Lessons From Controversy: Ovarian Cancer Screening and Serum Proteomics

David F. Ransohoff

In 2002 a study reported that a blood test, based on pattern-recognition proteomics mass spectroscopy analysis of serum, was nearly 100% sensitive and specific to detect ovarian cancer. Plans to introduce a commercial screening test by early 2004 were delayed amid concerns about whether the approach was reproducible and reliable. In this issue of JNCI, two commentaries discuss whether the initial results are reproducible and whether bias may account for results. This essay describes how threats to validity from chance and bias may cause erroneous results and inflated expectations in the kind of observational research being conducted in several "-omics" fields to assess molecular markers for diagnosis and prognosis of cancer. To address such threats and to realize the potential of new -omics technology will require application of appropriate rules of evidence in the design, conduct, and interpretation of clinical research about molecular markers. [J Natl Cancer Inst 2005;97:315–9]

BACKGROUND AND CONTEXT

Three years ago a study reported that a blood test, based on pattern-recognition proteomics analysis of serum, was nearly 100% sensitive and specific for detecting ovarian cancer and was possibly useful in screening *(1)*. After this and other supporting work *(2)*, commercial laboratories planned to market a test in late 2003 or early 2004 *(3,4)*, but plans were delayed by the U.S. Food and Drug Administration *(5–8)*. Questions were raised about whether the technology's results were reproducible and reliable enough for application in practice *(4,9–11)*. Although the chronology of claims, plans, and delay has been reported in professional journals and the lay press *(4,8,10,11)*, the scientific issues have not been thoroughly discussed in print. The question about whether the approach of discovery-based serum proteomics can accurately and reliably diagnose ovarian cancer—or any cancer—has not been resolved.

Two commentaries in this issue of the *Journal (12,13)* discuss details of the controversy and provide insight not only into whether serum proteomics may diagnose ovarian cancer but also into the larger process by which discovery-based "-omics" research is currently being explored and evaluated. Fields with names like genomics, transcriptomics, metabolomics, epigenomics, and ribonomics may face similar questions about reproducibility and validation. For example, in a special section on genomics *(14)* in a recent issue of *Science*, an essay entitled "Getting the noise out of gene arrays" noted that "[t]housands of papers have reported results obtained using gene arrays…. But are these results reproducible?" *(15)*. If similar questions about reproducibility are being asked in different "-omics" fields, there may be a larger problem in the process by which this kind of research is designed and assessed and by which conclusions and claims are made.

One way to state the problem is as follows: what rules of evidence are required to decide whether a result is accurate and reliable? We use rules already (even if not explicitly) to make decisions as investigators about how to design and report research, as reviewers and editors about whether to publish it, and as reviewers for funding agencies about whether to fund it. Although terms like rules of evidence and validation sound perhaps philosophical or arcane, they are central to understanding and improving the current situation in which claims and expectations may have exceeded data. Such rules have practical tangible consequences for design, conduct, and interpretation of "-omics" research. If we do not learn to explore and evaluate such fields properly, we will pay an opportunity cost of inflated expectations followed by disappointment and the inefficient use of investigators' and funding agencies' time and resources.

By way of disclosure, I have followed this topic closely for 3 years because of a long-standing interest in methods to evaluate diagnostic tests *(16,17)*, and I have had numerous discussions with both Drs. Baggerly and Petricoin (authors of the brief communication and counterpoint commentary *[12,13]*, respectively) about these and other issues. I have sent specimens to Dr. Petricoin for analysis. On the one hand, I am extremely hopeful that a pattern-recognition serum proteomics approach works, because a noninvasive blood test for cancer could be so useful to patients and physicians. On the other hand, I am by nature and training profoundly skeptical about most promising claims. Skepticism, however, can be convincingly addressed by evidence. A general consideration of "What rules of evidence are required?" cannot be provided in a brief essay and is discussed elsewhere in articles on molecular markers *(18–23)* and, more broadly, in textbooks about clinical epidemiology and clinical research design *(24–29)*. This commentary borrows from these sources and tries to put the current controversy and discussion in a larger context and to suggest practical lessons and next steps.

