U.PORTO

MESTRADO

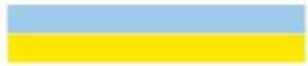
ONCOLOGIA - ESPECIALIZAÇÃO EM ONCOLOGIA MOLECULAR

Genome-wide association studies and clinical outcome in ovarian cancer patients: validation in an independent cohort

Ricardo Jorge Correia Pinto

M

2017



U.PORTO

Ricardo Jorge Correia Pinto. Genome-wide association studies and clinical outcome in ovarian cancer patients: validation in an independent cohort

M.ICBAS 2017

Genome-wide association studies and clinical outcome in ovarian cancer patients: validation in an independent cohort

Ricardo Jorge Correia Pinto

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

Ricardo Jorge Correia Pinto

Genome-wide association studies and clinical outcome in ovarian cancer patients: validation in an independent cohort

Dissertação de Candidatura ao grau de Mestre em Oncologia – Especialização em Oncologia Molecular, submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

Orientador – Professor Doutor Rui Manuel de Medeiros Melo Silva

Categoria – Professor Associado com Agregação

Afiliação – Coordenador do Grupo de Oncologia Molecular e Patologia Viral do Centro de Investigação do Instituto Português de Oncologia do Porto

Coorientador – Mestre Joana Isabel Gomes Assis

Categoria – Aluna de doutoramento

Afiliação – Faculdade de Medicina da Universidade do Porto; Grupo de Oncologia Molecular e Patologia Viral do Centro de Investigação do Instituto Português de Oncologia do Porto

INFORMAÇÃO TÉCNICA

TÍTULO:
Genome-wide association studies and clinical outcome in ovarian cancer patients: validation in an independent cohort
Dissertação de candidatura ao grau de Mestre em Oncologia – especialização em Oncologia Molecular, apresentada ao Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto
AUTOR:
Ricardo Jorge Correia Pinto

DATA:
Setembro de 2017

EDITOR: Ricardo Jorge Correia Pinto
CORREIO ELETRÓNICO: ricardojcpinto@hotmail.com
1ª EDIÇÃO: Setembro de 2017

Agradecimentos

Durante esta etapa da minha vida académica que agora termina, tive a oportunidade de alargar os meus conhecimentos numa área de eleição, assim como de crescer a nível pessoal. Para tal, foi essencial o contributo de algumas pessoas, às quais gostaria de endereçar o meu profundo agradecimento.

À Comissão de Coordenação do Mestrado em Oncologia, sob as pessoas da Professora Doutora Cármen Jerónimo e Professora Doutora Berta da Silva, atual e ex-diretoras, pela oportunidade de ingressar neste mestrado e de enriquecer os meus conhecimentos na área da oncologia.

Ao Professor Doutor Rui Medeiros, meu orientador, por me ter acolhido no seu grupo de investigação e por me ter dado a oportunidade de desenvolver este trabalho. Obrigado pela partilha de conhecimento e ideias e por todo o apoio dado para a realização deste projeto.

À Mestre Joana Assis, minha coorientadora, pelo incansável apoio dado ao longo deste caminho, pelos valiosos ensinamentos transmitidos e pela disponibilidade que sempre demonstraste. Obrigado por todos os conselhos dados, pela (muita) paciência e por acreditares em mim e neste trabalho. Referências que irei levar para a vida. Um enorme obrigado por tudo, Joana!

À Doutora Carina Pereira e ao Mestre Augusto Nogueira por toda a ajuda dada na realização deste trabalho, pela partilha de ideias que contribuíram para o seu enriquecimento e por todo o apoio dado.

À Doutora Deolinda Pereira, à Dr.ª Mariana Brandão, ao Dr. João Dias e à Dr.ª Sara Alves pelo apoio prestado na componente clinica, nomeadamente na revisão dos processos clínicos das doentes com cancro do ovário.

Aos elementos do Grupo de Oncologia Molecular e Patologia Viral, por me acolheram de braços abertos e por terem ajudado na minha integração. Pela vossa disponibilidade e por todos os momentos de descontração.

Aos meus amigos de sempre, especialmente à Alexandra, pela amizade, por me aturarem, por todos os momentos únicos que, apesar da menor frequência, mostram

realmente a verdadeira amizade que nos une. Por acreditarem em mim e sempre me apoiarem nesta longa caminhada.

Aos meus colegas do Mestrado em Oncologia, pelo espírito de grupo criado e pelos momentos de boa disposição.

Aos meus pais, pelo apoio constante e imprescindível, por acreditarem em mim e por tudo o que fizeram para que alcançasse todos os meus objetivos. Porque esta dissertação também é vossa. As palavras nunca serão suficientes para vos agradecer tudo o que fazem por mim.

Ao meu irmão Miguel, pelas brincadeiras e parvoíces, mas também pelo apoio, à tua maneira.

À minha família, por toda a força, por todos os momentos que me permitiram e permitem “recarregar baterias” e pela compreensão pelas alturas em que por vezes não estive tão presente.

Abbreviations

A

A	Adenine
AIM	Ancestry informative marker
AR	Amphiregulin
ATAD5	ATPase family, AAA domain containing 5

B

BABAM2	BRISC and BRCA1 A complex member 2
BNC2	Basonuclin 2
BOD1L1	Biorientation of chromosomes in cell division 1 like 1
BRCA1	BRCA1, DNA repair associated
BRCA2	BRCA2, DNA repair associated
BRISC	BRCC36 isopeptidase complex
BRE	Brain and reproductive organ-expressed
BTC	Betacellulin

C

C	Cytosine
CI	Confidence interval
CYP2C9	Cytochrome P450 family 2 subfamily C member 9

D

DFS	Disease-free survival
DNA	Deoxyribonucleic acid

E

EDTA	Ethylenediamine-tretaacetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EOC	Epithelial ovarian cancer
EPG	Epigen
EPI	Epiregulin
eQTL	Expressive quantitative trait loci

F

FDR	False discovery rate
FIGO	International Federation of Gynecology and Obstetrics
FSH	Follicle-stimulating hormone

G

G	Guanine
GWAS	Genome-wide association study

H

HB-EGF	Heparin-binding EGF
HDL-C	High density lipoprotein-C

HOXD1	Homeobox D1
HOXD3	Homeobox D3
HR	Hazard ratio

I

IARC	International Agency for Research on Cancer
IC_{50}	Half maximal inhibitory concentration

L

LCL	Lymphoblastoid cell line
LD	Linkage disequilibrium
LH	Luteinizing hormone

M

MAF	Minor allele frequency
MERIT40	BRISC and BRCA1 A complex member 1
MLH1	mutL homolog 1
mRNA	Messenger RNA
MSH2	mutS homolog 2
MSH6	mutS homolog 6

N

NRG	Neuregulin

O

OC	Ovarian cancer
OR	Odds ratio
OS	Overall survival

P

PCR	Polymerase chain reaction
PFS	Progression-free survival
PMS2	PMS1 homolog 2, mismatch repair system component
PSIP1	PC4 and SFRS1 interacting protein 1

S

SNCAIP	Synuclein alpha interacting protein
SNP	Single nucleotide polymorphism

T

T	Thymine
TF	Transcription factor
TGF-α	Transforming growth factor-α
TTC39B	Tetratricopeptide repeat domain 39B

V

VKORC1	Vitamin K epoxide reductase complex subunit 1

Index

Figure index

Table index

Abstract

***Introduction**:* Genome-wide association studies (GWAS) have allowed the discovery of novel and impacting findings concerning the association of single nucleotide polymorphisms (SNPs) with the susceptibility and clinical outcome of complex traits, namely in the field of oncology.

Ovarian cancer (OC) is the seventh most incidence cancer in woman worldwide. Despite the achieved improvements in diagnosis and treatment, OC is considered the deadliest gynecological cancer in the developed world. However, the identification of predictive biomarkers for OC first-line treatment remains a challenge and the results from candidate-gene studies have not reached the desired clinical implementation. In agreement, the research on this field might benefit from the accomplishment of genome-wide strategies. Since 2009, 15 OC GWAS have been performed, with the discovery of 49 SNPs associated with disease susceptibility and 46 with impact in the clinical outcome (P<5.00x10^{-2}). Despite the achieved results, they present limited implication and further validation is mandatory. So far, five validation studies have been conducted which could confirm the association of 12 OC susceptibility SNPs, although no clinical outcome associated variant was able to be validated.

Thereby, the purpose of this study was to select and validate the influence of GWAS-associated variants in an independent cohort of Epithelial Ovarian Cancer (EOC) patients from the North region of Portugal.

***Methods**:* Upon the collection of all OC GWAS-identified variants, we submitted all the clinical outcome associated polymorphisms to the SNP Prioritization Online Tool (SPOT) software, in order to select the most suitable variants to be studied, according to specific criteria. Moreover, we conducted a retrospective hospital-based cohort study gathering 339 EOC patients submitted to first-line treatment. Polymorphisms genotyping was performed by TaqMan® Allelic Discrimination methodology, using validated assays. Overall survival (OS) and disease-free survival (DFS) were the two clinical endpoints established in this study. All statistical tests were two-sided and a 5% level of significance was considered.

Results: Based on the prioritization rankings provided by the SPOT software, we select *Neuregulin 3* (*NRG3)* rs1649942 and *Brain and reproductive organ-expressed (BRE)* rs7572644 as two of the most top prioritized clinical outcome associated SNPs. Patients carrying the *NRG3* rs1649942 A allele presented a significantly longer OS when compared to GG genotype patients (log-rank test, P=0.011) in the FIGO IV stage subgroup. No impact was observed for early disease stage patients or considering DFS as outcome. We hypothesized that *NRG3* rs1649942 GG genotype might be associated with the expression of peripheral genes that promote the acquisition of an aggressive phenotype, namely with a pro-oncogenic role in the metastatic niche.

Regarding the *BRE* rs7572644 polymorphism, we observed that C allele carriers exhibit a decreased OS (*P*=0.014) and DFS (*P*=0.032), when compared to TT homozygous patients, in the subgroup of early stage disease patients (FIGO I/II). Moreover, multivariate Cox regression analysis revealed a three-fold increased risk of death (HR, 3.09; 95% CI, 1.25-7.66; *P*=0.015) and recurrence (HR, 3.33; 95% CI, 1.35-8.23; *P*=0.009) for FIGO I/II C allele carriers, after adjustment for hormonal status, histology, surgery extension and tumor grade. No significant impact was observed for late stage patients. For early disease stage patients submitted to first-line treatment, the presence of *BRE* rs7572644 C allele could lead to an improved ability to repair platinum-induced damages and the anti-apoptotic activity of EOC cells.

Conclusion: GWAS development will aid to rethink OC genomics much beyond the obvious and direct analysis, namely, through the identification of variants lying in regulatory regions of the genome with influence on complex genetic networks. Therefore, it is essential to analyze GWAS data to address the possible role of associated markers which, ultimately, could translate in clinical implementation. The *BRE* rs7572644 and *NRG3* rs1649942 GWAS-identified variants were validated in an independent cohort of EOC Portuguese patients, particularly in specific EOC subgroups considering FIGO staging. Further functional post-GWAS analyses are indispensable to understand the biological mechanisms underlying the observed results.

Keywords: Single nucleotide polymorphisms, GWAS, epithelial ovarian cancer, validation study, clinical outcome, Brain and reproductive organ-expressed, Neuregulin 3

Resumo

***Introdução**:* Os *genome-wide association studies* (GWAS) têm permitido a descoberta de novos e importantes resultados no que diz respeito à associação de polimorfismos de nucleótido único (SNPs) com a suscetibilidade e desfecho clínico de doenças complexas, nomeadamente na área da oncologia.

O cancro do ovário (CO) é o sétimo cancro mais incidente na mulher a nível mundial. Apesar das melhorias alcançadas no diagnóstico e tratamento, é considerado o cancro ginecológico mais letal nos países desenvolvidos. Contudo, a identificação de biomarcadores preditivos à primeira linha de tratamento continua a ser um desafio, e os resultados obtidos por estudos de genes candidatos não têm alcançado a implementação clínica desejada. Desta forma, a investigação nesta área poderá beneficiar da realização de estratégias *genome-wide*. Desde 2009, foram realizados 15 GWAS em CO, levando à identificação de 49 SNPs associados com a suscetibilidade para esta doença e 46 com impacto no desfecho clinico (P<5.00x10^{-2}). Apesar dos resultados obtidos, eles apresentam uma implicação limitada, sendo obrigatória a sua posterior validação. Até à data, cinco estudos de validação foram conduzidos, confirmando a associação de 12 SNPs com a suscetibilidade para CO, contudo nenhuma variante associada com o desfecho clínico foi alvo de validação.

Desta forma, o presente estudo tem como objetivo selecionar e validar a influência de duas variantes genéticas reportadas por GWAS numa coorte independente de doentes com cancro epitelial do ovário (CEO) da região Norte de Portugal.

Métodos: Após a compilação de todas as variantes genéticas reportadas por GWAS realizados em CO, submetemos todos os polimorfismos associados com desfecho clínico ao *software SNP Prioritization Online Tool (SPOT)*, de modo a selecionar as variantes genéticas a ser estudadas, de acordo com critérios específicos. Adicionalmente, conduzimos um estudo retrospetivo de base hospitalar do tipo coorte, envolvendo 339 doentes com CEO submetidas a tratamento de primeira linha. A genotipagem dos polimorfismos foi realizada recorrendo à metodologia de discriminação alélica TaqMan®, usando *assays* validados. Os dois desfechos clínicos estabelecidos neste estudo foram a sobrevivência global (SG) e a sobrevivência livre de doença (SLD). Todos os testes estatísticos foram bilaterais e um nível de significância de 5% foi considerado.

Resultados: Com base nos *rankings* de priorização fornecidos pelo *software* SPOT, selecionamos duas variantes de elevada priorização associadas com o desfecho clínico de doentes com CO, *Neuregulin 3* (*NRG3)* rs1649942 e *Brain and reproductive organ-expressed (BRE)* rs7572644. Doentes portadoras do alelo A do polimorfismo *NRG3* rs1649942 apresentaram uma SG significativamente superior quando comparadas com

doentes portadoras do genótipo GG (teste *log-rank*, *P*=0.011), no subgrupo de doentes em estadio FIGO IV. Nenhum impacto foi observado para doentes em estadios precoces da doença ou quando considerada a SLD como desfecho clínico. Assim, foi colocada a hipótese que o genótipo GG do polimorfismo *NRG3* rs1649942 poderá estar associado com a expressão de genes periféricos que promovem a aquisição de um fenótipo agressivo, nomeadamente com um papel pró-oncogénico no nicho metastático.

No que diz respeito ao polimorfismo *BRE* rs7572644, observamos que portadoras do alelo C exibiam uma menor SG (*P*=0.014) e SLD (*P*=0.032), quando comparadas com doentes homozigóticas para o alelo T, no subgrupo de doentes em estadios iniciais da doença (FIGO I/II). Ademais, a análise multivariada de regressão de Cox revelou um risco aumentado de morte (HR, 3.09; 95% CI, 1.25-7.66; *P*=0.015) e recorrência (HR, 3.33; 95% CI, 1.35-8.23; *P*=0.009) em cerca de três vezes, para doentes portadoras do alelo C em estadio FIGO I/II, após ajuste para o status hormonal, histologia, extensão da cirurgia e grau de diferenciação tumoral. Nenhum impacto significativo foi observado nas doentes em estadio avançado da doença. Em doentes em estadios precoces submetidos a tratamento de primeira linha, a presença do alelo C do polimorfismo *BRE* rs7572644 poderá conduzir a uma capacidade aumentada de reparação de danos induzidos pela quimioterapia e da capacidade anti-apoptótica das células de CEO.

Conclusão: O desenvolvimento de GWAS irá auxiliar no entendimento da genómica do CO muito além da análise óbvia e direta, nomeadamente, através da identificação de variantes localizadas em regiões regulatórias do genoma com influência em complexas interações genéticas. Consequentemente, é essencial analisar os dados obtidos por estes estudos de forma a compreender o possível papel dos marcadores associados, o que, em última instância, poderá resultar numa possível implementação clínica. As variantes previamente identificadas por GWAS *BRE* rs7572644 e *NRG3* rs1649942 foram validadas numa coorte independente de doentes portuguesas com CEO, particularmente em subgrupos específicos considerando o estadiamento FIGO. Análises funcionais pós-GWAS são indispensáveis de modo a perceber os mecanismos biológicos subjacentes aos resultados observados.

1. Introduction

1.1. Genome-wide association studies: a revolutionary tool in genetics research

With the scientific advances achieved in the last decades, it became clear that the human genetic variability is considerably higher than the initially expected. Despite some genetic variations present low impact in human health, it is considered that a fraction of these alterations is able to introduce phenotypic variations that, in a particular context, might have an impact in the development of a specific feature or in response to endo/exogenous stimulus [1, 2].

One of the most common genetic alterations are polymorphisms, i.e., DNA sequence variations where the minor allele is present at least in one percent of the population [3]. The simplest form of this variation corresponds to a single nucleotide substitution in DNA sequence, known as single nucleotide polymorphism (SNP) [3, 4]. This type of genetic alteration, which is expected to occur every 100-300 base pairs in the genome, represents almost 90% of all nucleotide variations. As SNPs are distributed throughout the entire genome, their biological effect depends upon their location, ranging from silent to gene expression or protein alterations [5].

The commonly occurring SNPs contrast with rare genetic variants (mutations), usually involved in monogenic disorders, as they introduce detrimental functional changes that ultimately lead to the pathological condition. Thus, the common disease/common variant hypothesis states that common disorders are possibly prompted by genetic variations that are also common in the general population. Consequently, common SNPs have low penetrance and the total genetic risk due to common genetic variants might be spread across numerous genetic factors [6].

Although family-based studies and genetic linkage analysis are successful approaches in the discovery of genes (and gene variants) associated with Mendelian traits, they present a limited impact in the mapping of common disease associated loci [7]. Thus, the development of population-based studies, allied to the advent of sequencing techniques and genome research methods, have allowed the discovery of important and impacting findings regarding the association of genetic markers with disease susceptibility and clinical outcome [8]. Accordingly, since 2005, a new era in genome mapping started with the accomplishment of the first genome-wide association study (GWAS) [9]. GWAS allow to test, in a cost-effective manner, the association of thousands of SNPs with a particular trait simultaneously in thousands of samples, which has revolutionized molecular and genetic research [10, 11]. GWAS are considered non-candidate gene studies that use a whole-genome approach to unravel traits such as anthropometries, pathologies or even response to drugs [12-14]. Despite its typical design based on observational studies, other methods

including lymphoblastoid cell lines (LCLs) and pooled DNA can be used, as untypical strategies.

It is estimated that nearly 10 million SNPs are inherited in haplotype blocks, meaning that a set of them (tagSNPs) are representative of the most SNPs present in each block [15]. Thereby, GWAS are based in the principle of linkage disequilibrium (LD), which represents the non-random association between alleles at different loci. Generally, loci close to each other show a stronger LD than loci distant in the chromosome. Thus, LD allows the determination of genetic markers necessary to identify a haplotype, being the number of such markers considerably lower than the total number of variants present in the population [7]. LD is generally reported in terms of r^2, a statistical measure of correlation. The higher the r^2 value, the greater the information shared by two SNPs, i.e., one allele of the first SNP is often carried together with one allele of the second SNP. In practice, it implies that there is the need to genotype only one of the two SNPs to detect the complete allelic spectrum. Consequently, the LD creates two alternative analyses as the functional SNP could be directly associated and correlated with the trait in opposite to the association of a tagSNP in high LD with the functional SNP, following an indirect approach [6] (Figure 1).

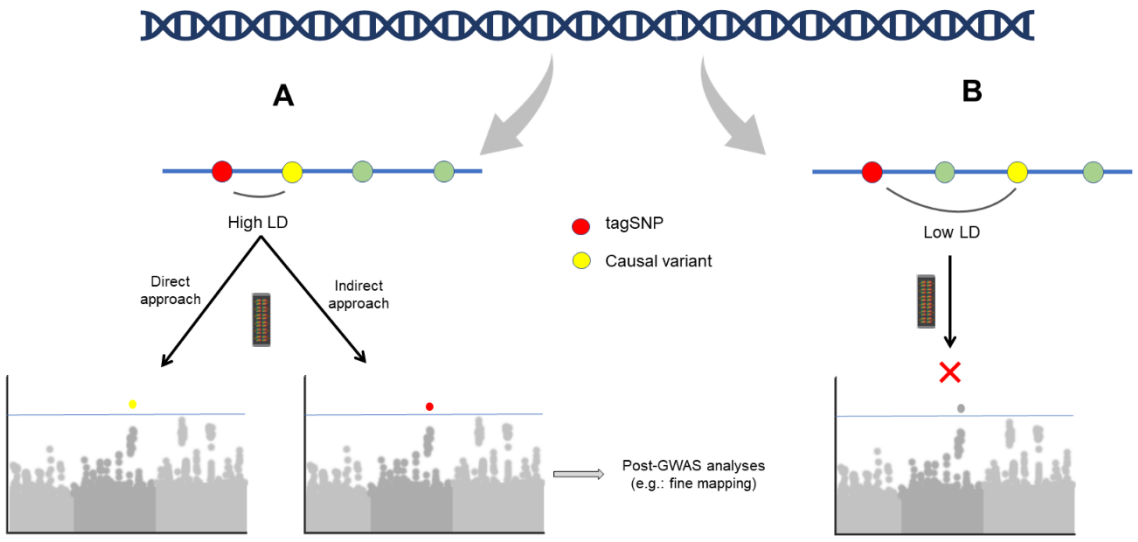


Figure 1 - Challenges beyond the use of tagSNPs in GWAS. (A): The direct correlation of a variant with a trait could occur if the causal variant is found to be associated in the GWAS analysis. On opposite, the causal variant might be in high LD with an associated tagSNP, and further analyses are needed to correlate it with a trait of interest. (B): If the causal variant is in low LD with the tagSNP, it might not be covered by GWAS analysis.

A consensus presupposes that an $r^2 \geq 0.8$ is sufficient for tagSNP mapping to reach a good coverage of untyped SNPs [16]. By the existence of these two approaches, a GWAS significant SNP cannot be assumed as the causal variant and may require further studies to map the precise location of the causal variant. Therefore, GWAS virtually allow for the putative research of the entire genome [17].

Another significant insight from GWAS include the assignment of putative risk regions in or adjacent to genes not previously predicted to be involved in the manifestation of a trait, associated loci shared by traits not previously related and the association with chromosomic regions characterized with low gene content [18]. However, the development of GWAS is also associated with some disadvantages, as addressed in Table 1.

Table 1 – Overview of the main advantages and disadvantages of GWAS.

