8)	3	(4.1)	0	(0)	11	(15)
Neutropenia	0	(0)	10	(13.6)	19	(26)	9	(12.3)	38	(52)
Trombocytopenia	4	(5.4)	2	(2.7)	2	(2.7)	1	(1.3)	9	(12.3)
Abdominal pain	14	(19.0)	10	(13.6)	2	(2.7)	0	(0)	26	(35.3)
Constipation	18	(24.6)	6	(8.2)	0	(0)	0	(0)	24	(32.8)
Diarrhea	15	(20.5)	4	(5.4)	1	(1.3)	0	(0)	20	(27.3)
Dyspepsia	3	(4.1)	2	(2.7)	0	(0)	0	(0)	5	(6.8)
Nausea	30	(41)	11	(15)	0	(0)	0	(0)	41	(56.1)
Vomiting	12	(16.4)	7	(9.5)	1	(1.3)	0	(0)	20	(27.3)
Asthenia	15	(20.5)	29	(39.7)	6	(8.2)	0	(0.0)	50	(68.4)
Fatigue	2	(2.7)	3	(4.1)	0	(0)	0	(0)	5	(6.8)
Mucosal inflammation	3	(4.1)	1	(1.3)	0	(0)	0	(0)	4	(5.4)
Pyrexia	9	(12.3)	1	(1.3)	0	(0)	0	(0)	10	(13.6)
Bronchitis	3	(4.1)	2	(2.7)	0	(0)	0	(0)	5	(6.8)
Urinary tract infection	1	(1.3)	0	(0)	0	(0)	0	(0)	1	(1.3)
ALT/GPT increased	4	(5.4)	1	(1.3)	1	(1.3)	0	(0)	6	(8.1)
AST/GOT increased	5	(6.8)	2	(2.7)	0	(0)	1	(1.3)	8	(10.9)
GGT increased	0	(0)	1	(1.3)	0	(0)	3	(4.1)	4	(5.4)
Decreased appetite	8	(10.9)	4	(5.4)	0	(0)	0	(0)	12	(16.4)
Hypoalbuminaemia	2	(2.7)	4	(5.4)	0	(0)	0	(0)	6	(8.2)
Hypogamnesaemia	7	(9.5)	0	(0)	0	(0)	0	(0)	7	(9.5)
Artralgia	4	(5.4)	1	(1.3)	0	(0)	0	(0)	5	(6.8)
Back pain	7	(9.5)	3	(4.1)	0	(0)	0	(0)	10	(13.6)
Dizziness	3	(4.1)	0	(0)	0	(0)	0	(0)	3	(4.1)
Dysgeusia	16	(21.9)	1	(1.3)	0	(0)	0	(0)	17	(23.2)
Cough	4	(5.4)	0	(0)	0	(0)	0	(0)	4	(5.4)
Dyspnea	6	(8.2)	4	(5.4)	1	(1.3)	0	(0)	11	(15)
Pulmonary embolism	1	(1.3)	1	(1.3)	4	(4.5)	0	(0)	6	(8.2)
Lymphedema	3	(4.1)	1	(1.3)	0	(0)	0	(0)	4	(5.4)

We studied the possible relationship between HRD status and response to treatment. No correlations were found between study treatment ORR or CBR, and HRR mutations. The different GI parameters (Supplementary Material and Methods) were compared with the mutation-based stratification. In the whole population, HRD status was associated with losses ($p = 0.0038$) and the percentage of the genomes affected by losses ($p = 0.034$) (Figure 3A,B). Considering that GI caused by HRR gene mutations has been principally described in the ovarian cancer population, we studied GI patterns according to cancer type (Supplementary Figure S4). The ovarian cancer cohort ($n = 38$) showed a significant correlation between HRD status and the total number of events ($p = 0.0053$), loss events ($p = 0.0012$), and percentage of the genome affected by losses ($p = 0.012$). Loss of heterozygosity (LOH) did not correlate with treatment response ($p = 0.091$) (Figure 3C–F). On the other hand, the endometrial cancer cohort ($n = 19$) did not show any significant results.

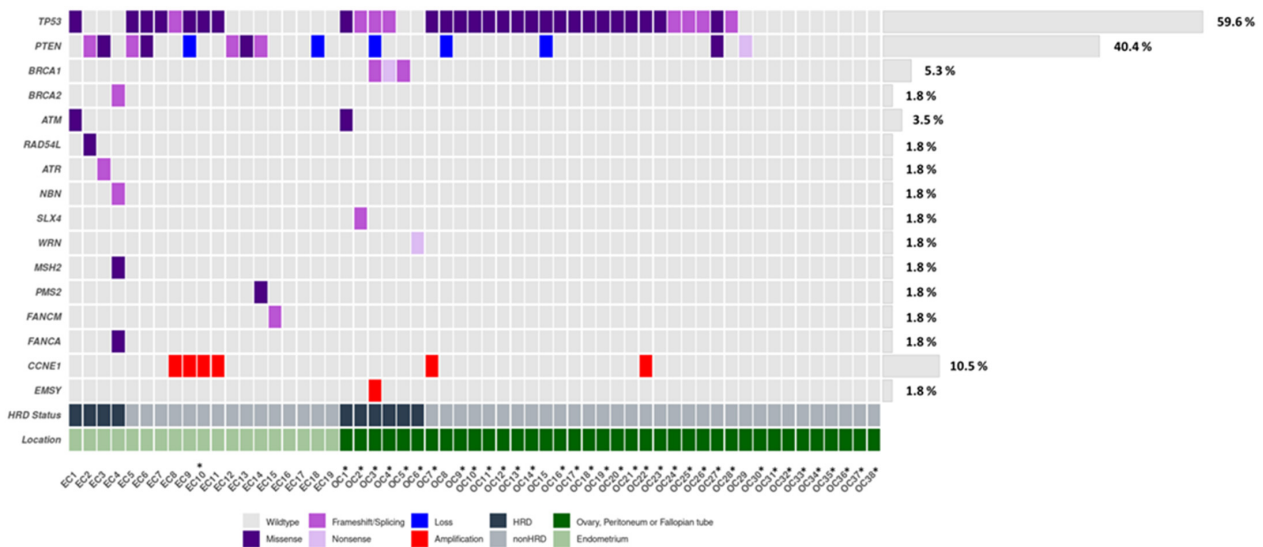


Figure 2. Oncoplot of genetic and genomic alterations across the 35 genes of the custom panel, related to HRR and other DNA repair pathways across the EC ($n = 19$) and OC ($n = 38$) cohorts. The Oncoplot shows SNVs and CN at the gene level in *CCNE1*, *PTEN*, and *EMSY*. * Serous histology.

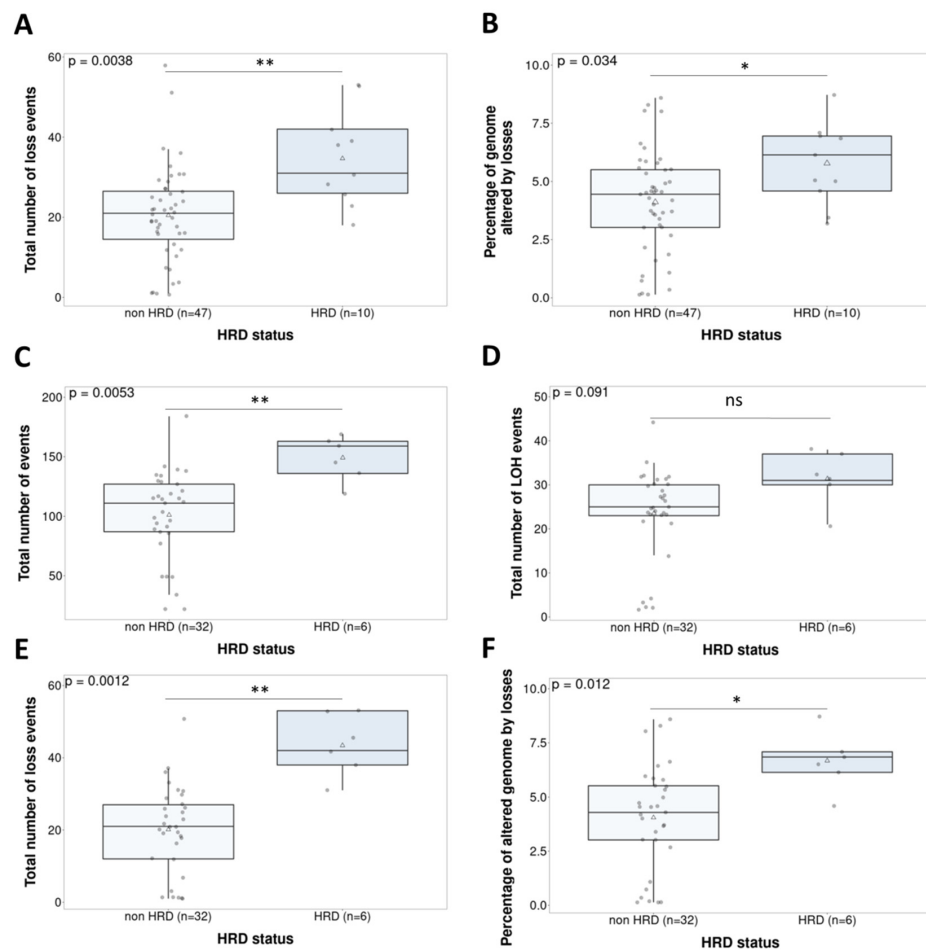


Figure 3. Non-parametric tests (Wilcoxon signed rank test) comparing the genomic instability parameters and HRD status in the whole population ($n = 57$)—loss events (Cohen’s $d = 1.2$) (A) and percent of altered genome by losses ($d = 0.8$) (B)—and in the ovarian cancer population ($n = 38$)—total number of events ($d = 1.4$) (C), LOH events ($d = 0.81$) (D), loss events ($d = 2$) (E), and percent of altered genome by losses ($d = 1.17$) (F). ns. Not significant; * $p < 0.05$; ** $p < 0.01$.

3.3.2. Characterization of Copy Number Patterns across the Clinical Trial Population: Clinical Impact of Genomic Instability-Based Classification

Finally, the GI parameters were evaluated as a predictive biomarker for the combination of olaparib and lurbinectedin. First, in terms of the response and duration of response, LTRs were assessed. When evaluating the ovarian cancer population specifically ($n = 27$), we observed a trend towards an association between LTRs and total LOH events, which did not reach statistical significance with the current sample size ($p = 0.055$) (Figure 4B). Second, the relationship between GI and ORR was also evaluated. The total number of LOH events was not associated with ORR ($p = 0.074$) (Supplementary Figure S5A). We observed a significant correlation between ORR and the percentage of genome altered by losses ($p = 0.021$), although only two cases qualified as responders with HGS histology (Figure 4A). In the endometrial cancer population, the percentage of the total genome that was altered was not associated with ORR ($p = 0.07$) (Supplementary Figure S5B). Finally, the classification of responses as CBR was studied, but did not yield significant associations, for example, with the total number of events ($p = 0.063$) and gains ($p = 0.088$) (Supplementary Figure S6). In the HGS population ($n = 38$), a higher number of events was significantly associated with longer PFS ($p = 0.041$) (Figure 4C). Although the GI parameters were correlated with the PARPi response in the non-parametric tests, only few parameters showed significance in the univariate survival analysis, and multivariate analysis was not significant. However, the results showed a correlation between higher GI and outcome, which raises the possibility of developing this parameter as a predictive marker.

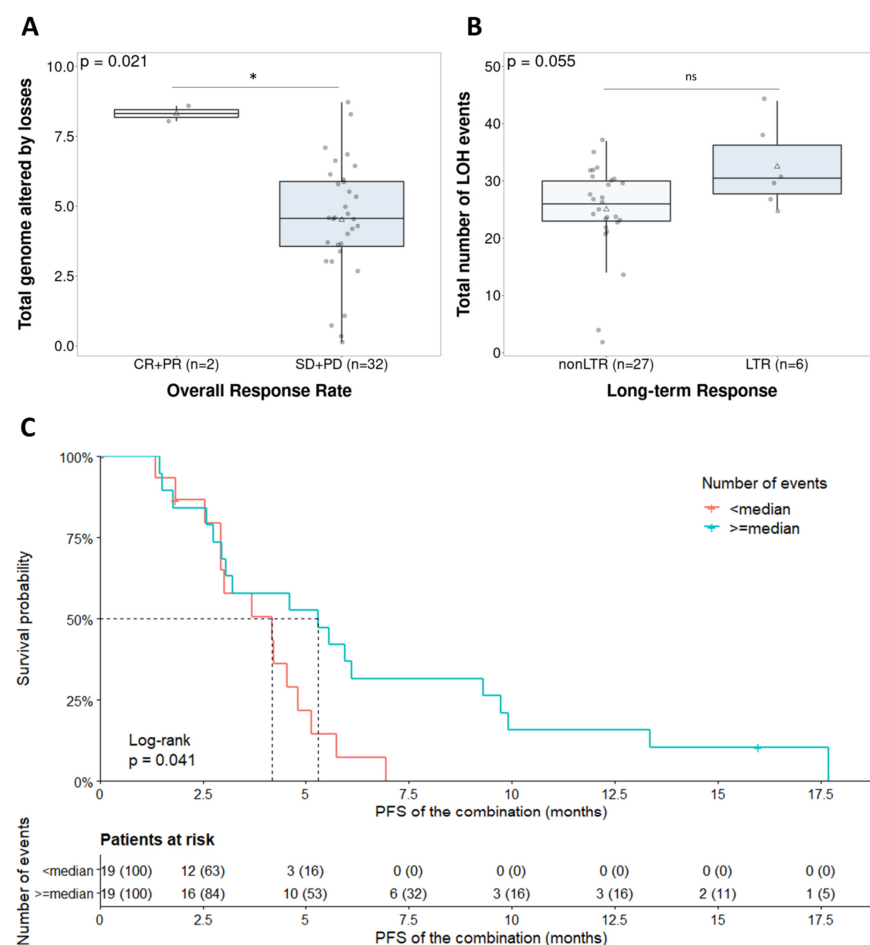


Figure 4. Clinical implication of the genomic instability parameters in the ovarian cancer population ($n = 38$). (A) Total number of events regarding overall response rate ($d = 0.69$); (B) total number of LOHs in long-term responders ($d = 0.31$); (C) survival analysis stratification due to the total number of events. Ns: Not significant; * $p < 0.05$.

