A THESIS SUBMITTED TO
THE UNIVERSITY OF WALES
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

TITLE
THE MOLECULAR GENETIC ETIOLOGY OF OVARIAN CANCER

BY
JOHNATHAN MARK LANCASTER

SEPTEMBER 2004
To Vanessa, Georgina and Sebastian
Summary

Ovarian cancer is the fifth leading cause of cancer death among women in Western Europe and the United States and has the highest mortality rate of all gynecologic cancers. Approximately 75% of cases of epithelial ovarian carcinoma are diagnosed at advanced-stage (III/IV) with disseminated intra-peritoneal metastases, such that the majority of patients succumb to the disease within 5 years. Mortality from the disease has changed little over the last several decades. Despite such dismal statistics, our understanding of the molecular etiology that underlies ovarian cancer development, progression and response to therapy remains incomplete. The recent development of DNA microarrays enables the simultaneous measurement of expression of thousands of genes in a single sample, providing a molecular phenotyping not evident by traditional clinical, molecular or histopathologic methods.

This thesis outlines the characterization of genome-wide expression patterns that underlie ovarian cancer development and metastasis, as well as clinical behavior relating to likelihood of optimal surgical resection, response to chemotherapy, and ultimate survival. Individual genes that contribute to the expression profiles are analysed further to delineate their specific role in ovarian cancer development and progression. Additionally, the contribution of a low penetrance polymorphic allele in the progesterone receptor gene as a risk factor for the development of the disease is examined in a large population-based case-control trial.

Our data suggest that microarray analysis can facilitate the characterization of the molecular basis to ovarian cancer development, metastasis, and response therapy. Specific genes identified in this analysis represent not only potential biomarkers for the presence and clinical behavior of ovarian cancers, but appealing therapeutic targets. Our findings suggest that gene-expression profiles can be developed that can be applied in the clinic to not only provide prognostic information, but predict response to specific chemotherapeutic agents, enabling treatments to be tailored to individual patients with ovarian cancer.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>3.</td>
</tr>
<tr>
<td>Declaration</td>
<td>5.</td>
</tr>
<tr>
<td>Chapter 1.</td>
<td>6.</td>
</tr>
<tr>
<td>Background: The Molecular Genetic Etiology of Ovarian Cancer</td>
<td></td>
</tr>
<tr>
<td>Chapter 2.</td>
<td>60.</td>
</tr>
<tr>
<td>Gene Expression Patterns That Characterize Advanced Stage Serous Cancers</td>
<td></td>
</tr>
<tr>
<td>Chapter 3.</td>
<td>81.</td>
</tr>
<tr>
<td>Patterns of Gene Expression That Characterize Long-Term Survival in Advanced Stage Serous Ovarian Cancers</td>
<td></td>
</tr>
<tr>
<td>Chapter 4.</td>
<td>109.</td>
</tr>
<tr>
<td>High expression of tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) is associated with favorable ovarian cancer survival.</td>
<td></td>
</tr>
<tr>
<td>Chapter 5.</td>
<td>123.</td>
</tr>
<tr>
<td>High expression of Insulin-like Growth Factor Binding Protein-2 (IGFBP2) in epithelial ovarian cancers produces elevated pre-operative serum levels.</td>
<td></td>
</tr>
<tr>
<td>Chapter 6.</td>
<td>139.</td>
</tr>
<tr>
<td>High insulin-like growth factor-2 (IGF-2) gene expression is an independent predictor of poor survival for patients with advanced stage serous epithelial ovarian cancer.</td>
<td></td>
</tr>
<tr>
<td>Chapter 7.</td>
<td>155.</td>
</tr>
<tr>
<td>Identification of genes associated with ovarian cancer metastasis using microarray expression analysis.</td>
<td></td>
</tr>
<tr>
<td>Chapter 8.</td>
<td>186.</td>
</tr>
<tr>
<td>Prediction of optimal versus suboptimal cytoreduction of advanced stage serous ovarian cancer using microarrays.</td>
<td></td>
</tr>
<tr>
<td>Chapter 9.</td>
<td>216.</td>
</tr>
<tr>
<td>Prediction of response to platinum-based chemotherapy for advanced-stage ovarian cancers using gene expression profiles.</td>
<td></td>
</tr>
<tr>
<td>Chapter 10.</td>
<td>244.</td>
</tr>
<tr>
<td>No Association Between Ovarian Cancer Risk and Progesterone Receptor Gene Polymorphism (PROGINS) in Population-Based, Case-Control Study in North America.</td>
<td></td>
</tr>
<tr>
<td>Chapter 11.</td>
<td>256.</td>
</tr>
<tr>
<td>Genetic testing for breast and ovarian cancer susceptibility: issues and dilemmas.</td>
<td></td>
</tr>
<tr>
<td>Statement of Work.</td>
<td>272.</td>
</tr>
<tr>
<td>Acknowledgements.</td>
<td>273.</td>
</tr>
</tbody>
</table>
DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ........................................... (candidate)
Date ............................................

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ........................................... (candidate)
Date ............................................

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed ........................................... (candidate)
Date ............................................

I hereby give consent for my thesis, if accepted to be available for photocopying and for inter-library loans after expiry of bar on access approved by the University of Wales on the special recommendation of the Constituent Institution/University College concerned.

Signed ........................................... (candidate)
Date ............................................
Chapter 1

Background: The Molecular Genetic Etiology of Ovarian Cancer

Published as:

JM Lancaster, L Havrilesky, A Berchuck. The genetic etiology of sporadic ovarian cancer. Chapter in "Diagnosis and Management of Ovarian Disorders" (Igaku-Shoin) 2nd Ed, Altcheck & Deligisch (eds), 2003
INTRODUCTION

Epithelial ovarian cancer is the deadliest gynecologic malignancy and a leading cause of cancer death in women. Malignant transformation of a normal ovarian epithelial cell is caused by genetic alterations that disrupt regulation of proliferation, programmed cell death and senescence. Although approximately 10% of epithelial ovarian cancers arise in women who have inherited mutations in cancer susceptibility genes such as BRCA1 or BRCA2, the vast majority of tumors result from the accumulation of genetic damage over the course of a lifetime – and are referred to as sporadic cancers. Despite significant advances in our knowledge of the alterations that cause sporadic ovarian cancers, much remains unknown regarding their molecular etiology. The recent completion of the initial phases of the human genome project provides the framework for studies that will elucidate our understanding of this complex disease. Hopefully, this knowledge will lead to the development of new approaches to early diagnosis, treatment and/or prevention that will decrease ovarian cancer mortality.

Sporadic Ovarian Cancer is a Genetic Disease

Ovarian cancers exhibit a high degree of genetic disruption that is manifest at both the chromosomal and molecular levels. Karyotype analysis was initially used to demonstrate large chromosomal gains and losses as well as complex translocations. In one study of 23 ovarian cancers, the average number of chromosomal alterations was seven (range 2-14).(1) More recently, studies using a technique called comparative genomic hybridization (CGH) have confirmed that most ovarian cancers have gains or losses of large segments of chromosomes. Of 44 ovarian cancers examined, thirteen areas of chromosomal gain and five areas of chromosomal loss were seen in at least 20% of cases.(2) The most common areas of chromosomal loss were 16q and 17pter-q21, and the most frequent areas of chromosomal gain were 3q25-26 and 8q24. In another study that included 20 sporadic ovarian cancers, the average number of alterations in each cancer was 7.5, and gains were more than twice as common as losses.(3) In agreement with the studies noted above,(2) losses on 17p and gains on
8q23-24 and 3q26 were confirmed in two other studies to be among the most frequent events in ovarian cancers.(3; 4)

Differences exist in the pattern of genetic alterations observed in serous, mucinous and endometrioid ovarian cancers.(5) Using CGH, gains at 1q have been identified frequently in endometrioid and serous tumors. Increased copy number at 10q was seen in endometrioid tumors only, whereas gains at 11q occurred mostly in serous tumors. In mucinous tumors, the most common copy number change was a gain at 17q. Although the findings of this small study, which included only 24 well or moderately differentiated cancers, cannot be considered definitive, they add weight to the theory that there are differences in the molecular pathogenesis of various histologic types of ovarian cancers.

Although using CGH, large chromosomal gains are detected more often than losses, smaller regions of allelic loss at specific genetic loci can frequently be detected using loss of heterozygosity (LOH) analysis. LOH has been demonstrated to occur at a high frequency on many chromosomal arms; including 5q,(6;7) 6q,(8-11) 7p,(6;12) 8p,(13)11p,(14) 11q,(15-19) 13q, (20),14q,(21) 16q,(19) 17p,(22) 17q,(19;23) 22q,(24) and others.(6;20;25) It is unclear whether the extent of these genetic alterations reflects the need to inactivate multiple tumor suppressor genes or is the result of generalized genomic instability. One consistent finding of various studies has been that poorly differentiated, advanced stage cancers have more genetic alterations than early stage, well differentiated or borderline cases.(2;20) (2) For example, in one CGH study, the average number of alterations was 5.4 in low grade ovarian cancers compared to 11.2 in high-grade cases.(2) This finding could be interpreted as reflective of accumulation of genetic changes with evolution of a cancer. On the other hand, it is equally plausible that advanced stage, poorly differentiated cancers are intrinsically more virulent, even early in their development, by virtue of their specific mutations and/or increased genomic instability. If this latter theory is correct, this could have significant implications for early diagnosis of ovarian cancer. Cancers that are inherently more virulent might metastasize rapidly and be less amenable to early detection.

One of the most exciting technologic advances in recent years has been the development of microarray technology, which provides the opportunity to rapidly
evaluate the expression of thousands of genes in a tissue specimen. Several groups have applied this technology to the analysis of ovarian cancers. (26-28) Ono et al. examined the expression patterns of more than 9000 genes in nine ovarian cancers. (26) Fifty-five genes were shown to be commonly up-regulated, and 48 genes down-regulated in cancers relative to normal cells. Differences in expression were identified between mucinous and serous tumors in 115 genes. Consequently, other groups have applied and extended the technology to describe expression patterns for larger numbers of ovarian tumors. (28) To aid interpretation of expression patterns, this group macrodissected normal samples into epithelial and stromal fractions, as well as hybridizing RNA from endothelial and activated B cells, in an effort to identify patterns of expression associated with infiltrating blood vessels and inflammatory cells. Cluster analysis was used to identify differences in patterns of gene expression between normal and tumor tissues, and the most informative genes included the immediate early genes c-fos, jun-B, and EGR-1, and estrogen responsive genes (under-expressed in tumor relative to normal tissue). Several smaller clusters of genes were under-expressed in normal relative to tumor tissue, and included HE4 and PRAME. This group was able to validate their results by RT-PCR of fragments of three of the genes that exhibited differential expression patterns on microarray expression analysis.

Despite the advantages of microarray techniques, the technology is still evolving. At this time, only a subset of genes can be included in any array, and the arrays may be subject to production error and batch-to-batch variability. It is, therefore, not yet clear how accurately the expression data produced reflects true expression, such that ongoing validation using other platforms is essential. Within a cancer specimen, infiltrating normal and inflammatory cells and blood vessels can influence the expression profile considerably, necessitating careful review of tumor histology prior to analysis. Finally, one of the greatest challenges of this technique is data analysis. The interpretation of expression data from over 12,000 genes in current microarrays is daunting, and requires the highest level of expertise in biology, statistics and computation. Future arrays may be even larger as reports from the human genome project estimate that human cells contain about 30,000 genes.
Etiology of genetic alterations

While the identification of specific genetic changes involved in ovarian carcinogenesis is an essential endeavor, an understanding of the etiology of this damage is necessary if we hope to develop effective prevention strategies. Like other sporadic cancers, most epithelial ovarian carcinomas are thought to develop due to the accumulation of genetic alterations with time. Though epidemiological and molecular studies have begun to shed some light on the etiology of ovarian cancer, the exact cause of these genetic alterations remains unclear. It is likely that genetic mutation is a multifactorial process, including effects of age and environmental exposures, on a predisposed genetic background. Despite evidence to suggest that sporadic ovarian cancer generally is a monoclonal disease that originates in the ovarian surface...

Pathogenesis of Ovarian Cancer

ovulation

early menarche +

nulliparity

late menopause

- pregnancy

breastfeeding

OCP

genetic damage

ovarian cancer

apoptosis

Figure 1. Schematic representation of the pathogenesis of ovarian cancer. Ovulation results in genetic damage and can lead to ovarian cancer. Factors that decrease the number of lifetime ovulatory cycles (pregnancy, breast feeding, OCP use) are protective against ovarian cancer, possibly via direct induction of apoptosis of surface epithelial cells. In contrast, early menarche, nulliparity and late menopause, are associated with an increased number of lifetime ovulatory cycles, and increase ovarian cancer risk.
epithelium or underlying inclusion cysts,(29;30) there is evidence that peritoneal tumors in women who carry mutations in BRCA1 or BRCA2, may be polyclonal. (31)

It has been suggested that ovulation may be associated with the development of genetic mutations in the ovarian epithelium (figure 1), and several lines of evidence link ovulation and epithelial ovarian cancer. First, most animals - such as rats and mice - ovulate only when stimulated appropriately, and have a low incidence of epithelial ovarian cancer, whereas chickens and humans ovulate repetitively and have the highest incidence of the disease. In contrast, women with Turner’s syndrome, who are anovulatory, rarely develop epithelial ovarian cancer. Additionally, the observation that pregnancy and oral contraceptive pill use, which decrease lifetime ovulatory cycles, are strikingly protective against ovarian cancer,(32) is consistent with the theory that ovulation is the main driving force underlying the accumulation of genetic damage in the ovarian epithelium (figure 1). Molecular-epidemiological evidence supports an association between the number of ovulatory cycles and somatic DNA damage, as manifest by accumulation of mutant p53 protein.(33) Ovarian cancers that over-express mutant p53, are associated with a greater number of lifetime ovulatory cycles (mean = 388) than those that do not demonstrate p53 overexpression (mean = 342).(33)

It is unclear exactly why repetitive ovulation facilitates the development of ovarian cancer, and several factors including stimulation by gonadotropins, may play a role. One appealing theory is that mutations in the epithelium result from errors in DNA synthesis during the proliferative repair of ovulatory defects, as spontaneous mutations are more likely to occur in cells that are proliferating relative to those at rest.(34) Although DNA repair genes exist to maintain a high degree of fidelity in DNA synthesis, it is estimated that spontaneous errors occur - approximately once every million base pairs - that can elude repair and become fixed in the genome. In addition, the efficiency of the DNA repair systems may vary between individuals due to inherited differences in the activity of DNA repair proteins. It is also possible that mutations in the ovarian epithelium may arise due to oxidative stress as mutagenic free radicals are produced by leukocytes that infiltrate the ovulatory site in the process of corpus luteum regression and repair of the ovarian surface.
Pregnancy and the oral contraceptive pill are protective against ovarian cancer. This effect is not simply due to a decreased rate of mutations, however. The magnitude of protection against ovarian cancer is much greater than would be predicted based simply on the number of ovulations, which were suppressed. Five years of oral contraceptive use decreases risk by approximately 50% while decreasing lifetime ovulatory cycles by only about 10-20%.(32) Recently, however, it has been shown that administration of the progestin levonorgestrel, either alone or in combination with estrogen, stimulates apoptosis of ovarian epithelial cells in macaques.(35) This suggests that the progestagenic milieu of pregnancy and the oral contraceptive pill might protect against ovarian cancer by increasing apoptosis of ovarian epithelial cells, thereby cleansing the ovary of cells that have acquired genetic damage.
MECHANISMS OF MALIGNANT TRANSFORMATION

The number of cells in a population is dependent upon a critical balance between cellular proliferation, senescence, and apoptosis. The mutations that lead to the development of ovarian and other cancers primarily target genes involved in these processes. Development of a cancer results from disruption of these complex regulatory pathways with the net effect being an increased number of cells. Mutations that inactivate DNA repair genes accelerate the accumulation of other cancer-causing mutations. In addition to growth of a primary tumor, cancers are characterized by acquisition of a metastatic phenotype.

<table>
<thead>
<tr>
<th>Table 1. Classes of genes involved in growth regulatory pathways and malignant transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Growth stimulatory (oncogenes)</strong></td>
</tr>
<tr>
<td>Peptide growth factors</td>
</tr>
<tr>
<td>epidermal growth factor (EGF) and transforming growth factor (TGFα)</td>
</tr>
<tr>
<td>heregulin</td>
</tr>
<tr>
<td>Insulin-like growth factors (IGF-I, IGF-II)</td>
</tr>
<tr>
<td>Platelet derived growth factor (PDGF)</td>
</tr>
<tr>
<td>Fibroblast growth factors (FGFs)</td>
</tr>
<tr>
<td>Macrophage-colony stimulating factor (M-CSF)</td>
</tr>
<tr>
<td>Cytoplasmic factors</td>
</tr>
<tr>
<td>Non-receptor tyrosine kinases</td>
</tr>
<tr>
<td>G proteins</td>
</tr>
<tr>
<td>Serine-threonine kinases</td>
</tr>
<tr>
<td>Nuclear factors</td>
</tr>
<tr>
<td>Transcription factors</td>
</tr>
<tr>
<td>Cell cycle progression factors</td>
</tr>
<tr>
<td>Growth inhibitory (tumor suppressor genes)</td>
</tr>
<tr>
<td>Extra-nuclear factors</td>
</tr>
<tr>
<td>Cell membrane factors</td>
</tr>
<tr>
<td>Cell adhesion factors</td>
</tr>
<tr>
<td>Phosphatases</td>
</tr>
<tr>
<td>nuclear factors</td>
</tr>
<tr>
<td>cell cycle inhibitors</td>
</tr>
</tbody>
</table>

Ovarian cancers have the ability to invade the surrounding stroma due to production of proteases that degrade connective tissue and produce angiogenic factors that stimulate the development of new blood vessels to support their growth and spread. Although these molecular pathways are integral to the process of cancer progression, there is little evidence to date to suggest that evolution of the metastatic phenotype is directly
attributable to mutations in genes that encode proteases or other molecules involved in invasion and metastasis.

**Proliferation**

The rate of proliferation is a major determinant of the number of cells in a population. To prevent excessive proliferation, DNA synthesis and cell division are ordinarily restrained. When proliferation is appropriate, these inhibitory mechanisms are inactive and growth stimulatory signals are generated. Malignant tumors are characterized by alterations in genes that control proliferation. There is increased activity of genes that stimulate proliferation (oncogenes) and loss or inactivation of growth inhibitory (tumor suppressor) genes (Table 1). In the past, it was thought that cancer might arise entirely because of more rapid proliferation and/or a higher fraction of cells proliferating. It is now clear that although increased proliferation is a characteristic of many cancers, the fraction of cancer cells actively dividing and the transition time of the cell cycle is not strikingly different than that seen in some normal cells. Increased proliferation is only one of several factors that contributes to cancerous growth.

The fraction of ovarian cancer cells that are actively proliferating can be measured using various techniques. One approach is to assess the DNA content of cells in a sample. This can be accomplished with flow cytometry using disaggregated nuclei or in frozen sections using image analysis. The fraction of cells with a DNA content consistent with S phase can be distinguished from those in the G1 or G2/M phases to calculate a proliferation index. In one study, about 25% of ovarian cancers had an S phase fraction below 5% and this correlated with early stage and favorable survival.(36) Proliferation can also be assessed using immunohistochemical techniques to identify cells that express Ki67 or PCNA, antigens that are expressed only in actively proliferating cells. In most such studies, there has been a correlation between higher proliferation indices (>5-15%) and more advanced stage, worse grade and poor survival.(37-39)
Apoptosis

Cells are capable of activating a suicide pathway of programmed cell death referred to as apoptosis. Apoptosis is an active, energy-dependent process that involves cleavage of the DNA and proteins by endonucleases and proteases. Morphologically, apoptosis is characterized by condensation of chromatin and cellular shrinkage. This is in contrast to the process of necrosis, which is characterized by loss of osmoregulation and cellular fragmentation.

Since the size of a population of cells is normally static due to a balance between the birth rate and the death rate, growth of a neoplasm theoretically could result due to either increased proliferation or decreased apoptosis. In addition to restraining the number of cells in a population, apoptosis may serve an important role in preventing malignant transformation by specifically eliminating cells that have undergone mutations. Following exposure of cells to mutagenic stimuli, including radiation and carcinogenic drugs, the cell cycle is arrested so that DNA damage may be repaired. If DNA repair is not sufficient, apoptosis occurs so cells that have undergone significant damage do not survive. This serves as an anti-cancer surveillance mechanism by which mutated cells are eliminated before they become fully transformed. The p53 tumor suppressor gene is a critical regulator of cell cycle arrest and apoptosis in response to DNA damage, but apoptosis may also be triggered via other pathways under different circumstances.

The molecular events that effect cell death in response to various stimuli have only been partially elucidated thus far, but it appears that a family of genes encoding proteins that reside in the mitochondrial membrane are directly involved. The bcl-2 gene was first of these genes to be identified at a translocation breakpoint in B-cell lymphomas. Expression of bcl-2 acts to inhibit apoptosis and, paradoxically, persistence of bcl-2 expression in ovarian cancers has been associated with favorable prognosis. The bcl-X<sub>L</sub> gene, a structural and functional homologue of bcl-2, also inhibits apoptosis and has been shown to play a role in preventing apoptosis of ovarian cancer cells in response to chemotherapy. Conversely, other related genes such as bax, and bcl-X<sub>S</sub> have pro-apoptotic activity. High bax expression has been reported in 60% of newly diagnosed ovarian cancers and was associated with a favorable response.
to therapy.\(^{45}\) The precise mechanism by which bcl-2 and these other mitochondrial proteins act to regulate apoptosis is unclear, but those that increase membrane permeability stimulate apoptosis while those that decrease permeability prevent apoptosis. Activation of a family of cytosolic proteolytic enzymes called caspases also occurs during apoptosis leading to breakdown of cellular proteins.

**Senescence**

Normal cells are only capable of undergoing division a finite number of times before becoming senescent. Recently, it has been shown that cellular senescence is due to shortening of repetitive DNA sequences (TTAGGG) called telomeres that cap the ends of each chromosome. Telomeres are thought to be involved in chromosome stabilization and in preventing recombination during mitosis. At birth chromosomes have long telomeric sequences that become progressively shorter each time a cell divides. Malignant cells appear to avoid senescence by turning on expression of telomerase activity that acts to lengthen the telomeres.\(^{46,47}\) Telomerase is a ribonucleoprotein complex and both the protein and RNA subunits have been identified. The RNA component serves as a template for telomere extension and the protein subunit acts to catalyze the synthesis of new telomeric repeats.

Because telomerase expression in most normal tissues is restricted to development, it has been suggested that telomerase might be useful diagnostic marker in patients with cancer. Several groups have shown that telomerase activity is detectable in most ovarian cancers.\(^{48,49;49;50;50}\) It has been suggested that persistence of telomerase activity in peritoneal washings after primary therapy for advanced ovarian cancer may predict the presence of microscopic residual disease in some cases despite negative cytologic washings and biopsies.\(^{49}\) Demonstration of the utility of this approach awaits the completion of more definitive studies.

**GROWTH STIMULATORY PATHWAYS:**

**THE ROLE OF ONCOGENES**

Oncogenes encode proteins normally involved in stimulating proliferation, but when these gene products are overactive they contribute to the process of malignant
transformation (Figure 2 and Table 3). Oncogenes can be activated via several mechanisms. In some cancers there is amplification of the number of copies of oncogenes with resultant overexpression of the corresponding protein. Some oncogenes may become overactive when affected by point mutations. Finally, oncogenes may be translocated from one chromosomal location to another and then come under the influence of promoter sequences that cause overexpression of the gene. This latter mechanism frequently occurs in leukemias and lymphomas, but has not been demonstrated in gynecologic cancers or other solid tumors.

**Figure 2.** Classes of oncogenes.

Peptide growth factors: Peptide growth factors in the extracellular space can stimulate a cascade of molecular events that leads to proliferation by binding to cell membrane receptors. Unlike endocrine hormones, which are secreted into the blood stream to act in distant target organs, peptide growth factors usually act in the local environment where they have been secreted. The concept that autocrine growth
stimulation might be a key strategy by which cancer cell proliferation becomes autonomous is intellectually appealing and has received considerable attention. In this model, it is postulated that cancers secrete stimulatory growth factors that then interact with receptors on the same cell. Although increased production of stimulatory growth factors may play a role in enhancing proliferation associated with malignant transformation, they also are involved in development, stromal-epithelial communication, tissue regeneration and wound healing.

It has been shown that ovarian cancers produce and/or are capable of responding to various peptide growth factors. For example, epidermal growth factor (EGF)(51) and transforming growth factor-alpha (TGF-α)(52) are produced by some ovarian cancers that also express the receptor that binds these peptides (EGF receptor).(53;54) Some cancers produce insulin-like growth factor-1 (IGF-1), IGF-1 binding protein, and express type 1 IGF receptor.(55) Further, growth of ovarian cancer cell lines is potently reduced by application of phosphorothioate antisense oligodeoxynucleotides (S-ODNs) to the IGF1 receptor.(56) Platelet derived growth factor (PDGF) also is expressed by many types of epithelial cells including human ovarian cancer cell lines, but these cells usually are not responsive to PDGF.(57-59) In addition, ovarian cancers produce basic fibroblast growth factor (FGF) and its receptor, which can form a mitogenic complex in some ovarian cancers.(60) Ovarian cancers produce macrophage-colony stimulating factor (M-CSF)(61) resulting in elevated M-CSF serum levels in some patients.(62) The expression of the M-CSF receptor (fms) in

![Figure 3. Inactivation of the p53 tumor suppressor gene by "dominant negative" missense mutation or by truncation mutation and deletion](image-url)
levels of M-CSF and IL-6 have been shown to be directly correlated, suggesting a co-regulation in production of these cytokines by tumor cells.(68) In addition to expression of peptide growth factors and their receptors, ascites of patients with ovarian cancer contain phospholipid factors that that stimulate proliferation of ovarian cancer cells.(69;70)

It has been shown that normal ovarian epithelial cells produce - and are responsive to - many of the same peptide growth factors as malignant ovarian epithelial cells.(54;71-73) Thus, despite circumstantial evidence demonstrating the potential for autocrine and paracrine growth regulation of ovarian cancer cells by peptide growth factors, it remains unclear whether alterations in expression of growth factors are critical in the development of ovarian cancer. Peptide growth factors may function as necessary co-factors rather than as the driving force behind malignant transformation.

**Growth factor receptors including the Epidermal Growth Factor Receptor family (EGF Receptor, HER-2/neu):** Cell membrane receptors that bind peptide growth factors are composed of an extracellular ligand binding domain, a membrane spanning region and a cytoplasmic tyrosine kinase domain. Binding of a growth factor to the extracellular domain results in aggregation and conformational shifts in the receptor and activation of the inner tyrosine kinase.(74;75) This kinase phosphorylates tyrosine residues on both the growth factor receptor (autophosphorylation) and targets in the cell interior leading to activation of secondary signals. For example, phosphorylation of phospholipase C leads to breakdown of cell membrane phospholipids and generation of diacylglycerol and inositol-tri-phosphate, both of which play a role in propagation of the mitogenic signal.

The role of the epidermal growth factor (EGF) receptor family of transmembrane receptors and their ligands in growth regulation and transformation has been a prominent focus in cancer research.(76) EGF is a peptide growth factor of 53 amino acids that maintains its secondary structure by virtue of disulphide bonds between cysteine residues. At least five other peptide growth factors, including transforming growth factor-beta (TGF-beta) also interact with and activate the EGF receptor. EFG, TGF-beta and other EGF receptor ligands are produced as pro-forms that are inserted
into the cell membrane. The membrane anchored growth factor can interact with receptors on adjacent cells, a phenomenon known as juxtacrine growth regulation. Alternatively, the active peptide then can be cleaved and released into the extracellular space. The free peptide may interact with receptors on the same (autocrine) or nearby cells (paracrine) to stimulate growth.

The EGF receptor is ubiquitously expressed in both epithelial and stromal cells and plays a role in growth stimulation of most cell types. The EGF receptor has been shown to be amplified in some squamous cancers, and the EGF receptor can be targeted therapeutically with monoclonal antibodies. EGF receptor is expressed in normal ovarian epithelium and although the level of expression varies between cancers this is not a strong predictor of clinical behavior.

The EGF receptor family is often termed the erbB family because the first member identified was the v-erbB oncogene. The second member of the family (erbB2) initially was called neu because it was found to be the transforming gene responsible for the generation of neuroblastomas in rats treated with a chemical carcinogen. This Human EGF Receptor-like molecule was named both HER-2/neu and erbB2 by investigators working in the field. The transforming activity of neu in the animal model was due to the presence of a mutation in the transmembrane portion of the molecule that results in constitutive activation of the inner tyrosine kinase domain. Biochemical studies of HER-2/neu have shown that activation of this receptor is not driven by ligand binding, but rather is dependent on activation of other members of the erbB family (erbB3, erbB4) that heterodimerize with erbB2 and activate its tyrosine kinase domain.

In contrast to EGF receptor, which normally is expressed in both stromal and epithelial cells, HER-2/neu is expressed primarily in epithelial cells. The level of HER-2/neu is increased in some human breast, ovarian and other cancers due to amplification. The SKOV3 ovarian cancer cell line and the SKBR3 breast cancer cell line both have amplification of this gene. In human cancers, HER-2/neu may also be overexpressed due to alterations in regulation of transcription in the absence of gene amplification. Regardless of the underlying mechanism, it has been shown that overexpression occurs in about 20% of ovarian cancers and 30% of breast cancers and
correlates with aggressive features. The level of overexpression in breast cancers generally is higher than in ovarian cancers, however, and some studies have not found overexpression of HER-2/neu to adversely affect prognosis in ovarian cancer.(82;83) It has been shown that transfection of HER-2/neu into normal ovarian epithelial cells can induce a malignant phenotype \textit{in vitro} including the ability of cells to grow in an anchorage independent fashion and to form tumors in nude mice.

As noted above, activation of the \textit{erbB}3 and 4 transmembrane receptors is requisite for HER-2/neu kinase activity. At least four families of ligands, collectively called neuregulins (e.g. Heregulin, \textit{neu} differentiating factor), bind to \textit{erbB}3 and 4.(79) Interestingly, there is considerable promiscuity between \textit{erbB} ligands and receptors. For example, amphiregulin can activate both the EGF receptor (\textit{erbB}1) and \textit{erbB}3. And one of the more recently described ligands (epiregulin) can activate heterodimers of any of the \textit{erbB} family members; and these heterodimers are more potent growth stimulators than homodimers of any individual \textit{erbB} receptor. Although their molecular signaling mechanisms have not yet been fully elucidated, the \textit{erbB} family of receptors have also been exploited as therapeutic targets. Monoclonal antibodies that interact with HER-2/neu can decrease growth of breast and ovarian cancer cell lines that overexpress this receptor.(84;85;85) In addition, these antibodies may enhance the sensitivity of cancers to cytotoxic chemotherapy by interfering with repair of DNA adducts.(86) Recently an anti-HER-2/neu antibody that induces breast cancer regression has been approved for clinical use by the FDA.(87) It is possible that this approach might also benefit some women whose ovarian cancers overexpress HER-2/neu.

**Other Kinases:** Following interaction of peptide growth factors and their receptors, secondary molecular signals are generated to transmit the mitogenic stimulus towards the nucleus. This function is served by a multitude of complex and overlapping signal transduction pathways that occur on the inner cell membrane and cytoplasm. Many of these signals involve phosphorylation of proteins by enzymes known as kinases.(88) Cellular processes other than growth are also regulated by kinases, but one family of kinases appears to have evolved specifically for the purpose of transmitting growth
stimulatory signals. These tyrosine kinases transfer a phosphate group from ATP to tyrosine residues of target proteins. Some kinases that phosphorylate proteins on serine and/or threonine residues are also involved in stimulating proliferation. Although several families of intracellular kinases have been identified that can elicit transformation when activated in vitro, it remains uncertain whether structural alterations in most of these molecules play a role in the development of human cancers. The activity of kinases is regulated by phosphatases, which act in opposition to the kinases by removing phosphates from the target proteins.(89) It has been shown that a number of phosphatases are expressed by ovarian cancers and that some of these oppose the kinase activity of HER-2/neu.(90)

AKT2 is a gene on chromosome 19q that encodes a serine-threonine protein kinase. AKT2 was found to be amplified and overexpressed in 2 of 8 ovarian cancer cell lines and 2 of 15 primary epithelial ovarian cancers.(91) This study was confirmed by a larger series of 132 primary ovarian cancers in which 14% had AKT2 amplification or overexpression. AKT2 overexpression was found have a statistically significant association with higher grade and worse survival.(92) More recently it has been shown that 36% (31/91) of ovarian cancers exhibit elevated AKT2 activity in vitro,(93) and that these tumors tend to be high grade and advanced stage. The same group demonstrated induction of apoptosis in ovarian cancer cells by inhibition of the PI3 kinase/AKT2 pathway by Wortmannin or LY294002. Further studies are needed, however, to confirm the functional significance of AKT2 overexpression in ovarian cancers.

Using comparative genomic hybridization, the region of chromosome 3p26 that includes the regulatory subunit of phosphatidylinositol 3-kinase (PIK3CA), has been shown to be amplified in some ovarian cancer cell lines and 40% of primary ovarian cancers.(94) Interestingly, the AKT2 gene is one of the downstream targets of PIK3CA. Thus, theoretically, amplification of either of these two genes leads to excessive activation of this mitogenic pathway.

Binding of cytokines and growth factors also activate the Janus kinase (JAK) family of protein-tyrosine kinases (PTKs). These kinases phosphorylate various signaling proteins including a unique family of transcription factors termed the signal
transducers and activators of transcription, or STATs. Phosphorylation of STATs results in their transactivation, nuclear translocation, DNA binding, and dimerization, and constitutively activated STATs have been described in breast and prostate cancer cell lines and primary tumors.\(^{95;96}\) Constitutive activation of STAT 3 has been shown to be present in the ovarian cancer cell lines MDAH 2774, OV-1063, Caov-3, and O.C. 22819, but absent in normal ovarian surface epithelium.\(^{97;98}\) Further, STAT3 activation can be blocked by JAK inhibitors such as AG490 and apoptosis induced.\(^{97}\) Interestingly, activation of both STATs and JAKs has also been demonstrated to be associated with enhanced expression of the breast/ovarian cancer susceptibility gene, BRCA1, in human prostate cancer cell lines.\(^{96}\) Such findings underscore the critical role of the JAK/STAT system in cell signaling and provide preliminary evidence for their involvement in ovarian tumorigenesis.

**G proteins:** G proteins represent another class of molecules involved in transmission of growth stimulatory signals in towards the nucleus.\(^{99;100}\) They are located on the inner aspect of the cell membrane and have intrinsic GTPase activity that catalyzes the exchange of GTP (guanine-tri-phosphate) for GDP (guanine-di-phosphate). In their active GTP bound form, G proteins interact with kinases that are involved in relaying the mitogenic signal. Conversely, hydrolysis of GTP to GDP, which is stimulated by GTPase activating proteins (GAPs), leads to inactivation of G proteins. The ras family of G proteins are among the most frequently mutated oncoproteins in human cancers (eg. gastrointestinal and endometrial cancers). Activation of ras genes usually involves point mutations in codons 12, 13 or 61 that result in constitutively activated molecules.

Mutations in the ras genes do not appear to be a common feature of invasive serous epithelial ovarian cancers.\(^{101;102;103;103}\) K-ras mutations have been noted more frequently in mucinous ovarian cancers, but these tumors comprise only a small fraction of epithelial ovarian cancers. In contrast, K-ras mutations are common in borderline epithelial ovarian tumors, occurring in 20-50% of cases.\(^{104;105}\) Indeed, mutations in K-ras have recently been found in mullerian inclusion cysts in some patients with serous borderline tumors, suggesting that these two entities may be related.\(^{106}\)
**Nuclear factors:** If proliferation is to occur in response to signals generated in the cytoplasm, these events must lead to activation of nuclear factors responsible for DNA replication and cell division. Expression of several genes that encode nuclear proteins increases dramatically within minutes of treatment of normal cells with peptide growth factors. Once induced, the products of these genes bind to specific DNA regulatory elements and induce transcription of genes involved in DNA synthesis and cell division. When inappropriately overexpressed, however, these transcription factors can act as oncogenes. Among the nuclear transcription factors involved in stimulating proliferation, amplification and/or overexpression of members of the *myc* family has most often been implicated in the development of human cancers. (107;108) Myc proteins are key regulators of mammalian cell proliferation and treatment of cells with myc antisense oligonucleotides inhibits proliferation. It has been shown that myc acts as part of a heterodimeric complex with the protein max to initiate transcription of other genes involved in cell cycle progression. (107)

Amplification of the c-myc oncogene occurs in some epithelial ovarian cancers. In five small studies, c-myc was reported to be amplified in a total of 24 of 77 cases (31%). (109-113) In a more recent study in which 51 epithelial ovarian cancers were analyzed, a similar incidence of c-myc overexpression was observed (37%). (114) c-myc overexpression was more frequently observed in advanced stage serous adenocarcinomas, suggesting a role in tumor progression.

**GROWTH INHIBITORY PATHWAYS:**

**TUMOR SUPPRESSOR GENES**

Tumor suppressor genes encode proteins normally involved in inhibiting proliferation and inactivation of these genes plays a role in the development of most cancers. Knudson’s “two-hit” model established the paradigm that both alleles must be inactivated in order to exert a phenotypic effect on tumorigenesis. (115) The location and the type of the inactivating mutations in tumor suppressor genes may vary from one cancer to the next. Frequently, mutations in tumor suppressor genes alter the base
sequence resulting in the production of a premature stop codon (TAG, TAA or TGA) and truncated protein product. Such truncated protein products may result from several types of mutational events, including nonsense mutations, in which a single base substitution changes that sequence from a specific amino acid to a stop codon. In addition, microdeletions or insertions of one or several nucleotides that disrupt the reading frame of the DNA (frameshifts) also lead to the generation of stop codons downstream in the gene. In some cases, missense mutations occur which change only a single amino acid in the encoded protein. The functional significance of such a change depends upon the amino acid alteration and the location within the gene. A mutation in one allele, whether germline or somatic, is revealed following somatic inactivation of the corresponding wild-type allele, typically by deletion of part or all of the chromosome. This loss of heterozygosity (LOH) has become recognized as the hallmark of tumor suppressor gene inactivation.

