121. Charboneau, A., East, L., Mulholland, N., Rohde, M. & Boudreau, N. Pbx1 is required for Hox D3-mediated angiogenesis. *Angiogenesis* **8**, 289–296 (2005).
122. Shen, W., Chrobak, D., Krishnan, K., Lawrence, H. J. & Largman, C. HOXB6 protein is bound to CREB-binding protein and represses globin expression in a DNA binding-dependent, PBX interaction-independent process. *J. Biol. Chem.* **279**, 39895–39904 (2004).
123. Li, X., Nie, S., Chang, C., Qiu, T. & Cao, X. Smads oppose Hox transcriptional activities. *Exp. Cell Res.* **312**, 854–864 (2006).
124. Abe, M. *et al.* Disordered expression of HOX genes in human non-small cell lung cancer. *Oncol. Rep.* **15**, 797–802 (2006).
125. Alami, Y., Castronovo, V., Belotti, D., Flagiello, D. & Clausse, N. HOXC5 and HOXC8 expression are selectively turned on in human cervical cancer cells compared to normal keratinocytes. *Biochem. Biophys. Res. Commun.* **257**, 738–745 (1999).
126. Lewis, E. B. Pseudoallelism and gene evolution. *Cold Spring Harb. Symp. Quant. Biol.* **16**, 159–174 (1951).
127. Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
128. Kaufman, T. C., Lewis, R. & Wakimoto, B. Cytogenetic Analysis of Chromosome 3 in DROSOPHILA MELANOGASTER: The Homoeotic Gene Complex in Polytene Chromosome Interval 84a-B. *Genetics* **94**, 115–133 (1980).
129. Scott, M. P. *et al.* The molecular organization of the Antennapedia locus of *Drosophila*. *Cell* **35**, 763–776 (1983).
130. Garber, R. L., Kuroiwa, A. & Gehring, W. J. Genomic and cDNA clones of the homeotic locus Antennapedia in Drosophila. *EMBO J.* **2**, 2027–2036 (1983).
131. Duboule, D., Baron, A., Mahl, P. & Galliot, B. A new homeo-box is present in overlapping cosmid clones which define the mouse Hox-1 locus. *EMBO J.* **5**, 1973–1980 (1986).
132. Hart, C. P., Awgulewitsch, A., Fainsod, A., McGinnis, W. & Ruddle, F. H. Homeo box gene complex on mouse chromosome 11: molecular cloning, expression in embryogenesis, and homology to a human homeo box locus. *Cell* **43**, 9–18 (1985).
133. Lonai, P., Arman, E., Czosnek, H., Ruddle, F. H. & Blatt, C. New murine homeoboxes: structure, chromosomal assignment, and differential expression in adult erythropoiesis. *DNA* **6**, 409–418 (1987).
134. Kongsuwan, K., Webb, E., Housiaux, P. & Adams, J. M. Expression of multiple homeobox genes within diverse mammalian haemopoietic lineages. *EMBO J.* **7**, 2131–2138 (1988).
135. Kamps, M. P., Murre, C., Sun, X. H. & Baltimore, D. A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. *Cell* **60**, 547–555 (1990).
136. Nourse, J. *et al.* Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. *Cell* **60**, 535–545 (1990).
137. Zeltser, L., Desplan, C. & Heintz, N. Hoxb-13: a new Hox gene in a distant region of the HOXB cluster maintains colinearity. *Development* **122**, 2475–2484 (1996).
138. Biondi, A., Cimino, G., Pieters, R. & Pui, C. H. Biological and therapeutic aspects of infant leukemia. *Blood* **96**, 24–33 (2000).
139. Yekta, S., Shih, I. H. & Bartel, D. P. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594–596 (2004).

**Competing interests statement**
The authors declare no competing financial interests.

**DATABASES**

**Entrez Gene:** http://www.ncbi.nlm.nih.gov/gene
*AFF1* | *Erbb2* | *EOS* | *HOTAIR* | *HOXA3* | *HOXA4* | *HOXA5* | *HOXA7* | *HOXA9* | *HOXB3* | *HOXB9* | *HOXB13* | *HOXC6* | *HOXD1* | *HOXD8* | *HOXD10* | *MLL* | *MLLT3* | *MLLT10* | *TP53* | *trithorax*

**National Cancer Institute Drug Dictionary:** http://www.cancer.gov/drugdictionary/Tamoxifen

**UniProtKB:** http://www.uniprot.org
AR | Caspase 2 | Caspase 8 | CHDH | ER | HOXA10 | HOXA11 | HOXB7 | HOXC4 | HOXC6 | HOXC8 | HOXC9 | HOXC11 | IL17RB | NUP98 | PARP1 | PBX1 | S100β

**ALL LINKS ARE ACTIVE IN THE ONLINE PDF**

**OPINION**

# Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer

*Vathany Kulasingam, Maria P. Pavlou and Eleftherios P. Diamandis*

Abstract | Despite widespread interest, few serum biomarkers have been introduced to the clinic over the past 20 years. Each approach to ovarian cancer biomarker discovery has its own advantages and disadvantages and it seems likely that a global biomarker discovery platform that mines all possible sources for biomarkers might be more useful. Such data could be combined with information from relevant microarray data, bioinformatic analyses and literature searches. This proposed integrated systems biology approach has the potential to yield promising ovarian cancer markers for diagnosis, prognosis and monitoring of patients during therapy.

Ovarian cancer occurs in 1 of 2,500 post-menopausal women in the United States and is the most lethal gynaecological malignancy, accounting for 5–6% of all cancer-related deaths. When ovarian cancer is diagnosed at early stages, the survival rate is close to 90%[1]; however, the vast majority of patients are identified when they have late-stage disease. This is primarily because ovarian cancer has few early or specific symptoms. Patients diagnosed with advanced disease are managed with surgical cytoreduction and chemotherapy, but many experience resistance to chemotherapy and relapse, yielding an overall 5-year survival rate of 10–30%[2,3].

One of the best ways to diagnose cancer early, aid prognosis and predict therapeutic response is by using diagnostic or prognostic serum and tissue biomarkers. A biomarker, according to the US National Cancer Institute, is a biological molecule found in blood, another body fluid or in tissues that is a sign of a normal or abnormal process. Generally, biomarkers are produced by either the tumour itself or other tissues, in response to the presence of cancer or other associated conditions. Tumour markers can be used for screening the general population, differential diagnosis in symptomatic patients and clinical staging of cancer. Additionally, they can be used to estimate the tumour volume, evaluate response to treatment, assess recurrence through monitoring and as prognostic indicators for disease progression[4]. Unfortunately, there are not many reliable serum biomarkers currently used in the clinic, and tissue-based markers require an invasive procedure to obtain samples for diagnostic purposes. The most studied marker for ovarian cancer is CA125 and determination of its concentration in circulation is essential for monitoring response to treatment for ovarian cancer. It has also been proposed as a possible screening test for this disease[5,6]. However, this marker has low sensitivity, as its expression is increased in fewer than 50% of early-stage ovarian cancers and it is not expressed by tumour cells in 20% of women diagnosed with ovarian cancer. It also demonstrates low specificity, as its expression is increased in many benign gynaecological diseases, such as endometriosis[7], and it is also expressed outside of the female genital tract in tissues such as lung, breast and prostate.