4. Discussion

The combination of an inhibitor of DNA damage repair, such as olaparib, with a DNA damaging agent, such as lurbinectedin, is an exciting approach to maximizing the effect of DNA damage. In preclinical models, the combination of olaparib and lurbinectedin has shown a synergistic effect with relevant antitumor activity [21]. However, overlapping toxicities make the combination difficult. Hematological toxicity is the major concern of this combination, since only lurbinectedin as therapy showed grade 3–4 neutropenia up to 85% when administered at a flat dose [19], though this was lower (57%) when the dose was adjusted to body surface area [23]. On the other hand, treatment with chemotherapy and continuous doses of olaparib is usually not feasible due to the high rate of hematologically adverse events. An intermittent schedule of olaparib is better tolerated than a continuous one when combined with chemotherapy [10,11]. In our study, the treatment with lurbinectedin and olaparib was tolerable. Compared with phase 1, no new threats to safety in the expanded phase 2 study were observed. The most common adverse events were hematological (38% of patients had neutropenia grade ≥ 3), and among the nonhematological events, the most common was asthenia, in 8.2%. Although dose modification of at least one drug due to adverse events was common (57.5%), only six patients (8.2%) discontinued treatment due to toxicity.

In the POLA phase 1 dose-escalation trial, we demonstrated that the combination of lurbinectedin adjusted to body surface area and a short course of olaparib had a safe and tolerable profile with an encouraging DCR (stable disease 60%) in a heavily pretreated population [24]. In this phase 2 study, we assessed the efficacy of the RP2D of lurbinectedin (1.5 mg/m² on day 1) with a short course of olaparib (250 mg twice a day on days 1–5) administered every three weeks. To our knowledge, this study is the first phase 2 trial that tests this combination in gynecological malignancies. We showed that this combination provides an ORR of 9.6%, below the pre-specified boundary of efficacy (40%) and even in historical controls (24%). However, DCR was 72.6%, and 12 (16.4%) of the 73 patients treated were considered LTRs, with a median PFS of 13.3 months. In our study, there is particular difficulty in estimating the efficacy across different tumor types and patient characteristics: 61.6% of the patients were heavily pretreated (three or more lines of treatment) since no limit on previous lines of therapy was established. This population had a worse prognosis than the populations included in the recent large phase 3 studies: ovarian cancer was limited to three previous lines, and endometrial cancer was limited to two previous line. Our study subgroup analyses were performed according to histology, and the number of previous lines of therapy showed that patients with three or more previous lines of therapy had a low probability of response (2.2%). In our study, ovarian cancer patients had an ORR of 6.6%, irrespective of histology (HGS or endometrioid) and HR status. Historically, the four drugs (pegylated liposomal doxorubicin, paclitaxel, gemcitabine, and topotecan) most often used as single agents in platinum-resistant ovarian cancer had similar response rates (ranging from 10% to 15%) [28]. In select patients, the addition of bevacizumab to chemotherapy increases their response rates [29], but in most of these trials, the patients had received only one or two lines of previous chemotherapy. In a phase 2 study, lurbinectedin showed significant improvement in ORR compared with topotecan [19]. Regarding PARPi, olaparib monotherapy was approved by the FDA for patients treated with three or more lines based on the results (ORR of 34% and median PFS of 7.9 months) of a series of patients with *BRCA* mutations and platinum-resistant disease [30]. In the ROLANDO trial [31], olaparib combined with pegylated liposomal doxorubicin was assessed in platinum-resistant ovarian cancer and showed an ORR of 29%. However, the number of prior therapy lines was limited to a maximum of four, with at least one previous platinum-sensitive relapse, and *BRCA* mutations were present in 16% of patients compared with 7% of patients in the present study. On the other hand, patients with endometrial carcinoma had an ORR of 15.4%, which is higher than that reported with other PARPis in monotherapy, such as niraparib (4%) [32]. For endometrial cancer, the most active chemotherapeutic agents identified have been doxorubicin and cisplatin, and

both alone or in combination with other agents have been tested in phase III trials, with ORRs ranging from 14.7% to 42% [29,33–36]. However, lurbinectedin plus doxorubicin has only been tested in a small phase I trial with an ORR of 42% [37]. However, the median chemotherapy lines for advanced disease in this population was 1 (range 0–2).

Historically, cancer treatments have been investigated without studying biomarkers of response or fully understanding the mechanisms underlying resistance to the treatment. However, recent trials have evidenced the role specific biomarkers have played in the development of new treatments. The POLA translational study was designed to describe the correlation between different GI parameters and the benefit of the treatment, establishing GI as a predictive biomarker in this clinical scenario. In total, 57 cases were evaluated at the gene and genomic levels, defining those that presented HRD based on the mutational status of HRR genes (10 patients, 17.5%). *BRCA1/2* tumor mutations were present in 10% of cases, which is below the 20% of germline and somatic cases reported in HGS ovarian carcinoma [30]. Due to the nature of the population, with patients who have been previously treated with PARPis and, therefore, potentially mutated and responsive patients being excluded, the incidence of cases with *BRCA* mutations suffered an evident decrease. Regarding the correlation between HRD classification based on mutational status and response, no significant association was found. However, we found a significant correlation between response and different GI parameters, such as loss events, mainly present in the ovarian cancer cohort and HRD population. The lack of predictive power of HRR gene mutations could be explained by differences regarding the characteristics of the population, given that the population is composed principally of patients with HGS ovarian cancer who have been heavily pretreated and a lower frequency of *BRCA* mutations compared with other series. Recent reports have evidenced that mainly cases harboring *BRCA* mutations and, marginally, other HRR genes, such as *RAD51C* conferred sensitivity to PARPi [31,32]. Clinical and methodological issues might also have an impact on the results. For instance, the fact that the genetic and genomic analysis was performed on the primary tumor and not at the moment of relapse, previously to study entry, could affect the concordance between HRD status and treatment response. The mutational/LOH patterns are not reverted when a tumor recovers HR function, so they may not be accurate in predicting PARPi sensitivity in patients who have previously received and progressed on DNA damaging chemotherapy, such as platinum. In addition, the variant selection, which includes in silico prediction of pathogenesis, could have an impact on sensitivity prediction, since some of those mutations may not have a real loss of function and, hence, may not present the HRD phenotype.

The lack of a gold standard for the definition and assessment of GI has motivated a wide number of studies to find an accurate approach [33]. However, only two of them have been commercially approved: Myriad MyChoice[®] and the one from Foundation Medicine[®]. Even if both approaches have been extensively validated [6,31], developing an in-house tool adapted to our requirements and being able to establish the GI based on different parameters were advantages. Additionally, for the mutational analysis, we assessed the whole-genome CNV phenotype and adjusted an in-house pipeline to interrogate and define the GI patterns with regard to the combined treatment response. Hence, we aimed to achieve the most suitable classifier, overcoming the possible caveats of available methodologies. Our approach showed a correlation between different GI parameters and better response to the studied combination. The results concerning the ovarian cancer population were particularly interesting, where a higher percentage of losses ($p = 0.021$) appeared to be correlated with ORR. At the same time, without reaching statistical significance, a trend was observed between the number of LOH events ($p = 0.055$) with LTRs. However, all these results should be carefully considered because of the limited sample size. In addition, the total number of events was also significant in the log-rank test (Figure 4). As several GI parameters are associated with better responses, our next steps will be focused on obtaining a model combining these pre-defined parameters, using response as the endpoint. Then, the predictive role of this GI model will need to be validated in a prospective trial specifically addressing this endpoint.

5. Conclusions

In conclusion, the POLA study provides evidence that the administration of 1.5 mg/m² of lurbinectedin on day 1 and 250 mg of olaparib twice a day on days 1–5 every 21 days is feasible with a DCR of 72.6% and tolerable safety profile in patients who have been heavily pretreated for gynecological cancer. Based on these results, the combination would be suitable for further research and offers a potential alternative for patients with relapsed ovarian and endometrial cancer irrespective of *BRCA* mutation status. This translational study showed a correlation between different GI parameters and a better response; however, its predictive impact should still be investigated in a larger randomized study.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/cancers14040915/s1>, Supplementary Material and Methods; Table S1: Best overall response rate by tumor type; Table S2: Long-term responders; Figure S1: Consort diagram; Figure S2: Kaplan–Meier curves for the whole population (A); Kaplan–Meier curves of progression-free survival according to histology (ovarian vs. endometrium) (B); Swimmer plot of progression-free survival by tumor type (C); Figure S3: Kaplan–Meier curves of overall survival for the whole population (A); Kaplan–Meier curves of overall survival according to histology (ovarian vs. endometrium) (B); Figure S4: Comparison of GI patterns between cancer types. (A) Total number of events, (B) total number of gains, and (C) total number of losses; Figure S5: Clinical implication of GI parameters regarding ORR in (A) the global population and (B) the OC population; Figure S6: Clinical implication of GI parameters in the OC population regarding CBR. (A) Total number of events and (B) total number of gains.

Author Contributions: Conceptualization, A.P., I.R., J.A.L.-G. and A.G.; methodology, A.P., I.R., J.A.L.-G. and A.G.; validation, A.P., I.R., J.A.L.-G., R.L.-R., A.F.-S. and A.G.; formal analysis, A.P., I.R., J.A.L.-G., R.L.-R., A.F.-S. and A.G.; investigation, A.P., R.L.-R., A.O., A.R., M.J.R., E.G., L.F.-M., A.G., V.R.-F., A.F.-S., O.J., I.R., J.A.L.-G.; resources, ALL Authors; data curation, ALL Authors; writing—original draft preparation, O.J.; writing—review and editing, ALL Authors; visualization, ALL Authors; supervision, A.P., I.R., J.A.L.-G., R.L.-R., A.F.-S. and A.G.; project administration, A.P., I.R., J.A.L.-G. and A.G.; funding acquisition, A.P., I.R., J.A.L.-G. and A.G. All authors have read and agreed to the published version of the manuscript.

Funding: This trial was sponsored by AstraZeneca and PharmaMar, including supply of the drugs used in this study.

Institutional Review Board Statement: This study (NCT02684318, EudraCT 2015-001141-08) was approved on 3 October 2015 by a centralized ethics committee (Hospital Universitari Vall d’Hebron) and was conducted following the Declaration of Helsinki, ICH Good Clinical Practice guidelines, and the current legislation.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data is contained within the article or Supplementary Material.

Acknowledgments: Special thanks to Arturo Soto and Guiomar Piera. We are also thankful to all of the patients without whose participation the trial could not have been performed. We also thank Pivotal SLU for its support in monitoring and managing this clinical trial and performing the statistical analysis. We are especially grateful for the work of the study coordinator Pedro Mallol, and the support of the Susana Baviera and Marisa Teruel. Finally, we appreciate the support of Asunción Perales.

Conflicts of Interest: A.P. declares a consulting/advisory role for AstraZeneca, PharmaMar, Roche, Clovis Oncology, Tesaro, and MSD and support for travel and accommodations from PharmaMar. R.L.-R. declares no competing interests. A.O. has served on advisory boards for AstraZeneca Farmaceutica Spain, SA; AstraZeneca KK; Clovis Oncology, Inc.; Corcept Therapeutics; Deciphera Pharmaceutical; Eisai Europe Limited; EMD Serono, Inc.; F. Hoffmann-La Roche; GlaxoSmithKline; GlaxoSmithKline, LLC.; Got It Consulting, SL; Immunogen, SL; KL Logistics; Medison Pharma; Merck Sharp & Dohme de España; Mersana Therapeutics; Novocure GmbH; Pharma Mar; prIME Oncology; ROCHE FARMA; Sattucklabs; Sutro Biopharma, Inc.; TESARO Bio GmbH; Tesaro Bio Spain, SLU; and TESARO Inc. A.R. reports honoraria and advisory/consultancy (MSD, AstraZeneca,

Roche, GSK, Clovis, PharmaMar, Lilly, and Amgen), research grant/funding to his institution (Eisai, PharmaMar, and Roche), travel/accommodation/expenses (AstraZeneca, Tesaro: A GSK Company, PharmaMar, and Roche), and speakers bureau (MSD, AstraZeneca, Roche, GSK, Clovis, and PharmaMar) outside the submitted work. M.J.R. declares a consulting/advisory role for AstraZeneca, PharmaMar, Roche, Clovis Oncology, GSK, and MSD and support for travel and accommodations from PharmaMar, Roche, GSK, and AstraZeneca-MSD. E.G. has received advisory/consultancy honoraria from AstraZeneca-MSD, Clovis Oncology, GSK-Tesaro, and PharmaMar, Roche; speaker bureau/expert testimony honoraria from AstraZeneca, MSD, PharmaMar, Roche, GSK-Tesaro, and Clovis; and travel/accommodation/expenses from Roche, TESARO: A GSK Company, and Baxter. L.F.-M. declares an advisory/consultancy role for TESARO and GSK and speaking fees from AstraZeneca, Roche, and MSD. A.G. reports having received honoraria (Clovis, MSD, AstraZeneca, GSK, PharmaMar, and Roche) and travel/accommodation/expenses (Merck Sharp & Dohme, PharmaMar, Roche, Eisai, Pfizer, Pierre-Fabre, and Tesaro-A GSK Company). V.R.-F. has no conflicts of interest to declare. A.F.-S. declares no competing interests. O.J. declares being an employee of Pivotal SLU. I.R. declares personal honoraria from Roche, PharmaMar, Tesaro, AstraZeneca, and Clovis. J.A.L.-G. declares an advisory/consultancy role for AstraZeneca and MSD; speaker fees from AstraZeneca and Sophia Genetics; and research funding (to the institution) from Generalitat Valenciana, Instituto de Salud Carlos III, PharmaMar, and AstraZeneca.

