Novel biomarkers associated with histotype and clinical outcome in early-stage ovarian carcinoma

Hanna Engqvist

Department of Oncology
Institute of Clinical Sciences
Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2020

Cover illustration: Hanna Engqvist

Novel biomarkers associated with histotype and clinical outcome in early-stage ovarian carcinoma

hanna.engqvist@gu.se

ISBN 978-91-7833-896-2 (PRINT)
ISBN 978-91-7833-897-9 (PDF)
http://hdl.handle.net/2077/63246

Printed in Borås, Sweden 2020
Printed by Stema Specialtryck AB

ABSTRACT

Ovarian cancer is a collective name for multiple malignancies deriving from or involving the ovary, mainly comprising five histotypes of epithelial origin (clear-cell (CCC), endometrioid (EC), high-grade serous (HGSC), low-grade serous (LGSC), and mucinous carcinomas (MC)) with varying clinical (*e.g.* risk factors, survival outcome, response to therapy) and molecular behavior (*e.g.* origin, genetic characteristics). Despite known differences in disease states, the majority of ovarian carcinomas are still treated as one entity with surgery, followed by chemotherapy. This treatment regimen is not adequate, which is reflected in relatively poor 5-year overall survival rates (55%) for ovarian cancer patients. Hence, there is a strong need for novel biomarkers for improved stratification of ovarian carcinoma patients based on a combination of individual molecular tumor characteristics and conventional clinicopathological features, which can further form the basis for the future development of novel targeted treatment options for ovarian cancer histotypes.

This doctoral thesis focuses on early-stage (stage I and II) ovarian carcinomas for which limited information is available regarding molecular profiles associated with the diagnosis and prognosis of the different histotypes. In the first work, novel mutation and gene signatures were associated with histotype, overall survival (*e.g.* the tumor suppressor *MTUS1*), ovarian cancer (*e.g.* gene expression patterns for the long non-coding RNA *MALAT1*), and tumor aggressiveness (*e.g. COL3A1*). In the second and third works, histotype-specific prognostic gene signatures were validated on the protein level using immunohistochemistry identifying 20 prognostic biomarkers (11 CCC-associated biomarkers (ARPC2, CCT5, GNB1, KCTD10, NUP155, PITHD1, RPL13A, RPL37, SETD3, SMYD2, and TRIO), three EC-associated biomarkers (CECR1, KIF26B, and PIK3CA), five MC-associated biomarkers (CHEK1, FOXM1, GPR158, KIF23, and PARPBP), and COL3A1 for the main histotypes). In the fourth work, a multi-omics approach (genome- and transcriptome-wide analyses) integrating DNA methylation, DNA copy number alteration, and RNA sequencing data was applied to identify novel putative oncogenes and tumor suppressor genes associated with the CCC, EC, HGSC and MC histotypes.

Taken together, the current doctoral thesis presents novel insights into molecular features associated with early-stage ovarian carcinoma that may improve patient stratification and subclassification based on histotype and clinical outcome.

Keywords: ovarian carcinoma, histotype-specific diagnosis and prognosis, molecular biomarker, outcome prediction, integrative analysis

SAMMANFATTNING PÅ SVENSKA

Äggstockscancer är ett samlingsnamn för ett stort antal maligniteter som härstammar från eller involverar äggstocken, och delas huvudsakligen in i fem histotyper som utvecklas från epitelial vävnad (klarcelligt (CCC), endometrioidt (EC), höggradigt seröst (HGSC), låggradigt seröst (LGSC) och mucinöst karcinom (MC)), med varierande kliniska (t.ex. riskfaktorer, överlevnad, behandlingssvar) och molekylära särdrag (t.ex. härkomst, genetiska särdrag). Trots mångårig vetskap om variabelt kliniskt utfall hos de olika undergrupperna, kvarstår i stort sett samma behandlingsstrategi med kirurgi följt av cellgiftsbehandling. För vissa tumörgrupper är denna behandling effektiv men för andra är den mindre verksam, vilket avspeglas i den relativt ringa 5-årsöverlevnaden för äggstockscancer på 55%. Bakgrunden till de olika behandlingssvaren är intensivt studerat men har ännu inte till fullo klarlagts. Det finns därför ett stort behov av nya biomarkörer som bättre kan stratifiera patienter med äggstockskarcinom baserat på en kombination av individuella molekylära särdrag hos tumören och traditionella kliniska och patologiska särdrag. Biomarkörerna kan vidare utgöra en kunskapsgrund för utvecklingen av nya framtida inriktade behandlingsalternativ för äggstockscancer.

Denna doktorsavhandling avser tidiga stadier (stadium I och II) av äggstockskarcinom för vilka det finns begränsat med information avseende molekylära profiler som förknippas med diagnos och prognos av de olika histotyperna. I det första arbetet sammankopplas nya mutations- och gensignaturer med histotyp, total överlevnad (t.ex. tumörsuppressorgenen *MTUS1*), äggstockscancer (t.ex. genuttryck för *MALAT1* som är ett långt ickekodande RNA), och tumöragressivitet (t.ex. *COL3A1*). I det andra och tredje arbetet validerades histotyp-specifika prognostiska gensignaturer på proteinnivå med hjälp av immunohistokemi, varvid 20 prognostiska biomarkörer identifierades (elva CCC-associerade biomarkörer (ARPC2, CCT5, GNB1, KCTD10, NUP155, PITHD1, RPL13A, RPL37, SETD3, SMYD2, och TRIO), tre EC-associerade biomarkörer (CECR1, KIF26B, och PIK3CA), fem MC-associerade biomarkörer (CHEK1, FOXM1, GPR158, KIF23, och PARPBP), samt COL3A1 för de mest förekommande histotyperna). I det fjärde arbetet integrerades data från flera olika typer av analyser (heltäckande genomiska och transkriptomiska analyser) som innefattade data från DNA-metylering, DNA-avvikelser och RNA-sekvensering för att identifiera nya möjliga onkogener och tumörsuppressorgener som är förknippade med histotyperna CCC, EC, HGSC och MC.

Sammantaget ger denna doktorsavhandling ökad förståelse kring molekylära särdrag som förknippas med tidiga stadier av äggstockskarcinom som kan förbättra klassificering och indelning i undergrupper baserat på histotyp och överlevnad.

LIST OF PAPERS

The thesis is based on the following studies (Paper I-IV), referred to in the text by their Roman numerals.

I. **Engqvist H**, Parris TZ, Werner Rönnerman E, Söderberg E, Biermann J, Mateoiu C, Sundfeldt K, Kovács A, Karlsson P, Helou K. Transcriptomic and genomic profiling of early-stage ovarian carcinomas associated with histotype and overall survival. *Oncotarget* (2018). DOI: 10.18632/oncotarget.26225

II. **Engqvist H**, Parris TZ, Kovács A, Nemes S, Werner Rönnerman E, De Lara S, Biermann J, Sundfeldt K, Karlsson P, Helou K. Immunohistochemical validation of COL3A1, GPR158 and PITHD1 as prognostic biomarkers in early-stage ovarian carcinomas. *BMC Cancer* (2019). DOI: 10.1186/s12885-019-6084-4

III. **Engqvist H**, Parris TZ, Kovács A, Werner Rönnerman E, Sundfeldt K, Karlsson P, Helou K. Validation of novel prognostic biomarkers for early-stage clear-cell, endometrioid and mucinous ovarian carcinomas using immunohistochemistry. *Frontiers in Oncology*, section Women's Cancer (2020). DOI: 10.3389/fonc.2020.00162

IV. **Engqvist H**, Parris TZ, Biermann J, Werner Rönnerman E, Larsson P, Sundfeldt K, Kovács A, Karlsson P, Helou K. Integrative genomics approach identifies molecular features associated with early-stage ovarian carcinoma histotypes. Submitted (2020).

The following publications are not included in the thesis but are of relevance to the field.

i. Larsson P, **Engqvist H**, Biermann J, Werner Rönnerman E, Forsell-Aronsson E, Kovács A, Karlsson P, Helou K, Parris TZ. Optimization of cell viability assays to improve replicability and reproducibility of cancer drug sensitivity screens. *Scientific Reports* (2020). In press.

ii. Biermann J, Nemes S, Parris TZ, **Engqvist H**, Werner Rönnerman E, Kovács A, Karlsson P, Helou K. A 17-marker panel for global genomic instability in breast cancer. *Genomics* (2019). DOI: 10.1016/j.ygeno.2019.06.029.

iii. Biermann J, Langen B, Nemes S, Holmberg E, Parris TZ, Werner Rönnerman E, **Engqvist H**, Kovács A, Helou K, Karlsson P. Radiation-induced genomic instability in breast carcinomas of the Swedish hemangioma cohort. *Genes, Chromosomes and Cancer* (2019). DOI: 10.1002/gcc.22757.

iv. Parris TZ, Rönnerman E, **Engqvist H**, Biermann J, Truvé K, Nemes S, Forssell-Aronsson E, Solinas G, Kovács A, Karlsson P, Helou K. Genome-wide multi-omics profiling reveals extensive genetic complexity in 8p11-p12 amplified breast carcinomas. *Oncotarget* (2018). DOI: 10.18632/oncotarget.25329.

v. Biermann J, Parris TZ, Nemes S, Danielsson A, **Engqvist H**, Werner Rönnerman E, Forssell-Aronsson E, Kovács A, Karlsson P, Helou K. Clonal relatedness in tumour pairs of breast cancer patients. *Breast Cancer Research* (2018). DOI: 10.1186/s13058-018-1022-y.

vi. Biermann J, Nemes S, Parris TZ, **Engqvist H**, Werner Rönnerman E, Karlsson P, Forssell-Aronsson E, Steineck G, Helou K. A novel 18-marker panel predicting clinical outcome in breast cancer. *Cancer Epidemiology, Biomarkers & Prevention* (2017). DOI: 10.1158/1055-9965.EPI-17-0606.

CONTENT

ABBREVIATIONS

CTLP	Chromothripsis-like pattern
CCC	Clear-cell ovarian carcinoma
CNA	Copy number alteration
DEG	Differentially expressed gene
DMP	Differentially methylated probe
DSS	Disease-specific survival
EC	Endometrioid ovarian carcinoma
FIGO	International Federation of Gynecology and Obstetrics
FISH	Fluorescence *in situ* hybridization
FFPE tumor block	Formalin-fixed paraffin-embedded tumor block
GATK	Genome analysis toolkit
GDC Data Portal	Genomic Data Commons Data Portal
HGSC	High-grade serous ovarian carcinoma
HRD	Homologous recombination deficiency
IHC	Immunohistochemistry
KM plotter	Kaplan-Meier plotter
LGSC	Low-grade serous ovarian carcinoma
MC	Mucinous ovarian carcinoma
OS	Overall survival
PARP inhibitor	Poly ADP ribose polymerase inhibitor
RNA-seq	RNA sequencing
SNP	Single nucleotide polymorphism
TCGA	The Cancer Genome Atlas
TMA	Tissue microarray
UPPMAX	Uppsala Multidisciplinary Center for Advanced Computational Science

INTRODUCTION

CANCER

Cancer initiation and progression are multi-step processes involving the accumulation of genetic alterations in the descendants of a single somatic cell [1]. Cancer development may, therefore differ between individuals depending on their genetic predisposition (inherited mutations), differences in acquired somatic mutations, and exposure to environmental and/or stochastic factors. Hence, cancer is a genetically complex disease, wherein multiple genes interact and different genes can give rise to the malignant tumor [2]. Cancer can be classified into approximately 200 different cancer types according to the tissue of origin, *e.g.* skin cancer starts in the cells of the skin [3]. Moreover, cancer is defined by its abnormal cell growth, local invasiveness and ability to spread to other parts of the body than the site of origin [4].

Worldwide, cancer is the first or second leading cause of death in the majority of countries (134/183 countries), with an estimated 9.6 million deaths in 2018 [5,6]. Both cancer incidence (number of new cases) and mortality (number of deaths) are increasing due to *e.g.* aging, improved detection, a growing population and exposure to cancer risk factors due to socioeconomic development. Lung cancer has the highest incidence (11.6% of all cancers) and mortality (18.4% of all cancers) in both sexes, followed by female breast cancer, and prostate cancer in view of incidence, and colorectal cancer, and stomach cancer in view of mortality [5]. In females, breast cancer has the highest incidence and mortality followed by colorectal and lung cancer (incidence) and inversely for mortality. In males, lung cancer remains as the most commonly diagnosed cancer and the main cause of death, followed by prostate and colorectal cancer for incidence and liver and stomach cancer for mortality [5]. However, cancer statistics may differ substantially between countries. Furthermore, it is important to keep in mind that the worldwide cancer incidence and mortality rates are to some extent estimates due to lack of high quality national incidence and mortality data for many countries, *e.g.* only 34/194 (17.5%) and 68/134 (50.7%) of the WHO member states provide high quality national incidence and mortality data, respectively [7]. In Sweden, the highest overall cancer incidence is reported for prostate cancer (16.3%) followed by breast cancer (14%) [8].

Cancer genetics and epigenetics

During cancer development and progression, significant genetic aberrations, *e.g.* point mutations, deletions, inversions, translocations, copy number alterations (CNA) or epigenetic modification, tend to affect three main types of genes, namely proto-oncogenes (*e.g. ERBB2* (HER2/*neu*)), tumor suppressor genes (*e.g. TP53*), and DNA repair genes (*e.g. BRCA1, BRCA2*) [9,10]. Proto-oncogenes are involved in driving normal cell growth and division; overexpression via mutations, structural rearrangement or DNA amplification can result in the activation of proto-oncogenes to oncogenes, *i.e.* a gene that lead to uncontrolled cell growth and resistance to cell death. Tumor suppressor genes are involved in inhibiting cell growth and division and if mutated may result in inactivation of the gene, which may lead to uncontrolled cell growth. DNA repair genes are involved in repairing the DNA of damaged cells. If they become mutated they tend to be genetic drivers of carcinogenesis, and since their repairing ability is disabled it may result in additional mutations in other genes [9,10]. There are two main groups of mutations associated with cancer, namely driver mutations and passenger mutations. Driver mutations are mutations that tend to give cells a selective growth advantage, whereas tumor-associated genomic instability gives rise to passenger mutations that are bystanders and do not influence tumor progression. A typical tumor is reported to comprise two to eight driver mutations [11,12]. The total number of mutations in different cancer types depend on *e.g.* patient age and cell division rates in different cell types (a higher cell division rate results in a higher accumulative risk of acquiring additional mutations) [13].

Epigenetic modifications are heritable alterations that do not alter the DNA sequence, but nevertheless have an effect on gene expression levels (enhanced or reduced) by altering chromatin organization and gene accessibility for the transcriptional machinery [14]. One type of epigenetic modulation is DNA methylation, whereby methyl groups are added or removed from DNA CpG sites, *i.e.* sites rich in CG nucleotides. In cancer, hypermethylation (more DNA methylation than normal) often occurs in promoter regions of tumor suppressor genes, resulting in gene silencing, whereas hypomethylation (less methylation than normal) often occurs in oncogenes, leading to increased expression thereof [15]. The complex cancer landscape of somatic structural rearrangements may accumulate over time in a multistep process or as recently described through one single catastrophic event called chromothripsis (2-3% in cancer). Chromothripsis refers to the shattering of chromosomes, which results in massive gene reshuffling followed by random reassembly characterized by DNA copy number status changes. The exact cause of chromothripsis is not known, but may be due to ionizing radiation [16,17].