THE DEBATE OVER REPRODUCIBILITY

Baggerly et al. *(12)* argue that the results of two major studies of serum proteomics and diagnosis of ovarian cancer do not demonstrate reproducibility in independent subjects and that, for that reason, the apparent discrimination may be explained simply by chance—for example, by the overfitting that occurs when a multivariable model is used to fit a very large number of possible predictors (such as mass spectroscopy peaks) to discriminate among a group of subjects with or without cancer *(21)*. Baggerly et al. further say that the particular jackknife (or leave-one-out)

Affiliations of author: Departments of Medicine and Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Correspondence to: David F. Ransohoff, MD, Departments of Medicine and Epidemiology, University of North Carolina at Chapel Hill, CB# 7080, Bioinformatics Bldg. 4103, Chapel Hill, NC 27599-7080 (e-mail: ransohof@med.unc.edu).

See "Notes" following "References."

method of model-making used in one study is particularly prone to overfitting *(12)*; details of how problems arise from suboptimal jackknife methods have been described elsewhere *(30)*. Demonstrating that overfitting has not occurred can be accomplished by reproducing results in a group of subjects not used to derive the discriminating model while holding technical features constant, as discussed below. Separately, Baggerly et al. say that a potential problem from procedural bias may have caused "pervasive differences between cancer and control spectra" *(12)*, causing erroneous signals to become hard-wired in data and resulting in the appearance of discrimination.

Liotta et al. *(13)* respond that an expectation of reproducibility is not warranted and was not claimed, because the published studies had different goals, such as testing different chip surfaces, laser energy, or the machine variance that may occur over time. Liotta et al. say that, while "Baggerly et al. conclude that reproducibility within the two posted data sets used in these studies has not been demonstrated," the "[c]omparisons of… reproducibility across time can only be applied to study sets in which all process variables, except time, are constant," and the "study sets … did not fall into this category… to judge reproducibility…." *(13)*.

IMPORTANCE OF DEMONSTRATING REPRODUCIBILITY: AVOIDING CONCLUSIONS DUE TO CHANCE

The primary point of contention—whether results are reproducible—has critical implications for interpreting research. Reproducibility is not a uniform concept in the sense that reproducibility can be assessed in different settings to answer very different questions. The most efficient way to answer the question of "Does chance explain results?" is in the setting of a single study, in which the researcher tries to reproduce those results using the same types of subjects (but different subjects) *(23)* while holding other features constant, such as laboratory equipment, chip surfaces, laser energy, and so on.

Showing reproducibility in this setting may seem narrow in the sense that it does not test reproducibility by using different equipment, reagents, and types of subjects, but it provides the cleanest way to test the specific question about whether overfitting or chance explains results. The failure to clearly demonstrate reproducibility in this kind of assessment constitutes an important "no-go" result, because a conclusion that chance might explain discrimination has major implications for decisions about conduct of further research, about claims made, and even about publication.

On the basis of the concerns raised by Baggerly et al. and the response of Liotta et al., it would seem that such reproducibility has not been clearly demonstrated for the pattern-recognition serum proteomics studies discussed herein, although other work, discussed below, bears on this question.

REPRODUCIBILITY IN OTHER "-OMICS" RESEARCH

Lessons may be learned from how the question of reproducibility has been effectively addressed in other "-omics" research. In a genomics study that used a discovery-based pattern-recognition approach to identify expression signatures that predicted the prognosis of lymphoma, all technical features were held constant among a group of 240 subjects. That group of 240 subjects was split into a group of 160 used as a training set to derive a pattern-recognition model and a group of the remaining 80 subjects, not used in any way in training, treated as an independent validation set *(31)*. This study effectively showed that chance, or overfitting, did not explain the results, by demonstrating reproducibility in different patients in a setting in which technical features were held constant.