Abbreviations

BID: twice a day; CBR: clinical benefit rate; CI: confidence interval; CN: copy number; CVN: copy number variation; CR: complete response; DCR: disease control rate; DNA: deoxyribonucleic acid; ECG: electrocardiogram; ECOG: Eastern Cooperative Group; FDA: Food and Drug Administration; FFPE: formalin-fixed paraffin-embedded; GI: genomic instability; HGS: high-grade serous; HRD: homologous recombination deficiency; HRR: homologous recombination repair; ICH: International Conference on Harmonisation; LOH: loss of heterozygosity; LTR: long-term responders; MLPA: multiplex ligation-dependent probe amplification; ORR: overall response rate; PARPi: poly (ADP-ribose) polymerase inhibitors; OS: overall survival; PD: progression disease; PFS: progression-free survival; PG: pharmacogenomics; PR: partial response; PS: performance status; RECIST: Response Evaluation Criteria in Solid Tumors; RP2D: recommended phase 2 dose; SD: stable disease; SNVs: single nucleotide variants; TEAE: treatment-emergent adverse event; ULN: upper limit of normal.

References








1. Fong, P.C.; Boss, D.S.; Yap, T.A.; Tutt, A.; Wu, P.; Mergui-Roelvink, M.; Mortimer, P.; Swaisland, H.; Lau, A.; O'Connor, M.J.; et al. Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers. *N. Engl. J. Med.* **2009**, *361*, 123–134. [\[CrossRef\]](#)
2. Gelmon, K.A.; Tischkowitz, M.; Mackay, H.; Swenerton, K.; Robidoux, A.; Tonkin, K.; Hirte, H.; Huntsman, D.; Clemons, M.; Gilks, B.; et al. Olaparib in patients with recurrent high-grade serous or poorly differentiated ovarian carcinoma or triple-negative breast cancer: A phase 2, multicentre, open-label, non-randomised study. *Lancet Oncol.* **2011**, *12*, 852–861. [\[CrossRef\]](#)
3. Franzese, E.; Centonze, S.; Diana, A.; Carlino, F.; Guerrera, L.P.; Di Napoli, M.; De Vita, F.; Pignata, S.; Ciardiello, F.; Orditura, M. PARP inhibitors in ovarian cancer. *Cancer Treat. Rev.* **2019**, *73*, 1–9. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Pujade-Lauraine, E.; Ledermann, J.A.; Selle, F.; GebSKI, V.; Penson, R.T.; Oza, A.M.; Korach, J.; Huzarski, T.; Poveda, A.; Pignata, S.; et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): A double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Oncol.* **2017**, *18*, 1274–1284. [\[CrossRef\]](#)
5. Mirza, M.R.; Monk, B.J.; Herrstedt, J.; Oza, A.M.; Mahner, S.; Redondo, A.; Fabbro, M.; Ledermann, J.A.; Lorusso, D.; Vergote, I.; et al. Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. *N. Engl. J. Med.* **2016**, *375*, 2154–2164. [\[CrossRef\]](#)
6. Coleman, R.L.; Oza, A.M.; Lorusso, D.; Aghajanian, C.; Oaknin, A.; Dean, A.; Colombo, N.; Weberpals, J.I.; Clamp, A.; Scambia, G.; et al. Rucaparib maintenance treatment for recurrent ovarian carcinoma after response to platinum therapy (ARIEL3): A randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **2017**, *390*, 1949–1961. [\[CrossRef\]](#)
7. Ledermann, J.A.; Pujade-Lauraine, E. Olaparib as maintenance treatment for patients with platinum-sensitive relapsed ovarian cancer. *Ther. Adv. Med. Oncol.* **2019**, *11*. [\[CrossRef\]](#)
8. Moore, K.; Colombo, N.; Scambia, G.; Kim, B.G.; Oaknin, A.; Friedlander, M.; Lisysanskaya, A.; Floquet, A.; Leary, A.; Sonke, G.S.; et al. Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer. *N. Engl. J. Med.* **2018**, *379*, 2495–2505. [\[CrossRef\]](#)

9. Matulonis, U.A.; Monk, B.J. PARP inhibitor and chemotherapy combination trials for the treatment of advanced malignancies: Does a development pathway forward exist? *Ann. Oncol.* **2017**, *28*, 443–447. [[CrossRef](#)]
10. Oza, A.M.; Cibula, D.; Benzaquen, A.O.; Poole, C.; Mathijssen, R.H.J.; Sonke, G.S.; Colombo, N.; Špaček, J.; Vuylsteke, P.; Hirte, H.; et al. Olaparib combined with chemotherapy for recurrent platinum-sensitive ovarian cancer: A randomised phase 2 trial. *Lancet Oncol.* **2015**, *16*, 87–97. [[CrossRef](#)]
11. van der Noll, R.; Jager, A.; Ang, J.E.; Marchetti, S.; Mergui-Roelvink, M.W.J.; de Bono, J.S.; Lolkema, M.P.; de Jonge, M.J.A.; van der Biessen, D.A.; Brunetto, A.T.; et al. Phase I study of intermittent olaparib capsule or tablet dosing in combination with carboplatin and paclitaxel (part 2). *Investig. New Drugs* **2019**, *38*, 1096–1107. [[CrossRef](#)]
12. Lampert, E.J.; Hays, J.L.; Kohn, E.C.; Annunziata, C.M.; Minasian, L.; Yu, M.; Gordon, N.; Sissung, T.M.; Chiou, V.L.; Figg, W.D.; et al. Phase I/Ib study of olaparib and carboplatin in heavily pretreated recurrent high-grade serous ovarian cancer at low genetic risk. *Oncotarget* **2019**, *10*, 2855–2868. [[CrossRef](#)] [[PubMed](#)]
13. Del Conte, G.; Sessa, C.; Von Moos, R.; Viganò, L.; Digena, T.; Locatelli, A.; Gallerani, E.; Fasolo, A.; Tessari, A.; Cathomas, R.; et al. Phase I study of olaparib in combination with liposomal doxorubicin in patients with advanced solid tumours. *Br. J. Cancer* **2014**, *111*, 651–659. [[CrossRef](#)]
14. Gray, H.J.; Bell-McGuinn, K.; Fleming, G.F.; Cristea, M.; Xiong, H.; Sullivan, D.; Luo, Y.; McKee, M.D.; Munasinghe, W.; Martin, L.P. Phase I combination study of the PARP inhibitor veliparib plus carboplatin and gemcitabine in patients with advanced ovarian cancer and other solid malignancies. *Gynecol. Oncol.* **2018**, *148*, 507–514. [[CrossRef](#)] [[PubMed](#)]
15. Coleman, R.L.; Fleming, G.F.; Brady, M.F.; Swisher, E.M.; Steffensen, K.D.; Friedlander, M.; Okamoto, A.; Moore, K.N.; Efrat Ben-Baruch, N.; Werner, T.L.; et al. Veliparib with First-Line Chemotherapy and as Maintenance Therapy in Ovarian Cancer. *N. Engl. J. Med.* **2019**, *381*, 2403–2415. [[CrossRef](#)] [[PubMed](#)]
16. Leal, J.F.M.M.; Martínez-Díez, M.; García-Hernández, V.; Moneo, V.; Domingo, A.; Bueren-Calabuig, J.A.; Negri, A.; Gago, F.; Guillén-Navarro, M.J.; Avilés, P.; et al. PM01183, a new DNA minor groove covalent binder with potent in vitro and in vivo anti-tumour activity. *Br. J. Pharmacol.* **2010**, *161*, 1099–1110. [[CrossRef](#)] [[PubMed](#)]
17. Belgiovine, C.; Bello, E.; Liguori, M.; Craparotta, I.; Mannarino, L.; Paracchini, L.; Beltrame, L.; Marchini, S.; Galmarini, C.M.; Mantovani, A.; et al. Lurbinectedin reduces tumour-associated macrophages and the inflammatory tumour microenvironment in preclinical models. *Br. J. Cancer* **2017**, *117*, 628–638. [[CrossRef](#)]
18. Trigo, J.; Subbiah, V.; Besse, B.; Moreno, V.; López, R.; Sala, M.A.; Peters, S.; Ponce, S.; Fernández, C.; Alfaro, V.; et al. Lurbinectedin as second-line treatment for patients with small-cell lung cancer: A single-arm, open-label, phase 2 basket trial. *Lancet Oncol.* **2020**, *21*, 645–654. [[CrossRef](#)]
19. Poveda, A.; Del Campo, J.M.; Ray-Coquard, I.; Alexandre, J.; Provansal, M.; Guerra Alía, E.M.; Casado, A.; Gonzalez-Martin, A.; Fernández, C.; Rodriguez, I.; et al. Phase II randomized study of PM01183 versus topotecan in patients with platinum-resistant/refractory advanced ovarian cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2017**, *28*, 1280–1287. [[CrossRef](#)]
20. Gaillard, S.; Oaknin, A.; Ray-Coquard, I.; Vergote, I.; Scambia, G.; Colombo, N.; Fernandez, C.; Alfaro, V.; Kahatt, C.; Nieto, A.; et al. Lurbinectedin versus pegylated liposomal doxorubicin or topotecan in patients with platinum-resistant ovarian cancer: A multicenter, randomized, controlled, open-label phase 3 study (CORAIL). *Gynecol. Oncol.* **2021**, *163*, 237–245. [[CrossRef](#)]
21. Ordóñez, J.L.; Amaral, A.T.; Carcaboso, A.M.; Herrero-Martín, D.; Del Carmen García-Macías, M.; Sevillano, V.; Alonso, D.; Pascual-Pasto, G.; San-Segundo, L.; Vila-Ubach, M.; et al. The PARP inhibitor olaparib enhances the sensitivity of Ewing sarcoma to trabectedin. *Oncotarget* **2015**, *6*, 18875–18890. [[CrossRef](#)] [[PubMed](#)]
22. Ávila-Arroyo, S.; Nuñez, G.S.; García-Fernández, L.F.; Galmarini, C.M. Synergistic Effect of Trabectedin and Olaparib Combination Regimen in Breast Cancer Cell Lines. *J. Breast Cancer* **2015**, *18*, 329–338. [[CrossRef](#)] [[PubMed](#)]
23. Cruz, C.; Llop-Guevara, A.; Garber, J.E.; Arun, B.K.; Perez Fidalgo, J.A.; Lluch, A.; Telli, M.L.; Fernández, C.; Kahatt, C.; Galmarini, C.M.; et al. Multicenter phase II study of lurbinectedin in BRCAMutated and unselected metastatic advanced breast cancer and biomarker assessment substudy. *J. Clin. Oncol.* **2018**, *36*, 3134–3143. [[CrossRef](#)] [[PubMed](#)]
24. Poveda, A.; Oaknin, A.; Romero, I.; Guerrero-Zotano, A.; Fariñas-Madrid, L.; Rodriguez-Freixinos, V.; Mallo, P.; Lopez-Reig, R.; Lopez-Guerrero, J.A. A phase I dose-finding, pharmacokinetics and genotyping study of olaparib and lurbinectedin in patients with advanced solid tumors. *Sci. Rep.* **2021**, *11*, 4433. [[CrossRef](#)] [[PubMed](#)]
25. Povysil, G.; Tzika, A.; Vogt, J.; Haunschmid, V.; Messiaen, L.; Zschocke, J.; Klambauer, G.; Hochreiter, S.; Wimmer, K. pan-eln.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics. *Hum. Mutat.* **2017**, *38*, 889–897. [[CrossRef](#)]
26. Talevich, E.; Shain, A.H.; Botton, T.; Bastian, B.C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput. Biol.* **2016**, *12*, e1004873. [[CrossRef](#)]
27. Telli, M.L.; Timms, K.M.; Tedi, J.; Hennessy, B.; Mills, G.B.; Jensen, K.C.; Szallasi, Z.; Barry, W.T.; Winer, E.P.; Tung, N.M.; 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*, 3764–3773. [[CrossRef](#)]
28. Davis, A.; Tinker, A.V.; Friedlander, M. “Platinum resistant” ovarian cancer: What is it, who to treat and how to measure benefit? *Gynecol. Oncol.* **2014**, *133*, 624–631. [[CrossRef](#)]
29. Pujade-Lauraine, E.; Hilpert, F.; Weber, B.; Reuss, A.; Poveda, A.; Kristensen, G.; Sorio, R.; Vergote, I.; Witteveen, P.; Bamias, A.; et al. Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: The AURELIA open-label randomized phase III trial. *J. Clin. Oncol.* **2014**, *32*, 1302–1308. [[CrossRef](#)]

30. Kaufman, B.; Shapira-Frommer, R.; Schmutzler, R.K.; Audeh, M.W.; Friedlander, M.; Balmaña, J.; Mitchell, G.; Fried, G.; Stemmer, S.M.; Hubert, A.; et al. Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation. *J. Clin. Oncol.* **2015**, *33*, 244–250. [[CrossRef](#)]
31. Perez-Fidalgo, J.A.; Cortés, A.; Guerra, E.; García, Y.; Iglesias, M.; Bohn Sarmiento, U.; Calvo García, E.; Manso Sánchez, L.; Santaballa, A.; Oaknin, A.; et al. Olaparib in combination with pegylated liposomal doxorubicin for platinum-resistant ovarian cancer regardless of BRCA status: A GEICO phase II trial (ROLANDO study). *ESMO Open* **2021**, *6*, 100212. [[CrossRef](#)] [[PubMed](#)]
32. Madariaga, A.; Garg, S.; Tchrakian, N.; Dhani, N.C.; Jimenez, W.; Welch, S.; Mackay, H.; Ethier, J.-L.; Gilbert, L.; Rodriguez, A.; et al. Phase II trial assessing niraparib with or without dostarlimab (anti-PD-1) in recurrent endometrial carcinoma. *J. Clin. Oncol.* **2021**, *39*, 5574. [[CrossRef](#)]
33. Tate Thigpen, J.; Brady, M.F.; Homesley, H.D.; Malfetano, J.; Dubeshter, B.; Burger, R.A.; Liao, S.; Mackey, D. Phase III trial of doxorubicin with or without cisplatin in advanced endometrial carcinoma: A gynecologic oncology group study. *J. Clin. Oncol.* **2004**, *22*, 3902–3908. [[CrossRef](#)]
34. McMeekin, S.; Dizon, D.; Barter, J.; Scambia, G.; Manzyuk, L.; Lisyanskaya, A.; Oaknin, A.; Ringuette, S.; Mukhopadhyay, P.; Rosenberg, J.; et al. Phase III randomized trial of second-line ixabepilone versus paclitaxel or doxorubicin in women with advanced endometrial cancer. *Gynecol. Oncol.* **2015**, *138*, 18–23. [[CrossRef](#)] [[PubMed](#)]
35. Makker, V.; Colombo, N.; Herráez, A.C.; Santin, A.; Colomba, E.; Miller, D.; Fujiwara, K.; Pignata, S.; Baron-Hay, S.; Ray-Coquard, I.; et al. A multicenter, open-label, randomized, phase III study to compare the efficacy and safety of lenvatinib in combination with pembrolizumab versus treatment of physician’s choice in patients with advanced endometrial cancer. *Gynecol. Oncol.* **2021**, *162*, S4. [[CrossRef](#)]
36. Moore, K.N.; Oza, A.M.; Colombo, N.; Oaknin, A.; Scambia, G.; Lorusso, D.; Konecny, G.E.; Banerjee, S.; Murphy, C.G.; Tanyi, J.L.; et al. Phase III, randomized trial of mirvetuximab soravtansine versus chemotherapy in patients with platinum-resistant ovarian cancer: Primary analysis of FORWARD I. *Ann. Oncol.* **2021**, *32*, 757–765. [[CrossRef](#)]
37. Kristeleit, R.; Moreno, V.; Boni, V.; Guerra, E.M.; Kahatt, C.; Romero, I.; Calvo, E.; Basté, N.; López-Vilariño, J.A.; Siguero, M.; et al. Doxorubicin plus lurbinectedin in patients with advanced endometrial cancer: Results from an expanded phase I study. *Int. J. Gynecol. Cancer* **2021**, *31*, 1428–1436. [[CrossRef](#)]