Successful screening strategies for ovarian cancer must demonstrate a sensitivity of 75% and a specificity of ≥99.6% to obtain a positive predictive value (PPV) of 10%[8]. A PPV of 10% equates to a situation in which only 1 of 10 surgical interventions leads to the diagnosis of ovarian cancer. Currently, there are no screening strategies with proven efficacy for the early detection of ovarian cancer, although there are several ovarian cancer screening trials currently underway that are based on transvaginal ultrasound, or serum concentration of CA125 combined with transvaginal ultrasound as part of a multimodal screening strategy[9]. The multimodal screening strategy does not provide optimal sensitivity for early detection but it does exhibit adequate specificity compared with annual screening with transvaginal ultrasound[9]. Therefore, new efforts have emerged to identify new serum markers to aid in the screening process. A major challenge is

finding a marker or a group of markers that will not only detect clinically evident ovarian cancer but also early disease, before it causes symptoms. Clearly, a panel of biomarkers would increase the sensitivity and specificity for diagnosis compared with a single marker. For example, a recent study examined the ability of six serum biomarkers to discriminate between disease-free individuals and patients with ovarian cancer[10]. Using a multiplex, bead-based immunoassay system, the combination of the six markers (leptin, prolactin, osteopontin, insulin-like growth factor 2, macrophage inhibitory factor and CA125) exhibited better discrimination compared with CA125 alone (95.3% sensitivity with 99.4% specificity).

The past decade has witnessed an impressive growth in the field of large-scale and high-throughput biology, which is attributed to an era of new technological development. Most of the proteomic technology platforms for biomarker discovery focus on the implementation of mass spectrometry techniques (BOX 1). A renewed interest from the proteomics and genomics communities in discovering new cancer biomarkers by using mass spectrometry and microarrays has emerged. However, few biomarkers have taken the path to the patient bedside. One of the major roadblocks is biomarker specificity. Although many potential markers with acceptable sensitivity have recently been identified using proteomic technologies, most lack sufficient specificity to be useful (that is, they are found to be nonspecific markers of inflammation, anaemia, cachexia, malnutrition or malabsorption, infections, angiogenesis and other pathologies that are associated with but are not specific to cancer). Given the lack of specificity of potential candidates that have been discovered by individual methods, the hypothesis is that combining high-throughput strategies might facilitate the delivery of more effective candidate molecules for cancer diagnosis and prognosis.

In this Opinion article, five phases for biomarker identification and development are discussed along with the limitations of many of the current discovery-based experiments. We use ovarian cancer as an example to highlight the advantages and disadvantages of different techniques that are used for general biomarker discovery and propose an integrated systems biology strategy for new marker identification. The development and validation of biomarkers for specific uses such as diagnosis, prognosis and prediction are not covered. The general biomarkers that emerge from this integrated approach could be instrumental in substantially reducing the burden of cancer through prevention, early diagnosis, individualized therapies and improved monitoring post-treatment.

## Limitations of the biomarker pipeline

Conceptually, there are five phases of biomarker development, which include a preclinical exploratory phase, clinical assay and validation stage, retrospective longitudinal study, prospective screening evaluation and randomized control trials[11]. In Phase 1, tumour and non-tumour specimens are compared to generate hypotheses for clinical tests for detecting cancer. Strategies such as gene expression profiling, mass spectrometry-based methods, as well as other means to biomarker discovery can be used to aid this phase. In Phase 2, a clinical assay is established using a specimen of choice (usually something that can be obtained non-invasively). The subjects assayed in this phase have established disease. Whether the assay can detect early disease is not addressed in this phase. In the retrospective longitudinal clinical repository studies phase (Phase 3), specimens collected and stored from a cohort of healthy subjects who were monitored for the development of cancer are used. Evidence for the capacity of the biomarker to detect preclinical disease is demonstrated in this phase. Criteria for 'positive' screening results are defined and used in Phase 4, which consists of prospective screening studies. In this phase, subjects are screened using the assay, and diagnostic procedures are applied to those who have screened positive, which enables the stage or nature of the disease at which the assay is most effective to be established. Finally, the objective of Phase 5 is to determine whether screening has reduced the burden of cancer on the population through randomized control trials.

The current biomarker discovery efforts (Phase 1) are highly variable, not only in methods of marker identification, but also in study design and patient selection. Interpatient heterogeneity and intra-tumour heterogeneity are important confounding factors for some sources. In addition, the danger of bias and the problems of overfitting the data, as well as the handling and storing of clinical specimens, are vital factors that need consideration before a study is conducted[12]. However, the most stringent criteria for sample size (statistical power), sample collection and accurate data analysis need to be applied in the later phases of biomarker development. For example, the clinical specimens to be used in Phase 2 (once candidate molecules have been selected to investigate further) should follow the prospective specimen collection retrospective blinded evaluation (PRoBE) study design[13]. The basic premise of the PRoBE study design is that specimens are collected prospectively from a clinically relevant and well-defined cohort in the absence of knowledge about patient outcome. After outcome status is known, cases and controls are randomly selected from the cohort and specimens are assayed for the biomarkers in a fashion that is blinded to case–control status. Therefore, one must clearly distinguish between studies that are looking for new biomarkers for disease (discovery-based platforms or Phase 1 as described above and which form the focus of this article) and studies seeking to validate a new biomarker (Phases 2–5). Phase 1 studies are not as dependent on the number of samples but are more dependent on the depth of analysis and efficient selection of the most promising molecules from the thousands of possible candidates. Certainly, new tumour marker tests — single or multiparametric — must undergo rigorous validation to assess their clinical value in Phases 2–5.

## Sources to mine for biomarker discovery

In the following section, six frequently used sources for identifying potential cancer biomarkers with an appropriate technology-based discovery platform are examined. For each source, the advantages and disadvantages, as well as a few examples of how this technology has been applied to discovering tumour markers for ovarian cancer, are discussed.