It is not yet clear, however, that RNA expression signatures work as well for predicting prognosis of other cancers; each study of this type requires appropriate consideration of chance as an explanation for results. In one study of breast cancer prognosis using RNA expression patterns *(32)*, subjects used in the validation set were not totally independent of subjects used in the training set *(33,34)*. A recent study that assessed expression of the same genes but in totally independent subjects found a degree of discrimination substantially lower than originally reported *(35)*. In another study reporting strong discrimination *(36)*, the independent group was too small to provide meaningful validation *(21,34)*. Further, there is a biological reason to worry that the degree of discrimination shown in lymphoma might not necessarily be found in a heterogeneous tumor such as those found in breast cancer. As one observer has noted *(37)*, "It is not by chance that the first applications of microarrays to the diagnosis of cancer were made on leukaemias and lymphomas." Because leukemias and lymphomas may behave in a clonal manner, sampling of even one cell may provide strong information about the biology of the entire tumor. In contrast, breast cancer is heterogeneous, involving "…variability in invasive potential between cancer cells within the same tumour. For example, prostate cancer is often present in many distinct foci, only one of which may have the potential to be invasive and dictate the outcome for the patient. Must each focus be analyzed separately? Within one focus there may only be a few cells with the potential to invade — will their gene-expression pattern be masked by the surrounding, less malignant cells?" *(37)*. A recent study about breast cancer prognosis described how discovery-based research helped to identify not a "signature" but rather 21 specific genes whose expression was hypothesized to provide information about prognosis and was then used to assess an entirely new group of subjects *(38)*. However, because the genes selected fell into four groups with biologically plausible function (proliferation, HER2, estrogen, and invasion) and because details of the selection process were not described, it is unclear to what extent a "discovery-based" approach contributed to gene selection. In any case, the genes selected showed reproducible discrimination in an independent group, thus demonstrating that chance does not explain results.

TECHNOLOGY AS A MOVING TARGET

Liotta et al. identify a difficult challenge affecting evaluation of many new technologies—the problem of a moving target in which, from study to study, technical features change. For serum proteomics technology, such features include type of mass spectroscopy machine, sample application, washing, chip surface, and patients *(13)*. Similar kinds of problems affect interpretation of genomics technology concerning RNA expression arrays *(15)*. Solving the moving-target problem requires holding the target still long enough to answer critical questions to build confidence in the technique and, by extension, in the field. The consequence of not taking time to solve this moving-target problem is that investigators or even entire fields may charge ahead in an effort that, in retrospect, turns out to have been based on work that was

insufficiently strong (i.e., that does not clearly avoid chance and bias) to support claims and expectations and further work. Although false starts are to be expected in science and the road to discovery is never straight, problems may be handled better than they are now.

THREATS TO VALIDITY FROM CHANCE AND BIAS

The two main threats to validity of nonexperimental (i.e., observational) research come from chance (21) and bias (23). As noted, problems caused by chance, such as overfitting that can cause apparent discrimination among compared groups but may not be reproducible, may be avoided by using a totally independent validation group while holding everything else constant (21). Unfortunately, this approach is not routinely used (21,39).

Threats from bias are more challenging because they are more numerous, not always obvious, and not straightforward to address (23). Bias occurs when subjects, specimens, or data in the groups being compared (i.e., cancer vs. not cancer) are inherently different or are handled differently in a way that systematically introduces a signal into data for one of the compared groups. Sources of bias in a study of a serum proteomics assay could include differences in the collection of specimens (types of collection tubes or length of time to spinning or freezing), specimen storage (number of thaw/freeze cycles or storage temperature), or the process of analysis on a mass spectroscopy machine. A single difference at a single step can cause signal to become hard-wired into data in a way that may not be undone or may even be undetectable by subsequent analysis (23). Because bias routinely receives little attention in reports of "-omics" research (23), one must wonder whether bias accounts for results, as suggested by Baggerly et al. (12). The reproducibility of results in different studies does not assure against bias, because repeated studies showing the same result may simply reproduce mistakes. As Sackett has said, "Bias times 12 is still bias" (40); the reasons are discussed elsewhere (23).

Bias is such a serious problem in nonexperimental research, like much discovery-based proteomic analysis, that a study is considered guilty of bias and erroneous results until proven innocent (23). Proving innocence relies on a process involving detailed effort and detailed reporting of that effort at every step of research, including in design to try to avoid bias, in conduct to measure whether it may have occurred, and in interpretation to determine whether bias could have affected results and conclusions (23). If this process sounds difficult and complicated, it is. As noted by Cole (40), bias is a "plague upon the house of epidemiology." This plague, like it or not, will be visited on "-omics" research about diagnosis and prognosis. The stakes are high if erroneous conclusions affect decisions about future research or patient care.