Article

The Scarface Score: Deciphering Response to DNA Damage Agents in High-Grade Serous Ovarian Cancer—A GEICO Study

Antonio Fernández-Serra ^{1,2,†} , Raquel López-Reig ^{1,2,†} , Raúl Márquez ³, Alejandro Gallego ⁴,
Luis Miguel de Sande ⁵, Alfonso Yubero ⁶, Cristina Pérez-Segura ⁷ , Avinash Ramchandani-Vaswani ⁸,
María Pilar Barretina-Ginesta ⁹, Elsa Mendizábal ¹⁰, Carmen Esteban ¹¹, Fernando Gálvez ¹² ,
Ana Beatriz Sánchez-Heras ¹³, Eva María Guerra-Alía ¹⁴, Lydia Gaba ¹⁵ , María Quindós ¹⁶, Isabel Palacio ¹⁷,
Jesús Alarcón ¹⁸, Ana Oaknin ¹⁹, Jessica Aliaga ²⁰, Marta Ramírez-Calvo ¹, Zaida García-Casado ¹ ,
Ignacio Romero ²¹ and José Antonio López-Guerrero ^{1,2,22,*} 

- ¹ Molecular Biology Lab, Molecular Biology Department, Instituto Valenciano de Oncología, 46009 Valencia, Spain
- ² Joint IVO-CIPF Cancer Research Unit, 46012 Valencia, Spain
- ³ Medical Oncology Department, MD Anderson Cancer Center, 28033 Madrid, Spain; raulmarquez@mdanderson.es
- ⁴ Medical Oncology Department, Hospital Universitario La Paz, 28046 Madrid, Spain
- ⁵ Medical Oncology Department, Hospital Universitario de León, 24008 León, Spain
- ⁶ Medical Oncology Department, Hospital Clínico Universitario Lozano Blesa, 50009 Zaragoza, Spain
- ⁷ Medical Oncology Department, Hospital de Sant Pau i Santa Tecla, 08025 Barcelona, Spain
- ⁸ Medical Oncology Department, Hospital Universitario Insular de Gran Canaria, 35016 Gran Canaria, Spain
- ⁹ Medical Oncology Department, Institut Català d'Oncologia Girona, 17007 Girona, Spain
- ¹⁰ Medical Oncology Department, Hospital General Universitario Gregorio Marañón, 28007 Madrid, Spain
- ¹¹ Medical Oncology Department, Hospital Virgen de la Salud, 45005 Toledo, Spain
- ¹² Medical Oncology Department, Complejo Hospitalario de Jaén, 23007 Jaén, Spain
- ¹³ Medical Oncology Department, Hospital General Universitario de Elche, 03203 Elche, Spain
- ¹⁴ Medical Oncology Department, Hospital Universitario Ramón y Cajal, 28034 Madrid, Spain
- ¹⁵ Medical Oncology Department, Hospital Clínic de Barcelona, 08036 Barcelona, Spain
- ¹⁶ Medical Oncology Department, Complejo Hospitalario Universitario A Coruña, 15006 A Coruña, Spain
- ¹⁷ Medical Oncology Department, Hospital Central Asturias, 33011 Oviedo, Spain
- ¹⁸ Medical Oncology Department, Hospital Universitario Son Espases, 07120 Palma de Mallorca, Spain
- ¹⁹ Medical Oncology Department, Hospital Universitari Vall d'Hebron, 08035 Barcelona, Spain
- ²⁰ Pathology Department, Instituto Valenciano de Oncología, 46009 Valencia, Spain
- ²¹ Medical Oncology Department, Instituto Valenciano de Oncología, 46010 Valencia, Spain
- ²² Department of Pathology, Catholic University of Valencia, 46001 Valencia, Spain
- * Correspondence: jalopez@fivo.org; Tel.: +34-961-114-337 or +34-961-104-039
- † These authors contributed equally to this work.



Citation: Fernández-Serra, A.; López-Reig, R.; Márquez, R.; Gallego, A.; de Sande, L.M.; Yubero, A.; Pérez-Segura, C.; Ramchandani-Vaswani, A.; Barretina-Ginesta, M.P.; Mendizábal, E.; et al. The Scarface Score: Deciphering Response to DNA Damage Agents in High-Grade Serous Ovarian Cancer—A GEICO Study. *Cancers* **2023**, *15*, 3030. <https://doi.org/10.3390/cancers15113030>

Academic Editors: Luis Montuenga, Elena Castro and Antonio Maraver

Received: 20 April 2023

Revised: 26 May 2023

Accepted: 30 May 2023

Published: 1 June 2023



Copyright: © 2023 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/>).

Simple Summary: The response of high-grade serous ovarian cancer (HGSOC) to DNA-damaging agents largely depends on tumor genomic instability (GI), a phenomenon that affects the entire genome. Nowadays, surrogate biomarkers of this phenomenon, such as *BRCA*-gene mutations, are used in clinical practice to identify patients harboring this characteristic. However, these approaches do not capture the entire picture of GI, mainly due to the lack of information on non-*BRCA* mutation causes and hence, leading to the misclassification of patients. Thus, considering the great interest in studying GI from a comprehensive perspective, this study aims to establish an integrative response-predictive classifier (Scarface Score) for DNA-damaging agents in the context of HGSOC. The Scarface score will support clinical decision-making by correctly selecting the subpopulation of patients with better responses and avoiding overtreatment of those with a low Scarface Score.

Abstract: Genomic Instability (GI) is a transversal phenomenon shared by several tumor types that provide both prognostic and predictive information. In the context of high-grade serous ovarian cancer (HGSOC), response to DNA-damaging agents such as platinum-based and poly(ADP-ribose) polymerase inhibitors (PARPi) has been closely linked to deficiencies in the DNA repair machinery by homologous recombination repair (HRR) and GI. In this study, we have developed the Scarface score, an integrative algorithm based on genomic and transcriptomic data obtained from the NGS analysis

of a prospective GEICO cohort of 190 formalin-fixed paraffin-embedded (FFPE) tumor samples from patients diagnosed with HGSOC with a median follow up of 31.03 months (5.87–159.27 months). In the first step, three single-source models, including the SNP-based model (accuracy = 0.8077), analyzing 8 SNPs distributed along the genome; the GI-based model (accuracy = 0.9038) interrogating 28 parameters of GI; and the HTG-based model (accuracy = 0.8077), evaluating the expression of 7 genes related with tumor biology; were proved to predict response. Then, an ensemble model called the Scarface score was found to predict response to DNA-damaging agents with an accuracy of 0.9615 and a kappa index of 0.9128 ($p < 0.0001$). The Scarface Score approaches the routine establishment of GI in the clinical setting, enabling its incorporation as a predictive and prognostic tool in the management of HGSOC.

Keywords: high-grade serous ovarian cancer; genomic instability; machine learning; PARPi; platinum-based chemotherapy

1. Introduction

The term ‘genomic instability’ (GI) describes the characteristic of cells to progressively accumulate genomic alterations. In recent years, because of its increasing importance in the field of oncology, GI has gained greater attention in translational research [1]. GI is a hallmark of cancer and is relevant not only as an intrinsic feature of tumor cells but also as a potential driving force of tumorigenesis [2]. Although GI is present in every cancer type, some tumors show a remarkable accumulation of alterations [3]. High-grade serous ovarian cancer (HGSOC) is of particular interest in this respect. HGSOC is a molecularly and clinically heterogeneous disease that is characterized by TP53 mutations and DNA damage homologous recombination repair (HRR) deficiency (HRD) in approximately 50% of patients [4]. Deficiencies in this pathway could have different molecular causes in addition to classically known *BRCA1/2* mutation, such as other HRR-genes mutations and epigenetic modifications [5]. The HRD phenotype represents a clear molecular subtype that is highly enriched in copy number alteration patterns, which play important roles in oncogenesis, progression, and metastasis [2,6]. The so-called HRD phenotype is defined as a clinical profile similar to tumors harboring *BRCA* gene alterations. That is, showing a higher progression-free survival treated mainly with platinum salts and PARP inhibitors, among other therapies [7]. Copy number alteration patterns can be classified by the presence of specific GI events, also called genomic scars, reflecting a loss of genome integrity [8]. These genomic scars may be reliable biomarkers for homologous recombination repair deficiency (HRD) and could potentially be used to identify patients who would benefit from specific types of anticancer therapies, such as platinum-based chemotherapies or poly(ADP-ribose) polymerase inhibitor (PARPi) therapy [9–11]—the clinical utility of which has been shown in several clinical trials, including PAOLA [12], PRIMA [13], VELIA [14] and ATHENA [15]. As such, GI is a potential predictive and prognostic biomarker [6]. Because of these clinical implications, researchers are attempting to define GI status in order to select patients who will benefit from these therapeutic approaches.

Classically, the determination of HRD status has relied on *BRCA1* and *BRCA2* genotyping [16], but the HRR pathway involves a vast range of proteins, most of which are reportedly mutated in tumor samples [17]. Today, the development of high-throughput techniques allows the integrative analysis of multiomic data to generate machine learning models, which can more comprehensively determine HRD status [18,19].

Based on the above, the aim of this study was to develop a methodologic and analytic approach to determining GI status in patients with HGSOC using a comprehensive strategy that integrates data from single-nucleotide variations, somatic copy number alterations, and transcriptomics. These data were used to build a model (the Scarface score) that could predict a patient’s response to DNA-damaging agents.

2. Materials and Methods

2.1. Patient Selection

The study used 190 formalin-fixed and paraffin-embedded (FFPE) HGSOc samples that were ambispectively collected from patients treated at multiple centers from 2007 to 2020 (BorNeO 1703). An ambispective study implies the combination of both retrospective and prospective data, including past, present, and future time points. All patients signed an informed consent form approved by the required ethics committees, and the study was approved by the ethics committee of Fundación Instituto Valenciano de Oncología in 2021 (LBM-02-20, SCARFACE). The informed consent of patients was obtained following institutional, ethical, and legal regulations. The inclusion criteria were age ≥ 18 years at inclusion, diagnosis with HGSOc, and previous first-line treatment with platinum-based chemotherapy.

2.2. Mutational and Copy Number Variants Analysis

DNA extraction was performed using three 20 μm -thick sections of FFPE tumor blocks and a QIAamp DNA FFPE tissue kit (Qiagen, Hilden, Germany). The final concentration was measured spectrophotometrically using NanoDrop ND-1000 (Eppendorf, Hamburg, Germany). Genomic concentration, DNA integrity, and fragment size were determined by using a TapeStation 4200 bioanalyzer (Agilent, Santa Clara, CA, USA).

Libraries were prepared using the SureSelectXT HS Target Enrichment Kit using the Magnis NGS Prep System (Agilent, Santa Clara, CA, USA). Briefly, 200 ng of extracted DNA was enzymatically fragmented to a size range of 150–200 base pairs. Each library was then hybridized with a SureSelectXT HS custom panel combined with Agilent OneSeq backbone 1 Mb according to the manufacturer's protocol. The custom panel analyzed the following DNA damage response genes: *BRCA1*, *BRCA2*, *BARD1*, *BRIPI1*, *CHEK1*, *CHEK2*, *FAM175A*, *NBN*, *PALB2*, *ATM*, *MRE11A*, *RAD51B*, *RAD51C*, *RAD51D*, *RAD54L*, *FANCI*, *FANCM*, *FANCA*, *ERCC1*, *ERCC2*, *ERCC6*, *REQL*, *XRCC4*, *HELQ*, *SLX4*, *WRN*, *ATR*, *PTEN*, *CCNE1*, *EMSY*, *TP53*, *MLH1*, *MSH2*, *MSH6*, and *PMS2*.

Although HRR genes were overrepresented in the panel, genes belonging to the base excision repair, nucleotide excision repair, and mismatch repair pathways were also incorporated into the design. The OneSeq backbone was used to obtain copy number variants (CNVs), consisting of 147,000 single-nucleotide polymorphisms (SNPs) homogeneously distributed along the genome. Pooled libraries were sequenced (100 bp paired-end) using the NextSeq 550 System (Illumina, San Diego, CA, USA). A secondary analysis was performed using HaplotypeCaller (Broad Institute, Cambridge, MA, USA) for variant calling and VariantStudio 4.0 for annotation (Illumina). Variants were selected after a filtering process based on the following analytical parameters: coverage $>100\times$ (covered in forward and reverse sense); allele frequency $>5\%$; and annotation of Pathogenic, likely pathogenic, or VUS with a prediction of pathogenicity with Varsome classifier. Germline *BRCA1/2* alterations were obtained from analyses carried out at each hospital of origin.