***Blood.*** Blood is the most commonly used biological fluid for biomarker analysis in clinical practice. The advantages of using blood, serum and plasma as a source to mine for biomarkers include that it can be obtained through a minimally invasive procedure, it is abundantly available and some constituents of blood reflect diverse pathological states[14]. However, the complexities of the plasma proteome far exceed the current capabilities of mass spectrometry in resolving the individual peptides of trypsinized plasma in a single analysis (BOX 1). It is known that the plasma proteins range in concentration over 12 orders of magnitude and that 99% of the protein mass is comprised of only 22 proteins[15]. This large dynamic range of analytes in blood is a major disadvantage for using this source for biomarker discovery. The removal of predominant proteins facilitates better detection of less abundant proteins, but such depletion can lead to the loss of informative molecules. In addition, although blood is supplied to all organs of the body, raising the possibility of

Box 1 | **The promise of mass spectrometry**

The discoverer of the electron Sir J. J. Thomson (who received the Nobel Prize in Physics in 1906) constructed the first mass spectrometer (then called a parabola spectrograph). The use of mass spectrometry (MS) was restricted to low-molecular-weight compounds that were thermally stable owing to the inability to effectively transfer the ionized molecules from the condensed phase to the gas phase without excessive fragmentation. The introduction in the late 1980s of two techniques (electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI)) allowed for large, non-volatile and thermally labile compounds to be converted into gas-phase ions. The efforts related to the development of both ESI (by John Fenn) and MALDI (by Koichi Tanaka) led to a shared 2002 Nobel Prize in Chemistry. These soft ionization techniques enabled the study of polypeptides and the trend towards MS as the technique of choice for identifying and probing biological proteins of interest was accelerated by the genome project. Today, MS is maturing as a powerful analytical tool for biomarker discovery.

New and powerful mass analysers with complex multistage capabilities and excellent sensitivity, accuracy and resolution have emerged to tackle the challenges of protein and proteome analysis. Routine experiments involving MS include tandem MS (MS/MS) whereby detailed structural features of the peptides can be inferred from the analysis of the masses of the resulting fragments. This facilitated the identification of whole proteomes of organisms and to the identification of thousands of proteins in any given experiment. MS is also being used to identify post-translational modifications, although there are some inherent limitations of MS for detecting high-molecular-weight heavily glycosylated proteins.

In a typical proteomic experiment, it is preferable to first enzymatically convert the proteins to peptides (usually by trypsin digestion). Ionized peptides with characteristic mass-to-charge (m/z) ratios are isolated in the mass spectrometer and further fragmented, usually by collision with inert gas molecules, into smaller product ions — a process referred to as collision-induced dissociation (CID). The m/z of the product ions is further monitored and the resulting mass spectra provide information pertaining to the amino acid sequence of the peptide. Using pattern matching algorithms and probability-based scoring methods, the sequence of a peptide can be identified on the basis of the MS/MS spectra and database searching.

encompassing a whole range of different biomarkers, these could potentially be less specific than those identified in tumour tissue, for example. With respect to sample preparation, blood can pose reproducibility problems because it contains active proteases, lipids, and other components that can vary with diet and lifestyle.

Despite its limitations, several investigators have used blood for the discovery of biomarkers for ovarian cancer. For example, proteomic-pattern profiling technology has received considerable attention in recent years. With this technology, a small amount of unfractionated serum is added to a protein chip, which is subsequently analysed by surface-enhanced laser-desorption ionization time-of-flight mass spectrometry (SELDI-TOF-MS) to generate a proteomic signature (BOX 1). The potential of proteomic pattern analysis was first demonstrated for the diagnosis of ovarian cancer, with a sensitivity that was claimed to be 100% (even for early-stage disease) with 95% specificity[16]. However, subsequent work concluded that these data were erroneous owing to biases in sample collection and handling, and bioinformatic artefacts[12,17–20]. Criticism of this platform includes the lack of identification of key peaks, difficulty in reproduction of the results and lack of incorporation of existing ovarian cancer biomarkers. More recent studies using SELDI-TOF-MS take into account the potential limitations and have taken precautionary steps towards identification and validation of the profiles identified, including identification of the key peaks and carrying out the validation studies using a different technology platform[21,22]. For a discussion on the various blood-based studies on ovarian cancer, see the recent review by Nossov *et al.*[23].

Another study, using serum, examined three biomarkers (apolipoprotein A1, a truncated form of transthyretin and a cleavage fragment of inter-α-trypsin inhibitor heavy chain H4) that were obtained from a multicentre serum proteomic expression analysis, for their diagnostic potential in ovarian cancer[24]. The 3 markers, along with CA125, had a sensitivity of 74% (compared with 65% for CA125 alone) and a specificity of 97%. The samples were analysed by SELDI-TOF MS, but the identified biomarkers are acute-phase reactants[25]. As such, the changes seen are epiphenomena that are likely to prove to be nonspecific for a particular type of cancer. Some new strategies emerging in the field of blood biomarker discovery include the study of other blood constituents such as peripheral blood mononuclear cells, platelets and plasma circulating tumour cells and/or DNA. Cell-free DNA has been investigated for quantity, fragmentation pattern and tumour-specific sequences in patients with various malignancies, including ovarian cancer, with the widespread availability of quantitative real-time PCR technology[26]. Although the origin of circulating DNA is largely unknown, it is thought that a small fraction of the DNA could be from the tumour itself[27]. As such, the rationale is that the detection of increased DNA levels and tumour-specific DNA sequences in blood could provide a non-invasive means to obtain diagnostic information. Some of the issues with cell-free DNA in blood include loss of DNA during purification procedures, delayed blood processing and overall DNA integrity.

***Other relevant biological fluids.*** Exploring biological fluids proximal to tumours is an attractive strategy of identifying tumour-secreted proteins. Ovarian cancer ascites contain many cells of tumour origin, in addition to many soluble factors released by tumour cells or the tumour microenvironment that have been associated with invasion and metastasis[28]. The advantages of mining this source include its availability in large quantities (litres) and the ability to study the ovarian cancer cell secretome in the context of the tumour microenvironment after ultrafiltration. A major disadvantage is contamination by highly abundant serum proteins[29]. Obtaining non-malignant ascites for comparison (such as from patients with cirrhosis of the liver) is feasible. The accumulation of fluid in the peritoneal cavity usually occurs in advanced disease; therefore, disease-specific proteins from this fluid might not be suitable for diagnosing early-stage disease.

A recent study by Gortzak-Uzan *et al.*[30] used four patients with ovarian cancer who had high-grade serous carcinoma to carry out an in-depth proteomic analysis of ovarian cancer ascites. Although the authors identified more than 2,500 proteins, only ~230 were identified in the soluble fraction (after removing cells). By combining their data with available body fluid and microarray data sets, they produced a list of 80 potential biomarkers. However, many of their preferred candidates were intracellular proteins that have not yet been validated, and CA125 was identified in one patient sample only, which raises the question of whether this approach to biomarker discovery will bear fruit.