To effectively consider possible sources of bias in "-omics" research requires interdisciplinary collaboration that itself may be difficult, but the questioning of reproducibility in highly visible research in several "-omics" fields suggests that effort at collaboration may be worthwhile. When investigators trained in molecular biology or biochemistry start to conduct research in diagnosis and prognosis, they are undertaking, perhaps unwittingly, observational epidemiology research (41) that involves serious threats from chance and bias. In much conventional laboratory research, these threats are seldom major in the sense of being managed effectively and routinely by methods such as randomization and blinding; however, those methods typically are not available at important steps of observational research (23). Epidemiologists and biostatisticians face parallel challenges and may be unfamiliar with, or intimidated by, the technical and biological methods used in the laboratory. Effective research at this interdisciplinary or translational interface requires simultaneous in-depth consideration of technical, biological, and epidemiological detail. How well do biologists or technologists appreciate the nature or seriousness of problems that can be introduced by methods of handling patients, samples, data, or analysis, as well as how to design, conduct, and interpret research that addresses those problems? How well do epidemiologists or biostatisticians understand technical details of specimen collection, handling, and analysis, in a way they can use to anticipate and manage specific sources of bias? If communication at this interface is not effective, the kinds of questions occurring now will multiply, leading to misdirected effort and delay in learning what technology works and what does not.

LESSONS AND NEXT STEPS

The current situation suggests lessons and specific next steps not only in how to study serum proteomics and ovarian cancer but also, and more importantly, in how to explore an increasing number of "-omics" fields.

Serum Proteomics and Ovarian Cancer

The question of whether a serum proteomics approach can accurately and reproducibly diagnose ovarian cancer is now so complicated, and so important, that it should be addressed by a fresh study. Such a study should provide detailed effort to address chance and bias at every step—in design, conduct, and interpretation—and to report that effort thoroughly and transparently (23). Among published studies about serum proteomics and ovarian cancer, one that perhaps comes closest to this goal used a discovery-based approach to identify specific proteins that were then tested in a totally independent group of subjects (42). This study clearly avoided problems of chance and probably avoided bias. Notably, though, the reported degree of diagnostic discrimination for ovarian cancer was far lower than the nearly 100% sensitivity and specificity originally reported (1,2). A study under way to assess serum proteomics to diagnose prostate cancer, designed by investigators in the National Cancer Institute's Early Detection Research Network, has features that should successfully address many issues of chance and bias (43), but results have not been published. If we do not explore new technologies appropriately, we may repeat errors of a generation ago, when serum carcinoembryonic antigen showed initial promise of nearly 100% sensitivity and specificity for colorectal cancer but was followed by major disappointment when further studies were done. Events might have been predictable and avoidable if rules of evidence had been applied to evaluations of diagnostic tests (16,44).

Exploring Other "-omics" Fields

The questions about discovery-based serum proteomics raised in this commentary and questions raised about genomics elsewhere (15,33) suggest there may be broader lessons about how to

explore the increasing number of "-omics" fields. The failure of reports of current "-omics" research to provide much evidence and detail needed to assess problems resulting from chance and bias has caused "…a breed of 'forensic' statisticians *(45,46)*" to emerge, who "doggedly detect and correct" possible errors in published reports *(47)*. In an ideal world, questions that were asked after the fact by these statisticians would have been asked and addressed beforehand by investigators in designing and reporting research and by reviewers and editors in judging it. Indeed, there is a lesson in recognizing that even this detailed discussion *(12)* about a research report could not have happened if the original investigators had not taken the unusual and commendable step of posting so much raw data on the Web *(1)*. In other words, one cannot even speculate about some problems if data are as limited as in a typical "-omics" research report. This means that many potential problems, such as those of bias discussed above, may remain unidentified and unaddressed. The solution is not to post more raw data on the Web, although doing so may be useful for some purposes. Rather, the solution is to provide appropriate attention to the process of design, conduct, and interpretation of research and to thoroughly report that process in rigorously reviewed journal articles. In other words, what needs to be reported is not raw data but rather the process—how patients, specimens, and analyses were handled—that lead to those data *(23)*. Rules of evidence may then be applied to determine whether bias or chance provides an alternate explanation for results.

