Does tumour biology determine surgical success in the treatment of epithelial ovarian cancer? A systematic literature review

J Borley1, C Wilhelm-Benartzi1, R Brown1,2 and S Ghaem-Maghami*1

1Epigenetics Unit, Department of Surgery and Cancer, Imperial College London, 4th Floor IRDB, Hammersmith Hospital, London W12 0NN, UK;
2Division of Molecular Pathology, Section of Medicine, Institute for Cancer Research, Sutton SM2 5NG, UK

BACKGROUND: Ovarian cancer is the most lethal gynaecological cancer. Progression-free and overall survival is significantly related to surgical success and residual disease volume. It is unclear whether this survival advantage is due to an intrinsic biological element of the tumour cells which enables successful surgery and improved prognosis, or alternatively the number of tumour sustaining cells remaining irrespective of differences in biology.

METHODS: A systematic review of the literature was performed identifying studies that have investigated the association between biomarkers and surgical outcomes. We attempted validation of these results using The Cancer Genome Atlas ovarian cancer data sets.

RESULTS: Thirty studies were identified of which sixteen determined protein expression, eight gene expression and one DNA methylation in association with surgical debulking. Individualised linear models adjusting for batch, stage and age identified only expression of the genes MTDH and insulin-like growth factor-1 receptor (IGF1R) to be significantly associated with debulking surgery (P < 0.05, false discovery rate (FDR) < 5%), although in the case of IGF1R this was in the opposite direction to previous findings.

CONCLUSION: The majority of studies are limited by design, include heterogeneous samples and lack adjustment for major confounding factors. High quality detailed clinical annotations should be routinely collected in future to more accurately evaluate biomarkers of surgical outcome.

Keywords: ovarian cancer; debulking surgery; molecular biomarkers

Ovarian cancer is the second commonest and most lethal of all gynaecological cancers. In the United Kingdom, ~6500 new cases are diagnosed every year with the majority presenting with advanced disease (Cancer Research UK, 2011). Over the past 30 years, despite advances in chemotherapeutic agents, there has been little change to overall survival (OS) with current 5-year survival at ~35% (Coleman et al., 2011). The standard treatment for epithelial ovarian carcinoma (EOC) is primary surgery, which usually includes total abdominal hysterectomy, bilateral salpingoophorectomy, omentectomy and lymphadenectomy in some cases, followed by adjuvant platinum-based chemotherapy with or without paclitaxel. Surgical success is an important prognostic factor and it is widely accepted that volume of residual disease after primary surgery influences OS following treatment (Pomel et al., 2008; Shih and Chi, 2010). This appears to be true independent of the type of chemotherapy used (Vergote et al., 2010). The Gynaecologic Oncology Group has classified cytoreductive surgery to describe the amount of residual disease at the end of surgery (Hoskins et al., 1994). Optimal debulking currently refers to a maximal diameter of residual tumour ≤1 cm, whereas suboptimal debulk describes residual tumours >1 cm in diameter. Meta-analysis of published data in post platinum era shows a positive correlation between surgical debulk status and survival in advanced disease (Bristow et al., 2002). For every 10% increase in maximal cytoreduction, there appears to be a 5.5% increase in median survival. There is now growing evidence that in fact total (or complete) macroscopic debulking, defined by an absence of any macroscopic tumour post-operatively, is one of the most important factors in survival (Winberger et al., 2007, 2010; Eisenhauer et al., 2008; Colombo et al., 2009) with a reported increase in OS of 14–28 months in patients with total debulking as compared with optimally debulked patients. This appears to be true for all histological subtypes including clear cell, a particularly aggressive form of EOC (Takano et al., 2006). There are two arguments for this survival benefit; (1) that with minimal residual disease there are less malignant cells present and thus improved subsequent ability for chemotherapy to reach the centre of the tumour and a decreased rate of chemotherapy resistance, (2) an intrinsic biological element of the tumour cells leads to a less aggressive, less invasive disease that incurs a survival advantage and additionally allows a more successful surgical outcome. The crucial question of whether there is evidence for a role of tumour biology in determining surgical outcome is still unanswered (Chi and Schwartz, 2008). In the United Kingdom, all patients with suspected EOC undergo preoperative assessment in a bid to plan and predict surgical treatment. This includes assessment of tumour load through level of tumour markers, ultrasound scans, computer tomography (CT) and magnetic resonance imaging (MRI). However, despite this ability to preoperatively assess the involvement of certain organs,
true extent of disease is often difficult to ascertain. Much emphasis has been placed on investigating the use of the tumour marker cancer antigen 125 (Ca125) and imaging modalities as a possible predictive model of surgical outcome but without success (Ibeanu and Bristow, 2010). If tumour biology did determine surgical success, then biomarkers could potentially be developed that predicted outcomes of surgery and aided clinical treatment decisions. In this systematic review, we aim to summarise the evidence to date and attempt to validate previous findings through the publically available The Cancer Genome Atlas (TCGA) gene expression and DNA methylation data sets from patients with high-grade and advanced stage serous EOC.

