

# High-resolution serum proteomic patterns for ovarian cancer detection

Lance A Liotta, Emmanuel F Petricoin<sup>1</sup>, Timothy D Veenstra<sup>2</sup>  
and Thomas P Conrads<sup>2</sup>

Laboratory of Pathology — National Cancer Institute Clinical Proteomics Program, National Cancer Institute Center for Cancer Research, Bethesda, MD 20892, USA

<sup>1</sup>Food and Drug Administration — National Cancer Institute Clinical Proteomics Program, Center for Biologics and Evaluation Research, Food and Drug Administration, Bethesda, MD 20892, USA

<sup>2</sup>National Cancer Institute Biomedical Proteomics Program, Laboratory of Proteomics and Analytical Technologies, SAIC-Frederick, Inc., National Cancer Institute at Frederick, Frederick, MD 21702, USA

(Requests for offprints should be addressed to T D Veenstra; Email: Veenstra@ncicrf.gov)

*Endocrine-Related Cancer* (2004) 11 585–587

We appreciate the attention given to our article (Conrads *et al.* 2004) by Baggerly *et al.* (see previous Letter to the Editor, this issue); however, we are confused by some of the points raised within their letter because the authors seem to indicate that they have discovered anew some type of insidious experimental design flaw that we were not aware of and that this design flaw endangers the validity of our conclusions. This insinuation is not the case. The experimental design element they highlight in their criticism was explicitly planned in the study we reported in *Endocrine-Related Cancer* to isolate critical variables for proteomic pattern analysis utilizing mass spectrometry (MS). While we believe their concern is well intentioned in nature, it is clearly misguided and unfortunate, serving in actuality to only confuse a reader of the article as well as those directly involved in research within this still emerging and maturing field of study. We have never claimed or intimated that the samples were randomized and/or co-mingled in the initial experimental design, and in fact the quality assurance/quality control (QA/QC) report that we developed in the paper under scrutiny, which is also available on our website, explicitly details the fact that the cancer and control samples were not randomized and/or co-mingled. The letter by Baggerly *et al.* goes on to cite this lack of randomization as a ‘serious problem’. Again, we disagree. In our report, we state clearly that the three goals of this study were to begin to collect data for understanding critical unanswered questions in the field of proteomic pattern diagnostics, as we detail below.

First, does the use of high-resolution time-of-flight (TOF) MS for gathering proteomic patterns from surface-enhanced laser/desorption ionization (SELDI) ProteinChip arrays yield better analytical and clinical sensitivity and specificity compared with low-resolution

instrumentation, at least for the set of serum samples analyzed within this study? According to National Committee for Clinical Laboratory Standards (NCCLS) experimental design criteria, a single variable was isolated (i.e. the type of mass spectrometer) to answer this question. Since we were able to analyze the exact same SELDI ProteinChip spot with two different mass spectrometers, a direct comparison could be attempted.

Secondly, which contributes the greatest source of variability: the heterogeneity within and between clinical serum sets (i.e. normal study set versus cancer study set) or the sample application or MS process itself? To address this question, reference standards were used and the variable to be tested was fixed and isolated. Thus, we could not randomize and co-mingle the cancers and controls at the same time since we wanted to determine the variability within the mass spectrometer/sample application within a given run cycle and test that variable on a common set of samples: we chose not to co-vary two independent variables at once (run date and phenotype). This experimental design allowed us to answer the important question and show that, for this study set, the variability within the process itself was greater than the variability within the sample sets.

Thirdly, does the development of spectral QA/QC procedures positively impact the modeling performance? In other words, does elimination of ‘bad-looking spectra’ contribute to better-performing models?

Baggerly *et al.* seem to indicate that running the cancer samples on the third day of spectral acquisition, where the majority of lesser-quality mass spectra appear, will negatively impact any findings and downstream modeling. This statement is also confusing to us, given our clearly stated goals and objectives of this work. We viewed this event, which would almost certainly occur in a

true clinical setting, as a fortuitous opportunity to determine the impact of in-process controls and QA/QC criteria on the modeling outcome. In fact, we do not view this as a weakness at all, but a major strength of this study. It could be argued that if this was not our goal, we would not have included these data or description in the paper to begin with, but simply described the results obtained post-QA/QC! Such an example is exactly what would be required in NCCLS and Food and Drug Administration design control for clinical test development and accelerated stability protocols. In our clearly stated design, modeling was performed before and after QA/QC was imparted on the process. Lesser-quality mass spectra or, more accurately, spectra with lower total record counts, average amplitude, etc. were eliminated by our QA/QC procedures. Elimination of these mass spectra, where only those spectra that ‘passed’ our QA/QC procedure were retained, made the modeling more robust. This analysis, of course, was one of the major points of this paper, along with understanding the source of the greatest variability — in the specimens (claims had been that the variability arose from specimens) — that is, collection method bias between the cancers and controls: this was tested by running them in batches to test variability within each set.

Some cancer SELDI mass spectra were clearly affected when the mass spectrometer was beginning to fail on the third day as detailed in our QA/QC report and in the publication itself. However we cannot detect whether the cancer data acquired on the previous day were convincingly negatively affected by the spectrometer failure. Since we were able to correctly classify the cancers in the blinded test set regardless of the day the mass spectra were acquired, we postulate that our QA/QC procedures appear to be able to eliminate those cancers that were affected at the end of the run, and those that pass our QA/QC procedures performed better in the modeling.

