

**Genomic Alterations in
Residual High-Grade Serous Ovarian Cancer
After Neoadjuvant Chemotherapy and
Their Impact on Survival**

David Octeau

Faculty of Medicine
Division of Experimental Medicine
McGill University
Montreal, Qc

April 2018

A thesis submitted in partial fulfillment of the requirements
of the degree of Masters of Science

© David Octeau, 2018

Table of Contents

Title Page	i
Table of Contents	ii
Abstract	iii
Résumé	iv
Acknowledgements	v
Preface	vi
List of Figures	vii
List of Abbreviations	viii
Introduction	
1. Clinical Features and Treatment Strategies	
1.1 Epidemiology, anatomy and histology	1
1.2 Clinical Features of HGSC	1
1.3 Current Standard Treatment	2
2. Targeted Treatment Ovarian Cancer	
2.1 Personalized medicine	4
2.2 The homologous recombination DNA repair pathway	4
2.3 HRD defects in HGSC	7
2.4 PARP inhibitors and targeting HRD defective cancer	9
3. Beyond HRD: The Other Molecular Features of HGSC	
3.1 TP53 inactivation	10
3.2 Cell cycle perturbations	11
3.3 Other pathways of interest	12
3.4 Summary	13
4. The Neoadjuvant Challenge	
4.1 Neoadjuvant challenge	15
Methods	
1. Patient cohort and clinical data	17
2. DNA and RNA analysis	17
3. Semi-supervised class prediction and consensus clustering	19
4. Statistical analyses	20
Results	
1. Patient cohort selection	22
2. Identification of two biologically defined survival subgroups in the NACT cohort	22
3. Multivariate survival analysis	26
4. Differential gene expression analysis	28
5. Gene set analysis	28
6. Mutation analysis	33
Discussion	
1. Summary	
1.1 Identification of novel molecularly defined outcome groups	35
1.2 Differential expression and gene set analysis	37
1.3 Mutational analysis	38
2. Future Directions	
2.1 Tissue-Microarrays	40
2.2 Matched pre-NACT, post-NACT and time of recurrence samples	41
2.3 Pre-clinical modeling	42
Appendix	
1. Information on the NanoString panel	43
2. Detailed normalization procedure	44
3. TP53 primers for Sanger sequencing	47
4. List of differentially expressed genes	48
References	51

Abstract

Background: The prognosis of patients afflicted with high-grade serous ovarian cancer remains poor and a deeper molecular understanding of the disease is required to improve treatment strategies. Tumor biomarkers, such as mutations in *BRCA1* and *BRCA2*, and aberrant expression profiles, have been linked to improved response in patients with high-grade serous ovarian cancers (HGSC) undergoing primary debulking surgery (PDS). Recently, the use of chemotherapy prior to surgery, neoadjuvant chemotherapy (NACT), has become more widely used. While large-scale genomic studies have detailed the molecular landscape of tumors from patients that underwent primary debulking surgery (PDS) and correlated the identified subgroups to survival, none have done the same for the group of patients with residual tumor after NACT. Residual HGSC after NACT may represent a subset of disease enriched for molecular features of resistance.

Objective: Describe the underlying genomics of NACT-treated HGSC and correlate them to patient progression-free survival (PFS) and overall survival (OS).

Methods: Tumor samples were collected from patients with stage III or IV HGSC before and following neoadjuvant chemotherapy (NACT cohort, $n = 57$ and PDS cohort, $n = 30$). Tumor content was validated by histologic examination. Gene expression analysis was performed using a tailored NanoString-based assay, while next generation sequencing was performed on the MiSeq platform. Semi-supervised clustering, gene set analyses, and the appropriate survival models were used to assess the associations between genetic alterations and survival.

Results: Semi-supervised, internally validated consensus clustering of gene expression data revealed two patient clusters with prognostic relevance. The association with overall survival remained significant after controlling for clinical variables (OS: $p=0.0003$). Pathway based analysis of the differentially expressed genes between the Good and the Bad outcome groups revealed high cell cycle and DNA repair gene expression, and actionable targets in the Bad outcome group. Mutational analysis revealed relatively low *TP53* mutation frequency in both NACT-treated group, but *TP53* mutation status was not associated with survival.

Conclusions: A partial genomic profiling of the intratumoral molecular characteristics observed after NACT in high grade serous ovarian cancers revealed molecular subtypes with prognostic significance.

Résumé

Contexte: Le pronostic des patientes atteintes du cancer de l’ovaire épithélial séreux de haut degré de malignité (COSHM) demeure obstinément faible et une connaissance moléculaire plus profonde de la maladie est nécessaire pour améliorer les stratégies de traitement. Les biomarqueurs de survie, tels que les mutations dans les gènes *BRCA1* et *BRCA2*, ainsi que des profils d’expressions de gènes aberrants, ont été liés à un meilleur taux de survie chez les patients souffrantes de COSHM traitées en première ligne avec une ablation chirurgical (AC) suivit de chimiothérapie. Récemment, l’utilisation de chimiothérapie avant la chirurgie, la chimiothérapie néoadjuvante (CN), est de plus en plus utilisée afin de réduire la taille des tumeurs avant la chirurgie. Tandis que le profil moléculaire de tumeurs dites chimio-naïves, ou provenant de femmes qui ont subit une AC en première ligne, est bien connu, très peu d’études ont tenté d’étudier les tumeurs résiduelles chez les patients qui ont reçu de la chimiothérapie néoadjuvante. Les tumeurs résiduelles après la CN pourraient être enrichies avec des facteurs moléculaires de résistance.

Objectif : Décrire la génomique sous-adjacente de tumeurs traités avec la CN et la mettre en corrélation avec la survie sans progression (SSP) et la survie totale (ST) des patientes.

Méthodes: Des échantillons de tumeurs furent collectés de patientes avec le COSHM de stage III ou IV après la chimiothérapie néoadjuvante (cohorte CN, n = 57). Le pourcentage de cancer dans chaque échantillon fût validé par un pathologiste gynécologue d’expérience. L’expression de 800 gènes fût capturée avec un panneau NanoString personnalisé et la présence de mutations fût évaluée avec la plateforme MiSeq. Des regroupements semi-supervisés, une analyse d’ensemble des gènes, ainsi que les modèles statistiques de survie appropriés ont été utilisés pour évaluer l’association des défauts moléculaires avec la survie des patientes.

Résultats : Des regroupements personnalisés semi-supervisés de l’expression de certains gènes ont révélé deux groupes de patientes au pronostic différent. L’association avec le survie totale est restée statistiquement importante après avoir contrôlé pour les variables cliniques (ST; $p = 0.0003$). L’analyse d’ensemble des gènes a révélé que le groupe avec le pire pronostic avait une haute expression de gènes jouant des rôles important dans le cycle des cellules, ainsi que dans la réparation de l’ADN. L’analyse des mutations démontre un bas taux de mutations dans le gène *TP53* dans les tumeurs traitée avec de la CN.

Conclusions : Une évaluation partielle du génome des tumeurs après le traitement avec de la CN a révélé des caractéristiques importantes dans les tumeurs ayant un mauvais pronostic.

Acknowledgements

First and foremost, I would like to thank Dr. Walter Gotlieb for the opportunity to work in his laboratory. His assistance, support and guidance were invaluable. Your mission, reflected in the work by the Division of Gynecologic Oncology, is truly inspiring. Improving outcomes and quality of life for patients is no easy task and requires much dedication. I hope to emulate work done by you, Dr. Salvador and Dr. Lau in my own practice of medicine, and I wish the Division, and particularly its laboratory, all the best in the future.

I am also grateful for the opportunity to work with the other members of the laboratory: Dr. Amber Yasmeen, Dr. Roy Kessous, Dr. Liron Kogan and Tahira Bolch. Thank you for your assistance and support in running laboratory experiments, as well as your friendship and confidence over the last two years. There is a true spirit of camaraderie and collaboration on this truly unique team that made going to work easy.

I would like to thank Dr. Patricia Tonin for her invaluable mentorship. You are truly a scientific role model to me, and I hope our research paths will cross again in the future. Please extend my thanks to the rest of your laboratory for their advice and friendship.

Thank you to the Division of Experimental Medicine and the Lady David Institute for their support, especially the members of my supervisory committee: Dr. Elliopoulos, Dr. Pantopoulos and Dr. Aloyz.

Finally, thank you to my partner (Annaliese), my parents (Michel and Marie-Chantal), and my sister (Daphné) for their support and love.

Contribution of Authors

Study design. The study was made possible by the clinical data accumulated prospectively over the years by numerous students and clinical fellows. The high-grade serous ovarian cancer database was established principally through the work of Dr. Ido Laskov and Dr. Roy Kessous. I updated the database at the end of 2016 and 2017, and added new patients to the database. Study approval was obtained by Dr. Amber Yasmeeen and Dr. Roy Kessous. I conceived and designed the present study. I was responsible for the coordination and gathering of the data and involved in all of its aspects.

Sample processing. The samples were identified and removed from the biobank by Dr. Kessous, Dr. Yasmeeen and myself. Together, we were also responsible for setting the samples in OCT media, cutting and sliding the samples for staining and pathological review. The staining was performed by the molecular pathology unit of the Lady Davis Institute, and the samples were reviewed by the trained gynecological pathologist of the Jewish General Hospital, Dr. Manuela Pelmus and Dr. Alex Ferenczy. DNA sequencing and RNA counting were performed in the laboratory of Dr. Leon Van Kempen of the Molecular Pathology Unit of the Lady Davis Institute. I performed the PCR reaction for subsequent Sanger sequencing of the *TP53* gene.

Analysis. Dr. Kathleen Klein performed the DNA alignments, determined the optimal RNA normalization method, and created the tumor content control variable using the principal component analysis of the alternative variant allele frequencies. I annotated the VCF files, and designed and performed the filtering procedures. After the first set of samples normalized by Dr. Klein, I performed all of the RNA normalization procedures. I performed all of the statistical procedures described in this manuscript under the advice and supervision of the other authors. For my work, I learned and perfected my knowledge of the R statistical environment, high-throughput statistics, and genomics methodology through readings, classes and practice sets. All authors contributed to the interpretation of the findings.

List of figures and Tables

Figure 1: Homologous recombination DNA double-strand break repair and associated genes.....	7
Figure 2: Synthetic lethal mechanism of PARP inhibition	10
Figure 3: Representations of the commonly altered genes and pathways in HGSC.....	14
Figure 4: Graphical abstract	16
Figure 5: Semi-supervised, internally validated consensus clustering.....	24
Figure 7: Heatmap of the 50 most differentially expressed genes	30
Figure 8: Gene set analysis of the differentially expressed genes	31
Figure 9: Representation of the <i>FOXMI</i> -mediated transcriptional network.....	32
Figure 10: Mutational analysis	34
Table 1: Neoadjuvant SAM analysis results.....	22
Table 2: Patient characteristics by outcome group.....	25
Table 3: Multivariate survival analysis	27
Supplementary Table 1	43
Supplementary Table 2	45
Supplementary Table 3	45
Supplementary Table 4	46
Supplementary Table 5	47
Supplementary Table 6	50

List of abbreviations

OC = Ovarian Cancer
HGSC = High-Grade Serous Ovarian Cancer
STIC = Serous Tubal Intraepithelial Carcinoma
NACT = Neoadjuvant Chemotherapy
PDS = Primary Debulking Surgery
PARP = Poly(ADP-Ribose) Polymerase
HRD = Homologous Recombination DNA Repair
DSB = Double Strand Break
NHEJ = Non-Homologous End Joining
ssDNA = single stranded DNA
FA = Fanconi Anemia
HBOC = Hereditary Breast and Ovarian Cancer
TCGA = The Cancer Genome Atlas
CNV = Copy Number Variation
SSB = Single-Strand Break
MMR = Mismatch Repair
FDR = False Discovery Rate
PAC = Proportion of Ambiguous Clusters
CI = Confidence Interval
HR = Hazard Ratio
CC = Cell Cycle
CR = Chromatin Remodeling
GSA = Gene Set Analysis
DE = Differential Expression
LOH = Loss of Heterozygosity
WT = Wild Type
TIL = Tumor Infiltrating Lymphocyte
TMA = Tissue Microarray
PDX = Patient Derived Xenografts

Introduction

1. Clinical Features and Standard Treatment Strategies

1.1 Epidemiology, anatomy and histology. Ovarian cancer (OC), encompasses cancers originating from the peritoneum, fallopian tubes and/or the ovaries and is the most lethal gynecologic cancer¹. In Canada, only 2.7% of new cancer cases are expected to be of ovarian origin in 2017, but the same disease is expected to be responsible for approximately 5% of all cancer-related death, which is one of the highest death-to-incidence ratio out of all types of cancer¹. Once treated as a single, homogeneous disease, clinicians now subdivide ovarian cancer cases into various subtypes with distinct clinical presentations, cells of origin, molecular features and treatment strategies^{2,3}. Cancers of epithelial origins, which account for approximately 90% of ovarian cancers, are the most common and can generally be classified into one of four histological subtypes: serous, endometrioid, clear cell and mucinous⁴⁻⁶. In addition to histological subclasses, ovarian cancer cells are graded based on their appearance compared to normal cells, also known as differentiation. Low grade cancer cells almost appear normal (well differentiated), tend to grow slowly and are less likely to spread. High grade cancer cells are poorly differentiated, grow quickly and are more likely to spread^{4,6}. Patient survival varies greatly depending on the pathological subclassification⁷. Unfortunately, the high grade serous ovarian cancer (HGSC) subtype is both the most commonly diagnosed and the deadliest, accounting for 70-80% of ovarian cancer deaths and showing little improvement in long term survival in the past thirty years^{7,8}.

1.2 Clinical features of HGSC. Early HGSC symptoms are often non-specific and attributed to other, non-life threatening causes. As a result, most HGSCs are diagnosed at a late stage (III or IV), which means that the disease has already metastasized beyond the pelvis. For years, the lack of precursor lesions on women's ovaries puzzled pathologists, and was attributed

to lesion disappearance over the course of carcinogenesis. In recent years, with the advent of risk reducing salpingo-oophorectomy in women with germline mutations in high risk genes, such as *BRCA1* and *BRCA2*^{9,10}, lesions were identified in the fallopian tubes and, more specifically, the fimbria¹¹. Serous tubal intraepithelial carcinomas (STIC) have been found in many advanced stage HGSCs, including those of non-hereditary origins, and are thought to be the precursor lesions to HGSCs¹¹. In addition, disease originating from the peritoneum is thought to arise from tubal epithelium as a result of endosalpingiosis, or the ectopic growth of tubal epithelium¹². The establishment of molecular clonal relationships between STIC lesions and primary HGSCs strengthens this hypothesis^{13,14}. However, additional studies are needed, since only about 60% of HGSC have identifiable STICs, and a recent phylogenetic analysis associated STICs with intraepithelial HGSC metastases to the fallopian tube, contradicting the assumption that STICs are always the precursor lesion to HGSC¹⁵. In any case, the lack of consensus, easily identifiable precursors to HGSC makes early detection improbable, and large scale randomized controlled trials failed to implement successful screening procedures for ovarian cancer^{16,17}. Due to the lack of efficacious screening procedures and the stagnant long-term survival rates despite an improved understanding of the clinical features of HGSC, much of the current research efforts focus on molecularly defining the primary disease in order to efficaciously and rationally treat it.

1.3 Current standard treatment. Over the last forty years, the introduction of cisplatin¹⁸ in 1978 and paclitaxel¹⁹ in 1992 as chemotherapeutic agents is partially responsible for the steady increase in five-year survival rates. Surgical advancements have also increased survival in patients with ovarian cancer. Residual disease after surgery, even as small as 1 mm in size, is inversely correlated with survival even in patients with a high tumor burden^{20,21}. Today, the current standard treatment for HGSC is primary surgical cytoreduction to no visible residual disease, paired with

adjuvant (i.e. given after surgery) carboplatin and paclitaxel combination chemotherapy. Despite these advancements in our understanding of HGSC treatment, twelve-year survival rates, remain relatively unchanged²² and new treatment strategies are sought every day.

Because of the significant survival disadvantage observed when optimal surgical cytoreduction is not reached, neoadjuvant chemotherapy (NACT) has become increasingly popular as an alternative in patients for whom a good surgical outcome is unlikely. Neoadjuvant chemotherapy is the administration of chemotherapy prior to surgical cytoreduction in order to reduce the tumor burden. Since achieving optimal cytoreduction is difficult in patients with advanced stage disease and high tumor burden, NACT is designed to lower that burden prior to surgery, and by that increase the rate of patients achieving optimal cytoreduction and reduce the complication rates. In two large scale, randomized controlled trials comparing primary surgical cytoreduction and neoadjuvant chemotherapy, no differences in median survival were observed between the two groups^{23,24}. NACT has also been linked to lower rates of surgical complications and side effects, making it a viable option for some women with increased tumor burden or women that are not stable enough for upfront aggressive surgery at the time of diagnosis²⁵. As a result, the use of NACT has doubled from 1990 to 2010 and is now used in 40% of women with HGSC as the primary treatment strategy²⁶.

In short, primary debulking surgery (PDS) followed by 6-8 cycles of carboplatin and paclitaxel combination chemotherapy remains the preferred treatment strategy for HGSC. In selected patients, 3-6 cycles of carboplatin and paclitaxel combination neoadjuvant chemotherapy followed by surgery and additional chemotherapy may be beneficial. Due to its aggressive nature and its lack of treatment options, HGSC is a prime candidate for new personalized treatment options that take the genomic features of the disease into consideration.