Personalized diagnosis and treatment

Traditionally, similar treatment regimens have generally been administered to patients exhibiting similar clinicopathological characteristics. However, two patients with similar characteristics may respond differently to the same treatment. Therefore, the development in recent years has moved towards establishing patient-specific diagnosis and treatment strategies to determine which patients would most benefit from a specific therapy to reduce tumor burden and optimize survival outcomes, while minimizing side effects. The advancements in high-throughput technologies covering *e.g.* genome- and transcriptome-wide analyses make it possible to characterize molecular tumor profiles of individual patients, which in combination with clinicopathological features used in the clinic, could improve correct diagnosis. Such profiles may further contribute to novel molecular findings, which may be the basis for novel targeted treatment options in the future [18,19].

OVARIAN CANCER

Ovarian cancer is a group of malignancies that derives from the ovary, fallopian tube or the peritoneum. In 2016, 541 women were diagnosed with ovarian cancer in Sweden [20]. Worldwide, ovarian cancer is the eighth most common cause of death among women, with an estimated 295,414 new cases (corresponding to 3.4% of all cancers in women) in 2018 [6]. Due to its asymptomatic disease progression and lack of effective screening strategies, ovarian cancer is often referred to as a silent killer with 62% of ovarian cancers diagnosed at late stages (stage III and IV) in Sweden [20]. This is further reflected in an overall unfavorable prognosis of ovarian cancer with 5-year survival rates of 55% [8]. Both incidence and mortality rates vary by region, with *e.g.* the highest incidence rates in Europe and the lowest in Africa [21].

Pathologic classification

There are over 30 different subtypes of primary ovarian cancer, mainly distributed in three subcategories depending on its cell of origin: epithelial (carcinomas, >90%), sex-cord stromal (5-6%), and germ cell ovarian cancers (2-3%) [22]. Epithelial cells line the surface (outer layer) of the ovary, stromal cells mainly have a supportive and hormone producing (theca cells) function, and germ cells produce eggs. Ovarian carcinomas are further subdivided into histotypes, wherein the five main histotypes constitute more than 95% of ovarian

carcinomas, namely clear-cell (CCC), endometrioid (EC), high-grade serous (HGSC), low-grade serous (LGSC) and mucinous ovarian carcinomas (MC) [23]. Ovarian carcinomas also include *e.g.* undifferentiated carcinomas and malignant Brenner tumors. The main histotypes may also be stratified into two groups based on their level of cell differentiation, *i.e.* how well they correspond to a normal differentiated cell. HGSC, which is poorly differentiated, is referred to as type II ovarian carcinoma and the remaining four histotypes are generally well differentiated (type I) [24].

Historically, it was thought that all ovarian carcinomas originated from the ovarian surface epithelium, which is the least common cell type in the ovary. This fact made researchers question the actual biological origin of ovarian carcinomas. In recent years, it has then been shown that ovarian cancers primarily originate from outside of the ovary and involve the ovary in secondary events. For example, a large proportion of HGSCs (up to 70%) originate from fallopian tube epithelium [25]. Furthermore, EC and CCC may derive from endometriosis tissue that originates in endometrial epithelial cells, whereas it is not currently known from where MC originates [6]. It is being hypothesized that MCs derive from colorectal mucosa or the tubal peritoneal junction [22]. There are different theories from where LGSC originates, *e.g.* ovarian surface epithelium or fallopian tube [26].

Risk factors

Similar to other cancer forms, age is a significant risk factor. Moreover, the number of menstrual cycles during a female's lifetime is an established risk factor for ovarian cancer, which is associated with increased cell division and number of spontaneous mutations due to the repair of the surface epithelium after each ovulation [27]. Further, ovulation may also contribute to ovarian cancer initiation by the release of cytokines and growth factors due to inflammation [28]. Factors that reduce the number of menstrual cycles, such as the use of contraceptives, pregnancies and young age at menopause, also reduces the risk of ovarian cancer [29]. The risk differs when stratified by histotype, wherein the use of oral contraceptives (>5 years, >10 years) has been linked to a lower risk (14-15%, 36-49%) of developing CCC, EC and serous ovarian carcinomas (SC, HGSC and LGSC), but not for MC. The largest reduction of risk due to parity (*e.g.* the number of pregnancies carried 20 weeks or longer) has been found for CCC and EC (about 50-65%), while a slightly lower reduction was found for MC (44%) and the lowest risk reduction for SC (about 20%). A further risk reduction of about 15% was found for each further full-term pregnancy. A 5-year later menopause was also

shown to be associated with increased risk for developing CCC, EC and SC, but not for MC. Smoking is a significant risk factor for developing MC [27].

Family history based on the genetic predisposition, *i.e.* the likelihood to develop ovarian cancer based on the heritable characteristics, is one of the most important risk factors for ovarian cancer with an elevated risk for all histotypes except for MC [30]. Germline *BRCA1* and *BRCA2* mutations, *i.e. BRCA1* and *BRCA2* mutations in germ cells that are inherited by offspring, constitute an increased risk of developing ovarian and breast cancer. These germline mutations are found in up to 15% of ovarian cancers and up to 23% in HGSC [31,32]. Although the risk of developing ovarian cancer may vary by mutation type and location within the *BRCA* gene. *BRCA1* and *BRCA2* are tumor suppressor genes that normally protect the genome from DNA-damage and therewith ensure stability of the genetic material. *BRCA1* mutation carriers have an increased risk (16-68%) of developing ovarian cancer, and 11-30% for *BRCA2* mutation carriers. Moreover, a family history of breast cancer is further associated with an increased risk of ovarian cancer [33]. Endometriosis, a condition in which endometrium grows outside of the uterus, *e.g.* on the ovaries or fallopian tubes, has been shown to increase the risk of developing CCC, EC and LGSC [34]. Moreover, Lynch syndrome has been found to result in an increased risk of developing EC and CCC due to germline mutations in DNA mismatch repair genes (*e.g.* MLH1, MSH2, MSH6, PMS2) [35].

Screening strategies

Many ovarian cancer patients are diagnosed at late stages. Incidence rates vary significantly among the different stages, with 27% of all ovarian cancer patients are diagnosed at stage I, 9% at stage II, 46% at stage III, and 16% at stage IV (2% at an unknown stage) in Sweden [20]. Prognosis is more favorable for patients diagnosed at an early stage, but unfortunately there are currently no effective screening strategies available for early detection of ovarian cancer. To date, screening strategies are based on transvaginal ultrasound imaging in combination with blood-based biomarkers, such as CA125. However, this screening combination has shown no significant reduction in mortality [36-38]. In order to improve diagnostics for ovarian cancer, recent research has focused on examining samples collected closer to the ovaries, *e.g.* from the uterine cavity by uterine lavage or with Pap smears from the cervix. More specifically, samples were screened for known mutations associated with ovarian cancer using sequencing technologies with a 60% and 41% detection rate, respectively [39,40]. Another study combined the examination of circulating tumor DNA from a liquid biopsy taken from blood with a Pap smear resulting in a detection rate of 63% [41].

Risk of malignancy index (RMI) is a diagnostic tool used in the clinic prior to surgery to determine the likelihood whether an adnexal mass is benign or malignant. The index is based on three variables; CA125, menopause status and ultrasound points (based on *e.g.* size of mass, ascites, bilateral tumors, metastasis) [20].

Prognosis

Ovarian cancer holds the highest mortality rates among gynecological cancers. Since the 1980s, a modest improvement in overall survival rates has been achieved, wherein the 5-year overall survival rates for ovarian cancer have increased from about 38% to 55% and 10-year survival rates from about 32% to 43% in Sweden [8]. Tumor stage at the time of diagnosis is currently the most important prognostic factor used in the clinic, wherein 5-year survival rates stratified by stage are 89% for stage I, 71% for stage II, 41% for stage III, and 20% for stage IV in the USA [42,43]. Furthermore, a more favorable prognosis is seen if the patient is macroscopically tumor-free after surgery [42,44]. According to the American Cancer Society, 5-year survival rates stratified by histotype were lowest for SC (43%; typically diagnosed at an advanced stage) and significantly better survival rates for CCC (66%), EC (82%) and MC (71%) (typically diagnosed at an early stage). A recent report examining the prognostic relevance for early-stage ovarian carcinomas further supports these findings with EC being the most favorable histotype, whereas HGSC and LGSC had the most unfavorable prognoses [45]. Patients with EC or MC histotype in stage Ia and Ib have also been reported to have a very favorable prognosis (10-year disease-specific survival (DSS) ≥ 95%) [46]. In the clinic, the only prognostic biomarkers that are currently used are for defects in homologous recombination deficiency (HRD), *i.e.* when the cell is unable to repair DNA double-stranded breaks using homologous recombination, *e.g.* caused by *BRCA1* or *BRCA2* mutations. Patients with HRD are eligible to be treated with Poly ADP ribose polymerase (PARP) inhibitors, which have been shown to prolong ovarian cancer survival, especially in HGSC patients that to a large extent (about 50%) have been reported to harbor mutations in the HR pathway [47,48].

Staging

Ovarian cancer is staged according to tumor size and/or metastatic spread of the cancer. Stage has treatment implications and may help in the prognostication of the disease. As stated above, a lower stage usually indicate a better prognosis with a more favorable patient outcome. There are currently two staging systems

available for gynecological cancer, *i.e.* The International Federation of Gynecology and Obstetrics (FIGO) and the American Joint Committee on Cancer (AJCC) TNM (Tumor, Nodes, Metastasis) staging system. In the TNM system, the tumor is classified separately for primary tumor (T) size and organ extension, spread to the lymph nodes (N) and distant metastasis (M), whereas the FIGO system summarizes these three parameters into stage I to IV with further stratification into substages (Figure 1) [22,49]. For gynecological cancers, the FIGO system is the most commonly used [20,50]. If possible, the primary site of the tumor, *i.e.* where the malignant tumor originated, is determined. Furthermore, the histotype should be determined at the time of staging [49].

Histology and molecular characteristics

The largest histotype group comprises HGSCs (about 70%), followed by EC (about 10%), CCC (5-10%), MC (about 3-4%) and LGSC (<5%) within epithelial ovarian cancers in Sweden [20]. Within early-stage ovarian cancers in a Canadian cohort, fewer samples were classified as HGSC (35.5%), whereas EC (26.6%), CCC (26.2%) and MC (7.5%) had higher incidence rates, and LGSC (1.9%) was relatively unchanged, in comparison with the overall incidence rates for stage I-IV [51]. There are further regional variations across the histotypes, *e.g.* in Japan, there was a lower overall proportion of SC (40.8%) and a higher proportion specifically for CCC (26.9%), but also for EC (19.2%) and MC (13.1%), in comparison to other regions [52]. As previously described, survival outcomes (prognosis) and genetic predisposition in terms of germline mutations differ between histotypes. Furthermore, the histotypes are histologically and molecularly distinct diseases with diverse genetic changes, *e.g.* somatic mutations (mutations that occur in any of the cells of the body besides the germ cells and are hence not inherited by offspring), DNA CNA, and epigenetic changes (Figure 2).

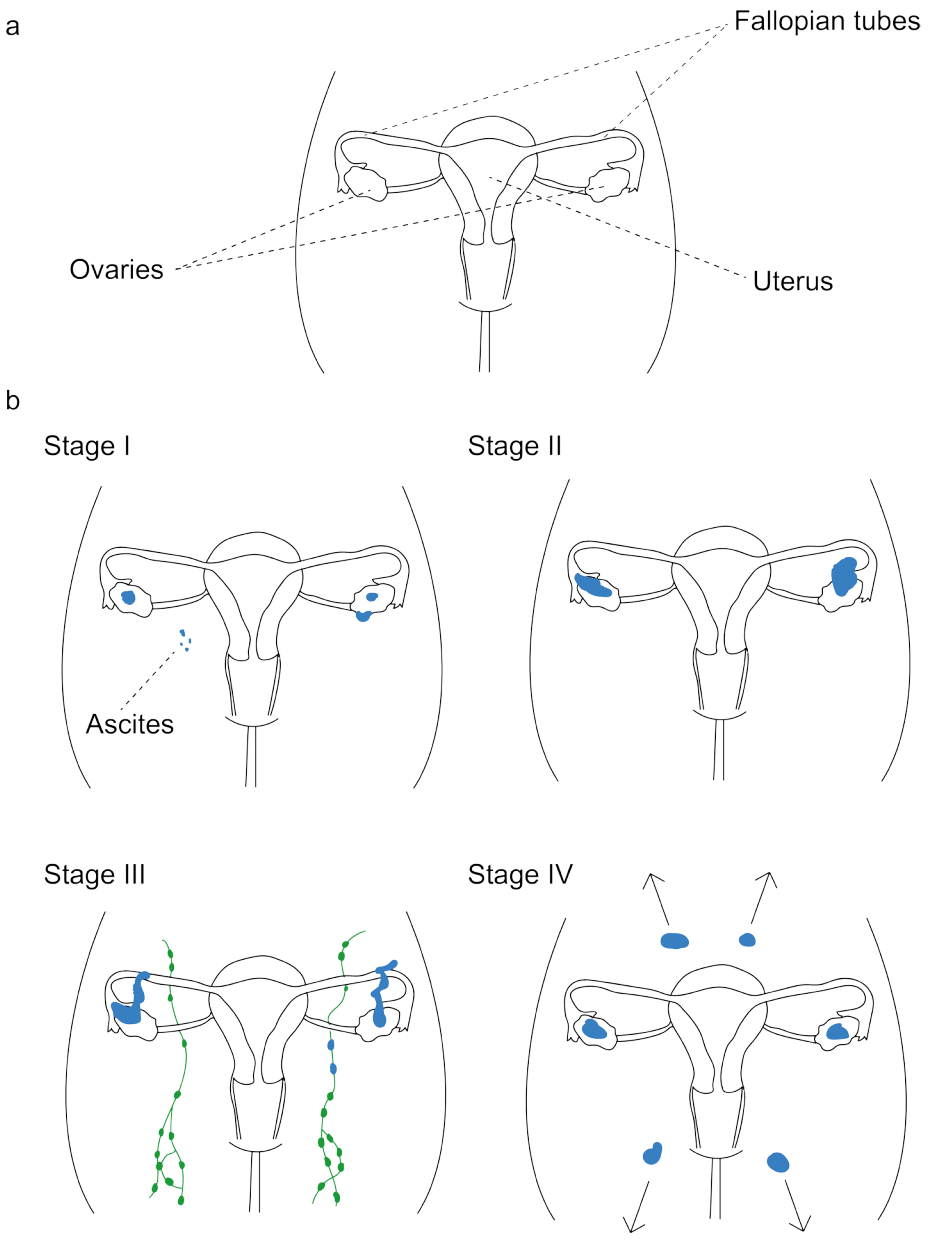


Figure 1. Illustration of ovarian cancer staging according to The International Federation of Gynecology and Obstetrics (FIGO) system. *Sketch of a healthy female reproductive system (uterus, fallopian tubes and ovaries) (a). The location of the primary tumor and metastatic spread (malignant tumor is illustrated in blue) for the different FIGO stages I to IV (b). In stage I, the cancer is confined to one or both ovaries/fallopian tubes. Malignant cells may also be found*

in the ascites, i.e. abnormal fluid in the peritoneal cavity. Ovarian cancer stage II implies metastatic spread to the uterus, fallopian tubes and/or other pelvic intraperitoneal tissues. Apart from one or both ovaries/fallopian tubes, the cancer can in stage III also be found in the peritoneum outside of the pelvis. The cancer may have also metastasized to the retroperitoneal lymph nodes (illustrated in green (healthy) and blue (cancerous)). In stage IV, the cancer has metastasized to distant organs, e.g. liver, lungs.