Tumor suppressor genes may also be inactivated by methylation of the promoter region of the gene,(116) which is proximal to the coding sequence and regulates whether or not the gene is transcribed from DNA into RNA. When the promoter is methylated it is resistant to activation and the gene is essentially silenced despite remaining structurally intact. Like oncogenes, tumor suppressor gene products are found throughout the cell (Table 2). In this section the various classes of tumor suppressor genes involved in sporadic epithelial ovarian cancers will be reviewed.

<table>
<thead>
<tr>
<th>Oncogenes</th>
<th>FUNCTION</th>
<th>ALTERATION</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>HER-2/neu</td>
<td>tyrosine kinase</td>
<td>overexpression</td>
<td>20%</td>
</tr>
<tr>
<td>K-ras</td>
<td>G protein</td>
<td>mutation</td>
<td>5%</td>
</tr>
<tr>
<td>AKT2</td>
<td>serine/threonine kinase</td>
<td>amplification</td>
<td>10%</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>serine/threonine kinase</td>
<td>overexpression</td>
<td>20-30%</td>
</tr>
<tr>
<td>c-myc</td>
<td>transcription factor</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tumor suppressor genes</th>
<th>FUNCTION</th>
<th>ALTERATION</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1</td>
<td>DNA repair</td>
<td>mutation/deletion</td>
<td>5%</td>
</tr>
<tr>
<td>p53</td>
<td>transcription factor</td>
<td>mutation/deletion</td>
<td>60%</td>
</tr>
<tr>
<td>p16</td>
<td>cdk inhibitor</td>
<td>homozygous deletion/methylation</td>
<td>715%</td>
</tr>
<tr>
<td>p27</td>
<td>cdk inhibitor</td>
<td>decreased expression</td>
<td>740%</td>
</tr>
</tbody>
</table>

Extra-nuclear tumor suppressor genes: Although most tumor suppressor gene products are nuclear proteins, some extra-nuclear tumor suppressors have been identified.
Theoretically, any protein that is normally involved in inhibition of proliferation could conceivably act as a tumor suppressor. In this regard, phosphatases that normally oppose the action of the tyrosine kinases by dephosphorylating tyrosine residues are appealing candidates.(89) Analysis of deletions on chromosome 10q23 in human cancers led to the discovery of the PTEN gene(117) In addition to its phosphatase activity, PTEN is homologous to the cytoskeleton proteins tensin and auxin and it has been postulated that PTEN might act to inhibit invasion and metastasis through modulation of the cytoskeleton.(118) Interestingly, it has been shown that PIK3CA and AKT2 kinase activity can be specifically opposed by the PTEN phosphatase. PTEN mutations are rare in serous ovarian cancers, perhaps because amplification of PIK3CA or AKT2 abrogates the need for loss of PTEN tumor suppressor function. In contrast, PTEN mutations occur in about 20% of endometrioid ovarian cancers.(119)Transforming growth factor-β (TGF-β): The transforming growth factor-beta (TGF-β) family of growth factors inhibits proliferation of normal epithelial cells.(120) It is thought that TGF-β causes cell cycle arrest in G₁ by triggering pathways that result in inhibition of cyclin dependent kinases. Several closely related forms of TGF-β have been discovered which are encoded by separate genes (TGF-β1, TGF-β2, TGF-β3 and Placental TGF-β). TGF-β is secreted from cells in an inactive form bound to a portion of its precursor molecule from which it must be cleaved to release biologically active TGF-β. Active TGF-β interacts with type I and type II cell surface TGF-β receptors and initiates serine/threonine kinase activity.(121) Prominent intracellular targets include a class of molecules called Smads that translocate to the nucleus and act as transcriptional regulators.(122) Although mutations in the TGF-β receptors and Smads have been reported in some cancers, this does not appear to be a feature of ovarian cancers.

Normal ovarian epithelial cells produce, activate and are growth inhibited by TGF-β.(123) however, most immortalized ovarian cancer cell lines have lost the ability to either produce, activate or respond to TGF-β.(58;123-127) This suggested that TGF-β might normally act as an autocrine growth inhibitory factor in normal ovarian epithelium, and that loss of this pathway might play a role in the development of some
ovarian cancers. Although convenient to work with, immortalized cell lines frequently have undergo profound genetic alterations in tissue culture. Examination of primary ovarian cancers obtained directly from patients revealed that in almost all cases cancers were sensitive to the growth inhibitory effect of TGF-β.(128) Thus, it remains unclear whether alterations in the TGF-β pathway play a role in the development of ovarian cancers.

**P53 tumor suppressor gene:** Mutation of the p53 tumor suppressor gene is the most frequent genetic event described thus far in human cancers (Figure 3 and 4).(129-131) The p53 gene encodes a 393 amino acid protein that appears to play a central role in the regulation of both proliferation and apoptosis.(132-134) In normal cells, p53 protein resides in the nucleus and exerts its tumor suppressor activity by binding to transcriptional regulatory elements of genes such as the cdk inhibitor p21 that act to arrest cells in G1. Beyond simply inhibiting proliferation, normal p53 is thought to play a role in preventing cancer by stimulating apoptosis of cells that have undergone

![Figure 4.](image)

Figure 4. Spectrum of p53 mutations in advanced ovarian cancers. Missense mutations cluster in exons 5-8, whereas truncation mutations are more evenly dispersed throughout the gene.
excessive genetic damage. In this regard, p53 has been described as the "guardian of the genome" since it delays entry into S phase until the genome has been cleansed of mutations. If DNA repair is inadequate, p53 may initiate apoptosis, thereby eliminating cells with genetic damage.

Many cancers have missense mutations in one copy of the p53 gene that result in substitution of a single amino acid in exons 5 through 8, which encode the DNA binding domains (Figure 4). Although these mutant p53 genes encode full-length proteins, they are unable to bind to DNA and regulate transcription of other genes. Mutation of one copy of the p53 gene often is accompanied by deletion of the other copy, leaving the cancer cell with only mutant p53 protein. If the cancer cell does retain one normal copy of the p53 gene, the mutant p53 protein can form a complex with wild-type p53 protein and prevent it from interacting with DNA. Because inactivation of both p53 alleles is not required for loss of p53 function, mutant p53 is said to act in a "dominant negative" fashion. While normal cells have low levels of p53 protein because it is rapidly degraded, missense mutations encode protein products that are resistant to degradation and over-accumulate in the nucleus; and overexpression of mutant p53 protein can be detected immunohistochemically. A smaller fraction of cancers have mutations in the p53 gene that encode truncated protein products. Whereas missense mutations in the p53 gene cluster in exons 5-8, truncation mutations are more evenly dispersed throughout the gene, presumably because they inactivate the p53 protein regardless of their location (Figure 4). In cases of p53 truncation mutations, deletion of the other allele occurs as the second event as is seen with other tumor suppressor genes.

Alteration of the p53 tumor suppressor gene is the most frequent genetic event described thus far in ovarian cancers. The frequency of overexpression of mutant p53 is significantly higher in advanced stage III/IV disease (40-60%) relative to stage I cases (10-20%). In addition, p53 inactivation is uncommon in borderline tumors. The higher frequency of p53 overexpression in advanced stage cases may indicate that this is a "late event" in ovarian carcinogenesis. Alternatively, the loss of p53 may confer a more aggressive metastatic phenotype. In advanced stage ovarian cancer, there is a suggestion that overexpression of p53 may be associated with
somewhat worse survival. (137;139-143) (145) The literature is not entirely consistent and most studies have not been large enough or optimally designed to yield reliable prognostic information. Finally, although there is a high concordance between p53 missense mutations and protein overexpression, about 20% of advanced ovarian cancers contain mutations that result in truncated protein products that generally are not overexpressed. (136) Overall, about 70% of advanced ovarian cancers have either missense or truncation mutations in the p53 gene.

The finding that overexpression of mutant p53 tumor suppressor genes is associated with high lifetime ovulatory cycles in women with ovarian cancer is consistent with the hypothesis that ovulation-associated proliferation may be the cause of these mutations in the ovarian epithelium. (33) In addition, most of the p53 point mutations in ovarian cancers are transitions rather than transversions. (22;146) which also suggests that these mutations occur spontaneously, rather than due to exogenous carcinogens.

It has been suggested that loss of p53 might confer a chemoresistant phenotype, because p53 plays a role in chemotherapy induced apoptosis. In this regard, several studies have examined the correlation between chemosensitivity and p53 mutation in ovarian cancers in vitro. (147-150) (42) Some have suggested a relationship between p53 mutation and loss of chemosensitivity, but in other equally valid studies such a relationship has not been observed. (151) The status of the p53 gene is probably one of a number of factors that determines sensitivity to chemotherapy.

**BRCA1 and BRCA2:** Inherited mutations of the BRCA1 and 2 genes on chromosome 17q and 13q, respectively, are the most frequent cause of hereditary ovarian cancers. Prior to the identification of BRCA1 and 2, it had been anticipated that somatic mutations in BRCA1/2 would be common in ovarian cancers, since more than half of these cancers exhibit loss of heterozygosity in the region of BRCA1, on chromosome 17q. (23) (152;153) Initially, two small studies reported somatic mutations in BRCA1 in about 10% of 54 ovarian cancers, (154;155) but somatic mutations were not seen in two larger studies. (153;156) In these initial studies, mutational screening was performed using single stranded conformation analysis.
More recently, a large study in which complete sequencing of the BRCA1 gene was performed in 103 ovarian cancers, somatic mutations were found in at least 7 cases.(157) In contrast to women with germline BRCA1 mutations whose median age at ovarian cancer diagnosis is typically in the mid 40s, the median age of women with somatic mutations was about 60. Similar to ovarian cancers with germline BRCA1 mutations, all of the ovarian cancers with somatic BRCA1 mutations were serous. In addition, loss of the wild-type BRCA1 allele invariably accompanied somatic BRCA1 mutations. This data is supportive of the hypothesis that loss of BRCA1 function occurs by way of the classic tumor suppressor paradigm with mutation of one copy and deletion of the other. The BRCA2 gene, which is responsible for some hereditary ovarian cancer cases, also has been examined for somatic mutations, but none have been identified.(158)

**DNA mismatch repair (MMR) genes:** Ovarian cancer is a feature of the hereditary non-polyposis colorectal cancer/Lynch II syndrome, which is an autosomal dominant condition caused by mutation of one of several DNA mismatch repair genes. Patients with HNPCC are at risk of cancers of the colon, endometrium, ovary, stomach, small intestine, liver and biliary tract, brain, and transitional cell carcinoma of the ureters and renal pelvis. The lifetime risk of ovarian cancer for women with HNPCC syndrome is approximately 10% (159,160).

Multiple genes have been implicated in this hereditary syndrome, including *hMSH2* (human mutS homolog 2) on chromosome 2p16; *hMLH1* (human mutL homolog 1) on chromosome 3p21; *hPMS1* and *hPMS2* (human postmeiotic segregation 1 and 2) on chromosomes 2q31 and 7q11, respectively; *hMSH6* on chromosome 2p16; and *hMSH3* on chromosome 5q11.2-q13.2 (161,162). However, *hMSH2* and *hMLH1* are thought to account for the majority of mutations of the MMR genes found in HNPCC families (161,162).

Estimates of the frequency of microsatellite instability (a marker for MMR defective genes) in sporadic ovarian cancers have been very variable, but overall do not support a significant role for the mismatch repair genes in sporadic ovarian carcinogenesis (163,164).
**Retinoblastoma tumor suppressor gene:** Initiation of the cell cycle with resultant cell division is dependent on progression through the G1 phase of the cycle into the DNA synthetic S phase. The Retinoblastoma gene (Rb), which was the first tumor suppressor gene discovered, plays a central role in actively regulating this process. In the early G1 phase of the cell cycle Rb protein binds to the E2F transcription factor and prevents it from activating transcription of other genes involved in cell cycle progression. When Rb is phosphorylated E2F is released and stimulates entry into the DNA synthesis phase of the cell cycle (Figure 5). Mutations in the Rb gene have been noted primarily in retinoblastomas and sarcomas, but rarely in other types of cancers. Loss of heterozygosity at the Rb locus occurs in about 30% of ovarian cancers, but mutations in the gene have not been detected and functional Rb protein is present despite loss of one copy of the gene.

![Diagram](attachment:image.png)

**Figure 5.** Regulation of cell cycle progression at G1/S by p53 cyclins, cdks, cdk inhibitors and Rb. In early G1, the Rb protein binds to E2F and prevents activation of other genes involved in progression. Rb phosphorylation by a family of cyclin dependent kinases and associated cyclins results in E2F release and stimulates entry into S phase of the cell cycle. A family of cdk inhibitors (p15, p16, p21, p27) prevent phosphorylation of Rb by cyclin-cdk complexes.
Cyclins, cyclin dependent kinases (cdks) and cdk inhibitors: Phosphorylation of Rb serves as a final common pathway with respect to initiation of proliferation; and this process is tightly controlled because of its critical importance (Figure 5). Rb is phosphorylated by a family of cyclin dependent kinases (cdk2, 4,6) and associated cyclins (cyclin D, E), which act as regulatory subunits. Conversely, a family of cdk inhibitors (p15, p16, p21, p27) has been described that prevent phosphorylation of Rb by cyclin-cdk complexes. Although many of the intricacies of regulation of G1 progression remain poorly understood, it is clear that inappropriately high activity of cyclins and cdks or loss of cdk inhibitors facilitates malignant transformation. Several alterations in these classes of genes have been described in human cancers including overexpression of cyclin D and loss of p16.

The p16 cdk inhibitor is the most frequently altered of the genes involved in regulating Rb phosphorylation.(116) The p16 gene on chromosome 9p21 encodes a protein that inhibits cdk4 or cdk6/cyclinD complexes from phosphorylating Rb. Initially it was noted that both copies of the p16 gene are deleted in a high fraction of immortalized cancer cell lines, including the ovarian cancer cell line, SKOV3.(169) The hypothesis that p16 loss plays a significant role in malignant transformation was strengthened by the finding that it is inactivated in some familial melanoma kindreds. Although the p16 gene also is inactivated by mutations or deletion of both alleles in some sporadic cancers, this occurs much less frequently than in immortalized cell lines. More commonly, however, the p16 gene appears to be silenced due to methylation of its promoter, which prevents transcription.

The role of the p16 gene in ovarian cancer has been studied extensively, but results are conflicting. Some studies have reported that the p16 gene is homozygously deleted or the promoter methylated in a fraction of cases,(170-172) whereas other studies have not found p16 deletions, mutations or methylation.(173-175) The inconsistency of these various reports likely reflects the technical difficulty of assaying promoter methylation and homozygous deletions in primary tumor samples. In addition, some groups have reported that some ovarian cancers have very high levels of p16 protein, (176) but the underlying mechanism and significance of this observation has not yet been elucidated. Finally, the p14arf protein arises from an alternative reading frame
of the p16 gene and has been shown to increase p53 expression by decreasing its
degradation.(177) Deletions of the p16 locus would lead to loss of p14arf expression,
which also could have significant consequences for regulation of G1 progression.

There is some evidence to suggest that decreased activity of other cdk inhibitors
may also play a role in the development of some cancers. The cdk inhibitor p21 binds
cdk2 and the proliferating cell nuclear antigen (PCNA) during the G1/S phase of the cell
cycle. Transfection of p21 into SKOV3 and OVCAR3 cell lines produced a reduction of
tumor cell growth and increased susceptibility to platinum induced apoptosis,
suggesting a possible role for p21 as an adjunct to platinum-based chemotherapy.(178)
Reduced expression of the p27 cdk inhibitor, which is encoded by a gene on
chromosome 12p, has been noted in some cancers due to increased p27 degradation.
About one-third of ovarian cancers have been noted to have decreased p27 expression
and this correlated with poor outcome.(179;180) In addition, overexpression of cyclins
D and/or cyclin E has been noted in some cancers. The cyclin D gene on chromosome
11q13 is translocated or amplified in some human cancers. Although the levels of
cyclin D appear to be high in some ovarian cancers, overexpression has not been
shown to be due to amplification or translocation.(181;175) Likewise, cyclin E levels are
high in some ovarian cancers, particularly clear cell tumors.(182)

Other cyclins, cdks and regulatory molecules such as the chk family are involved
in regulating progression from G2 to M. Alterations in these pathways clearly play a role
in the development of some cancers, but the intricacies of G2/M transition are less well
understood than those of G1/S. Studies of G2/M in ovarian cancers to date are
preliminary and have not yielded evidence of significant alterations.

Other genes: Several other known tumor suppressor genes including WT1 and APC
have been examined in ovarian tumors, but do not appear to be altered frequently in
these cancers. Recently, novel putative ovarian cancer tumor suppressor genes have
been described that are expressed in normal ovarian epithelial cells, but not in ovarian
cancers. One of these is a ras homologue named NOEY2 that was described by the
group at the MD Anderson Cancer Center.(183) The SPARC gene, which encodes an
extracellular matrix protein that is involved in adhesion,(184) and the DOC2 gene, which
is a GRB2 binding protein,(185) were described by the group at Brigham and Women’s Hospital. Finally, LOT-1, a transcription factor(186) and the OVCA1 and OVCA2 genes on chromosome 17p(187) were described by the group at Fox Chase Cancer Center. The role of these and other, as yet undiscovered, tumor suppressor genes in the development of ovarian cancers remains to be defined by future studies.
REFERENCES


region in ovarian cancers to a 300-kb segment of chromosome 6q27. *Cancer Res*, 56:5586-5589.


29. Jacobs, I.J., Kohler, M.F., Wiseman, R.W., Marks, J.R., Whitaker, R.,
(1992): Clonal origin of epithelial ovarian carcinoma: analysis by loss of
heterozygosity, p53 mutation, and X-chromosome inactivation. *J Natl Cancer
Inst*, 84:1793-1798.

unifocal origin of advanced epithelial ovarian cancer and for minor clonal

31. Schorge, J.O., Muto, M.G., Welch, W.R., Bandera, C.A., Rubin, S.C., Bell,
papillary serous carcinoma of the peritoneum in patients with germline BRCA1

to ovarian cancer risk. Collaborative analysis of twelve US case-control studies:
IV. The pathogenesis of epithelial ovarian cancer. *Am J Epidemiol*, 136:1212-
1220.

between lifetime ovulatory cycles and overexpression of mutant p53 in epithelial

34. Ames, B.N. and Gold, L.S. (1990): Too many rodent carcinogens:

35. Rodriguez, G.C., Walmer, D.K., Cline, M., Krigman, H.R., Lessey, B.A.,
Whitaker, R.S., Dodge, R.K., and Hughes, C.L. (1998): Effect of progestin on the
ovarian epithelium of macaques: Cancer prevention through apoptosis? *J Soc


17q11-21 in cancers of women who have both breast and ovarian cancer. *Am J Obstet Gynecol*, 172:908-913.


Chapter 2

Gene Expression Patterns That Characterize Advanced Stage Serous Ovarian Cancers

Published as:
JM Lancaster, HK Dressman, RS Whitaker, L Havrilesky, J Gray, JR Marks, JR Nevins, A Berchuck. Gene Expression Patterns That Characterize Advanced Stage Serous Ovarian Cancers., 11(1) p51-59,
Journal of the Society of Gynecologic Investigation, 2004
ABSTRACT

Objective: To identify gene expression patterns that characterize advanced stage serous ovarian cancers using microarray expression analysis.

Methods: We have used genome-wide expression analysis to compare a series of 31 advanced stage (III/IV) serous ovarian cancers from patients who survived either < 2 years or > 7 years with 3 normal ovarian epithelial samples. Array findings were validated by analysis of expression of the IGFBP2 and TRAIL genes using quantitative real-time PCR.

Results: Hierarchical clustering identified patterns of gene expression that distinguished cancer versus normal ovarian epithelium. Additionally, we have identified gene expression patterns that distinguish cancers on the basis of patient survival. These genes include many that are associated with immune function. Expression of IGFBP2 and TRAIL genes measured by array and QRT-PCR analysis demonstrated correlation coefficients of 0.63 and 0.78, respectively.

Conclusions: Global expression analysis can identify expression patterns and individual genes that contribute to ovarian cancer development and outcome. Many of the genes that determine ovarian cancer survival are associated with the immune response suggesting that immune function influences ovarian cancer virulence. With the generation of newer arrays with more transcripts, larger studies are possible to fully characterize genetic signatures that predict survival that may ultimately be used to guide therapeutic decision-making.
INTRODUCTION

Epithelial ovarian cancer is the deadliest gynecologic malignancy and is a leading cause of cancer death in women. It is believed that malignant transformation of normal ovarian epithelial cells is caused by genetic alterations that disrupt regulation of proliferation, apoptosis, senescence and DNA repair. The alterations that lead to ovarian cancer vary between patients and the striking clinical heterogeneity of ovarian cancer likely reflects an underlying molecular heterogeneity. Although approximately 10% of epithelial ovarian cancers arise in women who have inherited mutations in cancer susceptibility genes such as BRCA1 or BRCA2, the vast majority result from the sporadic accumulation of genetic damage over the course of a lifetime. Our group has shown for example, that the p53 tumor suppressor gene is frequently mutated in ovarian cancers, and that about 70% of advanced stage serous cases harbor p53 mutations (1-4). Other tumor suppressor genes implicated in the development of some ovarian cancers include p16, p21 and p27 (3;5). In addition, several putative ovarian cancer tumor suppressor genes that are expressed in normal ovarian epithelial cells, but not in ovarian cancers have been described - including NOEY2 (6), SPARC (7), DOC2 (8), LOT-1transcription factor (9), and OVCA1 and OVCA2 (10). Finally, oncogenes such as K-ras, HER2/neu, c-myc, PIC3KA and AKT2 are activated by mutation, amplification and/or over-expression in a fraction of ovarian cancers (11). However, when controlling for stage and grade, none of the single gene alterations identified in ovarian cancers to date correlate strongly with clinical outcome.

Recent advances in the use of DNA microarrays, that allow global assessment of gene expression in a single sample, has shown that gene expression profiles can provide molecular phenotyping that identifies distinct classifications not evident by traditional histopathological methods. Classifications of leukemias and lymphomas that have been achieved in recent analyses of gene expression patterns represent a significant step in the development of methodologies to phenotype tumors (12;13). Several groups have applied expression array technology to the analysis of ovarian cancers (14-20). These studies have demonstrated differences in gene expression patterns between normal ovarian epithelial cells and ovarian cancers and between various grades, stages and BRCA1/2 characteristics of ovarian cancers.
Most ovarian cancer deaths occur in women with advanced stage serous carcinomas and median survival among those with stage III/IV disease is only about 3 to 4 years. However, some women with advanced ovarian cancer survive much longer and it has been postulated that this may be attributable to less aggressive cancer growth, more favorable chemosensitivity and/or more effective host immune response. We have sought to characterize the global gene expression patterns that distinguish advanced stage serous epithelial ovarian cancers using microarray expression analysis.

MATERIAL AND METHODS

Tissues: We identified frozen cancer specimens from 46 women with advanced (FIGO stage III/IV) serous adenocarcinomas. In addition, normal ovarian surface epithelial cells from four women were grown in monolayer culture as described previously (21) and then harvested for array analysis. All of these tissues were collected under the auspices of a Duke IRB approved protocol with written informed consent. Microscopic slides from each cancer sample were examined and only those with >60% tumor (on a per cell basis), with few infiltrating lymphocytes or necrotic tissue, were selected for RNA extraction. All ovarian cancers were obtained at initial cytoreductive surgery from patients treated at Duke University Medical Center who then received platin-based chemotherapy. Twenty-four patients died within 2 years of diagnosis (short-term survivors), and 22 patients survived more than 7 years (long-term survivors). At primary surgery, 7/24 (29%) short-term survivors were optimally cytoreduced (maximum diameter of residual nodules less than 1cm) compared to 18/23 (78%) long-term survivors.

Microarrays: We analyzed 7,070 full-length human genes using Affymetrix HumanGeneFL arrays in an effort to characterize gene expression patterns in advanced stage serous ovarian cancers and to identify genes that impact survival in these patients. Approximately 30 mg of tumor tissue was added to a chilled BioPulverizer H tube (Bio101). Lysis buffer from the Qiagen RNeasy Mini kit was added and the tissue was homogenized for 20 seconds in a Mini-Beadbeater (Biospec Products). Tubes
were spun briefly to pellet the garnet mixture and reduce foam. The lysate was transferred to a new 1.5 ml tube using a syringe and 21 ga. needle, followed by passage through the needle 10 times to shear genomic DNA. Total RNA was extracted using the Qiagen RNeasy Mini kit (Qiagen Inc., Valencia, CA). RNA integrity was confirmed by visualization of the 28S:18S ribosomal RNA ratio on a 1% agarose gel. 10 micrograms of total RNA was used to develop the targets for Affymetrix DNA microarray analysis, prepared according to the manufacturer's instructions. Biotin-labeled cRNA, produced by in vitro transcription, was fragmented and hybridized to the HumanGeneFL (Affymetrix GeneChip) arrays at 45°C for 16 hr and then washed and stained using the GeneChip Fluidics. The array was scanned by a GeneArray Scanner and patterns of hybridization were detected as light was emitted from the fluorescent reporter groups that have been incorporated into the target and hybridized to oligonucleotide probes. The average difference measurements computed in the Affymetrix Microarray Analysis Suite (v.4.0) serve as a relative indicator of the level of expression (complete data is located at http://data.cgt.duke.edu/JSGI.php).

**Quantitative Real-time PCR (QRT-PCR):** Total RNA was extracted using the RNeasy extraction kit (Qiagen Inc., Valencia, CA). One microgram of total RNA was reverse transcribed with the Roche First Strand cDNA kit (Roche, Basel, Switzerland). Complementary DNA was quantified by fluorometry, and 5ng subjected to 40 cycles of QRT-PCR in the Roche Light Cycler (Roche, Basel, Switzerland), in the presence of intron-spanning cDNA primers to the house-keeping gene, beta-microglobulin, to confirm adequate cDNA normalization. In the same way, QRT-PCR was used to determine relative expression of the TRAIL and IGFBP2 genes in the presence of intron-spanning cDNA primers for each gene (TRAIL; 5'-AGACCTGCTGCTGATCGTG -3' and 5'-TTATTTGCGCCAGGCC -3', and IGFBP2; 5'- TGGAGGAGCCCAAGAGCT-3' and 5'-GGTTCACACACCAGCACTC-3'), 10 units of AdvanTaq Plus DNA Polymerase with PCR reaction buffer (Clontech, Palo Alto, CA), dNTP, and Syber Green. Serial dilutions of the most highly expressed samples were used to create a concentration curve and relative expression levels calculated for each sample.
**Statistical analysis:** Of the 46 tumors and 4 control samples processed, 31 tumors and 3 control samples passed quality control criteria and were used for subsequent data analysis. The remaining 16 samples produced poor quality scans. Scans were rejected if the proportion of the "present" calls was less than 35% and scaling factors were greater than 8. The average difference values determined by the Affymetrix Microarray Analysis Suite (v4.0) were used in the analysis. A hierarchical clustering method was used to group genes on the basis of similarity in the pattern with which their expression varied over all the samples (normal versus cancer tissue, and short versus long term survival). When comparing normal versus cancer tissues, we focused on 2,655 genes (about 38% of the 7,070 genes) whose transcripts exhibited the highest amount of variation as defined by the standard deviation and when comparing short versus long term survival, we focused in the same manner on 2,501 genes (about 35% of the 7,070 genes). Genes were centered based on adjusting the values of each gene to reflect their variation from some property of the series of the median observed value. Normalization was used by setting the magnitude (sum of the squares of the values) of a row/column vector to 1.0. Genes were clustered using the Pearson correlation coefficient as the distance metric, and clusters were determined by the average linkage. Hierarchical clustering was applied through the Cluster software program, and the results were visualized in Treeview (Eisen, M et al., PNAS 95:14863—14868 (1998); [http://www.microarrays.org/software](http://www.microarrays.org/software), for computational details). The expression level of each gene relative to the median expression level across all samples was represented by color, red represents expression greater than the mean, and green represents expression less than the mean, and the intensity of each color represents the magnitude of the deviation from the mean. Relative gene expression calculated by microarray and QRT-PCR were subject to linear model regression and correlation coefficient analysis, as well as analysis of variance (ANOVA) using Statgraphics Plus 5.0 software.
RESULTS

We have made use of Affymetrix DNA microarrays to profile the gene expression patterns that characterize the difference between advanced serous ovarian cancers and normal ovarian epithelium. In addition, we have also identified patterns that distinguish tumors on the basis of survival in these patients. Gene expression of the transcripts on the Human GeneFL array for all 34 samples are represented as the average difference value as computed by the Affymetrix Microarray Analysis Suite (v.4.0). The data is available online at: http://data.cgt.duke.edu/JSGI.php

Normal ovarian epithelium versus ovarian cancers

We compared a series of 3 normal ovarian epithelium samples against 31 advanced stage epithelial ovarian cancers. RNA was prepared from each and used to generate probes for hybridization to Affymetrix DNA microarrays. The hybridization results were normalized and analyzed as described in Methods. From this dataset, we applied hierarchical clustering and focused on 2655 genes (about 38% of the 7,070 genes) that exhibited a high degree of variability as defined by the standard deviation.

A display of the clustering results for the 2655 genes (Figure 1A) reveals a tight clustering of the normal ovarian epithelium samples when compared to ovarian cancers. It is possible to clearly identify gene expression patterns that distinguish the normal from cancer samples (Figure 1B). The cluster of genes which most clearly differentiate normal epithelium from cancer include a variety of genes previously shown to play a role in cell growth regulatory pathways or oncogenesis. Cell growth genes include TGFβ, MARKS, MIG2, IGFBP1, TNFα, and the p21 cyclin kinase inhibitor. Additional genes such as GLI, DOC1 (downregulated in ovarian cancer), SERPINE1, and thrombospondin 1 have been shown to participate in oncogenic events or to be deregulated in cancers.

Tables 1A & B list the genes demonstrating the highest level of up- and down-regulation in ovarian cancers compared to normal ovarian surface epithelium.
Figure 1 - Hierarchical cluster analysis of ovarian cancer versus normal ovarian samples. The figure demonstrates the genetic interactions among samples. The heatmap illustrates the expression levels of different genes across the samples. Grey represents lower expression levels, while red indicates higher expression levels. The clusters suggest distinct genetic profiles between normal and ovarian cancer samples.
<table>
<thead>
<tr>
<th>Gene name</th>
<th>Accession Number</th>
<th>Symbol</th>
<th>Fold change (control vs cancer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Ig rearranged gamma chain mRNA, V-J-C region and complete cds</td>
<td>M63438</td>
<td>IGKC</td>
<td>6734</td>
</tr>
<tr>
<td>Human omega light chain protein 14.1 (lg lambda chain related) gene,</td>
<td>M34516</td>
<td></td>
<td>6489</td>
</tr>
<tr>
<td>Human omega light chain protein 14.1 (lg lambda chain related) gene</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>folate receptor 1 (adult)</td>
<td>U20391</td>
<td></td>
<td>4686</td>
</tr>
<tr>
<td>kallikrein 6 (neurosin, zyme)</td>
<td>U62801</td>
<td>KLK6</td>
<td>2152</td>
</tr>
<tr>
<td>X76223</td>
<td>X76223</td>
<td>MAL</td>
<td>1495</td>
</tr>
<tr>
<td>Human MHC class II HLA-DP light chain mRNA, complete cds</td>
<td>M57466</td>
<td>HLA-DPB1</td>
<td>1361</td>
</tr>
<tr>
<td>kallikrein 7 (chymotryptic, stratum corneum)</td>
<td>L33404</td>
<td>KLK7</td>
<td>1143</td>
</tr>
<tr>
<td>ephrin-A1</td>
<td>M57730</td>
<td>EFNA1</td>
<td>1051</td>
</tr>
<tr>
<td>CD47 antigen (Rh-related antigen, integrin-associated signal transducer)</td>
<td>X69398</td>
<td>CD47</td>
<td>946</td>
</tr>
<tr>
<td>preferentially expressed antigen in melanoma</td>
<td>U65011</td>
<td>PRAME</td>
<td>928</td>
</tr>
<tr>
<td>keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Kobner/Weber-Cockayne types)</td>
<td>M21389</td>
<td>KRT5</td>
<td>923</td>
</tr>
<tr>
<td>defensin, beta 1</td>
<td>X92744</td>
<td>DEFB1</td>
<td>851</td>
</tr>
<tr>
<td>estrogen receptor 1</td>
<td>X03635</td>
<td>ESR1</td>
<td>847</td>
</tr>
<tr>
<td>SPARC-like 1 (mast9, hevin)</td>
<td>X86693</td>
<td>SPARCL1</td>
<td>841</td>
</tr>
<tr>
<td>small inducible cytokine subfamily B (Cys-X-Cys), member 10</td>
<td>X02530</td>
<td>CXCL10</td>
<td>555</td>
</tr>
<tr>
<td>KIAA0246 protein</td>
<td>D87433</td>
<td>STAB1</td>
<td>491</td>
</tr>
<tr>
<td>integrin, beta 2 (antigen CD18 (p95), lymphocyte function-associated</td>
<td></td>
<td>ITGB2</td>
<td></td>
</tr>
<tr>
<td>antigen 1; macrophage antigen 1 (mac-1) beta subunit</td>
<td>X64082</td>
<td></td>
<td>463</td>
</tr>
<tr>
<td>hemoglobin, beta</td>
<td></td>
<td></td>
<td>459</td>
</tr>
<tr>
<td>FXYD domain-containing ion transport regulator 3</td>
<td>X93036</td>
<td>FXYD3</td>
<td>448</td>
</tr>
<tr>
<td>lymphotoxin beta (TNF superfamily, member 3)</td>
<td>U89922</td>
<td>LTB</td>
<td>395</td>
</tr>
<tr>
<td>sodium channel, nonvoltage-gated 1 alpha</td>
<td>X76180</td>
<td>DC12</td>
<td>380</td>
</tr>
<tr>
<td>stratfin</td>
<td>X57348</td>
<td>SFN</td>
<td>371</td>
</tr>
<tr>
<td>adrenergic, alpha-2C-, receptor</td>
<td>U72648</td>
<td></td>
<td>367</td>
</tr>
<tr>
<td>U29195</td>
<td>U29195</td>
<td>NPTX2</td>
<td>364</td>
</tr>
<tr>
<td>envoplakin</td>
<td>U53786</td>
<td>EVPL</td>
<td>343</td>
</tr>
<tr>
<td>Gene name</td>
<td>Accession Number</td>
<td>Symbol</td>
<td>Fold change (control vs cancer)</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>------------------</td>
<td>-----------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Human vascular endothelial growth factor related protein</td>
<td>U43142</td>
<td>VEGFC</td>
<td>-294</td>
</tr>
<tr>
<td>Beta-1-Glycoprotein, Domains N And Ila, Pregnancy-Specific</td>
<td></td>
<td></td>
<td>-198</td>
</tr>
<tr>
<td>Beta-1-Glycoprotein 1, Pregnancy-Specific</td>
<td></td>
<td></td>
<td>-135</td>
</tr>
<tr>
<td>S77094, class A, 20 probes, 20 in S77094 1075-1621, nicotinic acetylcholine receptor alpha subunit</td>
<td>S77094</td>
<td>CHRNA1</td>
<td>-126</td>
</tr>
<tr>
<td>prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)</td>
<td></td>
<td>PTGS2</td>
<td>-91</td>
</tr>
<tr>
<td>Sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A</td>
<td>U52840</td>
<td>SEMA5A</td>
<td>-85</td>
</tr>
<tr>
<td>bone morphogenetic protein 2</td>
<td>M22489</td>
<td>BMP2</td>
<td>-62</td>
</tr>
<tr>
<td>EPICAN, ALT. SPlice 1</td>
<td></td>
<td></td>
<td>-58</td>
</tr>
<tr>
<td>ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)</td>
<td>M64595</td>
<td>RAC2</td>
<td>-55</td>
</tr>
<tr>
<td>GLIOMA PATHOGENESIS-RELATED PROTEIN</td>
<td>U16307</td>
<td>GLIPR1</td>
<td>-52</td>
</tr>
<tr>
<td>Y00081</td>
<td>Y00081</td>
<td>IL6</td>
<td>-48</td>
</tr>
<tr>
<td>potassium large conductance calcium-activated channel, subfamily M, alpha member 1</td>
<td>U11717</td>
<td>KCNMA1</td>
<td>-45</td>
</tr>
<tr>
<td>small inducible cytokine subfamily B (Cys-X-Cys), member 6 (granulocyte chemotactic protein 2)</td>
<td>U83303</td>
<td>CXCL6</td>
<td>-43</td>
</tr>
<tr>
<td>X04500</td>
<td>X04500</td>
<td>IL1B</td>
<td>-39</td>
</tr>
<tr>
<td>ONCOGENE TC21</td>
<td></td>
<td></td>
<td>-38</td>
</tr>
<tr>
<td>INTEGRIN, ALPHA 5 (FIBRONECTIN RECEPTOR, ALPHA POLYPEPTIDE)</td>
<td>X06256</td>
<td>ITGA5</td>
<td>-34</td>
</tr>
<tr>
<td>insulin-like growth factor binding protein 5</td>
<td>M62782</td>
<td>IGFBP5</td>
<td>-31</td>
</tr>
<tr>
<td>prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)</td>
<td>U04636</td>
<td>PTGS2</td>
<td>-31</td>
</tr>
<tr>
<td>fibroblast growth factor 2 (basic)</td>
<td>J04513</td>
<td>FGF2</td>
<td>-31</td>
</tr>
<tr>
<td>sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican)</td>
<td>X73608</td>
<td>SPOCK</td>
<td>-29</td>
</tr>
<tr>
<td>insulin-like growth factor binding protein 5</td>
<td>M65062</td>
<td>IGFBP5</td>
<td>-28</td>
</tr>
<tr>
<td>serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2</td>
<td>M31551</td>
<td>SERPINB2</td>
<td>-28</td>
</tr>
<tr>
<td>supenorcervical ganglia, neural specific 10</td>
<td>S82024</td>
<td>STMN2</td>
<td>-27</td>
</tr>
</tbody>
</table>
Gene expression patterns that characterize patient survival
We analyzed microarray expression data for the 31 advanced serous epithelial ovarian cancers to determine if specific gene expression patterns could be identified that reflected patient survival. Of the 31 samples, 14 represented long-term survivors (> 7 years) and 17 were short-term survivors (< 2 years). Although these samples did not segregate in the clustering based on analysis of the total gene expression profiles, it was nevertheless clear that gene expression clusters could be identified that coincided with this distinction in patient survival.