2. Targeted Treatment Ovarian Cancer

2.1 Personalized medicine. The current treatment of ovarian cancer, namely carboplatin and paclitaxel combination chemotherapy, is cytotoxic. This means that the drugs are toxic to all living cells, not just cancer cells. These agents preferentially kill cancer cells by taking advantage of their ability to divide rapidly. In other word, if a cell's DNA is damaged and tries to divide before repairing the damage caused by cytotoxic agents, it will die. On the other hand, healthy cells should be able to repair the damage before they divide and live. However, some healthy cells that divide rapidly, such as hair or gut cells, are heavily affected by cytotoxic agents, which explains side effects such as hair loss and extreme nausea experienced by some cancer patients. Targeted treatment, personalized care and precision medicine are synonyms for a single clinical aspiration: to target a faulty molecular pathway unique to cancer cells in order to specifically kill them while leaving healthy cells unscathed. Such treatments are made possible by the substantial biomedical understanding of carcinogenesis gained over the past twenty years, and the technological feats that allow scientists to read the genome of cancer cells on a scale and speed that is continuously expanding (reviewed in Reuter et al.²⁷).

For HGSC, the only successful and currently approved targeted therapy is the inhibition of poly(ADP-ribose) polymerase (PARP) in cancers with a known vulnerability in the homologous recombination DNA repair pathway (HRD)²⁸⁻³¹.

2.2 The homologous recombination DNA repair pathway¹. Faithful genome replication and repair is essential for genomic integrity within organisms. DNA breaks can take multiple forms. The most deleterious one is arguably the double-strand breaks (DSB), in which both strands

¹ For an exhaustive, recent review of the different DNA repair pathways, including DSB repair, please refer to Ranjha et al.³²

of the chromosome are cleaved, making the cell vulnerable to important losses of genetic information if left unrepaired, or if repaired inaccurately. They can be caused by ionizing radiation, reactive oxygen species and replication fork breakage during DNA replication. Homologous recombination-mediated repair is relatively error-free compared to the alternative non-homologous end joining (NHEJ) DSB repair pathway and is the preferred method of DSB repair in the G2 phase after DNA replication³³.

The HRD procedure can be separated into three different stages (figure 1). The pre-synapsis step requires the formation of single-stranded DNA at the breakage site to allow for homology search on the sister chromatid. For this purpose, DNA ends are resected by 5' to 3' degradation to form 3' single-stranded DNA (ssDNA) overhangs. This is accomplished by the MRN complex, composed in humans of MRE11³⁴, RAD50³⁵ and NBS1³⁶, and the other necessary accessory proteins EXO1³⁷, DNA2³⁸, CtIP³⁹ and BLM³⁸. After processing, binding of RPA is necessary to stabilize the ssDNA in order to avoid further degradation or the formation of secondary structures^{40,41}.

Following pre-synaptic processing, RPA is replaced by RAD51 filaments to initiate the homology search^{42,43}. However, this dynamic process is slow and limited by the large amount of RPA present on ssDNA that creates a barrier against RAD51^{43,44}. To counter this effect, multiple mediators have been identified that facilitate the RPA-Rad51 exchange on the ssDNA overhangs. BRCA2,^{45,46} its molecular partner PALB2⁴⁷, and, to a much lesser extent, RAD52⁴⁸, are a few of the identified facilitators of the RPA-RAD51 exchange.

The exact mechanism by which RAD51 finds the homologous sequence is poorly understood in eukaryotes. Much of the mechanistic insights come from research on its bacterial homolog, RecA^{49,50}. In short, following strand invasion and D-loop formation, homology to the

ssDNA sequence of interest is suggested to be tested in smaller groups of nucleotide triplets, quickly discarding non-homologous sequences until satisfactory homology is found in the sister chromatid⁵¹.

Following a successful homology search, the post-synaptic step involve strand elongation and ligation of the broken ends, forming a double Holliday Junction⁵². The exact mechanism by which Holliday Junctions are resolved in human is still debated. While HRD-mediated pathways may result in crossovers of genetic information in meiotic cell division⁵³, HRD DSB repair seems to favor a non-crossover product through a crossover suppression mechanism that is still under exploration⁵⁴⁻⁵⁷.

In short, DSB repair through HRD is an extremely delicate and intricate process that involves an ever-increasing number of molecular players (figure 1). In some instances, the proteins identified have multiple roles in the process. BRCA1, for example, is a master regulator of HRD that both accelerates ssDNA processing through direct interactions with CtIP in the pre-synaptic stage⁵⁸, and facilitates Rad51 loading by regulating BRCA2 through PALB2^{47,59,60}. In other instances, the signaling processes are multidimensional and require large complexes, such as the Fanconi Anemia (FA) complex composed of thirteen proteins, each playing an important role in DSB repair (reviewed by Moldovan and D'Andrea⁶¹; genes listed in figure 1). Due to its complexity, its dependence on its components, and its crucial role in keeping the genome's integrity, the HRD pathway is unsurprisingly faulty in many human cancers, including approximately 50% of all HGSC⁶².

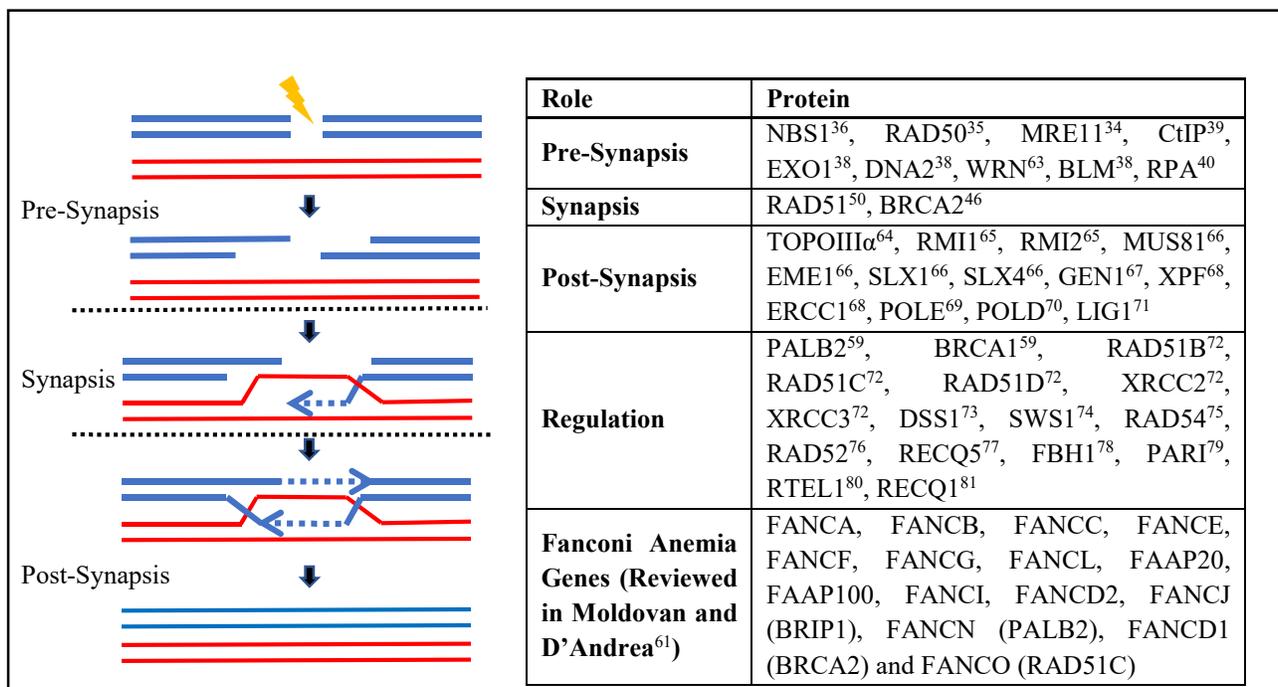


Figure 1: Homologous recombination DNA double-strand break repair and associated genes. Schematics of DNA double-strand break repair. Following damage, the DNA ends are resected by 5' to 3' degradation to form 3' ssDNA overhangs. With the help of the appropriate accessory protein, the ssDNA stretch invades the sister chromatid in search of homology, forming a D-loop. When the appropriate section is found, the broken DNA strands are polymerized and ligated to the free ends, creating a double Holliday Junction. The junctions are preferentially resolved in a non-crossover fashion, effectively ending the accurate repair of DNA DSB without the loss of genetic information. Next to the figure is a list of the main molecular players in DNA homologous recombination DSB repair. This list is not exhaustive.

2.3 HRD defects in HGSC. Pierre Paul Broca observed in the nineteenth century that some women from cancer-afflicted families were more prone to develop breast cancer.⁶³ This eventually led to the identification of rare germline mutations in the HRD pathway, or mutations arising in the germ cells, that could be transmitted from one generation to the next and predispose carriers to an increased risk of certain types of cancer⁶⁴. It is thought that one-fifth of ovarian cancer cases arise from this familial predisposition called Hereditary Breast and Ovarian Cancer (HBOC)^{65,66}. Women with HBOC have an estimated 40-59% risk of developing ovarian cancer by age 70^{67,68}, which is much higher than the 1.3% lifetime risk observed in non-carriers¹. However, while *BRCA1* and *BRCA2* may be the best-described HBOC susceptibility genes^{64,69}, they are thought to account for a mere 25% of all HBOC cases⁷⁰. While several genes have been found to confer

susceptibility to ovarian cancer^{66,71}, a significant proportion of families suffering from HBOC do not have an identifiable genetic explanation for the syndrome.

Most germline mutations occur at a single locus inherited from the mother or father. Thus, assuming allelic balance, or the equal transcription of the mutated and wildtype alleles, only half of the translated protein originating from that locus should be functional. In some cases, this results in an abnormal phenotype and such a gene is deemed haploinsufficient. *BRCA1* is thought to be haploinsufficient, making carriers of a pathogenic variant more likely to accumulate mutations and eventually develop cancer^{72,73}. In other cases, a single functional allele may be enough for proper molecular function. In those cases, cancer may arise as a result of the somatic loss of the wild type allele, a theory known as the “Two-Hit” hypothesis⁷⁴ that has since been described in ovarian cancer⁷⁵. However, somatic mutational driver events, which are non-hereditary mutations arising from cells in the body that are not germline, may also occur and be deleterious in the absence of germline mutations. Given that four fifths of ovarian cancers are sporadic, or arise without hereditary causes, much attention has been given to uncovering the somatic drivers of the disease.

Data from The Cancer Genome Atlas (TCGA) project, a large scale, multi-institution effort to unveil the somatic landscape of HGSC at the genetic, transcriptomic, epigenetic and proteomic levels, revealed that somatic mutations in HGSC are generally confined to very specific genes and pathways^{62,76}. HGSC is characterized by a relatively light mutational burden compared to other epithelial cancers, and recurrent mutations are rare and limited⁷⁷. *BRCA1*, *BRCA2*, and *NF1* are the top 3 recurrently somatically mutated genes in HGSC and are mutated in 5%, 3% and 6% of cases, respectively. Most other somatic mutations described in large scale analyses are found in $\leq 1\%$ of cases when analyzed on a gene-by-gene basis. However, when analyzed as a group of genes within specific molecular pathways, it was found that most of the mutations in ovarian

cancer, somatic or germline, occur in the homologous recombination DNA repair pathway and that such defects affected up to 37% of patients with HGSC.⁷⁶ Whether all mutations found actually affect the phenotype of the cancer, however, remains up for debate.

Despite a light mutational burden, HGSC is heavily affected by copy-number variations (CNV), epigenetic changes and other chromosomal rearrangements. Again, the HRD pathway is disproportionately affected by such changes. Promoter hypermethylation was not only found in *BRCA1* in 11.5% of cases, but also in *RAD51C* in another 8% of cases⁶². Other HRD-associated genes such as *EMSY* and *PTEN* were also found to be aberrantly amplified in 8% of case and deleted in 7% of cases, respectively⁶². When all of the different kinds of HRD deficiencies are taken into account, TCGA estimated that around 50% of HGSC cases were HRD-deficient. This observation corroborated a previous large-scale gene expression study of HGSC suggesting that many *BRCA*-wildtype tumor displayed a *BRCA* mutated-like, or an HRD deficient, gene expression profile⁷⁸.

2.4 PARP inhibitors and targeting HRD defective cancer. PARP is a protein necessary for single stranded break (SSB) DNA repair⁷⁹. It is responsible for the recognition and flagging of the breakage site through a process dubbed parylation, which is the formation of poly(ADP-ribose) chains at the breakage site. This process recruits the other molecular players necessary for successful repair (reviewed by Lord and Ashworth⁸⁰). If a single-strand break goes unrepaired, subsequent chromosome synthesis creates a double strand break (figure 2). If a cell is HRD-deficient and cannot repair the newly formed DSBs, they accumulate and the cell eventually undergoes apoptosis. PARP inhibitors, by disrupting SSB repair, lead to DSB that cannot be repaired in HGSC patients with *BRCA1* or *BRCA2* germline mutations. PARP inhibitors thus lead to synthetic lethality (figure 2), meaning they specifically target ovarian cancer cells, while not

affecting normal cells that have functional BRCA, and normal DSB repair. This has led to the approval of oral PARP inhibitors for the treatment of cancers with BRCA mutations^{28,29,81,82}. However, such germline mutations are only present in about 15% of HGSC cases and an additional 35% of patients with HRD-deficient cancers also benefit from such therapies⁸³. The lack of reliable predictors of somatic HRD deficiency beyond mutations remains a challenge, and none of them are currently approved for use in the clinic for HGSC.

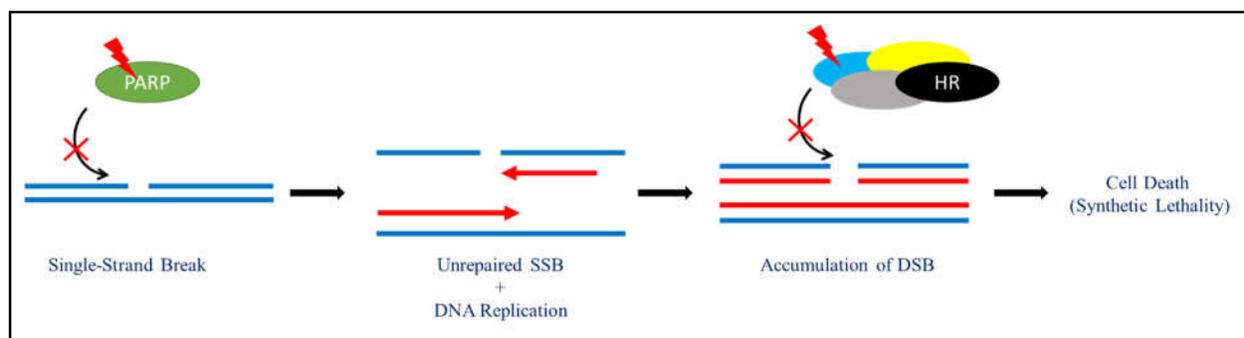


Figure 2: Synthetic lethal mechanism of PARP inhibition. Under normal circumstances, PARP signals the presence of SSBs. If a SSB is left unrepaired due to PARP inhibition, DNA replication creates a DSB. In turn, if the HRD machinery is damaged in any sort of way, the DSB are also left unrepaired, leading to their accumulation. Accumulation of cellular DSBs triggers apoptosis, and the cell dies. Cells with functional HRD are left relatively undamaged by PARP inhibition.

3. Beyond HRD: The Other Molecular Features of HGSC

3.1 TP53 inactivation. Biallelic loss of the tumor suppressor *TP53* are almost ubiquitous in HGSC (85 – 100%)^{62,84–86}. *TP53* encodes the p53 protein, which is most notably responsible for cell cycle arrest and apoptosis in response to stress (reviewed extensively by Bieging et al.⁸⁷). Its transcriptional targets arrest the cell cycle (p21, WAF1, Cip1), promote apoptosis (Bax) and block mitogenic growth factors (IGF-BP3). Moreover, due its wide range of targets and potentially damaging functions, p53 is regulated by a plethora of post-translational modifications and interactions, each with its own upstream trigger in response to unique stressors, and similarly unique downstream effects (reviewed by Kruse and Gu⁸⁸). In normal cells, p53 is transient and possesses a very short half-life, because it is also responsible for the transcription of its own

negative regulator, the E3-ligase MDM2 that targets it for ubiquitin-mediated proteosomal degradation⁸⁹⁻⁹¹. Its quaternary structure involves the formation of a tetramer involving four p53 molecules^{92,93}

Loss-of-function occurs through a combination of mutations in its DNA binding domain and chromosome 17 loss⁹⁴. Interestingly, while most tumor suppressor associated-mutations are frameshifts, *TP53* mutations are mostly point mutations in its DNA binding domain. Slight changes in p53 conformation because of a single amino acid shift entirely prevents it from binding its highly specific target DNA regions.

Because of its crucial role in regulating apoptosis and its sensitivity to point mutations, it comes as no surprise that p53 inactivation is one of the early events of carcinogenesis in multiple types of cancer. In fact, it is widely acknowledged as the most recurrently defective gene in human cancers^{87,95}. Mutations in *TP53* are an exclusive feature of the serous subtype in ovarian cancer, and its status is often used by pathologists to derive the final diagnosis of this particular subtype.