Improving the conduct and reporting of "-omics" research will take substantial effort but is worth it. An explosion of knowledge of biology, coupled with development of powerful measurement tools such as the polymerase chain reaction and mass spectroscopy, offers enormous opportunity to learn about etiology, diagnosis, and prognosis in multiple "-omics" fields. An important challenge is how to explore these fields efficiently. Although principal obstacles to this effort may include "…a lack of reagents and a lack of standards…" that could be addressed by "… focused work on reagents, technology, and informatics" *(48)*, a separate but critical obstacle is the process—currently suboptimal—by which exploration is conducted and by which results and claims are evaluated. As part of that process, chance and bias must be explicitly addressed in every study, even small ones. Indeed, important questions may be answered from small studies, even before problems such as standardization of technology and reagents are fully addressed. Although Liotta et al. argue that there is a "need for optimization and… standardization before analysis of reproducibility," a strong case can be made for just the opposite approach—i.e., that reproducibility should be demonstrated in small well-done studies that rigorously avoid chance and bias—before standardization and optimization are considered to be worth the effort. Chance and bias can be effectively addressed even in small studies, if they are done well. It is instructive that, for evaluations of therapy, one or two well-done clinical trials may provide much stronger evidence or proof of principle than a large number of less well-done (e.g., studies that are not randomized controlled clinical trials) studies. The strength of a randomized controlled clinical trial derives from its ability, through the experimental method, to avoid problems from chance and bias *(23)*. In summary, important questions can be addressed in small studies that avoid problems of chance and bias, even before larger problems of standardization and optimization have been solved.

This era provides unparalleled opportunity to learn how molecular markers may be used in diagnosis and prognosis. We know more basic biology than ever before and have powerful new tools to explore that biology. Exploration will be more efficient and successful if we apply appropriate rules of evidence to direct that exploration and to determine when results can support strong claims and high expectations.

REFERENCES

(1) Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359:572–7.

(2) Zhu W, Wang X, Ma Y, Rao M, Glimm J, Kovach JS. Detection of cancer-specific markers amid massive mass spectral data. Proc Natl Acad Sci U S A 2003;100:14666–71.

(3) Marcus A. Testing for ovarian cancer is on the way. Wall St J, October 1,2002: D1, D2.

(4) Pollack A. New cancer test stirs hope and concern. N York Times, February 3,2004:D1,D6.

(5) Food and Drug Administration. Letter to Correlogic Systems, Inc., February 18, 2004. Available at: http://www.fda.gov/cdrh/oivd/letters/021804-correlogic.html. [Last accessed: January 14, 2005.]

(6) Food and Drug Administration. Letter to Quest Diagnostics, March 2,2004. Available at: http://www.fda.gov/cdrh/oivd/letters/030204-quest.html. [Last accessed: January 14, 2005.]

(7) Food and Drug Administration. Letter to Laboratory Corporation of America, March 2, 2004. Available at: http://www.fda.gov/cdrh/oivd/letters/030204-labcorp.html. [Last accessed: January 15, 2005.]

(8) Wagner L. A test before its time? FDA stalls distribution process of proteomic test. J Natl Cancer Inst 2004;96:500–1.

(9) Diamandis E. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. J Natl Cancer Inst 2004;96: 353–6.

(10) Check E. Running before we can walk? Nature 2004;429:496–7.

(11) Garber K. Debate rages over proteomic patterns. J Natl Cancer Inst 2004;96:816–8.

(12) Baggerly KA, Morris JS, Edmonson SR, Coombes KR. Signal in noise: evaluating reported reproducibility of serum proteomics tests for ovarian cancer. J Natl Cancer Inst 2005;97:307–9.

(13) Liotta LA, Lowenthal M, Conrads TP, Veenstra TD, Fishman DA, Petricoin EF III. Misinformation generated by lack of communication between producers and consumers of publicly available experimental data. J Natl Cancer Inst 2005;97:310–4.

(14) Jasny BR, Roberts L. Solving gene expression. Science 2004;306:629.

(15) Marshall E. Getting the noise out of gene arrays. Science 2004;306:630–1.

(16) Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;299:926–30.