Advantages [16, 19]	Disadvantages [10, 11, 16]
In contrast to candidate gene studies, GWAS enable the identification of novel unsuspected susceptibility factors which allows a better comprehension of a variety of phenotypes.	GWAS are associated with a high rate of false-positive results
Typically, GWAS are based in a case-control study design, which make them less expensive and allow samples acquisition in a relatively simple manner when compared to the extensive pedigrees used in linkage studies.	Interpretation of obtained results can be problematic and require fine mapping of associated loci, as well as functional studies to understand the biological plausibility of certain findings
GWAS have higher statistical power for detection of slight genetic effects than genetic linkage studies.	By restricting statistical significance thresholds, GWAS need a substantial sample size.
Since this type of studies is based on LD principle, obtained results have a more restrict location than genetic linkage studies, leading to a rapid identification of pathological variants through narrowing regions that will be analyzed in subsequent functional studies.	Due to tagSNPs use, GWAS are incapable to detect rare susceptibility variants, beyond the lack of cost-effectiveness in low LD regions, which can represent almost 20% of the genome
GWAS are cost-effective due to tagSNPs use, which cover much of the genetic variation of a region of the genome.	It is necessary a high number of association tests (at least one per SNP)

Despite the initial enthusiasm about GWAS, the obtained results fell short of expectations with the recognition that the identified loci, alone or in aggregation, typically explained a limited proportion of trait heritability [20, 21]. This feature, known as “missing heritability”, reflects the usually small effect sizes conferred by trait-associated loci (Odds Ratio (OR) often < 1.50) [22]. Due to the modest effect sizes conferred by these variations, large sample sizes are required to achieve enough statistical relevance [23]. This limitation

is meaningful even for common traits, as cancer. Inclusively, effect sizes conferred by variants associated with lung, breast, or prostate cancers, known as the most incident cancers, are usually found to be modest, with OR values ranging from one to three [24-26].

In fact, the search for “missing heritability” has become an important challenge for GWAS. Yang and collaborators have proposed three major hypotheses as the source for missing heritability: 1) rare variants (frequency < 0.01%) may also have a role in heritability estimation, as they can have a great impact on phenotype; 2) common variants with subtle effects are not covered by the current available methodologies, prompting an increase in sample sizes; 3) heritability estimation found in family studies is frequently overestimated, by not avoiding shared environmental effects [27]. Recently, beyond the agreement with these three mutually compatible hypotheses, Bourrat and colleagues have proposed that the role of nongenetic factors (epigenetics) must also be considered [28]. Since GWAS focus exclusively on DNA, this genome-wide approach does not consider non-DNA information. Given that some epigenetic factors can be stably inherited (transgenerational effect), they could also respond to selection and should be incorporated in the definition of heritability [29].

Another drawback initially proposed for GWAS is that, in contrast to protein-coding alterations characteristic of Mendelian diseases, complex traits are mainly influenced by noncoding variants with a putative role in genetic regulation [30]. Namely, significant variants are broadly enriched in regions that are transcriptionally active (or with a role in transcription) in meaningful cell types, although they are absent from transcriptionally idle regions in those cell types. However, although some of the largest-effect variants are in genes or pathways with a direct role in a trait manifestation (core genes), SNPs that vastly contribute for heritability tend to be spread across the genome (peripheral genes). As proposed by Boyle and collaborators, for typical traits, an omnigenic model must be considered, assuming that regulatory networks are highly correlated, as the expression of peripheral genes might have an impact in the regulation or function of core genes. Thereby, a phenotype should be related to a dysfunction in associated tissues and a genetic variant will only be considered relevant if it has a putative regulatory role (and hence network impact) in those tissues. In summary, a phenotype manifestation might not be assigned to a single gene or genetic loci, being largely driven by peripheral genes with an indirect role in phenotype and propagated through complex regulatory networks for which only a small number of core genes have a direct role, which is consistent with most GWAS findings [30] (Figure 2).

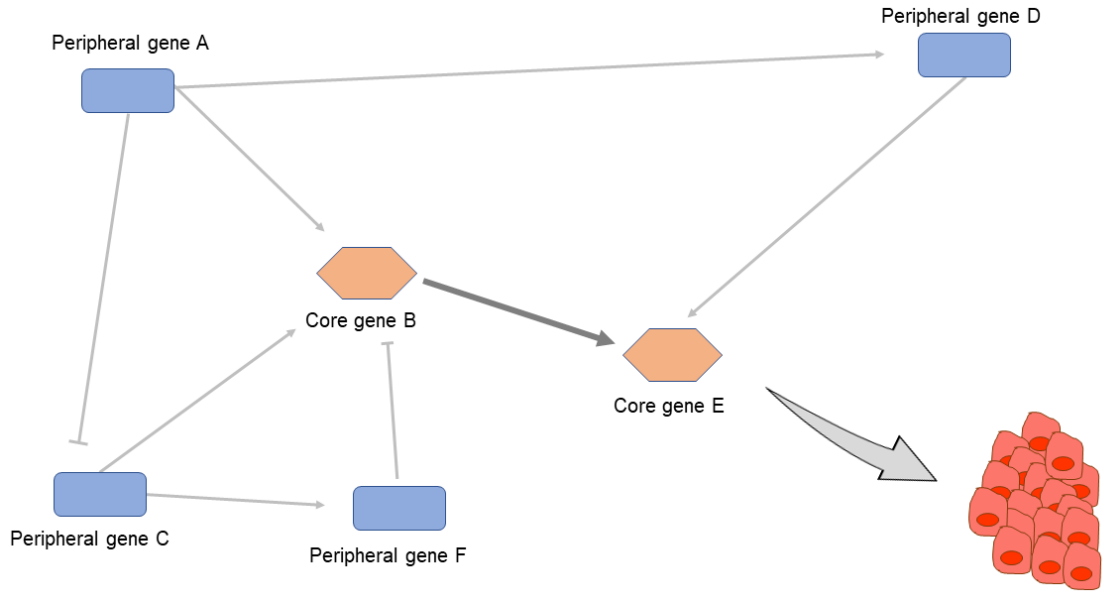


Figure 2 – Schematic representation of the omnigenic model. As postulated, several peripheral genes might contribute to the regulation of genes (core genes) with a direct role in a trait's phenotype. Thus, complex genetic networks could underlie the manifestation of a particular feature.

As in GWAS millions of SNPs are tested simultaneously for the association with a specific trait, each one with its own false positive probability, the cumulative likelihood to detect false positives is high, so multiple testing adjustment is required. A simpler approach is to perform a Bonferroni correction by the adjustment of the alpha value, generally set to $5.00x10^{-2}$, which assumes the independence of all performed tests ($\alpha = 0.05/k$, where k is the number of statistical tests conducted) [31]. Thus, a genome-wide significance threshold of $P<5.0x10^{-8}$ has generally been applied in the majority of GWAS. However, given the high number of analyses and outcomes addressed in a GWAS, even a *P*-value $\leq 10^{-10}$ might be applied to safely confirm an association [23, 32]. Additionally, an adjustment approach based in the determination of false discovery rate (FDR) can be used, which estimates the false positives among the significant results [33]. Permutation testing is another effective strategy widely applied in GWAS, where the response variable is repeatedly shuffled and an empirical *P*-value is registered [34, 35].

In the last years, GWAS development have demonstrated that common genetic variants might indicate underlying susceptibility loci to common diseases. Therefore, GWAS have been greatly explored in medical research leading to the identification of positive results in several pathologies including type I and II diabetes mellitus [36-43], inflammatory bowel disease (10-14) or coronary heart disease (25-27). A field of particular interest for

GWAS development is in oncology, with the identification of genetic markers associated with prostate cancer (15-20), breast cancer (21-23), colorectal cancer [44, 45], lung cancer [46, 47], gastric cancer [48], pancreatic cancer [49], bladder cancer [50]. A diagram of all cancer GWAS-identified variants is represented in Figure 3, according to their chromosomal localization.

Figure 3 – Diagram representing the chromosomal localization of all 864 cancer GWAS-identified variants (P<5.00x10^{-8}) [51].

The prompt increase in GWAS development provided a crucial opportunity to reveal the true impact of common genetic variations on complex traits. Besides the drawbacks associated with this genomic approach, GWAS findings might have potential clinical applicability, as the identification of risk/prognostic markers might lead to the prediction of high risk individuals or to the implementation of prophylactic strategies [52, 53]. Additionally, it is well known that patients exhibit distinct treatment response profiles, and this variability might be influenced by disease pathophysiology and drug metabolism, both features thought to be conditioned by individual's genetic background [54].

Although candidate gene studies have an important impact for personalized medicine, the development of GWAS might contribute with novel and less obvious targets to pharmacogenomics [55]. Furthermore, considering the expanded view from polygenic to omnigenic complex traits, the identification of peripheral genes (and SNPs) associated with drug response or toxicity profiles might be achieved in GWAS, contributing to expand our knowledge on variants with putative influence in treatment regulatory network. Actually,

these variants might be those that have escaped the negative selective pressures from recent and recurrent therapeutic drugs [52]. Thereby, the additional successful identification of genes involved in treatment response regulatory networks could be a major achievement in the era of treatment individualization [56-63].

Some GWAS findings are leading to clinical implementation, as the example of *CYP2C9* and *VKORC1* genetic variants (rs9923231 in *VKORC1* gene, rs1057910 and rs1799853 in *CYP2C9* gene), which have been considered by United States (US) Food and Drug Administration (FDA) as relevant biomarkers for warfarin dosage establishment [52, 64].

The oncology field is a paradigmatic example of this evident interaction between individuals' genetic profiles and treatment response phenotypes. In fact, the failure of response to a particular treatment is often observed in cancer patients and cytotoxic agents have a narrow therapeutic index, with potential for toxicity [14]. Thus, research on this field might benefit from genome-wide high-throughput approaches, with the identification of new genetic markers that could be integrated into predictive clinical nomograms.

1.2. Oncobiology and molecular epidemiology

Cancer is the leading cause of death worldwide, overcoming even de number of deaths caused by cardiovascular diseases. In 2012, according to the International Agency for Research on Cancer (IARC), 14,1 million new cases and 8,2 million deaths occurred worldwide due to this pathology [65]. Lung cancer (around 1,2 million new cases/year and 1,1 million deaths/year) and breast cancer (around 1,7 million new cases/year and 522 thousand deaths/year) are the most frequently diagnosed and deathly tumor types in man and woman, respectively. It is considered that the increase of cancer cases is due, especially, to the expansion and ageing of population, as well as the growing prevalence of risk factors associated with cancer susceptibility [66]. Thus, it is estimated that in 2030 about 22,2 million new cases will be diagnosed [67].

As a genome disease, originated by the deregulation of biological pathways and processes, cancer is characterized by a wide range of dynamic alterations (spontaneous or inherited) that affect multiple cellular systems, from molecular activity to cellular communication [68]. Namely, carcinogenesis is influenced by individual genetic background and is driven by the accumulation of multiple genetic events, as point mutations, translocations, and gene copy number variations. Epigenetic processes, as hypo and hypermethylation of specific genomic regions or variations in histones modification levels, frequently accompany these genetic alterations [68-70]. Despite the constant occurrence of

genetic and epigenetic events, only a part of them present selective growth advantage, being causal of neoplastic development [71].

Conceptually, tumorigenesis encompass three main stages: initiation, promotion and progression [72]. Initiation might occur upon a brief exposure to potent chemical, physical or biological carcinogenic agents, by the occurrence of epigenetic alterations and by the inheritance of germline alterations, which prompt non-lethal permanent genomic damage in normal cells, rendering them susceptible to both neoplastic promotion and development. Promotion is a reversible process in which the regular exposure to promoter agents prompt the proliferation of initiated cells, which favors the clonal expansion of cells with malignant potential. Consequently, the constant promotion of cellular proliferation enhances the propagation of damage caused by initiation, as well as increase the risk of additional mutations [72-74]. Finally, in the progression stage, additional mutations promote the expression of malignant features in propagated cells [73].

This complex process ultimately leads to modifications in the physiology of tumor cells responsible for the acquisition of particular hallmarks as self-sufficiency in growth factors, insensitivity to growth-inhibitory factors, altered cellular metabolism, evasion to apoptosis, unlimited replicative potential, sustained angiogenesis, tissue invasion and metastasis and capability to escape to immune response [69, 75]. Moreover, the acquisition of these hallmarks is also exacerbated by genomic instability and inflammation promoted by tumor cells [75]. However, a tumor cannot be seen as an inert mass of cells and the interaction with several other types of cells should be considered. Indeed, tumor microenvironment plays a relevant role, contributing with external signals to the development and manifestation of the malignant phenotype [75, 76]. Therefore, cancer development is considered as a dynamic process, where, in a first punctuated phase, random genomic alterations could occur, as well as the natural selection of specific clones of tumor cells, in subsequent phases [77]. Actually, a paradigmatic example of the dynamics in the tumor landscape can be ascertained by the adaptive capacity that cancer cells have to almost environmental change, with the unlimited ability to exhibit pathway superposition [78-80].

1.3. Ovarian cancer

Ovarian cancer (OC) is the seventh most common cancer among women worldwide, although represents the most lethal gynecological cancer in Western countries [66]. In 2012, almost 239 thousand newly diagnosed cases and 152 thousand associated deaths (3.6% of cases and 4.3% of deaths by cancer in women) were registered [65]. In Portugal, it is

estimated that around 620 new OC cases occurred with almost 380 deaths, with an age-standardized incidence and mortality rates of 8,2 cases and 4,4 deaths/100 000 women, respectively [81].

Based on the World Health Organization (WHO) criteria for gynecological tumors, ovarian neoplasms are classified according to its cellular heterogeneity and, therefore, might be integrated into three major groups (epithelial, sex cord and ovarian stroma, and germ cell tumors) [82]. Although the high cellular diversity, a great proportion (almost 90%) of ovarian tumors arise from the epithelial surface [83, 84]. Accordingly, epithelial ovarian tumors can be further divided into seven histological subcategories, namely serous, mucinous, endometrioid, clear cell, Brenner, seromucinous and undifferentiated (Figure 4). [82]. Within these histological subgroups, except for the undifferentiated subtype, epithelial tumors can be further characterized grounded on their behavior, being considered as benign, borderline or malignant [85, 86]. In fact, ovarian epithelial malignant tumors constitute almost 30% of epithelial neoplasms [85].

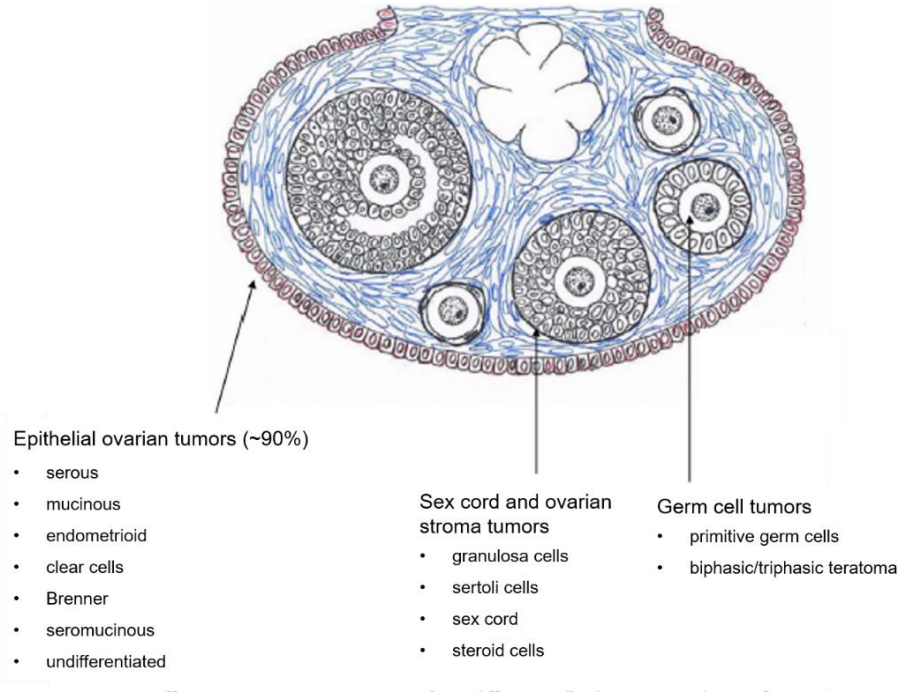


Figure 4 – Ovarian cancer cellular heterogeneity. Ovarian tumors can have an epithelial, sex cord and ovarian stroma, and germ cells origin (adapted from [113]).

Due to the great OC cellular heterogeneity, and considering that each histological subtypes exhibit distinct clinical and treatment patterns, the correct staging of the tumor is indispensable to a successful disease management. OC staging is surgical, being performed based on the International Federation of Gynecologists and Obstetricians (FIGO) guidelines [86]. According to these criteria, stage I tumors are confined to ovaries; stage II

tumors involve one or both ovaries, presenting pelvic invasion, or primary peritoneal tumors; stage III tumors involve one or both ovaries, with cytologically or histologically confirmed dissemination to the peritoneum outside the pelvis and/or metastasis to the retroperitoneal lymph nodes; stage IV tumors encompass those with distant metastasis (excluding peritoneal metastasis). Based on the FIGO criteria, OC staging not only provides essential information regarding the histopathology of the disease, but also reflects its dissemination pattern. In fact, OC dissemination occurs preferentially from primary organ to the peritoneal cavity, through exfoliation, transperitoneal migration and malignant cells deployment (transcavitary route). Alternatively, ovarian malignant cells might disseminate via lymphatic system, through lymph nodes, or hematological system, up to parenchyma of distant organs, as liver, lung or brain [87, 88]. However, the transcavitary propagation route is the most clinically relevant as it is an early event in the natural history of the disease which, in most of cases, has impact in prognosis. Due to this dissemination pattern and its early stage indolent nature, several organ systems are already affected at the time of diagnosis [89].

In the attempt to characterize the epithelial ovarian carcinogenesis and etiology, several theories have been proposed (Table 2). Despite none of these theories completely clarify epithelial ovarian cancer (EOC) etiology, it is likely that they act synergistically, evidencing the complex and multifactorial nature of ovarian tumors. Nevertheless the importance of the proposed theories, age is considered as a major risk factor for OC development: 80% of OC cases are diagnosed after the age of 45 years, with a peak of incidence after the menopause [90, 91].

Table 2 – Explicative hypothesis of epithelial ovarian cancer etiology (adapted from [92])

Hypothesis	Biological mechanism
Incessant Ovulation [93-100]	Ovarian surface epithelium is constantly exposed to repetitive cycles of damage (from the ovulation process) and repair (with subsequent cellular proliferation), which propitiates the acquisition of genomic lesions in ovarian epithelial cells. The accumulation of these lesions might lead to the development of inclusion cysts which subsequently can trigger OC.
Gonadotropins [101, 102]	Excessive stimulation of ovarian epithelial cells by gonadotropins follicle-stimulating hormone (FSH) and luteinizing hormone (LH) promotes cellular proliferation and neoplastic development.
Hormones [98, 100]	Through the direct interaction with the ovarian epithelium, estrogens and androgens might have a tumorigenic role and progestins may confer a protective effect against tumorigenesis.
Inflammation [98, 100]	Inflammatory factors resulting from the ovulatory process or concomitantly released with other pathological processes may damage ovarian epithelium and trigger tumorigenesis.

Moreover, a portion of OC cases (10-15%) arises in individuals with genetic predisposition, making it one of the key risk factors in ovarian tumorigenesis [98]. A family history of OC confers, in women with affected first-degree relatives, a three/four-fold increased risk of OC development [103]. Hereditary Breast and Ovarian Cancer Syndrome, which is associated with *BRCA1* (3-6%) and *BRCA2* (1-3%) gene mutations, is the most common cause of inherited OC [104, 105]. Despite the low frequency of these mutations, the lifetime risk of OC development for *BRCA1* mutation carriers is 40%, whereas for *BRCA2* mutation carriers is 20%, unlike to the 1,4% lifetime risk for general population [104, 106]. The second most common cause of inherited ovarian cancer is Lynch Syndrome, which is associated with germline mutations in mismatch repair genes *MLH1*, *MSH2*, *MSH6* and *PMS2* (1-2%) [104, 107]. Though, apart from the already identified syndromes, some studies point that the known susceptibility genes explain only a small fraction of the familiar risk [8, 108, 109].

Additionally, endocrine and reproductive factors seem to be important whereby the nulliparity, early menarche, late menopause and the exposure to hormone replacement therapy are considered as potential risk factors for OC development [84].

Due to the asymptomatic nature of the disease and the lack of methods for its early detection, nearly 70% of OC cases are diagnosed in an advanced stage. Furthermore, the development of resistance to current therapies is a common feature in these patients, which results in a 5-year survival rate of only 30% [110, 111]. However, a high 5-year survival rate (90%) is seen in patients with localized tumor at diagnosis [111, 112].

Over the past decades, significant advances have been achieved in the OC treatment. The combination of cytoreductive surgery followed by the doublet of platinum (carboplatin or cisplatin) and taxane (paclitaxel or docetaxel), every 21/21 days for 6 cycles, has been the standard regimen for EOC first-line treatment [113, 114]. Although EOC is considered a chemosensitive tumor, as most of patients respond to the first-line chemotherapy achieving tumor response rates up to 80%, a great percentage of them will relapse.

Disease stage, tumor size, histological subtype, differentiation degree and the extent of residual disease are considered as classic prognostic factors for OC. In opposite, the determination and selection of predictive biomarkers for OC first-line treatment has proved to be a challenge, which establish this field as a domain of high priority research.

1.4. Ovarian cancer as a study model for GWAS development

As previously mentioned, genetic factors are one of the key risk factors for OC [115]. Over the years, linkage analyses have concluded that several high penetrance genes have a crucial role in the malignant transformation of the ovary, although, with the postulation of the "common disease-common variant" hypothesis, common variants have also been associated with OC development [116]. Moreover, the identification of variants relevant to the response to treatment and survival in OC patients might contribute to a better understanding of prognosis, ultimately guiding the selection of improved chemotherapy schemes [117]. To date, several candidate gene studies have been performed allowing the recognition that individual genetic profiles have the potential to influence inter-patient variability in drug response and, hence, OC treatment outcome [92, 118-129]. Some of the obvious candidate genes encode drug metabolism enzymes or DNA repair mediators that can influence treatment response, toxicity profiles and, lastly, survival endpoints [130, 131]. Genome-wide strategies that encompass SNP genotypes, drug-induced cytotoxicity in cell lines and gene expression data are potential models for the identification of predictors of treatment and clinical outcome in OC patients [132].

To this date, 15 OC GWAS have been performed, most of them evaluating susceptibility loci (Supplementary Table 1), although the impact regarding clinical outcome has also been investigated (Supplementary Table 2) [133-145]. To highlight that the majority of the variants reviewed and presented are associated with a significant *P*-value ($P<5.00\text{x}10^{-2}$), despite only a small fraction reached the desired genome-wide threshold ($P<5.00\text{x}10^{-8}$).

Replicated and/or validated susceptibility and clinical outcome associated SNPs lying within genes are schematically grouped in Figure 5 according to the molecular pathway(s) they are involved. Each group was constituted based on information available in an online database, Reactome, which resulted in the establishment of seven main groups [146].