METHODS

Systematic literature search

A systematic literature search was performed in PUBMED using the search terms ((debulking surgery) OR (optimal suboptimal) OR (cytoreduct*)) AND (ovarian cancer) AND ((microarray) OR (gene expression) OR (protein expression) OR (DNA methylations)). Inclusion criteria for the studies were as follows: primary diagnosis of EOC, primary peritoneal or fallopian tube carcinoma, tissue taken from patients undergoing primary surgery, optimal and suboptimal debulking categories defined. Articles were excluded if they were not written in English, if the study was not performed on human tissue, if there were no specific results in relation to surgical outcome, if surgical debulking was not defined or if patients had neoadjuvant chemotherapy or interval debulking surgery. From inclusive dates of January 1990 to September 2011, the search initially yielded 144 articles in PUBMED. Inspection of abstracts to exclude publications lacking relevant clinical information narrowed the search to 41 potential studies. These articles were read in full and the reference lists were searched manually, yielding one further study. From a total of 42 studies, 2 publications were excluded as they were not written in English, 6 were excluded as they included tissue taken at recurrence or after chemotherapy, 2 were excluded as there was failure to define debulking surgery or residual disease, 1 was excluded as neither debulking nor primary surgery was defined and 1 was excluded as it was unclear whether tumours were epithelial ovarian in origin. The final 30 publications that met inclusion criteria are described in Supplementary Tables 1–3.

TCGA data set

Data consisting of 311 high-grade serous ovarian tumours were obtained from the TCGA data portal (http://cancergenome.nih.gov/dataportal). Level 2 expression data on Affymetrix HGU133A microarrays and level 3 methylation data on Illumina Human Methylation27 Beadchip and annotated clinical data were obtained. Methods including sample inclusion criteria, sample processing and quality control are previously published (Cancer Genome Atlas Research Network, 2011; Dai et al, 2011). Our analysis of these data was limited to the 279 patients with stage 3 or 4 serous EOC that contained details describing volume of residual disease. We performed power calculations to ensure the TCGA validation set was adequately powered to detect effects seen in previously published studies for markers we found to be significant (insulin-like growth factor-1 receptor (IGF1R) – An et al, 2009 and AEG1 – Li et al, 2011) and those we did not (HOXA11 – Fiegl et al, 2008). Individual generalised linear models were used to determine the association between gene expression and DNA methylation of specific loci with surgical debulking status adjusting for microarray batch, age at diagnosis and stage. Multiple comparisons were accounted for in the models using false discovery rate (FDR) estimation to calculate q values. Statistical analyses were performed using the R statistical package (version 2.10 at http://www.r-project.org) and power calculations were performed following previously published methods (Demidenko, 2007).