(17) Ransohoff DF. Challenges and opportunities in evaluating diagnostic tests. J Clin Epidemiol 2002;55:1178–82.

(18) Potter JD. At the interfaces of epidemiology, genetics and genomics. Nat Rev Genet 2001;2:142–7.

(19) Sullivan Pepe M, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;93:1054–61.

(20) Ransohoff DF. Developing molecular biomarkers for cancer. Science 2003;299:1679–80.

(21) Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer 2004;4:309–14.

(22) Ransohoff DF. Evaluating discovery-based research: when biologic reasoning cannot work. Gastroenterology 2004;127:1028.

(23) Ransohoff DF. Bias as a threat to validity of cancer molecular-marker research. Nat Rev Cancer 2005;5:142–9.

(24) Feinstein AR. Clinical epidemiology: the architecture of clinical research. Philadelphia (PA): WB Saunders, 1985.

(25) Hennekens CH, Buring JE. Epidemiology in medicine. Boston (MA): Little, Brown, 1987.

(26) Sackett DL, Haynes RB, Tugwell P, Guyatt GH. Clinical epidemiology: a basic science for clinical medicine. Boston (MA): Little, Brown, 1991.

(27) Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology: the essentials. 3rd ed. Baltimore (MD): Williams & Wilkins, 1996.

(28) Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. Designing clinical research: an epidemiologic approach. Philadelphia (PA): Lippincott Williams & Wilkins, 2001.

(29) Sackett DL, Haynes RB. The architecture of diagnostic research. Br Med J 2002;324:539–41.

(30) Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J Natl Cancer Inst 2003;95:14–8.

(31) Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 2002;346:1937–47.

(32) van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.

(33) Ransohoff DF. Gene-expression signatures in breast cancer. N Engl J Med 2003;348:1715–7; author reply 1716–17.

(34) Ransohoff DF. Discovery-based research and fishing. Gastroenterology 2003;125:290.

(35) Piccart M, Loi S, van't Veer L, Saghatchian-d'Assignies M, Glass A, Ellis P, et al. Multi-center external validation study of the Amsterdam 70-gene prognostic signature in node negative untreated breast cancer: are the results still outperforming the clinical-pathological criteria? Abstract presented at San Antonio Breast Cancer Symposium, December 8. Available at: http://www. abstracts2view.com/sabcs/search.php?queryxxxxxxx=Piccart&where=authors&intMaxHits=10&search=do. [Last accessed: December 23, 2004.]

(36) Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, et al. Gene expression predictors of breast cancer outcomes. Lancet 2003;361:1590–6.

(37) Masters JR, Lakhani SR. How diagnosis with microarrays can help cancer patients. Nature 2000;404:921.

(38) Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–26.

(39) Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. Lancet 2003;362:1439–44.

(40) Taubes G. Epidemiology faces its limits [news] [see comments]. Science 1995;269:164–9.

(41) Ransohoff DF. Research opportunity at the interface of molecular biology and clinical epidemiology. Gastroenterology 2002;122:1199.

(42) Zhang Z, Bast RC Jr, Yu Y, Li J, Sokoll LJ, Rai AJ, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res 2004;64:5882–90.

(43) Grizzle WE, Adam BL, Bigbee WL, Conrads TP, Carroll C, Feng Z, et al. Serum protein expression profiling for cancer detection: validation of a SELDI-based approach for prostate cancer. Dis Markers 2003;19:185–95.

(44) Sackett DL. Zlinkoff honor lecture: basic research, clinical research, clinical epidemiology, and general internal medicine. J Gen Intern Med 1987;2:40–7.

(45) Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A 2002;99:6562–6.

(46) Baggerly KA, Morris JS, Wang J, Gold D, Xiao LC, Coombes KR. A comprehensive approach to the analysis of matrix-assisted laser desorption/ionization-time of flight proteomics spectra from serum samples. Proteomics 2003;3:1667–72.

(47) Mehta T, Tanik M, Allison DB. Towards sound epistemological foundations of statistical methods for high-dimensional biology. Nat Genet 2004;36:943–7.

(48) Biomarkers and quality of care, key presentations at BSA meeting. NCI Cancer Bull, November 16, 2004;1:5.

NOTES