HGSC is characterized by frequent mutations in the *TP53* gene (96%). Low prevalence of recurrent somatic mutations in nine additional genes *e.g. BRCA1*, *BRCA2*, *RB1* and *CDK12*, have also been identified [53]. Apart from mutations in these genes, additional recurrent somatic mutations in oncogenes or tumor suppressor genes are relatively uncommon for HGSC. Instead HGSC is defined by genomic instability caused by *e.g.* widespread DNA CNA gains and losses such as in *CCNE1* (cyclin E1) (present in about 15% of HGSC) [53,54]. CNAs in *CCNE1* are reported to be an early event in HGSC tumorigenesis [55]. As previously described, about 50% of HGSCs are deficient in the HR DNA repair pathway [53]. This may be caused by germline, somatic and/or epigenetic aberrations in genes related to the HR pathway, such as *BRCA1* or *BRCA2*. HGSC may further constitute defects in genes of the Notch signaling pathway (about 22%), which is involved in multiple cellular processes such as cell proliferation, differentiation and apoptosis [54,56]. HGSC have been subdivided in four molecular types based on their gene expression profiles, namely C1/mesenchymal, C2/immune, C4/differentiated and C5/proliferative with differing clinical outcomes [53,57].

LGSCs are more stable and genomically homogeneous compared to HGSCs, and are characterized by mutations in the *KRAS*, *ERBB2* and *BRAF* oncogenes [20]. Moreover, the *BRAF* mutation has been reported to be a favorable prognostic factor for LGSCs [58]. In MC, *KRAS* mutations are commonly (40-50%) identified. Furthermore, *TP53* mutations (16-50%) and HER2*/neu* amplifications (20-30%) are also found. No association between MCs and *BRCA1* or *BRCA2* mutations has been reported [59]. EC and CCC are characterized by inactivating mutations (resulting in a gene product with less or no function) in *ARID1A* (about 30% and 50%, respectively), activating mutations (resulting in a gene product with enhanced function) in *PIK3CA* or inactivating mutations in *PTEN* [60]. In addition, EC may comprise *KRAS* mutations. Both EC and CCC have few *TP53* mutations and relatively stable genomes [6,20,23].

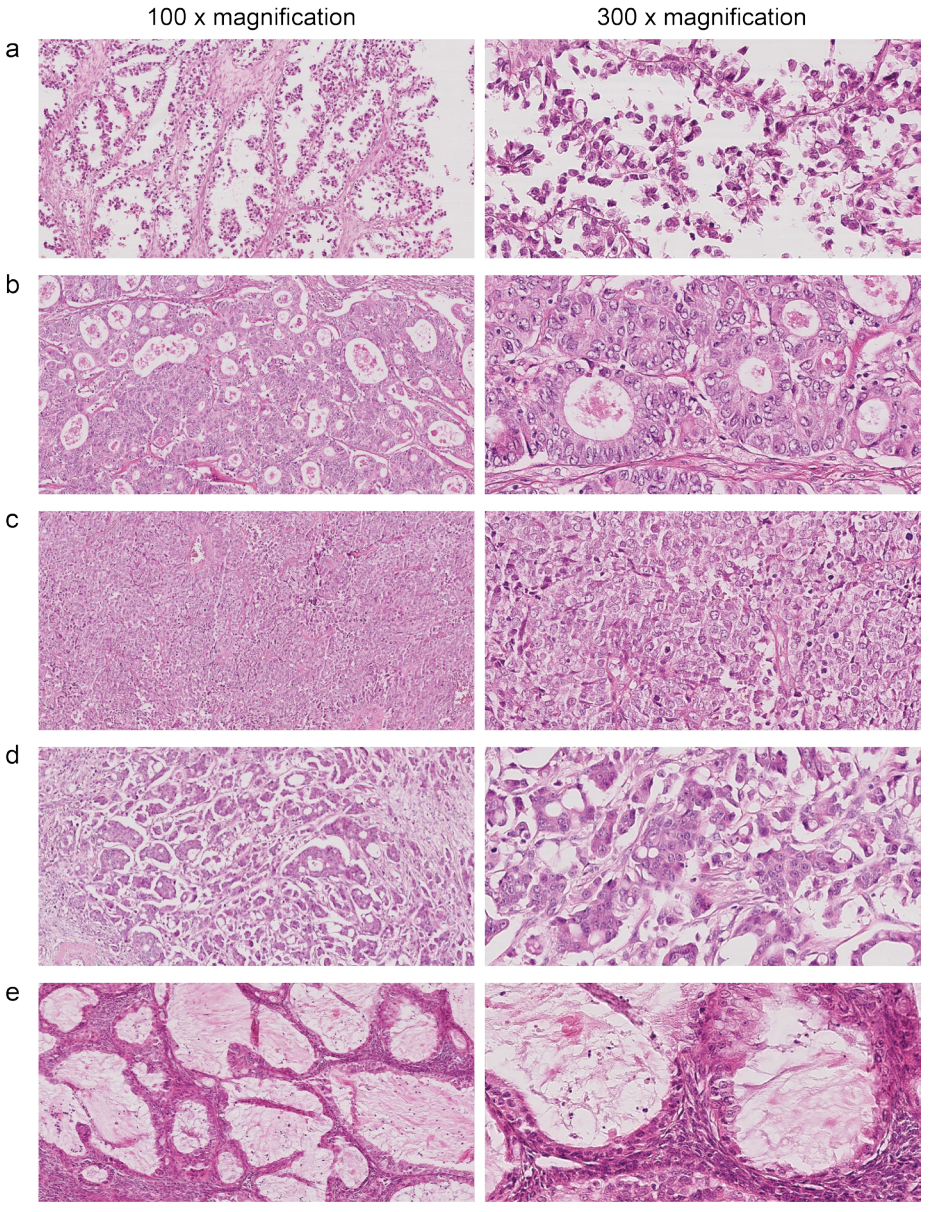


Figure 2. Eosin and hematoxylin stained tumor slides showing differences in molecular morphology representative for each of the main histotypes (CCC, EC, HGSC, LGSC, MC) with 100 x magnification (left column) and 300 x magnification (right column). *The cytoplasm surrounding the nucleus appears clear and white in CCC (a). Atypical irregular glandular structures are significant for EC (b). A zoom-in on tubule and cribriform structures is further shown for 300 x*

magnification. HGSC is characterized by a high nucleus/cytoplasm ratio, i.e. the nucleus is relatively large in comparison to the surrounding cytoplasm, with high mitotic activity (c). This *HGSC does not comprise any papillary or glandular structures. LGSC is defined by a low nucleus/cytoplasm ratio with less mitotic activity in comparison with HGSC (d). Mucin producing hollows are typical for the MC histotype (e).*

Treatment

Although the ovarian carcinoma histotypes are considered to be distinct diseases, the majority of all ovarian carcinoma patients are still treated with debulking surgery followed by platinum/taxane-based chemotherapy. To date, no alternative treatment regimens have alone proven superior to conventional therapy [61,62]. Treatment with PARP inhibitors is the first successful attempt to personalize therapy for ovarian carcinoma patients based on individual tumor characteristics [60]. PARP inhibitors are used in the clinic as a form of maintenance therapy for cancers deficient in the HR pathway, *e.g.* with *BRCA1* or *BRCA2* mutations [63]. PARP inhibitors block PARP enzymatic function, resulting in an accumulation of single-stranded breaks. This leads to the production of double-stranded breaks, which requires the HR pathway to repair DNA damage. Hence, cells deficient in the HR repair pathway die, while healthy cells survive [64]. Unfortunately, there are currently no other targeted therapies available in the clinic today for other ovarian carcinoma subgroups.

As previously described, HGSC are often poorly differentiated, and respond well to platinum/taxane-based chemotherapy. Nevertheless, patients with HGSC often relapse and treatment resistance occurs in 80 to 90% of patients that were initially diagnosed with metastatic disease, resulting in unfavorable outcomes [54]. *CCNE1* amplification, which has been associated with HGSC, has also been reported to contribute to chemotherapy resistance [65]. Furthermore, HGSCs show widespread HR deficiency, leading to eligibility for maintenance treatment with PARP inhibitors [66]. Overall, EC, LGSC and MC tend to be more well-differentiated and are less sensitive to chemotherapy, although with a more favorable prognosis in comparison with HGSC [6]. CCC also has a low response rate to chemotherapy, with intermediate survival outcome [6,67]. Despite the relative chemoresistance among LGSC patients, prognosis is favorable due to the tumor indolence and slow tumor growth. There are currently no targeted therapies available for LGSC. Recent clinical trials have however reported prolonged survival in some recurrent LGSC patients (15%) after treatment with MEK inhibitors [58]. EC and CCC may harbor mutations resulting in HRD, but not to the same extent as HGSC. The *PTEN* gene has for example been reported to confer HRD [62].

AIMS

The overall aim of this doctoral thesis was to provide novel insights into molecular features associated with early-stage ovarian carcinomas that may improve patient stratification and subclassification based on histotype and clinical outcome.

The specific aims were:

Paper I
To identify novel genetic biomarkers that may be of diagnostic and prognostic importance in early-stage ovarian carcinogenesis.

Paper II and III
To identify prognostic histotype-specific gene signatures and assess their clinical significance on the protein level in early-stage ovarian carcinomas.

Paper IV
To characterize genetic and epigenetic features for early-stage ovarian carcinoma histotypes, and identify putative oncogenes and tumor suppressor genes using an integrated multi-omics approach.

MATERIALS AND METHODS

PATIENTS AND TUMOR SAMPLES

The patient cohorts in **Paper I-IV** comprised early-stage primary invasive ovarian carcinoma patients, chosen for inclusion based on stage (I and II) and survival data (short-term survivors: 0-2 years, 2-5 years, 5-10 years; long-term survivors: >10 years). In **Paper II and III** inclusion criteria was further based on histotype (for FFPE samples not corresponding with the fresh-frozen samples used in **Paper I and IV**). The patients were diagnosed between 1994 and 2006 to enable at least ten years of follow-up time. Fresh-frozen tumor samples and formalin-fixed paraffin-embedded (FFPE) tumor blocks were retrieved from the tumor bank at the Sahlgrenska University Hospital Oncology lab (Gothenburg, Sweden) and the Departments of Clinical Pathology at hospitals in Western Sweden, respectively, in accordance with ethical approval (Regional Ethical Review Board, Gothenburg, Sweden, case number 767-14). Clinicopathological data corresponding to the tumor samples were obtained from the Cancer Registry at the National Board of Health and Welfare (Stockholm, Sweden) and the National Quality Registry at the Regional Cancer Center West (Gothenburg, Sweden).

Pathologists at Sahlgrenska University Hospital reclassified all tumors, using hematoxylin and eosin stained tumor slides, in accordance with current WHO criteria [22]. **Paper I** included 96 fresh-frozen tumor samples (17 CCC, 17 EC, 50 HGSC, 1 LGSC, 11 MC), 206 FFPE samples (95/206 FFPE samples corresponding with fresh-frozen tumor samples included in **Paper I**; 37 CCC, 46 EC, 94 HGSC, 29 MC) in **Paper II**, 112 FFPE samples (45/111 FFPE samples corresponding with fresh-frozen tumor samples included in **Paper I**; 37 CCC, 46 EC, 29 MC) in **Paper III**, and 96 fresh-frozen tumor samples (17 CCC, 17 EC, 51 HGSC, 11 MC) in **Paper IV** (Table 1). Tumor content (percentage of neoplastic cells in view of all cells present) were evaluated using representative imprints from fresh-frozen tumors stained with May-Grünwald Giemsa (Chemicon). Tumor samples comprising at least 50% neoplastic cell content were included in the analyses.

Table 1. Clinicopathological patient data for fresh-frozen (n=97) and FFPE (n=111) early-stage ovarian carcinoma samples. *Percentages of the number of patients are specified in the parentheses. Significant P values are marked in bold.*

	Fresh-frozen samples (n=97)						FFPE samples (n=111)				
	CCC (n=17)	EC (n=17)	HGSC (n=51)	LGSC (n=1)	MC (n=11)	***P* value**	CCC (n=20)	EC (n=29)	HGSC (n=44)	MC (n=18)	***P* value**
Patient age						NA					0.31
Mean	63	64	63	78	61		66	61	65	60	
Range	42-84	25-83	32-86	-	39-80		51-84	29-81	22-88	30-82	
Overall Survival						0.11					0.14
0-2y	2 (12)	1 (6)	2 (4)	1 (100)	3 (27)		3 (15)	2 (7)	5 (11)	3 (17)	
2-5y	3 (18)	5 (29)	17 (33)	-	2 (18)		7 (35)	4 (14)	9 (20)	1 (6)	
5-10y	7 (41)	5 (29)	19 (37)	-	3 (27)		1 (5)	2 (7)	10 (23)	4 (22)	
>10y	5 (29)	6 (35)	13 (25)	-	3 (27)		9 (45)	21 (72)	20 (45)	10 (56)	
Cause of death						**0.024**					**0.043**
Ovarian carcinoma	10 (59)	3 (18)	33 (65)	1 (100)	2 (18)		9 (45)	4 (14)	21 (48)	3 (17)	
Other cancer	0 (0)	3 (18)	7 (14)	-	3 (27)		2 (10)	3 (10)	1 (2)	2 (11)	
Other	6 (35)	6 (35)	5 (10)	-	4 (36)		1 (5)	4 (14)	5 (11)	5 (28)	
Not available	1 (6)	0 (0)	0 (0)	-	0 (0)		0 (0)	1 (3)	2 (5)	1 (6)	
Alive	0 (0)	5 (29)	6 (12)	-	2 (18)		8 (40)	17 (59)	15 (34)	7 (39)	
Stage						0.12					**0.029**
I	14 (82)	11 (65)	29 (57)	-	9 (82)		17 (85)	21 (72)	22 (50)	13 (72)	
II	3 (18)	6 (35)	22 (43)	1 (100)	2 (18)		3 (15)	8 (28)	22 (50)	5 (28)	
Tumor grade EC						NA					NA
FIGO grade I	NA	2 (12)	NA	-	NA		NA	9 (31)	NA	NA	
FIGO grade II	NA	9 (53)	NA	-	NA		NA	18 (62)	NA	NA	
FIGO grade III	NA	6 (35)	NA	-	NA		NA	2 (7)	NA	NA	
Dualistic model						**<0.001**					**<0.001**
Type I	17 (100)	17 (100)	0 (0)	1 (100)	11 (100)		20 (100)	29 (100)	0 (0)	18 (100)	
Type II	0 (0)	0 (0)	51 (100)	-	0 (0)		0 (0)	0 (0)	44 (100)	0 (0)	
CA125						0.13					**0.041**
<35	6 (35)	7 (41)	9 (18)	-	5 (45)		8 (40)	6 (21)	9 (20)	5 (28)	
35-65	1 (6)	0 (0)	29 (57)	-	2 (18)		7 (35)	7 (24)	31 (70)	6 (33)	
>65	10 (59)	10 (59)	13 (25)	1 (100)	4 (36)		5 (25)	15 (52)	4 (9)	7 (39)	
Not available	0 (0)	0 (0)	0 (0)	-	0 (0)		0 (0)	1 (3)	0 (0)	0 (0)	
Ploidy						0.14					0.28
Near diploid	1 (6)	7 (41)	15 (30)	-	2 (18)		4 (20)	10 (34)	7 (16)	5 (28)	
Aneuploid	16 (94)	9 (53)	36 (71)	1 (100)	8 (73)		14 (70)	17 (59)	34 (77)	11 (61)	
Not available	0 (0)	1 (6)	0 (0)	-	1 (9)		2 (10)	2 (7)	3 (7)	2 (11)	
Chemotherapy						NA					NA
Yes	17 (100)	17 (100)	49 (96)	1 (100)	11 (100)		20 (100)	25 (86)	42 (95)	16 (89)	
No	0 (0)	0 (0)	0 (0)	-	0 (0)		0 (0)	0 (0)	0 (0)	0 (0)	
Not available	0 (0)	0 (0)	2 (4)	-	0 (0)		0 (0)	4 (14)	2 (5)	2 (11)	

EXTERNAL COHORTS

A total of 30 normal samples were chosen from The Cancer Genome Atlas (TCGA) ovarian carcinoma cohort and used in **Paper I and IV**. Raw sequencing data (whole exome sequencing) from these patient samples were collected from the Genomic Data Commons Data Portal (GDC Data Portal) with approval through the database of Genotypes and Phenotypes (dbGaP; project #11044). The RNA sequencing (RNA-seq) analyses, including the analysis of the control samples, required large data capacity and were thus performed on external computational resources at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX, project number: b2015239). To be able to compare our data with the control samples, the raw data for the control samples were converted and compressed to FASTQ gzip format with the BEDTools (v. 2.25.0) and gzip module, and subsequently subjected to the same analyses as for our ovarian carcinoma samples (see below). In **Paper II and III**, Affymetrix gene expression microarray data for 1,657 ovarian carcinoma patients (HGSC n=1,232, EC n=62) was used to evaluate the clinical significance (overall survival) of putative histotype-specific prognostic biomarkers in the web-based Kaplan-Meier (KM) plotter tool (https://kmplot.com/analysis/; see below).