RESULTS

Protein expression associated with surgical outcomes

Sixteen out of the thirteen publications included data on protein expression in relation to clinical characteristics and surgical debulking outcomes. Supplementary Table 1 describes the main characteristics of each study; all used immunohistochemistry to determine levels of protein expression. Eleven of the studies were single studies investigating different proteins. Increased expression of Cyclin E (Rosen et al, 2006), c-erbB-2 (Simpson et al, 1995), Twist protein (Hosono et al, 2007), p63x (Jewell et al, 2009), ERCC1 (Lin et al, 2010), AEG-1 (Li et al, 2011) and P150Cas (Nick et al, 2011) have all been shown to be associated with suboptimal debulking. The majority of studies included all types of EOC histology and all stages of disease in analysis although this could lead to confounding and bias. For example, in one study increased expression of COX-2 protein was significantly related to suboptimal debulking surgery (defined as residual disease > 2 cm) in 64 EOC samples (P = 0.027) (Seo et al, 2004). However in subset analysis, using only serous and endometroid tumours (n = 46) there was no longer a significant association with debulking surgery (P = 0.743). Additionally, it should be noted that all the studies identified used univariate statistical analysis to find associations between protein expression and surgery. As surgical debulking is strongly associated with stage these results may indicate an association with stage rather than surgical outcome.

Four publications were identified that explored the relationship of p53 and residual disease, of which two found a significant association with p53 protein overexpression and suboptimal debulking in univariate analysis (n = 82, P = 0.01) (Dogan et al, 2005) (n = 83, P < 0.041) (Geisler et al, 1997), one demonstrated a trend to overexpression and suboptimal debulking (n = 136, P = 0.07) (Ferrandina et al, 1999) and one did not find a significant association (n = 79, P = 0.36) (Bar et al, 2001). These four studies used heterogeneous histological samples, variable clinical stages and differing definitions of suboptimal debulking. Results from a meta-analysis have highlighted the association between p53 expression and stage in ovarian cancer (de Graeff et al, 2009) and thus again there are likely to be important confounding factors.

Gene expression associated with surgical outcomes

There were 8 studies out of the 30 identified that investigated the specific gene expression to clinical parameters including debulking outcomes. Seven of these used reverse-transcriptase PCR (RT-PCR) of mRNA to determine gene expression, one used a different method of quantification, the RiboQuant Multi-Probe RNase protection assay system (mRNA electrophoresis) (Komiyama et al, 2011). The data are summarised in Supplementary Table 2. Increased gene expression of IGF1R (An et al, 2009), insulin-like growth factor-2 promoter transcripts (Lu et al, 2006), vascular endothelial growth factor C (VEGF-C) (Sinn et al, 2009), SRA1 (Leoutsakou et al, 2006), transforming growth factor-beta 1 (TGFβ1) (Komiyama et al, 2011), Cox5x2-adenosine receptor isoforms (CAR3/7 CAR4/7) (Reimer et al, 2007) and lower gene expression of the RNAse III enzyme Drosher (Merritt et al, 2008) have all been found to be associated with suboptimal debulking. It is of note that the studies vary in genes investigated promoter transcriptions and definitions of optimal and suboptimal debulking.
The use of microarray platforms has enabled a wider exploration of the potential association between gene expression patterns. Of the 30 identified studies, 3 publications have specifically explored this relationship with debulking status. The first published work (Berchuck et al., 2004) performed in 2004 used Affymetrix U133A GeneChip microarray to explore expression patterns in association with optimal (<1 cm residual disease) vs suboptimal debulking surgery (>1 cm residual disease). Using frozen tissue samples from 44 patients undergoing primary surgery for stage 3 or 4 serous EOC, 120 genes were associated with debulking status with a significance of \( P < 0.01 \). Seventeen of these exhibited >2-fold expression between the two groups. A statistical model correlation was used to define 32 genes, which could be incorporated into a predictive model for debulking accuracy. Genes included mitogen protein kinase (MAP) family; a metastasis suppressor gene, retinoic acid receptor-\( \beta \) (RARB) and p53 inducible protein P2X6. The authors concluded that the study supports the hypothesis that there are biological differences between tumours that are optimally vs suboptimally debulked. This analysis however has limitations through the absence of adjusting for multiple comparisons by FDR and a lack of independent data set validation. A similar study (Levine et al., 2004) has reported opposing results. Gene expression microarray (on the same U133A platform) was performed using 70 fresh frozen tissues from stage 3 and 4, high-grade serous ovarian cancers. Using a supervised class comparison, they found no differentially expressed genes between optimal (<1 cm) and suboptimal (>1 cm) debulking groups using an FDR of 10%. A follow-on study (Bonome et al., 2008), with a larger sample size of 185 serous histology tumours, concluded that expression profiling could not distinguish between optimally and suboptimally debulked tumours. Additionally, only 21 probes out of 22,283 were differentially expressed between the two groups.