One such example is shown in Figure 2 that illustrates a cluster of genes whose expression relates closely to the survival characteristics. Interestingly, this group of genes includes a substantial number of genes encoding immune system activities (Table 2). These include IL2 receptor, IL4 receptor, chemokine ligands (CCL4 and CCL5), T cell receptor (α and β), LSP1, CD53, CD48, CD14, CD37, and several interferon pathway activities (IFI30 and ISGF3/STAT1).
secreted phospholipase A2 (osteopontin), bone sialoprotein I, early T-lymphocyte activation 1

interleukin 1 receptor, common (severe combined immunodeficiency)

interleukin 1 receptor

small inducible cytokine A2 (monocyte chemotactic protein 1, homologous to mouse Sig-β)

HLA-I and HLCA-II human histocompatibility loci

human transcription factor ISRE-3, HRE sequence

interferon, common inducible protein 28

regulator of G-protein signaling 1

cytoskeleton 1, secretory granule

cytotoxic, alpha-hemolytic protein

integrin beta 2 (antigen CD18, p95), lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit

CD163 antigen

epithelial membrane protein 3

CD16 antigen

leukocyte-associated multiadhesive membrane protein-1

Fe receptor of IgE, high affinity receptor for; omega solvastide

intercellular beta 2 (antigen CD18, p95), lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit

CD53 antigen

complement component 1, q subcomponent, beta polypeptide

alpha-1-macroglobulin

Apo B100 dissociation inhibitor (BDI) beta

lincomycin SMG antigen partner-like 2

Rb104646 protein

heme oxygenase (decoulin) 1

collagen stimulation factor 1 receptor, formerly McDonough feline sarcoma viral (r-fms) oncogene homolog

arachidonic acid metabolizing activating protein

monokine induced by gamma interferon

major histocompatibility complex, class I, 20 alpha 1

H1B14, class I, 10 alpha, 10 in H1B14 protein OA77-2373; human class II histocompatibility antigen beta-2 chain, HLA DRB1

beta cell receptor beta locus

T cell receptor alpha locus

V00563

small inducible cytokine B4 (homologous to mouse Kip-2b)

small inducible cytokine B5 (GRANTS)

gamma 2 epsilon catalytic activator protein

tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide

Figure 2 – Hierarchical cluster analysis of ovarian cancers from long- (>84 months) versus short-term (<24 months) survivors. Forty-three genes are shown to display a gene expression patterns that reflect the relationship of the genes to short and long term survival. Each gene is represented by a single row and each sample is represented by a single column. The color red represents log expression ratios of overexpressed genes and the color green represents log expression ratios of under expressed genes. Black represents log expression ratio of 0 (similar expression in both samples), and gray represents insufficient data. This Figure demonstrates the genes that represented a pattern that had a pervasive order reflecting relationships among the genes in short term versus long term survival patients.
Validation studies using quantitative real time QRT-PCR

To validate differences in patterns of expression identified by microarray analysis, we used quantitative real time (QRT) PCR. Additional mRNA was available for 29 of the 31 samples used for microarray data analysis, and these 29 samples were subject to QRT-PCR analysis. The tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) and insulin-like growth factor binding protein 2 (IGFBP2) genes were chosen for further study on the basis of differential expression demonstrated by microarray analysis between normal and cancer tissue.

Figure 3A and 3B demonstrate the results of fitting a linear model to describe the relationship between TRAIL and IGFBP2 expression measured by (QRT) PCR and microarray analysis. P-values from ANOVA tables for expression of both genes measured by (QRT) PCR and microarray analysis were less than 0.01, indicating a statistically significant relationship between gene expression measured by (QRT) PCR and microarray analysis at the 99% confidence level. Comparison of relative expression levels for TRAIL using microarray expression analysis and QRT-PCR revealed correlation coefficients of 0.78, indicating a moderately strong relationship between the variables (Figure 3A). Similarly, comparison of relative expression levels for IGFBP2 using microarray expression analysis and QRT-PCR revealed correlation coefficients of 0.63 (Figure 3B).

**Figure 3:** Results of fitting a linear regression model to describe the relationship between TRAIL and IGFBP2 expression measured by (QRT) PCR and microarray analysis. P-values from ANOVA tables for expression of both genes measured by (QRT) PCR and microarray analysis were <0.01. **A)** Comparison of relative expression levels for TRAIL using microarray expression analysis and QRT-PCR (correlation coefficient, 0.78). **B)** Comparison of relative expression levels for IGFBP2 using microarray expression analysis and QRT-PCR (correlation coefficients, 0.63).
Figure 3A

IGFBP2 expression measured by QRT-PCR vs. Microarray analysis

Figure 3B

TRAIL expression measured by QRT-PCR vs. Microarray analysis
DISCUSSION

Traditional methods of phenotypic characterization of cancer are often limited and do not have the ability to discern subtle differences important in developing a better understanding of the natural history of neoplastic progression necessary for advancing therapeutic strategies. Since underlying molecular alterations are a main determinant of clinical behavior, there has been a significant focus on the development of independent molecular prognostic markers that would augment histopathologic and clinical staging techniques. Although some of the specific molecular changes associated with malignant transformation in various cell types have been elucidated, single gene alterations generally have not been sufficiently predictive to supplant or augment conventional clinical staging or prognostic systems. The development of arrays that assess expression of thousands of genes represents a more powerful approach to the analysis of cancer behavior.

Previous work has identified numerous genes that appear to be up or down-regulated in the process of ovarian epithelial malignant transformation. For example, Ono et al. examined the expression pattern of more than 9000 genes in nine ovarian cancers (18). Fifty-five genes were commonly up-regulated, and 48 genes down-regulated in cancers relative to normal cells. Differences in expression of 115 genes were identified between mucinous and serous tumors. Similarly, Wong et al compared expression profiles of 7 normal ovarian epithelial cell lines to 7 immortalized ovarian cancer cell lines using the MICROMAX human cDNA system, which contains 2,400 known genes (14). Shridhar et al., examined differences in 10,000 known genes and 13,000 ESTs using cDNA microarrays from Research Genetics, Inc.(16) The analysis included 21 early stage and 17 advanced stage ovarian cancers, as well as immortalized cell lines and normal ovarian epithelial cells. They confirmed up-regulation of several genes in ovarian cancers and most of these genes were up-regulated in both early and advanced stage cases. Jazaeri et a., (2002) used cDNA microarrays to compare gene expression patterns in BRCA1 and BRCA2 associated ovarian cancers with sporadic ovarian cancers (20). They identified 110 genes that exhibited statistically significant differences in expression levels between BRCA1- and BRCA2-associated ovarian cancers. This group of 110 genes segregated sporadic tumors into "BRCA1-
like" and "BRCA2-like," suggesting that BRCA1-related and BRCA2-related pathways are also involved in sporadic ovarian cancers.

Consistent with findings of other groups, our analysis identifies a variety of genes that are differentially expressed between normal ovarian epithelium and ovarian cancer, and many of these genes are known to play a role in cell growth regulatory pathways or oncogenesis (Table 1). Cell growth genes include TGFβ, MARKS, MIG2, IGFBP1, TNFα, and the p21 cyclin kinase inhibitor. Additional genes such as GLI, DOC1 (downregulated in ovarian cancer), SERPINE1, and thrombospondin 1 have been shown to participate in oncogenic events or to be deregulated in cancers. In our analysis, the VEGF-C gene was down-regulated in ovarian cancers compared to controls. Interaction between vascular endothelial growth factor (VEGF) and angiopoietins (Angs) is known to be critical in the remodeling of host vasculature, which support angiogenesis during tumor growth. As such, our finding of down-regulation of VEGF-C in cancer is somewhat surprising, and it is possible that this represents an artifact based on in-vitro growth of normal ovarian surface epithelial cells. It is also possible however, that remodeling of host vasculature to support angiogenesis during tumor growth involves a more complex interaction of the VEGF and Angs/Tie-2 system in host stroma endothelial cells than is currently appreciated.

Most importantly, we have successfully applied microarray expression technology to identify gene expression profiles that may be associated with survival in advanced stage serous epithelial ovarian cancers. Figure 2 illustrates a cluster of genes whose expression relates closely to the survival characteristics, and includes a substantial number of genes encoding immune system activities. This suggests a role for immune function genes in determining cancer virulence and ultimate patient survival (Table 2). These genes include IL2 receptor, IL4 receptor, chemokine ligands (CCL4 and CCL5), T cell receptor (α and β), LSP1, CD53, CD48, CD14, CD37, and several interferon pathway activities (IFI30 and ISGF3/STAT1). Previous reports have also implicated immune response genes in ovarian cancer pathogenesis. Soluble interleukin-2 receptors (sIL-2R) concentrations in patient sera have been shown to be associated with outcome in patients with ovarian cancer. Gebauer et al., demonstrated significantly higher concentrations of sIL-2R in preoperative sera of 119 patients with ovarian cancer.
compared to 130 patients with benign tumors (22). Further, for patients with advanced stage disease, elevated sIL-2R serum levels were associated with better prognosis (22). Chemokine CCL2, -3, -4, -5, -8, and -22 expression (and the corresponding proteins) have previously been demonstrated in ovarian cancer cell isolates from ascites, when a direct correlation was found between the CCL5 concentration and CD3(+) T-cell infiltration (23). Many of the immune system genes that we found to be differentially expressed are elevated in cancers from long-term survivors. It is possible to speculate that these patients have a more favorable clinical outcome because of a more highly active immune response. If the association between immune function and survival is confirmed, it raises the possibility that agents that enhance immune function may have utility as therapeutic tools.

We were able to satisfactorily validate our microarray expression findings for 2 genes (IGFBP2 and TRAIL), identified early in our analysis as demonstrating differential expression between normal ovarian surface epithelium and ovarian cancer, using QRT-PCR as a separate platform. We feel that the correlation coefficients of 0.63 and 0.78 for IGFBP2 and TRAIL gene expression measured by array and QRT-PCR offers some support for the integrity of our array analysis. As array technology is applied more widely such ongoing validation steps remain essential. We will continue to extend our QRT-PCR analysis of interesting candidate genes with the additional aim of identifying and characterizing novel ovarian cancer molecular markers.

We have clearly demonstrated that global expression analysis can identify individual genes that contribute to ovarian cancer virulence. Such information not only enhances our understanding of ovarian cancer biology, but also facilitates the identification of novel therapeutic avenues. This study also offers preliminary evidence that microarray expression analysis may be able to identify gene expression signatures that can predict clinical outcome and ultimately guide therapeutic decisions. The patterns of gene expression we identified that differentiate long and short-term survivors highlight the utility of microarray analysis in the elucidation of the molecular etiology of complex clinical phenotypes. As array technology rapidly evolves, we hope to extend the findings of the current study with newer generation arrays, larger numbers of cancers,
and predictive modeling strategies to identify distinct genetic signatures that can be used to predict clinical outcome from advanced stage epithelial ovarian cancer.
REFERENCES


Chapter 3

Patterns of Gene Expression That Characterize Long-Term Survival in Advanced Stage Serous Ovarian Cancers

Published as:

Note: Contributed to work as a co-investigator only
Abstract

Objective: A better understanding of the underlying biology of advanced ovarian cancer is critical for the development of early detection strategies and new therapeutics. The objectives of this study were to define gene expression patterns associated with favorable survival in advanced stage serous ovarian cancers.

Methods: RNA from 65 serous ovarian cancers was analyzed using Affymetrix U133A microarrays. This included 54 stage III/IV cases (30 short-term survivors who lived < 3 years and 24 long-term survivors who lived > 7 years) and 11 stage I/II cases. Genes were screened on the basis of their level of and variability in expression, leaving 7,821 for use in developing a predictive model for survival. We developed a composite predictive model that combines Bayesian classification tree and multivariate discriminant models. We use leave-one-out cross validation to select and evaluate models.

Results: Patterns of genes have been identified that distinguish short-term and long-term ovarian cancer survivors with 91% accuracy. The expression model developed for advanced stage disease classified all 11 early stage ovarian cancers as long-term survivors. This result suggests that gene expression patterns underlie differences in outcome; and an examination of the genes that provide this discrimination reveals that many have been implicated in cellular processes that define the malignant phenotype.

Conclusions: Differences in clinical outcome of advanced ovarian cancers are reflected by distinct patterns of gene expression. This biological distinction is further emphasized by the finding that early stage cancers share expression patterns with the advanced stage long-term survivors, suggesting a shared favorable biology.
Introduction

Epithelial ovarian cancers recapitulate all the histologic patterns observed in the mullerian derived structures of the genital tract - including the serous pattern of the fallopian tube, the endometrioid pattern of the endometrium and the clear cell and mucinous patterns of the endocervix. Despite the histological heterogeneity, most deaths are attributable to the serous type, which comprises about two-thirds of cases and has a propensity to present at an advanced stage. Only about 5% of women with invasive serous ovarian cancers have early stage (I/II) disease. Patients with early stage disease who undergo resection of the primary tumor and are not found to have metastatic disease in the peritoneal cavity or retroperitoneal lymph nodes have an 80-90% probability of survival (1). Advanced (stage III/IV) cases generally are disseminated throughout the peritoneal cavity at diagnosis and only a small fraction are cured (2,3) Primary surgical treatment includes removal of the ovarian tumor and resection of metastases. The goal is to leave minimal residual disease prior to administration of chemotherapy with platin and taxane drugs, as optimal surgical debulking is associated with longer survival (4).

Median survival of women with stage III/IV serous ovarian cancer is about 3 - 4 years, and more than 90% eventually relapse following initial therapy (2,3). Although the amount of residual cancer present after initial surgery is a strong prognostic factor (4), some optimally debulked ovarian cancers exhibit de novo resistance to chemotherapy and survival may be less than one year. At the other extreme, some suboptimally debulked cancers are exquisitely sensitive to chemotherapy and never relapse. This range of outcomes is thought to be attributable, at least in part, to variations in underlying biological characteristics of ovarian cancers. Host factors such as nutritional status and immune function also likely play a role in determining survival.

Prior studies that have sought to elucidate the molecular determinants of outcome in serous ovarian cancers generally have focused on single genes. Alterations in several oncogenes and tumor suppressor genes have been described in serous ovarian cancers (5). Mutation of the p53 tumor suppressor gene is the most common genetic alteration in ovarian cancers, occurring in at least 70% of advanced stage cases (6).
Among patients with advanced stage disease, however, neither p53 mutational status (6) nor other single gene alterations have proven strong predictors of outcome.

The development of microarray technology now permits analysis of expression levels of thousands of genes in a cancer. Our group and others have shown that expression microarrays can be used to identify unique characteristics of tumors that can predict clinical phenotypes and survival in breast cancer (7-9). In the present study, we used Affymetrix U133A microarrays containing over 22,000 probe sets to define expression patterns that distinguish between short-term (< 3 years) and long-term (> 7 years) survival in advanced serous ovarian cancers. Bayesian linear discriminant and classification tree analysis yielded predictive models with greater than 90% accuracy when subjected leave-one-out cross validation. In addition, these predictive models classified all early stage ovarian cancers as having gene expression patterns similar to that of the cancers from long-term survivors with advanced stage disease suggesting a shared favorable biology.

**Materials and Methods**

**Patients and tissue acquisition:** Serous ovarian cancer tissue was snap frozen at initial surgery prior to any chemotherapy in 65 women treated at Duke University Medical Center between 1988 and 2001 under the auspices of a protocol approved by the Duke IRB. This study included 54 advanced stage (III/IV) cases and 11 early stage (I/II) cases that were surgically staged. The stage III/IV cases all received platinum-based combination chemotherapy. Survival was less than 3 years in 30 cases and greater than 7 years in 24 cases. Median survival of the short-term survivors was 17.5 months. Median survival of the long-term survivors was 107.5 months. Twelve long-term survivors remain alive, seven of whom never developed recurrent disease. None of the patients in this study died of causes other than ovarian cancer. None of the 11 patients with stage I/II disease have died although one with stage IIC disease is in remission after second-line therapy of recurrent disease.
**Microarray analysis:** Frozen tissue samples were embedded in OCT medium and sections were cut and mounted on slides. The slides were stained with hematoxylin and eosin to assure that at least 60% of the cellular content was comprised of cancer cells. Approximately 30 mg of tissue was added to a chilled BioPulverizer H tube (Bio101 Systems, Carlsbad, CA). Lysis buffer from the Qiagen Rneasy Mini kit was added and the tissue homogenized for 20 seconds in a Mini-Beadbeater (Biospec Products, Bartlesville, OK). Tubes were spun briefly to pellet the garnet mixture and reduce foam. The lysate was transferred to a new 1.5 ml tube using a syringe and 21 gauge needle, followed by passage through the needle 10 times to shear genomic DNA. Total RNA was extracted using the Qiagen Rneasy Mini kit. Two extractions were performed for each tumor and the total RNA pooled at the end of the Rneasy protocol, followed by a precipitation step to reduce volume. Quality of the RNA was checked by an Agilent 2100 Bioanalyzer. Labeled probes for Affymetrix DNA microarray analysis were prepared according to the manufacturer’s instructions. Biotin-labeled cRNA, produced by *in vitro* transcription, was fragmented and hybridized to the Affymetrix U133A GeneChip arrays (22,283 probe sets, [http://www.affymetrix.com/support/technical/byproduct.affx?product=hgu133](http://www.affymetrix.com/support/technical/byproduct.affx?product=hgu133), Santa Clara, CA) at 45°C for 16 hr and then washed and stained using the GeneChip Fluidics. The arrays were scanned to a target intensity of 500 by a GeneArray Scanner and patterns of hybridization detected as light emitted from the fluorescent reporter groups incorporated into the target and hybridized to oligonucleotide probes.

**Statistical methodologies**
Expression was calculated using the robust multi-array average (RMA) algorithm (10) implemented in the Bioconductor ([http://www.bioconductor.org](http://www.bioconductor.org)) extensions to the R statistical programming environment (11). RMA generates log-2 scaled measures of expression using a linear model robustly fit to background-corrected and quantile-normalized (12) probe-level expression data and has been shown to have a better ability to detect differential expression in spike-in experiments (10). The 22,283 probe sets were screened to remove 68 control genes, those with a small variance and those expressed at low levels. Genes whose correlation with outcome was greater than 0.4 in
absolute value were used to develop predictive models for survival via binary tree (13,14) and linear discriminant (15,16) analyses. Similar approaches have been taken in other studies of various cancer types (17,18).

**Tree analysis:** Bayesian classification tree analysis recursively partitions the sample into subgroups according to levels of covariates and associates with each a predictive probability for the outcome variable (8,19). With categorical or continuous covariates, this is based on an underlying non-parametric model for the conditional distribution of predictor variables given outcome, consistent with the retrospective case-control design. Sequences of Bayes-factor-based tests of association were used to rank and select predictors that define significant splits of nodes (20), resulting in a conservative approach to forward generation of trees, producing trees that are effectively self-pruning. Multiple trees were systematically generated based on different choices among a large number of potential predictors and thresholds with the aim of finding classes of trees with high marginal likelihood. Predictions were based on model averaging, i.e., weighting predictions of trees by their implied posterior probabilities. Posterior predictive distributions were calculated at each node of each tree when evaluating or interpreting trees or averaging out-of-sample predictions across trees. In these analyses a single tree outperformed all remaining trees in each leave-one-out cross-validation; so we report one such tree, which appeared in most cross-validation runs as the tree with highest likelihood and strong out-of-sample predictive performance.

**Linear discriminant analysis:** Using leaps and bounds, the best regressions as measured by Bayes Information Criteria (BIC), constrained to involve no more than \( P \) predictors were identified. Best compositional regressions that include the outcome variable define the implied discriminant function. The following steps were implemented in context of leave-one-out cross-validation to choose a multivariate discriminant function: (i) set \( p=1 \), (ii) order genes in ascending order according to association with outcome, (iii) identify the best compositional model under this ordering, (iv) identify the ordering of genes minimizing bandwidth of the implied conditional covariance matrix of gene expression given outcome, (v) report the best compositional model under this ordering, (vi) repeat (iii) –(v) fifty times, (vii) order genes in ascending order according to
the number of models out of the 50 generated at step(vi) in which the gene is associated with outcome and report the best compositional model under this ordering, and (viii) report the five models out of those identified at steps (vi) and (vii) with highest marginal likelihood. Repeat for p=2 through p=P. The 'true value' of P depends on the ordering of genes chosen; step (iii) is taken to minimize this, the maximum number of independent variables in the compositional regressions. For each value of p, out-of-sample predictions were generated using a model average of the top 5 models identified at step (viii). The model average associated with the value of p achieving best out-of-sample predictive performance was reported.

**Combined tree and linear discriminant (hybrid) modeling:** Tree and linear discriminant function predictions were averaged to form composite predictions. Equal weighting achieves minimum mean squared prediction error among weighted averages assuming the models are unbiased and generate equally variable predictions; this result holds whether the models are correlated or not. This approach has the advantage that it is simple and agnostic and the optimality assumptions are plausible (but by no means necessary for the approach to perform well). 90% predictive interval estimates were calculated by averaging upper and lower limits of the 90% predictive intervals calculated under the two models. This procedure is conservative in so far as coverage probabilities exceed the nominal unless the predictions are perfectly positively associated (in which case they agree).

**Cross validation analyses:** Leave-one-out cross validation was used to select and evaluate models. This allows for choice among tree and linear discriminant models of varying complexities on the basis of ability to predict samples not used to fit the models. Here, both variable/feature selection -- the choice of which subset of the large list of potential expression and clinical measures to include in the model -- and estimation -- the determination of how to include those chosen -- were cross validated. For each model, out-of-sample predictive probabilities were thresholded at 1/2 to determine predictive classifications and out-of-sample classification accuracy was defined as the fraction of predicted and true classifications in agreement. Tree and linear discriminant models with best out-of-sample classification accuracy were chosen for inclusion in the predictive model. The two-sided Wilcoxon rank sum test for two sample comparisons
was used in descriptive analyses. Fold change was calculated as $2^d$, where $d$ is mean (log2 scaled) expression in the comparison group minus that in the baseline group; a fold change of 1 corresponds to no difference, 2 to a doubling, etc. Unless otherwise specified, short-term survivors serve as baseline.
<table>
<thead>
<tr>
<th>Stage</th>
<th>Median CA 125 Level (Range)</th>
<th>Long-term Survivors (n=24)</th>
<th>Short-term Survivors (n=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1.422 (60 - 16.510)</td>
<td>52 years</td>
<td>62 years</td>
</tr>
<tr>
<td>II</td>
<td>2 (7%)</td>
<td>1 (4%)</td>
<td>1 (4%)</td>
</tr>
<tr>
<td>III</td>
<td>7 (23%)</td>
<td>2 (8%)</td>
<td>2 (8%)</td>
</tr>
<tr>
<td>IV</td>
<td>21 (70%)</td>
<td>21 (70%)</td>
<td>21 (70%)</td>
</tr>
<tr>
<td>Other</td>
<td>Black</td>
<td>White</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Demographic and Clinical Characteristics of Women With Serious Ovarian Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage I</td>
<td>-</td>
</tr>
<tr>
<td>Stage II</td>
<td>-</td>
</tr>
<tr>
<td>Stage III</td>
<td>-</td>
</tr>
<tr>
<td>Stage IV</td>
<td>-</td>
</tr>
</tbody>
</table>
Results

Demographic and clinical features of the 54 women with advanced stage (III/IV) ovarian cancer and 11 with early stage (I/II) ovarian cancer are shown in Table I. Among the advanced stage cases, the short-term survivors had less favorable clinical features including somewhat higher grade and substage and a lower frequency of optimal debulking (largest residual tumors after primary surgery less than 1 cm in diameter). There was no difference in age at diagnosis, but pre-operative CA125 levels were somewhat higher among short-term survivors (p=0.06).

Gene expression data was generated from each cancer using Affymetrix GeneChip arrays. The microarray data was initially screened to discard those genes that did not vary or were expressed at low levels, leaving 7,821 for predictive model development. A table of these genes, ordered by association with survival category, their fold changes between long-term and short-term survivors and full descriptions appears in Supplementary Material. In addition, clinical features including age at diagnosis, pretreatment CA125 level and a binary indicator of the maximal size of the largest residual tumor nodule after primary surgery (optimal versus suboptimal debulking) were available for predictive modeling.

Utilization of classification and regression tree models to distinguish short-term and long-term survival

Our previous work has described the use of classification and regression trees together with Bayesian analysis to build predictors of recurrence in breast cancer (8). The most effective of these models makes use of a combination of clinical and gene expression data. We have thus used a similar approach to building a classifier that could distinguish long-term and short-term survival in ovarian cancer.

For this purpose, genes with absolute values of correlation to survival greater than 0.4 were used, netting 478 genes from the pool of 7,821 screened genes. The expression patterns of the 478 genes are shown in Figure 1.
These genes were then used to build a classifier/predictor that could accurately distinguish short-term and long-term survivors, making use of classification and regression tree analysis methods as we have previously described (19). The best tree models involved expression of two or three genes (and no clinical measures) and achieved 90.7% out-of-sample predictive accuracy, correctly classifying 20/24 (83.3%) long-term and 29/30 (96.7%) short-term survivors. An example tree from these analyses (Figure 2A) stratifies the patient population into three groups according to expression of CSTF3 (cleavage stimulation factor subunit 3) and ABCD3 (ATP-binding
cassette, sub-family D, member 3). These predictors appeared in the tree with the highest posterior probability in 98% of cross-validation analyses.

The gene expression based tree analysis was compared to the most highly predictive tree models involving only the clinical variables, including CA125. These trees, one for each left-out sample, achieved 67% cross validation predictive accuracy (63% among short-term survivors, 71% among long-term survivors) and involved only debulking status and age at diagnosis. An example of a tree that uses only clinical data is depicted in Figure 2B. Clearly, the ability to distinguish long-term and short-term survival was much improved by the gene expression data.

In examining gene expression in the early stage (I/II) cancers, the long-term and short-term survivors were used as a training set and the early stage cases as a test set. This analysis classifies all early stage cases as long-term survivors with an associated 93.6% average predicted probability. The tree with the highest posterior probability based on the training set (Figure 2C) divides the training cases into 4 groups according to expression of PDX1 (pyruvate dehydrogenase complex), mRNA clone DKFZp762G207 (DKFZ), and translocon-associated protein alpha (TAPα).
long-term survival. The models predicted a 99% probability of long-term survival for cases having low levels of both PD-L1 and PD-1 expression. The remaining 2% of cases were predicted to have a 99% probability of survival. The models predicted a 78% probability of long-term survival for cases having low levels of PD-L1 and high levels of PD-1 expression. The remaining 22% of cases were predicted to have a 78% probability of survival. The models predicted a 40% probability of long-term survival for cases having high levels of PD-L1 and low levels of PD-1 expression. The remaining 60% of cases were predicted to have a 40% probability of survival. The models predicted a 5% probability of long-term survival for cases having high levels of PD-L1 and high levels of PD-1 expression. The remaining 95% of cases were predicted to have a 5% probability of survival. The models predicted a 0% probability of long-term survival for cases having high levels of PD-L1 and high levels of PD-1 expression. The remaining 100% of cases were predicted to have a 0% probability of survival.
| Gene name                                                                 | Symbol | P-Value | Fold difference | Long/short (95%) |
|---------------------------------------------------------------------------|--------|---------|-----------------|----------------|----------------|
| TNF-alpha                                                                 | TNF    | 0.0000  | 1.00            | 0.50           |
| IL-1 beta                                                                 | IL1B   | 0.0000  | 1.89            | 0.50           |
| IFN-gamma                                                                | IFNG   | 0.0000  | 2.40            | 0.60           |
| CD45                                                                     | CD45   | 0.0000  | 1.20            | 0.40           |
| CD85J                                                                    | CD85J  | 0.0000  | 1.50            | 0.60           |

Table 2: Genes used in tree and linear discriminant analysis.
<table>
<thead>
<tr>
<th>LDM</th>
<th>Linear Discriminant Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y-V1 Transcription Factor</td>
<td>YY1</td>
</tr>
<tr>
<td>ATP Binding Protein Associated with Cell Differentiation</td>
<td>ATP</td>
</tr>
<tr>
<td>KIAA0624 Protein</td>
<td>KIAA0624</td>
</tr>
<tr>
<td>Embryonic Ectoderm Development Protein</td>
<td>EED</td>
</tr>
<tr>
<td>General Transcription Factor II, Polypeptide 1</td>
<td>GTF2H1</td>
</tr>
<tr>
<td>Net-6</td>
<td>NET-6</td>
</tr>
<tr>
<td>Lactase Net-G Protein</td>
<td>LACT1</td>
</tr>
<tr>
<td>Acetyl-Protein Kinase, Lyso phosphatase II</td>
<td>ACYKPL</td>
</tr>
<tr>
<td>Stress-Associated Endoplasmic Reticulum Protein 1</td>
<td>SERP1</td>
</tr>
</tbody>
</table>
Expression summaries for the genes utilized in the clinical and gene based trees involved in the cross-validation predictions of the long-term and short-term survivors are presented in Table 2.

Linear discriminant models: Although the tree based models performed well in discriminating long-term and short-term survivors, the models were not completely effective. As such, we have examined other methods for analysis. To complement the tree analysis, the strength of which is to identify abruptly nonlinear associations between predictors and outcome, a variant of linear discriminant analysis was employed. The approach taken differs from multivariate linear discriminant analysis in its use of predictive instead of fitted densities for classification and by incorporating a variable selection step with explicit complexity constraint. It identifies multivariate discriminant functions given a large pool of potential candidate predictors by factoring retrospective multivariate normal distributions of expression data given binary survival outcome compositionally (21) and carrying out model selection on each of the compositional regressions using the 'leaps and bounds' algorithm (22). The best multivariate linear discriminant functions correctly classified 19/24 (79.2%) long-term survivors and 27/30 (90%) short-term survivors, achieving 85.2% out-of-sample classification accuracy. These classifiers also predicted all 11 early stage cases to be long-term survivors. The discriminant functions involved 186 genes across leave-one-out analyses and reflected models allowing for simple conditional dependencies among gene expression measures given survival category (the case P=2). The top five genes in order of representation were clone T cell differentiation protein (MAL); APMCF1 protein (APMCF1); diphosphoinositol polyphosphate phosphohydrolase type 2 (NUDT4); plakophilin 4 (PKP4); and signal sequence receptor alpha (SSR1). Estimates of fold difference in expression of these genes and those appearing in 15 or more leave-one-out models between short-term and long term survivors with 95% confidence intervals are presented in Table 2. The MAL gene, which was the most highly upregulated in short-term survivors (3.0 fold), was included in 54 of 54 linear discriminant models used in leave-one-out analyses. None of the other genes was used in all 54 models.

Combined models: The fact that the tree-based model and the linear discriminant model identified different genes as predictors of long-term and short-term survival
suggested the possibility that they addressed different aspects of the underlying biology and thus that a combination of the two models would be more effective than either alone. The best tree and discriminant function predictions were combined through a simple predictive model average. This composite model correctly classified 20/24 (83.3%) long-term survivors and 29/30 (96.7%) short-term survivors and reached an overall out-of-sample predictive accuracy of 90.7%. Further, it classified all 11 early-stage cases as long-term survivors.

Figures 3(a) through 3(c) are plots of leave-one-out out-of-sample predictions generated by the best tree, discriminant and composite (hybrid) models. Predictive classifications are obtained by thresholding predictive probabilities at 0.5.

**Figure 3. Cross validation predictions of survival.** These figures plot individual leave-one-out out-of-sample predictions (on the x-axis) for the 54 advanced stage cancers by actual survival group for the linear discriminant (A), tree (B) and hybrid models (C). The short-term survivors are plotted in red and the long-term survivors are plotted in blue. In addition, out-of-sample predictions for the 11 early stage tumors are plotted in green (note that these samples were not used in 'model training'). Points are jittered in the direction of the y-axis to achieve their separation. A vertical line is plotted at 1/2 indicating the classification threshold used for determining out-of-sample predictive accuracy (cases with probability of long-term survival greater than 1/2 were declared long-term survivors; those less than 1/2 as short-term survivors. The predictive accuracy of the hybrid model using a threshold of 1/2 is 90.7%.
Tree Model

A.

Discriminant Model

B.

Hybrid Model

C.
The hybrid model combines two structurally different models, each with 90.7% leave-one-out out-of-sample classification accuracy, to form a composite model with 92.6% classification accuracy, suggesting that the tree and discriminant models provide complementary predictions.

In addition, the models trained on the advanced stage data were used to predict the survival outcomes of the 11 early stage cases. Both the linear discriminant model and tree models individually predicted all of these samples to be longer term survivors as did the hybrid model (Figure 3). These predictions suggest that aspects of the tumor biology of advanced stage cancers associated with long-term survival are shared by early stage ovarian cancers. Figure 4 presents a plot of out-of-sample point predictions using the combined model. This presents the estimated probability of a sample representing a long-term survivor together with 90% confidence intervals for the predictions.

![Prediction Point Estimates and 90% Interval Estimates](image)

**Figure 4. Out-of-sample point predictions with 90% error bars for the hybrid model.** Cases are ordered along the x-axis according to the rank of their point prediction, ordered smallest to largest. Short-term survivors are red, long-term survivors blue and early stage cases green.
Discussion

Approximately 10% of epithelial ovarian cancers are attributable to inherited mutations in high penetrance susceptibility genes (BRCA1 and BRCA2), while the majority of cases are sporadic. All ovarian cancers likely arise due to alterations that disrupt molecular pathways involved in regulation of proliferation, apoptosis and DNA repair and a number of genes have been identified that are involved in the development of some ovarian cancers (eg. p53, HER-2/neu, K-ras). These studies emphasize that substantial molecular heterogeneity exists between cancers (5). However, analysis of individual candidate genes based on their involvement in pathways related to carcinogenesis is inefficient and has yielded relatively few relevant genes. In addition, none of the genetic alterations described thus far has augmented the ability of conventional clinical staging systems based on extent of disease to predict survival of women with ovarian cancer.

Microarrays that assess global patterns of gene expression have proven useful in defining and predicting clinical phenotypes in a variety of cancer types (23). This includes studies of breast cancer in which several studies have described gene expression patterns that can serve as prognostic tools to define risk of recurrence (9,24) as well as lymph node involvement (8). The work of our own group has demonstrated the value in utilizing methods of analysis that sample multiple forms of data, both clinical and multiple gene expression patterns, so as to achieve a more precise discrimination and prediction of outcome for individual patients (8,19, 25). This same logic, of utilizing multiple forms of data, as well as methods of analysis, has been applied in the present study to more accurately achieve a classification and prediction of ovarian cancer survival.

Several groups have applied expression array technology to the analysis of ovarian cancers. Many of these studies have compared gene expression between normal ovarian epithelial cells and ovarian cancers. Numerous genes have been identified that appear to be up or down-regulated in the process of malignant transformation (26,27-29) In addition, microarrays have demonstrated patterns of gene expression that distinguish between histologic types (30) and stages (31). None of these prior studies has focused on using arrays to predict outcome.
In the present study we used expression arrays to develop models that are predictive of survival. The groups of patients utilized represent the extremes with respect to outcome - namely those who survived either less than 3 years or greater than 7 years. The availability of frozen tumor samples from a significant number of long-term survivors was a strength of this study as was the use of novel and complementary statistical approaches. The observation that no gene was more than 3 fold differentially expressed suggests that patterns will be more important than individual genes in predicting outcome. The multiple model approach generated predictive results with greater than 90% accuracy in this study in which the number of predictors is considerably larger than the number of samples. Both the tree models and the linear discriminant models demonstrate strong predictive ability in both out-of-sample and one-at-a-time cross validation contexts. By combining the results of the two approaches we exploit the relative strengths of the non-linear tree structures and the linear discriminant models to achieve better predictive accuracy.

Interestingly, the gene expression patterns of all early stage cases were characterized as similar to those of the long-term survivors. This provides compelling evidence that the favorable clinical outcome of both long-term survivors with advanced stage disease and early stage cases is attributable to a shared underlying pattern of causative molecular alterations. This has implications for screening as an approach to decreasing ovarian cancer mortality. Since cancers currently detected at an early stage all had patterns of gene expression similar to long-term survivors, early detection may only be feasible in this relatively small subset of the disease. Conversely, none of the early stage ovarian cancers exhibited patterns of gene expression similar to the presumably more aggressive cancers from short-term survivors, suggesting that these cancers may not progress or constitute the precursors of the more virulent forms.

