The search for predictive patterns in ovarian cancer: Proteomics meets bioinformatics

Proteomic patterns in serum that discriminate between malignant and benign ovaries may provide a powerful tool for screening in high risk women.

Epithelial ovarian cancer continues to be the leading cause of death from gynecologic cancers in the United States and accounts for 14,400 deaths per year (Ries et al., 2000). Ovarian cancer is associated with a heavy burden of morbidity and a high case-fatality rate despite high cure rates for early stage disease by surgical resection alone. This apparent contradiction is explained by the predominance of advanced stage disease at presentation. Close to 70% of women present with disease spread beyond the pelvis, a group whose long-term survival is only 29% (Piver et al., 1993). While progress in defining optimal adjuvant therapies for ovarian cancer has resulted in a modest improvement in disease-free and overall survival, an exponential magnitude of difference could be accomplished by downstaging the disease at diagnosis through the application of an effective screening tool. To date, despite extensive evaluation, the combination of physical examination, imaging with transvaginal ultrasound, and the serum marker CA-125 have not resulted in acceptable sensitivity and specificity levels, and the NIH Consensus Statement published in 1994 (NIH, 1994) found no evidence to recommend these screening modalities. A recent article in *Lancet* by Petricoin et al. describes what may be a major breakthrough in this stalemate with the identification of a proteomic pattern that discriminated ovarian cancer from non-cancer with a positive predictive value of 94% (Petricoin et al., 2002).

An appreciation of the epidemiology of epithelial ovarian cancer both supports the need to develop an effective screening strategy and identifies subsets of women who are likely to derive the most benefit from this strategy. The dominant risk factors, which include advancing age, a family history of ovarian and/or breast cancer, and reproductive events that promote uninterrupted ovulation, namely infertility and nulliparity, are not easily amenable to modification. The recent identification of the ovarian cancer susceptibility genes, *BRCA1*, *BRCA2*, and the HNPCC genes, has further defined a group of women whose lifetime risk for ovarian cancer may range from 16% to 63% (Easton et al., 1995; Streuwing et al., 1997). Several strategies have attempted to improve positive predictive values for ovarian cancer screening. CA-125, the most extensively studied biomarker in ovarian cancer, represents an antigenic determinant on a high molecular weight glycoprotein which is expressed in structures derived from coelomic epithelium. Elevated CA-125 levels are found in approximately 85% of patients with epithelial ovarian cancer, and this marker is routinely used to follow response to treatment. However, expression of CA-125 is highly correlated with tumor volume, and only 50% of patients with early stage disease have elevated levels, thus limiting its use as a screening tool (Jacobs and Bast, 1989). Skates et al. (1995) have proposed the application of statistical modeling to longitudinal serial CA-125 assays to create a predictive algorithm that minimizes assay fluctuations and relies on patterns over time rather than isolated levels to improve predictive probability. Others have suggested the combination of panels of tumor-associated markers to improve efficacy. Some candidate markers include ovarian carcinoma-associated antigen (OCA), macrophage colony-stimulating factor (MCS-F), a growth factor which stimulates monocyte proliferation, lysophosphatidic acid (LPA), a bioactive phospholipid with mitogenic and growth-factor-like activity, CYFRA 21-1, a soluble serum fragment of cytokeratin 19, and tumor-associated trypsin inhibitor, a 6 kDa polypeptide, among others (Menon and Jacobs, 2000; Ozols et al., 2000). However, to date, no single marker or combination of markers has emerged with a clear advantage over CA-125 alone, and the search for the optimal screening tool for ovarian cancer continues.

The approach taken by Petricoin et al. (2002) represents a new direction in this search, wherein a distinct pattern of proteins creates the discriminatory power. Proteomics is a new and emerging technology that can identify low molecular weight molecules in a high-throughput, nonbiased discovery approach using patient serum, plasma, urine, or other secretions such as ascites (Mills et al., 2001). Because of the vast number of data points this technology can generate, it is being linked to sophisticated computer generated algorithms to produce distinctive protein signatures with high discriminatory potential. The investigators used a dual-phase approach. In the Pattern Discovery phase, mass spectra generated by SELDI-TOF from serum of affected and high risk unaffected individuals were compared. Using a combination of genetic algorithms and cluster analysis, a small set of key protein values emerged that discriminated the cases from the unaffected controls. In the Pattern Matching phase, the optimum pattern defined in Phase I was applied to a new set of masked samples to assign each sample to a predictive category. All 50 cancer samples were correctly categorized, including all 18 stage I cancers, and 63 of 65 normal controls were correctly categorized, yielding a sensitivity of 100% and a specificity of 95%.

Epidemiological and clinical features of ovarian cancer, however, point out that without additional refinements, the proteomic results as reported will not be useful for screening the general population. Ovarian cancer is a relatively uncommon disease. The screening program must have a very high specificity, greater than 99%, to achieve an acceptable positive predictive value. A "positive" screening test will trigger a diagnostic procedure to determine if the woman has ovarian cancer. An exploratory laparotomy remains the gold standard to diagnose ovarian cancer and is associated with morbidity and, very rarely, mortality. It has been generally accepted that a screening test should result in no more than 10 laparotomies to diagnose one case of ovarian cancer. Since the prevalence of ovarian cancer is low (1 in 2500), a screening test which has a specificity of only 95% (the proteomic pattern) will result in far too many unnec-

essary laparotomies to make it clinically useful. The sensitivity of 100% in the proteomic assay is encouraging, but because of the low prevalence, a high sensitivity does not significantly influence the false-positive laparotomy rate. However, the false-positive rate can be improved by either increasing the specificity of the test or by identifying a high-risk population which will have a greater prevalence of ovarian cancer. It remains to be determined whether sequential proteomic assays, or the use of this assay in conjunction with other screening technologies, such as serum CA-125 levels or ultrasonography, will lead to improvements in specificity that will make the test clinically useful.