Bioinformatics analysis to obtain copy number events was performed using an in-house pipeline based on the CNVkit algorithm [20]. This pipeline was internally customized to ensure the suitability and reliability of the method (Supplementary Data S1 and Figures S1–S5). The CNVkit algorithm uses sequencing data from target and anti-target regions to infer copy number status. Circular binary segmentation was chosen for the segmentation step. The variant calling step was performed using Mutect2 (Broad Institute). Normalization was applied by using median read counts from a set of 10 control samples from healthy peritumoral ovarian tissue.

Independently, the panelcn.MOPS package (version 1.17.1) [21] was used to evaluate copy number changes at the gene level—particularly *CCNE1* amplification.

The presence of HRD-associated genomic scars (loss of heterozygosity (LOH), large-scale transitions, number of telomeric allelic imbalances, and a combined score (HRD score)) was assessed using the scarHRD package (version 0.1.1) for R [22].

The parameter settings and codes used for GI determination with CNVkit software and the script to extract analytical features are available at https://github.com/afernandezse/Pola_Phase2_GI_traslational (accessed 22 January 2022).

2.3. Transcriptomic Analysis

Gene expression analysis was performed using the HTG EdgeSeq System (HTG Molecular Diagnostics, Tucson, AZ, USA). This technique is based on RNA sequencing consisting of a prehybridization step with specific probes using a quantitative nuclease protection assay, followed by a standard next-generation sequencing (NGS) protocol. This technique requires a small input (i.e., 5 μ m FFPE section and an area of 15 mm²). The panel focuses on a selection of 2549 oncology-related mRNAs (the Oncology Biomarker Panel) rather than analyzing the entire transcriptome, obtaining the appropriate dynamic range in gene expression analysis. Gene expression data were parsed using HTG EdgeSeq Parser version 5.3.0.7184. Quality control was performed using HTG Reveal version 3.0 (HTG Molecular Diagnostics). Raw read counts were normalized according to the median [23].

2.4. Model Fitting

To improve the current detection of HRD-related GI, a data-mining model integrating several biological approaches was proposed. The model included genomic and transcriptomic data from 190 HGSOc samples, from which all data were available for 183 samples. The first layer of the model comprised 147,000 SNPs uniformly distributed along the entire genome at a resolution of 1 Mb. The second layer, comprised of GI parameters, was derived from CNVkit results. Finally, gene expression data obtained from targeted RNA sequencing of 2549 genes was the third layer. Because of the high number of SNP parameters, those that were less informative were removed under the criteria of a low number or near zero variance in total counts per SNP.

Briefly, the model fitting on the first and third layers consisted of three parts. First, feature selection was performed by extracting attributes using the ANOVA test, the signal-to-noise ratio, significant parameters identified from logistic regression analysis, recursive feature extraction [24], and the Boruta algorithm [25]. Second, model feeding was conducted. Each resulting set of features was tested to build three data-mining models using the following algorithms: support vector machine, random forest, and neural network (Supplementary Data S2). Third, specific hyperparameters were tuned. Second-layer building followed the same procedure but without feature extraction.

The final model consisted of an ensemble model (which was termed the Scarface score), in which the best-performing data-mining model was fed with its paired selected parameters. This model was benchmarked by studying its mean accuracy and kappa index from 500 bootstrapping iterations (detailed in Supplementary Data S2). Each model, including the ensemble model (the Scarface score), was trained and validated using two series, which were randomly selected from the total 183 HGSOc samples in a proportion of 70/30, respectively. The models were trained to discriminate between patients with a response to platinum-based chemotherapy \geq 12 months (responders) versus <12 months (non-responders).

2.5. Statistical Analysis

The chi-square and Fisher's exact tests were used to compare categorical GI and clinical and pathological variables. Non-parametric Wilcoxon and Kruskal–Wallis tests were used for continuous variables.

For time-to-event variables, survival analysis was performed using Kaplan–Meier estimation, and significance was obtained by log-rank testing. Univariate and multivariate Cox regression was also performed. Statistical significance was considered at $p < 0.05$. All tests were two-tailed. The time-to-event variables investigated were platinum-free interval (PFI), defined as the time between the end of platinum-based chemotherapy and relapse; progression-free survival (PFS) to PARPi, defined as the time between the start of PARPi treatment and disease progression; and overall survival (OS), defined as the time between diagnosis and death.

The performance of the models was evaluated using the ROCR and pROC packages from R version 4.1.2. Statistical analyses were performed using R studio version 2021.09.0.

3. Results

3.1. Study Population

FFPE tumor blocks from 190 patients with HGSOC were analyzed. Clinical parameters of the patient population are shown in Table 1. The median follow-up of the studied population was 31.03 months (range 5.87–159.27 months). Median PFI after first-line therapy was 16.28 months (range 0–83.33 months), the recurrence rate after first-line therapy was 52.11% (99/190), and the median PFS to PARPi was 11.03 months (range 1.03–64.63 months). Overall, 20.53% of patients had died at the time of data analysis.

Table 1. Main clinical, pathological, and treatment-related variables of the whole series.

Clinical Parameter		N	%	Clinical Parameter		N	%	
Histology	High-grade serous ovarian cancer	190	100	Surgery	Yes	167	87.9	
	IA	7	3.7		No	23	12.1	
Stage	IC1	6	3.2	Primary debulking surgery	Yes	114	68.3	
	IC2	9	4.7		No	53	31.7	
	IIA	4	2.1	Residual disease after primary debulking surgery	Yes	18	15.8	
	IIB	6	3.2		No	96	84.2	
	IIIA1	8	4.2	First-line platinum therapy	All	190	100.0	
	IIIA2	5	2.6		Yes	99	52.1	
	IIIB	10	5.3	Relapse after first-line therapy	No	91	47.9	
	IIIC	77	40.5		Yes	59	31.1	
	IVA	12	6.3	Received PARP	No	131	68.9	
	IVB	27	14.2		Yes	29	49.1	
	NA		19	10.0	Exitus	No	30	50.9
						Yes	39	20.5
Stage (aggregated)	Localized (I–IIB)	34	17.9	Clinical parameter	Median (range)			
	Locally Advanced (III–IVA)	120	63.2		Age at diagnosis, years	59.2	[34.1–83.9]	
Type of biopsy	Metastatic (IVB)	36	18.9	Platinum-free interval, months	16.3			
	Excisional	132	69.5		[0.0–83.3]			
BRCAg	Incisional	35	18.4	PFS to PARPi therapy, months	11.0			
	Tru-Cut	23	12.1		[1.0–64.6]			
	WT or benign/Likely benign	141	71.2		31.0			
BRCAg	Variant of unknown significance	13	6.8	Follow-up, months	[5.9–159.3]			
	Pathogenic	36	18.9		31.0			
				Overall survival, months	[5.87–159.27]			

PARPi, poly(ADP-ribose) polymerase inhibitor. NA, not available.

3.2. Mutational Distribution and Clinical Implications

Mutational analysis was performed based on the results of the NGS custom panel, which analyzed 35 DNA damage repair genes. As expected, the most frequently mutated gene was *TP53*, which was mutated in 72.11% (137/190) of samples, followed by *BRCA1* and *BRCA2*, with incidences of 16.84% (32/190) and 15.26% (29/190), respectively. Germline mutations were detected in 59.02% (36/61) of patients with *BRCA1/2*-mutated HGSOC.

Other HRR genes were also found to be altered, with a total incidence of 11.05% (21/190), some of them coexisting with *BRCA* mutations. In addition, alterations in other DNA damage repair genes were also identified (Figure 1). Mutational data were used to classify tumors as HRR-proficient or HRD, according to the mutational status of pathway-specific genes. Hence, 35.79% (68/190) of patients were considered HRR mutated (HRRmut).

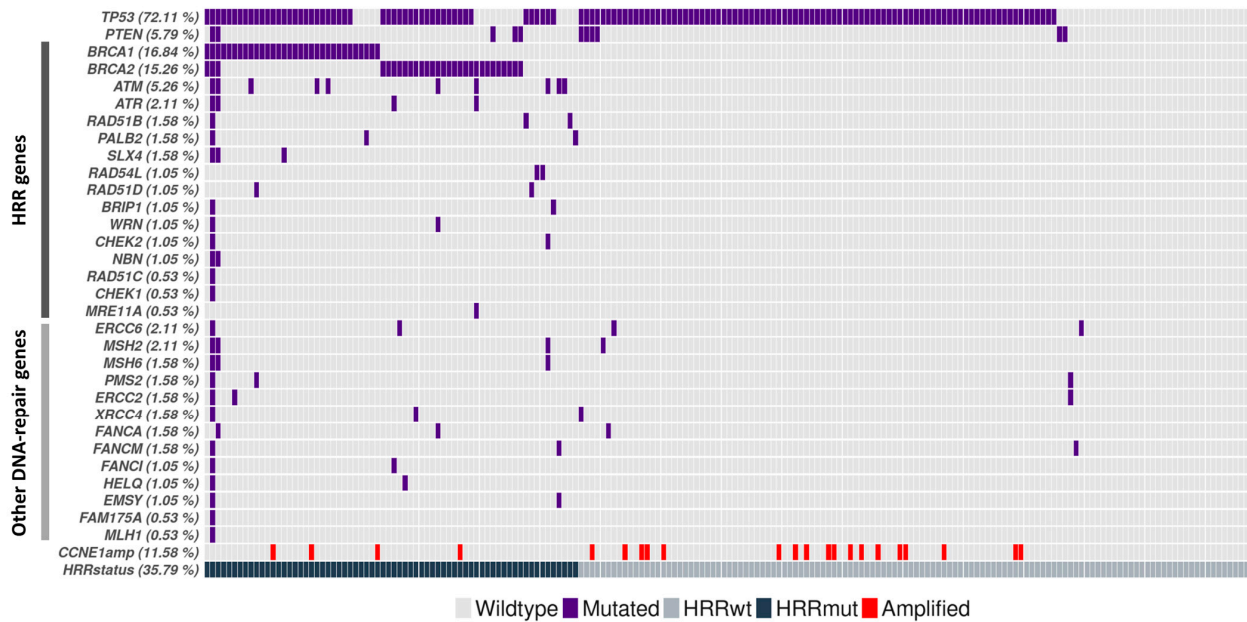


Figure 1. Distribution of mutations in DNA damage repair genes among 190 patients with high-grade serous ovarian cancer, stratified by HRR gene status. HRR, homologous recombination repair; mut, mutated; wt, wild-type.

Non-parametric and log-rank tests were used to evaluate the ability of HRR mutation status to predict response to DNA-damaging drugs (including platinum-based and PARPi therapies). The results revealed differences for tumors HRR wildtype (HRRwt) versus HRRmut with respect to both PFI ($p = 5 \times 10^{-8}$), with a median PFI of 15.3 and 72.1 months, and PFS to PARPi ($p = 0.00085$), with a median of 8.53 months for HRRwt and were not achieved by HRRmut, demonstrating the prognostic impact of HRR mutation status (Figure 2).

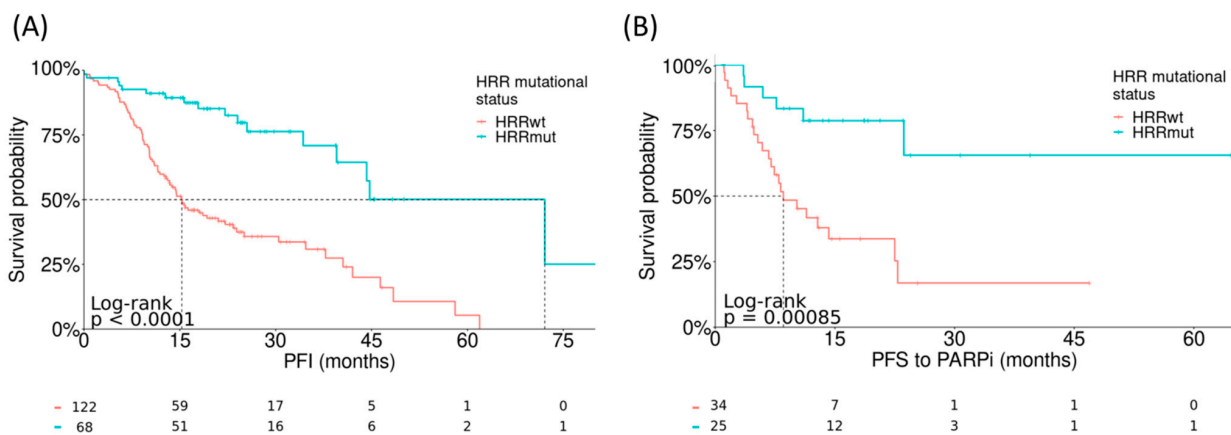


Figure 2. Log-rank test to evaluate the predictive ability of an HRR gene mutation-based classifier with respect to (A) PFI, HR = 0.25 (95% CI: 0.15–0.43) and (B) PFS to PARPi therapy, HR = 0.25 (95% CI: 0.1–0.62). HRR, homologous recombination repair; mut, mutated; PARPi, poly(ADP-ribose) polymerase inhibitor; PFI, platinum-free interval; PFS, progression-free survival; wt, wild-type.

CCNE1 has previously been implicated in the prognosis of patients with HGSOV [4], and therefore, the addition of the CCNE1 amplified cases in this series could increase the accuracy when classifying patients. For that reason, CCNE1 amplification was evaluated *in silico* in this series. Patients whose tumors harbored amplifications in CCNE1 (22/190, 11.58%) were classified as an independent subgroup to evaluate the prognostic implication of each genomic alteration. The addition of CCNE1 amplified cases as a new independent group showed significant differences in the log-rank tests for both PFI ($p < 0.0001$) and PFS to PARPi ($p = 0.00012$) (Figure S6). In the case of PFS to PARPi, the presence of CCNE1 amplification was associated with the worst-prognosis group, followed by HRRwt and, finally, HRRmut.

3.3. Copy Number Parameters and Their Clinical Implications

The applied NGS approach also includes 147,000 SNPs homogeneously distributed among the whole genome. These data facilitated the assessment of GI based on copy number analysis by using an in-house pipeline. Hence, we were able to establish GI profiles and quantify them using different predefined parameters (Supplementary Data S3). Each GI parameter was tested for associations with continuous and categorical response variables. GI parameters that were more significantly associated with PFI in non-parametric tests were the total number of LOH events of >15 Mb ($p = 0.019$) and the percentage of the genome that was altered by LOH of >15 Mb ($p = 0.016$) (Figure S7). However, there were also other GI parameters also resulted in significant correlation, as specified in Supplementary File.