3.2 Cell cycle pathway perturbations. As previously discussed, HRD deficiency is the predominant feature of about half of all HGSC cases. However, much less is known about the molecular drivers of the other half of cases, and no targeted therapies have been developed for them. What we do know is that they tend to fare worse than patients with HRD deficiencies under standard therapeutic conditions⁶². HRD-deficient tumors are more sensitive to cytotoxic agents due to their inability to repair the DSBs caused by the standard chemotherapeutic agent carboplatin, a DNA cross-linking agent⁹⁶.

One of the most cited defects in HRD-proficient HGSC cases is *CCNE1* focal amplification⁶². Cyclin E1 is one of the main regulators of G1 to S phase transition and is under tight regulation to prevent uncontrolled cell division⁹⁷. While the level of *CCNE1* mRNA is not

necessarily associated with survival, overexpression is mutually exclusive from HRD deficiency, probably due to a synthetic lethal relationship between CCNE1 and members of the HRD pathway⁹⁸. Selective targeting of CCNE1 in cell lines overexpressing this same gene as had some recent successes, but further validation in pre-clinical models is needed to exploit this new strategy in the clinical setting⁹⁹. Functionally, CCNE1 is thought to inhibit Rb, which is itself an inhibitor of E2F1, one of the main transcription factors driving the cell cycle (reviewed by Polager and Ginsberg¹⁰⁰). Such defects are thought to drive the cell cycle to bypass G1/S and G2/M checkpoints in cancer cells. Other recurrently overexpressed or amplified cell cycle genes are *CHEK1*, *CHEK2*, *ATR*, *ATM* and *FOXMI*⁶².

3.3 Other pathways of interest. Notch signalling was also found to be deregulated in 22% of HGSC cases in the TCGA cohort. *NOTCH3* and its activators *JAG1*, *JAG2*, and *MAMLI-3* were found to be amplified or to bear activating mutations that activated the signaling pathway⁶². PI3K and RAS signaling pathways were deregulated in an additional 45% of cases, notably through *PTEN* deletions and loss-of-function mutations, *PIK3CA* activating mutations or amplifications, *NF1* deletion or loss-of-function mutations, and *KRAS* amplifications⁶². Empowered by successes in melanoma and non-small cell lung cancer, immune signaling pathways and immunotherapy have garnered increased interest in oncology over the last few years (reviewed by Khalil et al.¹⁰¹). However, in HGSC, immunotherapy has enjoyed limited successes despite its potential in a subset of patients. For example, a higher mutational burden, a higher expression of PD1 and PDL1 proteins, and an increased number of tumor infiltrating lymphocytes have been shown to increase response to various immunotherapies¹⁰¹. In HGSC, patients with BRCA mutations and patients who received neoadjuvant chemotherapy were shown to possess a number of those characteristics, making them possible candidates for such therapies^{102,103}. However, the number of studies showing

such correlations are limited, and further investigation would strengthen the rationale behind immunotherapy use in a defined subset of patients with HGSC.

3.4 Summary. The most common, known defective molecular pathways in HGSC are summed up in figure 3A-C. While multiple pathways other than HRD have been described as deregulated in HGSC, little is known about how these defects are featured after neoadjuvant chemotherapy or at recurrence. A better understanding of how they may be selected in or out following chemotherapy may improve outcomes for some patients HGSC by guiding subsequent treatment modalities.

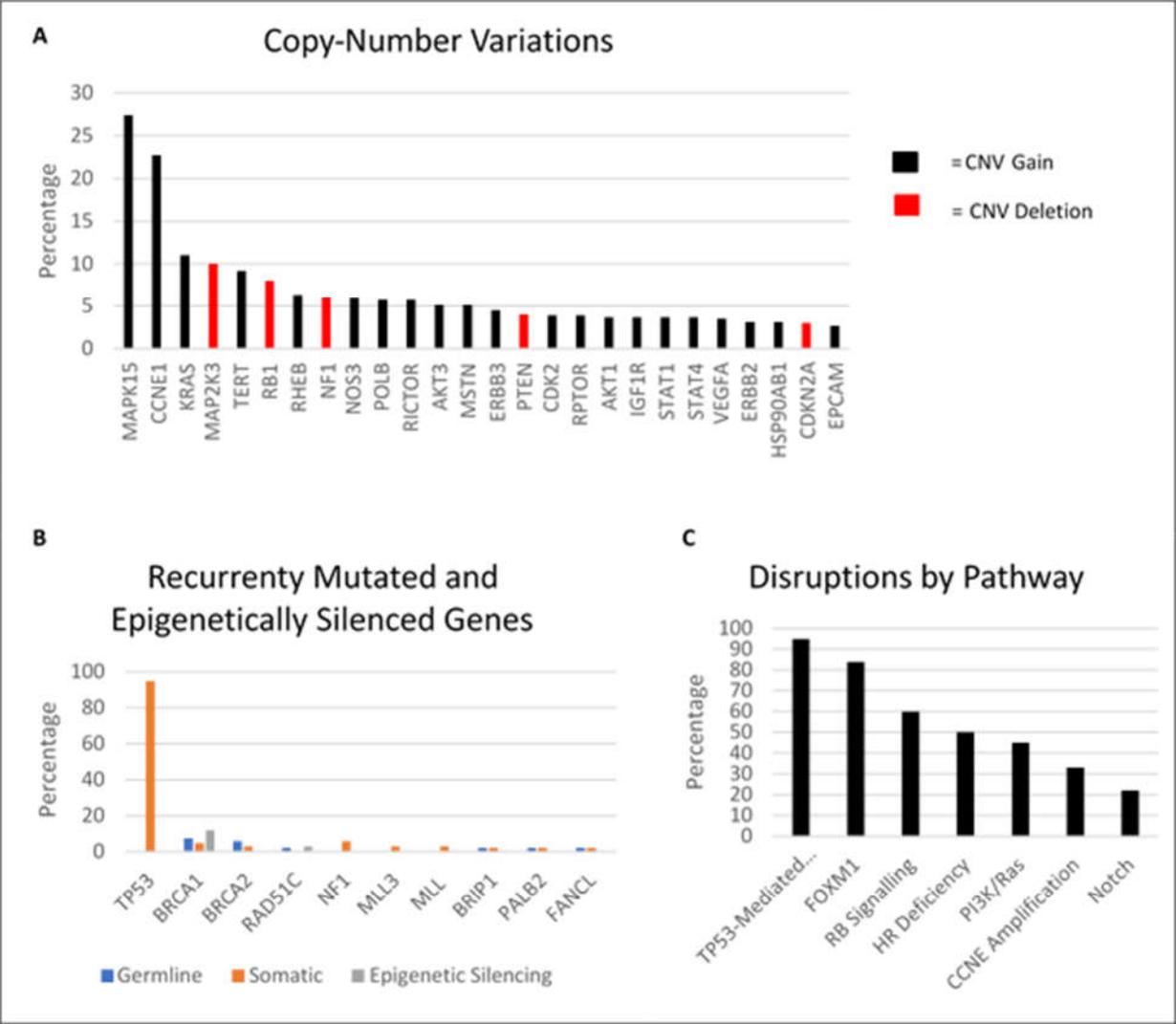


Figure 3: Representations of the commonly altered genes and pathways in HGSC. A) Commonly copy-number varied genes. Red bars represent copy-number losses and black bars represent copy-number gains. B) Recurrently mutated genes in the germline (blue bars) or somatic cells (orange bars). Grey bars represent epigenetic silencing of the gene of interest. C) Summary of the pathways recurrently altered in HGSC.

4. The Neoadjuvant Challenge

Critics of neoadjuvant chemotherapy use in HGSC suggest that tumor burden reduction through chemotherapy leads to microscopic disease sites that elude surgeons at the time of surgery, which may cause early recurrence. Others also claim that it may cause early resistance to platinum agents. Such claims, however, are largely unsubstantiated or anecdotal. NACT is a viable option for many patients, especially those with advanced stage disease, unresectable tumor burden or important comorbidities. In this new era of precision medicine, understanding the molecular drivers of residual HGSC after neoadjuvant chemotherapy may allow for the optimization of neoadjuvant and adjuvant treatment procedures.

Molecular knowledge of post-NACT HGSC samples, is lacking. Part of the reason is that widespread neoadjuvant administration is relatively new in HGSC treatment, and only recently has it seen an increase in use¹⁰⁴. One study reports that *BRCA1* or *BRCA2* mutated cells are preferentially killed during neoadjuvant chemotherapy¹⁰⁵. Instances of *BRCA1* germline mutation reversions in residual tumors¹⁰⁶ have also been reported, while others have observed changing levels of tumor-infiltrating lymphocytes¹⁰². It was also shown to lower the cellular levels of PARP in residual disease¹⁰⁷. However, to our knowledge, none have looked deeper into the genomics of post-NACT HGSC tumors or associated the post-NACT features to outcome.

The goal of the current study is to partially characterize the genome of residual tumor after neoadjuvant chemotherapy and correlate those characteristics to patient outcome (Figure 4).

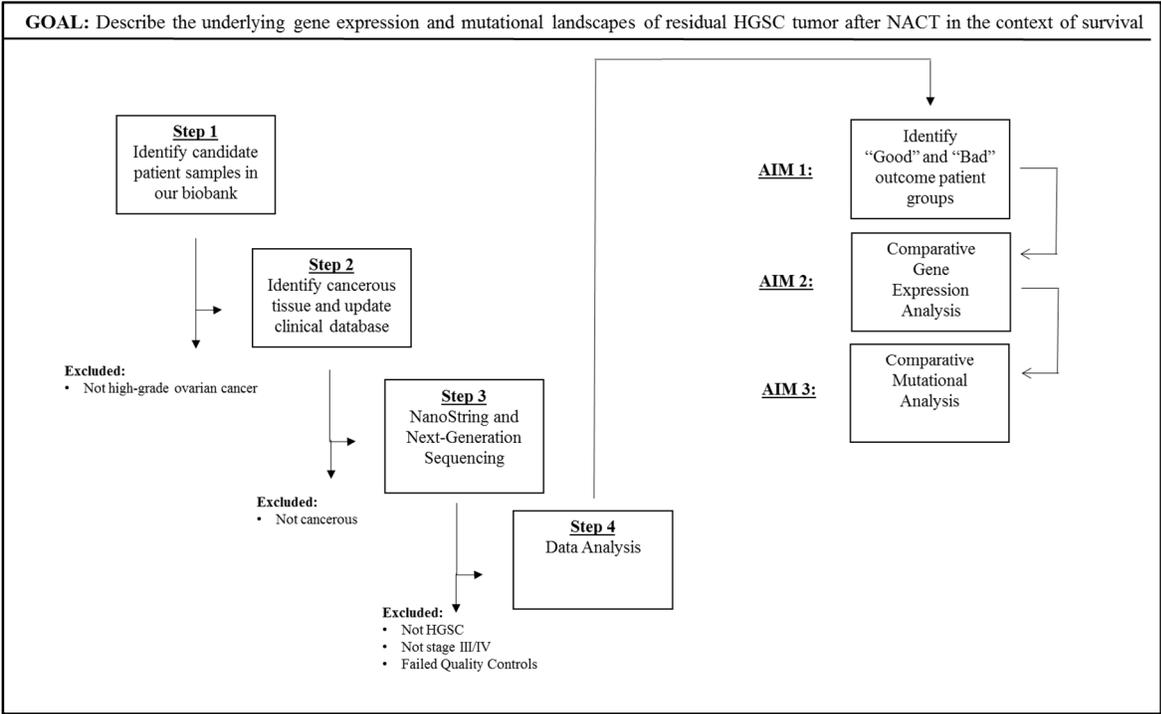


Figure 4: Graphical Abstract

Methods

1. Patient cohort and clinical data. For patients who were diagnosed and treated for HGSC at our institution between 2003 and 2015, tissue and blood samples were collected at the time of surgery and stored in the gynecologic oncology tumor bank (protocol #03-041). All patients participating in biobanking and research activities gave informed written consent. During the post-surgical surveillance period, follow-up examinations were performed at four-month intervals during the first two years from diagnosis, at six-month intervals during the fourth and fifth year, and yearly thereafter. For each patient, information such as age, body mass index (BMI), histologic type, tumor grade, FIGO stage, extent of cytoreduction, and chemotherapy treatment history, was collected in a prospectively maintained clinical database. Patient identifications were anonymized to prevent identification. This study was approved by the Jewish General Hospital Research Ethics Board (#15-070). Progression-free survival was defined as the time from diagnosis to evidence of recurrence by imaging, to the time of death or to the latest follow-up. Overall survival was defined as the time from diagnosis to the time of death or latest follow-up.

2. DNA and RNA Analysis. Approximately five hundred, fresh-frozen high-grade ovarian cancer samples were available in our biobank. Priority was given to those with matched complete clinical history, and approximately two hundred samples were selected for further processing. For each of them, 12 μ m sections were cut and subsequently stained with hematoxylin and eosin (H&E). Each slide was reviewed by a gynecological pathologist in order to verify histology and estimate the cancer content. Non-high grade ovarian cancer samples, and samples with low cancer content were removed from subsequent procedures. RNA and DNA were then serially extracted from the selected samples using the Bio Basic All-in-One DNA/RNA/Protein Mini-Prep Kit (Bio Basic Inc., Markham, ON, Canada). RNA and DNA concentrations were measured using the

NanoDrop ND-100 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). 132 samples had the appropriate mRNA and DNA quality and quantity.

Gene expression was measured using the NanoString PanCancer Pathway Panel (NanoString Technologies, Seattle, WA, USA). The panel contains probes against 770 genes implicated in carcinogenic pathways, curated from data by The Cancer Genome Atlas (TCGA). The nCounter SPRINT Profiler (NanoString Technologies, Seattle, WA, USA) was used to run the reactions. In addition to the 770 genes present on the default panel, 30 additional genes of interest were added for their potential involvement in ovarian cancer biology. Normalization of the raw NanoString data was performed using the nSolver Analysis Software v4.1 (NanoString Technologies, Seattle, WA, USA). The optimal normalization method was determined by testing each of them using technical replicates, and measuring the correlation between the replicates for each normalization method. In short, no background subtraction was performed and raw counts were normalized to the geometric mean of an optimal number of reference genes by the NSolver Software. The normalized counts were log₂ transformed for subsequent analyses (See sections 1 and 2 of the appendix for more details on the normalization procedure and the genes included in the analysis).

For DNA analysis, hotspot next generation sequencing was performed on the Illumina MiSeq (Illumina Inc., San Diego, CA, USA). The Roche Nimblegene TruSeqLT (Roche, Basel, SWI) preparation kit was used to create a library of 400 targeted regions in 168 genes of interest involved mostly in homologous recombination or DNA repair. Results were aligned against the hg38 UCSC Genome Browser assembly (NCBI Assembly ID: GRCh38, GCA_000001405.15). If multiple variants were identified at a single locus of the same sample, the variant with the maximum frequency value was used. The Ensembl Variant Effect Predictor was used to annotate

the resulting VCF files. In short, each gene was compared to its canonical transcript in the Ensembl database, and each variant was matched to existing cases in mutation repositories (e.g. COSMIC, ClinVar). Since drivers of carcinogenesis are thought to be sporadic in the population, only missense alleles with a population allele frequency below 1.5% in the gnomAD database¹⁰⁸ were kept for further downstream analysis. The raw BAM files were visualized using the Integrated Genome Viewer¹⁰⁹, if needed. Synonymous and intronic mutations were removed, unless the later occurred within three base-pairs of a coding exon to account for splice-region mutations. The pathogenicity of missense mutations was assessed *in silico* using the following tools: PolyPhen-2¹¹⁰, Sift¹¹¹, M-CAP¹¹², MutationAssessor¹¹³, and REVEL¹¹⁴. Mutations were kept if predicted to be pathogenic by at least three of the five predictor algorithms.

In *TP53*, next-generation sequencing was designed to capture exons 5 through 9, which captures approximately 80% of mutations occurring in the gene. Sanger sequencing was also performed on exons 4 and 10. The sequences of the primers used were generously provided by Dr. Patricia Tonin, and their sequence can be found in section 3 of the appendix. The DNA regions of interest were amplified using the Qiagen Hotstar Taq Plus DNA polymerase kit according to the manufacturers instructions (QIAGEN Inc., Toronto, ON). The reactions were run on an Eppendorf Mastercycler ep Gradient S (Eppendorf, Hamburg, GE) at 95°C for 5 minutes, 35 cycles of 94°C for 45 seconds, 60°C for 45 seconds and 72°C for 1 minute, and, finally, 72°C for 10 minutes. The sequencing reactions were performed by Genome Quebec.

3. Semi-supervised class prediction and consensus clustering. The semisupervised method of Bair et al.¹¹⁵ was used to identify biologically relevant survival sub-groups. First, the entire NACT cohort was randomly separated into either a training or testing cohort. Second, a cox proportional hazard analysis was performed using the Significance Analysis of Microarrays

(SAM) tool¹¹⁶. In short, SAM measured the strength of the relationship between each gene of our dataset and the overall survival of each patient in the training set. A thousand permutations of the survival times were performed by SAM to identify the significant genes at each permutation in order to estimate the False Discovery Rate (FDR). Genes with a q-value < 0.25 were kept for further analysis. SAM has the advantage of being designed specifically for biological experiments and controlling for multiple testing. Third, with the genes identified, consensus clustering¹¹⁷ was performed. In short, k-means clustering of the Euclidian distances between each sample was repeatedly performed using 80% of the data at random for each step. The proportion of times two samples occupy the same cluster was calculated for a set number of clusters ($k = 1$ to 4), creating a consensus score for each patient pair. A value of one represents two samples clustering together all the time and a value of zero represents two samples never clustering together. The optimal number of clusters was determined by evaluating the relative change in consensus between k and $k-1$ clusters, and the proportion of ambiguously clustered pairs (PAC) method¹¹⁸. Survival between the two identified groups was assessed by the Kaplan-Meier method, and the differences calculated using the Log Rank test. The same clustering method was then applied to the testing set to validate the robustness of the gene expression signature.