DATA ANALYSIS

Whole-transcriptome RNA sequencing analysis

Transcriptome-wide analysis was performed on a HiSeq2000 sequencer (Illumina, San Diego, CA, USA) and applied in **Paper I and IV** to identify genetic variants, fusion events, and changes in gene expression of early-stage ovarian carcinomas.

Principle

RNA-seq is a powerful tool that can provide an extensive overview of the transcriptome, *i.e.* information regarding RNA type (*e.g.* mRNA, non-coding RNAs or small RNAs) and quantity of transcript sequences that may be used to determine differences in gene expression between various conditions or sample groups, splicing patterns, and post-transcriptional modifications. Extracted and purified RNA (*e.g.* total RNA, mRNA) is converted into cDNA. Thereafter, the cDNA is fragmented, and adapter sequences are added to each fragment. The fragmented cDNA is subsequently sequenced on a high-throughput sequencer

with either single-end sequencing (sequencing of short sequences from one end) or paired-end sequencing (sequencing of short sequences from both ends) using a read depth of commonly 10-30 million reads per sample, yielding short sequences of ca 30-400 bp in length [68,69].

Method

RNeasy Lipid Tissue Mini Kit (Qiagen) was used to isolate total RNA from fresh-frozen tumor samples (n=96). The RNA concentration was measured using Nanodrop ND-1000 (Nanodrop Technologies) and QuBit (ThermoFisher Scientific), while RNA integrity was determined using the RNA 6000 Nano LabChip Kit with Agilent 2100 Bioanalyzer (Agilent Technologies). Total RNA from samples having an RNA integrity number greater than 6 were processed at the Science for Life Laboratory (National Genomics Infrastructure, Stockholm). A HiSeq2000 sequencer (Illumina) generated TruSeq strand-specific RNA libraries (Ribosomal depletion using RiboZero human) comprising 125 bp paired-end reads for each sample with approximately 10-22 million aligned reads per sample.

Computational analysis

The RNA-seq analyses in **Paper I** were performed on external computational resources at UPPMAX. Quality checks were performed with FastQC (default settings, v. 0.11.2) prior to and after removal of low quality bases and adapter sequences with TrimGalore (v. 0.4.0). Trimmed reads were thereafter aligned to the human reference genome hg19 with STAR aligner (v.2.5.0c) with 1-pass and 2-pass modes, respectively. Raw read counts (number of sequences mapped to the human reference genome hg19 assembly) were calculated with the htseq module (v. 0.6.1) on name sorted aligned reads (1-pass mode).

Genetic variants, *i.e.* mutations, were identified in **Paper I** following the Genome analysis toolkit (GATK) (v. 3.6) best practices (Broad Institute). In short, GATK SplitNCigarReads tool removed false positive calls and the HaplotypeCaller identified genetic variants from aligned reads (2-pass mode), and resulting genetic variants were subjected to filtering steps (*e.g.* quality by depth<2 were removed). Thereafter, the filtered genetic variants were annotated using ANNOVAR and common genetic variants in the human population were then removed by filtering against the 1000 Genomes Project dataset and dbSNP (minor allele frequency threshold=0.01), and matched with the Catalogue of Somatic Mutations in Cancer (COSMIC) database to identify known cancer-associated genetic variants [70-72]. Recurrent deleterious variants, *i.e.* harmful variants (resulting in increased likelihood of developing ovarian cancer) present

in at least 30% of the histotype or survival groups were identified and compared with normal controls.

Differentially expressed genes (DEGs) associated with survival (**Paper I**) and histotypes (**Paper IV**) were identified by comparing raw read counts for patients belonging to different survival groups (short-term survivors (0-2, 2-5, 5-10 years) in comparison with long-term survivors (>10 years)) and histotype groups (CCC vs MC, EC vs CCC, EC vs MC, HGSC vs CCC, HGSC vs EC, and HGSC vs MC) using DESeq2 (v. 1.14.0) in R/Bioconductor [73]. Fusion transcripts were identified in **Paper I** using FusionCatcher (v. 0.99.5a) with associated aligners (Bowtie, BLAT, STAR, Bowtie2) and annotation databases (ENSEMBL, UCSC, RefSeq). High probable false positive fusion transcripts were removed (marked with high or very high probability) and the oncogenic potential of identified fusion transcripts were determined with the Oncofuse tool [74]. Ingenuity Pathway Analysis (IPA) (Ingenuity Systems, Redwood City, USA) was employed in **Paper I** to associate identified genetic variants, fusion transcripts and DEGs with known cancer-related biological functions linked with cancer.

Fluorescence *in situ* hybridization analysis

Fluorescence i*n situ* hybridization (FISH) analysis was used in **Paper I** to validate commonly identified fusion transcripts involving the long non coding RNA *MALAT1*.

Principle

FISH is a cytogenetic technique to detect locus-specific chromosomal abnormalities and it can be used to locate specific DNA sequences on chromosomes, *e.g.* to evaluate possible gains, amplifications, deletions, inversions or translocations, such as fusions. Fluorescent DNA probes are synthesized and hybridized to the target DNA sequence of interest. The DNA probe is thereafter hybridized to a slide containing metaphase chromosome preparations and fluorescent signals can be observed on the chromosome(s) and in the cell nuclei to which the DNA probe has bound using a fluorescence microscope [75].

Method

Interphase FISH with two color-detection system (green/red) was used to identify commonly identified fusion transcripts involving the long-non coding (lnc) RNA *MALAT1*. More specifically, suitable bacterial artificial chromosome (BAC) clones covering each fusion partner were chosen using the University of California Santa Cruz (UCSC) Genome Browser and purchased from BACPAC

Resources Center (Children's Hospital Oakland Research Institute, CA, USA). Each BAC clone was tested on normal metaphase chromosomes to ensure the correct chromosomal location. The bacteria comprising the BAC clones were cultivated and the BAC DNA was subsequently extracted (Qiagen Plasmid Maxi kit), and separately labeled with biotin and digoxigenin using nick translation (Roche Diagnostics, Mannheim, Germany). Microscope slides with touchprints of corresponding fresh-frozen tumors were prepared. The labeled BAC DNA was hybridized to denatured interphase nuclei. The probes were detected using FITC avidin (green fluorescence color) and Rhodamine anti-dioxigenin (red fluorescence color) and subsequently counterstained with DAPI and mounted using an antifade solution (Vectashield DAPI, Vector Laboratories, Burlingame, CA, USA). A fluorescent microscope (Leica DMRA2, Leica Microsystems, Wetzlar, Germany) with attached camera (ORCA Hamamatsu CCD) was used for evaluation and image retrieval. A fusion transcript was indicated by a yellow hybridization signal, *i.e.* wherein the green-labeled and red-labeled BAC DNA are bound to the same specific chromosomal location.

Whole-genome Single nucleotide polymorphisms analysis

Genome-wide Single nucleotide polymorphisms (SNP) genotyping was performed in **Paper I** using Infinium HumanOmni2.5-8 Beadchips (Illumina, San Diego, CA, USA) to identify allelic imbalance and DNA copy number gains and losses.

Principle

Whole-genome SNP genotyping can identify structural changes in DNA, *e.g.* gains/losses on parts or whole chromosomes, and genetic variants. It is also possible to analyze allele-specific genetic alterations. Illumina genotyping BeadChips use one bead type with a two-color detection approach that can genotype between hundreds and millions of SNPs per sample. In this detection method, purified DNA is amplified, subsequently fragmented using enzymes, and precipitated and resuspended. Further, the DNA sample is hybridized to the BeadChip followed by enzymatic base extension, fluorescent staining, and detection of fluorescence intensities [76].

Method

DNA from nine tumor samples, comprising a verified fusion transcript using FISH, was extracted with the Wizard genomic DNA extraction kit (Promega) and subjected to phenol-chloroform purification. Purified DNA was processed at the Genomics DNA microarray resource center (SCIBLU, Department of oncology,

Lund University) using Infinium HumanOmni2.5-8 (v.1.3) with approximately 2.4 million markers corresponding to SNPs identified in the 1000 Genomes Project (Minor allele frequency (MAF)>2.5%) [77]. The analysis identified allelic imbalances and regions of DNA copy number gains and losses that could be compared with the identified genetic variants and fusion transcripts in **Paper I**.

Computational analysis

The circos package (v. 0.66) in UPPMAX was used to visualize SNP genotyping data (DNA copy number gains and losses) in view of RNA-seq data (genetic variants, fusion transcripts) [78].

Cox proportional hazard models

Candidate biomarkers with prognostic significance were identified in **Paper II and III** using Cox proportional hazard models associating RNA-seq read counts with survival data.

Principle

The association between survival time (overall (OS, the time from initial diagnosis to death from any cause) and disease-specific survival (DSS, the time from initial diagnosis to ovarian cancer-related death)) and raw RNA-seq counts (log2-values) can be analyzed with Cox proportional hazard regression models. The model follows the hazard function denoted by $h_1(t) = h_o\ (t)e^{\sum \beta x}$, which express the risk of dying at time t. $e^{\beta x}$ denotes the effect of individual probes, wherein patients with βx<0 correspond to patients with favorable prognosis and βx>0 with unfavorable prognosis.

Method

Univariable Cox proportional hazard models with Benjamini-Hochberg adjusted false discovery rates (*P* value<0.05) were applied to raw RNA-seq read counts for 95/96 patients stratified by histotype (CCC, EC, HGSC, and MC) from **Paper I** to associate histotype-specific gene expression data with survival data (OS and DSS). Time-dependent Area under the receiver operating characteristics curve [(AUC(t)] values were used to assess the predictive power, *i.e.* how well each regression model correlated with survival [79]. A concordance index (C-index) was determined for each biomarker that gave a measure on how well the biomarker could predict patient outcome. The C-index varied between 0.5 (no correlation between RNA-seq counts and survival) and 1 (perfect correlation between RNA-seq counts and survival).

Computational analysis

Candidate biomarkers for **Paper II and III** were selected among statistically significant genes with the highest C-index or log2 ratio, and with gene expression levels that could possibly be detected using immunohistochemistry (IHC).

Immunohistochemical analysis and evaluation

IHC on full-face FFPE sections and tissue microarrays (TMAs) were implemented in **Paper II and III** respectively, to validate the protein expression levels of candidate biomarkers with identified prognostic significance on the RNA level.

Principle

IHC is a method to detect proteins in cells of a tissue section. It is possible to detect whether a protein of interest is expressed or not and to what extent, *i.e.* the percentage of cells that express the protein, as well as the location of the protein, *e.g.* in the nuclei, membrane or cytoplasm of the cell. In combination with histopathological analysis of tumor morphology, IHC can be used to assist in determining accurate cancer diagnoses. The technique is based on the specific binding between monoclonal (specificity for a single epitope, *i.e.* a single segment of a protein) or polyclonal (specificity for several epitopes) antibodies and antigens present on the surface of the protein of interest. The antibodies are commonly labeled with a chromogen, which can be detected as a specific color (*e.g.* brown staining) upon antibody binding to the antigen. The resulting protein staining pattern in the tissue section is evaluated using a light microscope or digital pathology, *i.e.* evaluating a scanned image of the tissue section on the computer [80,81].

Method

In **Paper III**, TMAs were prepared for all tumors in the patient cohort. More specifically, a pathologist marked regions containing tumor tissue on hematoxylin and eosin stained FFPE sections for each tumor. Triplicate core biopsies (1 mm in diameter) from corresponding areas were then punched out, organized on a paraffin tumor block, and the tumor block was subsequently baked for one hour at 45°C. Four micrometer sections from full-face (**Paper II**) and TMA (**Paper III**) tumor blocks were prepared on microscope slides (FLEX IHC, Dako, Sweden). Suitable antibodies for each biomarker to be tested were primarily selected from the Human Protein Atlas. Each antibody was optimized on full-face FFPE sections to determine the proper antibody concentration and a positive control sample. Tumor sections were pretreated with the Dako PTLink system (pH 9), and immunostained with the respective antibodies with Dako

Autostainer Plus (Agilent Technologies) using DAB (3,3'-diaminobenzidine, brown color) as chromogen and hematoxylin (blue color) as counterstain. Lastly, the immunostained FFPE sections were rinsed (deionized water), dehydrated (ethanol series), cleared (xylene) and mounted. The IHC slides were evaluated by a board certified pathologist (two pathologists for **Paper II**), which did not have any knowledge about survival times. An H-score (immunoreactive score) was determined for each slide/TMA core, based on both the percentage of stained tumor cells and the staining intensities thereof (weak = 1, moderate = 2, strong = 3) [82]. The H-score for each tumor on the TMAs were based on the mean of triplicate cores.

Genome-wide DNA methylation and DNA copy number alteration analyses

Infinium MethylationEPIC BeadChips (Illumina, San Diego, CA, USA) were used in **Paper IV** to characterize genome-wide DNA methylation and DNA CNA patterns.

Principle

Illumina Infinium MethylationEPIC BeadChips is a comprehensive method to evaluate the DNA methylation status across the genome. It covers more than 850,000 CpG sites with single-nucleotide resolution. The DNA methylation detection method depends on bisulfite converted DNA, wherein the methylation state is transformed to a detectable genetic SNP difference, *i.e.* an unmethylated C is converted into a T and a methylated C remains a C. The CpG site detection on the array includes a combination of Infinium type I and II probes (each about 50 bases in length), wherein type I comprises two probes per CpG site binding to methylated or unmethylated sites, respectively, and type II comprises one probe binding to the CpG site relying on red and green fluorescent color detection [83,84]. Using computational analysis, it is also possible to retrieve DNA CNA information from the DNA methylation data.