One additional seminal study using expression profiling (Tothill et al., 2008) found 6 molecular subtypes following microarray gene expression profiling on 285 serous and endometroid tumours of the ovary, peritoneum and fallopian tube. Although not identified by the systematic literature search, this study found that out of the four subtypes which represented high-grade cancers there was a significant difference of residual disease volume between the groups. Approximately 50% of patients in the \( k \)-Means clustering subtype group \( C1 \) were suboptimally debulked compared with 11–27% of patients in the other high-grade subtype groups. Interestingly, this subtype group showed enhanced expression of the ‘stromal gene cluster’ that included pathways of enrichment of extracellular matrix production and remodelling, cell adhesion and angiogenesis. This difference in debulking surgery may however be attributed to the fact that a larger proportion of patients in the \( C1 \) group had primary peritoneal carcinoma, a disease that presents with a diffuse tumour pattern which is often not possible to remove completely at surgery.

Other translational studies

Other microarray-based technologies, such as comparative genomic hybridisation that provides a genome-wide survey of copy number changes, have also been utilised to study relationship between surgical outcome and biological factors. In one such study (Tan et al., 2011) investigating 50 ovarian clear cell carcinomas used hierarchical unsupervised clustering and determined 2 distinct groups based on copy number changes, both of which had an equal numbers of patients with residual disease <2 cm and >2 cm (\( P = 0.687 \)). The immunological response to ovarian cancer has also been studied in relation to debulking surgery (Barnett et al., 2010). Fresh-frozen EOC tumours were obtained from 232 women at primary debulking surgery, and tumour infiltrating T cells were determined through immunohistochemistry. The mean number of T-regulatory cells (CD4+CD25+ T cells) were observed to be higher in suboptimally (\( \geq 1 \) cm) debulked patients compared with those optimally debulked (<1 cm) (\( n = 145, P = 0.04 \)). Epigenetic alterations are being increasingly investigated as potential biomarkers for diagnosis, prognosis and treatment targets. Only one published study has described relationship of DNA methylation with surgical debulking outcome specifically. This study (Fiegel et al., 2008) using MethyLight assay of fresh-frozen EOC tumours demonstrated that HOXA11 methylation levels were found to be significantly different between tumours that had <2 cm residual disease vs >2 cm residual disease (\( n = 92, P = 0.002 \)). Again, these studies arise from heterogeneous tumour types with early and late stage disease and variable histopathological grades (including borderline tumours).

Validation through TCGA data set

Odds ratios were calculated from a sample of the included studies (Fiegel et al., 2008; An et al., 2009; Li et al., 2011) and were found to be 0.47, 0.13 and 0.23, respectively. These looked at the associations between having low levels (dichotomised at the median) of IGF1R, HOXA11 or AEG1, respectively, with sub-optimal debulking (as compared with optimal debulking). Using an \( \alpha \) of 0.05 and the TCGA sample size of 279, the power to detect previously seen effects in the TCGA data was 0.69, 0.99 and 0.98, respectively. Although it is possible that lack of validation could occur by chance, based on these examples we believe that the TCGA validation set was adequately powered to detect effects of the size seen in previously published studies.