Many of the genes that were critical components of the discriminatory patterns are known to affect the virulence of the malignant phenotype. The MAL gene (T-lymphocyte maturation associated protein) was the most differentially expressed (an average of 3 fold higher in cancers from short-term versus long-term survivors) and also was upregulated in short-term survivors relative to early stage disease (29 fold). MAL was included in 54 of 54 linear discriminant models used in leave-one-out analyses, while none of the other genes was used in all 54 models. Expression of MAL has been
demonstrated previously in ovarian cancers, most notably clear cell and serous cancers (30). The MAL protein has several hydrophobic domains and has been demonstrated to be a component of the protein machinery for apical transport in epithelial polarized cells (32,33). MAL also is a component of membrane rafts, which are microdomains that play a central role in signal transduction acting as a scaffold in which molecules of signal transduction pathways can interact (32). Further, MAL has been identified as a gene involved in resistance to cancer therapy. MAL was the most differentially expressed gene in microarray experiments that compared interferon-alpha sensitive and resistant cutaneous T-cell lymphoma cell lines (34). The emergence of interferon-alpha resistance in this in vitro model was accompanied by 5.5 fold upregulation of MAL. In addition, high MAL expression was found to correlate with poor response to therapy in a cohort of patients with cutaneous T-cell lymphoma (34).

Heat shock protein 27 (HSP27) is another gene implicated in resistance to therapy that was more highly expressed in cancers from short-term survivors relative to long-term survivors (1.37 fold) and early stage cases (1.24 fold). HSP27 is a member of the heat shock protein family of proteins that is normally upregulated in response to thermal injury and other stresses. Expression of HSP27 has been shown to augment the ability of cancer cells treated with cytotoxic chemotherapy to evade apoptosis and this enhances cell survival (35,36). This apoptotic resistance is thought to be due to inhibition of activity of the apoptosome and caspase activation complex as well as regulation of proteasome-mediated degradation of apoptosis-regulatory proteins. Several groups have found that high HSP27 expression increases chemoresistance of ovarian cancer cell lines in vitro (37,38) and this can be reversed by treatment with antisense HSP27 (38). In addition, high expression of HSP27 correlates with poor survival in several types of cancers including glioma (39), breast cancer (40) and ovarian cancer (38).

Lysophospholipase II (LYPLA2), elevated in cancers from patients with short-term survival may have a biochemical link to the disease process. This is one of a family of enzymes that can convert lysophospholipids to lysophosphatidic acid (LPA), a molecule commonly elevated in ovarian cancer ascitic fluid (41). LPA itself has been shown to have a number of biological effects including increasing proliferation, survival, invasiveness, and resistance to cisplatin (41,42). Ovarian cancer cells have been
shown to produce LPA and the increased levels of the LYPLA2 enzyme may be a significant component of this biosynthesis (41).

Finally, it has been suggested that androgens play a role in both the pathogenesis and growth of ovarian cancers (43). The SRD5A1 gene, which was more highly expressed in cancers of long-term survivors, is involved in converting testosterone to its active form dihydrotestosterone. Higher levels of SRD5A1 in ovarian cancers of long-term survivors may reflect maintenance of hormonal responsiveness in ovarian cancers with an inherently less aggressive phenotype, as is the case in breast and prostate cancers.

The value of gene expression based predictors of prognosis in advanced ovarian cancer will not be fully realized until additional therapies are available for those destined to have poor survival following conventional chemotherapy. In this regard, the expression profiles may not only predict the likelihood of long-term survival following platin chemotherapy, but also yield clues to individual genes involved in tumor development, progression and response to therapy. It is likely that some of the most differentially expressed genes, such as those discussed above, will represent appealing therapeutic targets.
REFERENCES


Chapter 4

High expression of tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) is associated with favorable ovarian cancer survival

Published as:
JM Lancaster, R Sayer, C Blanchette, B Calingaert, R Whitaker, J Schildkraut, J Marks, A Berchuck.
High expression of tumor necrosis factor-related apoptosis-inducing ligand is associated with favorable ovarian cancer survival. Vol 9; 762-766,
Clinical Cancer Research, 2003
Abstract

**Purpose:** The molecular determinants of survival in ovarian cancer are poorly understood. Using expression microarrays we recently found that high expression of the tumor necrosis factor-like apoptosis inducing ligand (TRAIL) gene is associated with prolonged survival in advanced ovarian cancer. TRAIL has also been shown to synergise with chemotherapeutic agents to induce apoptosis in ovarian cancer cell lines. We therefore sought to confirm the association between TRAIL expression and survival in a larger group of women with ovarian cancer.

**Experimental Design:** TRAIL expression was measured using quantitative real-time polymerase chain reaction in 120 epithelial ovarian cancers (11 stage I/II, 109 stage III/IV) and 8 normal ovarian surface epithelial samples.

**Results:** Ovarian cancers demonstrated 10-fold higher mean TRAIL expression than normal ovarian epithelial samples (p<0.001). Among ovarian cancers, high TRAIL expression was associated with prolonged survival and was 2.2-fold higher in cancers from patients who lived more than 5 years compared to patients who died within one year (p=0.03).

**Conclusions:** TRAIL expression is higher in ovarian cancers relative to normal ovarian epithelium. High TRAIL expression is associated with favorable ovarian cancer survival, which may be attributable to increased chemosensitivity of cancers that express the most TRAIL. The use of TRAIL to enhance sensitivity of ovarian cancers to therapy represents an appealing molecular therapeutic strategy worthy of further investigation.
INTRODUCTION

Epithelial ovarian cancer is among the leading causes of female cancer deaths in the United States and Western Europe. Similar to other types of human cancers, ovarian cancers are thought to arise due to accumulation of mutations in genes that regulate cellular proliferation and apoptosis. In this regard alterations in several genes, including the p53 tumor suppressor gene and HER-2/neu oncogene, have been identified in some ovarian cancers (1). A pattern of molecular alterations characteristic of all epithelial ovarian cancers has not been identified, however. The molecular pathogenesis of these cancers is heterogeneous and this is reflected in the variability of clinical characteristics such as histologic type, differentiation, potential for invasion and metastasis, response to therapy and outcome.

Most ovarian cancers are found to have metastasized extensively in the peritoneal cavity at diagnosis because effective screening and early detection techniques do not exist. About 70% of women with advanced ovarian cancer have a complete clinical response following surgical cytoreduction and platinum-based chemotherapy, but median survival is only about 3 years (2). There is considerable variation in the extent of response to therapy between patients, however. Some women succumb relatively quickly while others who are treated in an identical manner may live more than ten years. As noted above, the wide variation in outcome likely is determined by the heterogeneity of the molecular alterations that underlie the development of ovarian cancers.

Prior attempts to identify molecular prognostic markers have focused on single genes, but this approach is time consuming and relatively inefficient. The recent development of microarray technology has provided the opportunity to simultaneously evaluate the expression of thousands of genes in ovarian cancers. Several groups have used arrays to demonstrate differences in gene expression between normal ovarian epithelium and ovarian cancers, and also between various histologic types of ovarian cancer (3-7). We recently used microarrays to determine whether distinct patterns of gene expression could be identified that predict long or short-term survival in women with advanced stage serous ovarian cancers (8). Several individual genes involved in growth and apoptosis were shown to be differentially expressed. Expression
of the tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) gene was 7.4 fold higher in ovarian cancers than normal ovarian epithelium and 1.5 fold higher in cancers of long-term survivors compared to short-term survivors.

The TRAIL gene, which is located on chromosome 3q26, is a member of the TNF cytokine family. It was initially identified by homology to the c-terminal extracellular domain of other TNF family members, such as FasL, TNF-alpha, and lymphotoxin-alpha (9). Subsequently, TRAIL has been shown to induce apoptosis in a variety of human and mouse transformed or malignant cells, while normal cells are not affected (10). Selective induction of apoptosis in cancer cells suggests a potential therapeutic use for TRAIL (11,12). In ovarian cancers, Cuello et al found that a majority of twelve chemoresistant cell lines were resistant to TRAIL alone, but some were sensitive to the combination of TRAIL and cytotoxic drugs such as cis-platinum (13). In view of these in vitro data and our microarray data suggesting that high TRAIL expression correlates with favorable survival in advanced ovarian cancers, we sought to confirm the association between TRAIL expression and outcome in a larger cohort of patients.

METHODS

One hundred and twenty epithelial ovarian cancers and 8 normal ovarian surface epithelial samples were obtained with IRB approved informed consent from patients treated by the Division of Gynecologic Oncology at Duke University Medical Center. Eleven of the cancers were early stage (I/III), and 109 advanced stage (III/IV). Histopathologic subtypes included 95 serous, 6 endometrioid, 5 mucinous, 4 clear cell, and 10 mixed type. All patients with advanced stage disease were treated with primary surgical cytoreduction (52 optimal, 57 suboptimal) followed by platinum-based chemotherapy.

Total RNA was extracted using the RNAeasy RNA extraction kit (Qiagen Inc., Valencia, CA), and an aliquot of 1ug subject to a reverse transcriptase reaction with the Roche First Strand cDNA kit (Roche, Basel, Switzerland). Complementary DNA (cDNA) was quantified by fluorometry, and 5ng subject to 40 cycles of QRT-PCR in the Roche Light Cycler (Roche, Basel, Switzerland) according to manufacturers instructions in the presence of intron-spanning cDNA primers to the house-keeping gene, beta-microglobulin, to confirm adequate cDNA normalization. In the same way, QRT-PCR
was used to determine relative expression of the TRAIL gene in 120 epithelial ovarian cancers using the intron-spanning cDNA primers 5’- aga cct gcg tgc tga tgc tg-3’ and 5’- tta ttt tgc gcg cca gag cc-3’, 10 units of AdvanTaq Plus DNA Polymerase with PCR reaction buffer (Clontech, Palo Alto, CA), dNTP, and Syber Green. Serial dilutions of the sample demonstrating highest TRAIL expression were used to create a concentration curve and relative expression levels calculated for each sample from this curve. All reactions were performed in triplicate and mean values used. Statistical analysis was performed using the Students t-test and Kaplan-Meier log-rank survival and Spearman correlation coefficient analysis.

RESULTS

Quantitative real-time polymerase chain reaction (QRT-PCR) was used to determine relative expression of the TRAIL gene in 120 epithelial ovarian cancers and 8 normal ovarian surface epithelial samples (Table 1, Figure 1). For each sample, TRAIL expression was calculated as a relative value from a serial dilution curve of the cancer demonstrating the highest TRAIL expression level. Mean relative TRAIL expression in 120 ovarian cancers was approximately 10-fold higher than in 8 normal ovarian surface epithelial samples (1.28 vs. 0.13, p<0.001). Moderate and poorly differentiated (grade II/III) cancers expressed 3.1-fold greater TRAIL levels than well differentiated (grade I) cancers (p= 0.0001), although no difference was seen between moderate (grade II) and poorly differentiated (grade III) cancers (1.45 vs. 1.23, p=0.61).
**Figure 1.** Epithelial ovarian cancers and normal ovarian surface epithelial samples were subject to quantitative real-time polymerase chain reaction to determine relative expression of the TRAIL gene. Cancer and normal cDNA (left panel) was subject to PCR in real time (top right panel). Serial dilutions of the most highly expressed cancer were used to create a concentration curve (bottom right panel) and relative expression levels were calculated for each case.
**Table 1.** Relative expression of the TRAIL gene in epithelial ovarian cancers and normal ovarian surface epithelial (NOSE) samples determined by quantitative real-time polymerase chain reaction.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean TRAIL Expression</th>
<th>Difference (fold)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOSE</td>
<td>8</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ovarian cancers</td>
<td>120</td>
<td>1.28</td>
<td>10.1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Early stage</td>
<td>11</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced stage</td>
<td>109</td>
<td>1.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All ovarian cancers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survival &lt; 1 year</td>
<td>22</td>
<td>0.66</td>
<td>2.3</td>
<td>0.03</td>
</tr>
<tr>
<td>Survival &gt;5 years</td>
<td>36</td>
<td>1.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced serous cancers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survival &lt; 1 year</td>
<td>14</td>
<td>0.68</td>
<td>1.8</td>
<td>0.08</td>
</tr>
<tr>
<td>Survival &gt;5 years</td>
<td>29</td>
<td>1.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade I cancers</td>
<td>5</td>
<td>0.44</td>
<td>3.1</td>
<td>0.0001</td>
</tr>
<tr>
<td>Grade II/III cancers</td>
<td>108</td>
<td>1.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serous cancers</td>
<td>95</td>
<td>1.09</td>
<td>1.9</td>
<td>0.03</td>
</tr>
<tr>
<td>Non-serous cancers</td>
<td>25</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I/II</td>
<td>11</td>
<td>0.8</td>
<td>1.7</td>
<td>0.26</td>
</tr>
<tr>
<td>Stage III/IV</td>
<td>109</td>
<td>1.32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Although overall, serous cancers expressed higher levels of TRAIL relative to other histopathologic subtypes (1.09 vs. 0.57, p=0.03), no statistical difference was identified between TRAIL expression in serous compared to other individual histopathologic subtypes (endometrioid, mucinous, or clear cell). Expression in 11 early stage (I/II) and 109 advanced stage (III/IV) cancers was 6.1-fold (p=0.12) and 10.1-fold (p<0.0001) higher, respectively, than normal ovarian surface epithelium. In advanced stage (III/IV) cancers, TRAIL expression was 1.7-fold greater than early stage (I/II) cancers (p=0.26). There was no relationship between TRAIL expression and extent of cytoreductive surgery (optimal versus suboptimal).

High TRAIL expression measured by QRT-PCR was associated with prolonged survival. For all stages, relative TRAIL expression was 2.3-fold higher (1.52 vs. 0.66) in cancers from women who lived more than 5 years (n=36) than in cancers from those who died in less than one year (n=22) (Spearman correlation coefficient of 0.26, p=0.03). Among advanced stage (III/IV) cases, relative TRAIL expression was 2.2-fold higher in cancers of women who lived more than 5 years (n=32) than in cancers from

![Figure 2. Kaplan-Meier survival curve for 106 advanced stage (III/IV) ovarian cancers, grouped according to high (>0.5) or low (<0.5) relative expression of TRAIL (p=0.14, log-rank test).](image)
those who lived less than one year (n=20) (1.6 vs. 0.9, p=0.18). In figure 2, Kaplan-Meier survival analysis of the 109 patients with advanced stage ovarian cancer demonstrates the association between high TRAIL expression and favorable survival (p=0.14, log-rank test). In the subset of patients with advanced serous cancers, 29 long-term survivors (> 5 years) demonstrated 1.8-fold higher relative TRAIL expression relative to 14 short-term survivors (< 1 year) (1.52 vs. 0.68, p=0.08).

DISCUSSION

We found that ovarian cancer tissue samples express significantly more TRAIL mRNA than cultured normal human ovarian epithelial cells. Lower TRAIL expression in the normal ovarian epithelial cells examined in this study could be attributable to their increased proliferation in monolayer culture, whereas these cells normally are relatively quiescent in vivo. However, it also has been shown previously that TRAIL expression is higher in malignant breast cells compared to normal breast tissue (13). Immunostaining for TRAIL was seen in 52% of human breast cancers, but not in normal breast. In contrast, TRAIL mRNA was identified using QRT-PCR in essentially all normal and malignant breast tissues, suggesting that while TRAIL mRNA is expressed in all cases, translation to protein occurs predominantly in cancers. Although we identified TRAIL mRNA in both normal and malignant ovarian epithelial samples, significantly higher levels were seen in the cancers. The possibility that translation to protein also may be greater in cancers warrants further study.

TRAIL is a member of a family of “death ligands” including TNF and fas ligand that are capable of inducing apoptosis. Under most circumstances, normal cells are relatively resistant to TRAIL induced cell death, whereas cancer cells often are more sensitive (9,14). In addition, synergy between cytotoxic agents and TRAIL has been observed, which suggests the potential use of TRAIL as an adjunct to conventional cancer chemotherapy (11,12). In this regard, recombinant forms of TRAIL have been evaluated in animal models and shown to prolong overall survival in mice inoculated with breast cancer cells (16,17). Phase I human clinical trials using TRAIL are ongoing.
In ovarian cancers, Cuello et al. demonstrated that the combination of TRAIL and chemotherapeutic agents resulted in a significant increase in apoptosis and overall growth inhibition in ovarian cancer cells (13). We had previously found using expression arrays that high TRAIL expression appeared to predict favorable survival in 18 patients with advanced serous ovarian cancers (8). In the current paper, we have demonstrated in a much larger group of women with ovarian cancer that high TRAIL expression correlates with favorable survival. The association between TRAIL expression and survival was evident when comparing long-term survivors (>5 years) with short-term survivors (<1 year). In addition, among the entire group of patients, those whose cancers had high expression of TRAIL mRNA had more favorable median survival.

In view of the favorable effect of TRAIL expression on survival, it is somewhat surprising that TRAIL expression was higher in advanced cancers relative to early stage cases. Our sample size was not large enough to permit definite conclusions regarding these observed differences. It is possible, however, that low TRAIL expression may contribute to the lack of efficacy of adjuvant chemotherapy in early-stage cancers described in early studies (18). This must be interpreted with some caution, however, given the small number of early-stage cancers analysed in our set, and more recent data demonstrating a benefit to adjuvant chemotherapy in early stage disease (19).

The biological actions of TRAIL are regulated by an intricate array of cell membrane receptors and downstream signal transduction pathways (15,20). TRAIL mediates apoptosis by binding to multiple cell-surface receptors, including Death receptor (DR) 4/TRAIL-R1, Killer/DR5/TRAIL-R2/TRICK2, TRID/DcR1/TRAIL-R3/LIT and TRUNDD/DcR2/TRAIL-R4. TRAIL binding with DR4 and KILLER/DR5 induces activation of the caspase pathway and subsequent apoptosis. In contrast, TRID, which lacks a cytoplasmic death domain and TRUNDD, which has a truncated cytoplasmic death domain, are reported to act as non-functional “decoy receptors” that compete for TRAIL and protect cells from TRAIL-mediated apoptosis. To take full therapeutic advantage of this apoptotic cascade, characterization of each step in the pathway is essential. Our findings to date suggest that TRAIL may have some utility in the treatment of ovarian cancer, and that pharmacologic manipulation of the TRAIL pathway
may improve survival for patients with this disease. The use of TRAIL to enhance chemosensitivity of ovarian cancers represents an appealing molecular therapeutic strategy worthy of further investigation.
REFERENCES


Comparative hybridization of an array of 21,500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas. Gene, 238: 375-385.


Chapter 5

High expression of Insulin-like Growth Factor Binding Protein-2 (IGFBP2) in epithelial ovarian cancers produces elevated pre-operative serum levels

Submitted for Publication as:

JM Lancaster, R Sayer, C Blanchette, B Calingaert, R Whitaker, J Schildkraut, J Marks, A Berchuck. High expression of Insulin-like Growth Factor Binding Protein-2 (IGFBP2) in epithelial ovarian cancers produces elevated pre-operative serum levels, to *International Journal of Cancer, 2004*
ABSTRACT

Purpose: The molecular etiology of epithelial ovarian cancer remains unclear. Using microarray expression analysis, we recently reported that the expression of the insulin-like growth factor binding protein-2 (IGFBP-2) gene is elevated in advanced epithelial ovarian cancers. The aim of the current study was to further delineate the role of the IGFBP-2 gene in the pathoetiology of epithelial ovarian cancer, and determine if elevated ovarian cancer IGFBP-2 gene expression is reflected in serum levels of patients with ovarian cancer.

Experimental Design: Relative IGFBP-2 expression was measured using quantitative real-time polymerase chain reaction in 113 epithelial ovarian cancers and 6 normal ovarian surface epithelial samples. Pre-operative serum IGFBP-2 levels were measured by radioimmunoassay in 84 women (42 with ovarian cancer, 26 with benign gynecologic conditions, and 10 healthy female controls).

Results: Ovarian cancers demonstrated 38-fold higher mean IGFBP-2 expression than normal ovarian epithelial samples (p<0.01). Serum IGFBP-2 levels were elevated in serum of women with early- (p=0.02) and advanced-stage (p<0.01) ovarian cancer compared to controls and patients with benign gynecologic conditions (p=0.05 and p<0.01, respectively). Elevated serum IGFBP-2 levels were associated with poor survival (correlation coefficient, -0.65).

Conclusions: Epithelial ovarian cancers express high levels of IGFBP-2 relative to normal ovarian epithelium, and this is associated with elevated serum IGFBP-2 levels compared to both normal controls and patients with benign gynecologic disease. Our findings provide further support for the hypothesis that the IGF pathway plays a significant role in epithelial ovarian cancer pathogenesis. Further, IGFBP2 may represent an additional serum biomarker with utility in detection and monitoring of epithelial ovarian cancer.
INTRODUCTION

Epithelial ovarian cancer is the fourth leading cause of cancer deaths among women in the United States, and mortality from the disease has changed little over the last several decades (1). Since patients with early stage disease are most often asymptomatic, and currently no effective screening test exist, the vast majority of women are diagnosed with advanced stage disease, and available treatment modalities have limited efficacy. The pathogenesis of ovarian cancer remains poorly understood, and elucidation of the molecular basis to disease development and progression is essential if new targets for prevention, screening and treatment are to be identified.

The development of microarray expression analysis allows the simultaneous examination of expression levels of thousands of genes. This technology has recently been applied to ovarian cancer and has demonstrated differences in genetic expression patterns between normal ovarian epithelium and ovarian cancers, and also between various histologic subtypes (2-8). We have applied microarray analysis to characterize advanced stage serous ovarian cancers (9). In our analysis, expression of the insulin-like growth factor binding protein 2 (IGFBP-2) gene was 4.2-fold higher in ovarian cancers compared to normal ovarian epithelium and 2.7-fold higher in cancers of short-term survivors compared to long-term survivors. Previously published array studies have implicated IGFBP-2 and other members of the IGF pathway in development and progression of ovarian and other human cancers (10-12).

Expression of IGFBP-2 has previously been demonstrated in some, but not all, immortalized ovarian cancer cell lines (13, 14), and it has been reported that serum levels of IGFBP-2 in 20 women with ovarian cancer were elevated 2.5 fold compared to controls (15). In view of these preliminary data and our microarray findings suggesting that IGFBP-2 expression is elevated in advanced ovarian cancers with poor survival, we sought to confirm the role of IGFBP-2 in ovarian cancer pathogenesis in a larger cohort of patients.
MATERIAL AND METHODS

Tissue Samples: One hundred and thirteen fresh frozen epithelial ovarian cancers and 6 normal ovarian surface epithelial samples were obtained with IRB-approved informed consent from patients treated by the Division of Gynecologic Oncology at Duke University Medical Center. Tumor tissue was harvested at the time of initial surgery, snap frozen in liquid nitrogen, and stored at -80°C until RNA extraction. Hematoxylin and eosin staining of representative sections of each tumor was performed to determine the percentage of tumor in each specimen. Only samples consisting of more than 70% tumor were used for these studies.

Histopathologic subtypes included 90 serous, 5 endometrioid, 3 mucinous, 5 clear cell, and 10 mixed type. Ten of the cancers were early stage (I/II), and 103 advanced stage (III/IV). All patients with advanced stage disease were treated with primary surgical cytoreduction (59 optimal, 44 suboptimal) followed by platinum-based chemotherapy.

Quantitative RT-PCR: Total RNA was extracted using the RNAeasy RNA extraction kit (Qiagen Inc., Valencia, CA), and an aliquot of 1ug was subjected to a reverse transcriptase reaction with the Roche First Strand cDNA kit (Roche, Basel, Switzerland). Complementary DNA (cDNA) was quantified by fluorometry, and 5ng was subjected to 40 cycles of QRT-PCR in the Roche Light Cycler (Roche, Basel, Switzerland) according to manufacturer’s instructions in the presence of intron-spanning cDNA primers to the house-keeping gene, beta-microglobulin, to confirm adequate cDNA normalization. In the same way, QRT-PCR was used to determine relative expression of the IGFBP-2 gene in 113 epithelial ovarian cancers using the intron-spanning cDNA primers 5'-TGGAGGAGCCAGAAGCT-3' and 5'-GGTTACACACCAGCAGCCTC-3', 10 units of AdvanTaq Plus DNA Polymerase with PCR reaction buffer (Clontech, Palo Alto, CA), dNTP, and SYBR Green. Serial dilutions of the sample demonstrating highest IGFBP-2 expression were used to create a concentration curve and relative expression levels calculated for each sample from this curve. The standard curves were included in each run.

Serum Samples: Pre-operative serum IGFBP-2 levels were measured using radioimmunoassay in 84 women, including 42 patients with ovarian cancer (11 with
early- and 31 advanced-stage disease), 6 patients with borderline tumors, 26 patients 
with benign gynecologic conditions (12 ovarian cystadenomas, 4 corpus luteum cysts, 2 
endometriomas/osis, 2 tubo-ovarian abscesses, 3 mature teratomas, 1 leiomyomata 
uteri, 1 adhesive disease, 1 hemorrhagic ovarian cyst, 1 hydrosalpinx, 1 benign cystic 
teratoma with struma ovarii, and 1 para-tubal cyst), and 10 healthy female controls. 
Samples were obtained at Duke University Medical Center with IRB-approved informed 
consent. All sera samples were stored at –80°C until analyzed, and case and control 
samples were matched for length of time in storage. Serum Radioimmunoassay: Frozen 
sera were thawed, diluted and analyzed in duplicate for IGFBP-2 utilizing a 
radioimmunoassay kit according to the manufacturer’s instructions (IGFBP-2 RIA; DSL-
7100; Diagnostic Systems Laboratories, Inc., Foster, TX). The anti-IGFBP-2 antibody is 
not cross-reactive with IGFBP-3, -4, -5, or -6. To obtain an estimate of inter-assay 
variation, a group of 10 samples comprising low, medium, and high values were re-
assayed using a separate kit. The mean percentage agreement +/- standard deviation 
(SD) between the two estimates was 92.4 +/- 18.7%, yielding an inter-assay coefficient 
of variation of 20.2%. The average ratio of duplicate determinations (n = 32) within a 
randomly chosen assay was 0.995 +/- 0.05 (SD).

**Statistical Analysis:** Associations between clinicopathological parameters such 
as stage, grade, histology (serous and non-serous), and debulking status, and relative 
IGFBP-2 expression were analyzed by the Chi square test or the Fisher’s exact test, 
when appropriate. Only patients for whom the status of all variables was known were 
included in the multivariate regression models. Information on progression-free interval 
was not available for all patients, despite review of all available medical records. 
Progression-free interval was not, therefore, included in the analysis. Multivariate 
survival analyses were performed by constructing Kaplan-Meier survival curves and 
differences between curves were evaluated by the log-rank test. For further analysis, 
patients were divided into two groups, high IGFBP-2 expression and low IGFBP-2 
expression, using the mean value of relative IGFBP-2 expression for all samples as a 
cut-off. The differences between the survival curves between IGFBP-2 expression 
groups were tested for statistical significance by the log-rank test.
RESULTS

Quantitative real-time polymerase chain reaction (QRT-PCR) was used to determine relative expression of the IGFBP-2 gene in 113 epithelial ovarian cancers and 6 normal ovarian surface epithelial samples (Table 1). For each sample, IGFBP-2 expression was calculated as a relative value from a serial dilution curve of the sample demonstrating the highest IGFBP-2 expression. Mean relative IGFBP-2 expression in 113 ovarian cancers (Table 1) was approximately 37-fold higher in patients with ovarian cancer compared with 6 normal ovarian surface epithelial samples (0.223 vs. 0.006, p<0.001). Expression in 10 early-stage (I/II) and 103 advanced-stage (III/IV) cancers was 22-fold (p=0.02) and 39-fold (p<0.0001) higher, respectively, than normal ovarian surface epithelium. In advanced-stage (III/IV) cancers, IGFBP-2 expression was 1.7-fold greater than early-stage (I/II) cancers (p=0.05). Serous cancers showed a trend to higher levels of IGFBP-2 relative to other histopathologic subtypes, although this did not reach significance (0.23 vs. 0.17, p=0.15). IGFBP-2 expression was 1.7-fold higher in cancers from 95 Caucasian-compared to 14 African-American patients (0.2289 vs. 0.1357, p=0.05). There was no relationship between IGFBP-2 expression and histopathologic grade (well-, moderate-, or poorly-differentiated), survival or extent of cytoreductive surgery (optimal versus suboptimal).

Pre-operative serum IGFBP-2 levels were measured by radioimmunoassay in 42 patients with ovarian cancer, 26 patients with benign gynecologic disease, and 10 healthy controls (Table 2). Mean pre-operative serum IGFBP-2 levels were 5.2-fold higher in 42 patients with ovarian cancer compared to 10 healthy controls (1274.6 ng/ml vs. 243.8 ng/ml, p<0.001), and 3.2-fold higher compared to 26 patients with benign gynecologic disease (1274.6 ng/ml vs. 394.2 ng/ml, p<0.001). Mean pre-operative serum levels of IGFBP-2 in 11 patients with early-stage (I/II) ovarian cancer were 3.7-fold higher than healthy controls (911.8 ng/ml vs. 243.8 ng/ml, p=0.02), and 2.3-fold higher than patients with benign gynecologic disease (911.8 ng/ml vs. 394.2 ng/ml, p=0.05).

128
Table 1. Relative expression of the IGFBP-2 gene in ovarian cancers and normal ovarian epithelial (NOSE) samples determined by quantitative real-time polymerase chain reaction.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean IGFBP-2 Expression</th>
<th>Difference (fold)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal ovarian epithelium</td>
<td>6</td>
<td>0.0059</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ovarian cancers</td>
<td>113</td>
<td>0.2227</td>
<td>37</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Early stage</td>
<td>10</td>
<td>0.1292</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced stage</td>
<td>103</td>
<td>0.2291</td>
<td>1.8</td>
<td>0.05</td>
</tr>
<tr>
<td>Serous cancers</td>
<td>88</td>
<td>0.2353</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-serous cancers</td>
<td>25</td>
<td>0.1735</td>
<td>1.4</td>
<td>0.15</td>
</tr>
<tr>
<td>Caucasian</td>
<td>95</td>
<td>0.2289</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>14</td>
<td>0.1357</td>
<td>1.7</td>
<td>0.05</td>
</tr>
<tr>
<td>Residual disease &lt;1cm</td>
<td>59</td>
<td>0.2116</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual disease &gt;1cm</td>
<td>54</td>
<td>0.2349</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Serum IGFBP2 (ng/ml)</td>
<td>Fold-difference</td>
<td>p-value</td>
<td>Comment</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------</td>
<td>-----------------</td>
<td>---------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Cancer</td>
<td>42</td>
<td>1274.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>10</td>
<td>243.8</td>
<td>5.2</td>
<td>&lt;0.001 Cancer vs. Normal</td>
</tr>
<tr>
<td>Benign Disease</td>
<td>26</td>
<td>394.2</td>
<td>3.2</td>
<td>&lt;0.001 Cancer vs. Benign disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.6</td>
<td>0.01 Benign disease vs. Normal</td>
</tr>
<tr>
<td>Stage I/II</td>
<td>11</td>
<td>911.8</td>
<td>3.7</td>
<td>0.02 Stage I/II vs. Normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.3</td>
<td>0.05 Stage I/II vs. Benign disease</td>
</tr>
<tr>
<td>Stage III/IV</td>
<td>31</td>
<td>1403.3</td>
<td>5.7</td>
<td>&lt;0.001 Stage III/IV vs. Normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.5</td>
<td>&lt;0.001 Stage III/IV vs. Benign disease</td>
</tr>
</tbody>
</table>

Table 2. Pre-operative serum levels of IGFBP-2 in patients with ovarian cancer, benign gynecologic disease and normal controls.

Thirty-one patients with advanced stage (III/IV) ovarian cancer had mean pre-operative serum IGFBP-2 levels that were 5.7-fold (1403.3 ng/ml vs. 243.8 ng/ml, p<0.001) and 3.6-fold (1403.3 ng/ml vs. 394.2 ng/ml, p<0.001) higher than healthy controls and patients with benign gynecologic disease, respectively. For nine of the patients with advanced stage serous epithelial ovarian cancer we were able to correlate pre-operative serum IGFBP-2 levels with IGFBP-2 expression in corresponding cancer tissue measured by microarray analysis. For these nine patients, serum IGFBP-2 levels correlated with corresponding tissue IGFBP-2 expression measured by microarray analysis (correlation coefficient = 0.7, ANOVA p<0.05).
DISCUSSION

Evidence is accumulating in support of a critical role for the insulin-like growth factor pathway in regulation of cell proliferation, differentiation, apoptosis, and transformation in many human cancers (17-26). The insulin-like growth factor (IGF) family is composed of two polypeptide ligands (IGF-1 and IGF-2), two cell membrane receptors (IGF-1R and IGF-2R), six binding proteins (IGFBP-1 through IGFBP-6), and a large group of IGFBP proteases. The insulin-like growth factors, IGF-1 and IGF-2, are 7-kDa polypeptides with structural homology to each other and to proinsulin (16). IGF-1 and IGF-2 have been shown to act as potent mitogens for a variety of cancer cell lines, including sarcoma, leukemia, and cancers of breast, prostate, colon, lung, stomach, liver, esophagus, pancreas, thyroid, brain, kidney, endometrium, cervix and ovary (17-26).

Most of the circulating IGFs are produced by hepatocytes in response to growth hormone stimulation (27) and circulate in serum, bound (~99%) to IGFBPs (28,29). IGF binding proteins such as IGFBP-2 are thought to modulate IGF action by regulating their sequestration or release at the tissue level (30). In contrast, although sequestration of IGFs by IGFBPs are usually inhibitory to growth and proliferation, IGFBPs can augment the mitogenic effects of IGF by acting as a reservoir, presenting and slowly releasing IGF-1 for receptor interactions while protecting the receptor from down-regulation by high IGF-1 exposure (31). Finally, it has been suggested that IGFBP-2 may have a direct effect on growth and apoptosis (21).

IGFBP-2 is expressed in many malignancies including lung, colorectal, and Wilm’s tumor. Most recently, it has been shown using cDNA and tissue microarrays that over-expression of IGFBP-2 may be an important event in the development of hormone-refractory prostate cancer (32). Among 5184 genes analyzed by cDNA microarray, IGFBP-2 was consistently over-expressed in hormone-refractory, compared to hormone-sensitive, prostate cancers. Additionally, serum levels of IGFBP-2, the main IGFBP produced by prostate epithelial cells, have been shown to be elevated relative to healthy, age-matched controls (33,34).
In this study, we have investigated IGFBP-2 serum levels and tissue expression in normal ovarian epithelium and ovarian cancers. Several small studies have investigated the role of IGFBP-2 in ovarian cancer. Karasik et al. demonstrated high IGFBP-2 levels in malignant ovarian cyst fluid, but almost absent levels in benign cysts (35). Further, IGFBP-2 levels in sera of 11 women with epithelial ovarian cancer were higher compared to levels in 20 women with benign cysts (1.32 +/- 0.32 vs. 0.53 +/- 0.07 relative units; P = 0.004), implying to the authors that the elevated sera IGFBP-2 is produced locally (35). More recently, pre-operative IGFBP-2 serum levels were shown to be significantly elevated in serum samples from 30 patients with epithelial ovarian cancer compared to 12 controls undergoing surgery for non-ovarian pathology (902 +/- 58 ng/mL vs. 416 +/- 50 ng/mL; p<0.0001) (24). In this study, IGFBP-2 serum levels correlated positively with advanced stage, suboptimal debulking status, and early relapse, however, no correlation was seen with the serum tumor marker, CA-125 (36).

In our analysis of *IGFBP-2* expression in 113 ovarian cancers and 6 NOSE samples, we have demonstrated that ovarian cancers express 37-fold higher levels of IGFBP-2 mRNA than cultured normal human ovarian epithelial cells. It has previously been shown that *IGFBP-2* expression is higher in malignant breast cells compared to normal breast tissue (21). Immunostaining for IGFBP-2 was seen in 52% of human breast cancers, but not in normal breast. In contrast, *IGFBP-2* mRNA was identified using QRT-PCR in essentially all normal and malignant breast tissues, suggesting that while IGFBP-2 mRNA is expressed in all cases, translation to protein occurs predominantly in cancers. Although we identified *IGFBP-2* mRNA in both normal and malignant ovarian epithelial samples, significantly higher levels were seen in the cancers. The possibility that translation to protein also may be greater in cancers warrants further study.

Our analysis suggests that high *IGFBP-2* expression is associated with a more aggressive cancer phenotype. Significantly higher levels of *IGFBP-2* were demonstrated in advanced stage compared to early stage ovarian cancers. An association between high *IGFBP-2* expression and an aggressive phenotype would be consistent with our microarray findings, which demonstrated 2.7 fold higher expression
of IGFBP-2 in cancers of short-term (<2 years) survivors compared to long-term (>7 years) survivors.

Over the past several years, serological concentrations of the components of the IGF pathway have been shown to be differentially elevated or reduced in patients with breast, prostate, liver, and colorectal cancers (37-41). A general pattern of elevated serum IGFBP-2 with advanced disease has emerged.

Consistent with earlier small studies in ovarian cancer, we were able to demonstrate significantly higher pre-operative serum levels of IGFBP-2 in 42 patients with ovarian cancer compared to 10 healthy controls and 26 patients with benign disease (16,35). Additionally, our sample size was large enough to demonstrate a statistical difference in serum IGFBP-2 levels between patients with early and advanced stage disease, and also between controls, patients with benign disease, and early stage ovarian cancer. If IGFBP-2 is to have any utility as a clinical serum marker, the ability to differentiate healthy controls and patients with benign conditions from patients with early- (as opposed to advanced-) stage disease is essential. Of note, nutritional status is known to affect serum IGFBP-2 levels, and starvation has been shown to increase IGBP-2 levels significantly (42). It is therefore possible that nutritional deprivation may contribute to the higher IGFBP-2 levels seen in pre-operative serum from patients with ovarian cancer. However, given the higher IGFBP-2 levels in cyst fluid compared to corresponding sera (35), and the correlation we demonstrate between IGFBP-2 gene expression measured by microarray and serum protein levels, it is likely that elevated IGFBP-2 levels are attributable to increased production by cancer cells.

Our findings further implicate the IGF pathway in human epithelial ovarian cancer development and progression. Further, we have demonstrated that the application of microarray analysis can facilitate the identification of genes, genetic pathways, and proteins involved in the pathogenesis of epithelial ovarian cancer that may not only serve as novel serum markers, but also avenues for therapeutic intervention.