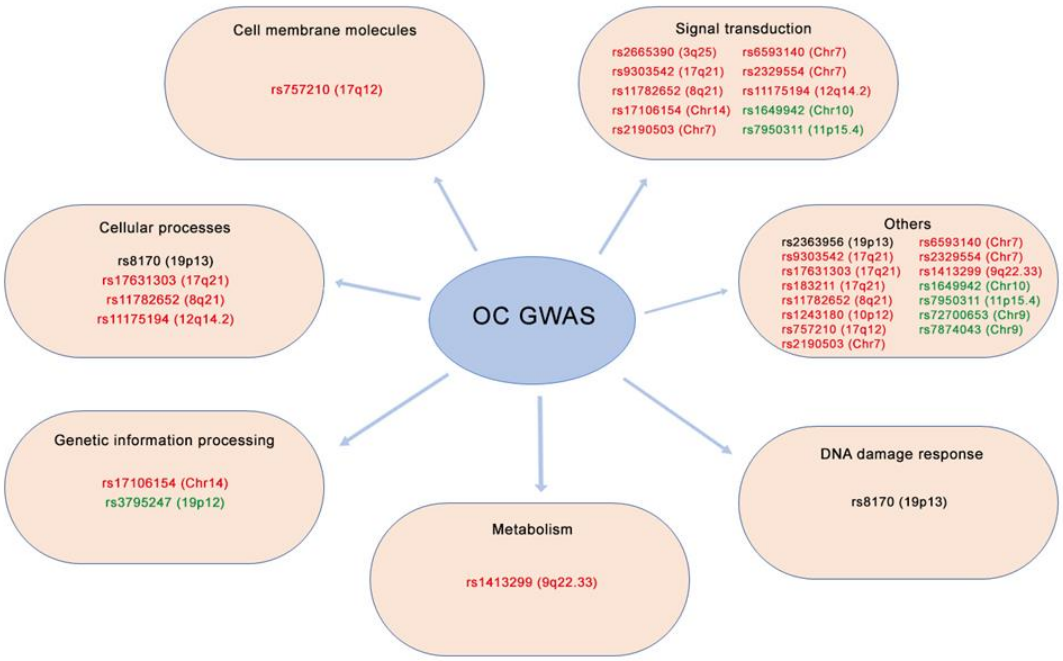


Figure 5 - Molecular pathways in which replicated and/or validated susceptibility and clinical outcome associated SNPs are involved, according to the information presented in Supplementary Tables 4 and 5. Highlighted in red are SNPs associated with OC susceptibility; highlighted in green are SNPs associated with OC clinical outcome and highlighted in dark are SNPs associated both with OC susceptibility and clinical outcome. Briefly, "signal transduction" set refers to SNPs in genes involved in signaling transduction pathways; "cell membrane molecules" set includes cell surface receptors and transporters subgroups; "cellular processes" set covers SNPs lying in genes involved in cell cycle, apoptosis, cellular senescence, cell motility and vesicle-mediated transport and catabolism; "genetic information processing" set encompass SNPs involved in gene expression and replication processes, as well as ubiquitin-mediated proteolysis processes; "metabolism" set enclose genetic variants associated with biomolecules and energy metabolism; in "DNA damage response" set, SNPs affecting DNA damage repair systems are presented and "others" set includes subgroups related to less representative pathways.

1.4.1. OC susceptibility GWAS

In 2009, the first OC GWAS was conducted by Song and collaborators concerning the identification of common OC susceptibility alleles [133]. The authors have evaluated more than 500 thousand SNPs, reporting the association of 12 SNPs at 9p22 locus with the risk of OC development ($P < 10^{-8}$). The most significant associated SNP (rs3814113) was also genotyped in a replication set which confirmed the association (combined OR=0.82; $P_{trend}=5.10x10^{-19}$). Moreover, the authors reported that the association is distinct regarding the histological subtype, being the strongest association obtained for serous OC subgroup (OR = 0.77; $P_{trend}=4.10x10^{-21}$). This SNP is localized near the *BNC2* gene, which encodes a zinc finger protein highly expressed in reproductive tissues and involved in DNA transcription (Supplementary Table 3). However, none of the top SNPs seem to be

associated with *BNC2* expression or to be near predictable or known transcription regulation elements, demanding for further evaluations [133].

After the publication of the first GWAS, ten additional OC susceptibility studies have been conducted [134, 135, 137-141, 143, 147, 148] (Supplementary Table 1). Most of these studies were case-control based, although studies conducted by Lu *et al*. [137] and Earp *et al.* [141] used a pooled DNA-technique in one of the GWAS phases. This approach allows to reduce the cost of genotyping, maintaining the study power, through the construction of two sets of pools composed by a combination of the same amount of DNA from cases and controls. Subsequently, these pools are genotyped and allelic frequencies calculated [149]. Briefly, 49 SNPs were found to be associated with susceptibility to EOC. Among them, 14 SNPs reached genome-wide significance for EOC risk, being the three most significant the genetic variants rs3814113 (P_{trend}=5.10x10^{-19}), rs2072590 (P=4.50x10^{-14}) and rs199661266 (P=1.00x10^{-9}) [133, 135, 143] (Supplementary Table 1). The SNP rs3814113 lies ~44kb upstream of *BNC2* gene, which encodes a protein highly expressed in reproductive tissues, being a potential regulatory protein of DNA transcription [133]. The genetic variant rs2072590 lies within a non-coding region downstream of *HOXD3* and upstream of *HOXD1* and it tags SNPs in the *HOXD3* 3' untranslated region, genes found to be implicated in cancer development [135]. The associated variant rs199661266 locates in intron 6 of *ATAD5*, a gene involved in DNA damage response [143] (Supplementary Table 3).

In this set of GWAS, were also found several markers associated with the susceptibility to particular EOC subtypes, highlighting the role of distinct molecular pathways underlying each histological subtype [134, 135, 138-141, 143, 147, 148]. In this context, 31 SNPs were found to have an influence in the susceptibility to OC serous subtype (top associated SNP: rs3814113; P_{trend}=4.10x10^{-21}) [133], 13 SNPs were associated with endometrioid subtype (top associated SNP: rs555025179; P=4.50x10^{-8}) [148], 13 SNPs with mucinous subtype (top associated SNP: rs112071820; P=1.50x10^{-13}) [148], seven SNPs with clear cell subtype (top associated SNP: rs757210; P=3.90x10^{-6}) [139] and one SNP with low-malignant potential serous subtype (top associated SNP: rs9609538; P=7,00x10^{-4}) [141]. Five additional SNPs were associated with less representative subtypes (top associated SNP: rs1413299; P=9.69x10^{-6}) [140] (Supplementary Table 1).

Regarding effect sizes conferred by associated markers, they have been found to have generally low to moderate effects, with ORs ranging from 0.67 to 2.19 (Supplementary Table 1). These values are in accordance with genetic variants associated with the susceptibility for other cancer types, as previously mentioned [24-26].

Concerning OC susceptibility GWAS, a highlight must be given to the study performed by Bolton and colleagues that reported that the BRCA1-interacting gene *MERIT40* might be a relevant gene underlying the genetic influence to EOC, supporting a

role of the 19p13 locus in OC susceptibility. These results revealed special interest by the fact that genetic variants in this region have already been identified as modifiers of breast cancer risk in *BRCA1*-mutated carrying patients [134].

1.4.2. OC clinical outcome GWAS

As OC patients may inherently vary in their ability to respond to treatment, genetic association studies have sought to identify variants with impact in clinical outcome. In 2010, Bolton and colleagues performed the first OC clinical outcome GWAS, evaluating more than 250 thousand SNPs, from which rs8170 (Hazard Ratio (HR)=1.11; P_{trend}=5.20x10^{-4}) and rs2363956 (HR=1.09; P_{trend}=5.60x10^{-4}), located in locus 19p13, were associated with OC survival (Supplementary Table 2). The presence of the variant allele was associated with a slightly decrease in patients' survival for both variants [134]. However, this first GWAS did not replicate any survival-associated SNPs.

In the total, five OC GWAS have evaluated the association of genetic markers with patients' clinical outcome [134, 136, 142, 144, 145] (Supplementary Table 2). Despite the case-control study design used by Bolton and collaborators, the other four clinical outcome GWAS were cohorts or used a LCLs approach. Briefly, in most cell line-based GWAS, LCLs are exposed to increasing concentrations of a drug, and individual cellular sensitivity to the agent is measured by cell growth inhibition or apoptosis ability [14]. The GWAS is simultaneously conducted, often including genome-wide genotype and gene expression evaluation [150]. In the context of a cell-based model, Huang and colleagues employed a GWAS to identify germline variants with clinical applicability. In this study, the risk allele of rs1649942 was significantly associated with an increased risk of disease progression and death in phase 1 patients submitted to carboplatin-based chemotherapy, with a greater genetic contribution among the subset of patients with optimally debulked tumors. Since clinical outcomes obtained from optimally debulked patients might represent the ideal treatment scenario for OC, in order to eliminate the confounder effect associated to the presence of residual disease, the role of rs16499942 was addressed in this subset of patients. However, the results were not replicated in phase two, which was attributable to a different categorization of residual disease and to the fact that patients were presumed, but not known, to have had standard doses of paclitaxel and carboplatin, which might reflect the impact of distinct clinical definitions across studies [136].

Clinical outcome in OC patients has been reported to be under the influence of 46 SNPs. The three most associated SNPs reported by these GWAS were rs185229225 (P_{meta}=2,2x10^{-7}), rs3842595 (P_{meta}=2,6x10^{-7}) and rs4910232 (P=4,7x10^{-7}), although any of the 46 associated SNPs have reached genome-wide significance [142, 145]. The

polymorphism rs185229225 lies within *BOD1L1*, a gene coding a protein found to be a protection factor of replication fork [145, 151] (Supplementary Table 4). SNP rs3842595 is located within *SNCAIP* gene, which is involved in ubiquitin mediated proteolysis system and in Parkinson's' disease [145] (Supplementary Table 4). With respect to variant rs4910232, it lies within a non-coding region [142] (Supplementary Table 4). Nevertheless, four SNPs were associated with clinical outcome in serous subtype, being rs7874043 the most significant SNP (P=7,3x10^{-5}), located within a putative regulatory element of *TTC39B* gene and, consequently, potentially involved in the regulation of High Density Lipoprotein-C (HDL-C) levels [144] (Figure 5, Supplementary Table 4). Moreover, rs8170 was associated with endometrioid subtype (P_{trend}=3.00x10^{-2}), and lies within *MERIT40*, a gene involved in DNA double strand break response pathway and with a role in cell cycle checkpoints [134] (Figure 5, Supplementary Table 4).

In concordance with the variants identified in susceptibility OC GWAS, SNPs that were found to be associated with OC patients' clinical outcome do not confer large effect sizes, despite the slightly increase in HR values (ranging from 1.07 to 1.91) (Table 3).

Regarding OC clinical outcome GWAS, a special focus must be given to the study performed by Johnatty and colleagues, which selected SNPs for replication in the presence of a good imputation quality ($r^2 \geq 0.9$) to minimize the risk of false positives. This study was one of the largest performed studies that evaluated the presence of genetic variation across the genome for a possible association with OC clinical outcome, either in regard to first-line standard treatment and regardless of treatment scheme. Once again, by the role of residual disease extension as predictive marker, patients were only included if they received a minimum of cytoreductive surgery and had available information on level of residual disease. Factors of appreciation of this study are that, beyond the SNPs prioritization on basis of good imputation quality, the final estimates were derived from meta-analysis of all available imputed data and genotypes from OC consortiums and the analysis was restricted to European invasive EOC patients with standardized clinical and pathological information [142].

Moreover, French and collaborators have identified two SNPs, located in a intronic region of *TTC39B*. Functional tests have showed that the likely functional SNP is rs7874043, which alters transcription factor (TF) binding and, ultimately, chromatin conformation (Figure 5, Supplementary Table 4). Moreover, the silencing of *PSIP1*, one of the targets of the regulatory element, significantly impaired DNA damage repair by homologous recombination in OC cell lines, suggesting that *PSIP1* is a potential target for a therapeutic approach as previously suggested for other tumors [152]. A high expression of *PSIP1* is also associated with high risk of recurrence proposing that altered *PSIP1* expression may be a functional consequence of associated SNP [153]. The strengths of the

study performed by French and colleagues are that they only included cases that received standard first-line treatment (duplet carboplatin/paclitaxel) and the progression-free survival (PFS) was the survival endpoint evaluation, rather than overall survival (OS) following exposure to multiple drugs [144].

1.5. OC and GWAS: the challenge of post-GWAS research

Since Song and collaborators conducted in 2009 the first OC GWAS, several loci have been identified as being associated with susceptibility and clinical outcome, although without reaching definitive and conclusive results. In fact, as previously mentioned, despite the relative success of GWAS, they are associated with some disadvantages, when compared to other genetic studies, as the GWAS' low statistical power in addition to high rates of false positive results. Consequently, a great number of association tests are necessary, as it is a wide study size due to strict statistical significance thresholds [10, 11, 16].

Additionally, the failure to conclude an association in OC GWAS might be caused by a variety of factors such disease heterogeneity as in most GWAS the initial analysis, that allows SNP selection for succeeding phases, combine all EOC histological subtypes whereas the subsequent phases might be performed based on a subgroup stratification. An example is the study performed by Bolton and colleagues [134]. Limitations in GWAS might also reflect heterogeneous clinical criteria and the use of self-reported ethnicity. However, the later limitation could be overcome using ancestry informative markers (AIMs) to define ethnicity, as exemplified in studies conducted by Song, Bolton and Goode [133-136].

Another factor that might explain the spurious results obtained by GWAS concerning to cancer treatment response is that, for most cancers, a wide variety of chemotherapeutic regimens is administered, which might contribute to heterogeneity in treatment response [154-156]. Moreover, the ability to obtain follow-up data for a suitable number of patients might restrict the detection of statistical relevance regarding survival endpoints. Accordingly, pharmacogenomics GWAS have failed to identify associations with relevant effect sizes for a genome-wide significance mainly due to the small samples sizes, as the low incidence of OC preclude the acquisition of a large quantity of samples, as well as cancer drug non-responders are often less frequent than responders. Moreover, drug response is not always measured in a completely quantitative manner, which prompt a heterogeneous phenotype definition [144]. Therefore, in order to overcome the small number of samples available, most of the performed OC GWAS have recruited patients from several consortiums [133-135, 139, 142-144].

Thereby, the major challenge in in this area of research is to discriminate the true associations from false positives through attempts to replicate positive findings in subsequent studies. Thus, it becomes essential to adopt strategies that enable researchers to support the obtained results. In this perspective, fine mapping, functional analyses or the study of populations from many geographical ancestries have become important approaches that will complement GWAS findings and will help to achieve a better comprehension of the role of the identified variants in disease genetics [10, 20, 157, 158].

Validation studies in an independent set or using different methodologies are another strategy with possible impact in GWAS evaluation. They allow the assessment and confirmation of previously reported results in order to validate them and/or to refute chance or possible bias as probable cause of association [159]. Moreover, they minimize false positive results observed in common low penetrance alleles, since only a limited number of variants are truly risk alleles [160-162]. Another reason why validation is important is that, in a first GWAS, the effect of an association is usually overestimated. Thereby, as validation studies are made, the estimate effect declines, in a phenome known as “winner’s curse” [31]. Consequently, the calculation of validation sample size is a major determinant to be taken in consideration [163].

Although the terms “replication” and “validation” are interchangeably used, they have different meanings: in a replication study, the replication sample is an independent dataset drawn from the original sample (discovery sample), in an attempt to confirm the effect in the GWAS target population [164] (Figure 6). As loci are replicated as truly causal, extension into multiple ethnicities is highly recommended to determine the generalizability and consistency of the proposed markers [165, 166]. On the other hand, in validation studies, the validation sample and the original sample are originated from different populations, meaning that an independent validation subset is used [165].

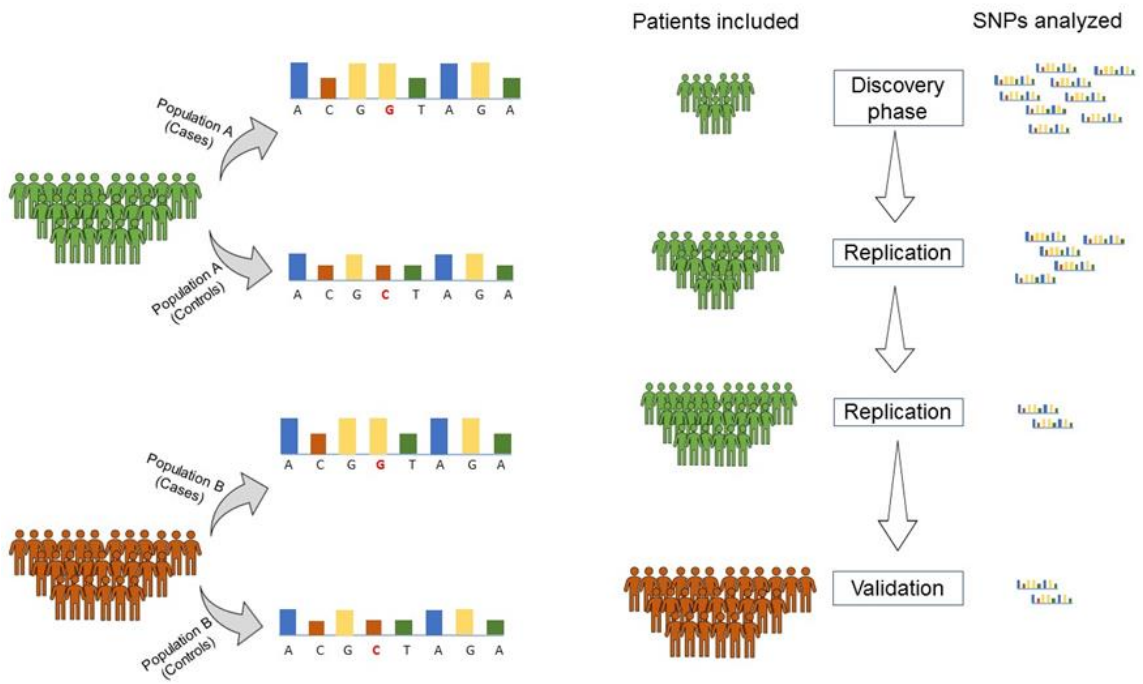


Figure 6 – Schematic representation of the rational beyond validation and replication studies. In a GWAS, hundreds of thousands of SNPs are genotyped in a limited number of samples in a discovery phase, in order to select variants associated with the interest trait. Then, these SNPs are genotyped in replication stages with larger sample sizes to confirm their association. Validation studies use a sample set recruited from a different population to confirm reported associations.

The intention of replication and validation studies is not only to provide further evidence to accept or refute the original association but also to systematic evaluate the potential sources of error or bias underlying the GWAS [166]. Validation studies have been made for several types of cancers, however, to the best of our knowledge, only five studies were conducted in OC [136, 137, 140, 141, 167]. As mentioned before, Huang and collaborators used a cell-based approach to identify SNPs associated with carboplatin sensitivity and performed, subsequently, a two-phase validation in clinical samples for SNPs found to be associated in the cell-based phase. However, no subsequent confirmation was performed for the second stage validation [136]. In the studies conducted by Lu *et al* (2012) and Earp *et al* (2014), a pooled DNA strategy was used and, to validate associated SNPs from the previous phase, they were genotyped in a large set of OC samples. Although in the later, all SNPs associated with OC susceptibility were validated, Lu and colleagues were not able to confirm any association [137, 141]. In a study conducted in Han Chinese women, Chen and collaborators tested the association of previously reported SNPs with OC risk and validated one SNP (rs9303542). In 2014, an independent validation study conducted by Mostowska and collaborators evaluated the association of seven GWAS-associated

SNPs in a Polish population regarding susceptibility to OC. They have statistically validated loci 9p22 and 8q24 for all OC patients and, more specifically, for serous subtype [167].

As previously mentioned, OC candidate gene studies are mainly conducted for variants in genes or pathways with an obvious role in the disease or in the response to platinum/taxane duplet. Besides their putative direct influence, no conclusive results have been achieved, since only a few solid associations were reported and, even less, were positively replicated and/or validated [116, 168]. The existence of non-definitive associations seems to be concordant with the OC GWAS results, as only a small part of associated GWAS variants are consistently replicated and/or validated and few robust associations have been reached ($P<5.00x10^{-8}$) (Figure 5).

2. Aims

2.1. Main aim

Independent validation of GWAS-identified *Neuregulin 3* (*NRG3)* rs1649942 and *Brain and reproductive organ-expressed* (*BRE)* rs7572644 genetic variants as predictive biomarkers, in a cohort of OC patients from the North region of Portugal.

2.2. Specific aims

- Literature review on OC GWAS;
- Selection of genetic variants previously reported to be associated with clinical outcome in OC patients;
- Validation of GWAS-associated *NRG3* rs1649942 and *BRE* rs7572644 in an independent cohort of OC patients.

3. Materials and methods

3.1. Study population description

We performed a retrospective hospital-based cohort study on European female patients with the histological confirmation of EOC, admitted between January of 1996 and December of 2012 in the departments of gynecology and oncology of the Portuguese Institute of Oncology, Porto, Portugal (IPO-Porto). From this group of patients were excluded those who were only admitted for a second opinion or to be submitted to specific treatment techniques, namely hyperthermic intraperitoneal chemotherapy or with follow-up in other institutions. A cohort of 339 conveniently sampled patients from the North region of Portugal and for which biological material was available was enrolled.

Tumor staging was performed according to the FIGO guidelines and the assessment of the tumor response to chemotherapy was based on Rustin criteria [169, 170]. Patients' clinicopathologic and follow-up data were obtained from their medical records. The mean age of included patients was 55 years (median = 54 years; minimum = 18 years; maximum = 80 years), from which 57.2% were post-menopausal women. The majority of patients was diagnosed with advanced disease stage (57.5% FIGO III/IV). The distribution considering the extent of residual disease occurred as follows: in 41.9% of the cases was achieved an optimal surgical resection whereas 11.8% and 26.2% presented residual disease ≤2 cm and >2 cm, respectively (no information available for 20.1% of the patients). Considering the histological subtype, 56.9% presented serous tumors, 12.1% clear cell, 10.0% mucinous, 10.0% endometrioid and the remaining 10.9% less common subtypes. Regarding the therapeutic strategy, 89.5% of patients were submitted to the standard regimen based on cytoreductive surgery followed by a combination of Paclitaxel (175mg/m^2) and Cisplatin (75 mg/m^2) or Carboplatin (Area under the curve 5-7.5), although doses were adjusted whenever severe toxicity was reported. Chemotherapy alone (5.1%), neoadjuvant chemotherapy (3.7%) or only surgery (1.7%) were also considered as first-line treatment options.

Follow-up information was reviewed from the initial diagnosis through December 2016 in 322 patients (95% of all patients). The mean follow-up of all participants enrolled in the study was 89.9 months (median = 68.5 months; minimum = 2 months; maximum = 246 months).

A written informed consent was obtained from each participant prior to their inclusion in this study, according to Helsinki Declaration principles. Furthermore, this study was approved by the ethics committee at IPO-Porto (CES IPO:286/2014).

3.2. Laboratory procedures

3.2.1. Sample collection and genomic DNA extraction

Peripheral venous blood samples were obtained with a standard technique and collected in ethylenediamine-tretaacetic acid (EDTA)-containing tubes.

Genomic DNA was extracted from peripheral blood samples using the extraction kit Qiagen®, QIAmp DNA Blood Mini Kit (Qiagen® 51106), as indicated by the manufacturer's procedure.