***Tumour tissues.*** Surgically removed or biopsy-obtained tissues are currently being considered as alternative sources for biomarker discovery. The rationale is that proteins originating from tissue could subsequently enter, and be measured in, the bloodstream. Leaky capillary beds, the local production

of proteases and the high rates of cell death in the tumour mass are expected to facilitate shedding or secretion of tumour-associated proteins into the bloodstream. One of the major advantages of using tissues is that the concentration of candidate biomarkers should be highest in tumour tissues and they should be a rich source for plasma biomarkers[31]. However, few tissue biomarkers have proven useful for serum analyses[32]. This could be due to differences in the dynamics of release and clearance of proteins from the circulation. For example, some differentially expressed proteins at the tissue level might be degraded by endogenous proteases or could be removed by the kidneys leading to undetectable levels in circulation. Laser capture microdissection (LCM) could improve the specificity of tissue biomarker discovery as it provides a means of extracting pure cell populations from the surrounding heterogeneous tissue, which enriches for the proteome of interest. Without LCM, there is much heterogeneity in the cellular and extracellular composition of tissues. Therefore, ovarian tumour cells could constitute a minor fraction of the whole cell population. It is known that tumour tissues contain various types of non-tumour cells, and vascular structures in tumour tissues can contain a large amount of plasma proteins. However, LCM is a labour-intensive process, requires fresh frozen tissues and expert users and yields low numbers of cells[31]. Tissues can also be difficult to obtain in sufficient quantities, especially normal counterparts for comparative analyses. To overcome this, some studies have used low malignant potential (LMP) ovarian tumours as controls because they have similar cytological features to invasive ovarian cancers but lack the capacity to invade[33]. Despite the disadvantages, the approach to using tissues as a source for biomarker discovery is popular. For example, Bengtsson *et al.*[34] used two-dimensional differential in gel electrophoresis (2D-DIGE) to decipher the proteome of 64 tissue samples, representing all stages of ovarian cancer. They identified ~220 proteins that were differentially expressed among normal, benign and malignant tumours, which could potentially be used to discriminate between benign and malignant tumours.

***Human cancer cell lines.*** Secreted proteins and membrane proteins shed from tumour cells are promising cancer biomarkers. The hypothesis is that conditioned media from cancer cell lines contains secreted proteins or proteins that are shed from the plasma membrane that might be found in the circulation of cancer patients. The advantages of such an approach are that a large number of cell lines representing various stages and histotypes are readily available and the analysis by mass spectrometry is straightforward, reproducible and in-depth, without the problems that are associated with high-abundance proteins. The limitations are that no single cell line can recapitulate the heterogeneity of human tumours and that it is an *in vitro* system, devoid of contributions of the host-tumour microenvironment. A recent review outlines in further detail the advantages and disadvantages of using a cell culture-based biomarker discovery platform[35]. An example of this approach is a study by Faca *et al.*[36] who used intact protein fractionation followed by trypsin digestion and liquid chromatography-MS/MS analysis to examine the cell-surface proteome and extracellular milieu of three ovarian cancer cell lines (CaOV3 and OVCAR3, which are a serous histotype, and ES2, which is a clear-cell histotype). They isolated more than 5,000 proteins, of which 3,300 were found in the conditioned media. The authors established an overlap between the proteins identified in these cell lines and the ones expressed by cancer cells isolated from ascites fluid. This extensive database should be a valuable source for future biomarker discovery.

***Animal models.*** Animal models have been used extensively in therapeutics but rarely in diagnostics. Some have argued that mouse models are a useful and underused resource for cancer biomarker discovery and validation[37]. The advantages of using an animal model to identify putative biomarkers include the fact that it is an *in vivo* model, which incorporates host–tumour interactions. Given the importance that host-tumour microenvironments have in cancer initiation and progression, the use of an *in vivo* system seems appealing. Furthermore, it is possible to implant xenografts of human cancer tissues into mice but recognizing that crucial features of the tumour microenvironment are altered in such cases, including the loss of nearby normal human tissues, vasculature and so on, and that tumour xenografts represent just one constituent of the tumour[38]. Inbred mouse models of cancer recapitulate many crucial features of human cancer[39]. Reduction in intra-individual variability can be achieved by matched cases and controls, in terms of genetic variability and environmental conditions. The mouse plasma proteome is as complex as the human[37], but in mice, blood specimens can be collected at any stage, before and during tumour development. The large ratio of tumour to plasma in mice, compared with humans, results in an increased concentration of potential biomarkers in plasma, making their detection easier. There is now a repository of tissues from ten different mouse models of human cancer[37]. However, there are many limitations of using animal models for biomarker discovery, such as it is unclear whether the same genetic alterations transform both mouse and human cells, and examining tissues and/or biological fluids in rodents has the same limitations as in humans. Nonetheless, genetically engineered mouse models of cancer are available and the advantages of such a strategy include the ability to reversibly control target gene expression with exogenous ligands[38]. A recent review outlines the advantages and disadvantages of the various mouse cancer models[38]. To our knowledge, only one animal model proteome has been deciphered for the identification of ovarian cancer biomarkers[40]. In this study, plasma from control and ovarian cancer-bearing mice were subjected to in-depth proteomic analysis to yield ~100 proteins that were increased in the plasma of tumour-bearing mice. Comparison of these proteins with those identified from a proteomic analysis of ovarian cancer cell lines and ascites fluid demonstrated that approximately half of these proteins were present in the mouse plasma proteome, including human epididymis protein 4 (HE4; also known as WFDC2) — a known ovarian cancer biomarker[40].

***Microarray profiling.*** Microarray technology has been used to compare gene expression profiles in ovarian cancers and normal ovaries. The aim is to identify genes that are differentially expressed between the two states, with the expectation that similar patterns could be seen for the respective proteins in serum[2]. Several studies have attempted to identify new molecular biomarkers for the early detection of ovarian cancer by gene expression profiling[8,41,42]. The advantages of this approach include high throughput and objective molecular subclassification. Gene expression levels reflect the cumulative effect of several underlying biological functions as DNA-microarray technology has enabled the simultaneous examination of thousands of genes, in contrast to studying the expression of single genes. Current microarray platforms are highly automated and enable parallel sample analysis. However, they lack the ability to identify protein expression levels and have considerable variability, use small sample sizes in many studies and yield different results depending on the statistical analysis used[43]. Tissue processing is not standardized, leading to the possibility of the contribution of surrounding non-tumour tissues to gene

expression profiles. Gene array studies also lack appropriate 'normal' or control specimens. The use of cultured surface epithelium is not an adequate control specimen.

Several studies have used microarray analysis to identify gene expression profiles that are associated with ovarian cancer. However, the gene signatures identified are often not overlapping. Spentzos *et al.*[44] used this technology to develop an ovarian cancer prognostic profile consisting of 115 genes. A new ovarian cancer biomarker discovered using cDNA microarrays is HE4. Quantification of HE4 in serum for diagnosis and monitoring is promising, and this marker was recently approved by the US Food and Drug Administration for monitoring patients with ovarian cancer[7,45]. In another study, oligonucleotide microarray analysis on fresh-frozen samples from ~60 patients with ovarian cancer was used to identify genes encoding proteins with evidence of secretion on the basis of an algorithm that was developed using results from previous studies[8]. Of the ~22,000 probe sets on the array, ~1,400 showed evidence of encoding secreted proteins. In serous ovarian carcinomas 275 of these were identified as differentially expressed (both overexpressed and underexpressed). Enzyme-linked immunosorbent assay (ELISA) was used to validate the findings from microarray analysis using an independent validation serum sample set consisting of 67 serous ovarian carcinomas, 67 healthy women and 15 patients with cystadenofibromas of the ovary. When combined with CA125 four proteins, matrix metallopeptidase 7 (MMP7), kallikrein 10 (KLK10), osteopontin (OPN) and secretory leukocyte peptidase inhibitor (SLPI), yielded a sensitivity of 95.7% and 100% specificity[8]. However, a major limitation of this study was that the number of patients with early-stage ovarian cancer was low ($n = 7$ for stage 1 and $n = 2$ for stage 2) and the distinction between early stage and the 15 patients with benign disease was not optimal.