ACKNOWLEDGMENTS
This research was supported by The Kathy Astrove Ovarian Cancer Research Fund and work grant # R03-HD39847 from the NICHD, NIH (DWS).
REFERENCES


22. Nickerson T, Pollak M, Huynh H. Castration-induced apoptosis in the rat ventral prostate is associated with increased expression of genes encoding insulin-like growth factor binding proteins 2, 3, 4 and 5. Endocrinology 1998;139:807-810.


40. Ma J, Pollak MN, Giovannucci E, Chan JM, Tao Y, Hennekens CH, Stampfer MJ. Prospective study of colorectal cancer risk in men and plasma levels of insulin-


Chapter 6

High insulin-like growth factor-2 (IGF-2) gene expression is an independent predictor of poor survival for patients with advanced stage serous epithelial ovarian cancer.

Published as:

Note: First & second authors contributed equally to this work
ABSTRACT

Objective: Epithelial ovarian cancer is the deadliest gynecologic malignancy, yet its molecular etiology remains poorly understood. Evidence is accumulating to support a role for the insulin-like growth factor family in human carcinogenesis, and recently using microarray expression analysis, we demonstrated over-expression of the insulin-like growth factor-2 (IGF-2) gene in advanced stage epithelial ovarian cancers. The purpose of the current study is to further elucidate the role of the IGF-2 gene in ovarian cancer development and progression.

Methods: Relative expression of IGF-2 was measured in 109 epithelial ovarian cancers and 8 normal ovarian surface epithelial (NOSE) samples, using quantitative real-time polymerase chain reaction. Associations with clinico-pathological parameters were examined.

Results: Expression of the IGF-2 gene was more than 300-fold higher in ovarian cancers compared with normal ovarian surface epithelium samples (p<0.001). High IGF-2 expression was associated with advanced stage disease at diagnosis (p<0.001), high grade cancers (p<0.05) and sub-optimal surgical cytoreduction (p=0.08). In multivariate analysis, relative IGF-2 expression was an independent predictor of poor survival.

Conclusions: Expression of the IGF-2 gene is significantly higher in ovarian cancers relative to normal ovarian surface epithelium. Further, high IGF-2 gene expression is associated with high grade, advanced stage, and is an independent predictor of poor survival in patients with epithelial ovarian cancer. As such IGF-2 is a molecular marker and potential therapeutic target for the most aggressive epithelial ovarian cancers.
Introduction

Epithelial ovarian cancer is the most lethal gynecologic malignancy. Currently, no effective screening or early detection techniques exist and approximately 80% of ovarian cancers have metastasized extensively throughout the peritoneal cavity at the time of diagnosis. Despite surgical cytoreduction and platinum-based chemotherapy, median survival is approximately 3 - 5 years; however, a minority of women survive much longer (1). To date, efforts to fully elucidate the molecular etiology of such survival differences have been unsuccessful (2,3). It is likely that the marked clinical differences in ovarian cancer stage at presentation, response to therapy and ultimately survival, are manifestations of a complex underlying molecular heterogeneity, such that traditional single-gene analysis techniques have proven inadequate. The recent development of microarray expression analysis technology has transformed the study of carcinogenesis by enabling the simultaneous study of thousands of genes in a single sample. Study of genome-wide expression is beginning to provide a more global understanding of the complex genetic changes that lead to malignant transformation in a variety of cancers. Our laboratory recently applied microarray technology to identify patterns of gene expression associated with outcome in women with advanced stage epithelial ovarian cancers (4). In that analysis, several members of the insulin-like growth factor family exhibited differential expression between ovarian cancers and normal ovarian surface epithelium samples. Using microarray analysis we demonstrated that the IGF-2 gene was 13-fold (p<0.01) more highly expressed in 31 advanced ovarian cancers compared to 3 normal ovarian surface epithelium samples. Previously, expression of the IGF-2 gene has been associated with development of cancers of the breast, colon, and prostate (5,6,7). Additionally, IGF-2 expression has been shown to be elevated in ovarian cancers versus normal ovarian tissue. Notably, Sawiris et al. recently analyzed a large series of microarray data in order to develop a gene chip specific to the study of ovarian cancer. The analysis identified 516 up-regulated genes with the IGF-2 gene being the most highly over-expressed of this group (8).
In light of this, and our recent array findings in advanced serous ovarian cancers, we sought to more fully elucidate the role of IGF-2 in ovarian carcinogenesis by applying quantitative real time polymerase chain reaction to measure the expression of the IGF-2 gene in a more comprehensive panel of epithelial ovarian cancers.

**MATERIALS AND METHODS**

**Study Population:** One hundred and nine fresh frozen epithelial ovarian cancers and 8 normal human ovarian surface epithelium samples were obtained with IRB-approved informed consent from patients treated by the Division of Gynecologic Oncology at Duke University Medical Center. Tumor tissue was harvested at the time of initial surgery, snap frozen in liquid nitrogen, and stored at -80C until RNA extraction. Hematoxylin and eosin staining of representative sections of each tumor was performed to determine the percentage of tumor in each specimen. Only samples consisting of more than 70% tumor were used for these studies. Histopathologic subtypes included 83 serous, 3 endometrioid, 7 mucinous, 6 clear cell, 1 undifferentiated and 9 mixed type. Ten of the cancers were early stage (I/II), and 99 advanced stage (III/IV). All patients with advanced stage disease were treated with primary surgical cytoreduction (49 optimal, 48 suboptimal, 2 unknown) followed by platinum-based chemotherapy.

**Tissue Preparation and Primer Design:** Total RNA was extracted using the RNAeasy RNA extraction kit (Qiagen Inc., Valencia, CA), and an aliquot of 1ug was subjected to a reverse transcriptase reaction with the Roche First Strand cDNA kit (Roche, Basel, Switzerland). Complementary DNA (cDNA) was quantified by fluorometry. Based on the published genomic sequence of IGF-2, two gene-specific intron-spanning primers were designed \( \text{(F5'-CTGTGCTACCCCGCCAAGT-3'} \) and \( \text{R5'-ACGTTTGGCCTCCTGAACG-3')}. \)

**Quantitative RT-PCR:** Quantitative RT-PCR was performed on the Roche LightCycler system (Roche, Basel, Switzerland) using SYBR Green I dye. For each sample, 5ng of cDNA was subjected to 40 cycles of QRT-PCR in a total volume of 20uL containing 10mM of each primer, 10 units of AdvanTaq Plus DNA Polymerase with PCR reaction buffer (Clontech, Palo Alto, CA), dNTP, and SYBR Green I dye according to manufacturer’s instructions in the presence of intron-spanning cDNA primers to the

142
house-keeping gene, beta-microglobulin, to confirm adequate cDNA normalization. Serial dilutions of the sample demonstrating highest IGF-2 expression were used to create a concentration curve and relative expression levels calculated for each sample from this curve. The standard curves were included in each run. Sample concentrations fell within the range of the serial dilution curve thus ensuring adequate assay sensitivity (Figure 1). All samples were run in triplicate and mean values were calculated.

**Figure 1.** Epithelial ovarian cancers and normal ovarian surface epithelial samples were subject to quantitative real-time polymerase chain reaction to determine relative expression of the IGF-2 gene. Serial dilutions of the most highly expressed cancer were used to create a concentration curve (bottom panel) and relative expression levels were calculated for each case from the real time increase in fluorescence demonstrated in the top panel.

**Statistical Analysis:** Statistical analyses were performed using the R statistical programming environment software (9). Associations between clinicopathological parameters such as stage, grade, histology (serous and non-serous), and debulking status, and relative IGF-2 expression were analyzed by the Chi square test or the Fisher's exact test, when appropriate. Only patients for whom the status of all variables was known were included in the multivariate regression models, therefore, a subset of
98 samples (81 serous and 17 nonserous) was analyzed in this way. Multivariate survival analyses were performed by constructing Kaplan-Meier survival curves and differences between curves were evaluated by the log-rank test. Cox proportional hazards models were used to estimate relative risk to determine whether the prognostic value of relative IGF-2 expression disappeared when other prognostic factors were considered. For further analysis, patients were divided into two groups, high IGF-2 expression and low IGF-2 expression, using the median value of relative IGF-2 expression for all samples as a cut-off. The differences between the survival curves between IGF-2 expression groups were tested for statistical significance by the log-rank test.

**Results**

Quantitative real-time polymerase chain reaction (QRT-PCR) was used to determine relative expression of the IGF-2 gene in 109 epithelial ovarian cancers and 8

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Mean IGF-2 Expression</th>
<th>Difference (Fold)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ovarian Cancers</td>
<td>109</td>
<td>0.43</td>
<td>330</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Normal ovarian epithelium</td>
<td>8</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survival &lt; 45 months</td>
<td>47</td>
<td>0.7</td>
<td>7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Survival &gt; 45 months</td>
<td>40</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survival &lt; 24 months</td>
<td>27</td>
<td>0.7</td>
<td>7</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Survival &gt; 60 months</td>
<td>29</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced Stage (III/IV)</td>
<td>99</td>
<td>0.47</td>
<td>5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Early Stage (I/II)</td>
<td>10</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade II/III</td>
<td>97</td>
<td>0.48</td>
<td>6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Grade I</td>
<td>6</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debunking Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suboptimal</td>
<td>48</td>
<td>0.2</td>
<td>1.8</td>
<td>0.08</td>
</tr>
<tr>
<td>Optimal</td>
<td>49</td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suboptimally Debulked</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survival &lt; 45 months</td>
<td>28</td>
<td>0.89</td>
<td>6.6</td>
<td>0.001</td>
</tr>
<tr>
<td>Survival &gt; 45 months</td>
<td>17</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimally Debulked</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survival &lt; 45 months</td>
<td>18</td>
<td>0.5</td>
<td>4</td>
<td>0.07</td>
</tr>
<tr>
<td>Survival &gt; 45 months</td>
<td>26</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
normal ovarian surface epithelial samples. Table 1 depicts the distribution of relative IGF-2 expression in the ovarian tissue samples in relation to the various clinicopathological variables of the study population.

Mean relative IGF-2 gene expression in 109 ovarian cancers was approximately 330-fold higher than in 8 normal ovarian surface epithelial samples (0.43 vs. 0.001, p<0.001). Advanced stage cancers demonstrated higher relative IGF-2 gene expression than early stage cancers. Relative IGF-2 gene expression from 99 patients with advanced (stage III/IV) disease was 5-fold higher than 10 cancers from patients with early (stage I/II) disease (0.47 vs. 0.09, p<0.001). Overall, relative expression of the IGF-2 gene in 97 moderate- and poorly-differentiated (grade II/III) cancers was 6-fold higher than in 6 well-differentiated (grade I) cancers (0.48 vs. 0.08, p<0.001). Within the group of patients with advanced (stage III/IV) disease, relative expression of the IGF-2 gene in 96 moderate- and poorly-differentiated (grade II/III) cancers was 3.6-fold higher than in 3 well-differentiated (grade I) cancers (0.5 vs. 0.14, p=0.03). Ovarian cancers that expressed high levels of IGF-2 were more likely to be incompletely resected at the time of initial surgery, i.e. high relative expression of the IGF-2 gene was associated with suboptimal tumor debulking (residual tumor >1cm at conclusion of surgery). Debulking status was available for 97/99 (98%) patients with advanced stage disease. Relative IGF-2 gene expression was 1.8-fold higher in 48 patients with residual tumor >1cm at conclusion of surgery (suboptimally debulked), compared to 49 patients with residual tumor <1cm (optimally debulked) at conclusion of surgery (0.62 vs. 0.33, p=0.08). Additionally, a survival detriment was associated with high IGF-2 gene expression in patients who were subject to optimal and suboptimal debulking at time of initial surgery. In the sub-optimally debulked category, 28 cancers from patients who died within 45 months (median survival for all patients in this study), of diagnosis exhibited approximately 6.6-fold higher relative IGF-2 gene expression than cancers from 17 patients who survived more than 45 months (0.89 vs. 0.13, p=0.001, 4 samples censored). Similarly, in patients who underwent optimal debulking, cancers from 18 patients who died within 45 months of diagnosis exhibited approximately 4-fold higher relative IGF-2 gene expression than cancers from 26 patients who survived more than 45 months (0.5 vs. 0.12, p=0.07, 5 samples censored).
Figure 2. Kaplan-Meier log-rank survival analysis of survival for patients with advanced (stage III/IV) epithelial ovarian cancer. The cut-off value used to determine high versus low IGF-2 gene expression was selected as the overall median IGF-2 gene expression value calculated for all 109 ovarian cancer samples. Chi-square = 3.3, DF=1, p=0.07.

Figure 3. Logistic regression models were applied to histological subgroups of patients using the median survival of 45 months as a cut-off. IGF-2 expression was found to be a significant predictor of survival in the subgroup of patients with serous cancers (p=0.0006), but not among non-serous cancers (p=0.484).
Multivariate analyses were performed to assess predictors of survival. Covariates considered included age, grade, stage, histology, and relative IGF-2 expression. A subset of 98 cancers were included in the analysis (10 samples missing complete covariate data were excluded). Considering all cancers, high relative IGF-2 gene expression was associated with poor outcome. Using the median survival as a cut-off, across all cancers the survival curves were significantly different (p<0.001) between those patients with a mean relative IGF-2 gene expression above or below the mean (Figure 2). Using multivariate Cox proportional hazards models with relative IGF-2 gene expression as the only predictor of overall survival, IGF-2 gene expression was highly significant across all cancers (p<0.001) and also in serous cancers alone (p<0.001); IGF-2 gene expression was not significant among the non-serous cancers (p=0.48) (Figure 3). Only IGF-2 expression (p<0.001) and stage at diagnosis (p<0.001) were found to be significant independent predictors of outcome overall and also among serous cancers alone.

Discussion

The insulin-like growth factor pathway is a complex system comprised of two growth factors (IGF-1 and IGF-2), two cell surface receptors (IGF-1R and IGF-2R), six specific high-affinity binding proteins (IGFBP-1 to IGFBP-6), IGFBP proteases as well as several other IGFBP-interacting molecules, which regulate and potentiate IGF actions.

The IGF ligands, IGF-1 and IGF-2, are single chain molecules of 70 and 67 amino acids, respectively. Both ligands interact with high affinity membrane bound receptors and are thought to work in an autocrine, paracrine and endocrine manner to mediate cell growth, differentiation, apoptosis, and transformation. Circulating levels are under complex physiological regulation. The majority of circulating IGFs are bound to high-affinity IGF binding proteins (IGFBPs), with >90% bound to a high molecular weight complex comprised of IGF binding protein 3 (IGFBP-3) and a separate protein known as the acid-labile subunit. IGFBPs appear to be responsible for protecting circulating IGFs and delivering them to their specific target sites. IGFBPs can also interact with IGFs in both an inhibitory and enhancing manner (10-12). Disruption of the
delicate balance of these complex processes may result in uncontrolled cell proliferation leading to malignant transformation, and evidence is accumulating that IGFs are important mitogens involved in the development of many types of malignancy (13).

To date, the role of IGFs in ovarian carcinogenesis and progression remains unclear. The most comprehensive study of the IGF pathway in ovarian cancer was presented by Conover et al. They found that primary ovarian epithelial cell lines derived from previously untreated ovarian cancers expressed all major components of the IGF system and were able to demonstrate functional responses to exogenous IGFs (14).

Using microarray analysis we recently demonstrated that the IGF-2 gene exhibits higher expression in advanced stage ovarian cancers compared to normal ovarian epithelium (4). As a result of this preliminary microarray data, we now report a quantitative analysis of insulin-like growth factor-2 gene expression levels in a set of 109 epithelial ovarian cancers, using quantitative real-time polymerase chain reaction. In this large study that includes both early and advanced stage cancers as well as serous and non-serous histopathologic subtypes, we have demonstrated more than 300-fold higher IGF-2 gene expression in ovarian cancers compared to normal ovarian surface epithelium. Such findings are consistent with other studies that have demonstrated increased IGF-2 gene expression in cancers compared to normal tissue (5-7).

Several mechanisms have been proposed to explain increased levels of IGF-2 including: loss of imprinting (LOI), loss of heterozygosity (LOH) with paternal duplication, excessive transcriptional activation, loss of transcriptional suppression, or alteration in IGF-binding proteins (15). The IGF-2 gene is one of the few known imprinted genes. In normal cells, IGF-2 is expressed only from the paternal copy of the gene, i.e. it is maternally imprinted. Loss of IGF-2 imprinting has been described in a number of cancers, including Wilm’s tumors, gastric adenocarcinomas, lung cancers, and colorectal cancers (16-19). When loss of imprinting occurs, biallelic expression of IGF-2 results, ultimately leading to overexpression of this potent growth factor.

Loss of IGF-2 imprinting has previously been implicated in the development of ovarian carcinomas. Chen et al determined the imprinting status of the IGF-2 gene in 43 ovarian cancers, 7 low malignant potential ovarian tumors, and their matched normal
tissues. Loss of imprinting (LOI) was found in 5 of 20 (25%) cancers. Overexpression of the IGF-2 gene was demonstrated in all five LOI samples. These results suggested that LOI may underlie overexpression of the IGF-2 gene resulting in development of some ovarian cancers (20).

Our analysis suggests that high IGF-2 expression is associated with a more aggressive ovarian cancer phenotype. High IGF-2 gene expression was associated with higher stage at diagnosis, tumor grade and, in multivariate analysis, shorter overall survival. Additionally, cancers exhibiting high IGF-2 gene expression demonstrated a trend towards being more likely to be incompletely resected at the time of primary surgery (p=0.08), and within this group of sub-optimally debulked patients, a high IGF-2 gene expression was associated with shorter survival (p<0.01). These findings are consistent with studies in other cancer types, which have demonstrated a correlation between increased IGF-2 expression and decreased survival. In colorectal carcinomas IGF-2 expression was associated with advanced Duke’s stage and poor clinical outcome (21). In our analysis, advanced stage ovarian cancers have 5-fold higher relative IGF-2 gene expression than early stage cancers, and moderate- and poorly-differentiated (grade II/III) cancers have 6-fold higher expression than well-differentiated (grade I) cancers. In a similar fashion, a correlation between high IGF-2 expression and poor prognostic clinical factors has been demonstrated in other cancers. IGF-2 expression was recently shown to be correlated with pathologic stage, lymph node metastasis, histologic differentiation and serum prostate-specific antigen (PSA) levels in 24 prostate carcinomas (22). Previous studies have shown that IGF-2 acts as a potent mitogen for a variety of cancer cell lines. Based on work in breast, lung, and colon cancers, a general pattern emerges that IGF-2 expression increases as the malignant potential of the tumor increases (23,24).

While mRNA levels do not account for regulatory events at the translational and post-translational level, several studies have demonstrated corresponding increases in bioactive IGF-2 protein levels in cancer cell lines. Nielsen et al have used radioimmunoassay to demonstrate a correlation of increased levels of IGF-2 mRNA expression and secretion of bioactive IGF-2 in human rhabdomyosarcoma (25). Additionally, IGF-2 protein levels have been shown to be differentially expressed in
gastric cancer cell lines when compared with normal gastric cell cultures (26). IGF-2 protein levels, and the interactions of IGF-2 with its receptors and binding proteins, have yet to be described in ovarian cancers.

Preliminary studies in breast, prostate, hepatocellular, pancreatic, and ovarian cancer cell lines suggest that agents that block the action of IGF-2 may have potential therapeutic utility (27-30), such that the ability to down-regulated IGF-2 expression could have significant clinical application in the management of patients with ovarian cancer in the future.

Our results support the hypothesis that insulin-like growth factor-2 plays an important role in ovarian carcinogenesis. Additionally, we have demonstrated that IGF-2 is a potential molecular marker of the most aggressive forms of epithelial ovarian cancer. Further elucidation of the role of the IGF pathway in ovarian cancer development and progression will increase our understanding of the biology of ovarian carcinogenesis, and may yield novel opportunities for cancer prevention and therapy.
REFERENCES


Chapter 7

Identification of genes associated with ovarian cancer metastasis using microarray expression analysis

Submitted for Publication, as:
Lancaster JM, Dressman HK, Henriott AH, Sayer RA, Gray J, Whitaker RS, West M, Marks JR, Nevins JR, Berchuck A. Identification of genes associated with ovarian cancer metastasis using microarray expression analysis, to Gynecologic Oncology, 2005
ABSTRACT
Although the transition from early to advanced stage ovarian cancer is a critical determinant of survival, little is known about the molecular underpinnings of ovarian metastasis. We utilized Affymetrix U95Av2 microarrays to characterize the molecular alterations that underlie the development of omental metastasis from 47 epithelial ovarian cancer samples collected from multiple sites in 20 patients undergoing primary surgical cytoreduction for advanced stage (IIIC/IV) serous ovarian cancer. Fifty-six genes demonstrated differential expression between ovarian and omental samples (p<0.01), and 20/56 differentially expressed genes have previously been implicated in metastasis, cell motility or cytoskeletal function. Ten of 56 genes are involved in p53 gene pathways. A Bayesian statistical tree analysis was used to identify a 27-gene expression patterns that could accurately predict the site of tumor (ovary vs. omentum). Nine of these 27 genes have previously been shown to be involved in oncogenesis and/or metastasis, and 10/27 genes have been implicated in p53 pathways. We conclude that gene expression patterns that distinguish omental metastasis from epithelial ovarian cancer can be identified and that many of the genes have functions that are biologically consistent with a role in oncogenesis, metastasis and p53 gene networks.
INTRODUCTION

Epithelial ovarian cancer is the most lethal gynecologic cancer in the United States and Western Europe, and mortality from the disease has improved little over the last fifty years (1). While other gynecologic cancers are diagnosed at an early stage because of effective screening programs (e.g. cervical cancer) or symptoms (e.g. endometrial cancer), more than 70% of patients with ovarian cancer present at stage III or IV (2). These patients usually have extensive metastatic disease throughout the peritoneal cavity including the omentum, and have a five-year survival less than 30% despite surgical cytoreduction and adjuvant chemotherapy (3). In contrast, patients diagnosed when disease is confined to the ovary (stage I/II), have a survival that exceeds 90% with surgery alone (4).

It is possible that the survival disparity observed between patients with early- and advanced-stage epithelial ovarian cancer reflects fundamental differences in the molecular alterations that underlie their development. Expression of several molecular markers, including the p53 and HER2/neu genes have been shown to be associated with advanced stage at diagnosis (5, 6). Although this suggests that the phenotypic differences observed between early- and advanced-stage ovarian cancers may be the result of underlying biologic differences, the genetic alterations that underlie the transition from early- to advanced-stage disease have yet to be fully elucidated. We hypothesize that microarray analysis of global gene expression patterns in primary ovarian cancers and metastatic omental implants can define a discrete set of genes that underlie the metastatic process in epithelial ovarian cancer. Characterization of such differences would provide greater insight into the genetic progression that occurs within individual patients as ovarian cancer metastasize, and may reveal novel avenues for therapeutic interventions.

MATERIALS AND METHODS

Forty-seven ovarian cancer samples, collected from multiple sites in 20 patients, were subject to global expression analysis of approximately 12,000 genes using
Affymetrix Human U95Av2 arrays. Expression data was subject to statistical analysis by hierarchical clustering, student’s t-test of log normalized signal intensity values, and Bayesian predictive statistical tree model analysis, including iterative out-of-sample, cross-validation predictions.

**Tissues:** We identified 47 frozen cancer specimens obtained from 20 women with advanced (FIGO stage III/IV) serous epithelial ovarian carcinoma, treated at Duke University Medical Center. Twenty-nine samples were resected from ovary (11 from left ovary, 16 from right ovary and 2 side-unspecified) and 18 samples were resected from omentum. All samples were obtained at primary cytoreductive surgery from patients treated at Duke University Medical Center under an IRB approved protocol with written informed consent. Microscopic slides from each cancer sample were examined and only those with >70% tumor (on a per cell basis) were selected for RNA extraction. Hybridization targets (probes for hybridization) were prepared from total RNA according to standard Affymetrix protocols.

**Sample Processing:** Approximately 30 mg of tumor tissue was added to a chilled BioPulverizer H tube (Bio101). Lysis buffer from the Qiagen RNeasy Mini kit was added and the tissue was homogenized for 20 seconds in a Mini-Beadbeater (Biospec Products). Tubes were spun briefly to pellet the garnet mixture and reduce foam. The lysate was transferred to a new 1.5 ml tube using a syringe and 21 ga. needle, followed by passage through the needle 10 times to shear genomic DNA. Total RNA was extracted using the Qiagen RNeasy Mini kit (Qiagen Inc., Valencia, CA). RNA integrity was confirmed by the Agilent 2100 Bioanalyzer.

**Microarray Analysis Hybridization procedures and parameters:** Total starting RNA for each reaction was 10 μg of total RNA. First strand cDNA synthesis was performed using a T7-linked oligo-dT primer, followed by second strand synthesis. An *in vitro* transcription reaction was performed to generate the cRNA containing biotinylated UTP and CTP, which was subsequently chemically fragmented at 95°C for 35 min. The fragmented, biotinylated cRNA was hybridized in MES buffer (2-(N-morpholino)ethansulfonic acid) containing 0.5 mg/ml acetylated bovine serum albumin to Affymetrix GeneChip Human U95Av2 arrays at 45°C for 16hr, according to the Affymetrix protocol (www.affymetrix.com and
www.affymetrix.com/products/arrays/specific/hgu95.affx). The arrays contain over 12,000 genes and ESTs. Arrays were washed and stained with streptavidin-phycoerythrin (SAPE, Molecular Probes). Signal amplification was performed using a biotinylated anti-streptavidin antibody (Vector Laboratories, Burlingame, CA) at 3 μg/ml, followed by a second staining with SAPE. Normal goat IgG (2 mg/ml) was used as a blocking agent.

Measurement data and specifications: Scans were performed with an Affymetrix GeneChip scanner and the expression value for each gene was calculated using the Affymetrix Microarray Analysis Suite (v5.0), computing the expression intensities in 'signal' units defined by software. Scaling factors were determined for each hybridization based on an arbitrary target intensity of 500. Files containing the computed single intensity value for each probe cell on the arrays (CEL files), experimental and sample information, and signal intensity values for each probe set, as derived from the Affymetrix Microarray Analysis Suite (v5.0) software (pivot files), can be found in the Supplementary Material on the project web site (http://data.cgt.duke.edu/mcr.php)

Array design: All assays employed the Affymetrix Human U95Av2 GeneChip. The characteristics of the array are detailed on the Affymetrix web site (www.affymetrix.com/products/arrays/specific/hgu95.affx). MIAME (minimal information about a microarray experiment)-compliant information regarding the analyses performed here, as defined in the guidelines established by MGED (www.mged.org), is detailed in the following sections.

Statistical analysis: Mean signal intensity values for each gene were calculated and compared between samples originating in ovary and omentum. Significance of median differences for each gene between both groups was established by application of student’s t-test. The goal of the predictive analysis was to identify those gene and gene expression patterns characteristic of tumor samples from different sites based on predictive tree models. The analysis of the microarray data obtained from the Affymetrix GeneChip Hu95Av2 arrays utilized the signal intensity values generated in the Affymetrix MAS 5.0 software that were quantile normalized and log base 2 transformed to develop predictive statistical tree models. The expression data was first
log base 2 transformed and an initial screen was applied which removed genes with expression that varied little between samples from ovary and omentum (standard deviation < 1 across all samples or range of expression < 4 in at least 90% of samples). This yielded a set of 4,302 genes whose expression values were quantile normalized prior to inclusion in subsequent analysis. Application of K-means clustering to these genes identified 300 groups or clusters of genes that shared certain expression characteristics. A singular value decomposition (SVD) was performed on the expression data of each cluster and the resulting dominant singular factor (principal component) from each SVD was used as a candidate predictive pattern or metagene in the tree modeling. (7-9). Formal predictive analysis using these 300 metagenes and the 4302 genes was performed using Bayesian classification tree models. This analysis defines specific probabilities within nodes of multiple trees, defined by recursively splitting the data within each node according to a threshold on a chosen predictor (7,10,11). Forward selection of tree models chooses node splits progressively "down" a tree based on optimizing an association measure over the possible (predictor, threshold) combinations and testing whether or not to split based on an assessment of significance of each split (Figure 1A and B). Our association testing uses probability models, computing Bayes' factors to test a null hypothesis of a common probability distribution within a node relative to a split into two subpopulations. This process generates multiple recursive partitions of the sample into subgroups (the "leaves" of the classification tree), and associates predictive probabilities of outcomes (site) with each subgroup. Multiple trees can be "spawned" at any node based on multiple choices of predictors and thresholds and this generates classes of trees for combinations in prediction. Overall predictions for an individual sample are then generated by averaging predictions, with appropriate weights, across many such tree models. From the set of 4302 gene and 300 metagene predictors, a series of tree models were designed that separated the specimens by probability of originating in ovary versus omentum on the basis of expression of an individual predictor (either gene or metagene) at branches (nodes) within a tree. This Bayesian classification tree analysis approach was utilized to develop a gene/metagene model that most accurately predicted specimen site (ovary versus omentum) during 47 iterative out-of-sample, cross-validation predictions: leaving
**Figure 1A. Development of a Bayesian Statistical Tree Model:** To validate the predictive model, 47 out-of-sample cross validation analyses were performed. This is a representative Bayesian statistical tree model used in the predictive model. In this example, 47 blinded samples are analysed.

The pattern of gene expression of a specific group of genes – or metagene (#74 in this example) – is examined in each blinded specimen and a judgement is made as to whether the expression pattern most closely resembles a specimen originating in ovary or omentum. This is the first branch (or node) in the tree.

In this particular tree, the expression pattern of 2 additional genes, KIAA0146 and STAR, are used to further split the specimens based on whether they most closely resemble a specimen originating in ovary or omentum. Using this tree analysis, eventually 44/47 of the blinded samples are categorized correctly. Many such trees are evaluated in the predictive model that accurately predicts specimen site (ovary versus omentum) in 87% of cases during 47 leave-one-out cross-validation analyses. The tree model includes Metagene #74, and four other single-gene predictors.

**Figure 1B. Cross-validation probability predictions of tumor site:** A summary of the predictions from the 47 cross-validation analyses with 397 covariates. Samples are plotted by index number, and the vertical locations of the sample numbers indicate the estimated predictive probabilities of omentum as the site of the tumor. Approximately 90% uncertainty intervals about these estimated probabilities are indicated by vertical dashed lines. The uncertainty intervals are generated by averaging over the predictions from different tree models. Each individual is predicted in an out-of-sample cross validation based on a model completely regenerated from the data of the remaining patients. Omentum samples are marked in red and ovarian samples are marked in blue. A sample is predicted as an omentum sample if the predicted probability is greater than 0.5 and is predicted as an ovarian sample otherwise. It is important to recognize that each sample in the data set, when assayed in this manner, constitutes a validation set that accurately assesses the robustness of the predictive model. This gene/metagene model accurately predicts tumor site in 87% of cases based on a simple threshold at 0.5 on the estimated probability in each case.
each tumor out of the data set one at a time, refitting the model (both the metagene factors and the tree structures) from the remaining tumors, and then predicting the hold-out case. This rigorously tests the predictive value of a model.

RESULTS

Genes Exhibiting Differential Expression between Ovaries and Omentum: We began by comparing the expression of genes from ovarian cancers and corresponding omental metastasis. Median expression values, calculated for genes in samples originating in either ovary, were compared to values for samples collected from omental metastasis, and led to the identification of 56 genes which demonstrated differential expression with p-values < 0.01, (Table 1A and B). Twenty-three genes were more highly expressed in omentum compared to ovary, and 33 genes were more highly expressed in ovary compared to omental metastasis. At least 20/56 (36%) genes differentially expressed between ovary and omentum (p<0.01) have previously been implicated in metastasis, cell motility, migration and cytoskeletal function, and include ANGPT1 (12), MMP1 (13), CSPG2 (14), MAPT (15), DPT (16), KIAA1775 (17), F2RL2/PAR3 (18), PCDH9 (17), ADAM22 (19), TBX1 (20), EVPL (21), GPR135 (22-24), CDC2 (25), POLYDOM (26), CAMK2G (27), CCL19 (28), DLX2 (29), PPAP2C (30), SAG (31) and HMG1C (32). Ten of 56 (18%) genes that were differentially expressed between ovary and omentum (p<0.01) have gene network relationships with the p53 tumor suppressor gene and include MJD/ATX3 (33, 34), STAR (34-36), NEF3 (37, 38), NR1H4 (38-40), EVPL (41-43), TMPO (43, 44), PIR51 (43,45,46), APOBEC2 (47), MMP1 (48), and CDC2 (49). The P53 gene demonstrated a 0.9 fold difference when comparing expression in the omentum versus ovary.
<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Function</th>
<th>Symbol</th>
<th>p value (FDR)</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-protein coupled receptor protein signaling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p55 gene network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcium Channel, Voltage-dependent, Beta 4 Subunit</td>
<td>Calcium Channel, Voltage-dependent, Beta 4 Subunit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CACNB4</td>
<td>DLX2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF03682</td>
<td>L07919</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.006</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation of Transcription, DNA-dependent</td>
<td>Calcium Ion Transport Activity, Calcium Channel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcium Ion Transport Activity, Calcium Channel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIA0691</td>
<td>TMSF1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.006</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transmembrane 4 Superfamily Member 1 (Plasminogen)</td>
<td>Syntenin 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIA0694</td>
<td>SYNGR1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.098</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ubiquitous Tetra-epitope containing protein, RoXaN</td>
<td>Al22326</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIA0694</td>
<td>HCG1827</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.009</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tryptophan 2</td>
<td>AL080168</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-Protocadherin, Tubby-like Protein 2</td>
<td>TUBP2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF012131</td>
<td>U82469</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mitochondrial Complex, Protein, Mitochondrial</td>
<td>FAM224</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.004</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mitochondrial Complex, Protein, Mitochondrial</td>
<td>C21 Steroid Hormone Biosynthesis, Cholesterol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Embryogenesis and Morphogenesis, RNA Polymerase</td>
<td>Glycerol N-Acetyltransferase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.009</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell Adhesion</td>
<td>GLYAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF023466</td>
<td>BAAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.007</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell Adhesion</td>
<td>L3001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U17260</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALAT1</td>
<td>U51704</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J03778</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>microtubule-associated protein tau</td>
<td>Apoptosis-Microtubule</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B cell growth factor [Homo sapiens]</td>
<td>Organization and Biogenesis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycine N-Acetyltransferase</td>
<td>C21 Steroid Hormone Biosynthesis, Cholesterol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(glycine N-Acetyltransferase)</td>
<td>Glycerol N-Acetyltransferase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steroid Hormone Biosynthesis, Cholesterol</td>
<td>Glycerol N-Acetyltransferase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MB183883</td>
<td>C21 Steroid Hormone Biosynthesis, Cholesterol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca2+-dependent activator protein for secretion</td>
<td>Ca2+-dependent activator protein for secretion</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
 " 7

=    > 3 +: 0 =    > A3 +A  '  

7

C 0 0 -*+

*)-*

  ++ ,

(

(

-(--

-(--*

! $ ! 

 "#! 

Y  % '"% %7

!  "   

 $   ""L

(

-(--

D!   $ # $

 1 / 

"



""%! 

"& 

! 

"' & " $ " L '

,)*

-(--

! ' ! ;

2;3
(

7'!/
0>-

" $  



!  

" " !"

 !