Method

DNA from 91 tumor samples was extracted and purified (see SNP analysis above). Purified DNA with 260/280 ratios greater than 1.8 (Nanodrop ND-1000 spectrophotometer, Nanodrop Technologies) was sent to SNP&SEQ Technology Platform (Uppsala, Sweden), for analysis on the EPIC array (MethylationEPIC v. 1.0, genomic build, v. 37).

Computational analysis

The package ChAMP (v. 2.14.0) in R/Bioconductor (v. 3.6.0) was used to perform the DNA methylation analyses [85,86]. The ratio of methylation, *i.e.* the β intensity values, were generated from IDAT files for each tumor sample. The intensity values were filtered using ChAMP default filtering steps, normalized with BMIQ normalization method to correct for probe type (I and II) bias and corrected for batch effects (array, slide) with the myCombat function in ChAMP. The intensity values for each CpG site were thereafter merged with probe information *e.g.* gene, type of region surrounding the CpG island and genomic regions (probe.features in ChAMP) and enhancer information (MethylationEPIC_v-1-0_B4_Manifest File.csv). Density plots of beta value distribution were made in ChAMP to examine the methylation data to identify possible outliers among the tumor samples. The β intensity values were filtered according to variance to identify the 1000 most variable probes within the data set, and visualized according to histotype in a Raw data, Descriptive, Inference statistics (RDI) plot with the yarr package (v. 0.1.5) and in a heatmap using pheatmap (v. 1.0.12) [87,88]. Hierarchical clustering was performed with the Ward's method and the Canberra distance measure (distance between samples), wherein a shorter distance indicated higher similarity. Histotype-specific differentially methylated probes (DMPs) were identified using both the limma (v. 3.40.2) and ChAMP packages, with Benjamini-Hochberg adjusted *P* value<0.05 and additionally>1.5 fold-change in the limma package [89].

The conumee package (v. 1.18.0) in R/Bioconductor was utilized to extract unsegmented CNA data for each probe and normalized using CNA data from healthy individuals (n=52, CopyNumber450kData, v. 1.8.0.) [90]. One part of the CNA analysis was performed in Nexus Copy Number (BioDiscovery, v. 7.5) with the Rank segmentation algorithm using default settings to identify genomic regions containing CNAs, *i.e.* homozygous loss/deletion (loss of two gene copies, log2 ratio≤-1), heterozygous loss (loss of one gene copy, log2 ratio<-0.3), and gain (≥3 gene copies, log2 ratio>+0.3). Recurrent significant CNAs were identified in at least 35% of the tumor samples (*P* value<0.05). Another part of the CNA analysis focused on evaluating genetic instability in the form of chromothripsis-like patterns (CTLPs) using the web-based CTLPScanner (http://cgma.scu.edu.cn/CTLPScanner/) with default settings. The input data (segmented CNA data) for this analysis was retrieved with ChAMP.

RESULTS AND DISCUSSION

Paper I

RNA sequencing and SNP genotyping data reveal novel genetic features in early-stage ovarian carcinoma associated with histotype and survival

Ovarian cancer is the most lethal of all gynecological cancers, and in 2016, the overall 5-year survival rate in Sweden was only 55% [8]. FIGO early-stage (stage I and II) ovarian carcinoma tumors are considered to be less aggressive with a more favorable prognosis (5-year survival rates above 71%), whereas late-stage (stage III and IV) tumors are more aggressive [43]. However, around 16% of stage I and II ovarian carcinomas behave aggressively, leading to the patient's death within five years. Since these patients are diagnosed with a tumor stage associated with generally good clinical outcome, they may be given inadequate treatment. In our patient cohort, we chose to only include early-stage ovarian carcinomas, since the genetic profiles of early-stage tumors are generally less complex compared to the later stages. This makes it possible to classify early events in ovarian carcinoma tumorigenesis, and to identify specific genomic alterations related to ovarian carcinoma. In recent years, it has been shown that ovarian carcinoma is not a single disease entity, but rather a group of diseases with different origins, genetic alterations, prognosis and clinicopathological features. This resulted in a reclassification of ovarian carcinoma into five distinct histotypes, namely HGSC, LGSC, EC, MC and CCC carcinomas [23,91]. The patients in our study cohort were diagnosed with ovarian carcinoma between 1994 and 2006. Hence, all patients were classified according to the older WHO classification systems from 1973 and 2003 [92,93]. It was therefore necessary to reclassify all tumors in the patient cohort according to current WHO criteria to determine histotypes according to the most recent stratification system [22]. It is further important to keep in mind that the patients included in our study cohort were diagnosed before it was praxis (year 2012) in Sweden to examine all patients for tumor spreading to the lymph nodes at the time of staging. Hence, we cannot say with 100% certainty that all patients were in fact in stage I or II, since a possible spreading to the lymph nodes were at the time not investigated, and if present, would have staged the patient in stage III.

In order to improve the survival rates of ovarian cancer, there is a need to identify novel biomarkers. Ideally, to enhance the ratio of benefit vs toxicity of the administered drug a selection of therapeutic drugs would be available for different types of ovarian cancer based on 1) the patient's clinicopathological

features and 2) the molecular features of the tumor. Unfortunately, few prognostic biomarkers and personalized treatment strategies are to date available in the clinic that can assist in the diagnosis and treatment of ovarian carcinomas. In **Paper I**, we used transcriptome- and genome-wide RNA-seq (n=96) and SNP (n=9/96) analyses to identify genetic variants, fusion transcripts, SNP genotyping and gene expression patterns associated with histotype and survival for early-stage ovarian carcinomas. SNP analysis was not performed for all samples in the patient cohort due to cost. Deleterious genetic variants, *i.e.* genetic variants that cause a change or loss in protein function, were identified among the tumor-specific coding variants (found in exonic regions). Mutation signatures comprising 38 and 49 deleterious variants were present in at least 30% (defined as recurrent deleterious variants) of the samples in either histotype group (CCC, EC, HGSC, MC) or survival group (0-2, 2-5, 5-10, >10 years), respectively. Few of the recurrent deleterious variants were found in normal ovarian tissue indicating tumor-specificity and those variants that were identified in normal tissue significantly differed in mutation rates. Hence, these variants could be used for diagnostic purposes. Furthermore, the majority of deleterious variants had not previously been linked to cancer (COSMIC database) indicating their novelty.

Fusion events may be strong driver mutations in tumorigenesis and may further promote genomic instability resulting in a high frequency of mutations [94]. It may be advantageous to therapeutically target a fusion gene if the fusion results in the oncogenic activation or repression of tumor suppressor genes. For example, BCR-ABL is an oncogenic fusion gene in chronic myelogenous leukemia, wherein inhibition of BCR-ABL resulted in improved clinical outcome [95]. A total of 3,344 fusion transcripts (1,503 unique fusion transcripts) were identified, wherein fusions between genes located on different chromosomes (interchromosomal fusions) were more common in comparison with fusions between genes located on the same chromosome (intrachromosomal fusions). Interestingly, the long non-coding (lnc) RNA *MALAT1* was involved in multiple fusion events. Further, *MALAT1* expression levels were significantly higher in tumors in the patient cohort compared to normal ovarian tissue. This is in line with previous reports demonstrating *MALAT1* overexpression with higher cell proliferation and invasion [96,97]. A recent report further confirmed the high occurrence of *MALAT1* in ovarian carcinoma-associated fusion genes, wherein *MALAT1* was reported to be involved in more than 60% of fusion genes identified in a cohort of HGSC and triple-negative breast cancer patients [98].

The genomic rearrangements of coding genetic variants, fusion events and CNAs were visualized in circos plots for each tumor (Figure 3a). All nine samples used in the SNP genotyping analysis comprised at least one fusion gene (as verified by FISH, Figure 3b). As expected, DNA breakpoints were often found on corresponding chromosome positions where a fusion gene was identified using RNA-seq. Moreover, DNA breakpoints may increase the risk of further mutations, which may for example be seen as a peak of exonic variants on chromosome 11 at the position of *MALAT1* which is involved in multiple fusions for sample OV155 (Figure 3a) [99]. Large CNAs have previously been associated with HRD. Whether a tumor is deficient in HR is of importance since it can have implications for clinical treatment, such as sensitivity to PARP inhibitors and platinum- and taxane-based chemotherapy [100,101].

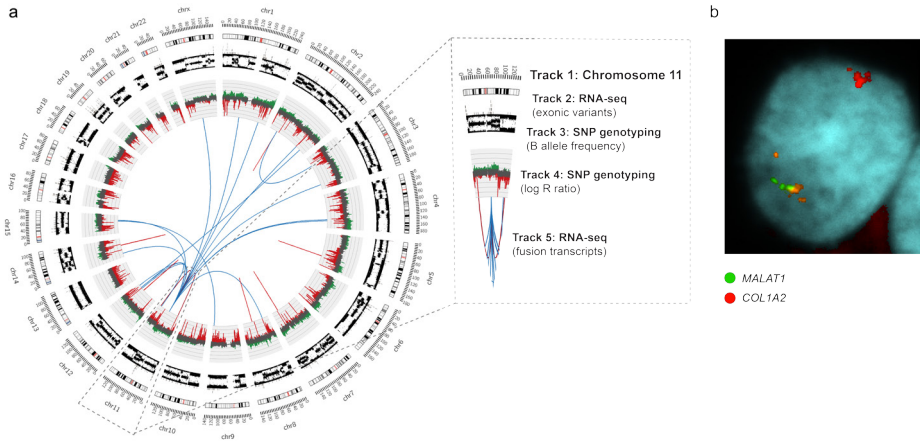


Figure 3. Genomic rearrangements for ovarian carcinoma sample OV155, and FISH validated fusion transcript. *Circos plot (a) showing the genomic rearrangements (coding genetic variants, copy number alterations, fusion genes) identified in ovarian carcinoma sample OV155. Chromosome 11 is shown as a legend with chromosome cytobands (Track 1), exonic genetic variants (Track 2, shown as dark gray bars) identified with RNA-seq data, B allele frequency (Track 3) and log R ratio of SNP genotyping data (Track 4), and fusion transcripts (Track 5) identified with RNA-seq data. Copy number gains are shown in green and losses in red. A total of 43 fusion genes were identified in OV155 with 16 intrachromosomal gene fusions (shown in red), and 27 interchromosomal gene fusions (shown in blue). FISH analysis image (b) illustrating evidence of a fusion gene (yellow) between MALAT1 gene (green) and COL1A2 gene (red) for OV336 sample. In a normal cell, MALAT1 gene is located on chromosome 11 and COL1A2 gene on chromosome 7.*

The gene expression analysis in **Paper I** was mainly focused on evaluating differences in gene expression patterns between the survival groups (0-2, 2-5, 5-10, >10 years) across the entire patient cohort (regardless of the tumor histotype) to identify genes associated with differences in the tumor aggressiveness. The analysis identified 23 genes that were differentially expressed between at least two survival groups. Few studies have previously been performed for early-stage ovarian carcinomas in view of histotype and clinical outcome. Therefore, **Paper I** presents an important addition to current research, wherein novel genetic profiles relating to mutations, fusion genes, CNA, and gene expression profiles were identified.

Paper II and III

Immunohistochemical validation of histotype-specific biomarkers for improved prognostication of early-stage ovarian carcinoma

Historically, ovarian carcinoma has been diagnosed and treated as one disease, but in recent years it has been shown that ovarian carcinoma can be subdivided into five main histotypes (CCC, EC, HGSC, LGSC, and MC) that significantly differ from one another [23,27,91]. HGSC is the most commonly diagnosed histotype, corresponding to about 70% of all ovarian cancers, which is also reflected by it being the most well-studied histotype [20,53]. In the clinic, determination of histotype is mainly done by histopathological analysis of tumor morphology. The protein expression of a panel of five biomarkers, namely Pax8, WT1, ER, PGR and p53, may further assist in accurate classification of the different histotypes using IHC [22]. It may in some cases be difficult to distinguish between EC and HGSC histotypes. WT1 may be helpful in these cases, although it does not perfectly discriminate between the histotypes since some EC tumors express WT1, whereas some HGSC lack in WT1-expression (Table 2). In addition to the patient cohort used in **Paper I**, additional tumor samples were included in the patient cohorts used in **Paper II and III** to increase the number of tumor samples within each histotype group and therewith improve the statistical power of the analyses. The number of tumor samples more than doubled from 95 to 206 samples in **Paper II** and from 45 to 112 samples in **Paper III**. The additional tumors were also diagnosed between 1994 and 2006, and hence reclassified according to current WHO stratification [22].

Table 2. Biomarkers for histotype-classification that are routinely used in the clinic. *Protein expression patterns of tumors (values shown in percentages) expressing each protein (Pax8, WT1, ER, PGR, and p53) [22]. The majority of histotypes (about 90%) can however be determined using solely microscopic evaluation by a pathologist.*

Histotype	Pax8	WT1	ER	PGR	p53
CCC	99	0	13	6	12
EC	84	4	86	72	11
HGSC	98	92	80	30	93
LGSC	100	100	96	50	0
MC	50-60	0	6	0	50

In the clinic, the most important prognostic clinicopathological factor at the time of diagnosis is tumor stage [42]. The 5-year survival rate for ovarian carcinoma patients (all epithelial subtypes) differs radically from 89% of the patients surviving at least 5 years with stage I disease, to only 20% with stage IV disease [43]. Histotype also has prognostic significance with differences in 5-year survival rates (serous 43%, EC 82%, MC 71%, CCC 66%). In an early-stage patient cohort, it has been reported that patients with EC and MC tumors in stage Ia or Ib have the most favorable prognosis with 10-year DSS of equal to or over 95% [46]. To date, few prognostic ovarian carcinoma biomarkers are implemented in the clinic. Within standard care, ovarian carcinoma patients with *BRCA*-mutated tumors can be treated with PARP inhibitors that may lead to prolonged survival [47]. Recent clinical trials have further shown increased survival for patients with HRD in general and not only HRD caused by *BRCA*-mutations [48]. Apart from testing the tumor for *BRCA* mutation status and HRD, no other prognostic biomarkers have been implemented for routine-testing in the clinic.

The biomarkers that were selected for protein expression analysis in **Paper II and III** were primarily selected from gene lists generated with Cox proportional hazard models that correlated RNA expression values (raw RNA-seq counts) with survival (OS/DSS) for tumors belonging to the same histotype group (CCC, EC, HGSC, MC). Due to lack of samples in the MC histotype group, Cox regression could only be performed for OS. Moreover, the RNA sequenced patient cohort (n=96) comprised only one LGSC sample, and it was hence excluded from these analyses. In **Paper II and III**, promising biomarkers (n=27 for **Paper II**, n=29 for **Paper III**) were chosen for IHC validation among significant genes with the highest C-index or highest log2 ratio. *MTUS1* and *COL3A1* that were identified in **Paper I** were further selected for validation in **Paper II**, leading to 29 biomarkers tested in each paper. IHC was chosen as validation method since it is presently used as a

standard method in the clinic for examining protein expression of specific biomarkers, which may enable easier implementation of testing novel biomarkers in the clinic. For a relatively large portion of the biomarkers (17/29) in **Paper II**, an optimized antibody dilution could not be determined, due to weak protein staining. To improve the number of biomarkers with detectable protein expression using IHC, biomarkers with overall higher RNA expression levels (RNA-seq counts>150) were selected in **Paper III**, which is *e.g.* reflected by the number of biomarkers with determined optimal antibody dilutions (28/29). The fact that not all biomarkers could be optimized may also depend on unsuitable antibodies and/or potential differences between RNA and protein expression levels due to *e.g.* chemical changes when a pre mRNA is converted to a mature mRNA and protein [102,103].