3.2.2. SNP selection

To select the variants to be validated in this study, we gathered all polymorphisms statistically associated with OC clinical outcome, identified in a recently accepted review on OC GWAS studies carried-out by our group [171], which were further submitted to the SNP Prioritization Online Tool (SPOT) software [172]. Based on the priority ranking returned by the SPOT software (which takes into account the *P*-value reported by the original study and the possible functional impact of each variant considering its genomic location), the minor allele frequency (MAF) in the Iberian population (>15%), the availability of the respective genotyping assay, the putative relevance in ovarian biological pathways and the previously accomplishment of validation studies, the *NRG3* rs1649942 and *BRE* rs7572644 SNPs were selected, among the top prioritized polymorphisms .

3.2.3. Polymorphisms genotyping

Genotyping for *NRG3* rs1649942 and *BRE* rs7572644 genetic variants was performed using TaqMan® Allelic Discrimination methodology (Figure 7), through the Real-Time Polymerase Chain Reaction (PCR) technique. The two validated assays were: C__29412070_10, targeting the rs7572644 polymorphism, whose VIC dye probe was associated with C allele and FAM dye probe was associated with T allele (TAAGAGCCATGGGGAACCATAGCTG[C/T]AGGGAAACCGTGATGCCTGCCAGCA) and C___8914657_10, targeting rs1649942 polymorphism, whose VIC dye probe was associated with A allele and FAM dye probe was associated with G allele (GCCCTGCGGTTGAGGGTTCTTGCCA[A/G]TTCGATTCTAATACATGAACACTTT).

Real-time PCR reactions were carried out using 6 µL reaction mixture, containing the following components: 2.5 µL of TaqPath™ ProAmp™ Master Mix (1x), 0.125 µL of

TaqMan® SNP Genotyping Assay mix, 2.375 µL of sterile water and 1 µL of genomic DNA. Thermal conditions were based in the activation of Taq DNA Polimerase at 95°C for 10 minutes, followed by 45 cycles at 92°C for 15 seconds to denature DNA chain and 60°C for 1 minute to primers pairing and extension.

Amplification was detected and data analyzed through the StepOne Plus Real-Time PCR system and StepOne Software (version 2.3 Applied Biosystems). In order to ensure the quality of genotyping, two negative controls were included in each amplification reaction (to avoid false positives) and double sampling was performed in, at least, 10% of the samples, with an accuracy above 99%. The genotype results were independently evaluated by two researchers, who were blinded to patient clinical status.

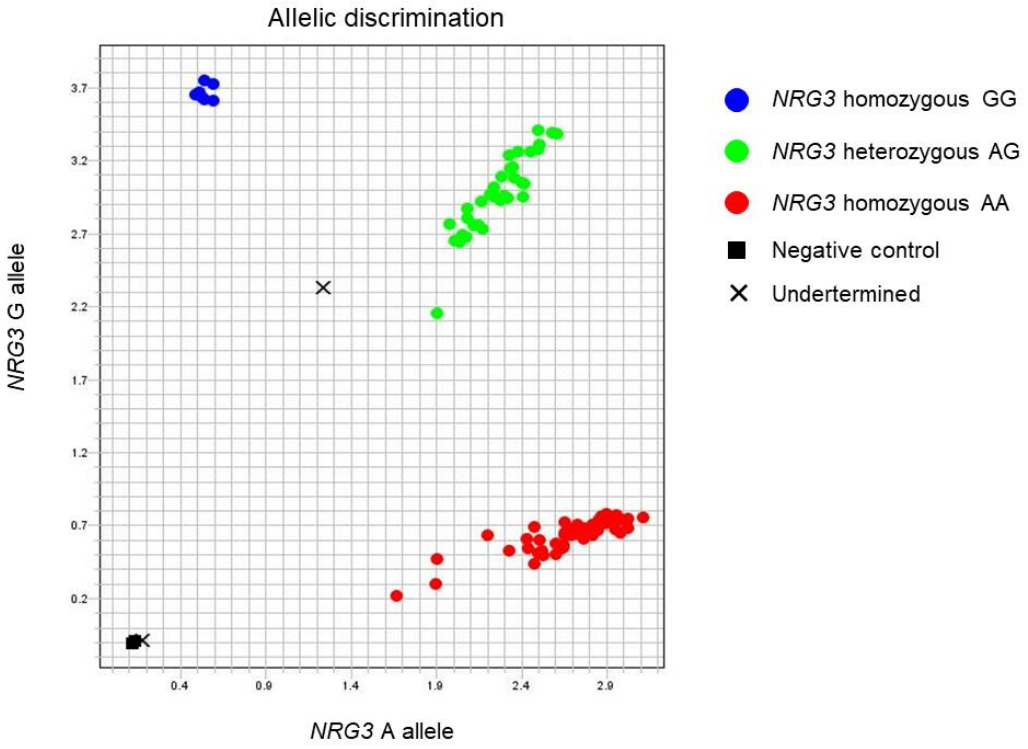


Figure 7 – Exemplification of an allelic discrimination plot for NRG3 rs1649942 polymorphism.

3.3. *In silico* analysis

In silico analysis was conducted to due to the lack of knowledge regarding the functional consequence of both intronic variants studied. Therefore, the Human Splicing Finder 3.0 was used to understand the possible influence of both SNPs in the respective

gene pre-mRNA splicing. Furthermore, to identify putative regulatory elements recognition binding sites created by the different alleles, the MotifMap online tool was also used.

3.4. Statistical analysis

Statistical analysis was performed resorting to the computer software IBM® SPSS® Statistics for Windows™ (version 24.0, SPSS Inc, 2016).

Associations between genetic polymorphisms and patients' clinicopathologic characteristics were assessed using chi-square test (χ^2), for categorical variables, whereas student's t-test was used for continuous variables (age).

Kaplan-Meier method was used to obtain survival curves and the log-rank test was used to compare the probabilities of survival. The most suitable genetic model for each variant was established after an initial comparison between Kaplan-Meier curves under the log-additive genetic model. Subgroup stratification was also performed according to FIGO stage (FIGO I/II vs FIGO III vs FIGO IV).

OS, defined as the interval of time between diagnosis and patients' death by EOC (EOC specific survival) or the last clinical evaluation, and disease-free survival (DFS), defined as the period from the date of diagnosis until the date of first recurrence or last clinical evaluation in patients with complete response to the first-line treatment, were the two clinical endpoints evaluated in this study. Endpoint definition was based on RECIST criteria [173].

The death and recurrence risk were estimated by a Cox proportional HR, along with 95% confidence interval (CI), adjusted for hormonal status (pre- vs post-menopausal), histologic subtype (serous vs others), surgery (complete vs others) and tumor grade (grade 1 vs grade 2 vs grade 3 vs grade 4). Cause of death was determined from the patients' medical records.

All tests were two-sided and a 5% level of significance was stablished.

4. Results

4.1. Descriptive statistics of *NRG3* rs1649942 and *BRE* rs7572644 polymorphisms

NRG3 rs1649942 genotype distribution (MAF, 28.3%) was 51.7% for homozygous AA (n=169), 40.0% (n=131) for heterozygous AG and 8.3% for homozygous GG (n=27) (4% of genotyping failure). Regarding the variant *BRE* rs7572644 (MAF, 21.1%), 63.2% of the patients were found to be homozygous for the T allele (n=208), 31.3% were heterozygous TC (n=103) and 5.5% homozygous for the C allele (n=18) (3% of genotyping failure).

Our data did not reveal significant statistical differences between the different genotypes of *NRG3* rs1649942 and *BRE* rs7572644 and patients clinicopathological characteristics, namely age (*P*=0.481 and *P*=0.968, respectively), FIGO stage (*P*=0.554 and *P*=0.402, respectively), histological subtype (*P*=0.060 and *P*=0.338, respectively), hormonal status (*P*=0.571 and *P*=0.271, respectively) and extent of residual disease (*P*=0.867 and *P*=0.863, respectively).

For the cohort involved in this study, the mean OS and DFS were 89.39 and 79.83 months, respectively.

4.2. Association of *NRG3* rs1649942 polymorphism with the clinical outcome of OC patients

Concerning the survival curves obtained using Kaplan-Meier method and log-rank test, no statistically significant differences were observed for the survival time of all patients according to *NRG3* rs1649942 genotypes (*P*=0.708), neither when considering a subgroup analysis restricted to early disease stage (FIGO I/II; *P*=0.644) or FIGO III patients (*P*=0.986). However, a significant impact of *NRG3* rs1649942 genotypes was observed for the FIGO IV stage patients (*P*=0.027). Considering a recessive genetic model (AA/AG genotypes vs GG genotype), no significant differences on survival were observed (*P*=0.699). Regarding the subgroup with distant metastasis at diagnosis, a consistent association was observed, as A allele carriers had a higher survival when compared to GG genotype carriers (*P*=0.011) (Figure 8). Namely, a mean OS time of 61.96 months was observed for patients with A allele that contrast with the 13.00 months reported for GG genotype patients. Anew, the protective effect of *NRG3* rs1649942 reference allele was not observed in early stages (FIGO I/II; *P*=0.377) or FIGO III stage patients (*P*=0.869).

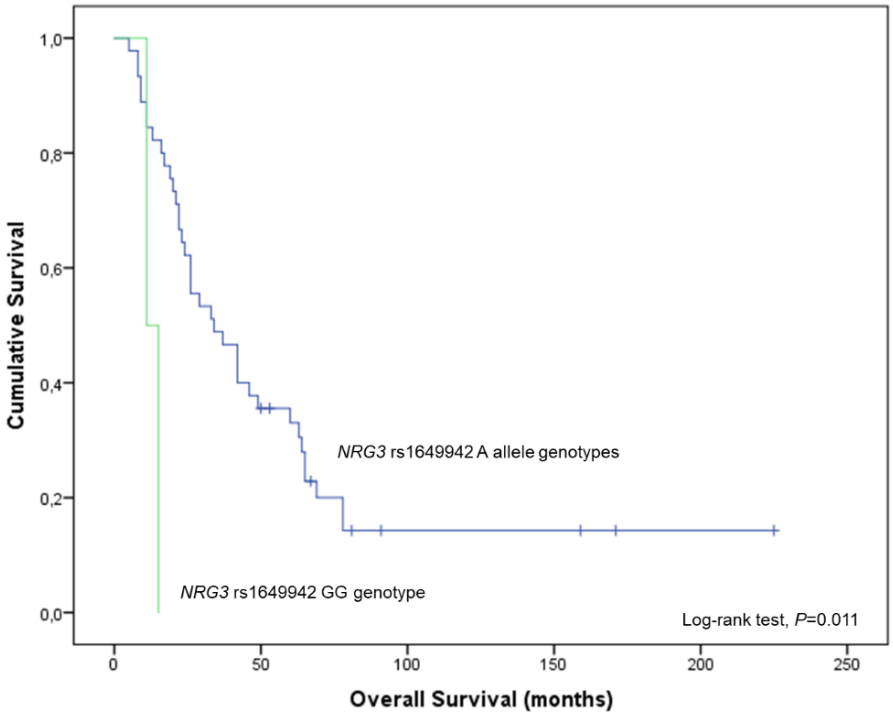


Figure 8 – Overall survival by Kaplan-Meier and log-rank test for the subgroup of EOC patients with FIGO IV stage disease at diagnosis, according to NRG3 rs1649942 polymorphism genotypes (recessive genetic model). The group of patients with A allele carrier genotypes (AA/AG genotypes) had significantly higher survival when compared with patients with GG genotype (*P*=0.011).

Regarding the impact of *NRG3* rs1649942 polymorphism with DFS for the entire cohort, no statistically significant associations were observed, considering either the log-additive and the recessive genetic model (*P*=0.356 and *P*=0.158, respectively). Additionally, no significant association was observed when the analysis was stratified by FIGO staging, independently of genetic model applied (data not showed). Furthermore, multivariate analyses for the risk of death and recurrence of EOC patients were calculated, adjusted for EOC prognostic variables (hormonal status, histological subtype, surgical extension and tumor grade). However, this analysis did not reveal a predictive impact of *NRG3* rs1649942 regarding the risk of death and recurrence of EOC patients (HR, 1.30; 95% CI, 0.63-2.67; *P*=0.477 and HR, 0.59; 95% CI, 0.24-1.45; *P*=0.251, respectively).

4.3. Association of *BRE* rs7572644 polymorphism with the clinical outcome of OC patients

Concerning the impact of *BRE* rs7572644 polymorphism on survival, no significant associations were observed either under a log-additive (TT vs TC vs CC; *P*=0.181) or

dominant genetic model analysis (CC/TC genotypes vs TT genotype; *P*=0.889). None the less, upon a stratified analysis, a statistically significant impact on survival time and *BRE* rs7572644 genotypes was observed in FIGO I/II and FIGO III stage patients (*P*=0.038 and *P*=0.040, respectively). No statistically significant difference was noticed for the later stage OC patients (*P*=0.139). Furthermore, under the dominant genetic model (CC/TC vs TT), we observed that TT homozygous genotype patients had an extended long-term survival of 39.53 months when compared to C allele carriers (214.46 vs 174.93 months; *P*=0.014), in the subgroup with early stage disease at diagnosis (FIGO I/II) (Figure 9). In opposite, for patients with FIGO III and IV stage, TT homozygous patients showed a shorter survival than those with CC/TC genotypes, even though these associations did not reach the significance level (mean survival time (months) for FIGO III stage patients: TT genotype = 100.52, C allele = 135.02; *P*=0.052; and FIGO IV stage patients: TT genotype = 44.99, C allele = 70.06; *P*=0.384, respectively).

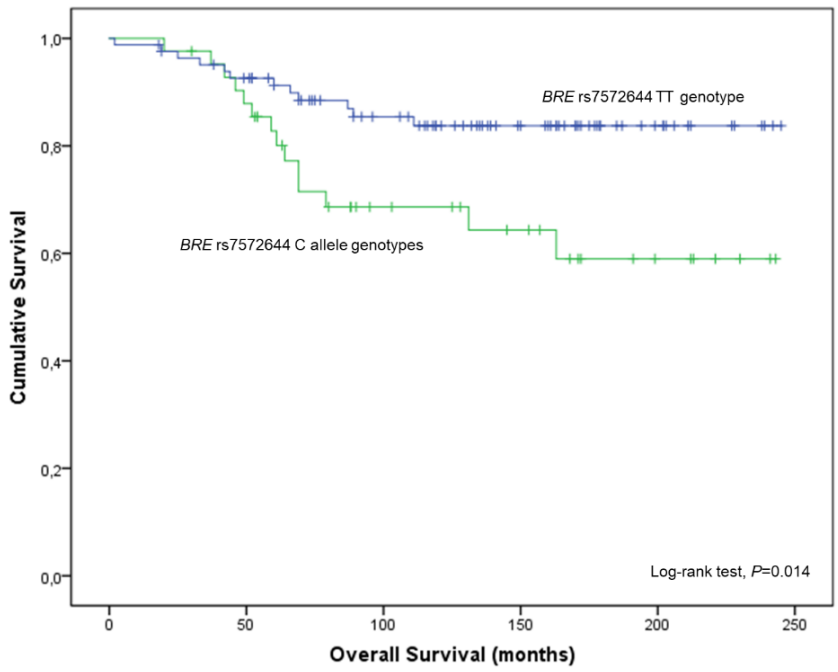


Figure 9 – Overall survival by Kaplan-Meier and log-rank test for the subgroup of EOC patients with FIGO I/II stage disease at diagnosis, according to BRE rs7572644 polymorphism genotypes (dominant genetic model). The group of patients with TT genotype had significantly higher survival when compared to C allele carrier genotype patients (*P*=0.014).

No statistically significant associations were observed considering the DFS as outcome, independently of the genetic model assumed (data not showed). However, for early stage patients, we observed that C allele carriers had a lower DFS time than TT genotype patients, with CC and TC genotype carriers presenting recurrence at 175.13

months and TT homozygous patients at 214.23 months (*P*=0.032) (Figure 10). No further statistical associations were observed for FIGO III (*P*=0.224) and FIGO IV stage patients (*P*=0.897).

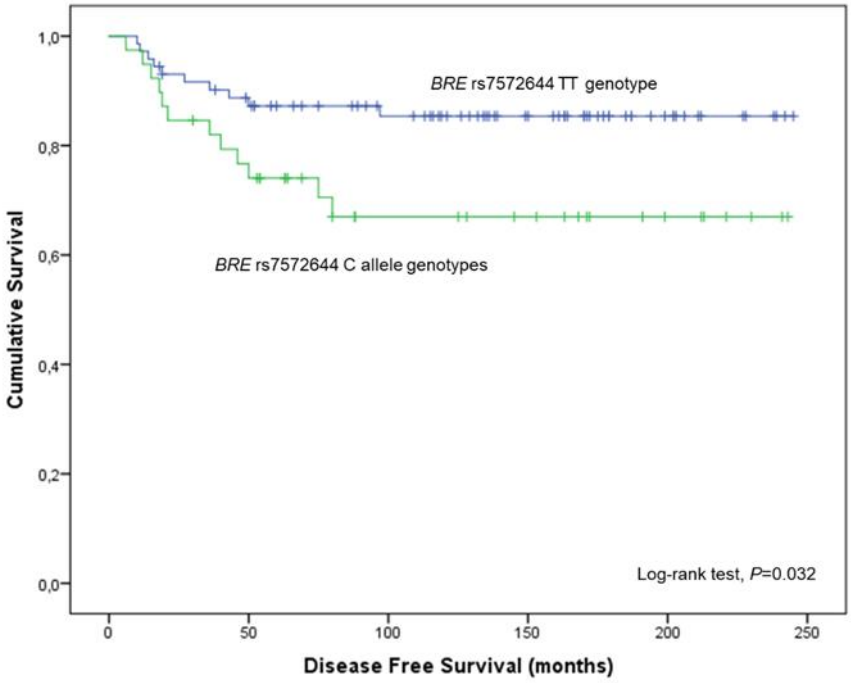


Figure 10 – Disease-free survival by Kaplan-Meier and log-rank test for the subgroup of EOC patients with FIGO I/II stage disease at diagnosis, according to BRE rs7572644 polymorphism genotypes (dominant genetic model). The group of patients with TT genotype had significantly higher survival when compared to C allele carrier genotype patients (*P*=0.032).

Furthermore, a multivariate analysis was performed to estimate the risk of death and recurrence of FIGO stages I/II patients, considering the *BRE* rs7572644 polymorphism and known EOC prognostic factors (hormonal status, histological subtype, surgical extension and tumor grade) (Table 3). We observed that C allele carriers had not only a threefold increased risk of disease recurrence but also death compared to TT homozygous genotype patients (HR, 3.33; 95% CI, 1.35-8.23; *P*=0.009 and HR, 3.09; 95% CI, 1.25-7.66; *P*=0.015, respectively) (Table 3).

Table 3 – Multivariate Cox regression analysis on the risk of recurrence and death in early stage patients (FIGO I/II) at diagnosis, considering several clinical and pathological variables.

	Risk of death			Risk of recurrence		
Variable	HR	95% CI	*P*-value	HR	95% CI	*P*-value
BRE rs7572644 (TT genotype vs C allele)	**3.09**	**1.25-7.66**	**0.015**	**3.33**	**1.35-8.23**	**0.009**
Hormonal status (Pre- vs Post-menopause)	1.17	0.98-1.39	0.086	**1.25**	**1.06-1.47**	**0.007**
Histology (serous vs others)	0.63	0.26-1.56	0.319	0.45	0.18-1.10	0.078
Surgery (complete vs other)	1.90	0.61-5.97	0.271	**3.37**	**1.21-9.36**	**0.020**
Tumor grade (Grade 1 vs 2 vs 3 vs 4)	1.05	0.91-1.21	0.489	1.00	0.86-1.15	0.946

Bold values are statistically significant.

5. Discussion

Besides OC low incidence, this gynecological cancer is a main factor of morbidity and mortality [118]. Namely, the high OC-associated lethality can be explained by the lack of specific and sensitive screening methods and by the anatomic location of this organ which allows the painless development of the tumor, both factors that preclude the early diagnosis of the disease [110, 111]. Although the implementation of platinum-based chemotherapy schemes has increased the response rates to first-line treatment up to 70%, the development of chemotherapy resistance by most patients remains a major hurdle, which also contributes for the slight percentage of individuals who survive 5 years after the initial diagnosis (around 30%) [120]. Thus, it becomes essential to identify biomarkers that contribute to the optimal selection of therapeutic schemes, namely by dose adjustment according to each individual's risk of relapse (risk stratification) [118].

The application of genome-wide strategies was faced with the potential to revolutionize the molecular oncology field. However, namely regarding OC, numerous results have been obtained by GWAS, although without reaching definitive conclusions. Therefore, it becomes essential to meticulously analyze these data to conclude the possible role of associated markers which, ultimately, could translate in clinical implementation [52]. To the best of our knowledge, five validation studies have been conducted in the attempt to validate OC GWAS associated markers, four of them evaluating susceptibility associated markers and the study conducted by Huang and colleagues being the only which attempted to validate clinical outcome associated variants [136, 137, 140, 141, 167]. Thus, it is evident the need to validate clinical outcome associated markers, not only because of the lack of this type of studies, but also due to the potential clinical applicability arising from truly causal variants. Furthermore, only the study conducted by Mostowska *et al*. was independently performed [167]. Thus, the present study was the first designed with the purpose to independently validate OC clinical outcome GWAS findings in a Caucasian population.

5.1. Association of *NRG3* rs1649942 polymorphism with the clinical outcome of OC patients

The genetic variant rs1649942 corresponds to an intronic variation that leads to the replacement of an adenine (A) by a guanine (G) in the *NRG3* gene [174]. This gene codifies a homonymous extracellular protein, being a member of the neuregulin family, which has been shown to have low expression levels in normal ovarian tissues [175]. This protein has the capacity to bind exclusively to ErbB4 and, so, to be involved in ErbB signaling, which, in turn, seems to be related to ovarian tumorigenesis [136, 175-179] (Figure 11).