4. Statistical analyses. Each gene was assigned to one or more molecular pathways by the NanoString bioinformatics team based on curated data from the literature. The canonical pathways represented were: homologous recombination DNA repair (HRD), mismatch DNA repair (MMR), DNA repair – Other (i.e. not HRD or MMR), Wnt signaling, Cell Cycle, Hedgehog signaling, Jak-STAT signaling, epithelial to mesenchymal transition, Notch signaling, TGF beta signaling, PI3K pathway, Ras signaling, transcription regulation, mitotic maintenance, and chromatin modification. A pathway score was derived for each pathway in each patient using

singular value decomposition of their expression data, as previously described by Tomfohr, Lu & Kepler¹¹⁹. Differential expression was measured using the negative binomial generalized linear model. To further control for the possible variable tumor content between each sample, a principal component analysis was performed on the somatic variant allele frequencies for all variants and extracted the eigenvector of the first principal component, as previously described¹²⁰. This was used as a control variable weighing the tumor content of each sample in order to model gene expression differences holding the tumor content even. The Pearson correlation coefficient between each pathway was measured and the difference in pathway scores assessed using an unpaired t-test. The pathway scores and the differential expression analyses were computed using the built-in functionalities of the NSolver software, while the heatmaps, correlation, boxplots and t-tests were generated in the R statistical environment. The p-values for the differential expression analysis and the t-statistics of the pathway scores were adjusted by the Benjamin-Hochberg FDR method.

Results

1. Patient cohort selection One hundred and thirty-two samples had previously been analysed using the NanoString gene expression assay and next-generation sequencing as part of a larger effort by our group to molecularly characterize the ovarian cancer samples from our tumor bank. For this study, 57 samples were included based on the following criteria: complete mRNA expression and DNA mutation data, stage III or IV cancer, and high-grade serous on final pathology. For the mutational analysis, 30 chemotherapy-naïve, primarily debulked, stage III/IV HGSC samples were included for mutational frequency comparisons.

2. Identification of two biologically defined survival subgroups in the NACT cohort.

The mRNA expression of 800 genes was measured in 57 stage III/IV high-grade serous ovarian

Gene	Score (d)	Q- Value
CDC25A	2.925	0
CDC25C	2.569	10.147
KIF2C	2.389	10.147
CDC6	2.317	10.147
CCNE2	2.275	10.147
CCNA2	2.272	10.147
BRIP1	2.224	10.147
CHEK1	2.136	10.147
VEGFA	2.108	10.147
BNIP3	2.096	10.147
PKMYT1	1.982	14.496
PTTG2	1.943	14.496
FOXM1	1.94	14.496
HIST1H3H	1.93	14.496
POLR2H	1.866	22.197
STMN1	1.841	22.197

Table 1: SAM Analysis Results.

List of genes used in the consensus clustering analysis. The score reflects the coefficient of the Cox Proportional Hazard regression analysis. The Q-value represents the FDR.

cancers previously exposed to neoadjuvant chemotherapy.

The data was normalized and randomly separated into a training (n = 28) and validation set (n = 29). After univariate ranking of the genes according to the strength of their relationship with overall survival in the training set, the genes most strongly associated with overall survival were prioritized (Table 1). Using those genes, consensus clustering separated the set into two groups (Figure 5A). The two groups did not have a significant difference in progression free survival (Figure 5B; log rank p = 0.69), but did show a significant difference in overall survival (Figure 5C; log rank p = 0.0061).

Using the same genes identified in the univariate gene ranking in the training set (Table 1), consensus clustering was then performed on the samples from

the testing set (Figure 5D). Similarly, the two identified groups had a non-significant difference in progression-free survival (Figure 5E; log rank $p = 0.083$), but a significant difference in overall survival (Figure 5F; log rank $p = 0.0084$).

By means of a semi-supervised, internally validated consensus clustering method, our cohort of patients treated with NACT was separated into a Good and a Bad outcome group (Table 2). Together, in a univariate analysis, the Good outcome group had better OS and PFS, and the rate of complete surgical cytoreduction was higher in the Good outcome group (Table 2).

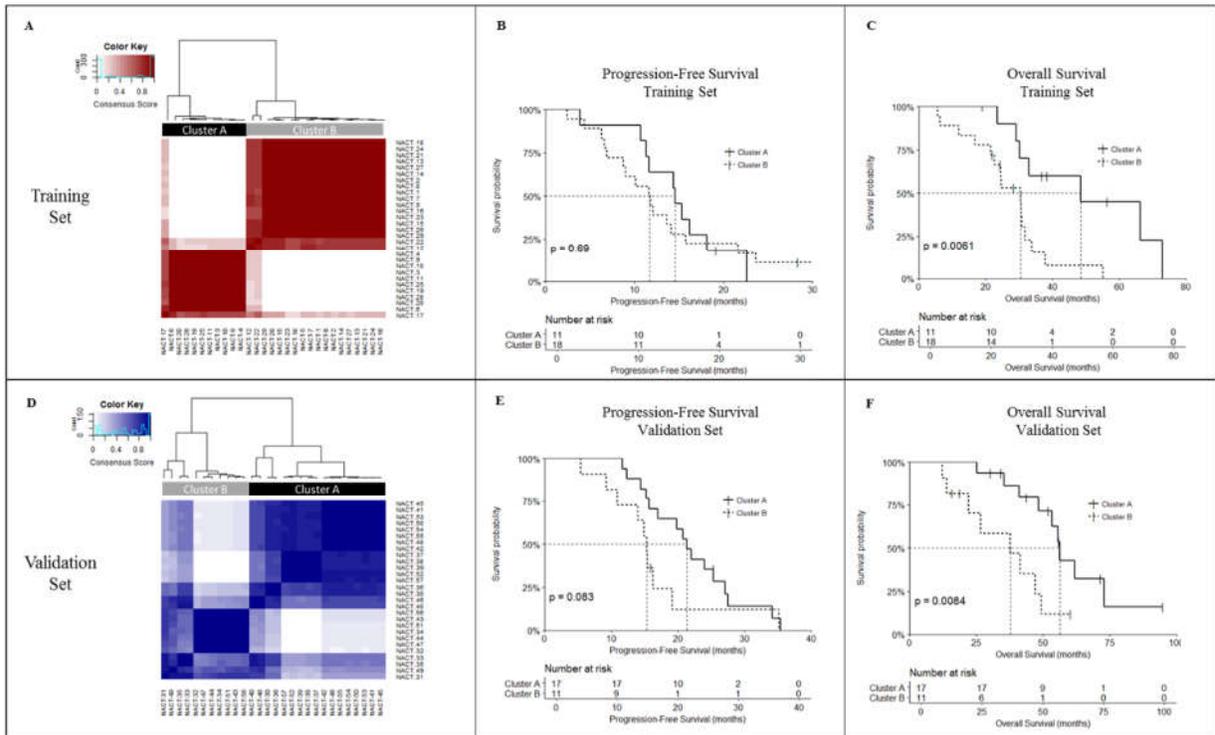


Figure 5: Semi-supervised, internally validated consensus clustering. A) Consensus clustering of samples in the training set revealed two molecularly distinct clusters. The survival of patients in the clusters uncovered in the training set were compared through a Kaplan-Meier analysis for B) progression-free survival and C) overall survival. D) The same analysis in a testing set divided the testing cohort in two. The patient clusters in the testing set did not have a significantly different E) progression-free survival, but did have significantly different F) overall survival.

	Good Outcome n = 28	Bad Outcome n = 29	p-value
Age			
< 50	8	4	0.4200
50 - 70	13	16	
>70	7	9	
Stage			
3C	26	25	0.6700
4	2	4	
No. NACT Cycles			
3	6	8	0.9300
4	17	16	
5+	5	5	
Progression Free Survival			
Median [IQR]	47 [37 - 61]	35 [23 - 41]	0.0160
Overall Survival			
Median [IQR]	117 [82 - 143]	63 [48 - 86]	0.0001
CA125 Reduction by NACT			
>95%	13	8	0.2000
75 - 95%	9	16	
50 - 75%	3	4	
<50%	3	1	
BRCA1/2 Germline Mutation			
Mutated	8	5	0.5900
Wild Type	16	20	
Not Tested	4	4	
Surgical Outcome			
Optimal	27	21	0.0250
Suboptimal	1	8	
NACT Treatment Type			
Carbo-Taxol	28	27	0.4900
Carbo Only	0	2	
Percent Cancer Tissue			
100%	18	22	0.6200
50 - 100%	7	5	
<50%	3	2	

Table 2: Patient characteristics by outcome group.

Characteristics of patients included in the comparative analysis of the NACT cohort, separated by predicted outcome group

* P-values were computed using the Fischer's exact test for categorical variables, the Wilcoxon rank-sum test for continuous variables, and the Log-Rank test for survival variables

† An optimal surgical outcome was defined as residual disease <1mm

‡ Progression Free Survival was defined as the time of diagnosis to evidence of recurrence by imaging or death

§ Overall survival was defined as the time from diagnosis to death or the latest date of follow-up

3. Multivariate survival analysis. Complete surgical reduction is a major clinical determinant of outcome in the treatment of HGSC. Other factors, such as germline *BRCA1/2* mutations may also affect survival. To control for their potential effects, a multivariate Cox Proportional Hazard survival analysis was performed, and the results are reported in Table 3. After controlling for external clinical variables, molecular cluster membership remained significantly associated with overall survival (HR = 6.4; 95% CI = [2.6, 15.5]; p = 0.0004). In addition to cluster membership, age was also found to be significantly associated with overall survival. However, the hazard ratio is very small, suggesting that although age might have a statistically significant effect on survival, that same effect may be small (HR = 1.03, 95% CI = [1.00, 1.1], p = 0.047). Molecular cluster membership was significantly associated with progression-free survival in a univariate analysis (Median: 47 months vs 35 months; p = 0.016; Table 2), but the association did not remain significant in the multivariate analysis. Nonetheless, the results suggest that molecular characteristics may explain the differences in overall survival better than clinical characteristics in our patient cohort.

<u>Variable</u>	<u>Progression-Free Survival</u>		<u>Overall Survival</u>	
	<u>HR [95% CI]</u>	<u>p-value</u>	<u>HR [95% CI]</u>	<u>p-value</u>
Group Membership (Bad vs Good)	1.7 [0.9,3.2]	0.08	4.8 [2,11.2]	0.0003
BRCA Mutated	1.4 [0.7,3]	0.32	0.89 [0.37,2.2]	0.8000
CA125 Reduction	1 [0.99,1]	0.97	0.99 [0.98,1.01]	0.4000
Stage (4 vs 3C)	1.8 [0.7,4.7]	0.20	1.5 [0.5,4.8]	0.5000
Surgical Outcomes (Suboptimal vs Optimal)	2.2 [0.9,5.2]	0.07	1.4 [0.5,3.7]	0.5000
Age	1.01 [1,1.04]	0.09	1.03 [1,1.1]	0.0470
Cycles of NACT	0.9 [0.7,1.3]	0.80	1.1 [0.7,1.7]	0.6000

Table 3: Multivariate survival analysis.

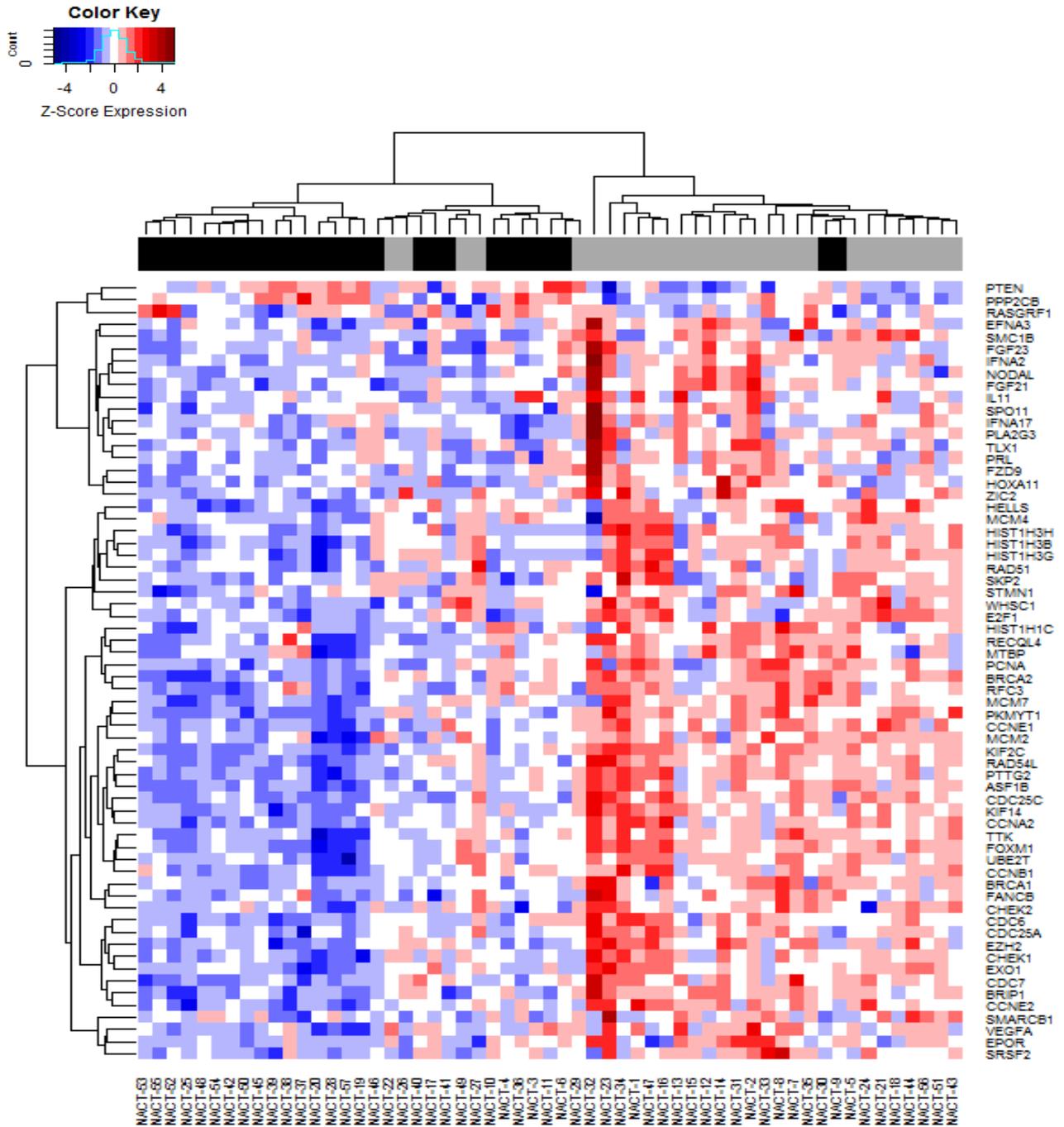
Evaluating the association of molecular cluster membership with overall survival and progression free survival using a Cox Proportional Hazard analysis in order to control for the effect of various clinical variables on survival.

4. Differential gene expression analysis. To gain a wider perspective of the molecular differences between each group, a differential expressions (DE) analysis was performed to compare the previously identified Good and Bad outcome groups. Seventy-two genes were found to be differentially expressed between our two survival groups. The top 50 most statistically significantly differentially expressed genes are shown in figure 7. Most of them were overexpressed in the Bad outcome group, whereas only 3 genes were found to be overexpressed in the Good outcome group (*PTEN*, *PPP2CB*, *RASGRF1*). The complete list of differentially expressed genes can be found in section 4 of the appendix.

5. Gene set analysis. To better understand the main biological processes that differ between our two patient groups and interpret the data in a biologically meaningful context, a gene set analysis (GSA) was performed. Pathway scores were computed for each patient, and the scores compared across the Good and Bad outcome group (Figure 8A). Expression of genes in the cell cycle (CC), homologous recombination (HRD) and mismatch repair pathway (MMR) were the most statistically significant differentially expressed pathways ($p < 0.0005$). The chromatin remodeling (CR) pathway was also found to be differentially expressed between the Good and Bad outcome groups ($p < 0.05$). In all cases, the pathways were overexpressed in the Bad outcome group. No significant differences in expression were observed in the other pathways tested. For method validation purposes, GSA was also carried out using the PANTHER Gene Ontology Biological Process sorting tool (Figure 8B). After corrections for multiple testing, the differentially expressed genes between our two groups mostly participate in cell proliferation ($p < 0.0005$), DNA recombination ($p < 0.005$), Meiosis ($p < 0.05$), DNA repair ($p < 0.0005$), mitosis ($p < 0.0005$), tyrosine kinase signaling ($p < 0.05$) and response to stress (0.005). Those pathways are closely

related to our manually defined gene sets, confirming the results we obtained using our pathway scoring algorithm.