We aimed to validate the published data on candidate genes and proteins using the TCGA data set. Data were available on 279 patients with stage 3 or 4 serous EOC who had details of residual disease volume. In all, 274 (98.2%) patients had high-grade disease. In all, 36 gene expression probes were found to cover the individual gene and proteins previously identified as being significantly associated with surgical debulking (Supplementary Tables 1 and 2). There was no coverage of the genes SRA or BCAR1 (encodes protein P130cas). The list of 120 genes significantly associated with debulking status found by Berchuck et al. was not validated as the gene expression probe IDs were not made publically available. Using individual generalised linear models adjusting for microarray batch, age and stage and adjusting for multiple comparisons using FDR, two probes; MTDH (probe ID ‘212250_at’) and IGF1R (probe ID ‘208441_at’) were found to be significantly differentially expressed (\( P < 0.05, FDR < 5\% \)) associated with optimal (<1 cm residual disease) or suboptimal (>1 cm residual disease) surgery. MTDH, which transcribes AEG-1 protein, demonstrated increased gene expression in those suboptimally debulked compared with those optimally debulked with an increased log expression level of 0.24 (\( P = 0.002, FDR 3.7\%, 95\% confidence interval (CI) 0.088, 0.392 \)). This correlates with the previous data that found increased protein expression of AEG-1 was associated with suboptimal debulking (Li et al., 2011). IGF1R expression was found to be significantly decreased in those with suboptimally debulked compared with those optimally debulked with a log expression level of 0.073 (\( P = 0.0004, FDR 1.4\%, CI = -0.112, -0.033 \)). This contrasts to previously reported data (An et al., 2009) that found a significant association with increased gene expression of IGF1R and suboptimal debulking (>1 cm residual disease).

Recent emphasis has been placed on achieving total macroscopic surgical debulking (microscopic residual disease) with clear survival benefits (Winter et al., 2008). Therefore, the analysis was repeated with surgical debulking groups defined as either microscopic (0 mm) residual disease or \( \geq 1 \) mm macroscopic residual disease. In this analysis, all three MTDH probes (probe ID ‘212251_at’, ‘212250_at’ and ‘212248_at’) were significantly associated with residual disease of \( \geq 1 \) mm (\( P < 0.05, FDR < 5\% \)) with their increased log expression (\( \beta = 0.414, P = 1.02 \times 10^{-5} \), \( \beta = 0.226, P = 0.005 \), and \( \beta = 0.104, P = 0.076 \), respectively).
Clinical Studies

emphasis is now placed in achieving total macroscopic debulking (Fiegl et al., 2008). Using individualised linear models adjusting for microarray batch, age and stage there was no significant difference in methylation between the optimal and suboptimal groups ($P = 0.681$) or between those with $\leq 2$ cm residual disease or $> 2$ cm residual disease ($P = 0.559$).

DISCUSSION

Significant differential expression at $P < 0.05$ and FDR $< 10\%$ was found only in 2 out of 36 available gene expression probes between optimal and suboptimal groups in the TCGA data set. Although it is possible that lack of validation could occur by chance, this is highly unlikely for most markers in TCGA where the power of the validation analysis is expected to be $> 0.9$. Only one of the significant probes, MTDH (also known as AEG-1) was found to correlate with the previous published study. Uregulation of MTDH has been found to inhibit apoptosis and increase the invasive ability of malignant cells (Emdad et al., 2007; Liu et al., 2010) and is involved in angiogenesis (Emdad et al., 2009). Despite this, it is apparent that there is still no clear well-validated evidence for a role of tumour biology on surgical success in the treatment for EOC. We would argue however that this has been due to limitations of study design. Previous studies have been heterogeneous in design without fully fulfilling REMARK (McShane et al., 2005) criteria. The REMARK report stated the importance of transparent and complete reporting in the publication of research on tumour markers, highlighting the need for clear descriptions of patient and specimen characteristics, assay methods, study design and statistical analysis.