It is extremely important to note that it cannot be proven that the features we identified are changing as a result of experimental bias since they also track with the biology. The argument cannot be used by either ‘biased’ side of the argument (biology versus bias) since both phenotype (cancer or control or biology) and run date (bias) are concordant and track together.

We are also perplexed that Baggerly *et al.* claim so dogmatically that other signs of bias have been found conclusively. Unfortunately the signs of systematic bias cited by the authors in two previous papers cannot be proven as ‘signs of bias’ since these ‘signs’ were based on the discovery of features with discriminatory power at mass/charge values that they deemed too small to make biologic sense. This claim is interesting since there are an

extraordinary large number of unknown metabolic and low-molecular-weight molecules that have yet to be identified and cataloged. Moreover, there exists a number of known molecules, such as lysophosphatidic acid (known to be an important molecule in ovarian cancer biology), with masses that would be contained within the same size range that have been deemed, without supporting scientific data, to be nothing but noise and devoid of biological information. It is our opinion that it is scientifically indefensible to simply label a mass/charge range as ‘noise’ and filled with ‘bias’ without *a priori* knowledge of the nature and identity of the biomolecule constituency contained within that mass/charge range. This sweeping assumption would in fact dismiss the entire field of metabonomics as lacking in any meaningful biological data!

It is also curious why Baggerly *et al.* would raise criticisms originally brought forth by others (Diamandis 2004) concerning the nature and identity of the features identified in this study as ‘valid’ when these issues are completely unproven experimentally and at present are speculative at best. The focus of this study was not to specifically address issues raised by Dr Diamandis and the inclusion of this matter by Baggerly *et al.* possibly reveals a potentially disturbing undercurrent of ‘bias’ in their interpretation, the motivation of which is unknown. The related statement that “the machine being used should allow for substantially easier identification of the peaks involved” is misleading and can be attributed to this groups’ possible lack of understanding of MS technology. The features identified in this study as being consistent within several of the diagnostic models with the highest diagnostic accuracy have low (i.e. approximately 2–4) signal-to-noise ratios. It would be impossible to isolate sufficient quantities of ions directly from the chip surface to obtain a reasonable tandem mass spectrum that would allow for even partial sequencing of these species. The authors may not be aware that it is physically impossible to even select for an ion with a mass-to-charge ratio greater than 3000 for tandem MS using the high-resolution mass spectrometer employed in this study.

We are equally confounded by the last sentence within the letter by Baggerly *et al.* that the “claims of 100% sensitivity and specificity are premature”. We certainly did identify a combination of features that were able to discriminate 100% of a separate testing set. This fact is clear and we are unsure how this claim is premature: we either attained this level of accuracy or we did not. We readily acknowledge that it would be premature to claim that we have identified a feature set that will have 100% accuracy as a clinical test. Since we are not developing a commercial clinical test, and at this time are conducting research to (1) identify sources of experimental variability,

(2) continue in platform development, and (3) develop and evaluate in-process controls (i.e. QA/QC), this claim has never been made. These three goals were described clearly within the report; however, the identification of a selected set of features that were to be used in diagnosis of ovarian cancer was never included as an aim for this study. We are unclear as to why this fact is being ignored by Baggerly *et al.*

Lastly, we should stress and highlight that the only true way to prove that any feature is devoid of impact from experimental bias — from sample collection bias all the way to MS analysis — is to identify the features. This focus is now one of our main goals. In fact, our joint discoveries that the diagnostic information content of the low-molecular-weight blood proteome exists mostly in a complexed state with high-abundance carrier proteins (Liotta *et al.* 2003, Mehta *et al.* 2003, Tirumalai *et al.* 2003, Zhou *et al.* 2004) now allows us to develop new methodologies where the bound information archive can be used as an input for high-throughput matrix-assisted laser/desorption ionization (MALDI)-TOF fingerprinting/pattern recognition and serves as the input for direct protein identification and MS-based tryptic fragment identification to generate a list of all bound protein fragments and low-molecular-weight protein biomarkers. This effort is currently underway and, as shown by the publications we have generated in this area, is moving forward at a rapid pace.

To reiterate, we want to make it abundantly clear that our studies do not detail a final clinical test, but are

research driven studies. The results of these studies may or may not be brought forward for further clinical validation. The paper makes no claims that the features found are being developed as a clinical test. We believe that the circumscribed goals of the paper were met with high rigor and the data and all of our analysis displayed in the most transparent and unrestricted way possible.

## References

- Conrads TP, Fusaro VA, Ross S, Johann D, Rajapakse V, Hitt BA, Steinberg SM, Kohn EC, Fishman DA, Whitely G *et al.* 2004 High resolution serum proteomic features for ovarian cancer detection. *Endocrine-Related Cancer* **11** 163–178.
- Diamandis EP 2004 MS as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Molecular and Cellular Proteomics* **3** 367–378.
- Liotta LA, Ferrari M & Petricoin EF 2003 Clinical proteomics: written in blood. *Nature* **425** 905.
- Mehta AI, Ross S, Lowenthal MS, Fusaro V, Fishman DA, Petricoin EF & Liotta LA 2003 Biomarker amplification by serum carrier protein binding. *Disease Markers* **19** 1–10.
- Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP & Veenstra TD 2003 Characterization of the low molecular weight human serum proteome. *Molecular and Cellular Proteomics* **2** 1096–1103.
- Zhou M, Lucas DA, Chan KC, Issaq HJ, Petricoin EF, Liotta LA, Veenstra TD & Conrads TP 2004 An investigation into the human serum ‘interactome’. *Electrophoresis* **25** 1289–1298.