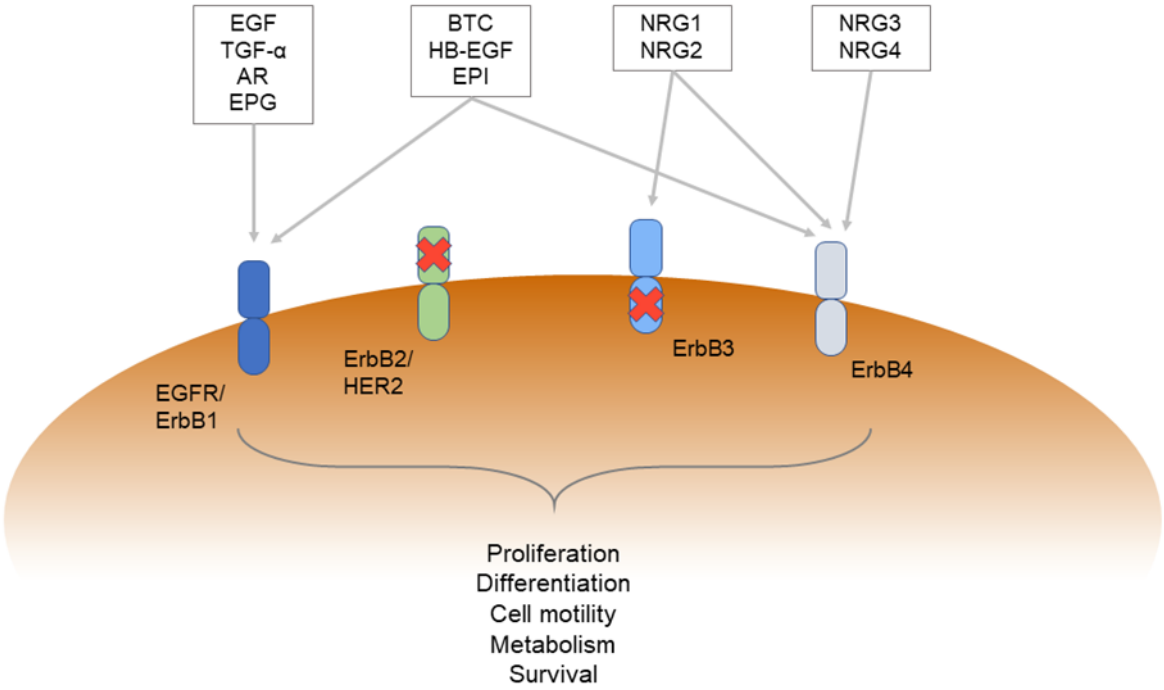


Figure 11 – Specificity of ligands to Epidermal Growth Factor Receptor (EGFR) family. Epidermal Growth Factor (EGF), Transforming Growth Factor-α (TGF-α), Amphiregulin (AR) and Epigen (EPG) are specific ligands for EGFR; Betacellulin (BTC), Heparin-binding EGF (HB-EGF) and Epiregulin (EPI) bind both EGFR and ErbB4; Neuregulin 1/2 (NRG1/2) have the capacity to bind ErbB3/4; Neuregulin 3/4 (NRG3/4) are exclusive ligands for ErbB4. No ligand is known to bind ErbB2, although it has the strongest kinase activity. Thus, it is the preferential heterodimerization partner of other ErbB receptors. In contrast, ErB3 lacks its kinase activity, so it needs to heterodimerize with other partners to have signaling potential (adapted from [180]).

This SNP was firstly identified to be associated with response to cytotoxic drugs in a study conducted by Huang *et al.* (2007) that aimed to discover genetic variants that contribute to cisplatin-induced cytotoxicity, concluding that rs1649942 was significantly associated with cisplatin IC_{50} [132]. Despite *NRG3* rs1649942 was first reported a decade ago it still is not well characterized [132]. Interestingly, although the most part of intronic variants lies within consensus regions of exon-intron junctions and directly affect splicing, some of them might lie in regions with a regulatory role in the maturation of pre-mRNA [181, 182]. Additionally, polymorphisms lying within regulatory regions of the genome may lead not only to alterations in TFs binding motifs, but also influence methylation patterns and the alternative splicing of target genes and, hence, affecting their transcription [183-185]. The *in silico* analyses' results provided by the Human Splicing Finder 3.0 and MotifMap bioinformatic tools revealed that the rs1649942 genetic variant is unlikely to be in a splicing or in a TF binding site. Nevertheless, this variant has already been shown to be associated with the expression levels of target genes, which suggest that its functional consequence

might be due to a regulatory network role [132, 136]. Effectively, some regions of the genome harbor genetic alterations capable of regulate the expression of near (cis-regulation) or distant genes (trans-regulation), being defined as expression quantitative trait loci (eQTL) [183]. Namely, it has been shown that these regulatory properties appear to clarify the role of a large proportion of non-coding variants [186].

Although the GWAS data was not validated in the overall analysis, our results indicate that *NRG3* rs1649942 polymorphism was significantly associated with OS in the subgroup of FIGO IV stage patients, either considering a log-additive (*P*=0.027) and a recessive genetic model analysis (*P*=0.011*).* Therefore, in the subgroup of individuals with advanced disease at diagnosis, a lower survival time was observed for patients carrying the GG genotype (13.00 months), in contrast to A allele carriers who presented a better OS (61.96 months). However, no impact on the time until tumor recurrence was noticed.

Huang *et al*. [136] reported, for the first time, the association of *NRG3* rs1649942 variant with the clinical outcome of OC patients, submitted to the first-line treatment. An association between this variant and PFS was observed for all the patients (*P*=0.008), being even more marked in the subgroup of optimal debulked patients (*P*=0.002), as GG genotype carriers presented a lower survival time when compared with AA/AG genotype patients. Thus, the negative impact assigned to GG genotype is consistently observed in our study. Furthermore, considering OS analysis, Huang and colleagues also reported a reduced survival time for GG homozygous patients compared with AA/AG genotype patients (*P*=0.014), although this association was only observed in the subset of optimally debulked patients. Moreover, the regulatory role of *NRG3* rs1649942 was demonstrated, namely by being associated with the expression levels of several genes, some of them associated with carboplatin IC_{50} [132, 136, 187]. One of these regulated genes, *KYNU*, known to be involved in tryptophan metabolism, was already reported to impact the OS of advanced stage serous OC patients submitted to the standard treatment and its overexpression was associated with an increase in the cellular sensitivity to carboplatin [136, 188, 189]. Taking together, these results suggest a regulatory role as the main functional consequence of this SNP.

In advanced disease stages, the metastatic process assumes a great preponderance being responsible for the high mortality rates [190]. Namely, the dissemination to the peritoneal cavity is an early phenomenon in the natural history of the ovarian disease since the malignant cells follow the peritoneal fluid, obeying to the intra-abdominal pressure variations, being able to avoid anoikis and, hence, having the ability to survive in suspension. The molecular mechanisms implied in OC cell dissemination are not fully described, though several biological pathways might be involved, with distinct gene expression profiles between tumor stages [110, 191-193]. Considering the previous

assumptions, we hypothesized that *NRG3* rs1649942 GG genotype might lead to the expression of peripheral genes that promote the acquisition of an aggressive phenotype, namely with a pro-oncogenic role in the metastatic niche. Consequently, GG genotype patients with distant metastatic disease at diagnosis have a lower survival time compared with A allele carriers. However, for less spread disease stages, the expression of core genes involved in the metastatic process might have a more direct and preponderant role when compared to the complex regulatory network in which the *NRG3* gene might be involved. In fact, neuregulins have already been associated with metastatic process, namely in medulloblastoma, breast and papillary thyroid tumors [194, 195]. Regarding OC, it was reported the relevant contribution of the interaction Neuregulin 1 (NRG1)-ErbB3 in omentun metastasis via a hematogenous route [196]. Consequently, the quantification of NRG3 circulatory levels might be considered in future studies. Despite the association with OS, the impact of this variant in the DFS time appears to be irrelevant, which might indicate a preponderant role of further clinical and pathological factors in the time until EOC recurrence.

5.2 Association of *BRE* rs7572644 polymorphism with the clinical outcome of OC patients

The polymorphism rs7572644 lies within a intronic region of the *BRE* gene and leads to the substitution in the DNA chain of a thymine (T) by a cytosine (C) [145, 174]. *BRE*, also known as *BABAM2* (BRISC and BRCA1 A complex member 2), codifies a protein involved in DNA damage response [145] (Figure 12). Namely, it is involved in the maintenance of the integrity of the BRCA1-A complex in the nucleus, although BRE could also act as an anti-apoptotic protein [145, 197, 198]. In the ovary, BRE deficiency was found to promote follicular atresia, through the enhancement of granulosa cells' apoptosis [199]. Regarding OC, some evidences point a possible role of BRE in disease susceptibility and chemotherapy response [145, 200]. As the functional impact of rs7572644 is not fully described yet, in *silico* analyses revealed a possible null impact in splicing mechanisms as well as no predicted TF binding site. However, the minor allele of this variant was shown to be associated with chemotherapy sensitivity [145].

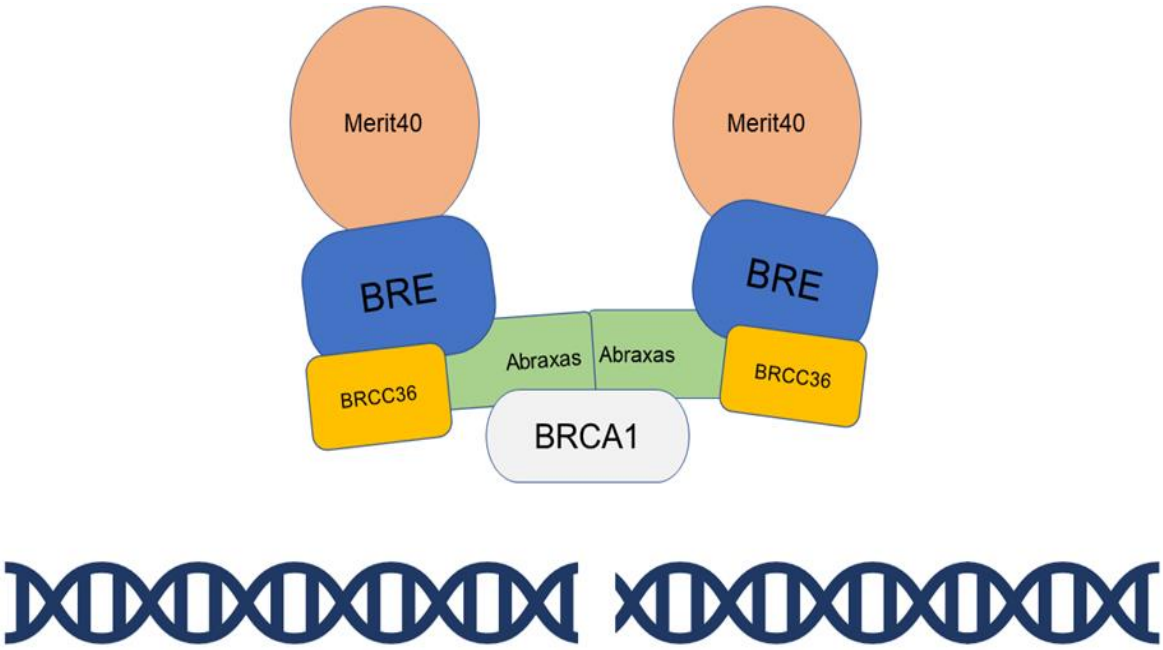


Figure 12 – Schematic representation of BRCA1-A complex. This DNA repair complex is structurally constituted by a dimer of heterotetramers composed by MERIT40, Abraxas, BRCC36 and BRE (adapted from [201]).

None significant results were observed in the overall cohort, according to the *BRE* rs7572644 genotypes. Upon subgroup analysis, we observed that TT genotype patients had a prolonged OS when compared to C allele carriers (dominant genetic model), for the subgroup of patients with early stage disease (FIGO I/II) at diagnosis (*P*=0.014). Moreover, despite the non-significant result for FIGO stage III subgroup (*P*=0.052), there is a trend for the impact of this variant in the survival of these subset of individuals. Moreover, in the subgroup of FIGO stages I/II patients, TT genotype was associated with improved DFS (*P*=0.032), meaning that C allele carrier patients would take less time until the emergence of recurrent disease. Additionally, using multivariate Cox regression models, and exclusively evaluating patients with localized tumor at diagnosis, we verified that C allele carriers have a three-fold increase in the risk of death and recurrence, adjusted for known OC prognostic factors. Namely, this polymorphism is the only variable significantly associated with the risk of death of this subset of patients. Thus, *BRE* rs7572644 polymorphism could be a useful predictive marker for early stage disease patients.

Considering the role of BRE in DNA damage repair, the presence of *BRE* rs7572644 C allele could lead to the translation of a more stable isoform with higher affinity for the BRCA1-A complex. Consequently, there might be an increase in the complex integrity which can promote an improved DNA damage repair response. Although early stage disease patients have tumor confined to ovaries (FIGO stage I) or with pelvic extension/primary peritoneal cancer (FIGO stage II), patients are submitted to platinum-based chemotherapy,

after cytoreductive surgery, meaning that cancer cells that have not been completely eradicated by surgery are exposed to the action of DNA damage agents. Considering the presence of *BRE* rs7572644 C allele, and hence an improved DNA repair damage response, residual cancer cells could promptly repair the platinum-induced damages, which could result in a precocious and higher risk of disease relapse. Moreover, BRE was also found to be an anti-apoptotic protein [198]. Namely, BRE was considered to be involved in lung cancer cisplatin resistance through its anti-apoptotic activity mediated by the protein kinase B signaling pathway [202]. Thereby, beyond the possible impact in DNA repair, we further hypothesized that *BRE* rs7572644 C allele is associated with an increased anti-apoptotic activity of EOC cells, promoting a decrease in the cellular sensitivity to chemotherapy. Supporting this assumption is the fact that PI3K/Akt signaling pathway (in which protein kinase B has a major effector role) activation has already been associated with cisplatin resistance in OC [202]. In agreement to what was hypothesized for *NRG3* rs1649942 polymorphism, *BRE* rs7572644 could also exert its effect through the regulation of target genes. Despite *in silico* analyses have not predicted the binding of any TF at this locus, this putative eQTL SNP could act through several other mechanisms, as mentioned before [183-185]. Therefore, the present variant could be potentially associated with the expression of peripheral genes with an additional negative role in the prognosis of OC patients. Moreover, this SNP could not be the functional variant itself, as it is in linkage disequilibrium (r^2>0.8) with other four intronic variants (rs7581813, rs2337700, rs55796876 and rs11691385) which could be causal of the disease.

On the other hand, the influence of *BRE* rs7572644 does not appear to be relevant in FIGO III and IV stage patients, although a trend was seen for FIGO III stage patients. Besides the undoubtedly importance of DNA repair mechanisms in OC susceptibility and treatment response, namely for the newly targeted therapies, its role in the metastatic process might not be so determinant, being that function fulfilled through additional and more relevant molecular networks [110]. In this sense, as disease dissemination increases less relevant might be the role of BRE in the maintenance of genomic integrity. Accordingly, BRE was reported to promote tumor cell growth but not metastasis of mouse Lewis lung carcinoma cells transfected with this protein in syngeneic mice models [203].

Variant *BRE* rs7572644 was firstly reported by Fridley *et al*., which used a cell-based model to associate drug response phenotypes with genetic variants [145]. The influence of each identified variant with any clinical endpoint was not assessed in the original study, although it was observed that *BRE* rs7572644 C allele carriers were sensitive to the carboplatin and paclitaxel combination. Therefore, intuitively, patients who are sensitive to chemotherapy might have an improved outcome. However, the data obtained in the present study appear to be in disagreement with the chemosensitive properties conferred by C allele

in the study of Fridley *et al*. [145]. Nevertheless, the results obtained in LCLs GWAS should be taken carefully as one of the limitations inherent to this approach is that it does not consider interactions between several biological pathways in response to chemotherapy, which can be a possible bias as it not fully represents the complexity of drug effect [14, 136, 145].

6. Conclusions and future perspectives

GWAS development will aid to rethink OC genomics much beyond the obvious and direct analysis. Namely, multiple small effect genetic variants, most of them lying in regulatory regions of the genome and potentially irrelevant to a particular disease, might indirectly exert its influence on trait phenotypes through the regulation of complex networks, affecting thereby the expression and activity of a restrict core set of genes [30]. This theory seems to be consistent with the results obtained for OC GWAS, as most of the associated variants are in regulatory sequences of genes with an unapparent direct role in this gynecological neoplasia. The combination of genomic knowledge might be the key to unravel OC genetics and further work is needed to underpin this assumption.

In this perspective, the present study aimed to validate two intronic variants associated with the clinical outcome of OC patients previously reported by GWAS. In fact, this is the first study which attempts to confirm the association of GWAS-identified variants with the clinical outcome of EOC patients, in an independent cohort, namely, in a Portuguese population. In sum, we were able to independently validate both *NRG3* rs1649942 and *BRE* rs7572644 variants in our population, as they seemed to have a role in the clinical outcome.

In our population, the two genetic polymorphisms previously highlighted in GWAS presented a FIGO staging-specific behavior. Although promising, these results, particularly regarding *NRG3* rs1649942 variant, should be taken carefully due to the under-power in FIGO stage IV A allele carriers. Though underpowered in the stratified analysis, this study represents one of the largest series of OC patients published so far. Furthermore, clinical-pathological characteristics of included patients are similar to the entire series admitted in our institution, meaning that we can accept the representativeness of our cohort. Moreover, a low number of individuals was lost to follow-up (attrition bias of 5%), although the extended period of follow-up time. Furthermore, the low range of 95% CIs determined in the multivariate analysis could demonstrate the low variance in the concluded outcomes.

We suggest that further validation studies should be conducted in larger cohorts in order to reinforce the study power. Additionally, the need for additional post-GWAS analyses is evident, in order to perceive the genetic context in which each variant is inserted and to understand their biological plausibility regarding OC. To achieve this purpose, future analysis should be oriented to fine map the regions where these variants lie, as well as for the accomplishment of functional studies. Further studies evaluating the regulatory network of each associated region should be conducted resorting to the analysis of chromatin markers which tag promoters or enhancers/silencers [144]. Additionally, eQTL studies might be informative of the influence of these polymorphisms in the expression levels of several genes and would contribute to the linking between identified variants and their target genes [158].

The success of GWAS suggests that this genomic approach will continue to be applied for the assessment of variants with probable impact in complex traits. However, their development should be complemented with post-GWAS analyses, in order to identify and confirm the most significant associated variants and to understand their potential biological involvement [10, 20, 157, 158]. Ultimately, GWAS findings might be of potential interest for clinical practice, in the era of personalized medicine, since some variants identified through these studies might be important independent prognostic markers or assume a predictive role of therapy response and, consequently, help in the adoption of treatment options suited to individual genetic profile.

7. References

1. Sachidanandam, R., D. Weissman, S.C. Schmidt, et al., *A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.* Nature, 2001. **409**(6822): p. 928-33.

2. Collins, F.S., L.D. Brooks, and A. Chakravarti, *A DNA polymorphism discovery resource for research on human genetic variation.* Genome Res, 1998. **8**(12): p. 1229-31.

3. Brookes, A.J., *The essence of SNPs.* Gene, 1999.

4. Shastry, B.S., *SNP alleles in human disease and evolution.* J Hum Genet, 2002. **47**(11): p. 561-6.

5. Monzo, M., A. Navarro, G. Ferrer, et al., *Pharmacogenomics: a tool for improving cancer chemotherapy.* Clin Transl Oncol, 2008. **10**(10): p. 628-37.

6. Bush, W.S. and J.H. Moore, *Chapter 11: Genome-wide association studies.* PLoS Comput Biol, 2012. **8**(12): p. e1002822.

7. Visscher, P.M., M.A. Brown, M.I. McCarthy, et al., *Five years of GWAS discovery.* Am J Hum Genet, 2012. **90**(1): p. 7-24.

8. Stadler, Z.K., P. Thom, M.E. Robson, et al., *Genome-wide association studies of cancer.* J Clin Oncol, 2010. **28**(27): p. 4255-67.

9. Klein, R.J., C. Zeiss, E.Y. Chew, et al., *Complement factor H polymorphism in age-related macular degeneration.* Science, 2005. **308**(5720): p. 385-9.

10. Manolio, T.A., *Genomewide association studies and assessment of the risk of disease.* N Engl J Med, 2010. **363**(2): p. 166-76.

11. Patel, J.N., H.L. McLeod, and F. Innocenti, *Implications of genome-wide association studies in cancer therapeutics.* Br J Clin Pharmacol, 2013. **76**(3): p. 370-80.

12. Manolio, T.A., L.D. Brooks, and F.S. Collins, *A HapMap harvest of insights into the genetics of common disease.* J Clin Invest, 2008. **118**(5): p. 1590-605.

13. Berndt, S.I., S. Gustafsson, R. Magi, et al., *Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture.* Nat Genet, 2013. **45**(5): p. 501-12.

14. Wheeler, H.E., M.L. Maitland, M.E. Dolan, et al., *Cancer pharmacogenomics: strategies and challenges.* Nat Rev Genet, 2013. **14**(1): p. 23-34.

15. Gabriel, S.B., S.F. Schaffner, H. Nguyen, et al., *The structure of haplotype blocks in the human genome.* Science, 2002. **296**(5576): p. 2225-9.

16. Wang, W.Y., B.J. Barratt, D.G. Clayton, et al., *Genome-wide association studies: theoretical and practical concerns.* Nat Rev Genet, 2005. **6**(2): p. 109-18.

17. Ku, C.S., E.Y. Loy, Y. Pawitan, et al., *The pursuit of genome-wide association studies: where are we now?* J Hum Genet, 2010. **55**(4): p. 195-206.

18. Hindorff, L.A., P. Sethupathy, H.A. Junkins, et al., *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9362-7.

19. Motsinger-Reif, A.A., E. Jorgenson, M.V. Relling, et al., *Genome-wide association studies in pharmacogenomics: successes and lessons.* Pharmacogenet Genomics, 2013. **23**(8): p. 383-94.

20. Manolio, T.A., F.S. Collins, N.J. Cox, et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.

21. Goldstein, D.B., *Common genetic variation and human traits.* N Engl J Med, 2009. **360**(17): p. 1696-8.

22. Altshuler, D., M.J. Daly, and E.S. Lander, *Genetic mapping in human disease.* Science, 2008. **322**(5903): p. 881-8.

23. Low, S.K., A. Takahashi, T. Mushiroda, et al., *Genome-wide association study: a useful tool to identify common genetic variants associated with drug toxicity and efficacy in cancer pharmacogenomics.* Clin Cancer Res, 2014. **20**(10): p. 2541-52.

24. Yang, I.A., J.W. Holloway, and K.M. Fong, *Genetic susceptibility to lung cancer and co-morbidities.* J Thorac Dis, 2013. **5 Suppl 5**: p. S454-62.

25. Peng, S., B. Lu, W. Ruan, et al., *Genetic polymorphisms and breast cancer risk: evidence from meta-analyses, pooled analyses, and genome-wide association studies.* Breast Cancer Res Treat, 2011. **127**(2): p. 309-24.

26. Ishak, M.B. and V.N. Giri, *A systematic review of replication studies of prostate cancer susceptibility genetic variants in high-risk men originally identified from genome-wide association studies.* Cancer Epidemiol Biomarkers Prev, 2011. **20**(8): p. 1599-610.

27. Yang, J., A. Bakshi, Z. Zhu, et al., *Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index.* Nat Genet, 2015. **47**(10): p. 1114-20.

28. Bourrat, P., Q. Lu, and E. Jablonka, *Why the missing heritability might not be in the DNA.* Bioessays, 2017. **39**(7).

29. Jablonka, E. and G. Raz, *Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution.* Q Rev Biol, 2009. **84**(2): p. 131-76.

30. Boyle, E.A., Y.I. Li, and J.K. Pritchard, *An Expanded View of Complex Traits: From Polygenic to Omnigenic.* Cell, 2017. **169**(7): p. 1177-1186.

31. Chanock, S.J., T. Manolio, M. Boehnke, et al., *Replicating genotype-phenotype associations.* Nature, 2007. **447**(7145): p. 655-60.

32. Ioannidis, J.P., G. Thomas, and M.J. Daly, *Validating, augmenting and refining genome-wide association signals.* Nat Rev Genet, 2009. **10**(5): p. 318-29.

33. Zeng, P., Y. Zhao, C. Qian, et al., *Statistical analysis for genome-wide association study.* J Biomed Res, 2015. **29**(4): p. 285-97.

34. Liu, J.Z., A.F. McRae, D.R. Nyholt, et al., *A versatile gene-based test for genome-wide association studies.* Am J Hum Genet, 2010. **87**(1): p. 139-45.