Paper II resulted in validation of the prognostic significance of COL3A1 protein expression with shorter OS in ovarian carcinoma (for all main histotypes), GPR158 expression with shorter OS in MC patients and PITHD1 expression with longer OS and DSS in CCC patients. However, PITHD1 protein expression did not correspond with its RNA expression, wherein *PITHD1* RNA expression correlated with shorter survival outcome. This discrepancy may be due to differences in measuring the expression levels, *i.e.* IHC only takes the expression levels of tumor cells into account, whereas RNA-seq measures the expression of all cells (*e.g.* tumor and stromal cells).

In **Paper III**, TMA were constructed with triplicate cores for each tumor sample in the patient cohort. It was cheaper and less time-consuming to perform IHC on TMAs in comparison with IHC on full-face FFPE sections, especially when testing multiple potential biomarkers on the same patient cohort. This may however only comply for research purposes. In the clinic, it may still be easier to use full-face FFPE sections for rapid diagnosis and prognosis, since it saves the step of constructing a TMA. Moreover, it is important to keep in mind that core biopsies from a tumor block only comprise selected areas of the entire tumor and not the full tumor section. In **Paper III**, we focused on identifying prognostic biomarkers for the CCC, EC, and MC histotypes. Potential biomarkers identified in HGSC were excluded due to relatively low prognostic potential in the Cox proportional hazard models. Ultimately, the prognostic significance of 17/29 proteins (10 CCC-associated biomarkers (ARPC2, CCT5, GNB1, KCTD10, NUP155, RPL13A, RPL37, SETD3, SMYD2, and TRIO), three EC-associated biomarkers (CECR1, KIF26B, and PIK3CA), four MC-associated biomarkers (CHEK1, FOXM1, KIF23, and PARPBP)) was validated on the protein level (Figure 4).

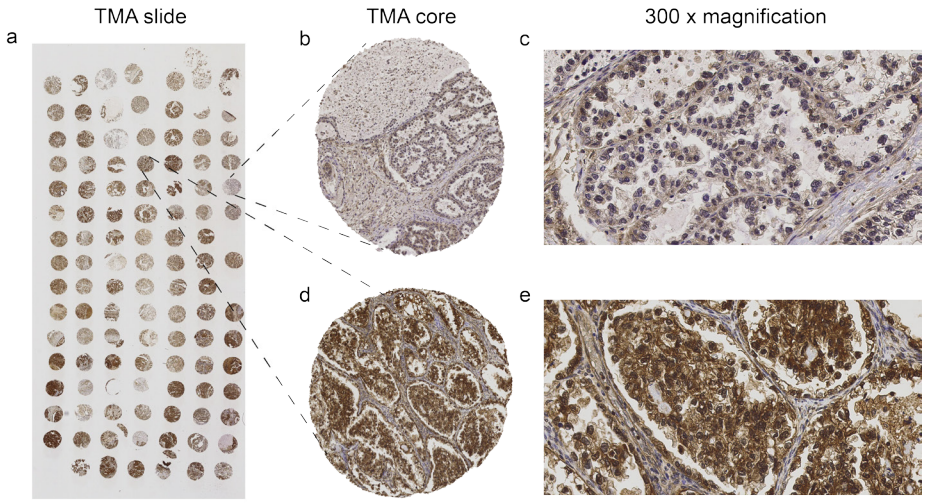


Figure 4. Protein expression of NUP155 using TMA. *TMA slide (a) for protein staining of NUP155 for 37 CCC samples with triplicate TMA cores. Representative IHC staining intensities for weak (b, c) and strong (d, e) NUP155 protein staining shown on a TMA core (60 x magnification, middle column; 300 x magnification, right column). The top section of the weakly stained TMA core (b) comprised of necrotic tissue, the left section with tissue that also stretches into the right section of the TMA core was comprised of stromal cells and the right section was comprised of tumor cells. The strongly stained TMA section (d) comprised of large glandular structures of stained tumor cells (brown color) and stromal cells (blue color) in between the stained tumor cells.*

It was revealed that all validated biomarkers (n=20) in **Paper II and III** could better predict the patients' survival time when combining the protein expression of the biomarkers with established clinicopathological parameters (age, stage, CA125, ploidy, and/or histotype) in comparison with models comprising only the established clinicopathological parameters. This further highlights the importance of these biomarkers. The KM plotter online tool was used to validate our findings in an external dataset on the RNA expression level. Unfortunately, KM plotter only comprises a few EC ovarian carcinoma samples and no CCC or MC samples. To date, there are no other public databases with CCC or MC expression data linked to survival data. A predictive model comprising the combined expression results from **Paper II and III** (GPR158 + MC-associated biomarkers, PITHD1 + CCC-associated biomarkers) in combination with established clinical parameters could even better predict the patient outcome.

In our models, the combination of individual biomarkers lead to an overall improved predictive power. However, it is important to keep in mind that in the clinic, immunohistochemical evaluation of multiple biomarkers may be too labor-intensive to perform and evaluate. Therefore, its clinical utility may therefore be hindered in daily pathology practice. In addition, some of the antibodies may not be widely available. The histotype-specific biomarker panels presented here may for these reasons need further testing to optimize and reduce the number of biomarkers in each panel, especially the panel with CCC-associated biomarkers in **Paper III** that contains the highest number of biomarkers. To the best of our knowledge, the protein expression of the biomarkers in **Paper II and III** have not previously been associated with prognosis in early-stage ovarian carcinomas.

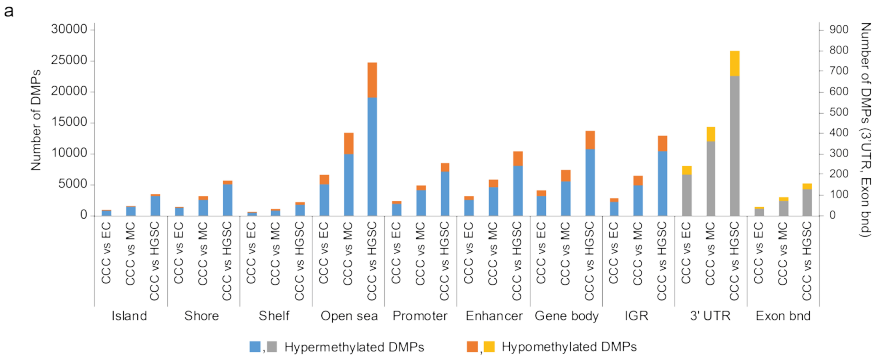
Paper IV

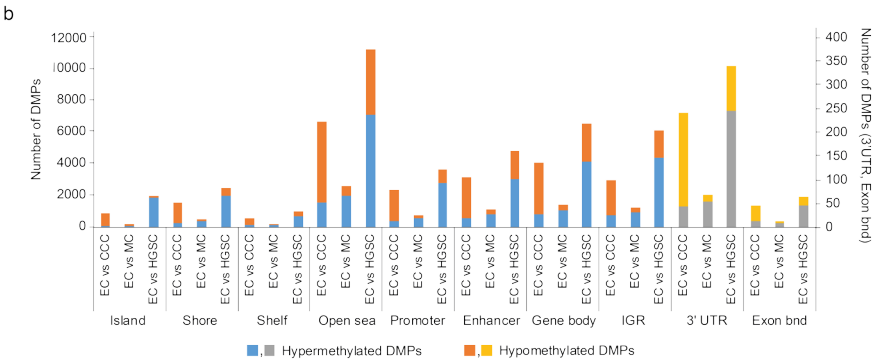
Genomic and epigenomic features associated with early-stage ovarian carcinoma histotypes drive transcriptomic deregulation

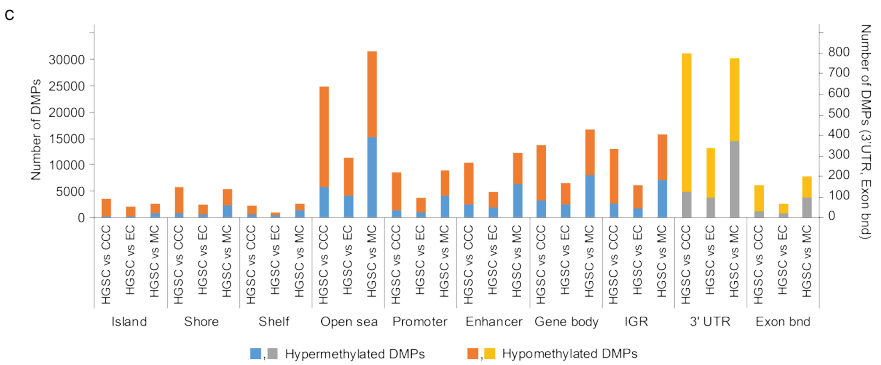
Sequencing technologies, such as next-generation sequencing (NGS), have revolutionized our understanding of the mechanisms behind cancer formation and progression through genome-scale analyses [19]. Cancer was long thought to be mainly based on genetic alterations, such as point mutations, CNA, as well as insertions and deletions. In recent years, it has however been shown that epigenetic changes also have a large impact on oncogenic transformation, and that genetic and epigenetic alterations may cooperatively influence carcinogenesis through *e.g.* mutations in epigenetic regulators [104]. Hence, the sum of genetic and epigenetic events may drive tumorigenesis [19]. Multi-omics approaches integrating various genome-wide sequencing techniques, such as DNA methylation and RNA-seq to study the entire genome (study all genes and their combined interactions in a single experiment), epigenome (study epigenetic regulation of all genes) and transcriptome (study the expression of all genes and their combined expression) enable us to a greater extent characterize individual tumors, or specific subtypes [105]. Multi-omics analyses have been suggested as a key to advancing personalized medicine into the clinic [105]. Along with affordable large-scale omics-wide approaches, widespread implementation in clinical practice may also be enabled, thereby allowing personalized treatment-decisions to be further based on the combination of tumor-associated genomic, transcriptomic and epigenomic features. Based on this knowledge, individualized treatment regimens may be available in the future [106].

In **Paper IV**, the aim was to better understand the mechanisms behind genetic, epigenetic and transcriptomic alterations that make up the disease state of early-stage ovarian carcinoma histotypes (CCC, EC, HGSC, MC). This paper comprises a comprehensive characterization of molecular histotype-specific features based on DNA methylation, DNA CNA and RNA-seq data, which may provide an important basis for further research advancing our understanding of various molecular aspects of the development and progression of individual ovarian carcinoma histotypes. More specifically, the RNA-seq data set (n=96) presented in **Paper I** was further analyzed in **Paper IV** to assess differences in gene expression patterns between different histotypes (CCC, EC, HGSC, and MC). **Paper IV** further presents a unique approach by integrating the RNA-seq data with DNA methylation and CNA data (n=91/96 of the RNA sequenced tumors in **Paper I** with sufficient tumor material remaining in the tumor bank), using the same patient cohort.

In general, the DNA methylation analysis revealed that highly methylated CpG sites were more common than unmethylated probes in the ovarian carcinoma cohort. In line with previous DNA methylation reports, CpG sites in regions related to gene expression regulation were generally unmethylated [107]. More specifically, unmethylated CpG sites were commonly identified in genomic regions, *e.g.* promoter, enhancer and exon regions, as well as in CpG islands and shores. A comparison of significant (Benjamini-Hochberg adjusted *P* value<0.05) differentially methylated probes (DMPs), *i.e.* differential probe-specific ratios of methylation, in different histotypes (CCC, EC, HGSC, and MC as case group vs the remaining histotype groups) revealed histotype-specific methylation patterns. Herein, CCC was generally found to be hypermethylated with beta intensity values greater than 0.8 in all histotype comparisons for all genomic regions as well as for regions surrounding CpG islands, and EC was more hypermethylated compared to MC and HGSC (Figure 5). HGSC and MC were mostly hypomethylated in all histotype comparisons. Similar promoter DNA methylation patterns have previously been shown for CCC, EC, and HGSC, but DNA methylation patterns in MC are less studied [108].


a
Number of DMPs
Number of DMPs (3'UTR, Exon bnd)
CCC vs EC
CCC vs MC
CCC vs HGSC
Island
Shore
Shelf
Open sea
Promoter
Enhancer
Gene body
IGR
3' UTR
Exon bnd
Hypermethylated DMPs
Hypomethylated DMPs


b
Number of DMPs
Number of DMPs (3'UTR, Exon bnd)
EC vs CCC
EC vs MC
EC vs HGSC
Island
Shore
Shelf
Open sea
Promoter
Enhancer
Gene body
IGR
3' UTR
Exon bnd
Hypermethylated DMPs
Hypomethylated DMPs


c
Number of DMPs
Number of DMPs (3'UTR, Exon bnd)
HGSC vs CCC
HGSC vs EC
HGSC vs MC
Island
Shore
Shelf
Open sea
Promoter
Enhancer
Gene body
IGR
3' UTR
Exon bnd
Hypermethylated DMPs
Hypomethylated DMPs

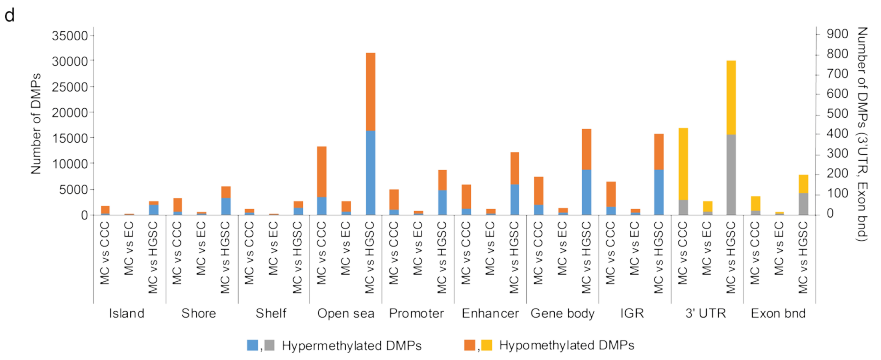

d
Number of DMPs
Number of DMPs (3'UTR, Exon bnd)
MC vs CCC
MC vs EC
MC vs HGSC
Island
Shore
Shelf
Open sea
Promoter
Enhancer
Gene body
IGR
3' UTR
Exon bnd
Hypermethylated DMPs
Hypomethylated DMPs

Figure 5. Histotype-specific DNA methylation patterns. *Column charts displaying the distribution of differentially methylated CpG sites in the different histotype comparisons, subdivided by regions surrounding CpG islands and genomic regions. In column chart (a) CCC, (b) EC, (c) HGSC, and (d) MC are compared with the other histotypes. The regions surrounding CpG islands includes CpG islands (defined as a genomic region being >200bp in length with >50% G and C nucleotide content), CpG shores (0-2kb from CpG islands), CpG shelves (2-4kb from CpG islands), and open sea (>4kb from CpG islands). The genomic regions include the promoter region, which is found 200bp-1500bp upstream of transcriptional start sites, comprising the 1st exon and 5' untranslated region (5' UTR). The genomic region further includes enhancers, gene bodies, intergenic regions (IGRs), 3' UTRs, and exons. 3' UTR and exons follow the right y-axis, whereas the remaining follow the left y-axis. Hypermethylated differentially methylated probes (DMPs) are defined as beta values ≥0.2 (shown in blue and gray) and hypomethylated DMPs are defined as beta values ≤-0.2 (shown in orange and yellow).*

Genome-wide DNA CNA profiles were evaluated in BioDiscovery Nexus Copy Number using single probe resolution CNA data extracted from the DNA methylation data with the conumee package. The normal CNA profile comprises two copies of a gene (one maternal and one paternal gene). In the CNA profile, the nucleus may comprise more copies than normal (CNA gain) or less copies than normal (CNA loss) [109]. A total of 51 recurrent copy number gains and 10 recurrent copy number losses were identified in at least 35% of the patient samples. Chromosome ideograms of chromosomes 1 to 22 revealed recurrent gains on all autosomal chromosomes except for chromosomes 9, 11 and 19, as well as prominent recurrent losses on chromosome arms 4q, 5q, 6p, 8p, 10-12p, 12q, 13q and 19q (Figure 6). The average number of CNAs per patient was significantly higher in HGSC compared to EC patients (Wilcoxon *P* value<0.05), which is supported by previous reports highlighting a high frequency of CNAs in HGSC [110]. No significant difference was found between the other histotypes.