The correlation between pre-established HRD scores, as previously described [26], and response variables was also evaluated. The highest significance for predicting PFI was seen with the LOH parameter stratified by its median value ($p = 0.0071$), followed by the HRD score stratified by its median value ($p = 0.031$). However, none of the pre-established HRD scores investigated was able to significantly predict PFS to PARPi (Figures S8 and S9).

Aiming to optimize the generated data, even though GI parameters on their own could work as a predictive biomarker and to improve the currently available biomarkers, the combination of them was used as a base to build a predictive model.

Finally, GI profiles, described by the presence of GI parameters, were determined to compare the different HRR mutational-based populations. As expected, a higher accumulation of GI was found in samples harboring mutations in the HRR pathway and was especially enriched for those with BRCA mutations (Figure S10).

3.4. Independent Model Fitting and Building of the Integrative Ensemble Model (Scarface Score)

In order to adjust a machine learning strategy to predict response to platinum-derived therapy, attributes from three different sources were used. The first model was derived from the raw coverage information of 147,000 SNPs, while the third model contained gene expression data from 2549 genes obtained from targeted RNA sequencing results. Feature selection was performed using several strategies, as described in the Materials and Methods. The second model included the most representative parameters of the GI phenomenon but was not subjected to feature selection because of a low number of features. Each set of selected parameters was tested and coupled with a data-mining algorithm. Every possible combination of the data-mining algorithm and selected features was tested.

The best performances were seen with a support vector machine with eight SNPs ('SNP model'; Table S1), a support vector machine with 28 GI parameters ('GI model'), and a neural network with the expression of seven genes ('HTG model'; Table S2). Selected features of each model are described in Supplementary Data S3. The performance of each model is shown in Table 2. Weights and main characteristics of the features included in each of the three models and the ensemble are included in Tables S3–S6. Among the three single-source models, the best performance was obtained with the GI model, which had an accuracy of 0.9038. Finally, an ensemble model (the Scarface score) was developed based on a support vector machine algorithm, using as an input the 43 attributes from the individual models described above. The ensemble model was trained with a bootstrapping of 500 iterations and

obtained an accuracy of 0.96 and a kappa index of 0.91, outperforming all three single-source models. All performance parameters were obtained from the validation series.

Table 2. Performance of the different predictive algorithms tested.

Model	TP/TN/FP/FN	Accuracy (95% CI)	Sensitivity	Specificity	Kappa
SNP model	29/13/5/5	0.8077 (0.6747–0.9037)	0.7222	0.8529	0.5752
HTG model	25/17/1/9	0.8077 (0.6747–0.9037)	0.9444	0.7353	0.6154
GI model	31/16/2/3	0.9038 (0.7897–0.968)	0.8889	0.9118	0.7903
Ensemble model	34/16/2/0	0.9615 (0.8679–0.9953)	0.8889	1.0000	0.9128

FP, false positive; FN, false negative; GI, genetic instability; SNP, single nucleotide polymorphism; TP, true positive; TN, true negative.

The clinical impact of each model was tested in the whole population of patients with HGSOC (n = 183) by using a log-rank test with PFI as a time-to-event variable. All four models, including the ensemble model, were able to distinguish responders from non-responders with significant differences in PFI (all $p < 0.0001$; Figure 3). The HTG-based model was found to be the most limited, while the highest statistical significance was obtained using the ensemble model ($p < 2 \times 10^{-16}$).

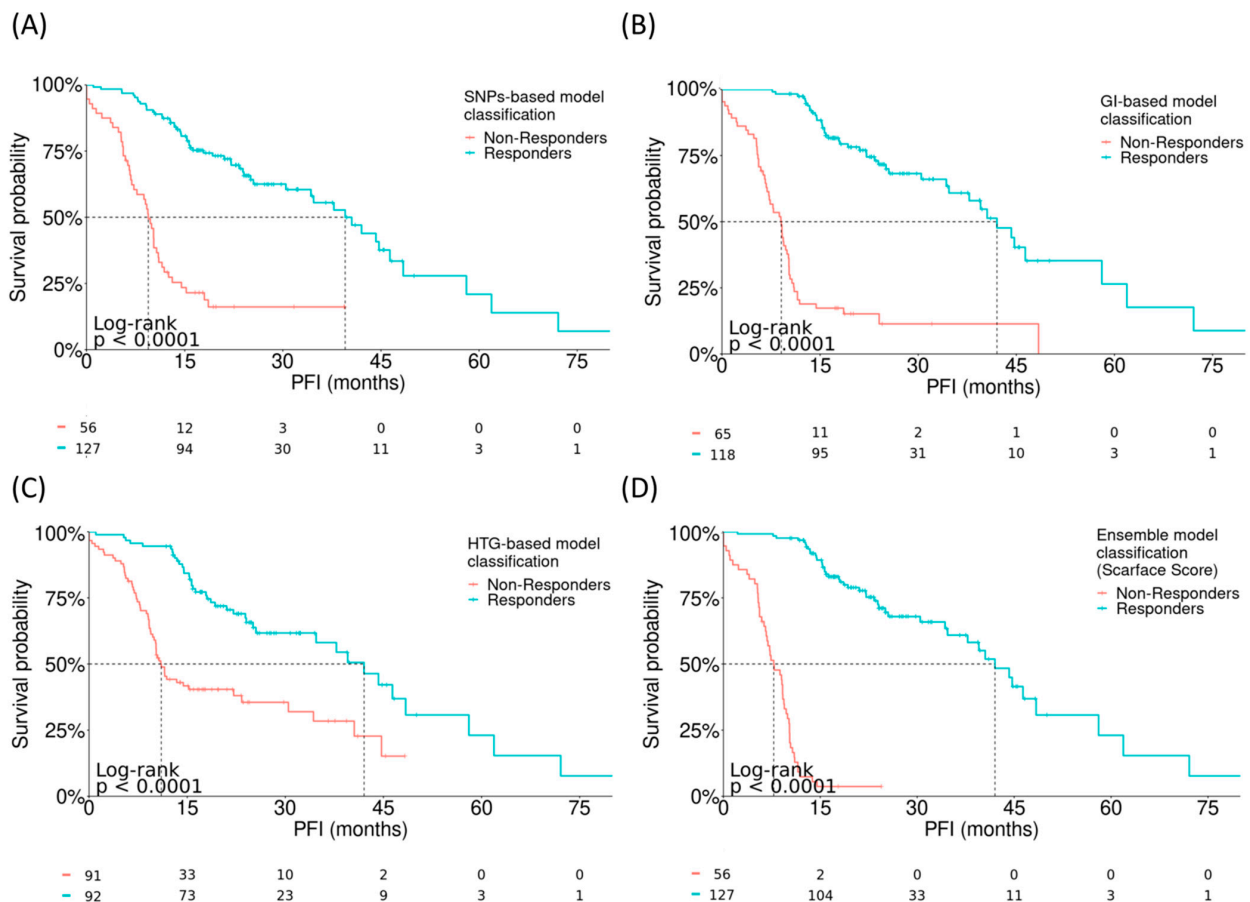


Figure 3. Correlation between the fitted models and PFI. Log-rank tests evaluating the performance of (A) SNP-based model, HR = 0.19 (0.12–0.29), (B) GI-based model, HR = 0.12 (95% CI: 0.08–0.19), (C) HTG-based model, HR = 0.34 (95% CI: 0.22–0.51), and (D) integrative ensemble model (Scarface Score).

score), HR = 0.046 (95% CI: 0.027–0.077). GI, genomic instability; PFI, platinum-free interval; SNP, single-nucleotide polymorphism.

The goodness-of-fit of each model was evaluated using receiver operating characteristic (ROC) curves, which showed how well each predictive model discriminated between patients with a PFI ≥12 versus <12 months. As expected, the highest discriminative power was obtained with the ensemble model, which had an area under the curve of 0.962, a sensitivity of 0.929, and a specificity of 0.945 (Figure 4A).

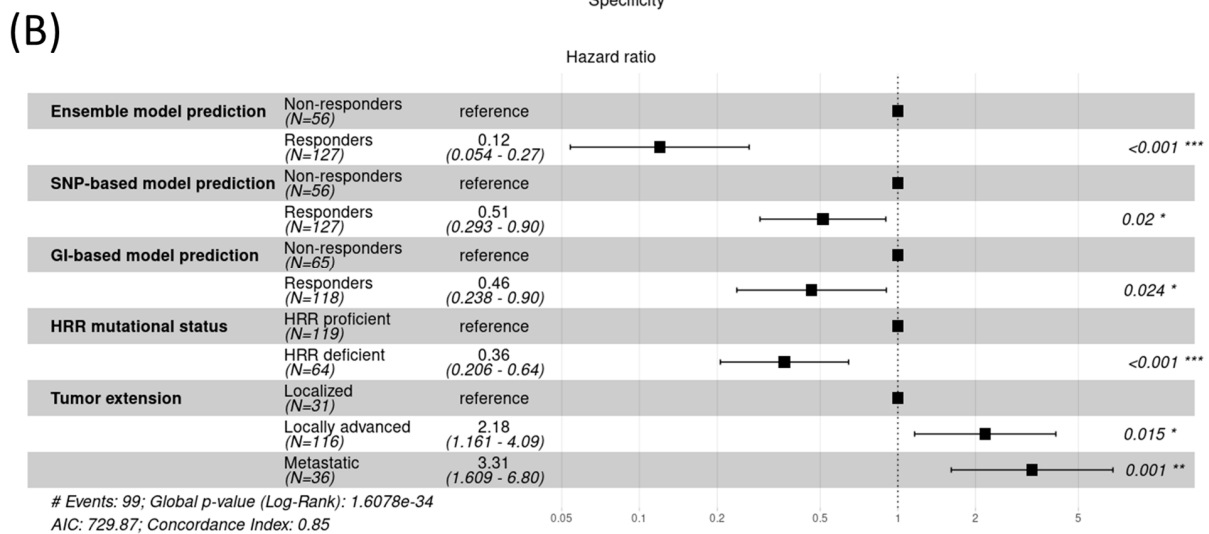
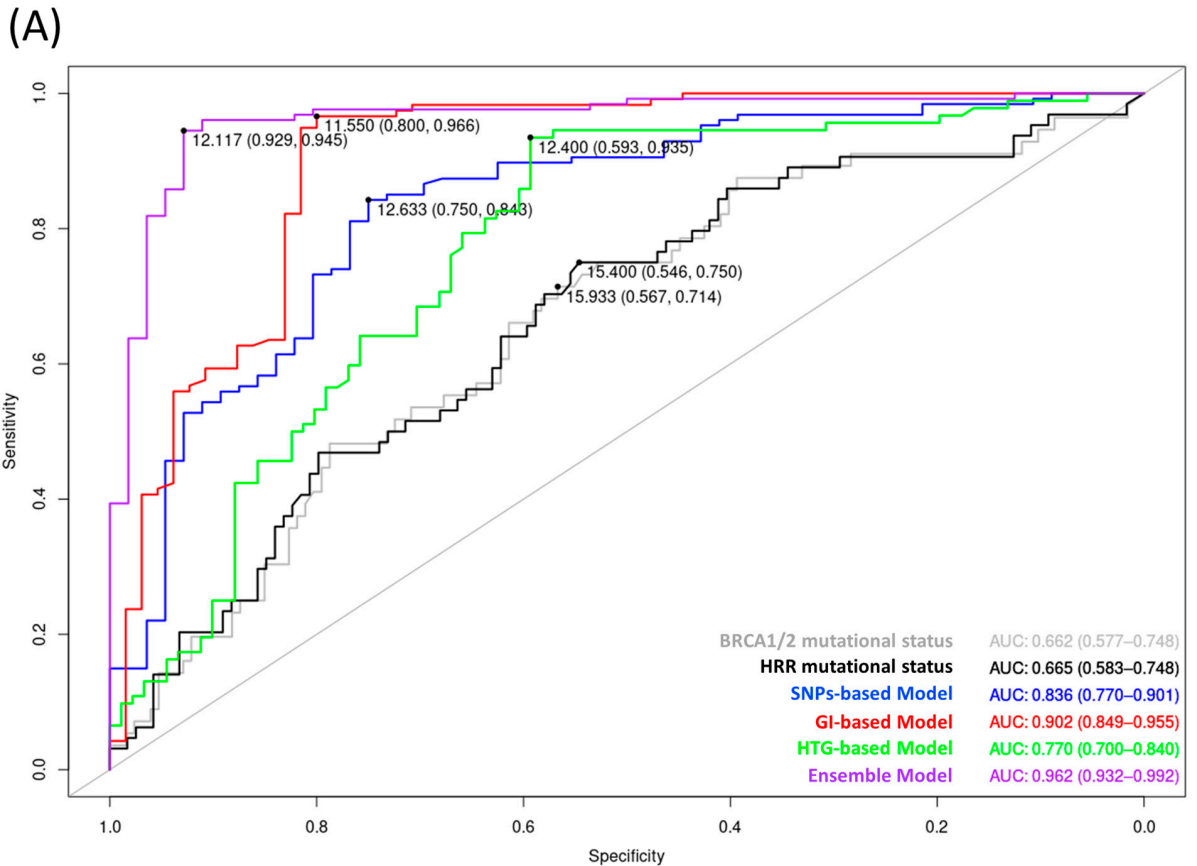


Figure 4. Performance of the predictive models. (A) ROC curves comparing the three predictive models, the ensemble model, and HRR-based classifications as categorical variables, with PFI as a continuous variable. (B) Multivariate Cox regression analysis for HRR mutation status, tumor extension, and the

performance of the models. Tumor extension was stratified based on stage: localized (I–IIb), locally advanced (III–IVa), or metastatic (IVb) regarding PFI. * p -value ≤ 0.05 , ** p -value < 0.01 and *** p -value < 0.001 . GI, genomic instability; HRR, homologous recombination repair; PFI, platinum-free interval; SNP, single-nucleotide polymorphism. # Characteristics of the regression.