This approach does have several potential advantages. It is minimally invasive and uses small amounts of biospecimens in a fast and cost-effective manner. The discriminatory power of the applied statistical modeling will continue to improve when the data set grows and the specificity may become acceptable. And apart from providing an effective screening tool, evaluation of the protein products that constitute the key cluster has the potential to elucidate the underlying molecular processes of ovarian carcinogenesis. Clearly, several future directions are indicated by this report. First, validation of the findings in a large prospective cohort with sufficient power to assess the discriminatory power of the analyses in subsets of women defined by risk parameters, genetic status, stage, and histopathologic type is needed. Reproducibility among mass spectroscopy techniques and across instrumentation must precede widespread application. Novel mathematical modeling approaches specifically tailored to proteomic technology may further advance the refinement of this approach. And the application of this model to other disease conditions could have unlimited possibilities, both in the diagnostic and treatment setting.

Recently, microarray technology has also been used to identify novel potential serum markers for early detection of ovarian cancer. Mok et al. (2001) used such a technique to identify overexpressed genes for secretory proteins as potential serum biomarkers. RNA was isolated from ovarian cancer cell lines and from normal human ovarian surface epithelial (HOSE) cells, and complementary DNA generated from these pools was hybridized to a microarray slide to identify genes overexpressed in cancer cells. Prostasin, a serine protease normally secreted by the prostate gland, was selected for further study. Prostasin was then measured in the serum using an enzyme-lined immunosorbent assay in 64 patients with ovarian cancer and in 137 normal subjects. The resulting sensitivity and specificity for prostasin in combination with serum CA-125 resulted in similar sensitivity (92%) and specificity (94%) as reported with the proteomic assay. At least 14 other candidate tumor markers have been identified through similar differential display technologies of gene expression (Mills et al., 2001). It is certainly possible that combinations of microarray technology and proteomics will lead to new algorithms for effective screening for ovarian cancer.

Perhaps the most promising outcome of the work reported by Petricoin et al. (2002) is the interface between proteomic technology and bioinformatics. The rapid explosion in the amount of data being generated by current genomic and proteomic technologies already exceeds the analytic capacity of the human mind. The need for increasing sophistication in data management and statistical interpretation is underscored by this paper, and attention to this aspect of the research is critical to its successful translation into clinical practice.

Mary B. Daly[1,3] and
Robert F. Ozols[2]

[1]Division of Population Science
[2]Division of Medical Science
Fox Chase Cancer Center
7701 Burholme Avenue
Philadelphia, Pennsylvania 19111
[3]E-mail: mb_daly@fccc.edu

Selected reading

Easton, D.F., Ford, D., Bishop, D.T., and the Breast Cancer Linkage Consortium. (1995). Am. J. Hum. Genet. *56*, 265–271.

Jacobs, I., and Bast, R. (1989). Hum. Reprod. *4*, 1–12.

Menon, U., and Jacobs, I. (2000). Tumor Markers. In Principles and Practice of Gynecologic Oncology, W. Hoskins, C. Perez, R. Young, eds. (Philadelphia: Lippincott Williams and Wilkins), pp. 165–182.

Mills, G., Bast, R., and Srivastava, S. (2001). J. Natl. Cancer Inst. *93*, 1437–1439.

Mok, S.C., Chao, J., Skates, S., Wong, K., Yiu, G.K., Muto, M.G., Berkowitz, R.S., and Cramer, D.W. (2001). J. Natl. Cancer Inst. *93*, 1458–1464.

NIH (1994). NIH Consens. Statement *12*, 1–30.

Ozols, R., Rubin, S., Thomas, G., and Robboy, S. (2000). Epithelial Ovarian Cancer. In Principles and Practice of Gynecologic Oncology, W. Hoskins, C. Perez, R. Young, eds. (Philadelphia: Lippincott Williams and Wilkins), pp. 981–1058.

Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D., Kohn, E.C., and Liotta, L.A. (2002). Lancet *359*, 572–577.

Piver, M.S., Fanning, J., and Craig, K.A. (1993). Ovarian Cancer. In Gynecologic Oncology, R.C. Knapp and R.S. Berkowitz, eds. (New York: McGraw Hill), p. 253.

Ries, L.A.G., Wingo, P.A., Miller, D.S., Howe, H.L., Weir, H.K., Rosenberg, H.M., Vernon, S.W., Cronin, K., and Edwards, B.K. (2000).Cancer *88*, 2398–2424.

Skates, S., Xu, F., Yu, Y., Sjovall, K., Einhorn, N., Chang, Y., Bast, R., and Knapp, R. (1995). Cancer *76*, 2004–2010.

Streuwing, J.P., Hartge, P., Wacholder, S., Baker, S.M., Berlin, M., McAdams, M., Timmerman, M.M., Brody, L.C., and Tucker, M.A. (1997). N. Engl. J. Med. *336*, 1401–1408.