35. Che, R., J.R. Jack, A.A. Motsinger-Reif, et al., *An adaptive permutation approach for genome-wide association study: evaluation and recommendations for use.* BioData Min, 2014. **7**: p. 9.

36. Todd, J.A., N.M. Walker, J.D. Cooper, et al., *Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes.* Nat Genet, 2007. **39**(7): p. 857-64.

37. Hakonarson, H., S.F. Grant, J.P. Bradfield, et al., *A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene.* Nature, 2007. **448**(7153): p. 591-4.

38. Sladek, R., G. Rocheleau, J. Rung, et al., *A genome-wide association study identifies novel risk loci for type 2 diabetes.* Nature, 2007. **445**(7130): p. 881-5.

39. Zeggini, E., M.N. Weedon, C.M. Lindgren, et al., *Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.* Science, 2007. **316**(5829): p. 1336-41.

40. Scott, L.J., K.L. Mohlke, L.L. Bonnycastle, et al., *A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.* Science, 2007. **316**(5829): p. 1341-5.

41. Saxena, R., B.F. Voight, V. Lyssenko, et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* Science, 2007. **316**(5829): p. 1331-6.

42. Steinthorsdottir, V., G. Thorleifsson, I. Reynisdottir, et al., *A variant in CDKAL1 influences insulin response and risk of type 2 diabetes.* Nat Genet, 2007. **39**(6): p. 770-5.

43. Zeggini, E., L.J. Scott, R. Saxena, et al., *Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes.* Nat Genet, 2008. **40**(5): p. 638-45.

44. Real, L.M., A. Ruiz, J. Gayan, et al., *A colorectal cancer susceptibility new variant at 4q26 in the Spanish population identified by genome-wide association analysis.* PLoS One, 2014. **9**(6): p. e101178.

45. Xu, W., J. Xu, K. Shestopaloff, et al., *A genome wide association study on Newfoundland colorectal cancer patients' survival outcomes.* Biomark Res, 2015. **3**: p. 6.

46. Lan, Q., C.A. Hsiung, K. Matsuo, et al., *Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia.* Nat Genet, 2012. **44**(12): p. 1330-5.

47. Hu, L., C. Wu, X. Zhao, et al., *Genome-wide association study of prognosis in advanced non-small cell lung cancer patients receiving platinum-based chemotherapy.* Clin Cancer Res, 2012. **18**(19): p. 5507-14.

48. Shi, Y., Z. Hu, C. Wu, et al., *A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1.* Nat Genet, 2011. **43**(12): p. 1215-8.

49. Willis, J.A., S.H. Olson, I. Orlow, et al., *A replication study and genome-wide scan of single-nucleotide polymorphisms associated with pancreatic cancer risk and overall survival.* Clin Cancer Res, 2012. **18**(14): p. 3942-51.

50. Matsuda, K., A. Takahashi, C.D. Middlebrooks, et al., *Genome-wide association study identified SNP on 15q24 associated with bladder cancer risk in Japanese population.* Hum Mol Genet, 2015. **24**(4): p. 1177-84.

51. [cited 2017 6th September]; Available from: https://www.ebi.ac.uk/gwas/.

52. Manolio, T.A., *Bringing genome-wide association findings into clinical use.* Nat Rev Genet, 2013. **14**(8): p. 549-58.

53. Polychronakos, C. and Q. Li, *Understanding type 1 diabetes through genetics: advances and prospects.* Nat Rev Genet, 2011. **12**(11): p. 781-92.

54. Salloum, N.C., M.J. McCarthy, S.G. Leckband, et al., *Towards the clinical implementation of pharmacogenetics in bipolar disorder.* BMC Med, 2014. **12**: p. 90.

55. Roden, D.M., R.B. Altman, N.L. Benowitz, et al., *Pharmacogenomics: challenges and opportunities.* Ann Intern Med, 2006. **145**(10): p. 749-57.

56. Azzato, E.M., P.D. Pharoah, P. Harrington, et al., *A genome-wide association study of prognosis in breast cancer.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(4): p. 1140-3.

57. Kiyotani, K., T. Mushiroda, T. Tsunoda, et al., *A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese.* Hum Mol Genet, 2012. **21**(7): p. 1665-72.

58. Rafiq, S., W. Tapper, A. Collins, et al., *Identification of inherited genetic variations influencing prognosis in early-onset breast cancer.* Cancer Res, 2013. **73**(6): p. 1883-91.

59. Shu, X.O., J. Long, W. Lu, et al., *Novel genetic markers of breast cancer survival identified by a genome-wide association study.* Cancer Res, 2012. **72**(5): p. 1182-9.

60. Tan, X.L., A.M. Moyer, B.L. Fridley, et al., *Genetic variation predicting cisplatin cytotoxicity associated with overall survival in lung cancer patients receiving platinum-based chemotherapy.* Clin Cancer Res, 2011. **17**(17): p. 5801-11.

61. Wu, C., B. Xu, P. Yuan, et al., *Genome-wide interrogation identifies YAP1 variants associated with survival of small-cell lung cancer patients.* Cancer Res, 2010. **70**(23): p. 9721-9.

62. Wu, X., Y. Ye, R. Rosell, et al., *Genome-wide association study of survival in non-small cell lung cancer patients receiving platinum-based chemotherapy.* J Natl Cancer Inst, 2011. **103**(10): p. 817-25.

63. Yang, J.J., C. Cheng, W. Yang, et al., *Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia.* Jama, 2009. **301**(4): p. 393-403.

64. Ni, X., W. Zhang, and R.S. Huang, *Pharmacogenomics discovery and implementation in genome-wide association studies era.* Wiley Interdiscip Rev Syst Biol Med, 2013. **5**(1): p. 1-9.

65. Ferlay, J., I. Soerjomataram, R. Dikshit, et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.

66. Torre, L.A., F. Bray, R.L. Siegel, et al., *Global cancer statistics, 2012.* CA Cancer J Clin, 2015. **65**(2): p. 87-108.

67. Bray, F., A. Jemal, N. Grey, et al., *Global cancer transitions according to the Human Development Index (2008-2030): a population-based study.* Lancet Oncol, 2012. **13**(8): p. 790-801.

68. Du, W. and O. Elemento, *Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies.* Oncogene, 2015. **34**(25): p. 3215-25.

69. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.

70. Tan, D.S., T.S. Mok, and T.R. Rebbeck, *Cancer Genomics: Diversity and Disparity Across Ethnicity and Geography.* J Clin Oncol, 2016. **34**(1): p. 91-101.

71. Vogelstein, B., N. Papadopoulos, V.E. Velculescu, et al., *Cancer genome landscapes.* Science, 2013. **339**(6127): p. 1546-58.

72. Ruddon, R.W., *Cancer Biology*. 4th ed. 2007: Oxford University Press.

73. Martinez, J.D., M.T. Parker, K.E. Fultz, et al., *Molecular biology of cancer.*, in *Burger's Medicinal Chemistry and Drug Discovery*, D.J. Abraham, Editor. 2003, John Wiley & Sons, Inc.

74. Kumar, V., A.K. Abbas, and J.C. Aster, *Robbins and Cotran Pathologic Basis of Disease*. 9th ed. 2015: Saunders/Elsevier.

75. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

76. Hui, L. and Y. Chen, *Tumor microenvironment: Sanctuary of the devil.* Cancer Lett, 2015. **368**(1): p. 7-13.

77. Greaves, M. and C.C. Maley, *Clonal evolution in cancer.* Nature, 2012. **481**(7381): p. 306-13.

78. Heng, H.H., S.W. Bremer, J.B. Stevens, et al., *Chromosomal instability (CIN): what it is and why it is crucial to cancer evolution.* Cancer Metastasis Rev, 2013. **32**(3-4): p. 325-40.

79. Heng, H.H., J.B. Stevens, S.W. Bremer, et al., *Evolutionary mechanisms and diversity in cancer.* Adv Cancer Res, 2011. **112**: p. 217-53.

80. Stevens, J.B., S.D. Horne, B.Y. Abdallah, et al., *Chromosomal instability and transcriptome dynamics in cancer.* Cancer Metastasis Rev, 2013. **32**(3-4): p. 391-402.

81. Ferlay, J., E. Steliarova-Foucher, J. Lortet-Tieulent, et al., *Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012.* Eur J Cancer, 2013. **49**(6): p. 1374-403.

82. Kurman, R.J., M. Carcangiu, C. Herrington, et al., *WHO Classification of Tumors of Female Reproductive Organs*. 4th ed. WHO Classification of Tumors. Vol. 6. 2014.

83. Levanon, K., C. Crum, and R. Drapkin, *New insights into the pathogenesis of serous ovarian cancer and its clinical impact.* J Clin Oncol, 2008. **26**(32): p. 5284-93.

84. Matulonis, U.A., A.K. Sood, L. Fallowfield, et al., *Ovarian cancer.* Nat Rev Dis Primers, 2016. **2**: p. 16061.

85. Devouassoux-Shisheboran, M. and C. Genestie, *Pathobiology of ovarian carcinomas.* Chin J Cancer, 2015. **34**(1): p. 50-5.

86. Prat, J. and F.C.o.G. Oncology, *Staging Classification for Cancer of the Ovary, Fallopian Tube, and Peritoneum: Abridged Republication of Guidelines From the International Federation of Gynecology and Obstetrics (FIGO).* Obstet Gynecol, 2015. **126**(1): p. 171-4.

87. Ng, J.S., J.J. Low, and A. Ilancheran, *Epithelial ovarian cancer.* Best Pract Res Clin Obstet Gynaecol, 2012. **26**(3): p. 337-45.

88. Roque, R., A. Pina, C. Soares, et al., *[Splenic metastasis in ovary serous adenocarcinoma].* Acta Med Port, 2007. **20**(6): p. 581-6.

89. Gallagher, D.J., J.A. Konner, K.M. Bell-McGuinn, et al., *Survival in epithelial ovarian cancer: a multivariate analysis incorporating BRCA mutation status and platinum sensitivity.* Ann Oncol, 2011. **22**(5): p. 1127-32.

90. Cramer, D.W., *The epidemiology of endometrial and ovarian cancer.* Hematol Oncol Clin North Am, 2012. **26**(1): p. 1-12.

91. Permuth-Wey, J. and T.A. Sellers, *Epidemiology of ovarian cancer.* Methods Mol Biol, 2009. **472**: p. 413-37.

92. Assis, J., D. Pereira, and R. Medeiros, *Ovarian cancer and DNA repair: DNA ligase IV as a potential key.* World J Clin Oncol, 2013. **4**(1): p. 14-24.

93. Whittemore, A.S., R. Harris, and J. Itnyre, *Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case-control studies. IV. The pathogenesis of epithelial ovarian cancer. Collaborative Ovarian Cancer Group.* Am J Epidemiol, 1992. **136**(10): p. 1212-20.

94. Koch, M., *Re: "Events of Reproductive Life and the Incidence of Epithelial Ovarian Cancer".* Am J Epidemiol, 1989. **129**(4): p. 861-3.

95. Hildreth, N.G., J.L. Kelsey, V.A. LiVolsi, et al., *An epidemiologic study of epithelial carcinoma of the ovary.* Am J Epidemiol, 1981. **114**(3): p. 398-405.

96. Ames, B.N. and L.S. Gold, *Too many rodent carcinogens: mitogenesis increases mutagenesis.* Science, 1990. **249**(4972): p. 970-1.

97. Schildkraut, J.M., E. Bastos, and A. Berchuck, *Relationship between lifetime ovulatory cycles and overexpression of mutant p53 in epithelial ovarian cancer.* J Natl Cancer Inst, 1997. **89**(13): p. 932-8.

98. Latha, T.S., K. Panati, D.S. Gowd, et al., *Ovarian cancer biology and immunotherapy.* Int Rev Immunol, 2014. **33**(5): p. 428-40.

99. Salazar, H., A.K. Godwin, M.B. Daly, et al., *Microscopic benign and invasive malignant neoplasms and a cancer-prone phenotype in prophylactic oophorectomies.* J Natl Cancer Inst, 1996. **88**(24): p. 1810-20.

100. Hunn, J. and G.C. Rodriguez, *Ovarian cancer: etiology, risk factors, and epidemiology.* Clin Obstet Gynecol, 2012. **55**(1): p. 3-23.

101. Cramer, D.W. and W.R. Welch, *Determinants of ovarian cancer risk. II. Inferences regarding pathogenesis.* J Natl Cancer Inst, 1983. **71**(4): p. 717-21.

102. Stadel, B.V., *Letter: The etiology and prevention of ovarian cancer.* Am J Obstet Gynecol, 1975. **123**(7): p. 772-4.

103. Stratton, J.F., P. Pharoah, S.K. Smith, et al., *A systematic review and meta-analysis of family history and risk of ovarian cancer.* Br J Obstet Gynaecol, 1998. **105**(5): p. 493-9.

104. Weissman, S.M., S.M. Weiss, and A.C. Newlin, *Genetic testing by cancer site: ovary.* Cancer J, 2012. **18**(4): p. 320-7.

105. Fasching, P.A., S. Gayther, L. Pearce, et al., *Role of genetic polymorphisms and ovarian cancer susceptibility.* Mol Oncol, 2009. **3**(2): p. 171-81.

106. Antoniou, A., P.D. Pharoah, S. Narod, et al., *Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies.* Am J Hum Genet, 2003. **72**(5): p. 1117-30.

107. Backes, F.J. and D.E. Cohn, *Lynch syndrome.* Clin Obstet Gynecol, 2011. **54**(2): p. 199-214.

108. Lichtenstein, P., N.V. Holm, P.K. Verkasalo, et al., *Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland.* N Engl J Med, 2000. **343**(2): p. 78-85.

109. Goldgar, D.E., D.F. Easton, L.A. Cannon-Albright, et al., *Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands.* J Natl Cancer Inst, 1994. **86**(21): p. 1600-8.

110. Yeung, T.L., C.S. Leung, K.P. Yip, et al., *Cellular and molecular processes in ovarian cancer metastasis. A Review in the Theme: Cell and Molecular Processes in Cancer Metastasis.* Am J Physiol Cell Physiol, 2015. **309**(7): p. C444-56.

111. Cho, K.R. and M. Shih Ie, *Ovarian cancer.* Annu Rev Pathol, 2009. **4**: p. 287-313.

112. Langhe, R., *microRNA and Ovarian Cancer.* Adv Exp Med Biol, 2015. **889**: p. 119-51.

113. Romero, I. and R.C. Bast, Jr., *Minireview: human ovarian cancer: biology, current management, and paths to personalizing therapy.* Endocrinology, 2012. **153**(4): p. 1593-602.

114. Huober, J., A. Meyer, U. Wagner, et al., *The role of neoadjuvant chemotherapy and interval laparotomy in advanced ovarian cancer.* J Cancer Res Clin Oncol, 2002. **128**(3): p. 153-60.

115. Jayson, G.C., E.C. Kohn, H.C. Kitchener, et al., *Ovarian cancer.* Lancet, 2014. **384**(9951): p. 1376-88.

116. Braem, M.G., L.J. Schouten, P.H. Peeters, et al., *Genetic susceptibility to sporadic ovarian cancer: a systematic review.* Biochim Biophys Acta, 2011. **1816**(2): p. 132-46.

117. Vella, N., M. Aiello, A.E. Russo, et al., *'Genetic profiling' and ovarian cancer therapy (review).* Mol Med Rep, 2011. **4**(5): p. 771-7.

118. Diaz-Padilla, I., E. Amir, S. Marsh, et al., *Genetic polymorphisms as predictive and prognostic biomarkers in gynecological cancers: a systematic review.* Gynecol Oncol, 2012. **124**(2): p. 354-65.

119. Paige, A.J. and R. Brown, *Pharmaco(epi)genomics in ovarian cancer.* Pharmacogenomics, 2008. **9**(12): p. 1825-34.

120. Caiola, E., M. Broggini, and M. Marabese, *Genetic markers for prediction of treatment outcomes in ovarian cancer.* Pharmacogenomics J, 2014. **14**(5): p. 401-10.

121. Medeiros, R., D. Pereira, N. Afonso, et al., *Platinum/paclitaxel-based chemotherapy in advanced ovarian carcinoma: glutathione S-transferase genetic polymorphisms as predictive biomarkers of disease outcome.* Int J Clin Oncol, 2003. **8**(3): p. 156-61.

122. Santos, A.M., H. Sousa, D. Pinto, et al., *Linking TP53 codon 72 and P21 nt590 genotypes to the development of cervical and ovarian cancer.* Eur J Cancer, 2006. **42**(7): p. 958-63.

123. Santos, A.M., H. Sousa, C. Portela, et al., *TP53 and P21 polymorphisms: response to cisplatinum/paclitaxel-based chemotherapy in ovarian cancer.* Biochem Biophys Res Commun, 2006. **340**(1): p. 256-62.

124. Pinto, D., D. Pereira, C. Portela, et al., *The influence of HER2 genotypes as molecular markers in ovarian cancer outcome.* Biochem Biophys Res Commun, 2005. **335**(4): p. 1173-8.

125. Costa, S., D. Pinto, D. Pereira, et al., *Importance of xeroderma pigmentosum group D polymorphisms in susceptibility to ovarian cancer.* Cancer Lett, 2007. **246**(1-2): p. 324-30.

126. Coelho, A., D. Pereira, A. Nogal, et al., *CXCL12 chemokine genotypes as predictive biomarkers of ovarian cancer outcome.* Mol Med Rep, 2009. **2**(1): p. 103-6.

127. Araujo, A.P., R. Ribeiro, D. Pereira, et al., *Ovarian cancer and genetic susceptibility: association of A61G polymorphism in the EGF gene.* Exp Biol Med (Maywood), 2009. **234**(3): p. 241-5.

128. Assis, J., D. Pereira, M. Gomes, et al., *Influence of CYP3A4 genotypes in the outcome of serous ovarian cancer patients treated with first-line chemotherapy: implication of a CYP3A4 activity profile.* Int J Clin Exp Med, 2013. **6**(7): p. 552-61.

129. Pereira, D., J. Assis, M. Gomes, et al., *Improvement of a predictive model in ovarian cancer patients submitted to platinum-based chemotherapy: implications of a GST activity profile.* Eur J Clin Pharmacol, 2016. **72**(5): p. 545-53.

130. Khrunin, A.V., D.V. Khokhrin, A.A. Moisseev, et al., *Pharmacogenomic assessment of cisplatin-based chemotherapy outcomes in ovarian cancer.* Pharmacogenomics, 2014. **15**(3): p. 329-37.

131. Yan, L., Y. Shu-Ying, K. Shan, et al., *Association between polymorphisms of ERCC1 and survival in epithelial ovarian cancer patients with chemotherapy.* Pharmacogenomics, 2012. **13**(4): p. 419-27.

132. Huang, R.S., S. Duan, S.J. Shukla, et al., *Identification of genetic variants contributing to cisplatin-induced cytotoxicity by use of a genomewide approach.* Am J Hum Genet, 2007. **81**(3): p. 427-37.

133. Song, H., S.J. Ramus, J. Tyrer, et al., *A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2.* Nat Genet, 2009. **41**(9): p. 996-1000.

134. Bolton, K.L., J. Tyrer, H. Song, et al., *Common variants at 19p13 are associated with susceptibility to ovarian cancer.* Nat Genet, 2010. **42**(10): p. 880-4.

135. Goode, E.L., G. Chenevix-Trench, H. Song, et al., *A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24.* Nat Genet, 2010. **42**(10): p. 874-9.

136. Huang, R.S., S.E. Johnatty, E.R. Gamazon, et al., *Platinum sensitivity-related germline polymorphism discovered via a cell-based approach and analysis of its association with outcome in ovarian cancer patients.* Clin Cancer Res, 2011. **17**(16): p. 5490-500.

137. Lu, Y., X. Chen, J. Beesley, et al., *Genome-wide association study for ovarian cancer susceptibility using pooled DNA.* Twin Res Hum Genet, 2012. **15**(5): p. 615-23.

138. Couch, F.J., X. Wang, L. McGuffog, et al., *Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk.* PLoS Genet, 2013. **9**(3): p. e1003212.

139. Pharoah, P.D., Y.Y. Tsai, S.J. Ramus, et al., *GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer.* Nat Genet, 2013. **45**(4): p. 362-70, 370e1-2.

140. Chen, K., H. Ma, L. Li, et al., *Genome-wide association study identifies new susceptibility loci for epithelial ovarian cancer in Han Chinese women.* Nat Commun, 2014. **5**: p. 4682.

141. Earp, M.A., L.E. Kelemen, A.M. Magliocco, et al., *Genome-wide association study of subtype-specific epithelial ovarian cancer risk alleles using pooled DNA.* Hum Genet, 2014. **133**(5): p. 481-97.

142. Johnatty, S.E., J.P. Tyrer, S. Kar, et al., *Genome-wide Analysis Identifies Novel Loci Associated with Ovarian Cancer Outcomes: Findings from the Ovarian Cancer Association Consortium.* Clin Cancer Res, 2015. **21**(23): p. 5264-76.

143. Kuchenbaecker, K.B., S.J. Ramus, J. Tyrer, et al., *Identification of six new susceptibility loci for invasive epithelial ovarian cancer.* Nat Genet, 2015. **47**(2): p. 164-71.

144. French, J.D., S.E. Johnatty, Y. Lu, et al., *Germline polymorphisms in an enhancer of PSIP1 are associated with progression-free survival in epithelial ovarian cancer.* Oncotarget, 2016. **7**(6): p. 6353-68.

145. Fridley, B.L., T.M. Ghosh, A. Wang, et al., *Genome-Wide Study of Response to Platinum, Taxane, and Combination Therapy in Ovarian Cancer: In vitro Phenotypes, Inherited Variation, and Disease Recurrence.* Front Genet, 2016. **7**: p. 37.

146. . http://www.reactome.org/

147. Kelemen, L.E., K. Lawrenson, J. Tyrer, et al., *Genome-wide significant risk associations for mucinous ovarian carcinoma.* Nat Genet, 2015. **47**(8): p. 888-97.

148. Phelan, C.M., K.B. Kuchenbaecker, J.P. Tyrer, et al., *Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer.* Nat Genet, 2017. **49**(5): p. 680-691.

149. Kirov, G., I. Nikolov, L. Georgieva, et al., *Pooled DNA genotyping on Affymetrix SNP genotyping arrays.* BMC Genomics, 2006. **7**: p. 27.

150. Wheeler, H.E. and M.E. Dolan, *Lymphoblastoid cell lines in pharmacogenomic discovery and clinical translation.* Pharmacogenomics, 2012. **13**(1): p. 55-70.

151. Higgs, M.R. and G.S. Stewart, *Protection or resection: BOD1L as a novel replication fork protection factor.* Nucleus, 2016. **7**(1): p. 34-40.

152. Daugaard, M., T. Kirkegaard-Sorensen, M.S. Ostenfeld, et al., *Lens epithelium-derived growth factor is an Hsp70-2 regulated guardian of lysosomal stability in human cancer.* Cancer Res, 2007. **67**(6): p. 2559-67.

153. Gyorffy, B., A. Lanczky, and Z. Szallasi, *Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients.* Endocr Relat Cancer, 2012. **19**(2): p. 197-208.