Although the algorithms were trained to predict response to platinum-based chemotherapy, the ultimate aim of the study was to develop a model that could identify patients who are candidates for PARPi therapies. Thus, the ability of the models to discriminate the best responders to PARPi therapies was also investigated using log-rank testing in a sub-cohort of 58 patients from the overall population who had received PARPi therapy in addition to first-line platinum-based chemotherapy. The performance of the models was compared with the stratification based on *BRCA* mutation, which is the current gold standard for selecting patients to receive PARPi therapy. The ensemble model was found to have a p -value of 0.00077 for non-responders versus responders, which outperformed *BRCA*-based classification ($p = 0.0048$) (Figure 5 and Figure S11), thus improving the discriminant power of the gold standard.

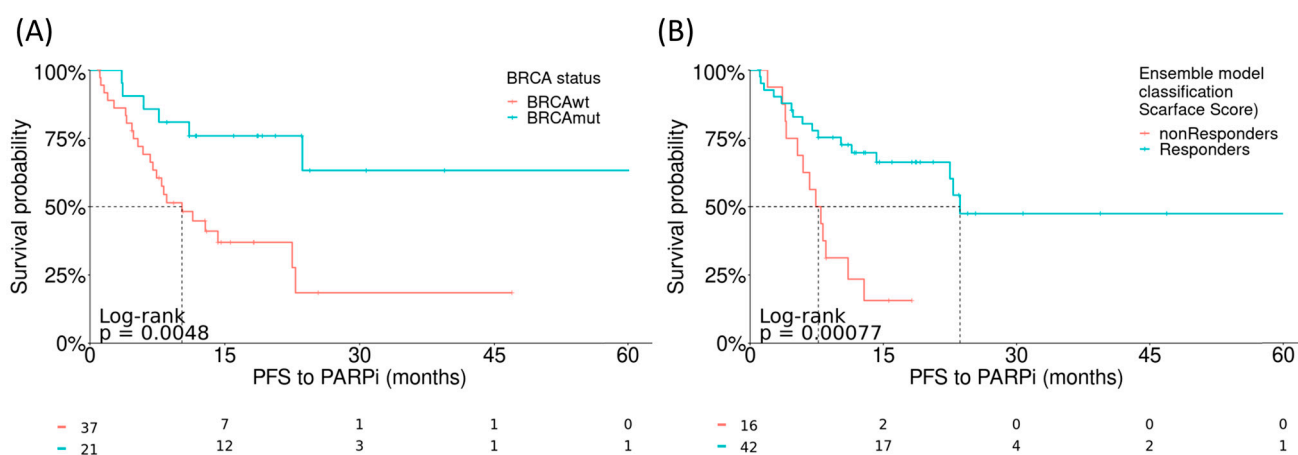


Figure 5. Correlation between the fitted models and PFS to PARPi. Log-rank tests evaluating the performance of: (A) *BRCA* mutation-based classification and (B) the integrative ensemble model (Scarface score). PARPi, poly(ADP-ribose) polymerase inhibitor; PFS, progression-free survival.

The ability of the models to predict overall survival was also evaluated. All models reached statistical significance, with the greatest significance seen for the ensemble model (Figure S12C–F). In contrast, classification based on *BRCA* or HRR gene status appeared unable to significantly predict overall survival (Figure S12A,B). Exact p -values and summarized survival analyses are shown in Table S7.

In addition to model performance, a multivariate analysis was performed to evaluate the ability of different clinicopathologic and mutational parameters to stratify patients according to overall survival. The most discriminant parameter was the ensemble model prediction (hazard ratio (HR) 0.12). However, other parameters, such as tumor extension (locally advanced, HR 2.18; metastatic, HR 3.31 and HRR mutation status (HR 0.36), also contributed to risk assessment (Figure 4B). Additional Cox analyses were performed evaluating a higher number of variables (Figure S13).

4. Discussion

GI, as a surrogate of HRD, has risen as a prognostic and predictive tool in HGSOc [27]. While HRR-based stratification, based on any alteration or effect in the genome, is widely recognized as essential, many efforts have been made to develop and clinically validate academic tools based on different approaches [28–30]. In this study, we developed three single-source models based on SNPs, GI, and RNA expression analysis, respectively, and an integrative ensemble model (the Scarface score) to predict response to DNA-damaging

agents—particularly platinum-based chemotherapy and PARPis. The Scarface model—which combined eight SNPs, 28 GI parameters, and the expression of seven genes—showed the best performance, with an accuracy of 0.9615 and a kappa index of 0.9128 in the validation series. However, the single-source models could also be suitable and efficient tools in a real-life clinical setting, helping to guide the clinical management of patients. The proposed models were built based on three layers: SNP deep NGS, a CNV profile using in silico algorithms, and targeted RNA sequencing using HTG EdgeSeq technology. Each layer has its strengths and limitations, but ultimately, each underpins the others. This design accounts for the different mechanisms by which HRD is produced and tries to mimic the complex biological context (e.g., genomic, transcriptomic). These different levels of biological information could be better represented by a multiomic approach. For this purpose, the capacity of machine learning to account for complex interactions in large datasets [31] made it optimal for the study of GI based on drug response. Several machine learning models (support vector machine, random forest, neural network, decision tree, and naïve Bayes) were adjusted with different parameters and hyperparameters, and the resulting models were benchmarked to rank the best performance for each layer.

Commercial solutions, such as MyChoice[®] CDx Plus (Myriad Genetics, Salt Lake City, UT, USA) and the FoundationOne[®] CDx (Foundation Medicine, Cambridge, MA, USA), which are based on identifying genomic scars, HRR gene mutations and LOH, have already shown their clinical benefit in clinical trials [32–34]. However, even if each model succeeds in predicting *BRCA1/2* status (for which they are trained), the fact that they do not cover other molecular mechanisms (e.g., CNV or gene expression) means that they do not provide information on other HRD-causing mechanisms independent of *BRCA* gene status [35,36]. In addition, there has not yet been a direct prospective comparison between the two tests. One study reported on the interchangeability of the MyChoice assay using LOH alone compared with the GI score (GIS) and showed poor agreement; among 3209 wild-type *BRCA* genes, 53% of those assigned as unstable by GIS (cut-off ≥ 33) were assigned as HRD-negative by %LOH criteria, while only 4% of unstable tumors assessed by %LOH were positive using GIS. Considering *BRCA1/2* and the official GIS cut-off of ≥ 42 , an agreement was 64.9% for positive cases and 96.6% for negative cases [37]. Similar discrepancies were also seen in a retrospective analysis that found that 23% of samples were classified as GI stable, with an LOH percentage of <16%, by FoundationOne harbored *BRCA1/2* germline mutations [38]. These facts, together with the high costs of these tests and long turnaround times for results, constitute the main limitations of both commercial tests.

With the Scarface model, we have integrated GI parameters—equivalent to HRD status—rather than HRR mutations to differentiate patients more accurately according to PFI. Information about gene expression is also provided, supporting the GI and contributing to responder–phenotype processes. This approach has the advantage of studying the GI phenomenon as a whole: at the genomic, chromosomal, and transcriptomic levels. Due to the impossibility of comparing our data with the gold standard, such as those mentioned above, since patients included in the study lack this type of determination, we compared the model with a classification based on HRR gene mutations (*BRCA1/2* only and all HRR genes) and scores from the scarHRD pipeline [22]. In our series, 35.8% of samples had HRR gene mutations, with *BRCA1* and *BRCA2* mutations in 16.84% and 15.26%, respectively. Additionally, amplification of *CCNE1* was performed in our series, with an incidence of approximately 12%. Co-occurrence of *BRCA1/2* mutations and *CCNE1* amplification were found in approximately 7% of *BRCA1/2* mutated cases, similar to the frequencies found in the OC-TCGA [4]. Even though these alterations are found together in a very low number of cases, there are not mutually exclusive. Those samples harboring mutations in HRR genes were classified as HRD for comparison. Both stratifications—based on *BRCA1/2* mutation and all HRR genes—were able to identify patients who would have an extended PFI (both $p < 0.0001$) and PFS to PARPi (*BRCA1/2*, $p = 0.0048$; all HRR genes, $p = 0.0013$). In this particular case, adding other HRR genes to *BRCA1/2* when classifying patients improved statistical power and increased the prognostic and predictive value. However, as recently

reported, they do not always overlap GI, suggesting higher accuracy of the GI score over an HRR gene panel to define an HRD phenotype [39,40]. For that reason, approaches at different levels, such as genomic scars, are gaining strength in the assessment of GI.

The scarHRD pipeline was applied to compare the performance of the classifiers. This pipeline has been trained to identify the genomic scars evaluated by the validated commercial solutions, LOH, large-scale transitions, telomeric allelic imbalances, and HRDscore. However, the results were not as good as expected. Differences in methodologic and analytic procedures caused a loss of statistical significance when analyzing our series, with several potential causes. First, in this approach, GI data were derived from NGS data covering a backbone and a medium-size panel, whereas the MyChoice kit was validated and calibrated using a comparative genomic hybridization array. Second, the CNVkit method was used with the parameters specifically tuned to our clinical scenario, including pre-analytical factors such as tumor burden in the sample. The best results were obtained when the series was stratified based on the median number of LOH events (PFI, $p = 0.0071$; PFS to PARPi, $p = 0.07$) and median HRD score (PFI, $p = 0.031$; PFS to PARPi, $p = 0.28$), but significance was only reached for PFI and not for PFS to PARPi. In contrast, the Scarface model achieved the highest statistical significance for both PFI ($p < 2 \times 10^{-16}$) and PFS to PARPi ($p = 0.00077$), improving the predictive performance above that of previously used classifiers.

As mentioned, the predictive algorithm was trained and validated in an ambispective, multicentric, real-life cohort of patients with HGSOE using PFI as an endpoint. Because of the real-life design, information regarding PFS to PARPi was not as accurate as expected; PFS to PARPi data were collected with respect to different lines of therapy (first-line therapy in 23 patients and second or later lines in 35 patients), different treatment combinations and schemes, and different PARPi drugs. As such, PFS to PARPi was not a suitable parameter for training and validating the model. The real-world nature of the series, which lacks centralized review, probably implies the misclassification of some studied cases. The concordance between the centralized review and the first diagnosis is approximately 70%, as previously presented in other works in OC [41]. This could be the cause of the low number of TP53 alterations found (72% in this cohort vs. more than 90% in other series [4]). The same cause could be responsible for the high number of BRCA1/2 mutated cases without TP53 alteration, uncommonly found in HGSOE. Representation of other histologies with different mutational patterns, such as the case of endometrial OC [42], could be influencing the results. Even if this fact constitutes a limitation of the study, it is also presented as a strength since it represents the reality of the clinical practice in which the model would be potentially used. Otherwise, another limitation of the study consists of the fact that the data sources are quite specific; thus, it is necessary to sequence the samples with the kit described in material and methods containing a backbone. Additionally, this fact limits the availability of data in public repositories. Therefore, although the presented algorithm showed that HRR mutations had predictive value for PFS to PARPi, the model should be further evaluated in a cohort with homogeneous PARPi response data to validate its clinical benefit. In addition, because this model addresses GI from different levels of regulation, it seems that it would be plausible to calibrate the model to predict response with different cut-offs in other tumors in which GI may play an important role in response to therapy, such as advanced prostate cancer with BRCA mutations or pancreatic cancer. Analogously, new optimal cut-offs for GIS and genomic LOH have been proposed in the VELIA and ARIEL2 clinical trials [5,39]. Thus, there is room for improvement in the exposed GI study approach.

5. Conclusions

The Scarface score constitutes a useful academic tool to predict response to DNA-damaging agents in HGSOE and, potentially, in other HRR-deficient tumors. This algorithm addresses the limitations of available and validated commercial solutions by looking at GI and the molecular biology of the tumor from a more comprehensive point of view.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers15113030/s1>, Figure S1: Differences in GI parameters according segmentation used in CNVkit pipeline; Figure S2: Differences in GI parameters adjusting p -value in CNVkit pipeline; Figure S3: Differences in GI parameters according tumor burden in CNVkit pipeline; Figure S4: GI parameters according to pre-filtering step in saasCNV pipeline; Figure S5: Comparison of GI parameters between implemented pipe-lines, highlighting differences for the assessment of each feature; Figure S6: Log-Rank test to evaluate the predictive information of HRR-gene mutation and CCNE1 amplification-based classification regarding (A) PFI and (B) PFS to PARPi; Figure S7: Correlation of GI parameters with PFI assessed by non-parametric tests; Figure S8: Log-Rank tests evaluating the implication of predefined HRD scars parameters from scarHRD package in correlation with PFI and PARPi response; Figure S9: Correlation of HRD score obtained on scarHRD package and time-to-event variables; Figure S10: Distribution of total number of LOH events > 15 Mb between BRCA-based populations; Figure S11: Log-Rank tests evaluating the implication of predictive models with PARPi response; Figure S12: Log-Rank tests evaluating the implication of mutational-based classifiers and predictive models with OS; Figure S13: Multivariate analysis performed by Cox regression for clinicopathological parameters, HRR alteration and three-source model performance in addition to ensemble model; Table S1: Selected parameters SNPs model; Table S2: Selected parameters HTG model; Table S3: Weights and main characteristics of the parameters included in the layer 1 of SNPs deep sequencing; Table S4: Weights and main characteristics of the parameters included in the layer 2 of Genomic Instability related parameters; Table S5: Weights and main characteristics of the parameters included in the layer 3 of gene expression; Table S6: Weights and main characteristics of the parameters included in the ensemble scarface model; Table S7: Log-rank test results for single-source and ensemble model. Additionally, PFI, BRCA (germline and somatic mutations) and HRR-based (all HR-genes interrogated in the panel, including BRCA1/2) classifications were added; Data S1: CN pipelines: personalization of working parameters; Data S2: Methods model fitting; Data S3: Selected parameters.