## An integrated discovery platform

For the past 10 years the expectation has been that the completion of several genome sequencing projects, the discovery of new oncogenes and tumour-suppressor genes, and recent advances in genomic and proteomic technologies, together with powerful bioinformatic tools, would have a direct and major impact on the way new cancer biomarkers are discovered. This gave rise to the premise that an exciting era of biomarker discovery was just about to begin. However, despite the wealth of technological and bioinformatic advances, the anticipated new diagnostic tools have been slow to emerge. One reason is that the concentration of biomarkers in serum and/or biological fluids is often too low (from ng per ml to pg per ml) and cannot be measured directly by the principal enabling technology of proteomic discovery — mass spectrometry (BOX 1), unless specific immunological reagents and highly sensitive ELISA methods are also available. Potential new tumour markers are expected to exist in the low ng–pg per ml concentration range. Therefore, in the initial discovery phase for new cancer biomarkers, a less complex sample (that is, one devoid of high abundance proteins) is essential. Some strategies to mine deeper into complex samples such as serum include multiple fractionation and immunoaffinity purification. However, the trade-off for in-depth coverage is low throughput.

As ovarian cancer is a complex and heterogenous disease, no single model or biological material (tissue, fluid or cell lines) is expected to emulate all aspects of the disease[46]. For this reason, an effective approach to new biomarker development should be well-conceived and play to the strengths of current technologies, while acknowledging and addressing the limitations[47]. Integrating several databases to mine, may complement proteomic analysis and aid in the identification of putative ovarian cancer biomarkers. A more effective biomarker discovery platform should focus on mining all possible sources (FIG. 1), indeed a disadvantage of one approach could be the advantage of another. A new framework for the identification and validation of markers should include not only different technologies, but also the integration of information on molecular and pathophysiological pathways. Combining data sets from these sources may prove to be a worthy investment. In the following sections we illustrate, using ovarian cancer as an example, how an integrated approach may yield promising cancer biomarkers.

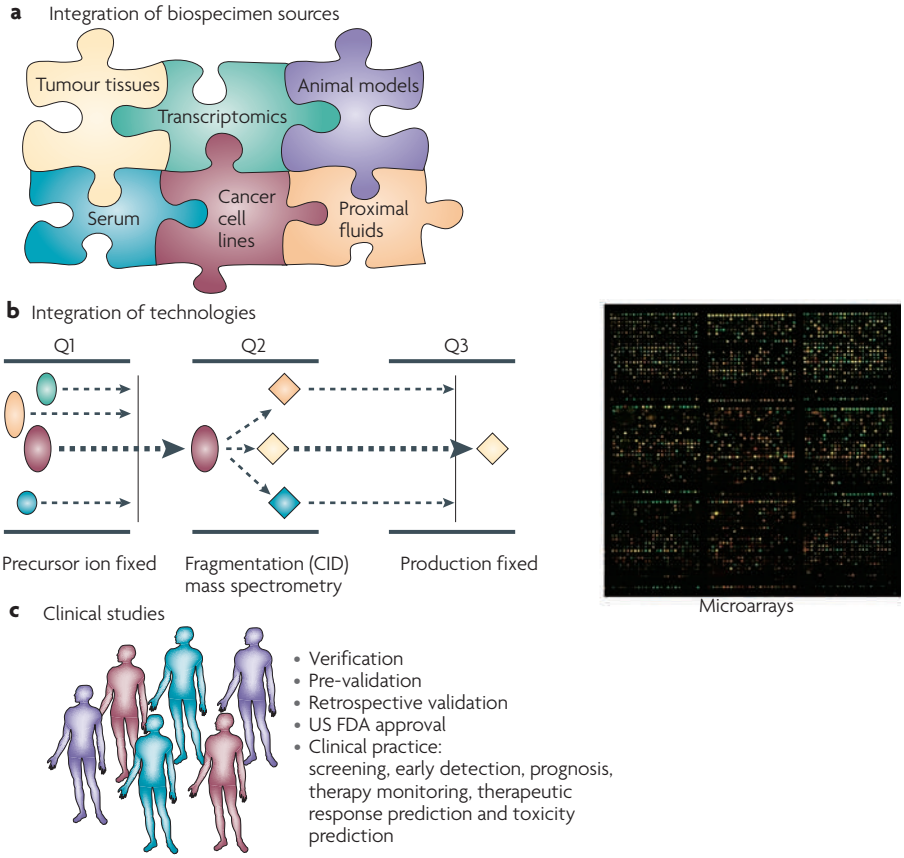


Figure 1 | **Integrated systems biology biomarker discovery platform. a** | This platform includes proteome and transcriptome comparisons of tissue, serum, proximal fluids, cancer cell lines and animal models to select the most promising cancer biomarkers. **b** | Some major technological tools, such as mass spectrometry and microarrays are needed to enable an integrated approach. **c** | Once shortlists of potential candidates are generated, clinical studies with large numbers of samples and quantitative assays, such as enzyme-linked immunosorbent assay and multiple reaction monitoring will be necessary to identify the best performing biomarkers for clinical practice. CID, collision-induced dissociation; FDA, Food and Drug Administration; Q, quadruple.

***The utility of combining two sources.*** In our laboratory, we delineated the subproteome of ovarian cancer ascites fluid, yielding approximately 450 proteins[48]. Using size exclusion chromatography and ultrafiltration to remove high-abundance proteins with molecular mass ≥30 kDa, an in-depth two-dimensional liquid chromatography-tandem mass spectrometry analysis was carried out. Among the 450 proteins, 25 were previously identified as ovarian cancer biomarkers, validating the effectiveness of identifying biomarkers with this approach. Some filtering criteria were applied to focus the most promising putative markers only. These included further considering secreted and membrane proteins only, as these proteins have the highest chance of reaching the circulation (289 of 450 proteins); the removal of all known high-abundance serum proteins (defined as a concentration of >5 ug per ml resulting in 159 of 289 proteins for further consideration); the removal of previously studied molecules as biomarkers in the serum of patients with ovarian cancer (identified by literature mining of the 159 proteins) and proteins with a single peptide hit only, therefore retaining 91 of 159 proteins. The final filtering criterion consisted of comparing the 91 proteins from the ascites proteome with the proteome of condition media of 4 ovarian cancer cell lines (also delineated in our laboratory, revealing >1,500 proteins)[49] and retaining common proteins only. Further validation was warranted for 52 candidates.