154. Daly, A.K., *Genome-wide association studies in pharmacogenomics.* Nat Rev Genet, 2010. **11**(4): p. 241-6.

155. Evans, W.E. and H.L. McLeod, *Pharmacogenomics--drug disposition, drug targets, and side effects.* N Engl J Med, 2003. **348**(6): p. 538-49.

156. Wang, L., H.L. McLeod, and R.M. Weinshilboum, *Genomics and drug response.* N Engl J Med, 2011. **364**(12): p. 1144-53.

157. Jallow, M., Y.Y. Teo, K.S. Small, et al., *Genome-wide and fine-resolution association analysis of malaria in West Africa.* Nat Genet, 2009. **41**(6): p. 657-65.

158. Freedman, M.L., A.N. Monteiro, S.A. Gayther, et al., *Principles for the post-GWAS functional characterization of cancer risk loci.* Nat Genet, 2011. **43**(6): p. 513-8.

159. Kraft, P., E. Zeggini, and J.P. Ioannidis, *Replication in genome-wide association studies.* Stat Sci, 2009. **24**(4): p. 561-573.

160. Chung, C.C., W.C. Magalhaes, J. Gonzalez-Bosquet, et al., *Genome-wide association studies in cancer--current and future directions.* Carcinogenesis, 2010. **31**(1): p. 111-20.

161. Huo, D. and O.I. Olopade, *Interpretation of genome-wide association study results.* Oncology (Williston Park), 2010. **24**(7): p. 643, 646.

162. De, R., W.S. Bush, and J.H. Moore, *Bioinformatics challenges in genome-wide association studies (GWAS).* Methods Mol Biol, 2014. **1168**: p. 63-81.

163. Zollner, S. and J.K. Pritchard, *Overcoming the winner's curse: estimating penetrance parameters from case-control data.* Am J Hum Genet, 2007. **80**(4): p. 605-15.

164. Bush, W.S., M.T. Oetjens, and D.C. Crawford, *Unravelling the human genome-phenome relationship using phenome-wide association studies.* Nat Rev Genet, 2016. **17**(3): p. 129-45.

165. Igl, B.W., I.R. Konig, and A. Ziegler, *What do we mean by 'replication' and 'validation' in genome-wide association studies?* Hum Hered, 2009. **67**(1): p. 66-8.

166. McCarthy, M.I., G.R. Abecasis, L.R. Cardon, et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges.* Nat Rev Genet, 2008. **9**(5): p. 356-69.

167. Mostowska, A., S. Sajdak, P. Pawlik, et al., *Replication study for the association of seven genome- GWAS-identified Loci with susceptibility to ovarian cancer in the Polish population.* Pathol Oncol Res, 2015. **21**(2): p. 307-13.

168. Bolton, K.L., C. Ganda, A. Berchuck, et al., *Role of common genetic variants in ovarian cancer susceptibility and outcome: progress to date from the Ovarian Cancer Association Consortium (OCAC).* J Intern Med, 2012. **271**(4): p. 366-78.

169. Prat, J. and F.C.o.G. Oncology, *FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum: abridged republication.* J Gynecol Oncol, 2015. **26**(2): p. 87-9.

170. Rustin, G.J., I. Vergote, E. Eisenhauer, et al., *Definitions for response and progression in ovarian cancer clinical trials incorporating RECIST 1.1 and CA 125 agreed by the Gynecological Cancer Intergroup (GCIG).* Int J Gynecol Cancer, 2011. **21**(2): p. 419-23.

171. Pinto, R., J. Assis, A. Nogueira, et al., *Rethinking ovarian cancer genomics: where GWAS stand?* Pharmacogenomics, in press.

172. Saccone, S.F., R. Bolze, P. Thomas, et al., *SPOT: a web-based tool for using biological databases to prioritize SNPs after a genome-wide association study.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W201-9.

173. Eisenhauer, E.A., P. Therasse, J. Bogaerts, et al., *New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1).* Eur J Cancer, 2009. **45**(2): p. 228-47.

174. ; Available from: http://www.ensembl.org.

175. Zhang, D., M.X. Sliwkowski, M. Mark, et al., *Neuregulin-3 (NRG3): a novel neural tissue-enriched protein that binds and activates ErbB4.* Proc Natl Acad Sci U S A, 1997. **94**(18): p. 9562-7.

176. Falls, D.L., *Neuregulins: functions, forms, and signaling strategies.* Exp Cell Res, 2003. **284**(1): p. 14-30.

177. Kogata, N., M. Zvelebil, and B.A. Howard, *Neuregulin 3 and erbb signalling networks in embryonic mammary gland development.* J Mammary Gland Biol Neoplasia, 2013. **18**(2): p. 149-54.

178. Maihle, N.J., A.T. Baron, B.A. Barrette, et al., *EGF/ErbB receptor family in ovarian cancer.* Cancer Treat Res, 2002. **107**: p. 247-58.

179. Sheng, Q. and J. Liu, *The therapeutic potential of targeting the EGFR family in epithelial ovarian cancer.* Br J Cancer, 2011. **104**(8): p. 1241-5.

180. Benzel, I., A. Bansal, B.L. Browning, et al., *Interactions among genes in the ErbB-Neuregulin signalling network are associated with increased susceptibility to schizophrenia.* Behav Brain Funct, 2007. **3**: p. 31.

181. Kwan, T., D. Benovoy, C. Dias, et al., *Genome-wide analysis of transcript isoform variation in humans.* Nat Genet, 2008. **40**(2): p. 225-31.

182. Coulombe-Huntington, J., K.C. Lam, C. Dias, et al., *Fine-scale variation and genetic determinants of alternative splicing across individuals.* PLoS Genet, 2009. **5**(12): p. e1000766.

183. Tak, Y.G. and P.J. Farnham, *Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome.* Epigenetics Chromatin, 2015. **8**: p. 57.

184. Albert, F.W. and L. Kruglyak, *The role of regulatory variation in complex traits and disease.* Nat Rev Genet, 2015. **16**(4): p. 197-212.

185. Gaffney, D.J., *Global properties and functional complexity of human gene regulatory variation.* PLoS Genet, 2013. **9**(5): p. e1003501.

186. Maurano, M.T., R. Humbert, E. Rynes, et al., *Systematic localization of common disease-associated variation in regulatory DNA.* Science, 2012. **337**(6099): p. 1190-5.

187. Gamazon, E.R., R.S. Huang, N.J. Cox, et al., *Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci.* Proc Natl Acad Sci U S A, 2010. **107**(20): p. 9287-92.

188. Crijns, A.P., R.S. Fehrmann, S. de Jong, et al., *Survival-related profile, pathways, and transcription factors in ovarian cancer.* PLoS Med, 2009. **6**(2): p. e24.

189. Harden, J.L., S.M. Lewis, S.R. Lish, et al., *The tryptophan metabolism enzyme L-kynureninase is a novel inflammatory factor in psoriasis and other inflammatory diseases.* J Allergy Clin Immunol, 2016. **137**(6): p. 1830-40.

190. Bast, R.C., Jr., B. Hennessy, and G.B. Mills, *The biology of ovarian cancer: new opportunities for translation.* Nat Rev Cancer, 2009. **9**(6): p. 415-28.

191. Halkia, E., J. Spiliotis, and P. Sugarbaker, *Diagnosis and management of peritoneal metastases from ovarian cancer.* Gastroenterol Res Pract, 2012. **2012**: p. 541842.

192. Lengyel, E., *Ovarian cancer development and metastasis.* Am J Pathol, 2010. **177**(3): p. 1053-64.

193. De Smet, F., N.L. Pochet, K. Engelen, et al., *Predicting the clinical behavior of ovarian cancer from gene expression profiles.* Int J Gynecol Cancer, 2006. **16 Suppl 1**: p. 147-51.

194. Atlas, E., M. Cardillo, I. Mehmi, et al., *Heregulin is sufficient for the promotion of tumorigenicity and metastasis of breast cancer cells in vivo.* Mol Cancer Res, 2003. **1**(3): p. 165-75.

195. Breuleux, M., *Role of heregulin in human cancer.* Cell Mol Life Sci, 2007. **64**(18): p. 2358-77.

196. Pradeep, S., S.W. Kim, S.Y. Wu, et al., *Hematogenous metastasis of ovarian cancer: rethinking mode of spread.* Cancer Cell, 2014. **26**(1): p. 77-91.

197. Feng, L., J. Huang, and J. Chen, *MERIT40 facilitates BRCA1 localization and DNA damage repair.* Genes Dev, 2009. **23**(6): p. 719-28.

198. Shi, W., M.K. Tang, Y. Yao, et al., *BRE plays an essential role in preventing replicative and DNA damage-induced premature senescence.* Sci Rep, 2016. **6**: p. 23506.

199. Yeung, C.K., G. Wang, Y. Yao, et al., *BRE modulates granulosa cell death to affect ovarian follicle development and atresia in the mouse.* Cell Death Dis, 2017. **8**(3): p. e2697.

200. Rebbeck, T.R., N. Mitra, S.M. Domchek, et al., *Modification of BRCA1-Associated Breast and Ovarian Cancer Risk by BRCA1-Interacting Genes.* Cancer Res, 2011. **71**(17): p. 5792-805.

201. Kyrieleis, O.J., P.B. McIntosh, S.R. Webb, et al., *Three-Dimensional Architecture of the Human BRCA1-A Histone Deubiquitinase Core Complex.* Cell Rep, 2016. **17**(12): p. 3099-3106.

202. Li, Y., K. Qi, L. Zu, et al., *Anti-apoptotic brain and reproductive organ-expressed proteins enhance cisplatin resistance in lung cancer cells via the protein kinase B signaling pathway.* Thorac Cancer, 2016. **7**(2): p. 190-8.

203. Chan, B.C., Q. Li, S.K. Chow, et al., *BRE enhances in vivo growth of tumor cells.* Biochem Biophys Res Commun, 2005. **326**(2): p. 268-73.

204. Cesaratto, L., E. Grisard, M. Coan, et al., *BNC2 is a putative tumor suppressor gene in high-grade serous ovarian carcinoma and impacts cell survival after oxidative stress.* Cell Death Dis, 2016. **7**(9): p. e2374.

205. Brachner, A., J. Braun, M. Ghodgaonkar, et al., *The endonuclease Ankle1 requires its LEM and GIY-YIG motifs for DNA cleavage in vivo.* J Cell Sci, 2012. **125**(Pt 4): p. 1048-57.

206. Ahmed, S., D. Bott, A. Gomez, et al., *Loss of the Mono-ADP-ribosyltransferase, Tiparp, Increases Sensitivity to Dioxin-induced Steatohepatitis and Lethality.* J Biol Chem, 2015. **290**(27): p. 16824-40.

207. McEwan, D.G. and I. Dikic, *PLEKHM1: Adapting to life at the lysosome.* Autophagy, 2015. **11**(4): p. 720-2.

208. Chen, S., Z. Yang, A.W. Wilkinson, et al., *The PZP Domain of AF10 Senses Unmodified H3K27 to Regulate DOT1L-Mediated Methylation of H3K79.* Mol Cell, 2015. **60**(2): p. 319-27.

209. Kai, F., J.P. Fawcett, and R. Duncan, *Synaptopodin-2 induces assembly of peripheral actin bundles and immature focal adhesions to promote lamellipodia formation and prostate cancer cell migration.* Oncotarget, 2015. **6**(13): p. 11162-74.

210. Hibar, D.P., H.H. Adams, N. Jahanshad, et al., *Novel genetic loci associated with hippocampal volume.* Nat Commun, 2017. **8**: p. 13624.

211. Hsieh, J., M. Koseki, M.M. Molusky, et al., *TTC39B deficiency stabilizes LXR reducing both atherosclerosis and steatohepatitis.* Nature, 2016. **535**(7611): p. 303-7.

212. Chen, T., B. Long, G. Ren, et al., *Protocadherin20 Acts as a Tumor Suppressor Gene: Epigenetic Inactivation in Nasopharyngeal Carcinoma.* J Cell Biochem, 2015. **116**(8): p. 1766-75.

213. Kim, K.M., H. Cho, K. Choi, et al., *A new MIF4G domain-containing protein, CTIF, directs nuclear cap-binding protein CBP80/20-dependent translation.* Genes Dev, 2009. **23**(17): p. 2033-45.

214. Houde, A.A., S.M. Ruchat, C. Allard, et al., *LRP1B, BRD2 and CACNA1D: new candidate genes in fetal metabolic programming of newborns exposed to maternal hyperglycemia.* Epigenomics, 2015. **7**(7): p. 1111-22.

215. Sato, Y., M. Akitsu, Y. Amano, et al., *The novel PAR-1-binding protein MTCL1 has crucial roles in organizing microtubules in polarizing epithelial cells.* J Cell Sci, 2013. **126**(Pt 20): p. 4671-83.

8. Appendix

Appendix 1

Supplementary Table 1 - Overview of OC susceptibility GWAS

Study	Associated SNPs		Population	No. cases/controls (combined)	MAF	Locus	Gene	Overall risk		Subtype-specific risk		
								OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	Histological subtype
Song *et al*. 2009	rs3814113 ‡, #		European ancestry	8761/11831	0.32	9p22.2	~44 kb upstream of *BNC2* gene	0.82 (0.79-0.86) [a]	5.10×10^{-19} [c]	0.77 (0.73-0.81) [a]	4.10×10^{-21} [c]	Serous
										0.86 (0.79-0.94) [a]	1.00×10^{-3} [c]	Endometrioid
										0.83 (0.76-0.91) [a]	6.60×10^{-5} [c]	Other
Bolton *et al*. 2010	rs8170 ‡		European ancestry	10496/13172	0.11 [e]	19p13	*MERIT40*	1.12 (1.07-1.17) [a]	3.60×10^{-6} [c]	1.18 (1.12-1.25) [a]	2.70×10^{-9} [c]	Serous
	rs2363956 ‡		European ancestry	10480/13176	0.46 [e]		*ANKLE1*	1.10 (1.06-1.15) [a]	1.20×10^{-7} [c]	1.16 (1.11-1.21) [a]	3.80×10^{-11} [c]	Serous
Goode *et al*. 2010	rs2072590 ‡		European ancestry	10406/16340	0.18 [e]	2q31	Non-coding region	1.16 (1.12-1.21) [a]	4.50×10^{-14}	1.20 (1.14-1.25) [a]	3.80×10^{-14}	Serous
			European ancestry							1.30 (1.17-1.44) [a]	7.30×10^{-7}	Mucinous
			European ancestry							1.13 (1.04-1.22) [a]	2.40×10^{-3}	Endometrioid
	rs2665390 ‡		European ancestry	10406/17369	0.07 [e]	3q25	*TIPARP*	1.19 (1.11-1.27) [a]	3.20×10^{-7}	1.24 (1.15-1.34) [a]	7.10×10^{-8}	Serous
			European ancestry							1.23 (1.08-1.40) [a]	1.90×10^{-3}	Endometrioid
	rs10088218 ‡		European ancestry	10462/16362	0.09 [e]	8q24	Gene desert region	0.84 (0.80-0.89) [a]	3.20×10^{-9}	0.76 (0.70-0.81) [a]	8.00×10^{-15}	Serous
			European ancestry							1.21 (1.05-1.40) [a]	1.00×10^{-2}	Clear cell
	rs1516982 ‡		European ancestry	10472/54111	0.16 [c]			0.86 (0.82-0.91) [a]	2.00×10^{-8}	0.81 (0.76-0.86) [a]	3.30×10^{-11}	Serous
	rs10098821 ‡, #		European ancestry	10414/16136	0.10 [c]			0.83 (0.78-0.89) [a]	4.70×10^{-9}	0.75 (0.70-0.81) [a]	2.30×10^{-13}	Serous
	rs9303542 ‡, #		European ancestry	10242/13091	0.32 [e]	17q21	*SKAP1*	1.11 (1.06-1.16) [a]	1.40×10^{-6}	1.14 (1.09-1.20) [a]	1.40×10^{-7}	Serous
Lu *et al*. 2012	No significant associations		White, non-Hispanic	342/643	-	-	-	-	-	-	-	-
			(White, non-Hispanic)	6195 (5620)/7854 (6966)								
Couch *et al*. 2013	BRCA1 mutation carriers	rs17631303 ‡	European ancestry	2273/11997	0.22	17q21	*PLEKHM1* [f]	1.27 (1.17-1.38) [a, g]	1.40×10^{-8} [c]			
		rs183211 ‡	European ancestry	2281/12070	0.26		*NSF* [f]	1.25 (1.16-1.35) [a, g]	3.10×10^{-8} [c]			
		rs4691139 ‡	European ancestry	2280/12070	0.52	4q32.3	Chr4:164987569	1.20 (1.17-1.38) [a, g]	3.40×10^{-8} [c]			

Supplementary Table 1 - Overview of OC susceptibility GWAS (cont.)

Study	Associated SNPs	Population	No. cases/controls (combined)	MAF	Locus	Gene	Overall risk		Subtype-specific risk		
							OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	Histological subtype
Pharoah *et al*. 2013	rs11782652 ‡	European ancestry	16283/23491	0.07	8q21	*CHMP4C*	1.19 (1.12-1.26) [a]	5.50×10^{-9}	1.24 (1.16-1.33) [a]	7.00×10^{-10}	Serous
	rs1243180 ‡	European ancestry		0.31	10p12	*MLLT10*	1.10 (1.06-1.13) [a]	1.80×10^{-8}	1.11 (1.07-1.15) [a]	1.40×10^{-7}	Serous
									1.08 (1.00-1.15) [a]	3.80×10^{-2}	Endometrioid
	rs757210 ‡	European ancestry		0.37	17q12	*HNF1B*	1.05 (1.02-1.09) [a]	9.00×10^{-4}	1.12 (1.08-1.17) [a]	8.10×10^{-10}	Serous
									0.80 (0.72-0.88) [a]	3.90×10^{-6}	Clear cell
									0.89 (0.81-0.99) [a]	2.70×10^{-2}	Mucinous
Earp *et al*. 2014	rs11108890 #	European ancestry	78/392 1483/21530	0.04	Chr12: 96137530	*LOC105369927* [f]			1.38 (1.16–1.66) [b]	4.00×10^{-4}	Mucinous
	rs933518 #	European ancestry	78/392 1483/21530	0.08	Chr16: 53079622				1.26 (1.11–1.45) [b]	8.00×10^{-4}	
	rs17106154 #	European ancestry	78/392 1483/21530	0.07	Chr14: 68230927	Lies within a ~150kb LD region of *ZFP36L1*			1.21 (1.06–1.39) [b]	5.80×10^{-3}	
	rs970651 #	European ancestry	78/392 1483/21530	0.16	Chr13: 47351705				1.16 (1.05–1.28) [b]	4.20×10^{-3}	
	rs7981902 #	European ancestry	78/392 1483/21530	0.13	Chr13: 47368792				1.15 (1.03–1.28) [b]	1.14×10^{-2}	
	rs2190503 #	European ancestry	114/392 2903/21528	0.13	Chr7: 50710111	Identify a locus upstream/intronic to *GRB10* gene			1.12 (1.04–1.22) [b]	5.00×10^{-3}	Endometrioid/ clear cell
	rs6593140 #	European ancestry	114/392 2903/21528	0.12	Chr7: 50765627				1.09 (1.03–1.17) [b]	6.00×10^{-3}	
	rs2329554 #	European ancestry	114/392 2903/21528	0.22	Chr7: 50842524				1.12 (1.03–1.22) [b]	6.00×10^{-3}	
	rs9609538 #	European ancestry	68/392 892/21529	0.24	Chr22: 31139832	~5 bp downstream *BPIL2* and ~500 bp upstream *C22orf28*			0.84 (0.76–0.93) [b]	7.00×10^{-4}	LMP Serous
Chen *et al*. 2014	rs1413299 ‡	Han Chinese	2496/3975	0.42	9q22.33	*COL15A1*	1.24 (1.15-1.33) [b]	1.88×10^{-8} [d]	1.25 (1.14–1.38) [b]	2.13×10^{-6} [d]	Serous
									1.30 (1.15–1.45) [b]	9.69×10^{-6} [d]	Other
	rs1192691 ‡	Han Chinese		0.41	10p11.21	245 kb upstream of exon 1 of *ANKRD30A*	0.81 (0.75-0.87) [b]	2.62×10^{-8} [d]	0.80 (0.72–0.87) [b]	1.13×10^{-6} [d]	Serous
									0.83 (0.73–0.92) [b]	1.16×10^{-3} [d]	Other

Supplementary Table 1 - Overview of OC susceptibility GWAS (cont.)

Study	Associated SNPs		Population	No. cases/controls (combined)	MAF	Locus	Gene	Overall risk		Subtype-specific risk		
								OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	Histological subtype
Chen *et al*. 2014 (cont.)	rs11175194 ‡		Han Chinese	2496/3975	0.37	12q14.2	*SRGAP1*	0.82 (0.76-0.88) [b]	1.14×10^{-7} [d]	0.83 (0.75–0.91) [b]	6.22×10^{-5} [d]	Serous
										0.73 (0.59–0.90) [b]	3.36×10^{-3} [d]	Endometrioid
										0.86 (0.76–0.97) [b]	1.40×10^{-2} [d]	Other
	rs633862 ‡		Han Chinese		0.42	9q34.2	5 kb upstream of *ABO* gene	0.83 (0.77-0.89) [b]	8.57×10^{-7} [d]	0.86 (0.79–0.95) [b]	1.95×10^{-3} [d]	Serous
										0.73 (0.59–0.90) [b]	2.50×10^{-3} [d]	Endometrioid
										0.80 (0.71–0.90) [b]	1.93×10^{-4} [d]	Other
Kuchenbaecker *et al*. 2015	rs56318008		European ancestry	15437/30845	0.15	1p36	*WNT4*	1.11 (1.06-1.15) [a]	8.00×10^{-7}	1.12 (1.06-1.17) [a]	6.00×10^{-6}	Serous
										1.24 (1.10-1.39) [a]	5.00×10^{-4}	Clear cell
	rs58722170		European ancestry		0.23	1p34.3	*RSPO1*	1.07 (1.03-1.11) [a]	2.00×10^{-4}	1.12 (1.07-1.17) [a]	4.00×10^{-7}	Serous
	rs17329882		European ancestry		0.24	4q26	*SYNPO2*	1.09 (1.06-1.13) [a]	3.00×10^{-7}	1.11 (1.07-1.16) [a]	3.00×10^{-7}	Serous
										1.09 (1.01-1.18) [a]	2.00×10^{-2}	Endometrioid
	rs116133110		European ancestry		0.31	6p22.1	*GPX6*	0.94 (0.91-0.97) [a]	9.00×10^{-5}	0.91 (0.87-0.94) [a]	3.00×10^{-7}	Serous
	rs6356		European ancestry		0.19	9q34.2	Upstream of *ABO*	1.12 (1.08-1.16) [a]	9.00×10^{-9}	1.13 (1.08-1.18) [a]	2.00×10^{-7}	Serous
										1.12 (1.03-1.21) [a]	7.00×10^{-3}	Endometrioid
										1.23 (1.10-1.38) [a]	3.00×10^{-4}	Mucinous
	rs199661266		European ancestry		0.28	17q11.2	*ATAD5*	0.90 (0.87-0.93) [a]	1.00×10^{-9}	0.90 (0.87-0.94) [a]	2.00×10^{-7}	Serous
										0.88 (0.82-0.95) [a]	5.00×10^{-4}	Endometrioid
										0.88 (0.80-0.98) [a]	2.00×10^{-2}	Clear cell
	BRCA1 mutation carriers	rs56318008	European ancestry	Affected/unaffected (2462/12790)	0.15	1p36	*WNT4*	1.15 (1.05-1.26) [a, g]	3.10×10^{-3}			
		rs58722170	European ancestry		0.23	1p34.3	*RSPO1*	1.14 (1.05-1.23) [a, g]	1.50×10^{-3}			
		rs17329882	European ancestry		0.24	4q26	*SYNPO2*	1.08 (1.00-1.17) [a, g]	4.20×10^{-2}			
		rs116133110	European ancestry		0.31	6p22.1	*GPX6*	0.92 (0.86-0.99) [a, g]	2.30×10^{-2}			
		rs6356	European ancestry		0.19	9q34.2	Upstream of *ABO*	1.11 (1.02-1.21) [a, g]	1.20×10^{-2}			

Supplementary Table 1 - Overview of OC susceptibility GWAS (cont.)