Gene expression from the HRD and CC pathways play a large role in differentiating the Good and Bad outcome groups in our cohort of patient that received NACT. We posited that some upstream, intersecting genes common to both pathways may account for most of the difference between the two groups. Manual inspection of the differential expression results in both pathways using volcano plots revealed that *FOXMI* belongs to both pathways and is strongly differentially expressed between the two groups (Figure 8C). With *FOXMI* as a central point, pathway reconstruction using information curated from the literature, GeneMania¹²¹ and PathwayNet¹²² revealed that *FOXMI* is a transcription factor responsible for the expression of most genes differentially expressed between our two groups (Figure 9). Together, our data suggests that *FOXMI* may play an important role in regulating subsequent response to chemotherapy following NACT and surgical debulking.



Column Annotation Key: = Samples in Good outcome group = Samples in Bad outcome group

Figure 7: Heatmap of the 50 most differentially expressed genes. The 50 most differentially expressed genes are depicted above. Columns represent patient samples, and the names were anonymized in the figure. Gene expression levels were normalized across all samples, and the representative Z-scores are represented by rows where dark blue represents a low expression, and red a high expression of a given gene. The column annotation represent the Good outcome group (black) and the Bad outcome group (grey).

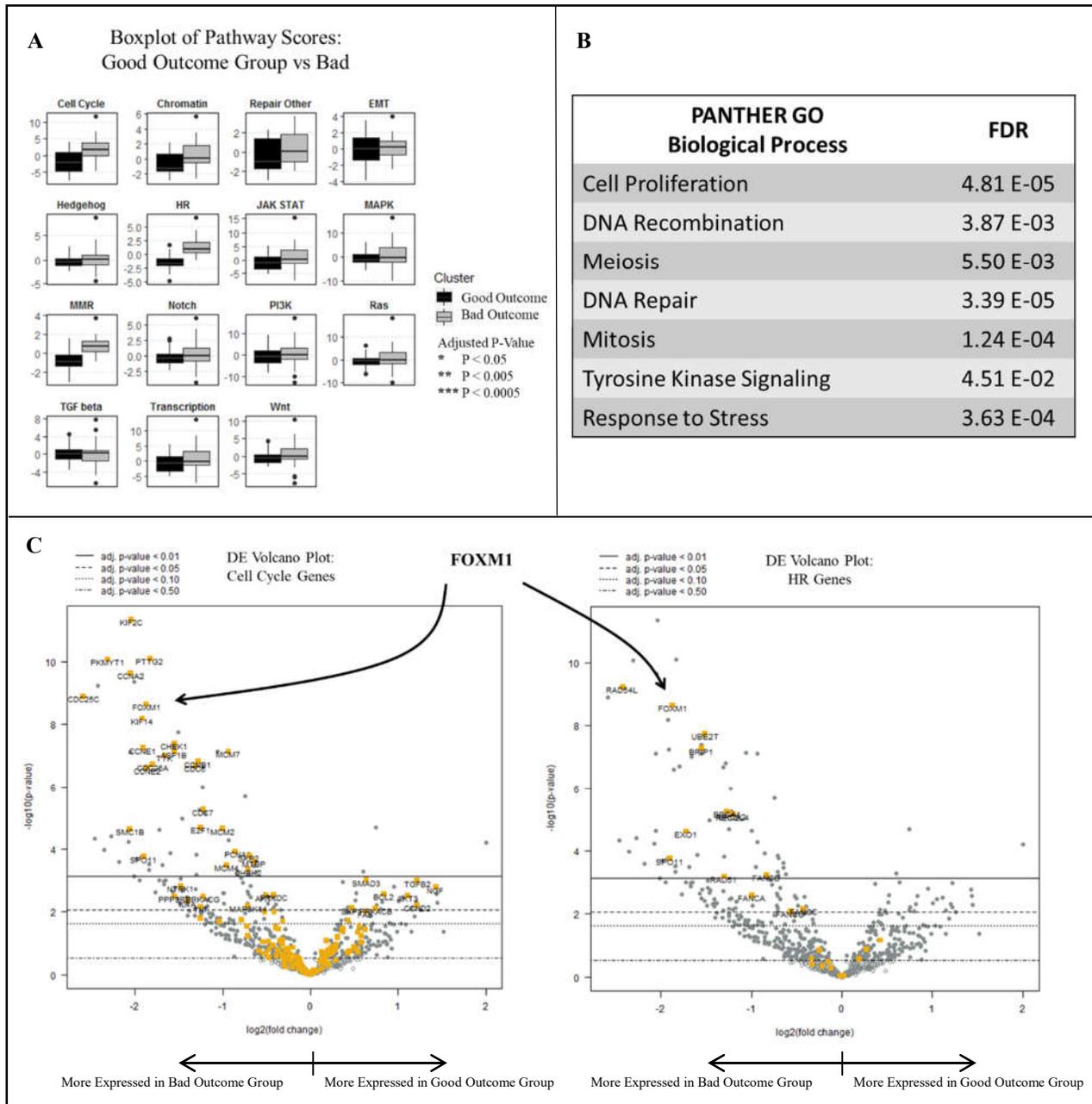


Figure 8: Gene set analysis of the differentially expressed genes. A) Results of the GSA using pathway scoring based on the singular value decomposition of the expression of each gene belonging to a given pathway. A higher score represents a higher expression of the genes in the represented pathway. Some genes may belong to more than just one pathway. B) Results of the GSA using the Protein Analysis Through Evolutionary Relationships (PANTHER). P-values are calculated using Fisher's exact test and adjusted via the Benjamin-Hochberg FDR method. C) Volcano plots of the differentially expressed genes in the Cell Cycle pathway (left) and the HRD pathway (right). Genes from the respective pathways are highlighted in yellow, and genes that do not belong to the pathway are in grey. Horizontal lines reference particular p-value thresholds.

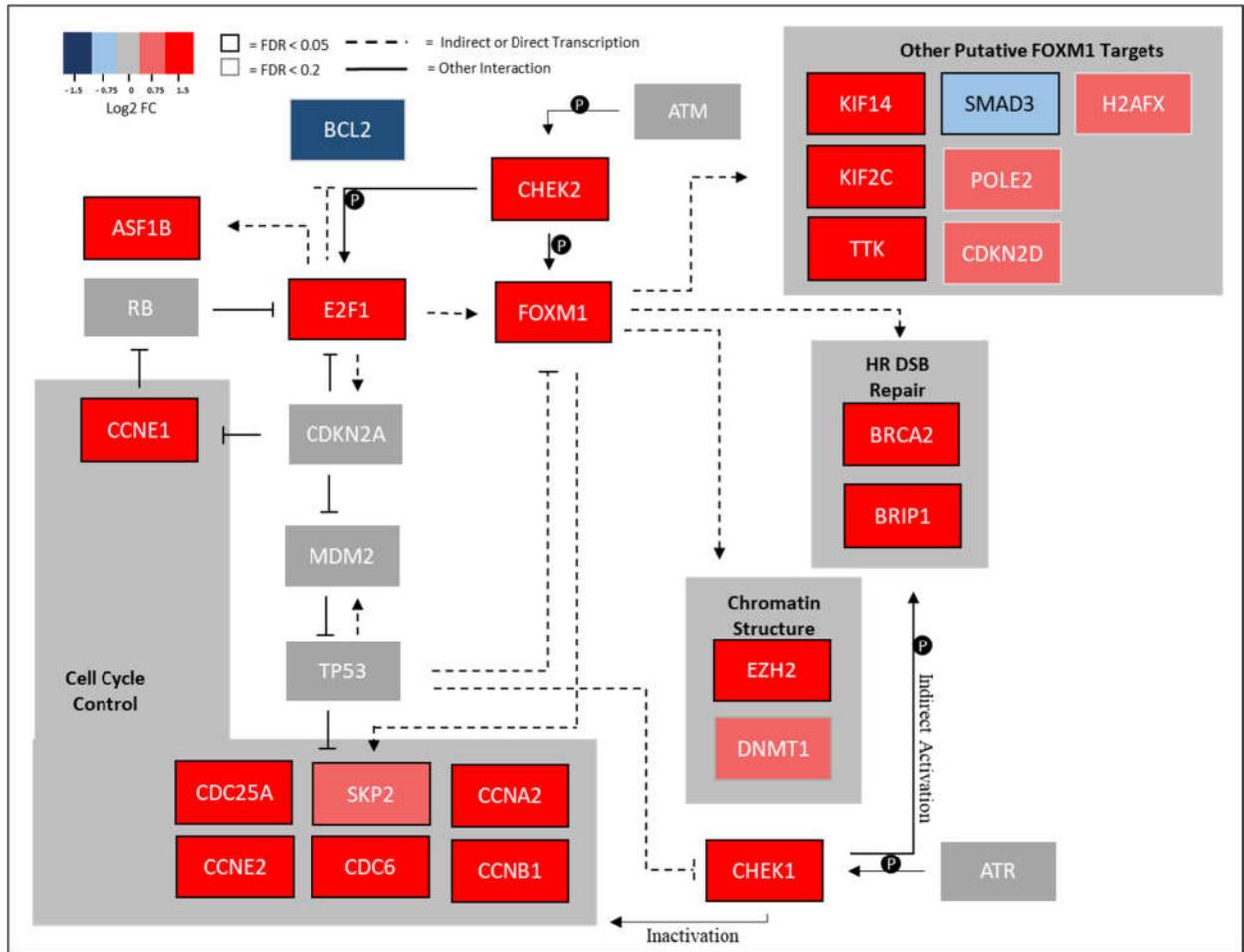
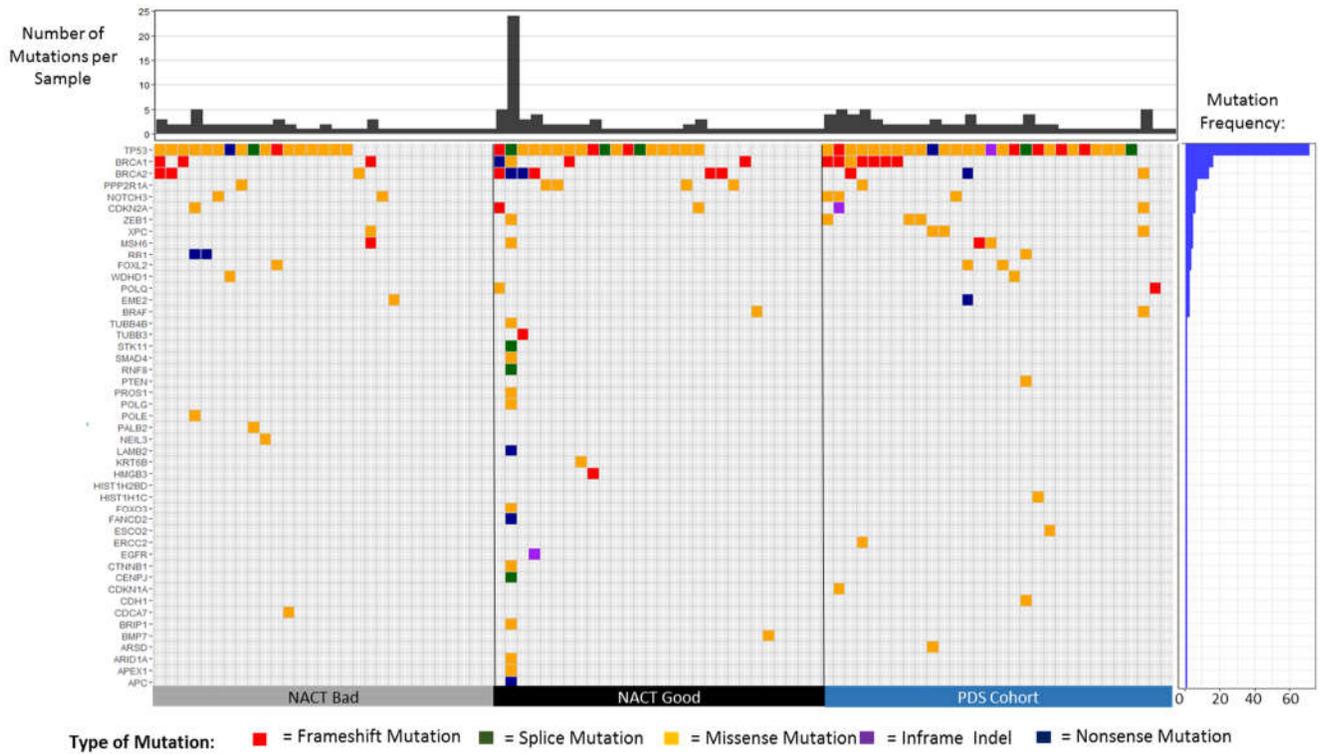


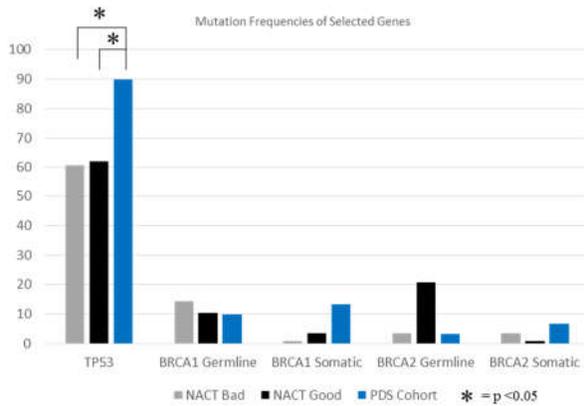
Figure 9: Representation of the *FOXM1*-mediated transcriptional network. The figure depicts a reconstruction of the relationship between some of the genes differentially expressed between our two groups. Each gene depicted was present on the expression panel. Colors represent the log₂ fold-change between the Good and Bad outcome groups, ranging from dark blue (overexpressed in the Good outcome group) to red (overexpressed in the Bad outcome group). A black box outline represents a FDR lower than 5%, and grey box outlines represent a FDR lower than 20%. Transcriptional relationships between genes or groups of genes are shown by dashes, while plain lines represent any other type of interactions (i.e. physical, phosphorylation, etc...). Transcriptional relationships may be indirect (i.e. through the transcription of a different transcription factor). *FOXM1* is a sequence-specific transcription factor, and studies have investigated its different targets. Grey boxes group genes according to the molecular pathway to which they belong.

6. Mutation analysis. In order to determine whether our NACT cohorts were enriched for specific mutations, next-generation sequencing was performed on approximately four-hundred mutational hotspots of one hundred and sixty-eight genes involved mainly in homologous recombination DNA repair. As a mutation frequency reference, a control group of thirty samples from chemotherapy-naïve, primarily debulked patients was also included. As expected, the most recurrently mutated genes were *TP53* (72%), *BRCA1* (16%), and *BRCA2* (12%) (Figure 10A). Most other genes were mutated at very low frequencies (<3%). The mutation frequency of *TP53* was lower than expected based on previous findings from the literature. However, when the mutation frequencies are inspected by group, the *TP53* mutation frequency is as expected in the PDS group (90%), but significantly lower in the NACT Good outcome group (62%) and the NACT Bad outcome group (61%) (Figure 10B). The frequency of *BRCA2* germline mutations seemingly spikes in the NACT Good outcome group (20%), but the difference was not statistically significant. When looking at the percentage of patients mutated in at least one homologous recombination gene in each group, 33% of patients in the PDS cohort, 24% of patients in the NACT Bad outcome group, and 29% of patients in the NACT Good outcome group were mutated in at least one HRD genes (Figure 10C).

A



B



C

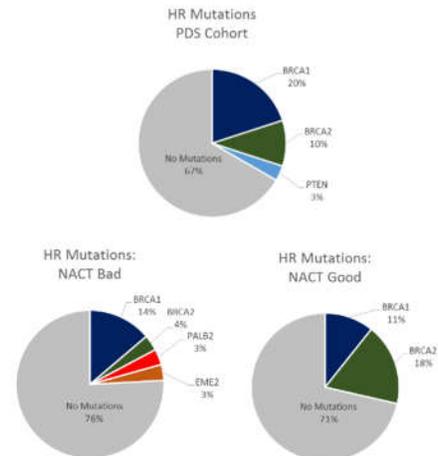


Figure 10: Mutation analysis results. A) Plot of the mutations in each patient (column) that are part of the NACT good outcome cohort, NACT bad outcome cohort or the PDS cohort, by gene (row). Also plotted is the total mutation burden for all samples, together, and the number of mutations found in each sample. Red squares indicate a frameshift mutation, green a splice site mutation, yellow a missense mutation, purple an inframe indel and darkblue a nonsense mutation. If two mutations were found in the same gene, the most deleterious was depicted. The graph was created using the GenVisR package. B) Plot of the mutation frequency of select genes, by patient cohort (NACT Good vs NACT Bad vs PDS). C) Pie charts representing the percentage of patients with at least one mutation in the homologous recombination pathway, separated by group.