Preliminary verification of some candidates ($n$ = 15) by ELISA (when available) in the serum of 100 healthy women, 100 women with benign gynaecological conditions and 100 women with ovarian carcinoma, highlighted nidogen 2, a basement membrane protein, as a candidate ovarian cancer biomarker[50]. With such a sample size, the power to detect a 20–50% difference between the means of groups with a 30–50% range of variation of means would be approximately 80%. Both serum nidogen 2 and CA125 concentrations in patients with ovarian cancer were increased, compared with sera from normal women ($p$<0.0001). In receiver operating characteristic (ROC) curve analysis, nidogen 2 had an area under the curve (AUC) of 0.73 but CA125 was superior with an AUC of 0.93 (FIG. 2), and there was no complementarity between the two markers. This example highlights the fact that new technological approaches could indeed identify potential biomarkers but that their superiority compared with existing ones is not a given. Therefore, combining two sources to mine (ovarian cancer ascites and cancer cell lines) revealed a new ovarian cancer biomarker, with a yield of approximately 7% (1 marker of the 15 tested). Consequently, the false discovery rate was 93%.

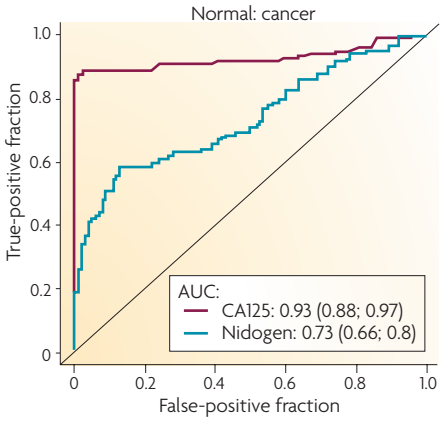


Figure 2 | **Performance characteristics of CA125 and nidogen 2.** Receiver operating characteristic curve for nidogen 2 and CA125 in the serum of non-cancer individuals and patients with ovarian cancer. The estimated area under the curve is shown (95% confidence interval in brackets). CA125 was superior and there was no complementarity between the two markers.

***Combination of multiple sources.*** Given that the proteomes of the relevant biomarker sources have been delineated for ovarian cancer, the possibility to combine these data sets has become feasible. We compared the proteomes of ovarian cancer cell line studies (Faca *et al.*[36] and Gunawardana *et al.*[49]), each identifying 961 and 383 extracellular and/or cell membrane-bound proteins, respectively. To this comparison, we added two proteomic studies on soluble ovarian cancer ascites fluid[30,48], each identifying 170 and 373 proteins, respectively. We also included 1 proteome of ovarian cancer tissues, identifying 69 proteins that were differentially expressed between malignant and benign tumour groups[34] and 1 microarray data set of 21 genes that were overexpressed in serous carcinoma[8] to generate a list of overlapping, prospective ovarian tumour markers.

There were no proteins common to all six data sets. Several proteins were identified in at least 2 of the data sets examined, and 33 proteins were identified in at least 3 data sets (TABLE 1). Interestingly, two proteins were found in four data sets (the two cancer cell line proteome studies[36,49], one ascites proteome[48] and in the microarray data set[8]). These two proteins are HE4 and granulin (GRN; also known as granulin-epithelin precursor (GEP), proepithelin, PC cell-derived growth factor and acrogranin). HE4 is made up of two whey acidic protein domains and four disulphide core domains and is overexpressed in ovarian cancer[7,45]. It belongs to a family of protease inhibitors that function in the immune response and is expressed particularly in serous and endometrioid ovarian cancer. Unlike CA125, which is increased in patients with advanced endometriosis, HE4 expression is not increased in many common benign gynaecological or other conditions[7,51]. One study examining women with benign ovarian tumours compared with women with invasive epithelial ovarian cancers has shown that the combination of HE4 and CA125 yielded a sensitivity of 76% at a specificity of 95%, which corresponds to a 33% increase in sensitivity of CA125 alone[52]. The isolation of HE4 through an integrated systems biology strategy provides strong credence to our hypothesis that promising new biomarkers can be identified through this approach.

The other protein identified in the four studies examined was GRN, a 68 kDa secreted protein[5354]. GRN is known to have a role in the regulation of inflammatory response and in wound healing and repair[54]. Interestingly, it is also a growth factor that has been shown to be highly expressed in invasive epithelial ovarian cancers (by using a cDNA library from microdissected tumour epithelium and further validated using RT-PCR and immunohistochemistry)[55,56], suggesting a possible role for GRN in ovarian tumour growth and invasiveness.

Recently, Pitteri *et al.*[40] carried out a study in which they integrated three biomarker discovery sources for ovarian cancer. The results from plasma proteomics of an ovarian cancer mouse model, analysis of proteins secreted by human ovarian cancer cell lines and proteomic analysis of fresh tumour cells enriched from ascites fluid were combined. In all three data sets 58 proteins were identified (including HE4), from which 25 proteins were selected (primarily on the basis of reagent availability) for preliminary verification studies. Eight proteins (including GRN) were found to be significantly increased in patients with ovarian cancer compared with controls. Furthermore, using an independent validation serum sample set, five of the eight proteins were found to have increased expression levels (GRN, IGFBP2, RARRES2, TIMP1 and CD14).

Clearly, the integration of various sources to mine for biomarkers can indeed yield promising candidates, which need to be properly validated according to Phases 2–5 in the biomarker development pipeline.