" " !"%" "

;;0;)#

-(--

'&'&%% '&'&  ' )#

(

-(--)

Q)*

-(--

(

0

(

-

@

.),,)-

0;

 ' !!?
;#$=

'  

A G
' %&  

+


<table>
<thead>
<tr>
<th>Function</th>
<th>Gene Name</th>
<th>Symbol</th>
<th>Number/Function</th>
<th>Comment/Change</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development/Protein binding</td>
<td>SLT3</td>
<td>1</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possession in Transport</td>
<td>KCNJ13</td>
<td>1</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possession Inward-Reaching channel, subfamily</td>
<td>M1309.J1</td>
<td>1</td>
<td>0.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matrix Metalloproteinase (9) (Interstitial collagene)</td>
<td>MMP-1</td>
<td>1</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collagen cascade, Issue Remodeling, metastasis</td>
<td>RNF4</td>
<td>1</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA binding</td>
<td>ZFR</td>
<td>1</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integral to plasma membrane</td>
<td>ADAM22</td>
<td>2</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phospholipid metabolism</td>
<td>PNIPAP</td>
<td>2</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcription activity</td>
<td>HS2ST1</td>
<td>1</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human CD2 gene involved in cell cycle control</td>
<td>CDC2</td>
<td>1</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemokine (C-C motif) ligand 19</td>
<td>CCL-19</td>
<td>1</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene Name</td>
<td>Accession</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIAA0506 protein</td>
<td>KIAA0506</td>
<td>likely ortholog of mouse polydactyl homolog 3 (mouse).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle development, actin filament, cell motility</td>
<td>0.002</td>
<td>ARK56.45 1.3 Homo sapiens adenosine monophosphate (HMP)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell transport</td>
<td>0.004</td>
<td>MALT72 1.2 Maltadoid, ankyrin repeat and leucine-rich repeat-containing protein (MALT)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>MYC 1.2 Mucin, a heavily glycosylated protein (MUC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>TIAO 1.2 TIAO, a protein tyrosine kinase (TIAO)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>MIF 1.2 Monocyte chemoattractant protein (MIF)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA repair, P53 network gene</td>
<td>0.003</td>
<td>NPEF 1.2 NPF, a nuclear protein (NPF)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.004</td>
<td>ZR2865 1.2 ZR2865, a zinc finger protein (ZR2865)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>OPT 1.2 OPT, a protein kinase (OPT)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA repair, P53 network gene</td>
<td>0.002</td>
<td>XPS2 1.2 XPS2, a protein phosphatase (XPS2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>DAX1 1.1 DAX1, a zinc finger protein (DAX1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>CAMK2G 1.0 CAMK2G, a calcium/calmodulin-dependent protein kinase (CAMK2G)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Predictive modeling using Bayesian statistical tree analysis: If the identified gene expression profiles between primary ovarian cancer and omental metastasis have biological significance, then gene expression profiles that are characteristic of either site should be robust enough to sustain interrogation by predictive modeling techniques. Following the initial screen to filter out non-varying genes, 4,302 genes were subject to K-means clustering and 300 groups or clusters of genes that shared certain expression characteristics were identified. These genes and metagenes were subsequently used in Bayesian classification tree analysis approach in which models were developed to predict the site of tumor. Examining the predictors which appeared in the high likelihood trees across cross-validation runs yielded a set of 19 genes and 1 metagene (containing 8 genes); when this set of 20 predictors (27 genes in total) was used as the predictor set for the tree analysis, the result were models that accurately predicted specimen site (ovary versus omentum) in 87% of cases during 47 leave-one-out cross-validation analyses (Figure 1A and B). Eight of these genes are within metagene #74, and 19 are single-gene predictors (Table 2). Twelve of 27 (44%) predictor genes (GRP135 (22-24), RHL-2 (50-52), GA (53-57), LAMC2 (58), MAGE-A10 (59), CDK5 (60), RPS6KB1 (61), PARD6B (62), SLIT3 (63), AIM2 (64), ANGPT1 (65), NR4A3 (66,67)) have previously been shown to be involved in oncogenesis, and 10/27 (37%) genes (LAMC2 (68,69), APOBEC2 (70), FHL (50), STAR (34-36), ELAVL1 (41-43), ANGPT1 (70,71), CDK5 (72,73), RPS6KB1 (36, 74), PARD6B (75,76), and RFXAP (77-79,69)) have been implicated in p53 pathways (Figure 2).
Figure 2. To characterize the gene network relationships underlying the gene expression patterns identified, we applied web-based Ingenuity Pathways Analysis software (http://www.ingenuity.com) to our array data. This is a web-based application that enables identification of gene networks from expression array data sets, using a knowledge base created from millions of individually modeled gene, protein, tissue, drug and disease relationships. The members of the 27-gene predictive profile that accurately predict site of origin (omentum/ovary) of specimen in 87% of cases were further analyzed using this software. The resulting modeled gene network/pathway was given a numerical score by the software according to the number of genes within the pathway that are represented in the gene-list being studied, and also by the biologic proximity of these genes within the pathways. The top-scoring gene network/pathway is displayed. Ten of 27 (37%) genes (represented in bold with shaded background) that represent the predictive profile are demonstrated here to have gene network relationships with the p53 tumor suppressor gene. B, Binding; A, Activation/deactivation; E, Expression; I, Inhibition; P, Proteolysis; M, Biochemical Modification; O, Other; P, Phosphorylation/dephosphorylation; T, Transcription.

© 2000-2005, Ingenuity Systems
Discussion
We have utilized expression microarrays to identify genes that differentiate primary epithelial ovarian cancer from metastatic deposits within the omentum. More than one third of the genes identified as differentially expressed between ovary and omental metastasis have previously been reported to be involved in metastasis, cell motility and cytoskeletal function; a finding consistent with analysis in several other cancers. For instance, in human bladder cancer cell lines, genes that demonstrated greatest differential expression with metastatic potential included genes for the matrix component, fibronectin (FN1), osteonectin/SPARC, collagen, and tissue inhibitor of metalloproteinase (80). In pulmonary metastasis from osteosarcoma, 53 genes were differentially expressed between highly metastatic (K7M2) and less metastatic (K12) dOSA tumor models (81). Functional studies were used to assign 10 of 53 (19%) differentially expressed genes to categories likely to be responsible for metastasis, though none are common to differential or predictive genes identified in our current study, suggesting that the molecular determinants of metastasis in ovarian carcinoma differ from those in osteosarcoma. Such molecular disparity may underlie the differences in metastatic spread patterns observed between the two cancers; ovarian cancer most commonly spreads throughout the peritoneal cavity, whereas osteosarcoma metastasizes to distant sites such as lung. Microarray expression analysis of primary and metastatic liver cancers identified 35 genes to be down-regulated in metastasis (82). Six of 35 (17%) down-regulated genes were involved in immune-related function, including immunoglobulin lambda locus, consistent with our findings of decreased expression of immunoglobulins lambda light and heavy chains in metastatic omental samples, compared to primary ovarian cancer. These findings offer further support to the hypothesis that immune function plays an important role in ovarian and other human cancers (83).
<table>
<thead>
<tr>
<th>Function</th>
<th>Value</th>
<th>Symbol</th>
<th>Expression</th>
<th>Number (continued)</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>MOF</td>
<td>437.27</td>
<td>0.001</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>ENOGL1</td>
<td>652.16</td>
<td>0.001</td>
<td>1</td>
</tr>
</tbody>
</table>

Known to be involved in 95 gene pathways; are represented in function column by bold text. Genes which have previously been shown to be involved in malfunctions are represented in gene name column by bold text. Genes which have previously been deemed necessary to include in expression column by bold text.
<table>
<thead>
<tr>
<th>Transaction high activity &amp; network</th>
<th>10.0</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R110-1</td>
<td>1.1</td>
<td>8</td>
</tr>
<tr>
<td>4M silenced (host)</td>
<td>1113</td>
<td></td>
</tr>
<tr>
<td>Immune response vaccine</td>
<td>1.0</td>
<td>111</td>
</tr>
<tr>
<td>4M2 births</td>
<td>1479</td>
<td></td>
</tr>
<tr>
<td>Signal transduction vaccine network</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>Approa. oncogenic &amp; network</td>
<td>0.1</td>
<td>11</td>
</tr>
<tr>
<td>Ads. CD54+R50+</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>Kras expression family</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>WxN2-FcR-1 in vivo antigen</td>
<td>1.0</td>
<td>9</td>
</tr>
<tr>
<td>Abs. expression kinetics, tumor</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>WxN2-FcR-1 in vivo antigen</td>
<td>1.0</td>
<td>9</td>
</tr>
<tr>
<td>Abs. expression kinetics, tumor</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>WxN2-FcR-1 in vivo antigen</td>
<td>1.0</td>
<td>9</td>
</tr>
<tr>
<td>Abs. expression kinetics, tumor</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>4M silenced (host)</td>
<td>1113</td>
<td></td>
</tr>
<tr>
<td>Immune response vaccine</td>
<td>1.0</td>
<td>111</td>
</tr>
<tr>
<td>Signal transduction vaccine network</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>Approa. oncogenic &amp; network</td>
<td>0.1</td>
<td>11</td>
</tr>
<tr>
<td>Ads. CD54+R50+</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>Kras expression family</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>WxN2-FcR-1 in vivo antigen</td>
<td>1.0</td>
<td>9</td>
</tr>
<tr>
<td>Abs. expression kinetics, tumor</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>WxN2-FcR-1 in vivo antigen</td>
<td>1.0</td>
<td>9</td>
</tr>
<tr>
<td>Abs. expression kinetics, tumor</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>4M silenced (host)</td>
<td>1113</td>
<td></td>
</tr>
<tr>
<td>Immune response vaccine</td>
<td>1.0</td>
<td>111</td>
</tr>
<tr>
<td>Signal transduction vaccine network</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>Approa. oncogenic &amp; network</td>
<td>0.1</td>
<td>11</td>
</tr>
<tr>
<td>Ads. CD54+R50+</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>Kras expression family</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>WxN2-FcR-1 in vivo antigen</td>
<td>1.0</td>
<td>9</td>
</tr>
<tr>
<td>Abs. expression kinetics, tumor</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>WxN2-FcR-1 in vivo antigen</td>
<td>1.0</td>
<td>9</td>
</tr>
<tr>
<td>Abs. expression kinetics, tumor</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>4M silenced (host)</td>
<td>1113</td>
<td></td>
</tr>
<tr>
<td>Immune response vaccine</td>
<td>1.0</td>
<td>111</td>
</tr>
<tr>
<td>Signal transduction vaccine network</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>Approa. oncogenic &amp; network</td>
<td>0.1</td>
<td>11</td>
</tr>
<tr>
<td>Ads. CD54+R50+</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>Kras expression family</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>WxN2-FcR-1 in vivo antigen</td>
<td>1.0</td>
<td>9</td>
</tr>
<tr>
<td>Abs. expression kinetics, tumor</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>WxN2-FcR-1 in vivo antigen</td>
<td>1.0</td>
<td>9</td>
</tr>
<tr>
<td>Abs. expression kinetics, tumor</td>
<td>1.0</td>
<td>1</td>
</tr>
</tbody>
</table>
In addition to identifying genes that display differential expression between the primary site (ovary) and the metastasis (omentum), we have applied Bayesian statistical tree analysis to further refine and validate our findings. To this end, we developed a 27-gene predictor (Table 2) involving 19 single genes and one metagene based on a cluster of 8 genes that in cross-validation analyses was highly significant in accurately predicting site of origin (ovary versus omentum) in 87% of cases during “leave-one-out” cross validation. This level of accuracy provides some confidence that the gene profiles identified are reproducible, and as such represent legitimate biologic phenomena.

More than a third of the genes within the 27-gene predictor have previously been implicated in carcinogenesis or metastasis (22-24, 50-67). For example, leukocyte platelet-activating factor receptor/G protein-coupled receptor 135 (GPR135), binds platelet-activating factor (PAF), which is known to play an important role in controlling cell migration, tumor growth, dissemination and metastasis (22-24). PAF has a metastasis-enhancing effect on murine melanoma pulmonary metastasis, stimulates in vitro migration of endothelial cells, promotes in vivo angiogenesis and acts as a potent inducer of tumor cell motility (84). In fact, platelet-activating factor receptor-binding antagonists have anti-metastatic properties (85). We have demonstrated 0.8-fold lower expression of the GPR-135 gene in omental metastasis compared to primary ovarian cancer (p=0.005), which not only implies a role for the PAF pathway in ovarian metastasis, but reveals a potential novel avenue for therapeutic intervention. The laminin gamma 2 chain/laminin B2 polypeptide, truncated/kalinin (LAMC2/LAMB2T/LAMNB2) gene was also a predictor of site (ovary versus omentum) in our predictive model. Laminin 5 is a basement membrane component that actively promotes adhesion and migration of epithelial cells, and laminin-5 gamma2 has previously been implicated in tumor cell migration, metastasis and poor outcome in multiple human cancers including colorectal, pancreatic, and esophageal carcinomas (86-88). Clearly, expression of the LAMC2 gene impacts metastatic potential, and our data suggests that the gene may also play a pivotal role in ovarian cancer metastasis. Four-and-a-half-LIM-only/Heart protein (FHL-2) gene exhibited 0.9-fold lower expression in omental implants compared to primary ovary (p=0.03). FHL-2 was
originally identified as a protein downregulated in rhabdomyosarcoma (89), and is highly expressed in heart and ovary, as well as breast, leukemia, cervical, colon, and lung cancer cell lines (50). FHL2, a p53-inducible protein, is known to interact with beta-catenin, which is involved in both cell-cell adhesion and in transcriptional regulation by the Wingless/Wnt signaling pathway. FHL-2 may function as a coactivator of Wnt-responsive genes, which play a critical role in oncogenesis (51). Recently, it has been shown that FHL2 interacts directly with the breast-ovarian cancer susceptibility gene, BRCA1, and that the BRCA1-FHL2 interaction may be involved in transcriptional regulation and play a significant role in cancer cell growth (52). In light of the interactions of FHL2 with other genes and gene pathways, including BRCA1, TP53, beta-catenin, all of which are implicated in ovarian carcinogenesis, it is perhaps not surprising that the FHL2 gene may play a central role in ovarian cancer metastasis.

Glutaminase (GA) is responsible for catabolic clearance of glutamine, an amino acid essential for the growth of normal and neoplastic cells, and has been previously implicated in cell growth control and proliferation of tumor cells (53,54). Our array analysis revealed that expression of the liver mitochondrial glutaminase (GA) gene was lower in omental metastasis than in primary ovarian cancer. Interestingly, reduction of glutamine levels using glutamine antagonists or bacterial GA has been utilized as a therapeutic strategy in some malignancies, while glutamine supplementation is known to promote a more malignant and less differentiated phenotype in human colon cancer cell lines (55,56). In contrast, Gomez-Fabre, et al., (2000) demonstrated that GA expression in breast cancer cells is dependent on a proliferative state, with highest levels at the beginning of the exponential growth phase (57). Further, they have shown that transfection with antisense constructs of GA decrease cell proliferation in correlation with diminished GA expression. The role of GA expression in malignant cell growth, therefore, appears to be complex. While it appears that the GA gene may be involved in the development and progression of ovarian and other cancers, it is unclear whether this is a direct effect or a consequence of changes in glutamine levels.

Prior to the development of genome-wide expression analysis technology, efforts to delineate the molecular underpinnings of ovarian cancer development and metastasis relied on single-gene strategies. In those studies, the p53 tumor suppressor gene was
the single molecular marker shown most consistently to be associated with ovarian cancers that present with metastatic disease (5, 6, 90,91). Currently, using genome-wide expression analysis, our data provide further evidence for a pivotal role for the p53 pathway in ovarian cancer pathoetiology. We have demonstrated that 10/56 (18%) of genes, differentially expressed between ovary and omentum have gene network relationships to the p53 gene, including *MJD, STAR, NEF3, NR1H4, EVPL, TMPO, PIR51, APOBEC2, CDC2 and MMP1* (33-49). For example, expression of both *APOBEC2* and *MMP1* genes have previously been shown to be directly p53-dependent (47,48). Of interest, NR1H4 binds RXRA, which is the therapeutic target of Bexarotene, a retinoid used to treat cutaneous T-cell lymphoma. Moreover, 10/27 (37%) genes (*LAMC2, APOBEC2, FHL, STAR, PARD6B, ELAVL1, ANGPT1, CDK5, RPS6KB1, and RFXAP*) within the 27-gene model, which accurately predicts site of origin of cancer (ovary versus omentum), have gene network relationships with *p53* (68-79). Expression of *APOBEC2* and *FHL2* genes is p53-dependent (47, 50), and p53 is known to modulate *MMP2* promoter transcription, which in turn interacts with LAMC2 protein (92,93). Such findings reinforce the importance of the p53 pathway in ovarian cancer development and progression, and illustrate how genome-wide analysis can refine our understanding of the complex gene network relationships that underlie previously identified single-gene:phenotype associations.
REFERENCES


178


use of oligonucleotides and tissue arrays identifies cancer/testis antigens as biomarkers in lung carcinoma. Cancer Res, 14: 3971-3979


Chapter 8

Prediction of optimal versus suboptimal cytoreduction of advanced stage serous ovarian cancer using microarrays

Published as:

Note: Contributed to work as a co-investigator only
Abstract

Objective: To define gene expression patterns associated with optimal versus suboptimal debulking of advanced stage serous ovarian cancers.

Study Design: RNA from 44 advanced serous ovarian cancers (19 optimal, 25 suboptimal) was interrogated using Affymetrix U133A microarrays that contain over 22,000 genes. Genes were screened on the basis of their association with debulking status to obtain the top 120 differentially expressed genes. These genes were then used to develop a predictive model for debulking status, which was subjected to out-of-sample cross validation.

Results: We found that patterns of expression of 32 genes can distinguish between optimal and suboptimal debulking with 72.7% predictive accuracy. Analysis of the data based on clusters of coordinately expressed genes resulted in only a marginal improvement in predictive accuracy (75%).

Conclusions: These data support the hypothesis that favorable survival associated with optimal debulking of advanced ovarian cancers is due, at least in part, to the underlying biological characteristics of these cancers
Introduction

Although cure rates for advanced stage ovarian cancer have not changed appreciably in the past 25 years, median survival has improved substantially. This is generally attributed to more complete surgical cytoreduction and the development of new chemotherapy drugs – most notably those of the platinum and taxane families (1). Whereas the superior efficacy of new chemotherapy regimens has been demonstrated in prospective randomized trials, the benefit of surgical debulking has been inferred from retrospective analyses (2,3). Currently, median survival is about 5 years in patients in whom residual tumor nodules are all equal to or less than 1 cm in maximal diameter (optimally debulked) (4), compared to only about 3 years in those with larger tumors (1).

In contrast to its widespread use in advanced ovarian cancer, cytoreductive surgery has a much more limited role in the management of other abdominal malignancies. Resection of metastatic tumor masses may be efficacious in ovarian cancer because disease usually is confined to the abdominal cavity at presentation – despite an extraordinarily large tumor burden. Although optimal debulking has been associated with improved survival in observational studies, it is unclear whether this is directly attributable to removal of the primary tumor and metastases. It is conceivable that outcome may be predetermined by the underlying tumor biology and that optimal debulking is more likely to be technically feasible in cancers that are inherently less virulent and/or more sensitive to chemotherapy.

All human cancers arise due to a series of molecular alterations in diverse pathways involved in regulation of cell proliferation, apoptosis and DNA repair (5). The pattern of alterations varies between cancers, and this molecular heterogeneity is thought to underlie observed differences in clinical phenotypes such as grade, histologic type, stage and survival. In the past, molecular correlative studies typically involved single genes, but microarrays now allow analysis of expression levels of thousands of genes. Arrays have proven useful in predicting clinical phenotypes in several types of solid tumors including breast (6-8) and prostate cancer (9). The goal of the current project was to determine whether patterns of gene expression can be identified that discriminate between optimally and suboptimally debulked advanced stage serous
ovarian cancers. If distinct molecular signatures exist, this would suggest that biological
differences, rather than the extent of cytoreductive efforts, are the primary determinant
of outcome.

**Materials and Methods**

**Patients:** Ovarian cancer tissue was snap frozen at initial surgery prior to any
chemotherapy in 49 women treated at Duke University Medical Center between 1988
and 2001 for serous ovarian cancer. This tissue was collected under the auspices of a
protocol approved by the Duke IRB. There were 44 advanced stage (III/IV) cases and 5
early stage (I/II) cases. These patients had previously been selected as having lived
less than 3 years or greater than 7 years for a study in which microarrays were used to
develop a predictive model for survival. The stage III/IV cases all received platinum-
based combination chemotherapy and were nearly equally divided between those who
survived less than 3 years (n=21) and those who lived more than 7 years (n=23). All of
the patients were under age 75 and none died of causes other than ovarian cancer.

**Microarray analysis:** Frozen tissue samples were embedded in OCT medium and
sections were cut and mounted on slides. The slides were stained with hematoxylin and
eosin to assure that samples included greater than 60% cancer. Approximately 30 mg
of tissue was added to a chilled BioPulverizer H tube (Bio101). Lysis buffer from the
Qiagen Rneasy Mini kit was added and the tissue homogenized for 20 seconds in a
Mini-Beadbeater (Biospec Products). Tubes were spun briefly to pellet the garnet
mixture and reduce foam. The lysate was transferred to a new 1.5 ml tube using a
syringe and 21 gauge needle, followed by passage through the needle 10 times to
shear genomic DNA. Total RNA was extracted using the Qiagen Rneasy Mini kit. Two
extractions were performed for each tumor and the total RNA pooled at the end of the
Rneasy protocol, followed by a precipitation step to reduce volume. Quality of the RNA
was checked by visualization of the 28S:18S ribosomal RNA ratio on a 1% agarose gel.
The targets for Affymetrix DNA microarray analysis were prepared according to the
manufacturer's instructions. Biotin-labeled cRNA, produced by *in vitro* transcription,
was fragmented and hybridized to the Affymetrix U133A GeneChip arrays at 45°C for
16 hr and then washed and stained using the GeneChip Fluidics. The Affymetrix
U133A GeneChip includes 22,283 full length genes as well as a set of human maintenance genes to facilitate the normalization and scaling of array experiments. The arrays were scanned by a GeneArray Scanner and patterns of hybridization detected as light emitted from the fluorescent reporter groups incorporated into the target and hybridized to oligonucleotide probes. Estimates of gene expression were calculated using Affymetrix’s Microarray Suite (MAS) version 5.0 algorithm (http://affymetrix.com/products/algorithms_technote.html).

**Statistics:** The 22,283 genes represented on the U133A array were screened to obtain a list of the top 120 genes associated with debulking status as measured by two-sample t-statistic. A retrospective statistical model was used to quantify the multivariate associations between expression of the 120 genes and debulking status. The posterior predictive distribution implied by the retrospective model was input to Bayes Formula to generate prospective predictions of debulking status given expression data and the prior probability that an individual in the sample has been optimally debulked. The posterior probability that an individual with a given expression profile is optimally debulked implied by this modeling approach is a nonlinear function of expression of a subset of the 120 genes and the prior probability of optimal debulking. The retrospective model has the form $Pr(X|S)$ where $X$ represents the ensemble of 120 gene expression variables and $S$ debulking (S)tatus ($S=1$ indicates optimally debulked, $S=0$ indicates otherwise). This model, input to Bayes formula, is used to predict debulking status, $S$, given a specific gene expression profile, $X=x$, and the prior probability that the individual is debulked $Pr(S=1)$. In particular, the probability the individual was debulked given the profile $x$ was predicted using the expression:

$$Pr(S=1 \mid X=x) = Pr(X=x \mid S=1)* Pr(S=1) / (Pr(X=x \mid S=1)* Pr(S=1) + Pr(X=x \mid S=0)* Pr(S=0))$$

This model, which leads to a classifying function akin to the discriminant function, was used for prediction in two settings. First, it was implemented in a leave-one-out cross validation analysis to develop and validate the model. Second, it was used to predict the debulking status of 5 early stage cases. In both cases, the prior distribution of debulking status, $Pr(S)$, was fixed to correspond to its sample distribution (i.e. $Pr(S=1) = 19/44$). An initial screen yields $n$ gene expression measures. In the present
analysis, n=120 and the expression measures correspond to expression of individual genes. We model the n expression variables, \( X_1 \ldots X_n \) in \( X \), as arising from a normal distribution given debulking status. This is accomplished by factoring the distribution of \( X \) given \( S \) as a product of regressions \( Pr(X_1 | X_2 \ldots X_n, S) \* Pr(X_2 | X_3 \ldots X_n, S) \* \ldots \* Pr(X_n | S) \). To maintain parameter identifiable and model parsimony each of the regressions is constrained to depend on a subset of at most \( P \) of the variables on the right hand side of the conditioning bar (\( | \) ). For each regression model, the subset of \( P \) or fewer predictor variables with the highest posterior probability as measured by Bayes Information Criterion (BIC) (Kass and Raftery, 1995), an approximation to the Bayes factor, is chosen. With \( P=1 \), each regression depends on only one variable, either \( S \) or one of the remaining expression variables. The implied model for \( Pr(X|S) \) in this case therefore involves a product of terms of which only a subset depend on \( S \). It is this latter subset that defines predictions; the terms that do not depend on \( S \) cancel out of the numerator and denominator of the predictive expression given above. In generating predictions, \( Pr(X|S) \) is evaluated as the product of conditional posterior predictive distributions associated with the regression models defined as above and that, in doing so, we average over uncertainty in regression parameter values. The \( P=1 \) retrospective model is one of conditional independence of gene expression measures given survival category. This may oversimplify the underlying biology of the problem. Incrementally more complex models are obtained as \( P \) is allowed to increase. As \( P \) gets larger, the covariance structure among gene expression measures slowly becomes more complex, better reflecting underlying biology. Because of the limited sample size and a large number of potential predictors to include in \( X \), \( P \) can only be increased to a certain point, after which models will be specified as over-parameterized relative to the data set. To avoid this, the predictive accuracy of the retrospective model is evaluated as a function of \( P \) using leave-one-out cross validation. In the present analysis, out-of-sample predictive accuracy increases from \( P=1 \) to 2, then decreases from 2 to 3 and again from 3 to 4. Hence, the predictive model is based on \( P=2 \). The procedure defined differs from multivariate linear discriminant analysis in two major ways. First, it reduces sensitivity of classification predictions to model parameter estimates by using predictive,
not fitted densities. Second, to prevent overfitting, it embeds a variable selection step with an explicit complexity constraint.

Results

The demographic and clinical features of the 44 women with advanced ovarian cancer are shown in Table I. Nineteen were optimally debulked with residual tumor nodules all less than or equal to 1 centimeter and 25 were suboptimally debulked and had one or more nodules greater than 1 cm. Among the optimal patients, only three (stage IIIA/IIIB) presented with metastatic disease less than 1 cm, whereas the others underwent resection of larger deposits of metastatic disease that rendered them optimal. Survival greater than 7 years was noted in 14/19 (74%) optimally debulked patients compared to only 9/25 (36%) suboptimally debulked patients. A non-parametric analysis was used to determine whether there was a difference in survival by debulking status within each survival category. Among short-term survivors who lived less than three years, median survival of 5 optimally debulked patients was 22 months compared to 16 months in 16 suboptimally debulked patients. The 2-sided p-value for a difference between the 2 survival distributions was 0.057 using the Wilcoxon rank sum test with continuity correction. Among long-term survivors, median survival was 104 months among 14 optimally debulked patients compared to 91 months in 9 suboptimally debulked patients (p = 0.925).

Raw expression measures from the microarrays were transformed to the log base-2 scale, then the data from each microarray was normalized to the distribution of the median expression values across chips and analyzed using the two-staged procedure described in the Methods section. In the first stage, the 22,283 genes were screened to obtain a list of the top 120 genes associated with debulking status in the stage III/IV cases as measured by two sample t-statistic (Figure 1A, Table II). These 120 genes all have p-values for association with debulking status of < 0.01 (30 genes have p values < 0.001). Absolute differences in mean expression between optimally and suboptimally debulked cancers in mean expression of genes represented on the arrays ranged from near zero to 4.65 fold (mean = 1.14). Among the top 120 genes there were
that exhibited more than a two-fold difference in expression between optimal and suboptimal cases (Table II).

<table>
<thead>
<tr>
<th>Table I – Demographic and clinical characteristics of women with advanced stage ovarian cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td><strong>Median age</strong></td>
</tr>
<tr>
<td><strong>Stage</strong></td>
</tr>
<tr>
<td>IIIA</td>
</tr>
<tr>
<td>IIIB</td>
</tr>
<tr>
<td>IIIC</td>
</tr>
<tr>
<td>IV</td>
</tr>
<tr>
<td><strong>Histologic grade</strong></td>
</tr>
<tr>
<td>Well differentiated</td>
</tr>
<tr>
<td>Moderately differentiated</td>
</tr>
<tr>
<td>Poorly differentiated</td>
</tr>
<tr>
<td>Gene (Chromosome)</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Homo sapiens double-stranded RNA-specific deaminase (ADAR3), mRNA</td>
</tr>
<tr>
<td>Homo sapiens mitochondrial protein kinase kinase 7 (MAP3K7), mRNA</td>
</tr>
</tbody>
</table>
Homo sapiens most2 protein (most2) mRNA

Consensus includes 5' probes 139376 IDEF+human DNA sequence from clone RP1-18T24 on chromosome 6p11.21-22.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.

Consensus includes gp96.157 IDEF+human DNA sequence from clone FLB4440 PRO25650 mRNA, complete cds.
Homo sapiens cold inducible RNA-binding protein (CRBP), mRNA
Homo sapiens CD44 FLJ11418, human chromosome 16 (16q22.2).
Homo sapiens CD1C antibody, C-polypeptide (CD1c), mRNA
Homo sapiens hypothalamic protein MRC3731 (MRC3731), mRNA
Homo sapiens cullin 4B (CUL4B), mRNA
Consensus includes q5:AV702810/FEA-EST/UG=HE.14579 SET transactivation (myeloid leukemia-associated)

Homo sapiens FGFR1 oncogene partner (FGP), mRNA
Homo sapiens mune leukemia viral (mlv) (mnu) oncogene homolog (Bmi1), mRNA
Consensus includes q5:AL051394/FEA=EST/UG=HS.61698 Myosin X
Consensus includes q5:BF061074/FEA=EST/UG=HS.131954 HSA histone family, member D
Consensus includes q5:B44148/FEA=EST/UG=HS.92943 KIAA0211 protein
Homo sapiens asencom-polypeptide-associated complex alpha-polypeptide (NACA), mRNA
Cluster info: AL606698:XM.114901 1X Homo sapiens CDNA 5' end
Consensus includes q5:AL135948:1/DEF=Homo sapiens mRNA:CDNA
DKFZp761e1812 (from DKFZp761e1812); complete cds.
Consensus includes q5:B223232/FEA=EST/UG=HS.172081 MKK4030 protein
Homo sapiens hypothalamic protein FLJ21919 (FLJ21919), mRNA

Homo sapiens CD44 FLJ21860, human chromosome 16 (16q22.2), similar to HS67491 human Golgi membrane glycoprotein MG160 (GCL1)
Homo sapiens CD44 FLJ21860, human chromosome 16 (16q22.2), similar to HS67491 human Golgi membrane glycoprotein MG160 (GCL1)
Homo sapiens CD44 FLJ21860, human chromosome 16 (16q22.2), similar to HS67491 human Golgi membrane glycoprotein MG160 (GCL1)
Homo sapiens chromosome 1 open reading frame 10 (C1orf10), mRNA

Consensus includes gb:AA002505 (DEF-Homo sapiens, Homo sapiens serotonin 2C receptor (HTR2C), mRNA.

Homo sapiens ARIP1 (atypical receptor 1, yeast homolog B (centrin-like beta) (ACTR1B), mRNA.

Homo sapiens FLJ20391, mRNA

Homo sapiens serologically defined colon cancer antigen 16 (DDCAG16), mRNA.

Homo sapiens host cell receptor 2 (HCTR2), mRNA.

Consensus includes gb:BJ779023/FLJ20391/1P, EST Hs.Hs.11319, mRNA.

Homo sapiens natriuretic peptide receptor B (NPPB), mRNA.

Homo sapiens N-myristoyltransferase (NMT1), mRNA.


Consensus includes gb:SF72814/1P, DEF-Homo sapiens, Homo sapiens, cardiac muscle (ACTC), mRNA.

Homo sapiens, cardiac muscle (ACTC), mRNA: complete cds.


Gene list (UCSC) or Gene bank (gb) number

Genes more highly expressed in suboptimally debulked ovarian cancers

p value
Human vascular-dependent calcium channel beta-1 subunit (CACNB1), mRNA.

Consensus includes g5:AK023823.1/DEP: Homosapiens CDNA FL13770.6/IG: Hs.332530 KIAA1067 protein.
In the second stage of the analysis a formal predictive model for debulking status was built using the subset of the ‘top’ 120 genes that best predict debulking status using the retrospective modeling approach described in the Methods section. Complexity of the model fit at this stage of the analysis is governed by the parameter, \( P \), that constrains the number of conditional correlations among the initial group of 120 genes and between those genes and debulking status. This parameter was varied over a discrete range of values associated with models of simple structure, carrying out leave-one-out cross validation for each value of this parameter. Optimum predictive ability was obtained using this class of models with \( P=2 \). The final predictive model was defined by 32 genes (Genes in Figure 1B correspond to those in bold type in Table II). They are ordered in terms of the magnitude of their association with debulking status as measured by 2 sample t-statistic. Genes above the horizontal line are positively associated with optimal debulking, while those below are negatively associated.
Figure 1 - Expression intensity of genes most highly differentially expressed between optimal and suboptimal ovarian cancers. A) top 120 genes according to t-statistic, B) 32 genes used in predictive model. For each gene, expression intensity is normalized to range from 0 to 1. Tumors in which a gene's expression is high are represented in yellow, those in which it is low are represented in blue with intermediate levels of expression represented in shades of dark yellow through green to light blue. Genes are plotted in rows; expression values for individual tumors are plotted in columns. Genes above the black horizontal line have positive t-statistics while those below have negative t-statistics. Associated p-values decrease as one moves away from the horizontal line to the top and bottom of the plot. Rows are identified by the corresponding Affymetrix probe name. Columns correspond to cancers, with those from optimally debulked patients depicted to the left of the vertical black line and those from suboptimally debulked patients to the right.
Figure 2 plots individual leave-one-out out-of-sample predictions for the 44 stage III/IV cancers by actual debulking status. Using this classification threshold, 5/25 (25%) suboptimally debulked and 7/19 (37%) optimally debulked samples were misclassified. Predictive accuracy is 32/44 (72.7%). There was no relationship between misclassified cases and clinical or pathologic features. Among misclassified optimally debulked cases, 2 were short-term survivors and 5 were long-term survivors, 3 were grade 2 and 4 were grade 3 and 5 were stage IIIC, 1 stage IIIB and 1 stage IV. Among misclassified suboptimally debulked cases, 3 were short-term survivors and 2 were long-term survivors, 1 was grade 2 and 4 were grade 3 and all were stage IIIC.

The predictive model for debulking based on patterns of gene expression was also

![Graph showing individual leave-one-out out-of-sample predictions for 44 stage III/IV cancers by debulking status. The 25 suboptimally debulked patients are plotted in red and the 19 optimally debulked patients are plotted in blue. In addition, out-of-sample predictions for five stage I/II cases are plotted in green. Points are jittered in the direction of the y-axis to achieve their separation. A vertical line is plotted at 1/2 indicating the classification threshold used for determining out-of-sample predictive accuracy (tumors with probability of being optimally debulked greater than 1/2 were declared optimal; those less than 1/2 as suboptimal). Cases that were misclassified are accompanied by their respective numbers.](image-url)
applied to 5 early stage serous ovarian cancers. None of these cancers have recurred and 3 were grade 2 and 2 were grade 3 (2 stage IA, 2 stage IC, 1 stage IIC). These samples were not used in 'model training.' Three of the cases were classified based on gene expression patterns with the optimally debulked advanced stage cancers. The exceptions were cancers 1762 (estimated probability of being optimally debulked near zero) and 1764 (probability 0.25). The 3 early stage tumors that had expression patterns similar to the optimally debulked advanced stage cases were grade 2 while the 2 cases classified by the model as suboptimal were grade 3 (1 stage IA, 1 stage IC).

An alternative to modeling expression of individual genes is to model composite measures of expression. One such option is to identify clusters of related genes and calculate and model cluster-specific expression measures (e.g. cluster averages). To this end, k-means of clustering were used to identify k=1,000 clusters and again to identify k=5,000 clusters among the 22,283 genes. For each clustering, cluster averages were calculated for each sample and subjected to the two-step modeling procedure described in the Methods. In the first step screen, the n clusters most highly correlated with debulking status were identified. With k=1,000, n=100 was used; with k=5,000, n=150 was used. In both cases, the highest out-of-sample predictive accuracy obtained was 0.75 under a model of conditional independence (i.e. \( P=1 \)); larger values of \( P \) lead to worse predictive performance in both cases. With the \( k=1,000 \) clustering \( n=1,000 \) was also use (i.e. the initial screening step was ignored). Computations limited us to the model of conditional independence in which an out of sample predictive accuracy of 0.705 was obtained. These models were rejected in favor of the model involving individual genes because the improvement in predictive accuracy was modest (corresponding to one sample in 44) at best and their interpretation more difficult.
Comment

Most ovarian cancer deaths are attributable to the serous histologic type, which has a propensity to spread extensively throughout the peritoneal cavity prior to diagnosis. Several theoretical lines of reasoning are invoked to support the utility of debulking advanced stage ovarian cancer prior to administration of chemotherapy. First, the ability of cytotoxic drugs to kill cancer cells is thought to be greatest in smaller masses, which generally have a high growth fraction (11). Larger masses are more likely to be necrotic and have hypoxic areas that contain cells that are nondividing or in a resting (G0) phase. Another important factor is that each chemotherapy treatment destroys a constant proportion of cells (11). Therefore, if the number of cells is lower at the initiation of chemotherapy fewer cycles of treatment should be needed to eradicate the cancer. In addition, the number of cells present at the onset of chemotherapy correlates with the likelihood of developing drug resistance during the course of treatment (12). Finally, it has been suggested that there is a less robust immunological response to the cancer in larger masses (13).

Although women whose cancers are optimally debulked live significantly longer on average, survival is poor in some cases – perhaps due to de novo resistance to chemotherapy. Conversely, some patients who are suboptimally debulked have cancers that are exquisitely sensitive to chemotherapy and never relapse after primary treatment. It is unclear whether outcome is determined primarily by the extent of surgical debulking, the underlying biology or whether both play a role. Examination of the molecular alterations present in ovarian cancers has the potential to enhance our understanding of heterogeneous clinical phenotypes such as grade, histologic type, stage, survival and debulking status. In the past, this has involved studies of individual genes involved in regulating proliferation, apoptosis and DNA repair (5). Approximately 10% of epithelial ovarian cancers arise in women who have inherited mutations in cancer susceptibility genes such as BRCA1 or BRCA2, but the vast majority of cases result from the accumulation of genetic damage over the course of a lifetime – and are referred to as sporadic cancers (5). The p53 tumor suppressor gene is the
most frequently mutated gene in serous ovarian cancers, and is associated with advanced stage disease (14,15). While p53 is mutated in about 20% of stage I/II cases, at least 70% of advanced stage serous ovarian cancers harbor p53 mutations (16-17). Although p53 mutation is common in advanced stage cases, it is not strongly associated with adverse outcome or debulking status. In this regard, we performed complete sequencing of the p53 gene in 109 advanced ovarian cancers treated on GOG protocols 114 (optimally debulked) and 132 (suboptimally debulked), and found no relationship between p53 mutational status and survival or debulking status (16). Mutations in p53 were present in 70% of suboptimal cases and 85% of optimal cases.

Microarrays with sequences complementary to thousands of genes have been created that allow global assessment of gene expression. Expression arrays have proven useful in predicting clinical phenotypes in several types of solid tumors. In breast cancer, it has been shown that tumors can be resolved using arrays into subsets based on BRCA mutation status (6) and estrogen receptor status (7). In addition, a panel of 70 genes defined by microarrays could discriminate good and poor-prognosis subsets of early breast cancer (8). In a prospective evaluation of 295 consecutive stage I/II breast cancers, those in the good-prognosis group had a 95% survival, compared to 55% in the poor-prognosis group. Several groups have applied expression array technology to the analysis of ovarian cancers. Many of these studies have compared gene expression between normal ovarian epithelial cells and ovarian cancers. Numerous genes have been identified that appear to be up or down-regulated in the process of malignant transformation (18-20). In addition, microarrays have demonstrated patterns of gene expression that distinguish between histologic types (21) and stages (22). None of the prior array studies in ovarian cancer have compared optimally versus suboptimally debulked ovarian cancers.