Author Contributions: Conceptualization, R.L.-R., A.F.-S., I.R. and J.A.L.-G.; methodology, R.L.-R. and A.F.-S.; formal analysis, R.L.-R. and A.F.-S.; data curation, R.M., A.G., L.M.d.S., A.Y., C.P.-S., A.R.-V., M.P.B.-G., E.M., C.E., F.G., A.B.S.-H., E.M.G.-A., L.G., M.Q., I.P., J.A. (Jesús Alarcón), A.O., J.A. (Jessica Aliaga), M.R.-C., Z.G.-C. and I.R.; writing—original draft preparation, R.L.-R., A.F.-S., I.R. and J.A.L.-G.; writing—review and editing, R.L.-R., A.F.-S., I.R. and J.A.L.-G.; visualization, R.L.-R. and A.F.-S.; supervision, R.L.-R., A.F.-S., I.R. and J.A.L.-G.; funding acquisition, R.L.-R., A.F.-S., I.R. and J.A.L.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by GVA Grants “Subvencions per a la realització de projectes d’i+d+i desenvolupats per grups d’investigació emergents (GV/2020/158)” and “Ayudas para la contratación de personal investigador en formación de carácter predoctoral” (ACIF/2016/008) and “Beca de investigación traslacional Andrés Poveda 2020” from GEICO group. This study was awarded the Prize “Antonio Llombart Rodríguez-FINCIVO 2020” from the Royal Academy of Medicine of the Valencian Community.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Fundació Instituto Valenciano de Oncologia (ACO-COE-3012-02, 2014 and LBM-02-20, SCARFACE, 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the Biobank of the Fundació Instituto Valenciano de Oncologia for providing the biological samples for the analysis.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shen, Z. Genomic instability and cancer: An introduction. *J. Mol. Cell Biol.* **2011**, *3*, 1–3. [[CrossRef](#)] [[PubMed](#)]
2. Kim, T.-M.; Xi, R.; Luquette, L.J.; Park, R.W.; Johnson, M.D.; Park, P.J. Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. *Genome Res.* **2013**, *23*, 217–227. [[CrossRef](#)] [[PubMed](#)]
3. Weir, B.; Zhao, X.; Meyerson, M. Somatic alterations in the human cancer genome. *Cancer Cell* **2004**, *6*, 433–438. [[CrossRef](#)] [[PubMed](#)]

4. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **2011**, *474*, 609–615. [[CrossRef](#)]
5. Ngoi, N.Y.L.; Tan, D.S.P. The role of homologous recombination deficiency testing in ovarian cancer and its clinical implications: Do we need it? *ESMO Open* **2021**, *6*, 100144. [[CrossRef](#)]
6. Zack, T.I.; Schumacher, S.E.; Carter, S.L.; Cherniack, A.D.; Saksena, G.; Tabak, B.; Lawrence, M.S.; Zhang, C.Z.; Wala, J.; Mermel, C.H.; et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **2013**, *45*, 1134–1140. [[CrossRef](#)]
7. Lord, C.J.; Ashworth, A. BRCAness revisited. *Nat. Rev. Cancer* **2016**, *16*, 110–120. [[CrossRef](#)]
8. Uzilov, A.V.; Ding, W.; Fink, M.Y.; Antipin, Y.; Brohl, A.S.; Davis, C.; Lau, C.Y.; Pandya, C.; Shah, H.; Kasai, Y.; et al. Development and clinical application of an integrative genomic approach to personalized cancer therapy. *Genome Med.* **2016**, *8*, 62. [[CrossRef](#)]
9. Kang, J.; D'Andrea, A.D.; Kozono, D. A DNA repair pathway-focused score for prediction of outcomes in ovarian cancer treated with platinum-based chemotherapy. *J. Natl. Cancer Inst.* **2012**, *104*, 670–681. [[CrossRef](#)]
10. Marquard, A.M.; Eklund, A.C.; Joshi, T.; Krzystanek, M.; Favero, F.; Wang, Z.C.; Richardson, A.L.; Silver, D.P.; Szallasi, Z.; Birkbak, N.J. Pan-cancer analysis of genomic scar signatures associated with homologous recombination deficiency suggests novel indications for existing cancer drugs. *Biomark. Res.* **2015**, *3*, 9. [[CrossRef](#)]
11. Watkins, J.A.; Irshad, S.; Grigoriadis, A.; Tutt, A.N. Genomic scars as biomarkers of homologous recombination deficiency and drug response in breast and ovarian cancers. *Breast Cancer Res.* **2014**, *16*, 211. [[CrossRef](#)] [[PubMed](#)]
12. Grimm, C.; Cropet, C.; Ray-Coquard, I. Maintenance olaparib plus bevacizumab (bev) after platinum-based chemotherapy plus bev in patients (pts) with newly diagnosed advanced high-grade ovarian cancer (HGOC): Efficacy by timing of surgery and residual tumor status in the Phase III PAOLA-1 trial. *Gynecol. Oncol.* **2020**, *159*, 19. [[CrossRef](#)]
13. González-Martín, A.; Pothuri, B.; Vergote, I.; DePont Christensen, R.; Graybill, W.; Mirza, M.R.; McCormick, C.; Lorusso, D.; Hoskins, P.; Freyer, G.; et al. Niraparib in patients with newly diagnosed advanced ovarian cancer. *N. Engl. J. Med.* **2019**, *381*, 2391–2402. [[CrossRef](#)] [[PubMed](#)]
14. Coleman, R.; Fleming, G.; Brady, M.; Swisher, E.; Steffensen, K.; Friedlander, M.; Okamoto, A.; Moore, K.; Ben-Baruch, N.; Werner, T.; et al. VELIA/GOG-3005: Integration of veliparib (V) with front-line chemotherapy and maintenance in women with high-grade serous carcinoma of ovarian, fallopian tube, or primary peritoneal origin (HGSC). *Ann. Oncol.* **2019**, *30*, v895–v896. [[CrossRef](#)]
15. Monk, B.J.; Parkinson, C.; Lim, M.C.; O'malley, D.M.; Oaknin, A.; Wilson, M.K.; Coleman, R.L.; Lorusso, D.; Bessette, P.; Ghamande, S.; et al. A randomized, phase III trial to evaluate rucaparib monotherapy as maintenance treatment in patients with newly diagnosed ovarian cancer (ATHENA-MONO/GOG-3020/ENGOT-ov45). *J. Clin. Oncol.* **2022**, *40*, 3952–3964. [[CrossRef](#)]
16. Faraoni, I.; Graziani, G. Role of BRCA mutations in cancer treatment with poly (ADP-ribose) polymerase (PARP) inhibitors. *Cancers* **2018**, *10*, 487. [[CrossRef](#)]
17. Wagener-Rydzek, S.; Merkelbach-Bruse, S.; Siemanowski, J. Biomarkers for homologous recombination deficiency in cancer. *J. Pers. Med.* **2021**, *11*, 612. [[CrossRef](#)]
18. Dong, F.; Davineni, P.K.; Howitt, B.E.; Beck, A.H. A BRCA1/2 Mutational Signature and Survival in Ovarian High-Grade Serous Carcinoma. *Cancer Epidemiol. Biomark. Prev.* **2016**, *25*, 1511–1516. [[CrossRef](#)]
19. Nik-Zainal, S. From genome integrity to cancer. *Genome Med.* **2019**, *11*, 4. [[CrossRef](#)]
20. Talevich, E.; Shain, A.H.; Botton, T.; Bastian, B.C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput. Biol.* **2016**, *12*, e1004873. [[CrossRef](#)]
21. Povysil, G.; Tzika, A.; Vogt, J.; Haunschmid, V.; Messiaen, L.; Zschocke, J.; Klambauer, G.; Hochreiter, S.; Wimmer, K. panelcn.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics. *Hum. Mutat.* **2017**, *38*, 889–897. [[CrossRef](#)] [[PubMed](#)]
22. Sztupinszki, Z.; Diossy, M.; Krzystanek, M.; Reiniger, L.; Csabai, I.; Favero, F.; Birkbak, N.J.; Eklund, A.C.; Syed, A.; Szallasi, Z. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *NPJ Breast Cancer* **2018**, *4*, 16. [[CrossRef](#)] [[PubMed](#)]
23. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]
24. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [[CrossRef](#)]
25. Kursa, M.B.; Rudnicki, W.R. Feature selection with the Boruta package. *J. Stat. Softw.* **2010**, *36*, 1–13. [[CrossRef](#)]
26. Telli, M.L.; Timms, K.M.; Reid, J.; Hennessy, B.; Mills, G.B.; Jensen, K.C.; Szallasi, Z.; Barry, W.T.; Winer, E.P.; Tung, N.M.; 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*, 3764–3773. [[CrossRef](#)]
27. da Cunha Colombo Bonadio, R.R.; Fogace, R.N.; Miranda, V.C.; Diz, M.d.P.E. Homologous recombination deficiency in ovarian cancer: A review of its epidemiology and management. *Clinics* **2018**, *73*, e450s. [[CrossRef](#)]
28. Despierre, E.; Moisse, M.; Yesilyurt, B.; Sehoul, J.; Braicu, I.; Mahner, S.; Castillo-Tong, D.C.; Zeillinger, R.; Lambrechts, S.; Leunen, K.; et al. Somatic copy number alterations predict response to platinum therapy in epithelial ovarian cancer. *Gynecol. Oncol.* **2014**, *135*, 415–422. [[CrossRef](#)]
29. Bogush, T.A.; Basharina, A.A.; Bogush, E.A.; Scherbakov, A.M.; Davydov, M.M.; Kosorukov, V.S. The expression and clinical significance of ER β /ER α in ovarian cancer: Can we predict the effectiveness of platinum plus taxane therapy? *Ir. J. Med. Sci.* **2022**, *191*, 2047–2053. [[CrossRef](#)]

30. Staropoli, N.; Arbitrio, M.; Salvino, A.; Scionti, F.; Ciliberto, D.; Ingargiola, R.; Labanca, C.; Agapito, G.; Iuliano, E.; Barbieri, V.; et al. A Prognostic and Carboplatin Response Predictive Model in Ovarian Cancer: A Mono-Institutional Retrospective Study Based on Clinics and Pharmacogenomics. *Biomedicines* **2022**, *10*, 1210. [[CrossRef](#)]
31. Stevens, L.M.; Mortazavi, B.J.; Deo, R.C.; Curtis, L.; Kao, D.P. Recommendations for Reporting Machine Learning Analyses in Clinical Research. *Circ. Cardiovasc. Qual. Outcomes* **2020**, *13*, e006556. [[CrossRef](#)] [[PubMed](#)]
32. De Picciotto, N.; Cacheux, W.; Roth, A.; Chappuis, P.O.; Labidi-Galy, S.I. Ovarian cancer: Status of homologous recombination pathway as a predictor of drug response. *Crit. Rev. Oncol./Hematol.* **2016**, *101*, 50–59. [[CrossRef](#)]
33. Bartl, T.; Paspalj, V.; Grimm, C. Homologous recombination deficiency in epithelial ovarian cancer. *Memo-Mag. Eur. Med. Oncol.* **2020**, *13*, 367–370. [[CrossRef](#)]
34. Abkevich, V.; Timms, K.M.; Hennessy, B.T.; Potter, J.; Carey, M.S.; Meyer, L.A.; Smith-McCune, K.; Broaddus, R.; Lu, K.H.; Chen, J.; et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br. J. Cancer* **2012**, *107*, 1776–1782. [[CrossRef](#)] [[PubMed](#)]
35. Hoppe, M.M.; Sundar, R.; Tan, D.S.P.; Jeyasekharan, A.D. Biomarkers for Homologous Recombination Deficiency in Cancer. *J. Natl. Cancer Inst.* **2018**, *110*, 704–713. [[CrossRef](#)] [[PubMed](#)]
36. Miller, R.; Leary, A.; Scott, C.; Serra, V.; Lord, C.; Bowtell, D.; Chang, D.; Garsed, D.; Jonkers, J.; Ledermann, J.; et al. ESMO recommendations on predictive biomarker testing for homologous recombination deficiency and PARP inhibitor benefit in ovarian cancer. *Ann. Oncol.* **2020**, *31*, 1606–1622. [[CrossRef](#)] [[PubMed](#)]
37. Judkins, T.; LeClair, B.; Bowles, K.; Gutin, N.; Trost, J.; McCulloch, J.; Bhatnagar, S.; Murray, A.; Craft, J.; Wardell, B.; et al. Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk. *BMC Cancer* **2015**, *15*, 215. [[CrossRef](#)]
38. Fuh, K.; Mullen, M.; Blachut, B.; Stover, E.; Konstantinopoulos, P.; Liu, J.; Matulonis, U.; Khabele, D.; Mosammamarast, N.; Vindigni, A. Homologous recombination deficiency real-time clinical assays, ready or not? *Gynecol. Oncol.* **2020**, *159*, 877–886. [[CrossRef](#)]
39. Pujade-Lauraine, E.; Brown, J.; Barnicle, A.; Wessen, J.; Lao-Sirieix, P.; Criscione, S.W.; Bois, A.D.; Lorusso, D.; Romero, I.; Petru, E.; et al. Homologous Recombination Repair Gene Mutations to Predict Olaparib Plus Bevacizumab Efficacy in the First-Line Ovarian Cancer PAOLA-1/ENGOT-ov25 Trial. *JCO Precis. Oncol.* **2023**, *7*, e2200258. [[CrossRef](#)]
40. Pellegrino, B.; Mateo, J.; Serra, V.; Balmana, J. Controversies in oncology: Are genomic tests quantifying homologous recombination repair deficiency (HRD) useful for treatment decision making? *ESMO Open* **2019**, *4*, e000480. [[CrossRef](#)]
41. Lopez-Guerrero, J.A.; Gutierrez Pecharroman, A.; Palacios, J.; Romero, I.; Cristobal Lana, E.M.; Hardisson, D.; Vera-Sempere, F.; Illueca, C.; Vieites, B.; Garcia, A.; et al. Central pathology review of early-stage ovarian carcinoma: Description and correlation with follow-up—A study by the Spanish Group for Ovarian Cancer Research (GEICO). *Am. Soc. Clin. Oncol.* **2014**, *32*, 5583. [[CrossRef](#)]
42. Hollis, R.L.; Thomson, J.P.; Stanley, B.; Churchman, M.; Meynert, A.M.; Rye, T.; Bartos, C.; Iida, Y.; Croy, I.; Mackean, M.; et al. Molecular stratification of endometrioid ovarian carcinoma predicts clinical outcome. *Nat. Commun.* **2020**, *11*, 4995. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.