Table 1 | **Proteins identified in at least three ovarian cancer studies**

| Gene | Protein name | Cell line[36] | Cell line[49] | Ascites[48] | Ascites[30] | Microarray[8] | Tissue[34] |
|---|---|---|---|---|---|---|---|
| *CFB* | Complement factor B | | ✓ | ✓ | ✓ | | |
| *SPP1* | Osteopontin | | ✓ | ✓ | | ✓ | |
| *TGFBI* | Transforming growth factor-β-induced protein ig-h3 | | ✓ | ✓ | | ✓ | |
| *IGHG1* | Immunoglobulin heavy constant gamma 1 (G1m marker) | ✓ | | ✓ | ✓ | | |
| *ANXA2* | Annexin A2 | ✓ | ✓ | | | | ✓ |
| *RNASET2* | Ribonuclease T2 | ✓ | ✓ | ✓ | | | |
| *C3* | Complement C3r | ✓ | ✓ | ✓ | | | |
| *COL1A2* | Prepro-α2(I) collagen | ✓ | ✓ | ✓ | | | |
| *ATP6AP2* | Renin receptor | ✓ | ✓ | ✓ | | | |
| *SECTM1* | Secreted and transmembrane protein 1 | ✓ | ✓ | ✓ | | | |
| *FBLN1* | Fibulin 1 | ✓ | ✓ | ✓ | | | |
| *LGALS1* | Galectin 1 | ✓ | ✓ | ✓ | | | |
| *LTBP1* | Latent transforming growth factor-β binding protein, isoform 1L | ✓ | ✓ | ✓ | | | |
| *COL6A1* | Collagen alpha 1(VI) chain | ✓ | ✓ | ✓ | | | |
| *CLU* | Clusterin | ✓ | ✓ | ✓ | | | |
| *SERPING1* | Plasma protease C1 inhibitor | ✓ | ✓ | ✓ | | | |
| *CFI* | Complement factor I | ✓ | ✓ | ✓ | | | |
| *PROS1* | Vitamin K-dependent protein S | ✓ | ✓ | ✓ | | | |
| *NUCB1* | Nucleobindin 1 | ✓ | ✓ | ✓ | | | |
| *THBS1* | Thrombospondin 1 | ✓ | ✓ | ✓ | | | |
| *IGFBP2* | Insulin-like growth factor binding protein 2 | ✓ | ✓ | ✓ | | | |
| *COL1A1* | Collagen α1(I) chain | ✓ | ✓ | ✓ | | | |
| *VTN* | Vitronectin | ✓ | ✓ | ✓ | | | |
| *NPC2* | Epididymal secretory protein E1 | ✓ | ✓ | ✓ | | | |
| *COL5A1* | Collagen α1(V) chain | ✓ | ✓ | ✓ | | | |
| *IGFBP4* | Insulin-like growth factor binding protein 4 | ✓ | ✓ | ✓ | | | |
| *AGRN* | Agrin | ✓ | ✓ | ✓ | | | |
| *VASN* | Vasorin | ✓ | ✓ | ✓ | | | |
| *FOLR1* | Splice isoform 1 of folate receptor-α | ✓ | ✓ | ✓ | | | |
| *A2M* | Alpha-2-macroglobulin | ✓ | ✓ | ✓ | | | |
| *SERPINA1* | Alpha-1-antitrypsin | ✓ | ✓ | ✓ | | | |
| *WFDC2* | WAP four-disulphide core domain protein 2 | ✓ | ✓ | ✓ | | ✓ | |
| *GRN* | Granulin | ✓ | ✓ | ✓ | | ✓ | |

## Conclusion

Given the large number of potential target proteins identified in discovery phase platforms for ovarian and other cancers (Phase 1), the immediate challenge is to select the most promising candidates for further investigation. We have shown that through the integration of all available databases for biomarker discovery (genomic and proteomic), promising putative molecules can be identified and further validated.

There is now a need to bridge biomarker discovery with further verification, assay optimization, validation and commercialization steps so that the biomarkers can reach the clinic[57]. A major bottleneck to clinical translation is the need for analyte-specific reagents (such as ELISA assays) for validation and the requirement of large sample sets, preferably collected prospectively[13]. It is hoped that quantitative and highly sensitive proteomic techniques will soon bridge the gap between biomarker discovery, verification and validation[58]. In the future, multiple reaction monitoring methods, with and without prior enrichment, might enable quantification in serum without the need for analyte-specific reagents[59].

*Vathany Kulasingam, Maria P. Pavlou and Eleftherios P. Diamandis are at the Department of Laboratory Medicine and Pathobiology, University of Toronto, and the Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto M5G 1X5, Ontario, Canada.*

*Eleftherios P. Diamandis is also at the Department of Clinical Biochemistry, University Health Network, Toronto M5G 2C4, Ontario, Canada.*