Discussion

1. Summary

1.1 Identification of novel molecularly defined outcome groups. Previous studies on the association of gene expression levels with survival identified clinically relevant molecular subgroups in chemotherapy-naïve, primary-derived OC tumors¹²³. However, none have done the same for patients with residual diseases after chemotherapy. Patients who respond incompletely to NACT represent a minority of patients who may harbor chemotherapy-resistant clonal populations. Therefore, studying the underlying genomics of these tumors in the context of survival may provide insight into the mechanism of resistance to subsequent therapy and early recurrence. In breast cancer, for example, NACT has been shown to change the molecular classification of the disease with a direct impact on the treatment strategies¹²⁴. Moreover, many post-neoadjuvant biomarkers have been associated with survival^{125,126,127}. From tumor infiltrating lymphocytes¹²⁸ or mutations post-NACT¹²⁹, much data supports the idea of treatment adjustment post-NACT based on the “new” molecular features of the disease¹³⁰. To our knowledge, we are the first to report the molecular features of post-NACT HGSC in the context of survival.

First, candidate patients were identified in our prospectively maintained clinical database. Ovarian cancer is a clinically heterogeneous disease, and outcomes differ based on the clinical and histopathological characteristics of the primary disease. For the purpose of this study, only patients with stage III or IV high-grade serous ovarian cancer were considered. Since cancer content can have a high impact on expression results, DNA and RNA were only extracted from fresh-frozen tumor tissue with a validated high cancer content upon pathology review by an experienced gynecological pathologist. We also statistically controlled for tumor content using our mutation

data. RNA expression was assessed for 800 genes using a NanoString panel, and mutations were assessed in approximately 400 mutational hotspots of 168 genes using next-generation sequencing.

Using an internally validated, semi-supervised method we identified two different outcome groups in the neoadjuvant-treated population. Unsupervised methods have previously been used with some success in identifying patient groups with distinct survival outcomes. However, the studies captured a much larger molecular snapshot than we did through whole genome and whole exome analyses. Moreover, there is no guarantee that identified subgroups in unsupervised analyses will have optimal clinical relevance, which was one of our group's priority. On the other hand, supervised group identification that only uses clinical data (i.e. survival cutoffs) to identify groups may hide subtle molecular differences. Since patient survival is most likely due to a mix of clinical and molecular characteristics, a semi-supervised method similar to that used in previous studies was the favored approach^{62,123,131}.

Independent survival model evaluation is important to assess the strength of discovery methods. By randomly splitting our data in a training and testing set, we were able to not only discover new survival groups, but also make sure that our method was reproducible in an independent dataset. Nowadays, thousands of molecular profiles are available online for ovarian cancer to cross-validate gene expression findings. Unfortunately, none are available for patients that underwent NACT. In other words, all data available comes from chemotherapy-naïve tumors, which represent a different patient population to ours. Moreover, even though those molecular profiles are available, there is a growing concern about the lack of reliable clinical data, sample processing standards and batch effects, which may explain why no single, validated and reproducible gene expression signature has emerged from this wealth of data.

1.2 Differential expression and gene set analysis. A differential expression analysis between the Good and Bad outcome groups allowed us to uncover genes whose level generally changed depending on the outcome. Our results suggest that the cell cycle and homologous recombination pathway gene expression may be an important factor driving worse survival.

Cell cycle genes overexpression in cancer is common. Aberrant cell cycle is necessary for escaping the cell's checkpoints, evading apoptosis, and thus driving tumorigenesis. *CCNE1* amplification is a hallmark of cell cycle deregulation in HGSC⁶². It is synthetic lethal to HRD deficiency⁹⁸, and is thus a characteristic of worse-outcome cases, although it is not directly related to survival by itself¹³². As expected, it is overexpressed in our worse outcome group. *FOXM1*, *CHEK1* and *CHEK2* are also important genes overexpressed in the worse outcome group, and are interesting due to their roles affecting both the cell cycle and HRD pathways. *CHEK1* and *CHEK2* relay the stress responses due to double strand breaks from *ATR* and *ATM*, respectively, arresting the cell cycle and allowing HRD-mediated repairs in the G2 phase. *FOXM1*, similarly to *CCNE1*, drives the cell cycle by transcribing genes necessary for G2/M phase transition. Together, *FOXM1* and *CHEK1/2* may seem to possess antagonistic functions, but cell cycle regulation is complex and the interplay between the different proteins remains poorly understood. For example, *FOXM1* was shown to induce *CHEK1* transcription¹³³, and *CHEK2* may help stabilize the *FOXM1* protein¹³⁴. Interestingly, targeting *FOXM1*, *CHEK1* or *CHEK2* in HRD-proficient HGSC is an increasingly studied possibility. In HGSC cell lines, *FOXM1* downregulation was shown to confer increased susceptibility to cytotoxic agents¹³⁵. In a recent Phase II study, *CHEK1* and *CHEK2* inhibition in HRD-proficient, heavily pre-treated HGSC cases showed promising clinical activity¹³⁶. Based on our results, such agents may represent early alternatives to traditional

cytotoxic agents in patients with high expression of those genes, such as those identified in our worse outcome group.

Expression of the HRD pathway genes has been related to survival in HGSC before, principally with *BRCA1*, where lower expression due to epigenetic silencing conferred a survival advantage⁶². Interestingly, in our cohort, *BRIP1*, was the most differentially expressed HRD gene between the Good and the Bad outcome group, suggesting that survival-related changes in gene expression may not be limited to *BRCA1* and *BRCA2*. Germline mutations in *BRIP1* have been shown to predispose women to ovarian cancer¹³⁷, but its expression has not been described in the context of survival before.

1.3 Mutational Analysis. Interestingly, *BRCA1* and *BRCA2* germline mutations were not associated with better survival in our neoadjuvant patient cohort, which was unexpected since each of those mutations have been thoroughly validated as risk-inducing, pathogenic mutations. However, this may be explained by the nature of our cohort. Patients who respond very well to neoadjuvant chemotherapy are not represented in our cohort, because there is no tumor left at the time of surgery. Fallopian tube tissue removed as part of the standard of care total hysterectomy and bilateral salpingo-oophorectomy at the time of debulking surgery harbors small, microscopic sites of residual disease in most patients who underwent NACT, but those are usually found by the pathologist after setting in FFPE. The DNA and mRNA of FFPE-embedded tissue is often degraded and not useable for high-throughput gene expression or sequencing assays. Therefore, our tumor bank only contains samples that are visible to the naked eye of the surgeon at the time of surgery and removed for biobanking purposes. Although a link between a good response to neoadjuvant chemotherapy and survival has not been established in HGSC, it has been shown that *BRCA1/2*-mutated cells are selectively killed during neoadjuvant chemotherapy, especially if the

mutation is accompanied with somatic LOH of the WT allele¹⁰⁵. It has also been shown that *BRCA1/2* mutation-driven loss of HRD functionality may be overcome by alternative molecular mechanism and clinical mechanisms, such as desmoplastic reactions^{86,138}. Therefore, our cohort may be enriched for patients whose germline mutations in *BRCA1* or *BRCA2* may not participate in HRD deficiency in the remaining cell population. Our samples were not found to be enriched for any other particular kind of mutations.

The lower rate of *TP53* mutations in the NACT cohort is intriguing given the very high prevalence of such mutations in HGSC. Our first concern was the possibility that the analyzed tissue only contained stromal cells and no cancer cells, which would explain the absence of mutations. However, we were extremely rigorous in our filtering criteria, and the control group of chemotherapy-naïve tumors harbored the appropriate number of mutations. All tissue samples, NACT-treated or chemotherapy-naïve, were analyzed by the same gynecologic pathologist who was blinded to the patients' treatments. Moreover, the mutations were analyzed together and subjected to the same filtering criteria. As further controls, three cell lines with known mutation profiles were included to ensure that all of their validated mutations were found at the end of the filtering steps (data not shown). For these reasons, we are confident that the difference in mutation frequencies is due to the nature of the patient cohorts. If the difference was due to technical errors, we would have seen unexpected mutational profiles in the PDS and cell lines as well.

The loss of *TP53* mutations has previously been observed in a study looking at matched pre- and post-NACT breast cancer tissue¹²⁵. Unlike our results, loss of *TP53* mutations in post-NACT tissue was associated with improved survival. However, given the wide range of p53 molecular targets and its lack of clear association with survival in other studies, this may be an indirect correlation explained by other molecular factors that the authors did not investigate. In one of the

only studies of pre- and post-NACT HGSC, loss of *TP53* mutations was observed in 5 out of 11 initially *TP53*-mutated samples¹⁰⁵. Interestingly, gain of *TP53* mutations was also observed in 3 out of 11 *TP53*-wild type samples. This may be explained by the clonal heterogeneity of HGSC, and the Sanger sequencing methods used by the authors, which may be too insensitive to capture small mutant allele frequencies. Nonetheless, *TP53* mutation status, which is a defining feature of HGSC to many, may play a role in the chemotherapy induced selection during NACT and should be further investigated.

2. Future Directions

2.1 Tissue-Microarrays. A tissue microarray (TMA) contains tissue samples from multiple sources, allowing for simultaneous histologic analysis of markers on a single slide. Differences in mRNA expression levels are only relevant insofar as the related protein levels are similarly affected. In some instances, mRNA levels correlate poorly with associated protein levels¹³⁹. Alternative splicing, post-transcriptional mRNA regulation and post-translational modifications may all play a role in how much of the mRNA is actually translated. Analysis of the protein levels of genes of interests, such as *FOXMI*, *CHEK1*, *CHEK2* and *CCNE1*, would allow us to validate our findings and determine whether they warrant further investigation. In other words, if no protein levels are detected, there is little incentive to try to inhibit a particular target with small molecules or antibodies.

Beyond validation, it would be interesting to see whether the levels of tumor infiltrating lymphocytes (TIL) and immune checkpoint proteins differ between the Good and Bad outcome groups in the NACT cohort, and between the PDS and NACT cohorts as wholes. Ovarian cancer lags behind other types of cancer in the field of immunotherapy. However, it has been shown that NACT increases TIL infiltration and PD1/PDL1 expression in epithelial ovarian, and that, in turn,

increased TIL count and PD1/PDL1 expression correlates with improved response to checkpoint inhibitors^{140,141}. In addition, CD8+ TILs are associated with long-term survival in ovarian cancer and may help explain why a subgroup of patients undergoing NACT respond very well to treatment¹⁴². Results from this investigation may help provide a rationale for the use of immune checkpoint inhibitors in a subset of patients with HGSC.

2.2 Matched pre-NACT, post-NACT and time of recurrence samples. Like many solid tumors, HGSC is molecularly heterogeneous¹⁴³. One weakness of high-throughput genomic analyses is the inability to capture the entire molecular landscape of a particular cancer. In fact, it has been shown that even on the same primary tumor, different areas show different molecular profiles, and may potentially respond differently to chemotherapy¹⁴⁴. Nevertheless, personalized treatment based on somatic biomarkers, even if they may change from one area of the disease to the next, have been successful in clinical trials^{145,146} although none have been successful in ovarian cancer, yet. A better understanding of the resistant sub-clonal populations may further increase outcomes for cancer patients. Logically, neoadjuvant administration should select for chemotherapy-resistant sub-clones. Therefore, sampling the post-NACT tissue may be more informative because it would allow for better molecular identification of the agents of resistance for which alternatives treatments exist. Since systematic interval cytoreductive surgery or tumor biopsy is uncommon, very little is known about the evolution of HGSC from the time of diagnosis, to the time of surgery post-NACT, and to the time of recurrence. In the context of our findings, it would be interesting to see whether recurrent disease is more molecularly similar to the post-NACT sample than the pre-NACT sample, which would indicate that NACT selected the chemo-resistant disease that is most likely to recur later. Since HGSC is heavily afflicted by chromosomal

rearrangements^{62,86,147}, future studies should look deeper into chromosomal abnormalities instead of mutations and gene expression.

2.3 Pre-clinical modeling. Much of the failure to recapitulate pre-clinical findings in the clinical settings may be attributable to the aforementioned molecular heterogeneity. Immortalized HGSC cell lines are the workhorses of pre-clinical studies for personalized medicine, but their use is limited to the background genomics characterizing a particular cell line. Recently, there has been a large effort in improving patient-derived xenograft (PDX) techniques in mice and, most importantly, making sure that the models recapitulate the histological and genomic characteristics of the tumour. A number of PDX models have been described for high-grade serous ovarian cancer that succeeded in recapitulating the intra-tumour molecular heterogeneity and the clinical response to cytotoxic agents of the donor^{148,149}. Modeling the pre-NACT, post-NACT and recurrence samples in mice may be a favorable alternative to doing so in patients, since it would allow to control for outside factors, repeat experiments and avoid unnecessary potential harm to patients. Of course, this technique would only account for the contribution of intra-tumour heterogeneity to chemotherapy response, and would not model the effect of the tumour microenvironment. Nevertheless, it would provide some much-needed answers on the molecular features of post-NACT samples

Appendix

Section 1. Information on the NanoString pathways

Category/Pathway	Number of Genes	Panel Name	Representative Hallmarks
Hedgehog pathway	28	PanCancer Pathways	Resisting cell death
Cancer driver genes	124	PanCancer Pathways	—
Cell cycle and apoptosis	137	PanCancer Pathways	Resisting cell death
Chromatin modification	22	PanCancer Pathways	Genome instability and mutation
DNA repair	50	PanCancer Pathways	Genome instability and mutation
JAK-STAT pathway	86	PanCancer Pathways	Sustaining proliferative signaling
MAPK pathway	157	PanCancer Pathways	Sustaining proliferative signaling
Notch pathway	24	PanCancer Pathways	Enabling replicative immortality
PI3K pathway	201	PanCancer Pathways	Deregulating cellular energetics
RAS pathway	142	PanCancer Pathways	Sustaining proliferative signaling
TGF-pathway	51	PanCancer Pathways	Evading growth suppressors
Transcriptional misregulation	101	PanCancer Pathways	Genome instability and mutation
Wnt pathway	78	PanCancer Pathways	Enabling replicative immortality

Supplementary Table 1: List of molecular pathways included on the NanoString panel and the number of genes associated with each pathway. Each gene was assigned one or more pathway membership by the NanoString bioinformatics team. For our purposes, I subdivided the DNA repair pathway genes into their subcategories, namely: homologous recombination, mismatch repair, and other (i.e. all other DNA repair processes). Each of the 30 custom genes added by our group was assigned to one of the pathways as well. The complete list of genes can be found online at www.nanostring.com/products/gene-expression-panels

Genes added to the panel: ZEB1, XRCC5, ASF1B, EMSY, CLK2, EXO1, FAM134B, FOXM1, HIST1H1C, INSIG1, KIF14, KIF2C, MME, MSH4, MTBP, NUP205, CDK12, PARP1, PINX1, PMS1, RAD23A, RAD51C, RAD54B, RAD54L, RECQL, RECQL4, TERC, TERT, TNKS, SPO11

Section 2. Detailed normalization procedure

Each NanoString panel contains a number of pre-assigned genes, positive control probes and negative control probes for normalization purposes. First, negative control probes are used to eliminate background signals. Second, positive control probes are used to adjust for technical variability. Finally, a number of probes against genes that are not supposed to vary across conditions are included to control for mRNA input (housekeeping genes). To optimize our normalization procedure, eight technical replicates were included. Six different normalization procedures are suggested by the NanoString bioinformatics team (Supplementary Table 2), and each was tested on the eight pairs of technical replicates². The Pearson correlation between the gene expression of each pair was computed and can be found in Supplementary Table 3. The low correlation between the replicates for sample 5 were attributed to mislabeling and removed from further evaluation. The mean concordance coefficient for the seven remaining samples was computed across all six normalization procedures. Method 3 was determined to be the optimal procedure.

First, no background subtraction was performed. Our data had very little noise and most negative control probes had a final count of zero. Second, no positive control normalization was performed. Positive controls are a set number of artificial probes spiked at defined levels within each sample to adjust for variations across lanes, samples, cartridges, and date. Since the positive controls and the endogenous controls serve similar purpose, this step is often optional. Finally, the geometric mean of the housekeeping genes was used to control for sample input variability.

² The method optimization work was done by Dr. Kathleen Klein. Subsequent normalization steps were performed by myself.

Samples with a normalization outside the 0.3-3 range were removed from further analyses. An example workflow can be found in Supplementary Table 4.

	codeCount	background	sampleContent
method_1	sum	none	housekeeping.geo.mean
method_2	geo.mean	none	none
method_3	none	none	housekeeping.geo.mean
method_4	sum	mean.2sd	housekeeping.geo.mean
method_5	none	mean.2sd	housekeeping.geo.mean
method_6	geo.mean	mean	housekeeping.geo.mean

Supplementary table 2. Parameters for normalization methods for NanoString

	method_1	method_2	method_3	method_4	method_5	method_6
rep_1 PT05	0.15694	0.63047	0.17245	0.14264	0.18577	0.15021
rep_2 PT102	0.99514	0.95371	0.99514	0.99232	0.99455	0.99393
rep_3 PT26N	0.96392	0.89922	0.96307	0.96622	0.96572	0.96633
rep_4 PT57	0.8054	0.80098	0.8054	0.75848	0.75961	0.79167
rep_5 PT64	0.98222	0.99827	0.98222	0.9945	0.99657	0.99059
rep_6 PT66	0.97315	0.99727	0.96963	0.91407	0.89103	0.93186
rep_7 PT74	0.99192	0.98497	0.99169	0.98971	0.99095	0.99173
rep_8 PT75	0.99014	0.97616	0.99014	0.99335	0.99214	0.99405
MEAN	0.95741	0.94437	0.95676	0.94409	0.94151	0.95145

Supplementary Table 3. Concordance Correlation Coefficients of six recommendation methods for NanoString data

	Sample 1	Sample 2	Sample 3
Gene_A	24573	21007	21856
Gene_B	6948	6414	6589
Gene_C	2123	1826	1932
Gene_D	432	363	425
Gene_E	52	68	53
Gene_F	49	38	52
GeoMean	858	783	829
Mean	823		
Norm. Factor	0.96	1.05	0.99

Supplementary Table 4: Example of a normalization workflow using the NanoString housekeeping genes. First, the geometric mean of the housekeeping genes is computed for all samples. Second, the mean of the geometric means is computed, and, finally, this mean is divided by each individual geometric mean value. The end result is a normalization factor that is different for each sample, and each gene expression value is multiplied by that value for a given sample.