In the present study we examined expression levels of over 22,000 genes using Affymetrix arrays. We built a predictive model by focusing on the 120 genes that were most significantly differentially expressed between optimally and suboptimally debulked cancers. Through complex statistical modeling a panel of 32 genes was defined that had a 72.7% predictive accuracy when subjected to
internal cross validation using a one-at-a-time out of sample method. As we continue to examine array data from additional advanced stage serous ovarian cancers, some improvement in the predictive efficiency of expression array-based models may be attained due to increased statistical power. Regardless of whether or not a pattern of gene expression can be identified that reliably distinguishes between optimally and suboptimally debulked cancers, the results of this study clearly demonstrate that there are differences in expression of some genes between the groups. Several of the 120 genes most differentially expressed between optimal and suboptimal cancers have been implicated in malignant transformation. This includes the retinoic acid receptor-beta (RARB) and p53 inducible protein P2X6, (2) which were more highly expressed in suboptimally debulked cases. The retinoic acid receptor-beta (RARB) gene is a tumor suppressor that has been implicated in mediating both the antiproliferative (24) and chemopreventive effects (25) of retinoids in benign and malignant ovarian epithelium. Interaction of retinoids with RAR-beta induces differentiation, which may decrease chemosensitivity (26), and this may contribute to the poor outcome of suboptimal cancers that express high levels of this receptor. Conversely, the fibroblast growth factor receptor 3 (FGFR3)(27) and the fibroblast growth factor receptor1 oncogene partner (FOP)(28) were more highly expressed in optimally debulked cancers.

Perhaps the most intriguing finding is that expression of two members of the mitogen-activated protein kinase family (MAP2K4 and MAP3K7) were higher in optimally debulked cases. The MAP2K4 gene, which was the second most strongly associated with optimal debulking in this study, has already been implicated as a metastasis-suppressor gene in ovarian cancer (29). Levels of MAP2K4 have been shown previously to be downregulated in metastatic ovarian cancers relative to normal ovarian epithelium; and transfection of MAP2K4 into the SKOV3 cell line reduced the development of metastatic deposits in an intraperitoneal mouse model by 90%. It is possible that the association between high levels of MAP2K4 and optimal debulking status is due to these cancers having less ability to metastasize relative to cancers with decreased expression. The mechanism by which MAP2K4 acts as a metastasis-suppressor is not well
understood, but a slower rate of development of metastases or the presence of a smaller number of lesions might well render such cancers more amenable to optimal debulking.

Although the ability to develop a microarray based model with 72.7% accuracy in discriminating between cancers that were optimally or suboptimally debulked suggests that biological differences exist, the model misclassified a significant fraction of cases when subjected to internal cross-validation. This may reflect the fact that debulking status is not preordained in all cases by the underlying tumor biology. The efforts of the surgeon also are a significant determinant of whether optimal debulking is achieved. The finding that the model did not segregate all of the early stage ovarian cancers, which by definition have no metastatic disease, into the optimal group also could be interpreted as suggesting that debulking status is not entirely determined by the underlying biology.

There are several methodological issues to consider in interpreting the validity of the predictive model we have developed. First, several different Gynecologic Oncologists operated on the patients in this study and there may have been some differences in the extent to which various surgeons attempted to achieve optimal cytoreduction. In addition, the patients in this study do not represent a random sample of all those with stage III/IV serous ovarian cancer. They had either survived less than 3 years or greater than 7 years and more of the optimally debulked patients were in the long-term survival group, as would be anticipated. Furthermore, all of the patients were under age 75 and none died of causes other than ovarian cancer. Exclusion of the elderly and infirm, who are less likely to be subjected to vigorous attempts at maximal cytoreduction, was designed to avoid a confounding factor that might obscure efforts to identify patterns of gene expression associated with debulking status. However, this potentially limits the utility of the resulting predictive model to individuals with good performance status. Finally, three of the optimally debulked patients (stage IIIA/IIIB) had metastatic disease less than 1 cm at presentation and one was well differentiated, whereas all of the suboptimal cases were stage IIIC or IV and moderate or poorly differentiated. Ideally, these clinicopathologic features would
be equally distributed in the two groups, but these differences in the distribution of prognostic factors are characteristic of optimal versus suboptimal cases.

Another consideration in developing an array based model that predicts debulking status is that the fraction of patients with advanced stage disease who are optimally debulked varies considerably between institutions. We have previously reported that 27% of advanced ovarian cancers at Duke University were optimally debulked (30). This is relatively representative of other institutions in the United States, in which optimal debulking rates in most series have ranged from 17-40% (2,31-37). The model we have developed might have a similar accuracy in predicting debulking status in settings in which a similar fraction of patients are optimally debulked, but some studies have reported optimal debulking rates as high as 87% (38). This may be related to differences between patient populations, but also likely is attributable to differing practice patterns between surgeons at various institutions. The technical feasibility of resecting portions of the bowel, urinary tract, diaphragm, enlarged lymph nodes and splenectomy has been demonstrated, but some surgeons are more likely to perform these procedures than others (39-44). In some studies with the highest optimal debulking rates median survival of the optimal group has been noted to be much inferior to that seen in studies with lower rates of optimal debulking (38). This has been interpreted as suggesting that outcome is predetermined by the underlying biology more than by the extent of debulking. It is possible, however, that moderate debulking efforts would improve outcome, even if more vigorous efforts afford little additional benefit.

Even after accounting for differences between institutions, the degree of surgical effort required to achieve surgical debulking is an issue in developing predictive models. Some patients are optimal simply because minimal metastatic disease was found, while others may require extensive resections to render them optimal. In addition, optimal status encompasses patients with a single small nodule as well as those with hundreds of nodules less than one centimeter in diameter. Thus, debulking status does not always correlate well with either initial or residual tumor volume. Perhaps the ideal groups in which to develop array-
based predictive models would be suboptimally debulked patients versus those who had large amounts of cancer initially and were then optimally debulked.

In conclusion, using 44 advanced serous ovarian cancers, we have begun to define patterns of gene expression that distinguish between cancers that are optimally versus suboptimally debulked. This initial study is supportive of the hypothesis that there are biological differences between these two groups of cancers. Currently, essentially all healthy patients suspected of having advanced ovarian cancer undergo surgical exploration with the intent of performing debulking. It is conceivable that array analysis of a pre-operative biopsy sample could be used in the future to predict the likelihood of achieving optimal cytoreduction. This could facilitate more rational selection of patients for debulking surgery, whereas others less likely to benefit from this approach might undergo less extensive surgical debulking or neo-adjuvant chemotherapy.
REFERENCES


Chapter 9

Prediction of response to platinum-based chemotherapy for advanced-stage ovarian cancers using gene expression profiles

Submitted for Publication as:

ABSTRACT
Approximately 30% of patients with advanced (stage III/IV) epithelial ovarian cancer will fail platinum-based therapy. Current empiric-based treatment strategies result in many patients receiving multiple cycles of toxic platinum-based therapy without success, prior to initiation of therapy with active agents. In an effort to individualize therapy for patients with advanced stage ovarian cancers, we sought to develop gene expression profiles that predict response to platinum-based chemotherapy using microarray expression analysis. Predictive models made use of a classification tree methodology that samples multiple gene expression patterns (metagenes) for those best able to distinguish those patients likely to respond to therapy. Predictive models achieved an accuracy of 93% in leave-one-out cross-validation studies. We conclude that gene expression profiles can accurately predict the clinical response to platinum-based chemotherapy for advanced stage serous ovarian cancers, offering the potential to guide therapeutic decision-making for patients.
INTRODUCTION

Ovarian cancer is the fifth leading cause of cancer death among women in the United States and Western Europe and has the highest mortality rate of all gynecologic cancers. Currently, platinum drugs are the most active agents in epithelial ovarian cancer therapy (1-3). Consequently, the standard treatment protocol used in the initial management of advanced-stage ovarian cancer is primary cytoreductive surgery, or “debulking”, followed by primary chemotherapy with a platinum-based regimen that usually includes a taxane. Approximately 70% of patients will have a complete clinical response to this initial therapy, with absence of clinically detectable residual disease on clinical exam, radiologic imaging, or serum CA125 tumor marker (5,6). The remaining 30% of patients will demonstrate residual or progressive disease at conclusion of primary therapy with platinum and taxane. Currently, empiric-based treatment strategies result in these patients receiving multiple cycles of platinum-based therapy without a complete response. This approach results in a significant delay in initiation of alternate therapies or enrolment in clinical trials of novel agents.

The clinical heterogeneity of ovarian cancer, resulting from the acquisition of multiple genetic alterations that contribute to the development of the cancer, likely underlies the clinical heterogeneity of response to chemotherapy (7). Although a variety of gene alterations have been identified, no single gene marker can reliably predict response to therapy and outcome (8-12). Recent advances in the use of DNA microarrays that allow global assessment of gene expression in a single sample have shown that expression profiles can provide molecular phenotyping that identifies distinct classifications not evident by traditional histopathological methods, including ovarian cancer (13-20). We have now applied these methodologies to identify characteristics of primary ovarian cancers that predict therapeutic response. Further, understanding the biologic basis of chemo-respons will enhance our ability to develop agents that may improve chemo-sensitivity, and hence response rates.

MATERIAL AND METHODS

Tissues. We measured expression of 22,283 genes in 55 advanced (FIGO stage III/IV) serous epithelial ovarian adenocarcinomas using Affymetrix U133A GeneChips. All ovarian cancers were obtained at initial cytoreductive surgery from patients treated at
Duke University Medical Center who then received primary platinum-based chemotherapy. Thirty-three patients were considered "complete responders" to platinum therapy and 22 were considered "incomplete responders" by criteria outlined below. All tissues were collected under the auspices of a Duke IRB approved protocol with written informed consent.

**Platinum response.** Response to platinum at the completion of primary therapy was assessed from the medical record (57). Platinum response was considered complete if there was no evidence of residual disease on exam or abdominal/pelvic CT scan and the serum CA 125 level was < 20 U/ml. Those who demonstrated progression of disease during platinum therapy or had evidence of persistent disease by exam, CT scan or CA 125 (> 20 U/ml) at the completion of therapy were considered incomplete responders (58-59).

**Microarray analysis.** MIAME (minimal information about a microarray experiment)-compliant information regarding the analyses performed here, as defined in the guidelines established by MGED (www.mged.org), is detailed in the following sections. Frozen tissue samples were embedded in OCT medium and sections were cut and mounted on slides. The slides were stained with hematoxylin and eosin to assure that samples included greater than 70% cancer. Approximately 30 mg of tissue was added to a chilled BioPulverizer H tube (Bio101). Lysis buffer from the Qiagen RNeasy Mini kit was added and the tissue homogenized for 20 seconds in a Mini-Beadbeater (Biospec Products). Tubes were spun briefly to pellet the garnet mixture and reduce foam. The lysate was transferred to a new 1.5 ml tube using a syringe and 21 gauge needle, followed by passage through the needle 10 times to shear genomic DNA. Total RNA was extracted using the Qiagen RNeasy Mini kit. Two extractions were performed for each cancer and the total RNA pooled at the end of the Rneasy protocol, followed by a precipitation step to reduce volume. Quality of the RNA was measured using an Agilent 2100 Bioanalyzer to determine the 28S:18S ribosomal RNA ratio. The targets for Affymetrix DNA microarray analysis were prepared according to the manufacturer's instructions. Biotin-labeled cRNA, produced by in vitro transcription, was fragmented and hybridized to the Affymetrix U133A GeneChip arrays (www.affymetrix.com_products_arrays_specific_Hu133A.affx) at 45°C for 16 hr and then washed and stained using the GeneChip Fluidics. The arrays were scanned by a GeneArray Scanner and patterns of hybridization detected as light emitted
from the fluorescent reporter groups incorporated into the target and hybridized to oligonucleotide probes.

**Statistical analysis:** Expression was calculated using the robust multi-array average (RMA) algorithm (60) implemented in the Bioconductor (http://www.bioconductor.org) extensions to the R statistical programming environment (61). RMA generates log-2 scaled measures of expression using a linear model robustly fit to background-corrected and quantile-normalized probe-level expression data and has been shown to have a better ability to detect differential expression in spike-in experiments (62). The CEL files for each sample used in the analysis are found on the web site, http://data.cgt.duke.edu/platinum.php. The 22,283 probe sets were screened to remove 68 control genes, those with a small variance and those expressed at low levels. The core methodology uses statistical classification and prediction tree models, and the gene expression data (RMA values) enter into these models in the form of metagenes. As previously described 21,22, metagenes represent the aggregate patterns of variation of subsets of potentially related genes. In this example, metagenes are constructed as the first principal components (singular factors) of clusters of genes created by using k-means clustering. Bayesian methods of analysis are used to fit multiple candidate classification tree models, each candidate model based on varying the selection of predictor variables, and trees were individually generated by using a forward selection process. Predictions are based on weighted averages across multiple candidate tree models, and the combinations of genomic and clinical predictor variables appearing in highly weighted tree models provide insights on the interactions of risk factors determining the predictions. Full details of the statistical approach, including creation of metagenes, are described previously (21,22).

**RESULTS**

Measurement of clinical response: We evaluated 55 advanced serous epithelial ovarian carcinomas obtained from patients treated at Duke University Medical Center between 1988 and 2000. Response to therapy was evaluated from the medical record and patients were classified as either “complete responders” or “incomplete responders”, by criteria described above. A summary of the clinico-pathologic and chemo-response data is provided in Table 1. From the group of 55 patients analyzed, 33 were classified as “complete responders” and 22 as “incomplete responders".
Table 1: A summary of clinico-pathologic and response data for 55 patients and ovarian cancer samples studied.

<table>
<thead>
<tr>
<th></th>
<th>Platinum complete responders</th>
<th>Platinum incomplete responders</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>33</td>
<td>22</td>
</tr>
<tr>
<td>Mean age (Yrs)</td>
<td>60</td>
<td>54</td>
</tr>
<tr>
<td>Stage (n)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>30</td>
<td>19</td>
</tr>
<tr>
<td>IV</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Grade (n)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>22</td>
<td>10</td>
</tr>
<tr>
<td>III</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Surgical Debulking (n)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimally (&lt;1cm)</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>Suboptimal (&gt;1cm)</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>Chemotherapy (n)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platinum/Cytoxan</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Platinum/Taxol</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>Single Agent Platinum</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mean Serum CA125 (u/ml)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-platinum</td>
<td>2514</td>
<td>1480</td>
</tr>
<tr>
<td>Post-platinum</td>
<td>20</td>
<td>575</td>
</tr>
<tr>
<td>Mean Survival (Mnths)</td>
<td>85+</td>
<td>22</td>
</tr>
</tbody>
</table>
Statistical tree models using multiple metagenes to classify platinum response: Our recent work in breast cancer has described the development of predictive models that make use of multiple forms of genomic and clinical data to achieve more accurate predictions of individual risk of recurrence of disease (21,22). The logic is that no one single gene expression pattern can effectively dissect the population cleanly into responder or non-responder; as in other examples of clinical phenotypes, there is simply too much heterogeneity for one single marker, whether it is a protein assay, a gene mutation, or a gene expression pattern, to resolve the heterogeneity. But, by making use of additional gene expression patterns to continue resolving the heterogeneity at each split, it becomes possible to achieve a quite accurate segregation of the population of patients.

The method for selecting multiple gene expression patterns, that we term metagenes, makes use of Bayesian-based classification and regression tree analysis. Metagenes are derived from a clustering of the original gene expression data in which genes with similar expression patterns are grouped together. The expression data from the genes in each cluster are then summarized as the first principal component of the expression data, i.e., the metagene for the cluster. The metagenes are sampled by the classification trees to generate partitions of the samples into more and more homogeneous subgroups that in this case reflect the response to platinum-based therapy. At each node of a tree, the subset of patients is divided in two based on a threshold value of a chosen metagene. An illustration of the use of multiple metagenes to dissect the group of ovarian cancer patients based on response to platinum therapy is shown in Figure 1. This figure depicts an example of one of many trees that illustrates the capacity of a single pattern (metagene) to separate the patients into platinum “complete responder” or “incomplete responder” groups. In this example tree, a low value on metagene 326 (Mg326) largely defines complete responders (16 complete responders, 1 incomplete responder) while the most of the platinum incomplete responders exhibit a high value on Mg326
(21 incomplete responders, 16 complete responders). Thus, this initial split using a single gene expression pattern does distinguish the two groups, but it is clear that heterogeneity remains, particularly within the group with a high value on Mg326.

![Figure 1](image)

**Figure 1.** Use of successive metagene analysis to improve predictions of response to platinum therapy. Predictive accuracy is assessed by leave-one-out cross-validation in which the analysis is repeatedly performed – one ovarian cancer sample is removed at each reanalysis and the response to therapy is predicted for that one case. Thus in this tree example, 54 samples are illustrated. Gene expression patterns are shown as standard intensity images that relate to splits in the patient population based on selected metagenes. The top image shows the expression pattern of the genes in Mg326 (ordered vertically by their weight within the metagene) on the entire group of 55 patients. Samples are ordered (horizontally) by the value of Mg326 and the vertical black line indicates the split of patients into two subgroups based on optimal classification of response. The subgroup of platinum incomplete response defined by this initial split is then further split with Mg337. A third metagene (Mg134) further resolved heterogeneity.

This is a result seen in many such studies where gene expression profiles have been applied to the analysis of cancer phenotypes and simply reflects the fact that no one single gene expression profile can capture all of the heterogeneity within a group of cancers. To resolve the heterogeneity we look for additional gene expression patterns that can further dissect the initial splits. As shown in Figure 1, the initial split made on Mg326 that identified the largely platinum incomplete responder population (high value on Mg326) was further
split by a second metagene (Mg337) that could identify two subgroups that resolved much of the heterogeneity. A third metagene (Mg134) further resolved heterogeneity.

The value in the use of multiple metagenes to distinguish the platinum complete and incomplete responders is further illustrated by a scatterplot (Figure 2) of the collection of cancers against two of the most highly weighted metagenes (Mg326 and Mg337). As seen in this analysis, Mg326 achieves a separation of most of the platinum incomplete responders (blue symbols) but leaves many of the complete responders (red symbols) mixed with the incomplete responders. The second metagene (Mg337) then effectively resolves most of this mixture. Cancers with a high value on both metagenes (upper right compartment) are primarily platinum incomplete responders whereas the complete responders separate into the other two compartments.

Figure 2. Scatterplot of the 55 ovarian samples according to expression on the two highly weighted metagenes. The values for each sample on the indicated metagene is plotted as the sample index metagene number with complete responders tumors indicated in red and incomplete responders tumors in blue. The first split in the tree shown in Figure 1 is represented on the Y axis (Mg326) and the dotted line represents the threshold value that splits the complete responders and incomplete responders. The X axis (Mg337) represents the second split in the tree that further splits the heterogeneity. The vertical dotted line is drawn at that threshold, further resolving complete from incomplete responders.
Thus, while neither metagene alone is effective in separating the population into platinum incomplete and complete responders, a combination of the two does effectively resolve the heterogeneity seen within samples.

**Predicting response to therapy based on tree model summaries**

The major impetus for this study is the development of a predictive tool that accurately identifies the platinum incomplete responders at the time of initial diagnosis, and that could be applied to the management of these patients. Predictive accuracy is assessed by leave-one-out cross-validation in which the analysis is repeatedly performed – one ovarian cancer sample is removed at

**Figure 3.** Platinum response predictions from the metagene tree model. Estimates and approximate 95% confidence intervals for platinum response probabilities for each patient. Each patient is predicted in an out-of-sample cross validation based on a model completely regenerated from the data of the remaining patients. Patients indicated in red are those that had are platinum complete responders and those in blue are incomplete responders. The interval estimates for a few cases that stand out are wide, representing uncertainty due to disparities among predictions coming from individual tree models that are combined in the overall prediction.
each reanalysis and the response to therapy is predicted for that one case. Importantly, the entire model building process, including the selection of metagenes and their combination in sets of trees to be weighted by the data analysis, is repeated for each prediction. At any given node of a tree model, there may be several metagenes defining significant subgroups, so it is important to consider multiple tree models. A resulting set of tree models is evaluated statistically by computing the implied value of the statistical likelihood function for each tree; the set of likelihood values are then converted to tree probabilities by summing and normalizing with respect to all selected trees. Predictions are based on all trees in combination, via weighted averages of predictions from individual trees with the tree probabilities acting as weights. This “model averaging” is well known to generally improve prediction accuracy relative to choosing one “best” model $^{23,24}$, especially when several or many models fit the data comparably. In exploring and evaluating trees, several hundreds are generated and weighted; very low probability trees are discarded and the remaining are summarized and averaged to compute resulting predictions.

Figure 3 displays a predictive summary for the samples of ovarian cancers. The predicted probability of response is plotted for each patient along with the statistical uncertainty in the prediction. The latter derives from the uncertainties evident across the array of candidate trees generated in the analysis.

An examination of the estimated receiver operator characteristic (ROC) curves for response indicates a capacity to achieve up to 97% sensitivity with 86% specificity in predicting platinum complete responders (Supplementary Figure 1). These numbers serve to indicate a high degree of accuracy. Importantly, there was only one complete responder (case #36) that was inaccurately predicted as an incomplete responder. This is obviously the major concern in such analyses, since this individual clearly did benefit from the primary platinum based therapy.
Genes that predict resistance to platinum-based therapy include activities directly involved in carboplatin action

The top metagenes providing the capacity to predict platinum response include a total of 64 genes represented by 72 Affymetrix probe sets (Table 2). Two of these genes have been directly linked to the response to platinum-based therapy. Decorin is a component of metagene 326, the most highly weighted in the predictive models. Expression of this matrix proteoglycan is reduced in ovarian cancers and decorin acts synergistically with carboplatin against ovarian cancer cell lines in vitro (25). Additionally, FADD, a component of metagene 66, has been shown to be induced by carboplatin and associated with FAS-dependent apoptosis in response to carboplatin (26). It would thus appear that the capacity to predict the response to platinum-based therapy in ovarian cancer patients may reflect a variation in function of genes that are directly involved in the response to platinum. Further analysis indicates that many of the genes that provide prediction of response to platinum are components of a c-Myc network, including the FADD gene (Supplementary figure 2).

**Supplementary Figure 1.** Receiver Operating Characteristic (ROC) curve depicting the accuracy of the prediction of response to platinum-based therapy. This is a plot of the true positive rate against the false positive rate for varying cut-points of predicting response to platinum-based therapy. The curve is represented by the blue line; the closer the curve follows the left axis followed by the top border of the ROC space, the more accurate the assay. The red circle corresponds to sensitivity and specificity when a probability of 0.5 is used to determine prediction of complete responders and incomplete responders based on genomic profile predictions used in Figure 3.

Thus the response indicates a capacity to achieve up to 97% sensitivity with 86% specificity in predicting platinum complete responders. False positive rate (1 - specificity) is represented on the X axis, and the True positive rate (sensitivity) is represented on the Y axis.
Supplementary Figure 2. A c-myc network relating platinum-response genes.

To characterize the gene network relationships underlying the gene expression patterns identified, we utilized the web-based Ingenuity Pathways Analysis software (http://www.ingenuity.com) that uses a curated knowledge base to identify functional relationships. The top-scoring gene network/pathway from the genes involved in platinum response prediction is displayed and genes previously implicated in the chemotherapeutic response of platinum are highlighted in red. B, Binding; A, Activation/deactivation; E, Expression; I, Inhibition; L, Proteolysis; M, Biochemical Modification; O, Other; P, Phosphorylation/dephosphorylation; T, Transcription.
<table>
<thead>
<tr>
<th>GenBank Symbol</th>
<th>Gene Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN-023942</td>
<td>MGC3036</td>
<td>MYC network</td>
</tr>
<tr>
<td>8G144295</td>
<td>MKP1</td>
<td>MKP1 caspase inhibitor</td>
</tr>
<tr>
<td>4A087464</td>
<td>LOC149603</td>
<td>MKP1 caspase inhibitor</td>
</tr>
<tr>
<td>AK021881</td>
<td>HIF3A</td>
<td>Hypoxia-inducible factor 3</td>
</tr>
<tr>
<td>AF042263</td>
<td>FXR2</td>
<td>Farnesoid X receptor</td>
</tr>
<tr>
<td>NM_001458</td>
<td>FLNC</td>
<td>Filamin C, gamma (actin binding protein 280)</td>
</tr>
<tr>
<td>AV189015</td>
<td>EFNA3</td>
<td>Ephrin-A3</td>
</tr>
<tr>
<td>NM_001414</td>
<td>DOK1</td>
<td>dual-specificity phosphatase</td>
</tr>
<tr>
<td>NM_001418</td>
<td>CST3</td>
<td>Caspase 3</td>
</tr>
<tr>
<td>AV101546</td>
<td>CLDN9</td>
<td>Claudin 9</td>
</tr>
<tr>
<td>NM_014054</td>
<td>COX7A1</td>
<td>Chromosome 20 open reading frame 40</td>
</tr>
<tr>
<td>NM_003824</td>
<td>FADD</td>
<td>Fas (TNFRSF6)-associated via death domain</td>
</tr>
<tr>
<td>NM_003823</td>
<td>MLN1</td>
<td>Metastatin S100 protein</td>
</tr>
<tr>
<td>NM_010722</td>
<td>FOSL1</td>
<td>Fos-like antigen 1 (drosophila)</td>
</tr>
<tr>
<td>NM_014042</td>
<td>DFFP54M02</td>
<td>DFFP54M02 protein</td>
</tr>
<tr>
<td>NM_021785</td>
<td>RAI2</td>
<td>Retinoic acid induced 2</td>
</tr>
</tbody>
</table>

Table 2: Metagene and genes that comprise the platinum-response predictive model.
<table>
<thead>
<tr>
<th>Regulation of transcription</th>
<th>LIM domain binding</th>
<th>Myc network</th>
<th>Myc network</th>
<th>Myc network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adhesion molecule 1</td>
<td>Similar to alphodentin</td>
<td>217</td>
<td>154</td>
<td>63</td>
</tr>
<tr>
<td>AF021774</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TpSB2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TpSB2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TpSB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TpSB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tP21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tP21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDM1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDK6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRKAR1B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRKAR1B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZNF356</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZNF356</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENAG6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFP110</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMT1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMT2A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myc network</td>
<td>Function</td>
<td>Gene ID</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_000532</td>
<td>Fatty acid metabolism</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_007713</td>
<td>RNA polymerase II transcription factor A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_004280</td>
<td>EIF1A translation elongation factor 1 epsilon 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_001920</td>
<td>Decamovirus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_017874</td>
<td>RPRK-like 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_003228</td>
<td>TBP26L2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AB014533</td>
<td>Synovial sarcoma transcription gene on chromosome 8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_20062</td>
<td>STCL24A regulator</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_201342</td>
<td>SMZB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AV02483</td>
<td>Ribosomal protein S21</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_20026</td>
<td>Oxysteryl binding protein-2-like 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_201702</td>
<td>Nup210</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_00024</td>
<td>DNA replication and base excision</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_022105</td>
<td>Death associated transcription factor 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AL096282</td>
<td>Chromosome 20 open reading frame 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AB012919</td>
<td>Chromosome 20 open reading frame 14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NM_01786</td>
<td>Chromosome 20 open reading frame 11</td>
<td></td>
</tr>
<tr>
<td>Protein ID</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIAA0679</td>
<td>G protein-coupled receptor 64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIAA0657</td>
<td>C-terminal binding protein 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIAA0657</td>
<td>C-terminal binding protein 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>669985</td>
<td>GPR64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00756</td>
<td>C-terminal binding protein 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>001328</td>
<td>C-terminal binding protein 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3379</td>
<td>B3F37195</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>337830</td>
<td>AV495300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>018256</td>
<td>WRD12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>03810</td>
<td>TXNDC5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>014873</td>
<td>THEM2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>003191</td>
<td>TAR5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>006345</td>
<td>SR1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>016620</td>
<td>SR1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>034003</td>
<td>SLCO1A4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>089330</td>
<td>SLCO1A4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>006398</td>
<td>RNASEP1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>006648</td>
<td>PRP4B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>006117</td>
<td>PECI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development</td>
<td>NM_005663</td>
<td>WHSC2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>-------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>AF-101434</td>
<td></td>
<td>WHSC2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transport</td>
<td>L1_1669</td>
<td>TETRAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYC network</td>
<td>NM_002899</td>
<td>RBP1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Development
- MYC network
- AF-101434
- Transport
- L1_1669
- NM_002899
- NM_005663
- WHSC2
- WHSC2
- TETRAN
- RBP1
DISCUSSION

Up to 30% of patients with advanced stage epithelial ovarian cancer fail to demonstrate a complete response to primary platinum-based therapy, and often remain on chemotherapy for much of the remainder of their lives. As such, many of the challenges that women with ovarian cancer face are related to the chemotherapeutics they receive. Current empiric-based treatment strategies result in patients with chemo-resistant disease receiving multiple cycles of toxic therapy without success, prior to initiation of therapy with other potentially more active agents, or enrolment in clinical trials of new therapies. Throughout treatment for ovarian cancer, prolongation of survival and the successful maintenance of quality of life remain important goals, and improving our ability to manage the disease by optimizing the use of existing drugs and/or developing new agents is essential. In view of this, it is important that the choice of chemotherapy be individualized to each patient to reduce the incidence and severity of toxicities that could not only potentially limit quality of life, but also the ability to tolerate further therapy. To this end, individualizing treatments by identifying patients who are most likely to respond to specific agents, will not only increase response rates to those agents, but also limit toxicity and therefore improve quality of life for patients with non-responsive disease.

Using genome-wide measures of expression and novel Bayesian statistical computational techniques, we have developed a predictive tool for response to platinum-based chemotherapy for ovarian cancer. Out-of-sample validation of the predictor demonstrates 93% predictive accuracy overall. Moreover, the accuracy for predicting patients with disease that will not respond completely to platinum is 97% with only one patient misclassified. Accurate identification of patients likely not to completely respond to current front-line, platinum-based therapy is critical to successful clinical application of predictive profiles. Review of the medical record for the patient that was misclassified by the gene-predictor failed to reveal any obvious phenotypic difference that would explain the misclassification.

Currently, patients may receive multiple cycles of platinum-based primary therapy before it becomes clear that they are not completely responding. These patients may experience detriment to bone marrow reserve, quality of life and a delay in
timely initiation of alternate therapies. Patients clinically determined to have ovarian cancer that does not completely respond to platinum therapy can be treated with second-line (salvage) agents, which include topotecan, liposomal doxorubicin, gemcitabine, cyclophosphamide and oral etoposide, or enrolled in clinical trials. A genomic-based predictor can identify patients likely not to completely respond to primary platinum-based therapy earlier in their treatment, such that they can be initiated on second-line therapy sooner, or enrolled in clinical trials of new agents. In this way, it is anticipated that expression-profile directed therapy may lead to a significant improvement in the care of patients with advanced stage epithelial ovarian cancer in the future. Moreover, understanding the biologic underpinnings of response to therapy will facilitate the development of novel agents that may be active against platinum-resistant disease and thus improve survival rates for patients with ovarian cancer.

Sixty four genes comprise the platinum-response predictive profile, and include several that have been previously implicated in the chemotherapeutic response of platinum. One such gene, FAS-associated via death domain (FADD), is a universal apoptosis adapter protein that mediates signaling of death domain-containing members of the TNF receptor superfamily (26). Expression of other members of the TNF family (e.g. TRAIL) have been associated with survival from advanced stage epithelial ovarian cancer, and this survival difference may be a reflection of underlying platinum sensitivity (27). FADD is a member of the CD95 signaling pathway, which is known to be activated by cytotoxic drugs via up-regulation of the receptor and/or ligand; an essential factor in the early phase of drug-induced cell death (28,29). Further, FADD has been shown to be induced by carboplatin and to be associated with FAS-dependent apoptosis in response to carboplatin 30.

Another member of the platinum-response predictive profile is Decorin (DCN), a connective tissue proteoglycan, which has been shown to suppress metastasis (31) and growth of cancer cell lines via EGFR, TGF/β MAP kinase, and p21 mechanisms (32-34). Decorin has been shown to inhibit growth of ovarian cancer cell lines, SKOV3 and 2774 and to synergize with carboplatin to inhibit the growth of ovarian cancer cells (25).

Mesothelin (MSLN), a component of metagene 66, is a cell surface molecule that is expressed in the mesothelial lining of the body cavities, and also in many cancer
cells. CA125 is a tumor antigen that is routinely used for diagnosis of ovarian cancer and to monitor the recurrence after therapy (35). It has recently been shown that mesothelin is a novel CA125 binding protein and thus might contribute to the metastasis of ovarian cancer to the peritoneum by initiating cell attachment to the mesothelial epithelium via binding to the mesothelin (36).

Molecular determinants of the activity of platinum compounds have been examined from the publicly available database of the National Cancer Institute, which contains gene expression profiles of the 60 cell lines for which drug cytotoxicity patterns already exist (37). In this study, 19 genes from the list were shown to significantly correlate to the cytotoxicity of carbo- and cisplatin. None of these genes overlapped with our predictive gene lists, which may reflect the characteristics only found in cell lines and not in primary cancers.

The response to chemotherapy has been linked to apoptosis, although the precise mechanisms of this response have not been defined. It has been shown previously that regardless of p53 gene status, apoptosis increased after exposure to platin-based therapy, suggesting that the apoptosis induced by platinum occurred through the p53-dependent and -independent pathways (38). One such independent pathway may include c-myc. c-myc expression has recently been demonstrated to have a strong correlation with cisplatin-chemo-response for ovarian cancer (39). Our data demonstrate that 9 of the platinum-response predictor genes (FADD, CTBP1, DCN, LIG1, RNF110, RBP1, C20ORF14, LDB1, MSLN) have c-myc network relationships (Supplementary Figure 2) (40-54). For example, it has been shown that c-myc activates the CD95/Fas-FADD-mediated death signal (55). c-myc has previously been described as a single molecular marker implicated in the development of a subset of ovarian cancers, and as a predictor of chemotherapy response and a prognostic factor in patients with the disease (56). Our current findings reinforce the importance of the c-myc pathway in ovarian cancer response to therapy, and illustrate how genome-wide analysis can refine our understanding of the complex gene network relationships that underlie previously identified single-gene:phenotype associations.

The ability to predict response to therapy for human cancer will enable tailored therapies to be established for individual patients on the basis of cancer expression
profiles. As such, response rates can be improved, non-active toxic agents avoided, bone marrow spared, and quality of life enhanced. Ultimately, defining the biologic underpinnings of response to therapy will facilitate the development of more active agents that may improve survival for women with ovarian cancer.

**Acknowledgements**

The authors would like to acknowledge The University of Alabama ovarian cancer Specialized Program of Research Excellence (SPORE) and the Ovarian Cancer Research Fund, as sources of funding to support this research.
REFERENCES


40. Miyake, S, et al. (1993): Fission yeast genes nda1+ and nda4+, mutations of which lead to S-phase block, chromatin alteration and Ca2+ suppression, are members of the CDC46/MCM2 family. Mol Biol Cell, 4:1003-1015.


Chapter 10

No Association Between Ovarian Cancer Risk and Progesterone Receptor Gene Polymorphism (PROGINS) in Population-Based, Case-Control Study in North America

Published as:

**JM Lancaster, B Calingaert, R Wenham, JR Marks, PG Moorman, S Halabi, R Bentley, A Berchuck, JM Schildkraut.** No Association Between Ovarian Cancer Risk and Progesterone Receptor Gene Polymorphism (PROGINS) in Population-Based, Case-Control Study in North America, in Vol 12(3) p226-7, *Cancer Epidemiology, Biomarkers & Prevention, 2003*
ABSRACT

Objective: The protective effect of pregnancy and oral contraceptive use on ovarian cancer risk may be attributable to the action of progestins on the ovarian epithelium, and it is unclear whether a polymorphism, named PROGINS, in the progesterone receptor gene, is associated with increased risk of ovarian cancer. We investigated the hypothesis that the PROGINS allele is associated with increased ovarian cancer risk in the North American population.

Method: Three hundred and nine ovarian cancer cases and 397 controls were enrolled through a population-based, case-control study in a 48-county region in North Carolina. Demographic, epidemiologic and medical information was obtained by in-person, trained nurse interview using a standardized questionnaire. Leukocyte DNA was extracted and subject to PROGINS allelotyping using a PCR-based assay. Odds ratios and 95% confidence intervals were calculated using unconditional logistic regression, and crude and adjusted odds ratios reported from a multivariate logistic regression model.

Results: Overall, 29.1% of cases and 28.2% of controls had the PROGINS allele. Adjusting the odds ratios for age, race, and menopause did not lead to significant changes in odds ratios. Although some increased risk associated with the PROGINS allele was seen among oral contraceptive non-users, no significant association was found between the PROGINS allele and ovarian cancer risk among any other subgroup.

Conclusions: We find no evidence to support an association between the PROGINS allele and increased risk of ovarian cancer, in the North American population.
INTRODUCTION

Although the exact epidemiologic and molecular underpinnings of ovarian cancer have yet to be fully elucidated, pregnancy and oral contraceptive use are known to reduce risk, and it has been postulated that this protective effect may be attributable to the action of progestins on the ovarian epithelium (1,2). Germline polymorphisms in steroid receptor genes can alter ligand/receptor interactions and modify such protective effects. Recently, a TaqI restriction fragment length polymorphism in intron G of the human Progesterone Receptor (hPR) gene was shown to be associated with ovarian cancer in a pooled German/Irish population (3). This polymorphism, named PROGINS, is due to an Alu sequence insertion within the ligand-binding domain of the gene, and is linked to a G to T change in exon 4, causing a Valine to Leucine amino acid substitution in the hinge region of the receptor, and a C to T substitution in exon 5. A study of BRCA1 and BRCA2 mutation carriers found that PROGINS increased ovarian cancer risk by 2.4 fold among the subgroup who had never used oral contraceptives (4). In contrast, no association of PROGINS and sporadic ovarian cancer risk has been identified in several studies of Caucasians from North-America, Southern England, Australia and Austria (5-8).

