Author's response to reviews

Title: Do serum biomarkers really measure breast cancer?

Authors:

Jonathan L Jesneck (JonathanL_Jesneck@dfci.harvard.edu)
Sayan Mukherjee (sayan@stat.duke.edu)
Zoya Yurkovetsky (yurkovetskyz@upmc.edu)
Merlise Clyde (clyde@stat.duke.edu)
Jeffrey R Marks (marks003@mc.duke.edu)
Anna E Lokshin (marks003@mc.duke.edu)
Joseph Y Lo (Joseph.Lo@duke.edu)

Version: 3 Date: 16 April 2009

Author's response to reviews: see over
April 16, 2008

Rikki Graham, PhD
Senior Assistant Editor
BMC-series journals

Dear Dr. Graham,

Please find the revised version of manuscript 1153141180219503, “Do serum biomarkers really measure breast cancer?” according to the referees’ comments. My coauthors and I have addressed all referees’ comments, as described in the enclosed “Response to referee” section.

On behalf of all co-authors, I thank you for your consideration. Please let me know if any additional information is needed.

With my best regards,

Jonathan Jesneck, Ph.D.
Research Fellow
Broad Institute and
Dana-Farber Cancer Institute
44 Binney St., Room D640
Boston, MA 02215
Tel.: 617.632.6782
Fax.: 617.632.4410
Email: JonathanL_Jesneck@dfci.harvard.edu
or jesneck@broad.mit.edu

Encl: Response to referees
Response to referee

We thank Dr. Petricoin for his insightful comments. Our responses are listed below.

Note: Manuscript changes are shown by the revision-tracking feature of Microsoft Word.

Minor essential revisions:

1. This is a good suggestion. Although in our previous submission we initially intended to simplify the manuscript by including classification results for only cross-validation, we have now included the classification results for a train/test split of the data as well. Additional plots have been added to Figures 1 and 2, and we have added text to the Methods, Results, and Discussion. See
   a. Page 12, section “Evaluating classification performance”, sentences 3-4: “we first defined … cross-validation (LOOCV),”
   b. Page 13, section “Classifier performance”, sentence 1: “AUC=0.77 … for LOOCV).”
   c. Page 16, section “Discussion”, paragraph 1, sentences 2-3: “These classification results … our data set.”
   d. Page 29, Figure 1 legend, sentences 2-3: “The models were … benign lesions (C,F).”
   e. Page 29, Figure 2 legend, sentences 1-2: “Bayesian model averaging … benign lesions (C,F).”

2. The reviewer posts several very serious and important questions, for which there is no single answer. Different analytes have different CVs that depend on the reagents used to measure their concentrations. In general, there are three sources of variability of Luminex data: (i) Accuracy of individual measurements (Luminex reader). This error is negligible provided instruments are calibrated according to SOPs and sufficient number of beads (not less than 100) is counted. (ii) Variability in sample preparation and operator factor; and (iii) Variability in antibodies and/or concentration standards from different lots (quality of reagents). This factor contributes the most to the variability. Below, we address reviewer’s points to the best of our ability.
   a. Complete information about characteristics of individual assays including inter- and intra-assay CVs is available from the manufacturers of assays (see Supplementary Table 2) and from the Luminex Core website (http://www.upci.upmc.edu/luminex/sources.cfm). In summary, the intra-assay CV for different analytes was in the range of 0.7-11%
(typically < 5%) and inter-assay 3.7-19% (<11% for same lot reagents).

b. One replicate per patient sample was performed with reactions from 100 beads measured and averaged. There are two reasons of performing assays in several replicates. One is to obtain accurate information on reaction kinetics by averaging several reactions = wells on ELISA plate, and the other is to minimize potential operator error. In our case, the estimate of reaction variability is derived from 100 reactions/beads. Average operator error in Dr. Lokshin’s laboratory is <1%. This conclusion is based on the results obtained from the study of multiple biomarkers in over 50 quality control samples analyzed in over 300 different runs over the course of 3 years.

c. Based on our analysis of assays performed monthly within one-month interval for 3 months using the same lot of reagents, the calculated inter-assay CV was less than 11%. When different lots of reagents are used (all reagents used for this study were research grade), we perform additional normalization using quality control samples. These quality control samples have been aliquoted from large volumes of sera from healthy men and women and from patients with pancreatic and ovarian cancers. The samples were selected so that concentrations of individual analytes across the samples would span dynamic ranges of most or all individual analytes. Control samples are included in each run and can be used as reference points for normalization. At least 8 control samples are included in each run. This way, all measurements are normalized regardless of whether they utilize the same of different lots of reagents. Ideally, this normalization procedure should reduce effects of lot-to-lot reagents variability below stochastic CV of quality control samples. In practice, the combined inter-assay CV for assays performed using different lots of reagents and normalized using quality controls samples varied between 1% and 19% depending on analyte.

d. The above points reflect our general approach. The initial biomarker discovery phase is performed using large number of biomarkers and research grade reagents in single replicates. After the most robust classification panel is identified and validated in a blinded validation set, the reagents for this panel are re-developed according to clinical grade standards. For clinical trials, all tests will be performed in triplicates.

e. Briefly summarizing some of this information in the text, we have updated the manuscript. See pages 6-7, section “Measuring serum protein levels with ELISA”, paragraph 1, sentences 6-9: “One replicate per patient … same lot of reagents.”