Study	Associated SNPs		Population	No. cases/controls (combined)	MAF	Locus	Gene	Overall risk		Subtype-specific risk		
								OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	Histological subtype
Kuchenbaecker *et al*. 2015 (cont.)	*BRCA2* mutation carriers	rs58722170	European ancestry	631/7580	0.23	1p34.3	*RSPO1*	1.35 (1.17-1.57) [a, g]	$5.20x10^{-5}$			
Keleman et *al*. 2015	rs752590		European ancestry	1644 (mucinous subtype) /21693	0.21	2q13	*PAX8*			1.34 (1.21-1.49) [a]	$3.30x10^{-8}$	Mucinous
	rs711830		European ancestry		0.32	2q31.1	*HOXD3*			1.30 (1.20-1.40) [a]	$7.50x10^{-12}$	Mucinous
	rs688187		European ancestry		0.32	19q13.2	*IFNL3*			0.67 (0.60-0.75) [a]	$6.80x10^{-13}$	Mucinous
Phelan et *al*. 2017	rs112071820		European ancestry	25509/40941	0.33	3q22.3	Non-coding region			1.29 (1.20-1.37) [a]	$1.50x10^{-13}$	Mucinous
	rs9870207		European ancestry		0.27	3q28	Non-coding region			1.19 (1.12-1.27) [a]	$4.50x10^{-8}$	Serous
	rs13113999		European ancestry		0.44	4q32.3	Non-coding region			1.23 (1.14-1.32) [a]	$4.70x10^{-8}$	Serous
	rs555025179		European ancestry		0.44	5q12.3	*MAST4*			1.18 (1.11-1.26) [a]	$4.50x10^{-8}$	Endometrioid
	rs150293538		European ancestry		0.01	8q21.11	*LINC01111*			2.19 (1.65-2.90) [a]	$2.00x10^{-9}$	Serous
	rs320203		European ancestry		0.12	9q31.1	*LINC00587*			1.29 (1.18-1.41) [a]	$1.70x10^{-8}$	Mucinous
	rs7902587		European ancestry		0.12	10q24.33	Non-coding region			1.29 (1.18-1.41) [a]	$4.00x10^{-8}$	Serous
	rs8098244		European ancestry		0.31	18q11.2	*LAMA3*			1.19 (1.12-1.27) [a]	$3.90x10^{-8}$	Serous
	rs6005807		European ancestry		0.09	22q12.1	*MIR5739*			1.17 (1.11-1.23) [a]	$4.50x10^{-9}$	Serous
	BRCA1/2 mutation + non-mutation carriers	rs2165109	European ancestry	*BRCA1* mutation carriers (2933 affected/16103unaffected) *BRCA2* mutation carriers (954 affected/11458 unaffected)	0.25	2q13	*LOC400997*			1.09 (1.05-1.12) [a]	$4.20x10^{-8}$	Serous + *BRCA1/2*
	BRCA1/2 mutation + non-mutation carriers	rs9886651	European ancestry		0.46	8q24.21	*MIR1204*			1.08 (1.05-1.11) [a]	$3.50x10^{-9}$	Serous + *BRCA1/2*
	BRCA1/2 mutation + non-mutation carriers	rs7953249	European ancestry		0.42	12q24.31	Non-coding region			1.08 (1.06-1.06) [a]	$1.10x10^{-9}$	Serous + *BRCA1/2*

The results shown in this table are the combined phases results of each study or the validation phases results (when conducted).a: per-allele; b: log-additive model; c: P_{trend}; d: P_{meta}; e: MAF values for all populations obtained on "Ensembl" database; f: data obtained from "NCBI" database; g: HR values. The primary endpoint in this analysis was the age at ovarian cancer diagnosis. Mutation carriers were followed until the age of ovarian cancer diagnosis, or risk-reducing salpingo-oophorectomy or age at last observation.
ǂ: replicated SNPs; #: validated SNPs

Appendix 2

Supplementary Table 2 - Overview of OC clinical outcome GWAS

Study	Associated SNPs		Population	No. cases/deaths (combined)	MAF	Locus	Gene	Overall risk		Subtype-specific risk		
								HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	Histological subtype
Bolton *et al*. 2010	rs8170 ‡		All populations	8946/3354	0.11 [d]	19p13	*MERIT40*	**OS** 1.11 (1.04-1.17) [a]	5.20×10^{-4} [b]	1.28 (1.02–1.60) [a]	3.00×10^{-2} [b]	Endometrioid
	rs2363956 ‡		All populations	8900/3342	0.46 [d]		*ANKLE1*	**OS** 1.09 (1.04-1.14) [a]	5.60×10^{-4} [b]	1.09 (1.03-1.16) [a]	5.20×10^{-3} [b]	Serous
Huang *et al*. 2011	rs1649942 ‡		European ancestry Non-Hispanic white	1703 cases	0.24	Chr10	*NRG3*	**PFS** 1.25 (1.03-1.52) [a]	2.30×10^{-2}			
								PFS Optimally debulked patients 1.43 (1.12-1.81) [a]	4.00×10^{-3}			
								OS Optimally debulked patients 1.48 (1.10-2.00) [a]	9.00×10^{-3}			
Johnatty *et al*. 2015	"All chemotherapy" group	rs6674079 ‡	European ancestry	4426 cases	0.28	1q22	*RP11–284F21.8*	**OS** 1.15 (1.08-1.23) [a]	7.10×10^{-6}			
				4095 cases				**PFS** 1.07 (1.01-1.13) [a]	2.80×10^{-2}			
		rs7950311 ‡	European ancestry	4426 cases	0.48	11p15.4	*HBG2*	**OS** 1.10 (1.04-1.17) [a]	1.70×10^{-3}			
		rs4910232 ‡	European ancestry	4426 cases	0.32	11p15.3	*RP11-179A10.1*	**OS** 1.12 (1.05-1.19) [a]	9.40×10^{-4}			
				4095 cases				**PFS** 1.17 (1.10-1.24) [a]	4.70×10^{-7}			
		rs2549714 ‡	European ancestry	4426 cases	0.06	16q23	*RP11–314O13.1*	**OS** 1.20 (1.06-1.36) [a]	3.40×10^{-3}			
				4095 cases				**PFS** 1.14 (1.01-1.28) [a]	2.80×10^{-2}			
		rs3795247 ‡	European ancestry	4426 cases	0.08	19p12	*ZNF100*	**OS** 1.16 (1.04-1.30) [a]	8.80×10^{-3}			
				4095 cases				**PFS** 1.26 (1.14-1.40) [a]	1.05×10^{-5}			
	"Standard chemotherapy" group	rs7950311 ‡	European ancestry	1799 cases	0.48	11p15.4	*HBG2*	**OS** 1.28 (1.16-1.42) [a]	6.80×10^{-7}			
		rs4910232 ‡	European ancestry	1799 cases	0.32	11p15.3	*RP11-179A10.1*	**OS** 1.20 (1.08-1.33) [a]	5.30×10^{-4}			
				1598 cases				**PFS** 1.24 (1.12-1.56) [a]	1.20×10^{-5}			

Supplementary Table 2 - Overview of OC clinical outcome GWAS

Study	Associated SNPs			Population	No. cases/deaths (combined)	MAF	Locus	Gene	Overall risk		Subtype-specific risk		
									HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	Histological subtype
Johnatty *et al*. 2015 (cont.)	"Standard chemotherapy" group	rs2549714 ‡		European ancestry	1799 cases	0.06	16q23	*RP11–314O13.1*	**OS** 1.53 (1.28-1.84) [a]	5.00×10^{-6}			
					1598 cases				**PFS** 1.29 (1.08-1.55) [a]	5.60×10^{-3}			
		rs3795247 ‡		European ancestry	1799 cases	0.08	19p12	*ZNF100*	**OS** 1.34 (1.13-1.60) [a]	9.70×10^{-4}			
					1598 cases				**PFS** 1.39 (1.18-1.65) [a]	9.20×10^{-5}			
French *et al*. 2016	rs72700653 ‡			European ancestry	1244 cases	0.02	Chr9	*TTC39B*			**PFS** 1.91 (1.36-2.69) [a]	2.20×10^{-4}	Serous
	rs7874043 ‡			European ancestry		0.02	Chr9				**PFS** 1.90 (1.38-2.61) [a]	7.30×10^{-5}	Serous
											OS 1.56 (1.09-2.23) [a]	1.50×10^{-2}	Serous
Fridley *et al*. 2016	Paclitaxel	MTT	rs185229225	European ancestry	74 cases	0.02	Chr4: 13609129	*BOD1L1*	- [e]	2.20×10^{-7} [c]			
		Caspase	rs3842595	European ancestry		0.14	Chr5:121778606	*MGC32805/SNCAIP*	- [e]	2.60×10^{-7} [c]			
	Carboplatin	MTT	rs150303591	European ancestry		0.29	Chr4:79009309	*FRAS1*	+ [f]	5.90×10^{-7} [c]			
	Combination	MTT	rs201023017	European ancestry		0.41	Chr3:143103669	*SLC9A9*	+ [f]	6.00×10^{-7} [c]			
			rs66696671	European ancestry		0.23	Chr10:121366953	*TIAL1*	- [e]	7.30×10^{-7} [c]			
		Caspase	rs12025262	European ancestry		0.39	Chr1:247356732	*ZNF731P*	- [e]	6.60×10^{-7} [c]			
			rs10674174	European ancestry		0.42	Chr13:61892075	*PCDH20*	- [e]	8.20×10^{-7} [c]			
	Paclitaxel	MTT	rs35067965	European ancestry		0.33 [d]	Chr18:45539 6	*COLEC12*	- [e]	2.20×10^{-5}			
		Caspase								3.80×10^{-5}			
	Carboplatin	MTT	rs8091660	European ancestry		0.28 [d]	Chr18:46087936	*CTIF*	- [e]	8.90×10^{-6}			
		Caspase	rs113867814	European ancestry		Del [d]	Chr18:46259604			1.20×10^{-5}			
		MTT	rs2748151	European ancestry		0.28 [d]	Chr20:60133486	*CDH4*	+ [f]	4.70×10^{-5}			
		Caspase	rs113594423	European ancestry		0.10 [d]	Chr20:60379048			2.40×10^{-5}			
	Carboplatin	Caspase	rs5830067	European ancestry		Del [d]	Chr2:28537890	*BRE*	+ [f]	1.70×10^{-5}			
	Combination	Caspase	rs7572644	European ancestry		0.28 [d]	Chr2:28320033		- [e]	5.80×10^{-6}			

Supplementary Table 2 - Overview of OC clinical outcome GWAS

Study	Associated SNPs			Population	No. cases/deaths (combined)	MAF	Locus	Gene	Overall risk		Subtype-specific risk		
									HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	Histological subtype
Fridley *et al*. 2016 (cont.)	Paclitaxel	Caspase	rs75314082	European ancestry	74 cases	0.09 [d]	Chr2:55087315	*EML6*	- [e]	$7.90x10^{-5}$			
	Combination	MTT	rs17046344	European ancestry		0.17 [d]	Chr2:55023600		+ [f]	$4.90x10^{-5}$			
	Paclitaxel	Caspase	rs72817940	European ancestry		0.09 [d]	Chr2:58998563	*LINC01122*	+ [f]	$6.40x10^{-5}$			
	Carboplatin	Caspase	rs4233974	European ancestry		0.38 [d]	Chr2:59295043		- [e]	$2.60x10^{-5}$			
	Carboplatin	MTT	rs17261321	European ancestry		0.39 [d]	Chr2:80197843	*CTNNA2*	+ [f]	$3.60x10^{-5}$			
	Combination	MTT	rs6719499	European ancestry		0.31 [d]	Chr2:80193386		- [e]	$6.00x10^{-5}$			
	Paclitaxel	MTT	rs1525599	European ancestry		0.12 [d]	Chr2:141778702	*LRP1B*	+ [f]	$8.60x10^{-5}$			
	Combination	Caspase	rs13020675	European ancestry		0.26 [d]	Chr2:142212928		- [e]	$6.20x10^{-5}$			
	Paclitaxel	Caspase	rs201083182	European ancestry		Del [d]	Chr6:65736914	*EYS*	- [e]	$2.30x10^{-6}$			
	Combination	Caspase	rs2064701	European ancestry		0.17 [d]	Chr6:65676556		+ [f]	$3.60x10^{-5}$			
	Paclitaxel	Caspase	rs550987	European ancestry		0.22 [d]	Chr6:124905510	*NKAIN2*	- [e]	$4.10x10^{-5}$			
	Combination	Caspase	rs670616	European ancestry		0.28 [d]	Chr6:124885773		+ [f]	$7.80x10^{-5}$			
	Carboplatin	Caspase	rs10230114	European ancestry		0.29 [d]	Chr7:47705506	*C7orf65*	+ [f]	$2.40x10^{-5}$			
	Combination	Caspase	rs11771997	European ancestry		0.43 [d]	Chr7:47712495		+ [f]	$2.40x10^{-5}$			
	Paclitaxel	Caspase	rs12572446	European ancestry		0.49 [d]	Chr10: 47665906	*ANTXRL*	+ [f]	$4.30x10^{-5}$			
	Combination	Caspase	rs10906942	European ancestry		0.50 [d]	Chr10: 47670851		+ [f]	$4.90x10^{-5}$			
	Carboplatin	Caspase	rs10999018	European ancestry		0.06 [d]	Chr10: 71654602	*COL13A1*	+ [f]	$2.40x10^{-5}$			
	Combination	Caspase	rs77535242	European ancestry		0.06 [d]	Chr10: 71652985		+ [f]	$3.50x10^{-5}$			
	Paclitaxel	Caspase	rs77438645	European ancestry		0.07 [d]	Chr12:130304313	*TMEM132D*	- [e]				
	Carboplatin	Caspase	rs1451904	European ancestry		0.39 [d]	Chr12:130166947		+ [f]	$6.50x10^{-5}$			
	Carboplatin	Caspase	rs690089	European ancestry		0.40 [d]	Chr18: 8845223	*MTCL1*	- [e]	$7.80x10^{-5}$			
	Combination	Caspase	rs35765215	European ancestry		0.16 [d]	Chr18:8839469		- [e]	$6.00x10^{-5}$			

The results shown in this table are the combined phases results of each study or the validation phases results (when conducted).
a: per-allele; b: P_{trend}; c: P_{meta}; d: MAF values obtained on "Ensembl" database; e: (-), carriers of the minor/variant allele are, on average more sensitive to chemotherapeutic agents (lower IC50 or EC50); f: (+), carriers of the minor/variant allele are, on average more resistant to chemotherapeutic agents (higher IC50 or EC50).
ǂ: replicated SNPs; #: validated SNPs

Appendix 3

Supplementary Table 3 - Overview of molecular pathways which susceptibility associated SNPs are known to be involved

Study	SNP	Gene	Molecular pathway [a]	Functional consequence
Song *et al*. 2009	rs3814113	~44 kb upstream of *BNC2* gene		Deletion of 5 kb surrounding rs3814113 decreased *BNC2* expression levels [204]
Bolton *et al*. 2010	rs8170	*MERIT40*	DNA double strand break response; Cell cycle checkpoints	
	rs2363956	*ANKLE1*	Human lymphocyte development [205]	
Goode *et al*. 2010	rs2665390	*TIPARP*	Aryl hydrocarbon receptor signaling [206]	
	rs9303542	*SKAP1*	B-cell receptor signaling pathway; RAS signaling pathway	
Couch *et al*. 2013	rs17631303	*PLEKHM1*	Osteoclast function regulation [207]; Bone resorption [207]; Endocytic and autophagy pathways [207]	
	rs183211	*NSF*	Delta508-CFTR (cystic fibrosis transmembrane conductance regulator) traffic / Sorting endosome formation in CF (cystic fibrosis); Trafficking of AMPA receptors; Vasopressin-regulated water reabsorption	
Pharoah *et al*. 2013	rs11782652	*CHMP4C*	HIV life cycle; MTOR signaling; Endocytosis	Association between rs11782652 and *CHMP4C* overexpression in primary EOC tissues and LCLs [139].
	rs1243180	*MLLT10*	Leukemogenesis [208]	Association between rs1243180 and *C10orf114* and *SKIDA1* expression in primary EOC tissues [139].
	rs757210	*HNF1B*	Regulation of β-cell development; Type II Diabetes Mellitus; Hepatic ABC transporters	Association between the minor allele of rs757210 and overexpression of *HNF1B* in serous EOC tissues [139].
Earp *et al*. 2014	rs17106154	~150 kb in LD region of *ZFP36L1* (*BRF1*) gene	Validated targets of C-MYC transcriptional repression; PI3K / Akt signaling; Translational control	
	rs2190503	Identify a locus upstream/intronic to *GRB10* gene	IGF1 pathway; Signaling events regulated by Ret tyrosine kinase; Insulin pathway	
	rs6593140			
	rs2329554			

Supplementary Table 3 - Overview of molecular pathways which susceptibility associated SNPs are known to be involved

Study	SNP	Gene	Molecular pathway [a]	Functional consequence
Earp *et al*. 2014 (cont.)	rs9609538	~5 bp downstream of *BPIL2* gene and ~500 bp upstream of *C22orf28*		The minor allele of this SNP is predicted to alter transcription factor binding site activity and miRNA binding site activity [141].
Chen *et al*. 2014	rs1413299	*COL15A1*	Protein digestion and absorption; Collagen biosynthesis and modifying enzymes; Degradation of the extracellular matrix	
	rs11175194	*SRGAP1*	Signaling by Robo receptor; Regulation of RhoA activity; Signaling by Slit	
Kuchenbaecker *et al*. 2015	rs56318008	*WNT4*	Negative regulation of TCF-dependent signaling by WNT ligand antagonists; WNT ligand biogenesis and trafficking; Mesenchymal Stem Cell differentiation pathways and lineage-specific markers	No effect on *WNT4* transcription in OC cells [143]
	rs58722170	*RSPO1*	Regulation of FZD by ubiquination; WNT signaling; Signaling by GPCR	
	rs17329882	*SYNPO2*	Actin binding protein [209]	
	rs116133110	*GPX6*	Folate metabolism; Detoxification of Reactive Oxygen Species; Selenium metabolism and selenoproteins	
	rs199661266	*ATAD5*	DNA damage response [143]	
Keleman et al	rs752590	*PAX8*	ID signaling pathway; Thyroid cancer; TSH signaling pathway	
	rs711830	*HOXD3*	Activation of HOX genes during differentiation; Developmental Biology	
	rs688187	*IFNL3*	Peginterferon alpha-2a/Peginterferon alpha-2b Pathway; all-trans-Retinoic Acid Mediated Apoptosis; RIG-I/MDA5 mediated induction of IFN-alpha/beta pathways	
Phelan *et al*. 2017	rs555025179	*MAST4*	Microtubule scaffolding [210]	
	rs8098244	*LAMA3*	Syndecan-family-mediated signaling events; Validated transcriptional targets of AP1 family members Fra1 and Fra2; Alpha 6 Beta 4 signaling pathway	

a: data obtained from “GeneAnalytics” database (exceptions are referenced)

Appendix 4

Supplementary Table 4 - Overview of molecular pathways which clinical outcome associated SNPs are known to be involved

Study	SNP	Gene	Molecular pathway	Functional consequence
Bolton *et al*. 2010	rs8170	*MERIT40*	DNA double strand break response; Cell cycle checkpoints	
	rs2363956	*ANKLE1*	Human lymphocyte development [205]	
Huang *et al*. 2011	rs1649942	*NRG3*	ErbB4 signaling events; Signaling by ErbB2; Agrin Interactions at Neuromuscular Junction	This SNP is associated with baseline expression of 18 genes [136]
Johnatty *et al*. 2015	rs7950311	*HBG2*	Factors involved in megakaryocyte development and platelet production; IL-2 pathway; p70S6K signaling	
	rs3795247	*ZNF100*	Gene expression	
French *et al*. 2016	rs72700653	*TTC39B*	Mediation of the association of HDL-regulating proteins [211]	The minor alleles of these SNPs enhance expression of the non-canonical *TTC39B* promoter [144]
	rs7874043			
Fridley *et al*. 2016	rs185229225	*BOD1L1*	Replication fork protection factor [151]	
	rs3842595	*SNCAIP/MGC32805*	Putative ubiquitin pathway; Parkinson's disease pathway; Parkin-Ubiquitin proteasomal system pathway	
	rs150303591	*FRAS1*	Phospholipase-C pathway; Integrin pathway; ERK signaling	
	rs201023017	*SLC9A9*	Sweet taste signaling; Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds	
	rs66696671	*TIAL1*	Formation of the HIV-1 Early Elongation Complex; Apoptosis and autophagy; Translational control	
	rs10674174	*PCDH20*	WNT signaling (Antagonist) [212]	

Supplementary Table 4 - Overview of molecular pathways which clinical outcome associated SNPs are known to be involved

Study	SNP	Gene	Molecular pathway	Functional consequence
Fridley *et al*. 2016 (cont.)	rs35067965	*COLEC12*	Binding and uptake of ligands by Scavenger receptors; Phagosome; Vesicle-mediated transport	
	rs8091660	*CTIF*	Translation initiation [213]	
	rs113867814			
	rs2748151	*CDH4*	CDO in myogenesis; Natural Killer cell receptors; S-1P stimulated signaling	
	rs113594423			
	rs5830067	*BRE*	Apoptosis and survival-caspase cascade; DNA double strand break response [145]; TWEAK pathway	
	rs7572644			
	rs75314082	*EML6*	Microtubules dynamics regulation [145]	
	rs17046344			
	rs1525599	*LRP1B*	Metabolic health programming [214]	
	rs13020675			
	rs17261321	*CTNNA2*	CDO in myogenesis; Adhesion [145]; Adherens junction	
	rs6719499			
	rs10999018	*COL13A1*	Articular cartilage extracellular matrix pathway; T Cell co-signaling pathway: ligand-receptor interactions; Natural Killer cell receptors	
	rs77535242			
	rs690089	*MTCL1*	Microtubule dynamics regulation [215]	
	rs35765215			

a: data obtained from “GeneAnalytics” database (exceptions are referenced)

Appendix 5

A paper entitled *Rethinking ovarian cancer genomics: where GWAS stand?* has been submitted and accepted for publication in the scientific journal *Pharmacogenomics*.