In the majority of previous studies, early FIGO stage disease has been included in analysis. By its very definition, of disease confined to the ovary or pelvis (FIGO, 1987), stage 1 and 2 disease can always or almost always be completely excised. Consequently, a much higher proportion of optimally or total macroscopic debulked tumours will occur in those with early stage disease (Wimberger et al., 2007). This is not adjusted for in the majority of analysis and thus any differences may more be by association with stage and metastasis rather than surgical outcomes. It is now agreed that EOC is not a single disease, but a general term applied to a distinct set of cancers that share an anatomical location (Vuong et al., 2007). There are even distinct diseases within the same histological subtype, a high-grade serous tumours being clinically and molecularly distinct from a low-grade serous tumour (Bowtell, 2010). Thus, the majority of previous studies have extremely heterogeneous samples included. Interestingly, the study with the most homogenous of tumour samples; high-grade, high-stage serous EOC (Bonome et al., 2008) did not find any statistical significant differences in gene expression microarray between patients who were optimally and suboptimally debulked.

There is also discrepancy between studies when defining surgical groups, with suboptimal debulking described as residual disease $> 1$ or $2$ cm. Furthermore, there may be significant measurement error with regards to tumour estimation (Préfontaine et al., 1994) leading to bias in analysis. Much emphasis is now placed in achieving total macroscopic debulking (microscopic residual disease) with clear survival benefits (Winter et al., 2008). We propose that this would be a more suitable classification for molecular studies because a category of presence or absence of macroscopic residual disease is likely to be less susceptible to errors in measurement. This may explain the addition of two further MTDH gene probes in TCGA analysis, which were found to be significantly differentially expressed when surgical groups were defined as macroscopic residual disease or $\geq 1$ mm macroscopic residual disease.

It is clear that there are many clinical factors that will impact surgical success. Debunking rates are known to differ dependent on surgical expertise, institution and country (Nguyen et al., 1993; Crawford et al., 2005; Skirnisdottir and Sorbe, 2007). Personal and institutional philosophy on what accounts as unresectable disease also differs (Chi and Schwartz, 2008) and will undoubtedly affect debulking rates. It has been shown that the incorporation of extensive upper abdominal procedures such as diaphragm peritoneectomy, splenectomy and partial liver resection increased successful cytoreductive outcomes and consequently significantly improves survival (Chi et al., 2009). Thus, detailed descriptions on surgical effort involved to achieve debunking outcomes should also be considered in analysis. Surgical effort could be simply quantified through a grid-scoring system similar to a previously published ranking system for measurement of metastatic deposits before surgery (Eisenkop et al., 2003; Fotopoulou et al., 2010). The abdomen divided into anatomical segments, for example, right upper quadrant (diaphragm, liver, adjacent peritoneal surfaces), left upper quadrant (spleen, stomach, transverse colon and splenic flexure), pelvis (including recto-sigmoid), etc., with a numerical score assigned dependent on surgical resection at that site, a higher score given for resection of major organs as compared with excision of small nodules or plaques. Anecdotally, there are also clinically heterogeneous tumours described at surgery. For instance, some tumours present as widespread miliary disease with small deposits dispersed throughout the abdomen whereas others are more bulky with large solitary deposits at specific sites. This pattern of disease spread is likely to affect debulking outcomes in the majority of departments and thus should also be recorded and included in multivariate analysis. Additionally, patient’s performance status, age and comorbidity can also influence surgical outcome (Wimberger et al., 2007). When all this is considered it appears naive to attribute a molecular signature to a surgical outcome without detailed knowledge of these important clinical characteristics.

SUMMARY

The relationship between surgical outcome and tumour biology is complex and remains unclear. To date, there are no specific predictive models for surgical success that are clinically useful and the majority of previous studies have limitations in design making their interpretation difficult. We suggest that further effort is made into collecting detailed clinical and surgical information in ovarian cancer to allow for important clinical characteristics that are known to affect surgical outcomes. This should include patient performance status, accurate sizing and location of disease before and after surgery and the extent of surgical effort performed.

Supplementary Information accompanies the paper on British Journal of Cancer website (http://www.nature.com/bjc)

REFERENCES


This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.