In light of these conflicting reports, we sought to investigate the hypothesis that the PROGINS allele was associated with increased ovarian cancer risk in a North American population, utilizing a large population-based, case-control study in North-Carolina.

MATERIALS AND METHODS

Study participants are enrolled through the North Carolina Ovarian Cancer (NCOC) study, an ongoing population-based, case-control study, which uses rapid case ascertainment to identify primary epithelial ovarian cancer cases diagnosed between ages 20-74 years of age. Eligible women are those with newly diagnosed epithelial ovarian cancer since January 1, 1999 who are residents in a 48-county region in North Carolina and have had no history of ovarian cancer prior to the current diagnosis. Physician permission is obtained
before any case is contacted. Once permission to contact is obtained, potential ovarian cancer cases are mailed a package of material describing the study and then are contacted by phone by nurse interviewers and invited into the study. Surgical pathology reports and tumor blocks are obtained for all cases and were reviewed by a single expert gynecologic pathologist to confirm diagnosis and study eligibility.

Controls are identified from the same 48-county area in North Carolina. Control subjects are required to have least one intact ovary, and are frequency matched to cases on the basis of race (black vs. non-black) and age (5-year intervals). Identification of controls occurs through either list-assisted (Genesys Sampling Systems, Fort Washington, PA) random digit dialing (RDD) or Health Care Financing Administration (HCFA) phone lists (women 65 to 74 years of age only). HCFA phone lists were used early in the data collection period; however enrollment of control women using this method was hindered due to the lack of telephone numbers on the HCFA computer tapes and was therefore suspended. RDD identified controls are initially screened for eligibility (age, county of residence, at least one intact ovary, and English speaking). Those who passed the initial eligibility screening are then asked whether they would permit study personnel to contact them and provide them with additional information. As with the case subjects, potential control subjects are mailed a package of material describing the study and then are subsequently contacted by phone by nurse interviewers and invited into the study.

Three hundred and nine cases and 397 controls were enrolled between April 1999 and July 2001 and are included in this analysis. For the current analysis, we limited the study subjects to Caucasians and African-Americans. Response rates were 85 percent for cases and 39 percent for HCFA controls. Seventy-three percent of controls identified by RDD, who passed the eligibility screening, agreed to be contacted and sent additional study information. Among those sent additional study information the response rate was 71 percent.
In-person interviews are conducted in the home of each study participant by a trained nurse interviewer using a ninety-minute standardized questionnaire. Information on known and suspected ovarian cancer risk factors including family history of cancer, menstrual characteristics, pregnancy and breastfeeding history, hormone use, and lifestyle characteristics such as smoking history and exposure, alcohol consumption, talc use, sunlight exposure, physical activity, medical history, and occupational history is obtained. Additionally, anthropometric descriptors (height, weight, waist and hip circumference) were measured and a blood sample (30 mls) was collected from each participant at the time of the interview. Within 48 hours all blood samples were centrifuged, and the buffy coat, red blood cells, and plasma were separated. All plasma was stored at -70°C. DNA was extracted from the buffy coat using an ABI nucleic acid extractor.

Extracted genomic DNA was assayed as previously described (5). Each 20uL PCR reaction contained 50 ng DNA, 4 mM magnesium chloride, 1X Gibco PCR buffer (part # Y02028), 0.5 units Taq DNA polymerase (GIBCO cat# 10342-020), 200 uM dNTPs, 2 uM forward primer (5'-GCCAGAAAGGAAAATAAAAGA-3'), and 2 uM reverse primer (5'-AAAGTATTCTTGCTAAATGTC-3'). PCR conditions consisted of an initial denaturing step at 95°C for 3 minutes, 30 cycles of 94°C for 45 seconds, 57.0°C for 45 seconds, and 72°C for 1 minute, an extension step at 72°C for 10 minutes, then at 4°C hold until analyzed. The PCR samples were then resolved on a 2% agarose gel stained with ethidium bromide. In this manner, the wild type allele is seen as a 149-bp band, and the polymorphic insertion-containing allele as a 455-bp band.

All statistical analyses were performed using SAS software (SAS Institute Inc., Cary, NC). Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated using unconditional logistic regression. We report crude ORs as well as adjusted ORs from a multivariate logistic regression model, which included race and age. The potential confounders including menopause status, tubal ligation, oral contraceptive use, family history of breast or ovarian cancer in first
and second degree relatives, and parity were individually tested to see if they changed the crude OR by 10%. Any that did were added to the multivariate model. For the overall analysis, the study has 80% power to detect an OR of 1.6 or greater for carriers heterozygous for the PROGINS allele and an OR of 2.6 or greater for carriers homozygotes for the PROGINS allele compared to noncarriers, for risk of ovarian cancer at an alpha=0.05 level.

Table 1:
Selected characteristics of epithelial ovarian cancer cases and controls

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cases (N=309)</th>
<th>Controls (N=397)</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>(%)</td>
<td>n</td>
</tr>
<tr>
<td>Age at diagnosis/interview</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-49 years</td>
<td>105 (34)</td>
<td>137 (35)</td>
<td>1.0</td>
</tr>
<tr>
<td>50-74 years</td>
<td>204 (66)</td>
<td>260 (65)</td>
<td>1.0</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>276 (89)</td>
<td>346 (87)</td>
<td>1.0</td>
</tr>
<tr>
<td>African-American</td>
<td>33 (11)</td>
<td>51 (13)</td>
<td>0.8</td>
</tr>
<tr>
<td>Post Menopausal</td>
<td>190 (61)</td>
<td>235 (59)</td>
<td>1.1</td>
</tr>
<tr>
<td>Oral Contraceptive Use Duration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never Used</td>
<td>107 (35)</td>
<td>133 (34)</td>
<td>1.0</td>
</tr>
<tr>
<td>&lt;2 years</td>
<td>90 (29)</td>
<td>90 (23)</td>
<td>1.2</td>
</tr>
<tr>
<td>2-5 years</td>
<td>36 (12)</td>
<td>53 (13)</td>
<td>0.8</td>
</tr>
<tr>
<td>&gt;5 years</td>
<td>73 (24)</td>
<td>116 (29)</td>
<td>0.8</td>
</tr>
<tr>
<td>User of Unknown Duration</td>
<td>3 (1)</td>
<td>5 (1)</td>
<td></td>
</tr>
<tr>
<td>1° Family History of Breast or Ovarian Cancer</td>
<td>51 (17)</td>
<td>66 (17)</td>
<td>1.0</td>
</tr>
<tr>
<td>Pregnancies resulting in live birth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>58 (19)</td>
<td>53 (13)</td>
<td>1.0</td>
</tr>
<tr>
<td>1</td>
<td>58 (19)</td>
<td>64 (16)</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>104 (34)</td>
<td>146 (37)</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(17)</td>
<td>4+</td>
</tr>
<tr>
<td>---------------</td>
<td>-----</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>Had Tubal Ligation</td>
<td>76</td>
<td>(25)</td>
<td>132</td>
</tr>
</tbody>
</table>

**Tumor Behavior**

**Borderline**
76 (25)

**Invasive**
233 (75)

**Tumor Stage***

**I/II**
133 (43)

**III/IV**
175 (57)

**Tumor Grade**

**Well differentiated**
35 (12)

**Moderately differentiated**
68 (23)

**Poorly differentiated**
100 (34)

**Undifferentiated**
11 (4)

**Not Applicable***
79 (27)

**Tumor Histology**

**Serous**
182 (59)

**Endometrioid**
38 (12)

**Mucinous**
40 (13)

**Clear cell**
15 (5)

**Other**
34 (11)

*1 missing tumor stage

**16 missing tumor grade

***borderline and carcinosarcomas (malignant mixed mullerian tumor), neuroendocrine carcinoma, and clear cell do not typically receive grades
RESULTS

The distributions epidemiologic risk factors for the cases and controls are shown in Table 1. Cases and controls are similar in age (mean: cases 54.3, controls 54.9) and race, with 11% of cases and 13% of controls being African American. The distributions of menopause status, OC use, and the family history of breast or ovarian cancer are also similar. Controls tend to have had more pregnancies than cases and compared to cases, also are more likely to have had tubal ligations (33% vs. 25%). Additionally, the distributions of ovarian tumor characteristics, including stage, histology and tumor behavior, are found in Table 1. Over 50% of cases were diagnosed with stage III/IV cancer, had invasive tumors, or were of the serous histologic subtype.

Overall, 29.1% of cases and 28.2% of controls had the PROGINS (T2) allele (Table 2). Slightly more cases were heterozygous 25.9% vs. 23.9% but slightly more of the controls were homozygous 4.3% vs. 3.2%. Crude ORs for being heterozygous and homozygous for the rare allele compared to the reference group of non-carrier were 1.1 (95% CI = 0.8 – 1.5) and 0.8 (95% CI = 0.3 – 1.7), respectively. These results remained virtually identical when limiting the cases to invasive tumors only and also when limiting cases to invasive tumors of the serous histologic subtype. Simultaneously adjusting the ORs for age, race, and menopause did not lead to significant changes in any of these ORs. Within the subgroup of women who had never used oral contraceptives, 107 patients with ovarian cancer and 133 control subjects, we found that there was a significant increased risk of ovarian cancer among PROGINS carriers (Table 2). Those heterozygous for the PROGINS allele had an adjusted OR of 1.7 (95% CI = 0.9 – 3.4) and those homozygous for the PROGINS allele had an adjusted OR of 2.2 (95% CI = 0.5 – 9.9). The sample size of these subgroups did not provide sufficient power for either of these ORs to be statistically significant, however, when we combined the carriers (one or more PROGINS alleles), there was sufficient power to identify a borderline significant increased risk, adjusted OR of 1.8 (95% CI = 1.0 – 3.3).
### Table 2: Odds ratio (ORs) and 95% confidence intervals (CIs) for the association between the PROGINS allele in 309 cases of epithelial ovarian cancer and 397 matched controls from 48 counties in North Carolina.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Cases (N=309)</th>
<th>Controls (N=397)</th>
<th>OR</th>
<th>95% CI</th>
<th>OR*</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1T1</td>
<td>219 (71)</td>
<td>285 (72)</td>
<td>1.0</td>
<td>referent</td>
<td>1.0</td>
<td>referent</td>
</tr>
<tr>
<td>T1T2</td>
<td>80 (26)</td>
<td>95 (24)</td>
<td>1.1</td>
<td>(0.8 - 1.5)</td>
<td>1.1</td>
<td>(0.7 - 1.5)</td>
</tr>
<tr>
<td>T2T2</td>
<td>10 (3)</td>
<td>17 (4)</td>
<td>0.8</td>
<td>(0.3 - 1.7)</td>
<td>0.7</td>
<td>(0.3 - 1.6)</td>
</tr>
<tr>
<td>T1T2/T2T2</td>
<td>1.0</td>
<td>(0.8 - 1.5)</td>
<td></td>
<td></td>
<td>1.0</td>
<td>(0.7 - 1.4)</td>
</tr>
</tbody>
</table>

**Oral contraceptive users**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Cases (N=309)</th>
<th>Controls (N=397)</th>
<th>OR</th>
<th>95% CI</th>
<th>OR*</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1T1</td>
<td>146 (72)</td>
<td>180 (68)</td>
<td>1.0</td>
<td>referent</td>
<td>1.0</td>
<td>referent</td>
</tr>
<tr>
<td>T1T2</td>
<td>51 (25)</td>
<td>70 (27)</td>
<td>0.9</td>
<td>(0.6 - 1.4)</td>
<td>0.9</td>
<td>(0.6 - 1.3)</td>
</tr>
<tr>
<td>T2T2</td>
<td>5 (2)</td>
<td>14 (5)</td>
<td>0.4</td>
<td>(0.2 - 1.3)</td>
<td>0.4</td>
<td>(0.2 - 1.2)</td>
</tr>
<tr>
<td>T1T2/T2T2</td>
<td>0.8</td>
<td>(0.6 - 1.2)</td>
<td></td>
<td></td>
<td>0.8</td>
<td>(0.5 - 1.2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Cases (N=309)</th>
<th>Controls (N=397)</th>
<th>OR</th>
<th>95% CI</th>
<th>OR*</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1T1</td>
<td>73 (68)</td>
<td>105 (79)</td>
<td>1.0</td>
<td>referent</td>
<td>1.0</td>
<td>referent</td>
</tr>
<tr>
<td>T1T2</td>
<td>29 (27)</td>
<td>25 (19)</td>
<td>1.7</td>
<td>(0.9 - 3.1)</td>
<td>1.7</td>
<td>(0.9 - 3.4)</td>
</tr>
<tr>
<td>T2T2</td>
<td>5 (5)</td>
<td>3 (2)</td>
<td>2.4</td>
<td>(0.6 - 10.3)</td>
<td>2.2</td>
<td>(0.5 - 9.9)</td>
</tr>
<tr>
<td>T1T2/T2T2</td>
<td>1.7</td>
<td>(1.0 - 3.1)</td>
<td></td>
<td></td>
<td>1.8</td>
<td>(1.0 - 3.3)</td>
</tr>
</tbody>
</table>

**Parous**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Cases (N=309)</th>
<th>Controls (N=397)</th>
<th>OR</th>
<th>95% CI</th>
<th>OR*</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1T1</td>
<td>179 (71)</td>
<td>246 (72)</td>
<td>1.0</td>
<td>referent</td>
<td>1.0</td>
<td>referent</td>
</tr>
<tr>
<td>T1T2</td>
<td>65 (26)</td>
<td>84 (24)</td>
<td>1.1</td>
<td>(0.7 - 1.6)</td>
<td>1.0</td>
<td>(0.7 - 1.5)</td>
</tr>
<tr>
<td>T2T2</td>
<td>7 (3)</td>
<td>14 (4)</td>
<td>0.7</td>
<td>(0.3 - 1.7)</td>
<td>0.7</td>
<td>(0.3 - 1.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Cases (N=309)</th>
<th>Controls (N=397)</th>
<th>OR</th>
<th>95% CI</th>
<th>OR*</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1T1</td>
<td>40 (69)</td>
<td>39 (74)</td>
<td>1.0</td>
<td>referent</td>
<td>1.0</td>
<td>referent</td>
</tr>
<tr>
<td>T1T2</td>
<td>15 (26)</td>
<td>11 (21)</td>
<td>1.3</td>
<td>(0.5 - 3.3)</td>
<td>1.2</td>
<td>(0.5 - 3.2)</td>
</tr>
<tr>
<td>T2T2</td>
<td>3 (5)</td>
<td>3 (6)</td>
<td>1.0</td>
<td>(0.2 - 5.1)</td>
<td>0.9</td>
<td>(0.2 - 5.2)</td>
</tr>
</tbody>
</table>

*adjusted odds ratio

252
Among women with a history of oral contraceptive use, carriers had similar risk to noncarriers (adjusted OR = 0.8; 95% CI = 0.5-1.2), although there was some suggestion of a protective effect among the subgroup homozygous for the PROGINS allele (adjusted OR = 0.4; 95% CI = 0.2-1.2).

Neither subsetting the data by age, parity, nor race resulted in any significant associations. Among African-Americans, none were homozygous for the PROGINS allele and only 1 case (3%) and 4 controls (8%) were heterozygous. While this resulted in an OR of 0.4 (95% CI = 0.0 - 4.9) for carriers, the sample was very small and thus the 95% CI was quite wide. We also examined the association between the PROGINS genotype and cancer stage and grade. The PROGINS allele was distributed similarly among those with stage I/II disease and those with stage III/IV. The PROGINS distribution was also similar between invasive cases with undifferentiated/poorly differentiated cancer and those with moderately/well differentiated tumors.

**DISCUSSION**

This study is the first population-based, case-control study to examine the relationship between the PROGINS allele and ovarian cancer among a United States population. Furthermore, the existence of epidemiologic data as well as information on cancer stage and histology, allow us to examine the significance of PROGINS within specific subgroups while controlling for potential confounding factors. Our study does not support the conclusion that the PROGINS allele is associated with ovarian cancer risk in the overall sample, although our sample size provides sufficient statistical power to detect an association of the level previously observed (3). There did appear to be some increased risk associated with the PROGINS allele among the subpopulation who had no history of oral contraceptive use. This is consistent with the previous finding by Runnebaum et al. (4). We should point out that this increased risk may be an artifact of our
sample since the PROGINS distribution among cases who were OC non-users exactly matches the PROGINS distribution among controls who are OC users. No significant association was found between the PROGINS allele and ovarian cancer risk among any other subgroup within our sample.

The data did provide some suggestion that homozygosity for the PROGINS allele may be protective among OC users, and carrying a PROGINS allele may be protective among African-Americans. However the number of these genotypes within our sample prohibits definitive conclusions from being drawn. We find no evidence to support an association between the PROGINS allele and increased risk of ovarian cancer, in the North American population.

ACKNOWLEDGEMENTS
This study was supported by the following grants: National Institutes of Health (NIH) (grant number 1-R01-CA76016) and the Department of Defense (DOD) (grant number DAMD17-98-1-8656). The authors would like to thank Robin Berger, Lantoya Hobbs, and Nancy Fisher for conducting the interviews.
REFERENCES


Chapter 11

CONCLUSIONS:

Genomics in Ovarian Cancer Research and Clinical Practice
Introduction

This thesis has described a range of molecular genetic analyses aimed at improving our understanding of epithelial ovarian cancer— including that of disease risk, development, progression, and response to therapy.

The major focus of the work is the application of genome-wide expression arrays to characterize epithelial ovarian cancer. Specifically, the thesis begins by detailing the use of microarray analysis to characterize advanced stage (III/IV) serous ovarian cancers and normal ovarian epithelial samples. Differentially expressed genes, identified in the initial study using first-generation gene-chips, were then subject to an expanded expression analysis in a larger collection of ovarian cancer samples. In doing so, several genes were identified that may have potential future clinical utility as either serum screening and/or monitoring biomarkers (IGFBP2), markers of favorable (TRAIL) and poor prognosis (IGF2), and as novel therapeutic targets (TRAIL, IGF2, and IGFBP2). Using later generation microarrays, the gene expression patterns that underlie ovarian cancer metastasis were explored and described. Many of the genes within the profiles that appear to underlie ovarian cancer metastasis have previously described functions in oncogenesis, metastasis and p53 gene networks. These findings reinforce the importance of the p53 pathway in ovarian cancer development and progression, and illustrate how genome-wide analysis can refine our understanding of the complex gene network relationships that underlie previously identified single-gene:phenotype associations. Further, this thesis has described how microarray analysis was applied to primary ovarian cancer specimens, successfully identifying profiles that predict response to platinum-based adjuvant therapy for advanced ovarian cancer. To our knowledge, this represents one of the first steps toward the application of gene-expression array technology to tailor therapies for women with ovarian cancer. Finally, the genetic basis of risk for the development of ovarian cancer was explored in a population-based case-control study of approximately 700 women in North America in an effort to determine whether a polymorphism in the progesterone receptor gene is associated with increased risk of ovarian cancer.
Although the thesis describes a selection of molecular genetic techniques for the analyses of epithelial ovarian cancers, the main focus of the work has been application of microarray technology to elucidate the molecular underpinnings of ovarian cancer development, progression, response to therapy, and outcome.

**Genomic Expression Analysis:** Genomic expression analysis is revolutionizing the way human disease research is performed throughout the world. In the very near future, it is anticipated that it will have a similar impact on the way clinical medicine is practiced. Globally, expression array data can be interpreted in two ways. The first, enables us to view the expression of tens of thousands of genes in a single sample much like a large scale Northern blot. As such, the interactions and relationships of each gene can be studied, and assumptions can be made about the biologic basis of the specific phenotype. Such approaches will provide critical insights into the biology that drives human disease development, and will hopefully reveal novel therapeutic avenues for the development of more active therapies. The second interpretation of array data enables us to view the gene expression patterns as entities in themselves. That is to say, that each discrete gene profile represents an independent biomarker that has potential application in clinical practice.

**Genomic Medicine**

**Application to Human Disease:** Molecular profiles for predictive phenotyping represent a critical component of our efforts to truly implement personalized medicine – health planning, treatment strategies and drugs customized to the individual patient rather than broad populations. The value, and simultaneously the key challenge, in genomic data now accessible is its scale and complexity; the information content has the potential to identify unique characteristics of the individual that will define customized health care strategies. The principle challenges reflect the need to improve methodologies to identify, characterize and deliver clinically and scientifically useful information from masses of genomic
data, and to progressively refine, evaluate and validate genomic predictors that play roles in overall clinical risk assessments and prognostic models. Human cancer provides a striking example of the need for individualized treatment. For example, a woman with an apparent early stage ovarian cancer will undergo surgical staging with resection of the primary tumor. Significant debate exists regarding the benefit of chemotherapy in this group of women with confirmed early stage disease. To date, no clinical tools exist to predict which patients are cured by surgery alone and which patients have residual microscopic disease (which will likely give rise to clinical recurrence, and which might benefit from adjuvant chemotherapy). As a result, it is possible that some women who had no ovarian cancer cells remaining at the conclusion of their initial surgery will receive adjuvant chemotherapy unnecessarily. These women are needlessly exposed to all of the potential toxicities and side-effects of therapy. Moreover, it is possible that other women, who have microscopic disease remaining at the conclusion of surgery may have benefited from adjuvant chemotherapy, will not receive it and will ultimately experience an incurable recurrence. Multiple clinical risk factors relate to risk of disease recurrence, but individually or combined do not reach the necessary precision in their predictive ability to accurately segregate women who will benefit from more aggressive therapy from those who will not. It is in this context that genome-scale molecular data, such as gene expression profiles, offer potential for substantially improved resolution towards personalized prognosis. It has now been clearly demonstrated that expression assays of primary tumor can reliably identify differences that reflect tumor behavior and ultimate outcome of disease. The need now is for broader development of prognostic methods that summarize expression data and combine it with traditional clinical and genetic information in prognostic models, coupled with expansion of studies to understand, evaluate, improve and validate such models in broader, diverse populations.
Ovarian cancer

The importance of molecular genetics in ovarian cancer: An incomplete understanding of the molecular etiology that underlies ovarian cancer risk, development, progression and response to therapy is a major impediment to improving outcome from the disease. Our inability to fully define the molecular genetic events associated with ovarian cancer risk and the malignant transformation of the ovarian epithelium represent an obstacle to the development of clinical protocols focused on women most likely to develop ovarian cancer. This underlies, in part, our limited capacity to diagnose early stage and often curable disease. Our lack of understanding of the molecular basis of the disease means that physicians cannot use molecular information to make more informed clinical decisions. Currently available diagnostic tools do not provide accurate prognostic information relating to treatment outcome or survival, nor identify which patients might benefit from surgery versus neoadjuvant chemotherapy. Moreover, selection of chemotherapeutic agents continues to be made on the basis of empiric algorithms and statistical response rates. Chemotherapeutic decisions are governed by statistics, age, weight and renal function, rather than by the unique biologic characteristics of the tumor or the patient. Epithelial ovarian cancer is a highly clinically heterogeneous disease, likely due to a highly heterogeneous molecular etiology. Until we characterize the various molecular characteristics that contribute to the clinical phenotype we will be unable to optimally utilize biologic data in clinical decision-making.

Single gene studies in ovarian cancer: It is believed that malignant transformation of normal ovarian epithelial cells is caused by genetic alterations that disrupt regulation of proliferation, apoptosis, senescence and DNA repair. The alterations that lead to ovarian cancer vary between patients and the striking clinical heterogeneity of ovarian cancer likely reflects an underlying molecular heterogeneity. Although approximately 10% of epithelial ovarian cancers arise in women who have inherited mutations in cancer susceptibility genes such as
BRCA1 or BRCA2, the vast majority occur from the sporadic accumulation of genetic damage over the course of a lifetime. Our group has shown that the p53 tumor suppressor gene is frequently mutated in ovarian cancers, and that about 70% of advanced-stage serous cases harbor p53 mutations (1-4). Other tumor suppressor genes implicated in the development of some ovarian cancers include p16, p21 and p27 (3,5). In addition, several putative ovarian cancer tumor suppressor genes that are expressed in normal ovarian epithelial cells, but not in ovarian cancers have been described - including NOEY2 (6), SPARC (7), DOC2 (8), LOT-1 transcription factor (9), OVCA1 and OVCA2 (10). Finally, oncogenes such as K-ras, HER2/neu, c-myc, PIC3KA and AKT2 are activated by mutation, amplification and/or over-expression in a fraction of ovarian cancers (11). Despite such information, single-gene markers that reliably predict individual ovarian cancer clinical phenotypes have not been identified.

**Genome expression analysis of ovarian cancers:** Global technologies are now being used to analyze genomic, epigenetic, transcriptional profiling and proteomic patterns in tissue and blood samples from patients with ovarian cancer. These technologies have great potential for identification of indicators for prognosis, response to therapy, and new therapeutic targets. For instance, high resolution comparative genomic hybridization analysis (CGH) has shown that histologic subtypes of ovarian cancer have distinct patterns of genomic change (12). Several groups have applied microarray technology to the analysis of ovarian cancers (13-15). These studies have demonstrated differences in gene expression patterns between normal ovarian epithelial cells and ovarian cancers and between various grades, stages and BRCA1/2 characteristics of ovarian cancers. Such global assessments of gene expression in ovarian cancer specimens have shown that gene expression profiles can provide molecular phenotyping that identifies distinct classifications not evident by traditional histopathologic methods. Despite the enormous potential of array technology to change the way we classify cancers from histopathologic to molecular categories, to date, no reports exist documenting the application of such
technology in directing clinical management decisions for patients with ovarian cancer.

**Genomic medicine and ovarian cancer - the need for “tailored therapy”:** In the United States approximately 26,000 new cases of epithelial ovarian cancer were diagnosed in 2004 and 16,090 women died of the disease (16). The majority of newly diagnosed patients (approximately 17,500 patients) will receive adjuvant chemotherapy, and the most of these patients (~14,000) will experience persistent or recurrent disease requiring treatment with salvage therapies. Review of SEER data suggests that the prevalence of ovarian cancer is approximately 200,000 cases per year (17) and it is therefore likely that approximately 25% of these women (approximately 50,000 in the U.S.) are undergoing active chemotherapy annually.

Patients with persistent disease at the conclusion of primary therapy, or those that develop platinum-resistant recurrent disease are treated with salvage agents that have response rates of 10-20%. This response rate is dependent on the response to initial platinum therapy, the disease-free interval before recurrence, previous agents used, existing cumulative toxicities, and the patient’s performance status. Although choice of salvage agent is made based-upon all of these factors, no reliable clinical predictor of response to therapy exists, such that the majority of patients are treated somewhat empirically. In light of the response rates observed in this setting, between 1:5 and 1:10 patients receive benefit from each agent, yet all agents have some toxicity to patients. Many of the agents used have cumulative toxicities, which not only impact quality of life, but can compromise a patient’s tolerance of consequent therapies and therefore impact survival. Many patients with ovarian cancer will be on therapy for the remainder of their lives, and for this reason, it would be highly desirable to be able to individualize the choice of chemotherapy to each patient, in order to reduce the incidence and severity of toxicities that could potentially limit quality of life and compromise tolerance of subsequent therapies.
Throughout diagnosis and treatment for ovarian cancer, prolongation of survival and the successful maintenance of quality of life remain important goals. Improving our ability to diagnose and manage the disease by optimizing existing - and developing new – screening, diagnostic, and therapeutic methodologies is essential in this endeavor. To this end, individualizing medicine by the development and integration of genomic, genetic and clinical data to define comprehensive models of disease susceptibility, onset, progression, response and outcomes will enable physicians to define women at risk, focus screening and prevention strategies, and identify which patients will - and will not - respond to specific chemotherapeutic agents. By tailoring therapy to individual patients on the basis of biologic markers of response, it may be possible to increase response rates to existing agents and identify patients who might respond to other agents, including novel biological therapies such as gene therapy, radioimmunotherapy and anti-angiogenesis agents. If successful, we will diagnose disease at an earlier and more curable stage, increase therapy response rates, reduce the incidence and severity of toxicities from non-active agents that could not only limit quality of life, but also limit ability to tolerate further therapies, and therefore reduce survival.

**Future directions:** The major focus of the work described in this thesis is the application of genome-wide expression arrays to characterize epithelial ovarian cancer. We have applied genome-wide expression analysis techniques to better understand ovarian cancer development, progression, and response to therapy. We believe that expression array technologies will play an increasingly important role in biomedical research and clinical practice, and our ongoing and future research plans reflect this.

As previously mentioned, expression array data can be interpreted in two ways. The first views expression of tens of thousands of genes in a single sample much like a large scale Northern blot revealing details of the biology that underlies a phenotype. The second interpretation of array data views the gene expression profile as an independent biomarker, with less focus on the individual
genes that contribute to the profile, but with potential to make future phenotypic predictions of unknown samples. The future direction of our research will take advantage of both of these interpretations.

Currently, in the first regard, we are exploring ways to manipulate the expression of individual genes within the profiles that underlie the biology of platinum-response and ovarian cancer metastasis. We are developing small interfering RNA (siRNA) approaches to achieve a targeted in-vitro knock-down of expression of individual genes within complex gene networks with the objective of changing the phenotype. If successful, we hope to be able to manipulate the genetic profile of chemo-resistant ovarian cancer cells to increase platinum sensitivity. We aim to take a similar approach to targeted manipulation of genes that appear to be critical in the complex profiles that underlie ovarian cancer metastasis.

In regard to the former objective of reversing platinum-resistance using siRNA strategies, we have promising preliminary results. As described in chapter 9, using genome-wide microarray expression analysis we have defined a gene expression profile that predicts response to platinum-based therapy for ovarian cancer. Many of the genes within this profile have gene network relationships with the c-myc gene. In ongoing preliminary studies, using real-time PCR we have demonstrated that the c-myc gene is more highly expressed in platinum-resistant ovarian cancer cell lines, and that expression of the c-myc gene in platinum-resistant cells can be inhibited by RNA interference (RNAi) technology (Fig 1A). Further, we have preliminary data to suggest this in-vitro targeted knock-down of c-myc gene expression may reverse platinum-resistance (Fig 1B).
Figure 1A: Western blot analysis demonstrating targeted knock-down of c-myc gene expression using RNA interference (preliminary findings): C13 platinum-resistant ovarian cancer cell lines were grown to 50% confluence, and transfected with c-myc siRNA at 3 different concentrations 50nM, 100nM and 150nM, and two RNAi-negative controls (mock). Following cell counts, Western blot analysis was performed to analyse c-myc protein levels in cells subject to c-myc RNAi transfection, and revealed significant decrease in c-myc protein level in cells transfected with (50nM and 150nM) c-myc siRNA. Actin control.

Figure 1B: Reversal of platinum resistance by c-myc siRNA (preliminary findings): Ovarian cancer cell line C13 viability with cisplatin treatment (0.1, 0.5, 1.0, 10, 100 uM) both in presence (solid line) and absence (broken line) of c-myc siRNA transfection.
If these preliminary findings are validated in ongoing work, then siRNA targeted gene silencing may be a developed further as a potential clinical therapeutic adjuvant to currently available therapy. It may be possible to assay an ovarian cancer specimen resected at the time of primary surgical cytoreduction, and determine if the cancer is likely to respond to primary platinum therapy. If the gene expression profiles predict that the specimen is likely to be platinum resistant, then addition of siRNA-based gene silencing therapy may be used to reverse this resistant phenotype. Although our preliminary work has focused on decreasing expression of a gene (c-myc) which we have demonstrated to be over-expressed in platinum-resistant cells using siRNA, it is equally feasible to decrease expression of other genes, or increase expression of genes that are under-expressed, in resistant cells using a variety of expression vector-based molecular genetic techniques.

We are pursuing a similar approach to manipulation of gene expression within the profiles that we have demonstrated to underlie ovarian cancer metastasis (chapter 7). Use of siRNA and expression-vector approaches will enable us to decrease or increase expression of genes within key gene networks. The effect on cell invasion and (thus possibly) metastatic potential can be measured using commercially available cell migration assays. It may soon become feasible to take advantage of advances in drug delivery systems (e.g. nanotechnology-based liposomal capsules) to safely and efficaciously deliver siRNAs and expression vectors to tumor sites, with the promise of enhancing our ability to treat ovarian cancer. If successful, it is possible that this research will reveal a potentially novel and clinically valuable approach to reversing the metastatic potential of an ovarian cancer cell.

In the second regard (viewing the gene expression profile as an independent biomarker), in ongoing research we are developing gene expression profiles that predict response to platinum-based primary therapy, and also to salvage agents. We are expanding the numbers of ovarian cancer samples (including non-serous histopathologic subtypes) in our retrospective analysis of the gene expression profiles that predict response to platinum-based
chemotherapy, in an effort to refine and improve the predictive ability of the profile described in chapter 9. Additionally, we are prospectively collecting ovarian cancer samples to be used as an independent external validation set for our predictive model. In a similar way we are using tumor bank samples from patients who developed recurrent/platinum-resistant disease and went on to receive treatment with salvage agents (liposomal doxorubicin, topotecan etc) in an effort to develop gene expression profiles that are predictive of response to these agents. These profiles will then be validated in prospectively collected samples.

If gene expression profiles can reliably predict response to primary and salvage therapy, as described above, prospective clinical trials will be required to demonstrate the benefit of expression profile-directed therapy. While reasonably straightforward in concept for patients with recurrent disease, for whom few successful therapeutic options exist, design of such trials will more challenging for patients undergoing primary therapy. It may be ethically difficult to withhold platinum in the primary therapy setting from a patient on the basis of a gene expression profile, when, statistically, no other agent has efficacy close to that of platinum for advanced epithelial ovarian cancer. However, in patients predicted to have platinum-resistant disease at primary diagnosis it might be possible to trial adjunct agents (such as platinum-sensitising siRNAs and other novel biological therapies), in addition to platinum.

The objective of these studies will be to tailor therapy to individual patients on the basis of biologic markers of response. If successful, we may increase therapy response rates and reduce the incidence and severity of toxicities from non-active agents that could not only limit quality of life, but also limit ability to tolerate further therapies, and therefore reduce survival.

Limitations of Microarray Analysis: Despite the advantages of microarray techniques, the technology is still evolving. At this time, only a subset of genes can be included in any single array, and the arrays may be subject to production error and batch-to-batch variability. It is, therefore, not yet clear how accurately
the expression data produced reflects true expression, such that ongoing validation using other platforms is essential. Within a cancer specimen, infiltrating normal and inflammatory cells and blood vessels can influence the expression profile considerably, necessitating careful review of tumor histology prior to analysis. As microarray technology evolves and moves into the clinical arena, consistency within and between centers will become of prime importance. Subtle differences in the way specimens are collected, handled and processed can have a dramatic impact on RNA quality and yield, and subsequent gene-chip hybridization. Such differences can significantly decrease the reliability, and therefore clinical application, of array predictions. Finally, one of the greatest challenges of this technique is data analysis. The interpretation of expression data from up to 40,000 genes in current microarrays is daunting, and requires the highest level of expertise in biology, statistics and computation.

**Summary:** The work described in this thesis provides novel insights into the molecular underpinnings of ovarian cancer development, progression and response to therapy. Further, this work reveals how application of expression array technologies may not only enhance our ability to tailor currently available therapies for ovarian cancer, but may aid in the identification of novel therapeutic approaches. We are convinced that use of genomic technologies in the laboratory and the clinic will lead to an improvement in our ability to care for patients with advanced stage epithelial ovarian cancer.
REFERENCES


STATEMENT OF WORK

The work presented in this thesis is my own, unless otherwise stated on the chapter title page. For the 8 chapters that list my name as first author, I personally: designed the experiments, identified patients/samples for the study, extracted RNA and DNA from tumor samples (with the aid of Ms Regina Whitaker, the senior technician in the Gynecologic Oncology Laboratory at Duke University Medical Center), extracted and collated all clinical data from electronic and paper patient records for samples analysed, established and maintained the clinico-pathologic databases, designed QRT-PCR primers, designed PROGINS polymorphism genotyping assay, performed QRT-PCR/genotyping reactions, performed data and statistical analysis for QRT-PCR studies, coordinated sample assay and data analysis for microarray experiments in which core facilities were used, collaborated with biostatisticians in analysis and interpretation of microarray gene expression data, authored the manuscripts/chapters. For the 3 chapters that list my name as co-author, I contributed to study design, contributed to completion of QRT-PCR assays, collaborated on data analysis and interpretation, and co-authored the manuscript/chapter.
Acknowledgements

To my wife Vanessa and our twins, Georgina and Sebastian, for all their love and support. And for all the times I wasn’t at home.

To my parents and grandparents for giving me the opportunity.

To Stephen Bassett for being my friend.

To Ian Jacobs for lighting a flame in me, and for sending me to America. And for forgiving me when I never returned.

To Roger Wiseman and Charles Cochran, for your kindness and patience in turning me into a scientist.

To Andrew Berchuck. My friend and mentor. For helping me become who I am.

To Mike, Larry, Angeles, Laura and Rob. For making it fun.

To Regina, for everything.

To Joe Nevins and Holly Dressman for the science, the beer and the pool.

To Jonathon Gray, for outstanding PhD supervision!

To Sharon. For reminding me to do it.

To all my colleagues, past and present, at the University of Wales College of Medicine, the National Institute of Environmental Health Sciences, Duke University Medical Center, and the H. Lee Moffitt Cancer Center & Research Institute. Thank you all for your teaching, support, encouragement and friendship.

And most of all to my patients, who inspire me to do what I do.