Section 3: TP53 primers for Sanger sequencing

TP53 exon	Forward	Reverse	Annealing Temperature	Size (BP)
2 and 3	TGGAAGAGAGAATGTGAAGC	AGAGCAGTCAGAGGACCAG	60°C	496
4	CTGGTAAGGACAAGGGTTG	CAGGAGTCAGAGATCACACAT	55°C	551
5 and 6	GGTGTAGACGCCAACTCTC	ACCCATTTACTTTGCACATC	55°C	662
10	TAGGTACTTGAAGTGCAGTT TCT	CTGGGACCCAATGAGATG	55°C	368
11	TTGATTTGAATTCCCGTTG	AACCCTTAACTGCAAGAACA	55°C	486

Supplementary Table 5: List of TP53 primers used in the study.

Section 4. List of Differentially expressed genes

	Log2 fold change	std error (log2)	Lower confidence limit (log2)	Upper confidence limit (log2)	Linear fold change	Lower confidence limit (linear)	Upper confidence limit (linear)	P-value	BH.p.value
KIF2C	-2.04	0.231	-2.49	-1.59	0.243	0.178	0.333	4.53E-12	3.42E-09
PTTG2	-1.83	0.227	-2.27	-1.38	0.281	0.207	0.383	8.17E-11	2.12E-08
PKMYT1	-2.31	0.287	-2.87	-1.74	0.202	0.137	0.298	8.43E-11	2.12E-08
CCNA2	-2.05	0.263	-2.56	-1.53	0.242	0.169	0.346	2.31E-10	4.36E-08
HIST1H3B	-2.01	0.264	-2.53	-1.49	0.249	0.174	0.356	4.48E-10	6.77E-08
RAD54L	-2.42	0.322	-3.05	-1.79	0.187	0.121	0.289	5.99E-10	7.54E-08
CDC25C	-2.59	0.354	-3.28	-1.89	0.166	0.103	0.269	1.28E-09	1.38E-07
FOXM1	-1.87	0.261	-2.38	-1.35	0.274	0.193	0.391	2.26E-09	2.13E-07
KIF14	-1.92	0.279	-2.47	-1.37	0.264	0.181	0.386	6.52E-09	5.47E-07
UBE2T	-1.51	0.228	-1.95	-1.06	0.352	0.258	0.48	1.82E-08	1.37E-06
CHEK1	-1.55	0.243	-2.03	-1.08	0.341	0.245	0.474	4.06E-08	2.79E-06
BRIP1	-1.55	0.245	-2.02	-1.07	0.343	0.246	0.478	5.28E-08	3.24E-06
CCNE1	-1.91	0.302	-2.5	-1.31	0.267	0.177	0.403	5.59E-08	3.24E-06
EZH2	-1.06	0.17	-1.39	-0.726	0.48	0.381	0.604	7.25E-08	3.47E-06
MCM7	-0.934	0.15	-1.23	-0.64	0.523	0.427	0.642	7.47E-08	3.47E-06
ASF1B	-1.55	0.249	-2.04	-1.06	0.342	0.244	0.479	7.60E-08	3.47E-06
HIST1H3G	-2.06	0.331	-2.71	-1.41	0.24	0.153	0.377	7.82E-08	3.47E-06
TTK	-1.66	0.27	-2.19	-1.13	0.316	0.219	0.456	9.92E-08	4.16E-06
CCNB1	-1.28	0.211	-1.69	-0.861	0.413	0.31	0.551	1.52E-07	6.05E-06
CDC25A	-1.8	0.302	-2.4	-1.21	0.286	0.19	0.432	1.92E-07	7.26E-06
CDC6	-1.31	0.22	-1.74	-0.876	0.404	0.3	0.545	2.14E-07	7.70E-06
CCNE2	-1.86	0.315	-2.48	-1.24	0.275	0.179	0.422	2.43E-07	8.34E-06
HIST1H3H	-1.23	0.224	-1.67	-0.795	0.425	0.313	0.576	1.03E-06	3.37E-05
RFC3	-0.746	0.14	-1.02	-0.471	0.596	0.493	0.722	2.08E-06	6.53E-05

	Log2 fold change	std error (log2)	Lower confidence limit (log2)	Upper confidence limit (log2)	Linear fold change	Lower confidence limit (linear)	Upper confidence limit (linear)	P-value	BH.p.value
CDC7	-1.22	0.243	-1.7	-0.748	0.428	0.308	0.595	5.55E-06	0.000162
BRCA1	-1.27	0.251	-1.76	-0.773	0.416	0.296	0.585	5.58E-06	0.000162
BRCA2	-1.21	0.243	-1.68	-0.734	0.432	0.311	0.601	6.81E-06	0.000189
RECQL4	-1.2	0.242	-1.68	-0.73	0.434	0.312	0.603	7.01E-06	0.000189
VEGFA	-1.46	0.3	-2.05	-0.87	0.364	0.242	0.547	1.05E-05	0.000274
E2F1	-1.25	0.269	-1.78	-0.726	0.42	0.291	0.604	2.08E-05	0.000508
PPP2CB	0.745	0.16	0.432	1.06	1.68	1.35	2.08	2.09E-05	0.000508
MCM2	-1	0.216	-1.43	-0.579	0.499	0.372	0.669	2.24E-05	0.000529
SMC1B	-2.06	0.444	-2.93	-1.19	0.24	0.131	0.439	2.33E-05	0.000533
EXO1	-1.72	0.372	-2.45	-0.989	0.304	0.183	0.504	2.42E-05	0.000537
PLA2G3	-2.27	0.507	-3.26	-1.27	0.208	0.104	0.414	4.01E-05	0.000865
HOXA11	-2.46	0.554	-3.54	-1.37	0.182	0.0859	0.387	4.63E-05	0.000971
HELLS	-0.931	0.212	-1.35	-0.515	0.525	0.393	0.7	5.27E-05	0.00107
FGF23	-2.07	0.476	-3	-1.14	0.238	0.125	0.454	5.93E-05	0.00118
RASGRF1	2	0.46	1.09	2.9	3.99	2.13	7.45	6.34E-05	0.00123
IL11	-2.35	0.561	-3.45	-1.25	0.196	0.0914	0.42	0.000103	0.00195
PCNA	-0.855	0.206	-1.26	-0.451	0.553	0.418	0.732	0.00012	0.00221
FGF21	-1.6	0.389	-2.36	-0.841	0.329	0.194	0.558	0.00013	0.00233
HIST1H1C	-0.705	0.172	-1.04	-0.367	0.613	0.485	0.775	0.000146	0.00255
EPOR	-1.3	0.319	-1.93	-0.676	0.406	0.263	0.626	0.00015	0.00257
SKP2	-0.698	0.172	-1.04	-0.36	0.617	0.488	0.779	0.000165	0.00276
PRL	-1.93	0.48	-2.87	-0.993	0.262	0.137	0.503	0.000176	0.00284
SPO11	-1.9	0.471	-2.82	-0.974	0.268	0.142	0.509	0.000177	0.00284
IFNA17	-1.98	0.498	-2.96	-1.01	0.253	0.129	0.497	0.000204	0.00321
MTBP	-0.645	0.164	-0.967	-0.323	0.64	0.512	0.799	0.000246	0.0038
IFNA2	-2.18	0.557	-3.27	-1.09	0.221	0.104	0.471	0.00026	0.00392
FZD9	-1.84	0.477	-2.78	-0.909	0.279	0.146	0.533	0.000299	0.00443
NODAL	-1.69	0.44	-2.56	-0.831	0.309	0.17	0.562	0.000318	0.00462
MCM4	-0.953	0.248	-1.44	-0.466	0.517	0.369	0.724	0.000329	0.00468
WHSC1	-0.643	0.168	-0.973	-0.312	0.641	0.51	0.805	0.000352	0.00492
TLX1	-1.84	0.493	-2.81	-0.875	0.279	0.143	0.545	0.000453	0.00613
CHEK2	-0.708	0.19	-1.08	-0.336	0.612	0.473	0.792	0.000455	0.00613
PTEN	0.599	0.162	0.281	0.917	1.51	1.22	1.89	0.000516	0.00683
STMN1	-0.822	0.223	-1.26	-0.384	0.566	0.418	0.766	0.000539	0.00698
SRSF2	-0.619	0.169	-0.95	-0.289	0.651	0.518	0.818	0.000546	0.00698
FANCB	-0.832	0.228	-1.28	-0.385	0.562	0.412	0.766	0.000599	0.00754
SMARCB1	-0.704	0.194	-1.08	-0.325	0.614	0.472	0.798	0.000613	0.00759

	Log2 fold change	std error (log2)	Lower confidence limit (log2)	Upper confidence limit (log2)	Linear fold change	Lower confidence limit (linear)	Upper confidence limit (linear)	P-value	BH.p.value
RAD51	-1.3	0.358	-2	-0.594	0.407	0.25	0.663	0.000653	0.00795
ZIC2	-2.01	0.561	-3.11	-0.912	0.248	0.116	0.531	0.000723	0.00867
EFNA3	-0.792	0.224	-1.23	-0.352	0.578	0.426	0.783	0.00085	0.01

Supplementary Table 5: List of significantly differentially expressed genes between the good and bad outcome groups. The generalized linear model was used to determine differential expression. The p-values were corrected for multiple testing by the Benjamin-Hochberg method. Significant genes were selected if the corrected p-value was below 0.01.

References

1. Canadian Cancer Statistics Advisory Committee. Canadian Cancer Statistics 2017. (2017).
2. Cho, K. R. & Shih, I.-M. Ovarian Cancer. *Annu. Rev. Pathol. Mech. Dis.* **4**, 287–313 (2009).
3. Matulonis, U. A. *et al.* Ovarian cancer. *Nat. Rev. Dis. Primer* **2**, 16061 (2016).
4. Mutch, D. G. & Prat, J. 2014 FIGO staging for ovarian, fallopian tube and peritoneal cancer. *Gynecol. Oncol.* **133**, 401–404 (2014).
5. Duska, L. R. & Kohn, E. C. The new classifications of ovarian, fallopian tube, and primary peritoneal cancer and their clinical implications. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **28**, viii8-viii12 (2017).
6. Meinhold-Heerlein, I. *et al.* The new WHO classification of ovarian, fallopian tube, and primary peritoneal cancer and its clinical implications. *Arch. Gynecol. Obstet.* **293**, 695–700 (2016).
7. Ovarian Cancer - Cancer Stat Facts. Available at: <https://seer.cancer.gov/statfacts/html/ovary.html>. (Accessed: 11th January 2018)
8. Bowtell, D. D. *et al.* Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat. Rev. Cancer* **15**, 668–679 (2015).
9. Carcangiu, M. L. *et al.* Atypical epithelial proliferation in fallopian tubes in prophylactic salpingo-oophorectomy specimens from BRCA1 and BRCA2 germline mutation carriers. *Int. J. Gynecol. Pathol. Off. J. Int. Soc. Gynecol. Pathol.* **23**, 35–40 (2004).
10. Piek, J. M. *et al.* Dysplastic changes in prophylactically removed Fallopian tubes of women predisposed to developing ovarian cancer. *J. Pathol.* **195**, 451–456 (2001).

11. Kindelberger, D. W. *et al.* Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: Evidence for a causal relationship. *Am. J. Surg. Pathol.* **31**, 161–169 (2007).
12. Bijron, J. G. *et al.* Fallopian tube intraluminal tumor spread from noninvasive precursor lesions: a novel metastatic route in early pelvic carcinogenesis. *Am. J. Surg. Pathol.* **37**, 1123–1130 (2013).
13. Kuhn, E. *et al.* TP53 mutations in serous tubal intraepithelial carcinoma and concurrent pelvic high-grade serous carcinoma--evidence supporting the clonal relationship of the two lesions. *J. Pathol.* **226**, 421–426 (2012).
14. Salvador, S. *et al.* Chromosomal instability in fallopian tube precursor lesions of serous carcinoma and frequent monoclonality of synchronous ovarian and fallopian tube mucosal serous carcinoma. *Gynecol. Oncol.* **110**, 408–417 (2008).
15. Eckert, M. A. *et al.* Genomics of Ovarian Cancer Progression Reveals Diverse Metastatic Trajectories Including Intraepithelial Metastasis to the Fallopian Tube. *Cancer Discov.* **6**, 1342–1351 (2016).
16. Jacobs, I. J. *et al.* Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *The Lancet* **387**, 945–956 (2016).
17. Buys, S. S. *et al.* Effect of Screening on Ovarian Cancer Mortality: The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA* **305**, 2295–2303 (2011).
18. Wiltshaw, E. & Kroner, T. Phase II study of cis-dichlorodiammineplatinum(II) (NSC-119875) in advanced adenocarcinoma of the ovary. *Cancer Treat. Rep.* **60**, 55–60 (1976).

19. Einzig, A. I., Wiernik, P. H., Sasloff, J., Runowicz, C. D. & Goldberg, G. L. Phase II study and long-term follow-up of patients treated with taxol for advanced ovarian adenocarcinoma. *J. Clin. Oncol.* **10**, 1748–1753 (1992).
20. Horowitz, N. S. *et al.* Does Aggressive Surgery Improve Outcomes? Interaction Between Preoperative Disease Burden and Complex Surgery in Patients With Advanced-Stage Ovarian Cancer: An Analysis of GOG 182. *J. Clin. Oncol.* **33**, 937–943 (2015).
21. du Bois, A. *et al.* Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: A combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials. *Cancer* **115**, 1234–1244 (2009).
22. Sopik, V., Iqbal, J., Rosen, B. & Narod, S. A. Why have ovarian cancer mortality rates declined? Part II. Case-fatality. *Gynecol. Oncol.* **138**, 750–756 (2015).
23. Vergote, I. *et al.* Neoadjuvant Chemotherapy or Primary Surgery in Stage IIIC or IV Ovarian Cancer. *N. Engl. J. Med.* **363**, 943–953 (2010).
24. Kehoe, S. *et al.* Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. *The Lancet* **386**, 249–257 (2015).
25. Yang, L. *et al.* Neoadjuvant chemotherapy versus primary debulking surgery in advanced epithelial ovarian cancer: A meta-analysis of peri-operative outcome. *PLOS ONE* **12**, e0186725 (2017).
26. Wright, J. D. *et al.* Comparative effectiveness of upfront treatment strategies in elderly women with ovarian cancer. *Cancer* **120**, 1246–1254 (2014).
27. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-Throughput Sequencing Technologies. *Mol. Cell* **58**, 586–597 (2015).

28. Ledermann, J. *et al.* Olaparib Maintenance Therapy in Platinum-Sensitive Relapsed Ovarian Cancer. *N. Engl. J. Med.* **366**, 1382–1392 (2012).
29. Audeh, M. W. *et al.* Phase II trial of the oral PARP inhibitor olaparib (AZD2281) in BRCA-deficient advanced ovarian cancer. *J. Clin. Oncol.* **27**, 5500–5500 (2009).
30. Audeh, M. W. *et al.* Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *The Lancet* **376**, 245–251 (2010).
31. Fong, P. C. *et al.* Poly(ADP)-Ribose Polymerase Inhibition: Frequent Durable Responses in BRCA Carrier Ovarian Cancer Correlating With Platinum-Free Interval. *J. Clin. Oncol.* **28**, 2512–2519 (2010).
32. Ranjha, L., Howard, S. M. & Cejka, P. Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma* 1–28 (2018).
doi:10.1007/s00412-017-0658-1
33. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Mol. Cell* **47**, 497–510 (2012).
34. Usui, T. *et al.* Complex Formation and Functional Versatility of Mre11 of Budding Yeast in Recombination. *Cell* **95**, 705–716 (1998).
35. Dolganov, G. M. *et al.* Human Rad50 is physically associated with human Mre11: identification of a conserved multiprotein complex implicated in recombinational DNA repair. *Mol. Cell. Biol.* **16**, 4832–4841 (1996).
36. Paull, T. T. & Gellert, M. Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes Dev.* **13**, 1276–1288 (1999).