*Correspondence to E.P.D.*
*e-mail: ediamandis@mtsinai.on.ca*

doi:10.1038/nrc2831
Published online 12 April 2010

1. Jemal, A. *et al.* Cancer statistics, 2008. *CA Cancer J. Clin.* **58**, 71–96 (2008).
2. Konstantinopoulos, P. A., Spentzos, D. & Cannistra, S. A. Gene-expression profiling in epithelial ovarian cancer. *Nature Clin. Pract. Oncol.* **5**, 577–587 (2008).
3. Cannistra, S. A. Cancer of the ovary. *N. Engl. J. Med.* **351**, 2519–2529 (2004).
4. Diamandis, E. P., Fritsche, H. A., Lilja, H., Chan, D. W. & Schwartz, M. K. *Tumor Markers: Physiology, Pathobiology, Technology, and Clinical Applications.* (AACC Press, Washington DC., 2002).
5. Kohn, E. C., Mills, G. B. & Liotta, L. Promising directions for the diagnosis and management of gynecological cancers. *Int. J. Gynaecol. Obstet.* **83** (Suppl. 1), 203–209 (2003).
6. Rustin, G. J., Nelstrop, A. E., Tuxen, M. K. & Lambert, H. E. Defining progression of ovarian carcinoma during follow-up according to CA 125: a North Thames Ovary Group Study. *Ann. Oncol.* **7**, 361–364 (1996).
7. Huhtinen, K., *et al.* Serum HE4 concentration differentiates malignant ovarian tumours from ovarian endometriotic cysts. *Br. J. Cancer* **100**, 1315–1319 (2009).
8. Meinhold-Heerlein, I. *et al.* An integrated clinical-genomics approach identifies a candidate multi-analyte blood test for serous ovarian carcinoma. *Clin. Cancer Res.* **13**, 458–466 (2007).
9. Menon, U. *et al.* Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncol.* **10**, 327–340 (2009).
10. Visintin, I. *et al.* Diagnostic markers for early detection of ovarian cancer. *Clin. Cancer Res.* **14**, 1065–1072 (2008).
11. Pepe, M. S. *et al.* Phases of biomarker development for early detection of cancer. *J. Natl. Cancer Inst.* **93**, 1054–1061 (2001).
12. Ransohoff, D. F. Lessons from controversy: ovarian cancer screening and serum proteomics. *J. Natl. Cancer Inst.* **97**, 315–319 (2005).
13. Pepe, M. S., Feng, Z., Janes, H., Bossuyt, P. M. & Potter, J. D. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J. Natl. Cancer Inst.* **100**, 1432–1438 (2008).
14. Jacobs, I. J. & Menon, U. Progress and challenges in screening for early detection of ovarian cancer. *Mol. Cell Proteomics.* **3**, 355–366 (2004).
15. Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell Proteomics* **1**, 845–867 (2002).
16. Petricoin, E. F. *et al.* Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572–577 (2002).
17. Karsan, A. *et al.* Analytical and preanalytical biases in serum proteomic pattern analysis for breast cancer diagnosis. *Clin. Chem.* **51**, 1525–1528 (2005).
18. Banks, R. E. *et al.* Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. *Clin. Chem.* **51**, 1637–1649 (2005).
19. Diamandis, E. P. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J. Natl. Cancer Inst.* **96**, 353–356 (2004).
20. Diamandis, E. P. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol. Cell Proteomics* **3**, 367–378 (2004).
21. Simpkins, F., Czechowicz, J. A., Liotta, L. & Kohn, E. C. SELDI-TOF mass spectrometry for cancer biomarker discovery and serum proteomic diagnostics. *Pharmacogenomics* **6**, 647–653 (2005).
22. Su, F. *et al.* Validation of candidate serum ovarian cancer biomarkers for early detection. *Biomark. Insights* **2**, 369–375 (2007).
23. Nossov, V. *et al.* The early detection of ovarian cancer: from traditional methods to proteomics. Can we really do better than serum CA-125? *Am. J. Obstet. Gynecol.* **199**, 215–223 (2008).
24. Zhang, Z. *et al.* Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **64**, 5882–5890 (2004).
25. Diamandis, E. P. Validation of breast cancer biomarkers identified by mass spectrometry. *Clin. Chem.* **52**, 771–772 (2006).
26. Chang, H. W. *et al.* Assessment of plasma DNA levels, allelic imbalance, and CA 125 as diagnostic tests for cancer. *J. Natl. Cancer Inst.* **94**, 1697–1703 (2002).
27. Stroun, M. *et al.* Neoplastic characteristics of the DNA found in the plasma of cancer patients. *Oncology* **46**, 318–322 (1989).
28. Adam, R. A. & Adam, Y. G. Malignant ascites: past, present, and future. *J. Am. Coll. Surg.* **198**, 999–1011 (2004).
29. Tamsma, J. T., Keizer, H. J. & Meinders, A. E. Pathogenesis of malignant ascites: Starling's law of capillary hemodynamics revisited. *Ann. Oncol.* **12**, 1353–1357 (2001).
30. Gortzak-Uzan, L. *et al.* A proteome resource of ovarian cancer ascites: integrated proteomic and bioinformatic analyses to identify putative biomarkers. *J. Proteome Res.* **7**, 339–351 (2008).
31. Kondo, T. Tissue proteomics for cancer biomarker development: laser microdissection and 2D-DIGE. *BMB. Rep.* **41**, 626–634 (2008).
32. Good, D. M. *et al.* Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future. *J. Proteome Res.* **6**, 4549–4555 (2007).
33. Jones, M. B. *et al.* Proteomic analysis and identification of new biomarkers and therapeutic targets for invasive ovarian cancer. *Proteomics* **2**, 76–84 (2002).
34. Bengtsson, S. *et al.* Large-scale proteomics analysis of human ovarian cancer for biomarkers. *J. Proteome Res.* **6**, 1440–1450 (2007).
35. Kulasingam, V. & Diamandis, E. P. Tissue culture-based breast cancer biomarker discovery platform. *Int. J. Cancer* **123**, 2007–2012 (2008).
36. Faca, V. M. *et al.* Proteomic analysis of ovarian cancer cells reveals dynamic processes of protein secretion and shedding of extra-cellular domains. *PLoS ONE* **3**, e2425 (2008).
37. Kelly-Spratt, K. S., Kasarda, A. E., Igra, M. & Kemp, C. J. A mouse model repository for cancer biomarker discovery. *J. Proteome Res.* **7**, 3613–3618 (2008).
38. Frese, K. K. & Tuveson, D. A. Maximizing mouse cancer models. *Nature Rev. Cancer* **7**, 645–658 (2007).
39. Maser, R. S. *et al.* Chromosomally unstable mouse tumours have genomic alterations similar to diverse human cancers. *Nature* **447**, 966–971 (2007).
40. Pitteri, S. J. *et al.* Integrated proteomic analysis of human cancer cells and plasma from tumor bearing mice for ovarian cancer biomarker discovery. *PLoS ONE* **4**, e7916 (2009).
41. Bignotti, E. *et al.* Differential gene expression profiles between tumor biopsies and short-term primary cultures of ovarian serous carcinomas: identification of novel molecular biomarkers for early diagnosis and therapy. *Gynecol. Oncol.* **103**, 405–416 (2006).
42. Welsh, J. B. *et al.* Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. *Proc. Natl. Acad. Sci. USA* **100**, 3410–3415 (2003).
43. Tan, D. S., Lambros, M. B., Natrajan, R. & Reis-Filho, J. S. Getting it right: designing microarray (and not 'microawry') comparative genomic hybridization studies for cancer research. *Lab. Invest.* **87**, 737–754 (2007).
44. Spentzos, D. *et al.* Gene expression signature with independent prognostic significance in epithelial ovarian cancer. *J. Clin. Oncol.* **22**, 4700–4710 (2004).
45. Moore, R. G. *et al.* A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol. Oncol.* **112**, 40–46 (2009).
46. Vargo-Gogola, T. & Rosen, J. M. Modelling breast cancer: one size does not fit all. *Nature Rev. Cancer* **7**, 659–672 (2007).
47. Rifai, N., Gillette, M. A. & Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnol.* **24**, 971–983 (2006).
48. Kuk, C. *et al.* Mining the ovarian cancer ascites proteome for potential ovarian cancer biomarkers. *Mol. Cell Proteomics* **8**, 661–669 (2009).
49. Gunawardana, C. G. *et al.* Comprehensive analysis of conditioned media from ovarian cancer cell lines identifies novel candidate markers of epithelial ovarian cancer. *J. Proteome. Res.* **8**, 4705–4713 (2009).
50. Kuk, C. *et al.* Nidogen-2: a new serum biomarker for ovarian cancer. *Clin. Biochem.* **43**, 355–361 (2009).
51. Hellstrom, I. *et al.* The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. *Cancer Res.* **63**, 3695–3700 (2003).
52. Moore, R. G. *et al.* The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass. *Gynecol. Oncol.* **108**, 402–408 (2008).
53. Davidson, B. *et al.* Granulin-epithelin precursor is a novel prognostic marker in epithelial ovarian carcinoma. *Cancer* **100**, 2139–2147 (2004).
54. He, Z. & Bateman, A. Progranulin (granulin-epithelin precursor, PC-cell-derived growth factor, acrogranin) mediates tissue repair and tumorigenesis. *J. Mol. Med.* **81**, 600–612 (2003).
55. Jones, M. B., Spooner, M. & Kohn, E. C. The granulin-epithelin precursor: a putative new growth factor for ovarian cancer. *Gynecol. Oncol.* **88**, S136–S139 (2003).
56. Jones, M. B. *et al.* The granulin-epithelin precursor/PC-cell-derived growth factor is a growth factor for epithelial ovarian cancer. *Clin. Cancer Res.* **9**, 44–51 (2003).
57. Simpson, R. J., Bernhard, O. K., Greening, D. W. & Moritz, R. L. Proteomics-driven cancer biomarker discovery: looking to the future. *Curr. Opin. Chem. Biol.* **12**, 72–77 (2008).
58. Diamandis, E. P., Hanash, S., Lopez, M., Carr, S. & Petricoin, E. F. Protein quantification by mass spectrometry: is it ready for prime time? *Clin. Chem.* **55**, 1427–1430 (2009).
59. Addona, T. A. *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nature Biotechnol.* **27**, 633–641 (2009).



**Competing interests statement**
The authors declare no competing financial interests.


## DATABASES

**UniProtKB:** http://www.uniprot.org
CA125 | GRN | HE4 | nidogen 2

## FURTHER INFORMATION

**Eleftherios P. Diamandis' homepage:** http://www.acdclab.org/

**ALL LINKS ARE ACTIVE IN THE ONLINE PDF**