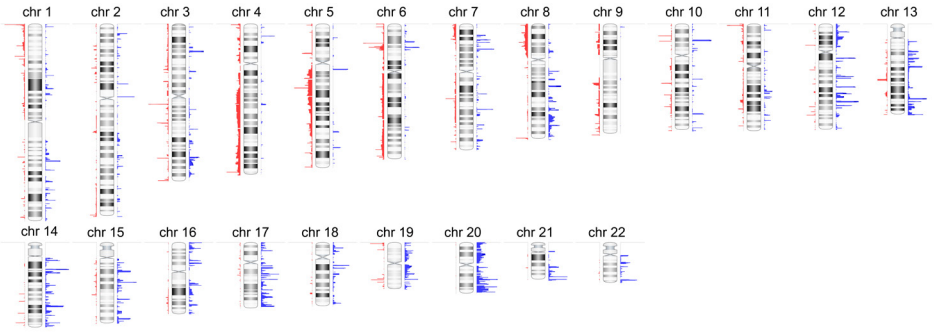


Figure 6. Genome-wide CNA changes in early-stage ovarian carcinoma. *Chromosome ideograms of chromosomes 1 to 22 showing genome-wide CNA changes in the patient cohort as a whole (all histotypes). Genomic gains and losses are shown in blue and red, respectively.*

The extent of chromothripsis-like patterns (CTLPs), wherein a single catastrophic event may have shattered one or more chromosomes, thereby generating genomic rearrangements of the shattered chromosome pieces when the cell attempts to repair the damage, was evaluated using the CTLPScanner with CNA segments as input data (derived from ChAMP) [17]. The CTLPs identified here comprised more than 20 CNA status changes, *e.g.* changes from gain to normal or normal to loss with a log2 ratio difference of at least 0.3. A total of 64 CTLPs, all of which were genomic gains, were identified. The CTLPs were most commonly found on chromosomes 1, 3, 17 and 19 (Figure 7).

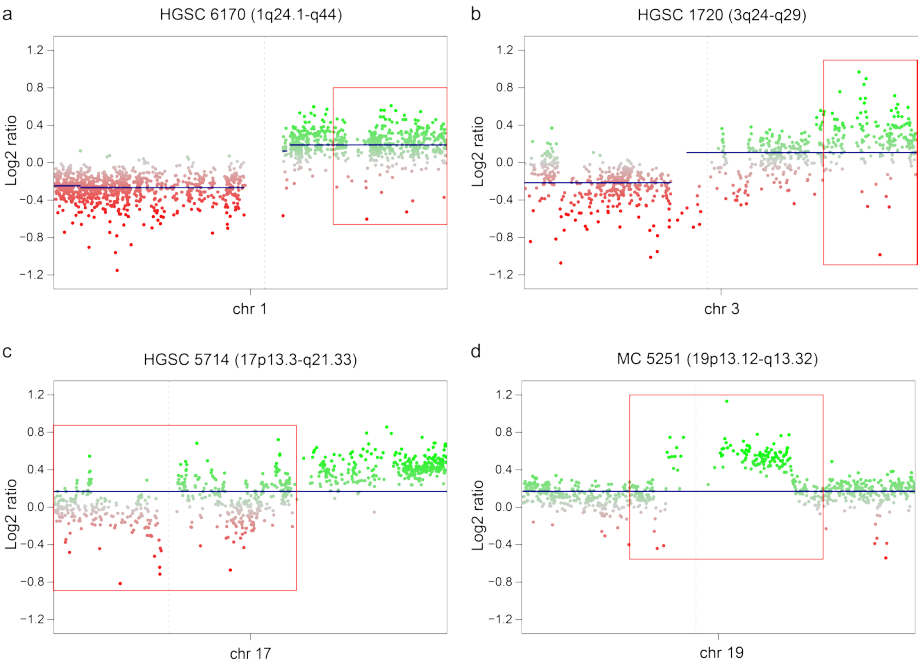


Figure 7. Genomic imbalance in early-stage ovarian carcinomas demonstrated by chromothripsis-like events. *Representative DNA CNA zoom-in plots (a-d) showing chromothripsis-like patterns (CTLPs) on chromosomes 1, 3, 17 and 19 comprising the highest frequency of CTLPs. Red boxes highlight the CTLP region, which is also specified in the parentheses. Genomic gains and losses are shown in green and red, respectively. The centromere dividing the p-arm (left of centromere) and q-arm (right of centromere) is shown as a vertical dashed line.*

RNA expression analysis generally showed lower expression in ovarian carcinoma in comparison with normal ovarian tissue, as well as expression patterns specific to the different histotypes (CCC, EC, HGSC, MC). If a gene has been found to be altered in cancer tissue through multiple mechanisms, *e.g.* altered gene expression, DNA methylation profiles and/or DNA CNA, it may suggest that the gene is important in the initiation or progression of cancer growth [111]. These molecular mechanisms may work together, *e.g.* to make sure the gene is enhanced/silenced. In cancer, oncogenes that promote abnormal cell growth are often overexpressed, and tumor suppressor genes that slow down cell growth may be underexpressed and can thereby contribute to cancer development [112]. Here, we identified 46 putative oncogenes that were overexpressed, hypomethylated and showed genomic gain, and three putative tumor suppressor genes that were underexpressed, hypermethylated and showed genomic loss. Functional studies, *e.g.* overexpression of the candidate gene and knock-out studies, need to be performed in order to validate the identified oncogenes and tumor suppressor genes.

Despite newly developed techniques for comprehensive omics-wide analyses, surprisingly few integrative analyses have been performed in the field of ovarian cancer research. A report from 2019 investigated DNA methylation patterns in different histotypes. For a subgroup of patients (n=47/162), CNA data was also analyzed, and gene expression analysis for 512 genes using NanoString assay was performed for HGSC patients (n=61/162), with only 13 patients analyzed with all three methods [113]. Hence, **Paper IV** is unique in which it analyses early-stage ovarian carcinomas (n=96) using multiple analysis methods on the same patient cohort (with the exception of five samples). This study presents novel ovarian carcinoma histotype-specific data on not only the most commonly occurring and well-studied HGSC histotype, but also the rarer histotypes (CCC, EC, and MC), further restricting the analysis to early-stage tumor.

CONCLUSIONS AND FUTURE PERSPECTIVE

Overall, this doctoral thesis presents novel insights into molecular characteristics associated with early-stage ovarian carcinoma that may improve patient stratification and subclassification based on histotype and clinical outcome. Valuable knowledge is provided for a relatively large patient cohort considering the rarity of the disease and the fact that early-stage tumors are less commonly diagnosed than late-stage tumors. The identified genetic and epigenetic aberrations (*e.g.* genetic variants, fusion genes, CNA, gene expression profiles, DNA methylation patterns, oncogenes, and tumor suppressor genes) may help to better classify ovarian carcinomas by histotype and stratify patients by prognosis using not only clinicopathological features but also molecular tumor features. Furthermore, the identified genetic and epigenetic aberrations may help to determine which patients are in most need of more aggressive treatment regimens and/or targeted therapy. Additional studies need to be performed to validate the clinical significance of the identified aberrations in relation to the stratification and subclassification of the ovarian carcinoma patients using *e.g.* immunohistochemistry and functional studies.

Prognostic signatures relating to individual histotypes were also identified on the transcriptomic level and validated on the protein level. We provide evidence for improved outcome prediction in view of conventional clinicopathological features used in the clinic. The identified signatures may improve our understanding of ovarian tumor progression and may help with prognostication at the time of diagnosis. The prognostic signatures may further assist in the development of future individualized therapeutic strategies for ovarian carcinoma patients. However, additional testing needs to be performed using larger cohorts, and cohorts further comprising late-stage tumors, to further confirm our findings. Moreover, the number of proteins to be tested in each histotype-specific biomarker panel needs to be optimized further and reduced to enable more practical implementation in the clinic.

ACKNOWLEDGEMENTS

There are several people that I would like to thank for making this thesis possible. I would like to direct a special thanks to the people named below.

First of all, my main supervisor **Khalil**. Thank you for this opportunity. I have really enjoyed working in your group. You surround yourself with kind people – just like you. Your door has always been open and I cannot remember a single time where you did not have time to answer any of my questions. Further, your knowledge has been invaluable. **Toshima**, you are simply THE BEST. Please always remember that! You are the heart of the group and you always find all the answers relating to research. Moreover, you are kind, enthusiastic and you help everyone around you. These years would not have been half as fun without you. **Elisabeth**, thank you for being so kind, trustworthy and helpful. It was a lot of fun working with you! **Anikó**, you are the most fast working and effective person I know. Somehow, you always found time to help me, which I really appreciated. **Karin**, thank you for wanting to be involved in my PhD project as my co-supervisor and for your valuable clinical input regarding ovarian cancer. **Per**, thank you for our group meetings and for being interested in new findings. **Alexandra**, thank you for managing practical issues and for being so quick in replying to emails. Thank you **Szilárd**, for your invaluable statistics support, and **Shahin**, for all your help with preparing FFPE sections. **Peter** and **Daniella**, thank you for your company, you are both very nice and positive people**.**

I am going to miss all of you!

My family and friends, especially my **mum**, **dad** and **brother**, and of course my own family **Dominic** and **Juni**.

REFERENCES

1 Vogelstein, B. & Kinzler, K. W. The multistep nature of cancer. *Trends Genet* **9**, 138-141 (1993).

2 Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: implications for targeted therapeutics. *British journal of cancer* **108**, 479-485, doi:10.1038/bjc.2012.581 (2013).

3 Song, Q., Merajver, S. D. & Li, J. Z. Cancer classification in the genomic era: five contemporary problems. *Human genomics* **9**, 27, doi:10.1186/s40246-015-0049-8 (2015).

4 National Cancer Institute, National Institute of Health, Bethesda, MD, USA, 2019

5 Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394-424, doi:10.3322/caac.21492 (2018).

6 Stewart, W. B., Wild, P. C. & Weiderpass, E. (World Health Organization, International Agency for Research on Cancer, Lyon, 2020).

7 Antoni, S., Soerjomataram, I., Møller, B., Bray, F. & Ferlay, J. An assessment of GLOBOCAN methods for deriving national estimates of cancer incidence. *Bulletin of the World Health Organization*, 174-184 (2016).

8 Bergman O., F. L., Hont G., Johansson E., Ljungman P., Munch-Wikland E., Nahi H., Zedenius J. (Cancerfonden, Socialstyrelsen, 2018).

9 Albertson, D. G., Collins, C., McCormick, F. & Gray, J. W. Chromosome aberrations in solid tumors. *Nat Genet* **34**, 369-376, doi:10.1038/ng1215 (2003).

10 Romero-Laorden, N. & Castro, E. Inherited mutations in DNA repair genes and cancer risk. *Current problems in cancer* **41**, 251-264, doi:10.1016/j.currproblcancer.2017.02.009 (2017).

11 Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546-1558, doi:10.1126/science.1235122 (2013).

12 Pon, J. R. & Marra, M. A. Driver and passenger mutations in cancer. *Annu Rev Pathol* **10**, 25-50, doi:10.1146/annurev-pathol-012414-040312 (2015).

13 Milholland, B., Auton, A., Suh, Y. & Vijg, J. Age-related somatic mutations in the cancer genome. *Oncotarget*, 24627–24635, doi:10.18632/oncotarget.5685 (2015).

14 Klymenko, Y. & Nephew, K. P. Epigenetic Crosstalk between the Tumor Microenvironment and Ovarian Cancer Cells: A Therapeutic Road Less Traveled. *Cancers* **10**, doi:10.3390/cancers10090295 (2018).

15 Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1**, 239-259, doi:10.2217/epi.09.33 (2009).

16 Luijten, M. N. H., Lee, J. X. T. & Crasta, K. C. Mutational game changer: Chromothripsis and its emerging relevance to cancer. *Mutation research* **777**, 29-51, doi:10.1016/j.mrrev.2018.06.004 (2018).

17 Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27-40, doi:10.1016/j.cell.2010.11.055 (2011).

18 Rogowski, W. *et al.* Concepts of 'personalization' in personalized medicine: implications for economic evaluation. *Pharmacoeconomics* **33**, 49-59, doi:10.1007/s40273-014-0211-5 (2015).
19 Schweiger, M. R., Barmeyer, C. & Timmermann, B. Genomics and epigenomics: new promises of personalized medicine for cancer patients. *Briefings in functional genomics* **12**, 411-421, doi:10.1093/bfgp/elt024 (2013).
20 Borgfeldt, C. (Regional Cancer Centers, 2019).
21 Bhatla, N. & Jones, A. (World Ovarian Cancer Coalition, 2018).
22 Kurman, R. J., International Agency for Research on Cancer. & World Health Organization. *WHO classification of tumours of female reproductive organs*. 4th edn, (International Agency for Research on Cancer, 2014).
23 Prat, J. Ovarian carcinomas: five distinct diseases with different origins, genetic alterations, and clinicopathological features. *Virchows Arch* **460**, 237-249, doi:10.1007/s00428-012-1203-5 (2012).
24 Kurman, R. J. & Shih Ie, M. The Dualistic Model of Ovarian Carcinogenesis: Revisited, Revised, and Expanded. *Am J Pathol* **186**, 733-747, doi:10.1016/j.ajpath.2015.11.011 (2016).
25 Park, K. J. *et al.* Observations on the origin of ovarian cortical inclusion cysts in women undergoing risk-reducing salpingo-oophorectomy. *Histopathology* **72**, 766-776, doi:10.1111/his.13444 (2018).
26 Fadare, O. *Precancerous Lesions of the Gynecologic Tract*. (2016).
27 Wentzensen, N. *et al.* Ovarian Cancer Risk Factors by Histologic Subtype: An Analysis From the Ovarian Cancer Cohort Consortium. *J Clin Oncol* **34**, 2888-2898, doi:10.1200/jco.2016.66.8178 (2016).
28 Yang-Hartwich, Y. *et al.* Ovulation and extra-ovarian origin of ovarian cancer. *Sci Rep* **4**, 6116, doi:10.1038/srep06116 (2014).
29 Tung, K. H. *et al.* Effect of anovulation factors on pre- and postmenopausal ovarian cancer risk: revisiting the incessant ovulation hypothesis. *American journal of epidemiology* **161**, 321-329, doi:10.1093/aje/kwi046 (2005).
30 Zheng, G. *et al.* Familial risks of ovarian cancer by age at diagnosis, proband type and histology. *PLoS One* **13**, e0205000, doi:10.1371/journal.pone.0205000 (2018).
31 Reid, B. M., Permuth, J. B. & Sellers, T. A. Epidemiology of ovarian cancer: a review. *Cancer biology & medicine* **14**, 9-32, doi:10.20892/j.issn.2095-3941.2016.0084 (2017).
32 Alsop, K. *et al.* BRCA mutation frequency and patterns of treatment response in BRCA mutation-positive women with ovarian cancer: a report from the Australian Ovarian Cancer Study Group. *J Clin Oncol* **30**, 2654-2663, doi:10.1200/jco.2011.39.8545 (2012).
33 Kuchenbaecker, K. B. *et al.* Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *Jama* **317**, 2402-2416, doi:10.1001/jama.2017.7112 (2017).
34 Pearce, C. L. *et al.* Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. *Lancet Oncol* **13**, 385-394, doi:10.1016/s1470-2045(11)70404-1 (2012).