37. Farah, J. A., Cromie, G. A. & Smith, G. R. Ctp1 and Exonuclease 1, alternative nucleases regulated by the MRN complex, are required for efficient meiotic recombination. *Proc. Natl. Acad. Sci.* **106**, 9356–9361 (2009).
38. Zhu, Z., Chung, W.-H., Shim, E. Y., Lee, S. E. & Ira, G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* **134**, 981–994 (2008).
39. Sartori, A. A. *et al.* Human CtIP promotes DNA end resection. *Nature* **450**, 509–514 (2007).
40. Alani, E., Thresher, R., Griffith, J. D. & Kolodner, R. D. Characterization of DNA-binding and strand-exchange stimulation properties of γ -RPA, a yeast single-strand-DNA-binding protein. *J. Mol. Biol.* **227**, 54–71 (1992).
41. Heyer, W.-D., Ehmsen, K. T. & Liu, J. Regulation of Homologous Recombination in Eukaryotes. *Annu. Rev. Genet.* **44**, 113–139 (2010).
42. Sung, P. & Robberson, D. L. DNA strand exchange mediated by a RAD51-ssDNA nucleoprotein filament with polarity opposite to that of RecA. *Cell* **82**, 453–461 (1995).
43. Ma, C. J., Gibb, B., Kwon, Y., Sung, P. & Greene, E. C. Protein dynamics of human RPA and RAD51 on ssDNA during assembly and disassembly of the RAD51 filament. *Nucleic Acids Res.* **45**, 749–761 (2017).
44. Sung, P. Function of yeast Rad52 protein as a mediator between replication protein A and the Rad51 recombinase. *J. Biol. Chem.* **272**, 28194–28197 (1997).
45. Jasin, M. Homologous repair of DNA damage and tumorigenesis: the BRCA connection. *Oncogene* **21**, 8981–8993 (2002).
46. Mizuta, R. *et al.* RAB22 and RAB163/mouse BRCA2: proteins that specifically interact with the RAD51 protein. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6927–6932 (1997).

47. Xia, B. *et al.* Control of BRCA2 Cellular and Clinical Functions by a Nuclear Partner, PALB2. *Mol. Cell* **22**, 719–729 (2006).
48. Baumann, P. & West, S. C. Heteroduplex Formation by Human Rad51 Protein: Effects of DNA End-structure, hRP-A and hRad52. *J. Mol. Biol.* **291**, 363–374 (1999).
49. Ogawa, T., Yu, X., Shinohara, A. & Egelman, E. H. Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science* **259**, 1896–1899 (1993).
50. Shinohara, A., Ogawa, H. & Ogawa, T. Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* **69**, 457–470 (1992).
51. Qi, Z. *et al.* DNA sequence alignment by microhomology sampling during homologous recombination. *Cell* **160**, 856–869 (2015).
52. Bzymek, M., Thayer, N. H., Oh, S. D., Kleckner, N. & Hunter, N. Double Holliday junctions are intermediates of DNA break repair. *Nature* **464**, 937–941 (2010).
53. Schwacha, A. & Kleckner, N. Identification of double Holliday junctions as intermediates in meiotic recombination. *Cell* **83**, 783–791 (1995).
54. Ira, G., Malkova, A., Liberi, G., Foiani, M. & Haber, J. E. Srs2 and Sgs1–Top3 Suppress Crossovers during Double-Strand Break Repair in Yeast. *Cell* **115**, 401–411 (2003).
55. Plank, J. L., Wu, J. & Hsieh, T.-S. Topoisomerase III α and Bloom’s helicase can resolve a mobile double Holliday junction substrate through convergent branch migration. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11118–11123 (2006).
56. Wu, L. & Hickson, I. D. The Bloom’s syndrome helicase suppresses crossing over during homologous recombination. *Nature* **426**, 870–874 (2003).
57. Johnson, R. D. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J.* **19**, 3398–3407 (2000).

58. Cruz-García, A., López-Saavedra, A. & Huertas, P. BRCA1 Accelerates CtIP-Mediated DNA-End Resection. *Cell Rep.* **9**, 451–459 (2014).
59. Zhang, F. *et al.* PALB2 Links BRCA1 and BRCA2 in the DNA-Damage Response. *Curr. Biol.* **19**, 524–529 (2009).
60. Sy, S. M. H., Huen, M. S. Y. & Chen, J. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proc. Natl. Acad. Sci.* **106**, 7155–7160 (2009).
61. Moldovan, G.-L. & D'Andrea, A. D. How the Fanconi Anemia Pathway Guards the Genome. *Annu. Rev. Genet.* **43**, 223–249 (2009).
62. Network, T. C. G. A. R. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
63. Broca, P. *Traité des tumeurs*. (P. Asselin, 1869).
64. Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66–71 (1994).
65. Campeau, P. M., Foulkes, W. D. & Tischkowitz, M. D. Hereditary breast cancer: new genetic developments, new therapeutic avenues. *Hum. Genet.* **124**, 31–42 (2008).
66. Nielsen, F. C., Hansen, T. van O. & Sørensen, C. S. Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat. Rev. Cancer* **16**, 599–612 (2016).
67. Chen, S. & Parmigiani, G. Meta-analysis of BRCA1 and BRCA2 penetrance. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **25**, 1329–1333 (2007).
68. Mavaddat, N. *et al.* Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. *J. Natl. Cancer Inst.* **105**, 812–822 (2013).

69. Wooster, R. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789–792 (1995).
70. Kast, K. *et al.* Prevalence of BRCA1/2 germline mutations in 21 401 families with breast and ovarian cancer. *J. Med. Genet.* **53**, 465–471 (2016).
71. Walsh, T. *et al.* Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18032–18037 (2011).
72. Sedic, M. *et al.* Haploinsufficiency for BRCA1 leads to cell-type-specific genomic instability and premature senescence. *Nat. Commun.* **6**, 7505 (2015).
73. Konishi, H. *et al.* Mutation of a single allele of the cancer susceptibility gene BRCA1 leads to genomic instability in human breast epithelial cells. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17773–17778 (2011).
74. Knudson, A. G. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 820–823 (1971).
75. Abkevich, V. *et al.* Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br. J. Cancer* **107**, 1776–1782 (2012).
76. Kanchi, K. L. *et al.* Integrated analysis of germline and somatic variants in ovarian cancer. *Nat. Commun.* **5**, (2014).
77. Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science* **349**, 1483–1489 (2015).
78. 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. *J. Clin. Oncol.* **28**, 3555–3561 (2010).

79. Satoh, M. S. & Lindahl, T. Role of poly(ADP-ribose) formation in DNA repair. *Nature* **356**, 356–358 (1992).
80. Lord, C. J. & Ashworth, A. PARP inhibitors: Synthetic lethality in the clinic. *Science* **355**, 1152–1158 (2017).
81. Spriggs, D. R. & Longo, D. L. PARP Inhibitors in Ovarian Cancer Treatment. *N. Engl. J. Med.* **375**, 2197–2198 (2016).
82. Mirza, M. R. *et al.* Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. *N. Engl. J. Med.* **375**, 2154–2164 (2016).
83. Swisher, E. M. *et al.* Rucaparib in relapsed, platinum-sensitive high-grade ovarian carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial. *Lancet Oncol.* **18**, 75–87 (2017).
84. Vang, R. *et al.* Molecular Alterations of TP53 are a Defining Feature of Ovarian High-Grade Serous Carcinoma: A Rereview of Cases Lacking TP53 Mutations in The Cancer Genome Atlas Ovarian Study. *Int. J. Gynecol. Pathol. Off. J. Int. Soc. Gynecol. Pathol.* **35**, 48–55 (2016).
85. Cole, A. J. *et al.* Assessing mutant p53 in primary high-grade serous ovarian cancer using immunohistochemistry and massively parallel sequencing. *Sci. Rep.* **6**, 26191 (2016).
86. Patch, A.-M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).
87. Bieganski, K. T., Mello, S. S. & Attardi, L. D. Unravelling mechanisms of p53-mediated tumour suppression. *Nat. Rev. Cancer* **14**, 359–370 (2014).
88. Kruse, J.-P. & Gu, W. Modes of p53 Regulation. *Cell* **137**, 609–622 (2009).
89. Haupt, Y., Maya, R., Kazaz, A. & Oren, M. Mdm2 promotes the rapid degradation of p53. *Nature* **387**, 296–299 (1997).

90. Honda, R., Tanaka, H. & Yasuda, H. Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS Lett.* **420**, 25–27 (1997).
91. Kubbutat, M. H., Jones, S. N. & Vousden, K. H. Regulation of p53 stability by Mdm2. *Nature* **387**, 299–303 (1997).
92. Stenger, J. E., Mayr, G. A., Mann, K. & Tegtmeier, P. Formation of stable p53 homotetramers and multiples of tetramers. *Mol. Carcinog.* **5**, 102–106 (1992).
93. Friedman, P. N., Chen, X., Bargonetti, J. & Prives, C. The p53 protein is an unusually shaped tetramer that binds directly to DNA. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3319–3323 (1993).
94. Tavassoli, M. *et al.* Whole chromosome 17 loss in ovarian cancer. *Genes. Chromosomes Cancer* **8**, 195–198 (1993).
95. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
96. Telli, M. L. *et al.* Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Clin. Cancer Res.* **22**, 3764–3773 (2016).
97. Hwang, H. C. & Clurman, B. E. Cyclin E in normal and neoplastic cell cycles. *Oncogene* **24**, 2776–2786 (2005).
98. Etemadmoghadam, D. *et al.* Synthetic lethality between CCNE1 amplification and loss of BRCA1. *Proc. Natl. Acad. Sci.* **110**, 19489–19494 (2013).
99. Au-Yeung, G. *et al.* Selective Targeting of Cyclin E1-Amplified High-Grade Serous Ovarian Cancer by Cyclin-Dependent Kinase 2 and AKT Inhibition. *Clin. Cancer Res.* **23**, 1862–1874 (2017).

100. Polager, S. & Ginsberg, D. p53 and E2f: partners in life and death. *Nat. Rev. Cancer* **9**, 738–748 (2009).
101. Khalil, D. N., Smith, E. L., Brentjens, R. J. & Wolchok, J. D. The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. *Nat. Rev. Clin. Oncol.* **13**, 273–290 (2016).
102. Lo, C. S. *et al.* Neoadjuvant Chemotherapy of Ovarian Cancer Results in Three Patterns of Tumor-Infiltrating Lymphocyte Response with Distinct Implications for Immunotherapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **23**, 925–934 (2017).
103. Mesnage, S. J. L. *et al.* Neoadjuvant chemotherapy (NACT) increases immune infiltration and programmed death-ligand 1 (PD-L1) expression in epithelial ovarian cancer (EOC). *Ann. Oncol.* **28**, 651–657 (2017).
104. Melamed, A. *et al.* Trends in the use of neoadjuvant chemotherapy for advanced ovarian cancer in the United States. *Gynecol. Oncol.* **143**, 236–240 (2016).
105. Sokolenko, A. P. *et al.* Rapid selection of BRCA1-proficient tumor cells during neoadjuvant therapy for ovarian cancer in BRCA1 mutation carriers. *Cancer Lett.* **397**, 127–132 (2017).
106. Afghahi, A. *et al.* Tumor BRCA1 Reversion Mutation Arising During Neoadjuvant Platinum-Based Chemotherapy in Triple-Negative Breast Cancer Is Associated with Therapy Resistance. *Clin. Cancer Res.* clincanres.2174.2016 (2017). doi:10.1158/1078-0432.CCR-16-2174
107. Marques, M. *et al.* Chemotherapy reduces PARP1 in cancers of the ovary: implications for future clinical trials involving PARP inhibitors. *BMC Med.* **13**, 217 (2015).

108. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
109. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
110. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
111. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
112. Jagadeesh, K. A. *et al.* M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **48**, 1581 (2016).
113. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118–e118 (2011).
114. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
115. Bair, E. & Tibshirani, R. Semi-Supervised Methods to Predict Patient Survival from Gene Expression Data. *PLoS Biol.* **2**, (2004).
116. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5116–5121 (2001).
117. Monti, S., Tamayo, P., Mesirov, J. & Golub, T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Mach. Learn.* **52**, 91–118 (2003).
118. Şenbabaoğlu, Y., Michailidis, G. & Li, J. Z. Critical limitations of consensus clustering in class discovery. *Sci. Rep.* **4**, 6207 (2014).

119. Tomfohr, J., Lu, J. & Kepler, T. B. Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics* **6**, 225 (2005).
120. Kessous, R. *et al.* Distinct homologous recombination gene expression profiles after neoadjuvant chemotherapy associated with clinical outcome in patients with ovarian cancer. *Gynecol. Oncol.* **148**, 553–558 (2018).
121. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, W214–W220 (2010).
122. Park, C. Y. *et al.* Functional knowledge transfer for high-accuracy prediction of understudied biological processes. *PLoS Comput. Biol.* **9**, e1002957 (2013).
123. Tothill, R. W. *et al.* Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **14**, 5198–5208 (2008).
124. Gonzalez-Angulo, A. M. *et al.* Gene expression, molecular class changes, and pathway analysis after neoadjuvant systemic therapy for breast cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **18**, 1109–1119 (2012).
125. Jiang, Y.-Z., Yu, K.-D., Bao, J., Peng, W.-T. & Shao, Z.-M. Favorable Prognostic Impact in Loss of TP53 and PIK3CA Mutations after Neoadjuvant Chemotherapy in Breast Cancer. *Cancer Res.* **74**, 3399–3407 (2014).
126. Yu, K.-D. *et al.* Identification of prognosis-relevant subgroups in patients with chemoresistant triple-negative breast cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **19**, 2723–2733 (2013).

127. Magbanua, M. J. M. *et al.* Serial expression analysis of breast tumors during neoadjuvant chemotherapy reveals changes in cell cycle and immune pathways associated with recurrence and response. *Breast Cancer Res.* **17**, 73 (2015).
128. Dieci, M. V. *et al.* Prognostic value of tumor-infiltrating lymphocytes on residual disease after primary chemotherapy for triple-negative breast cancer: a retrospective multicenter study. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **25**, 611–618 (2014).
129. Balko, J. M. *et al.* Molecular Profiling of the Residual Disease of Triple-Negative Breast Cancers after Neoadjuvant Chemotherapy Identifies Actionable Therapeutic Targets. *Cancer Discov.* **4**, 232–245 (2014).
130. Penault-Llorca, F. & Radosevic-Robin, N. Biomarkers of residual disease after neoadjuvant therapy for breast cancer. *Nat. Rev. Clin. Oncol.* **13**, 487–503 (2016).
131. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
132. Sapoznik, S., Aviel-Ronen, S., Bahar-Shany, K., Zadok, O. & Levanon, K. CCNE1 expression in high grade serous carcinoma does not correlate with chemoresistance. *Oncotarget* **8**, 62240–62247 (2017).
133. Tan, Y. *et al.* Two-fold elevation of expression of FoxM1 transcription factor in mouse embryonic fibroblasts enhances cell cycle checkpoint activity by stimulating p21 and Chk1 transcription. *Cell Prolif.* **43**, 494–504 (2010).
134. Tan, Y., Raychaudhuri, P. & Costa, R. H. Chk2 Mediates Stabilization of the FoxM1 Transcription Factor To Stimulate Expression of DNA Repair Genes. *Mol. Cell. Biol.* **27**, 1007–1016 (2007).

135. Westhoff, G. L., Chen, Y. & Teng, N. N. h. Targeting Foxm1 Improves Cytotoxicity of Paclitaxel and Cisplatinum in Platinum-resistant Ovarian Cancer. *Int. J. Gynecol. Cancer* **27**, 1602–1609 (2017).
136. Lee, J.-M. *et al.* Prexasertib, a cell cycle checkpoint kinase 1 and 2 inhibitor, in BRCA wild-type recurrent high-grade serous ovarian cancer: a first-in-class proof-of-concept phase 2 study. *Lancet Oncol.* **19**, 207–215 (2018).
137. Rafnar, T. *et al.* Mutations in *BRIP1* confer high risk of ovarian cancer. *Nat. Genet.* **43**, 1104–1107 (2011).
138. Bunting, S. F. *et al.* 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* **141**, 243–254 (2010).
139. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
140. Daud, A. I. *et al.* Programmed Death-Ligand 1 Expression and Response to the Anti-Programmed Death 1 Antibody Pembrolizumab in Melanoma. *J. Clin. Oncol.* **34**, 4102–4109 (2016).
141. Daud, A. I. *et al.* Tumor immune profiling predicts response to anti-PD-1 therapy in human melanoma. *J. Clin. Invest.* **126**, 3447–3452 (2016).
142. Garsed, D. W. *et al.* Homologous Recombination DNA Repair Pathway Disruption and Retinoblastoma Protein Loss Are Associated with Exceptional Survival in High-Grade Serous Ovarian Cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **24**, 569–580 (2018).
143. Salomon-Perzyński, A., Salomon-Perzyńska, M., Michalski, B. & Skrzypulec-Plinta, V. High-grade serous ovarian cancer: the clone wars. *Arch. Gynecol. Obstet.* **295**, 569–576 (2017).

144. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
145. Westbrook, C. A. *et al.* Clinical significance of the BCR-ABL fusion gene in adult acute lymphoblastic leukemia: a Cancer and Leukemia Group B Study (8762). *Blood* **80**, 2983–2990 (1992).
146. Hauschild, A. *et al.* Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. *The Lancet* **380**, 358–365 (2012).
147. Hillman, R. T., Chisholm, G. B., Lu, K. H. & Futreal, P. A. Genomic Rearrangement Signatures and Clinical Outcomes in High-Grade Serous Ovarian Cancer. *JNCI J. Natl. Cancer Inst.* **110**, (2018).
148. Topp, M. D. *et al.* Molecular correlates of platinum response in human high-grade serous ovarian cancer patient-derived xenografts. *Mol. Oncol.* **8**, 656–668 (2014).
149. Liu, J. F. *et al.* Establishment of patient-derived tumor xenograft models of epithelial ovarian cancer for pre-clinical evaluation of novel therapeutics. *Clin. Cancer Res.* clincanres.1237.2016 (2016). doi:10.1158/1078-0432.CCR-16-1237