35 Vierkoetter, K. R. *et al.* Lynch Syndrome in patients with clear cell and endometrioid cancers of the ovary. *Gynecol Oncol* **135**, 81-84, doi:10.1016/j.ygyno.2014.07.100 (2014).
36 Buys, S. S. *et al.* Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *Jama* **305**, 2295-2303, doi:10.1001/jama.2011.766 (2011).
37 Jacobs, I. J. *et al.* Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet* **387**, 945-956, doi:10.1016/s0140-6736(15)01224-6 (2016).
38 Henderson, J. T., Webber, E. M. & Sawaya, G. F. Screening for Ovarian Cancer: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *Jama* **319**, 595-606, doi:10.1001/jama.2017.21421 (2018).
39 Maritschnegg, E. *et al.* Lavage of the Uterine Cavity for Molecular Detection of Mullerian Duct Carcinomas: A Proof-of-Concept Study. *J Clin Oncol* **33**, 4293-4300, doi:10.1200/jco.2015.61.3083 (2015).
40 Kinde, I. *et al.* Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. *Sci Transl Med* **5**, 167ra164, doi:10.1126/scitranslmed.3004952 (2013).
41 Wang, Y. *et al.* Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers. *Sci Transl Med* **10**, doi:10.1126/scitranslmed.aap8793 (2018).
42 Måsbäck, A. in *Swedish Society of pathology* (Lund, Sweden, 2019).
43 Torre, L. A. *et al.* Ovarian cancer statistics, 2018. *CA Cancer J Clin* **68**, 284-296, doi:10.3322/caac.21456 (2018).
44 Dahm-Kahler, P., Palmqvist, C., Staf, C., Holmberg, E. & Johannesson, L. Centralized primary care of advanced ovarian cancer improves complete cytoreduction and survival - A population-based cohort study. *Gynecol Oncol* **142**, 211-216, doi:10.1016/j.ygyno.2016.05.025 (2016).
45 Leskela, S. *et al.* The Frequency and Prognostic Significance of the Histologic Type in Early-stage Ovarian Carcinoma: A Reclassification Study by the Spanish Group For Ovarian Cancer Research (GEICO). *Am J Surg Pathol*, doi:10.1097/pas.0000000000001365 (2019).
46 Kobel, M. *et al.* Tumor type and substage predict survival in stage I and II ovarian carcinoma: insights and implications. *Gynecol Oncol* **116**, 50-56, doi:10.1016/j.ygyno.2009.09.029 (2010).
47 Madariaga, A., Lheureux, S. & Oza, A. M. Tailoring Ovarian Cancer Treatment: Implications of BRCA1/2 Mutations. *Cancers* **11**, doi:10.3390/cancers11030416 (2019).
48 *PARP Inhibitors Show Promise as Initial Treatment for Ovarian Cancer,* <https://www.cancer.gov/news-events/cancer-currents-blog/2019/parp-inhibitors-ovarian-cancer-initial-treatment> (2019).
49 Prat, J. Abridged republication of FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum. *Cancer* **121**, 3452-3454, doi:10.1002/cncr.29524 (2015).

50 McCluggage, W. G., Hirschowitz, L., Ganesan, R., Kehoe, S. & Nordin, A. Which staging system to use for gynaecological cancers: a survey with recommendations for practice in the UK. *J Clin Pathol* **63**, 768-770, doi:10.1136/jcp.2010.080978 (2010).
51 Kobel, M. *et al.* Differences in tumor type in low-stage versus high-stage ovarian carcinomas. *Int J Gynecol Pathol* **29**, 203-211, doi:10.1097/PGP.0b013e3181c042b6 (2010).
52 Machida, H. *et al.* Trends and characteristics of epithelial ovarian cancer in Japan between 2002 and 2015: A JSGO-JSOG joint study. *Gynecol Oncol* **153**, 589-596, doi:10.1016/j.ygyno.2019.03.243 (2019).
53 Bell, D. *et al.* Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609-615, doi:10.1038/nature10166 (2011).
54 Bowtell, D. D. *et al.* Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nature reviews. Cancer* **15**, 668-679, doi:10.1038/nrc4019 (2015).
55 Karst, A. M. *et al.* Cyclin E1 deregulation occurs early in secretory cell transformation to promote formation of fallopian tube-derived high-grade serous ovarian cancers. *Cancer Res* **74**, 1141-1152, doi:10.1158/0008-5472.can-13-2247 (2014).
56 Groeneweg, J. W., Foster, R., Growdon, W. B., Verheijen, R. H. & Rueda, B. R. Notch signaling in serous ovarian cancer. *Journal of ovarian research* **7**, 95, doi:10.1186/s13048-014-0095-1 (2014).
57 Konecny, G. E. *et al.* Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. *J Natl Cancer Inst* **106**, doi:10.1093/jnci/dju249 (2014).
58 Kaldawy, A. *et al.* Low-grade serous ovarian cancer: A review. *Gynecol Oncol* **143**, 433-438, doi:10.1016/j.ygyno.2016.08.320 (2016).
59 Ricci, F., Affatato, R., Carrassa, L. & Damia, G. Recent Insights into Mucinous Ovarian Carcinoma. *Int J Mol Sci* **19**, doi:10.3390/ijms19061569 (2018).
60 Wang, Y. K. *et al.* Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. *Nat Genet* **49**, 856-865, doi:10.1038/ng.3849 (2017).
61 Della Pepa, C. *et al.* Ovarian cancer standard of care: are there real alternatives? *Chin J Cancer* **34**, 17-27, doi:10.5732/cjc.014.10274 (2015).
62 Konstantinopoulos, P. A., Ceccaldi, R., Shapiro, G. I. & D'Andrea, A. D. Homologous Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian Cancer. *Cancer Discov* **5**, 1137-1154, doi:10.1158/2159-8290.cd-15-0714 (2015).
63 Patel, J. N. *et al.* Characterisation of homologous recombination deficiency in paired primary and recurrent high-grade serous ovarian cancer. *British journal of cancer* **119**, 1060-1066, doi:10.1038/s41416-018-0268-6 (2018).
64 Liu, Q., Lopez, K., Murnane, J., Humphrey, T. & Barcellos-Hoff, M. H. Misrepair in Context: TGFbeta Regulation of DNA Repair. *Front Oncol* **9**, 799, doi:10.3389/fonc.2019.00799 (2019).

65 Patch, A. M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489-494, doi:10.1038/nature14410 (2015).
66 Tomao, F. *et al.* Parp inhibitors as maintenance treatment in platinum sensitive recurrent ovarian cancer: An updated meta-analysis of randomized clinical trials according to BRCA mutational status. *Cancer Treat Rev* **80**, 101909, doi:10.1016/j.ctrv.2019.101909 (2019).
67 Takano, M., Tsuda, H. & Sugiyama, T. Clear cell carcinoma of the ovary: is there a role of histology-specific treatment? *J Exp Clin Cancer Res* **31**, 53, doi:10.1186/1756-9966-31-53 (2012).
68 Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nature reviews. Genetics* **20**, 631-656, doi:10.1038/s41576-019-0150-2 (2019).
69 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics* **10**, 57-63, doi:10.1038/nrg2484 (2009).
70 Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**, D805-811, doi:10.1093/nar/gku1075 (2015).
71 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164, doi:10.1093/nar/gkq603 (2010).
72 Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
73 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
74 Kumar, S., Vo, A. D., Qin, F. & Li, H. Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. *Sci Rep* **6**, 21597, doi:10.1038/srep21597 (2016).
75 O'Connor, C. Fluorescence in situ hybridization (FISH). *Nature Education* **1**, 171 (2008).
76 in *Illumina* (San Diego, California, USA, 2017).
77 *Infinium Omni2.5-8 Kit*, <https://www.illumina.com/products/by-type/microarray-kits/infinium-omni25-8.html> (2020).
78 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645, doi:10.1101/gr.092759.109 (2009).
79 Heagerty, P. J. & Zheng, Y. Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92-105, doi:10.1111/j.0006-341X.2005.030814.x (2005).
80 Matos, L. L., Trufelli, D. C., de Matos, M. G. & da Silva Pinhal, M. A. Immunohistochemistry as an important tool in biomarkers detection and clinical practice. *Biomarker insights* **5**, 9-20, doi:10.4137/bmi.s2185 (2010).
81 Duraiyan, J., Govindarajan, R., Kaliyappan, K. & Palanisamy, M. Applications of immunohistochemistry. *Journal of pharmacy & bioallied sciences* **4**, S307-309, doi:10.4103/0975-7406.100281 (2012).
82 McCarty, K. S., Jr., Miller, L. S., Cox, E. B., Konrath, J. & McCarty, K. S., Sr. Estrogen receptor analyses. Correlation of biochemical and immunohistochemical

methods using monoclonal antireceptor antibodies. *Archives of pathology & laboratory medicine* **109**, 716-721 (1985).
83 Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45**, e22, doi:10.1093/nar/gkw967 (2017).
84 in *Illumina* (San Diego, California, USA, 2019).
85 Tian, Y. *et al.* ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* **33**, 3982-3984, doi:10.1093/bioinformatics/btx513 (2017).
86 Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428-430, doi:10.1093/bioinformatics/btt684 (2014).
87 Kolde, R. pheatmap: Pretty Heatmaps. R package version 1.0.12 (2019).
88 Phillips, N. YaRrr!: The Pirate's Guide to R. R package version v. 0.1.5 (2017).
89 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
90 Hovestadt, V. & Zapatka, M. conumee: Enhanced copy-number variation analysis using Illumina DNA methylation arrays., R package version 1.9.0 (2017).
91 Leskela, S. *et al.* The Frequency and Prognostic Significance of the Histologic Type in Early-stage Ovarian Carcinoma: A Reclassification Study by the Spanish Group for Ovarian Cancer Research (GEICO). *Am J Surg Pathol* **44**, 149-161, doi:10.1097/pas.0000000000001365 (2020).
92 Serov, S. F. & Scully, R. E. International Histological Classification of Tumors: Histological Typing of Ovarian Tumors, World Health Organization, Geneva. (1973).
93 Tavassoli, F. & Devilee, P. World Health Organization Classification of Tumors: Pathology and Genetics of Tumors of the Breast and Female Genital Organs, IARC, Lyon. (2003).
94 Parker, B. C. & Zhang, W. Fusion genes in solid tumors: an emerging target for cancer diagnosis and treatment. *Chin J Cancer* **32**, 594-603, doi:10.5732/cjc.013.10178 (2013).
95 Roy, L. *et al.* Survival advantage from imatinib compared with the combination interferon-alpha plus cytarabine in chronic-phase chronic myelogenous leukemia: historical comparison between two phase 3 trials. *Blood* **108**, 1478-1484, doi:10.1182/blood-2006-02-001495 (2006).
96 Zou, A., Liu, R. & Wu, X. Long non-coding RNA MALAT1 is up-regulated in ovarian cancer tissue and promotes SK-OV-3 cell proliferation and invasion. *Neoplasma* **63**, 865-872, doi:10.4149/neo_2016_605 (2016).
97 Zhou, Y. *et al.* The Long Noncoding RNA MALAT-1 Is Highly Expressed in Ovarian Cancer and Induces Cell Growth and Migration. *PLoS One* **11**, e0155250, doi:10.1371/journal.pone.0155250 (2016).
98 Eismann, J. *et al.* Transcriptome analysis reveals overlap in fusion genes in a phase I clinical cohort of TNBC and HGSOC patients treated with buparlisib and

olaparib. *Journal of cancer research and clinical oncology* **146**, 503-514, doi:10.1007/s00432-019-03078-9 (2020).
99 Cannan, W. J. & Pederson, D. S. in *Reference Module in Life Sciences* (2017).
100 Wang, Z. C. *et al.* Profiles of genomic instability in high-grade serous ovarian cancer predict treatment outcome. *Clin Cancer Res* **18**, 5806-5815, doi:10.1158/1078-0432.ccr-12-0857 (2012).
101 Abkevich, V. *et al.* Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *British journal of cancer* **107**, 1776-1782, doi:10.1038/bjc.2012.451 (2012).
102 Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature reviews. Genetics* **13**, 227-232, doi:10.1038/nrg3185 (2012).
103 Wang, D. Discrepancy between mRNA and protein abundance: insight from information retrieval process in computers. *Computational biology and chemistry* **32**, 462-468, doi:10.1016/j.compbiolchem.2008.07.014 (2008).
104 Shen, H. & Laird, P. W. Interplay between the cancer genome and epigenome. *Cell* **153**, 38-55, doi:10.1016/j.cell.2013.03.008 (2013).
105 Olivier, M., Asmis, R., Hawkins, G. A., Howard, T. D. & Cox, L. A. The Need for Multi-Omics Biomarker Signatures in Precision Medicine. *Int J Mol Sci* **20**, doi:10.3390/ijms20194781 (2019).
106 LeBlanc, V. G. & Marra, M. A. Next-Generation Sequencing Approaches in Cancer: Where Have They Brought Us and Where Will They Take Us? *Cancers* **7**, 1925-1958, doi:10.3390/cancers7030869 (2015).
107 Aran, D. & Hellman, A. DNA methylation of transcriptional enhancers and cancer predisposition. *Cell* **154**, 11-13, doi:10.1016/j.cell.2013.06.018 (2013).
108 Earp, M. A. & Cunningham, J. M. DNA methylation changes in epithelial ovarian cancer histotypes. *Genomics* **106**, 311-321, doi:10.1016/j.ygeno.2015.09.001 (2015).
109 Delaney, J. R. & Stupack, D. G. *Genomic Copy Number Alterations in Serous Ovarian Cancer*. (IntechOpen, 2018).
110 Testa, U., Petrucci, E., Pasquini, L., Castelli, G. & Pelosi, E. Ovarian Cancers: Genetic Abnormalities, Tumor Heterogeneity and Progression, Clonal Evolution and Cancer Stem Cells. *Medicines (Basel, Switzerland)* **5**, doi:10.3390/medicines5010016 (2018).
111 Santarius, T., Shipley, J., Brewer, D., Stratton, M. R. & Cooper, C. S. A census of amplified and overexpressed human cancer genes. *Nature reviews. Cancer* **10**, 59-64, doi:10.1038/nrc2771 (2010).
112 Jeong, H. M., Kwon, M. J. & Shin, Y. K. Overexpression of Cancer-Associated Genes via Epigenetic Derepression Mechanisms in Gynecologic Cancer. *Front Oncol* **4**, 12, doi:10.3389/fonc.2014.00012 (2014).
113 Bodelon, C. *et al.* Molecular Classification of Epithelial Ovarian Cancer Based on Methylation Profiling: Evidence for Survival Heterogeneity. *Clin Cancer Res* **25**, 5937-5946, doi:10.1158/1078-0432.ccr-18-3720 (2019).