

Research article

Open Access

Do serum biomarkers really measure breast cancer?

Jonathan L Jesneck^{1,2}, Sayan Mukherjee^{3,4,5}, Zoya Yurkovetsky^{6,7},
Merlise Clyde³, Jeffrey R Marks^{4,8}, Anna E Lokshin^{6,7} and Joseph Y Lo^{*9,10,11}

Address: ¹Cancer Program, Broad Institute of Harvard University and Massachusetts Institute of Technology, Cambridge, MA 02142, USA, ²Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA, ³Department of Statistical Science, Duke University, Durham, NC 27708, USA, ⁴Institute for Genome Sciences and Policy, Duke University, Durham, NC 27708, USA, ⁵Department of Computer Science, Duke University, Durham, NC 27708, USA, ⁶Division of Hematology/Oncology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA, ⁷University of Pittsburgh Cancer Institute, Pittsburgh, PA 15213, USA, ⁸Department of Experimental Surgery, Duke University Medical Center, Durham, North Carolina 27710, USA, ⁹Duke Advanced Imaging Labs, Department of Radiology, Duke University Medical Center, Durham, North Carolina 27710, USA, ¹⁰Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA and ¹¹Medical Physics Graduate Program, Duke University, Durham, NC 27708, USA

Email: 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 - loksax@UPMC.EDU; Joseph Y Lo* - Joseph.Lo@duke.edu

* Corresponding author

Published: 28 May 2009

Received: 6 September 2008

BMC Cancer 2009, 9:164 doi:10.1186/1471-2407-9-164

Accepted: 28 May 2009

This article is available from: <http://www.biomedcentral.com/1471-2407/9/164>

© 2009 Jesneck et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Because screening mammography for breast cancer is less effective for premenopausal women, we investigated the feasibility of a diagnostic blood test using serum proteins.

Methods: This study used a set of 98 serum proteins and chose diagnostically relevant subsets via various feature-selection techniques. Because of significant noise in the data set, we applied iterated Bayesian model averaging to account for model selection uncertainty and to improve generalization performance. We assessed generalization performance using leave-one-out cross-validation (LOOCV) and receiver operating characteristic (ROC) curve analysis.

Results: The classifiers were able to distinguish normal tissue from breast cancer with a classification performance of $AUC = 0.82 \pm 0.04$ with the proteins MIF, MMP-9, and MPO. The classifiers distinguished normal tissue from benign lesions similarly at $AUC = 0.80 \pm 0.05$. However, the serum proteins of benign and malignant lesions were indistinguishable ($AUC = 0.55 \pm 0.06$). The classification tasks of normal vs. cancer and normal vs. benign selected the same top feature: MIF, which suggests that the biomarkers indicated inflammatory response rather than cancer.

Conclusion: Overall, the selected serum proteins showed moderate ability for detecting lesions. However, they are probably more indicative of secondary effects such as inflammation rather than specific for malignancy.

Background

The high prevalence of breast cancer motivates the development of better screening and diagnostic technologies. To complement mammography screening, which has

moderate sensitivity [1] and specificity [2-4] and low positive-predictive value in younger women [1], researchers have investigated the efficacy of detecting breast cancer using proteins. Proteins offer detailed information about

tissue health conditions, allowing the identification of cancer type and risk, and thereby prompting potentially better-targeted and more effective treatment. Some studies have correlated breast cancer prognosis with proteins in the tumor [5], such as hormone receptors, HER-2, urokinase plasminogen activator, and plasminogen activator inhibitor 1 [6,7], and caPCNA [8]. However, accessing these local proteins requires biopsies, which is not practical for a screening regimen. We therefore consider serum proteins. Serum and plasma protein-based screening tests have already been developed for many diseases, such as Alzheimer's disease [9], cardiovascular disease [10], prostate cancer [11], and ovarian cancer [12].

For breast cancer, however, there are currently very few serum markers used clinically. Some studies have identified as possible breast cancer markers the proteins CA 15.3 [13-15], BR 27.29, tissue polypeptide antigen (TPA), tissue polypeptide specific antigen (TPS), shed HER-2 [15], and BC1, BC2, and BC3 [16,17]. However, other studies found a lack of sufficient diagnostic ability in serum proteins, including CA 15.3 [13,17-20], CA 125 [20], CA 19.9 [20], CA 125 [20], BR 27.29 [13,18], and carcinoembryonic antigen (CEA) [13]. The European Group of Tumor Markers identified the MUC-1 mucin glycoproteins CA 15.3 and BR 27.29 as the best serum markers for breast cancer, but they could not recommend these proteins for diagnosis due to low sensitivity [18].

Cancer biomarkers are valued according to their specificity and sensitivity, suiting them for different clinical roles. For example, general population screening requires high sensitivity but not necessarily high specificity, if a low-cost secondary screen is available. Conversely, disease recurrence monitoring requires high specificity but not necessarily high sensitivity, if a more sensitive secondary test is available. Therefore to optimize clinical utility, it is important to measure the sensitivity and specificity of any proposed biomarker. Because solid tumors cause many changes in the surrounding tissue, it is likely that some potential biomarkers measure the body's response to cancer rather than the cancer itself. Cancer marker levels may be increased due to various secondary factors, such as therapy-mediated response [21-23] and benign diseases. For example, CA 15-3 levels increase in chronic active hepatitis, liver cirrhosis, sarcoidosis [24], hypothyroidism [25], megablastic anemia [26], and beta-thalassaemia [27]. To help measure biomarker specificity in our study, we included normal, benign, and malignant samples.

In general, breast cancer biomarker studies have found individual circulating proteins, but it is important to consider covariation of multiple protein levels. It is useful to know which combinations of proteins may yield high diagnostic performance, even though each protein individually might yield low performance [28,29]. Some stud-

ies have detected discriminatory breast cancer biomarkers using mass spectrometry. [16,30-36] Because of mass spectrometry's peak interpretation and reproducibility challenges [37,38], scientists have searched for breast cancer biomarkers from predefined collections of known candidate markers using newer multiplex technologies, such as reverse-phase protein microarray [39]. To our knowledge, ours is the first study to assess a large set of serum proteins collectively by a sensitive and specific multiplex assay in order to identify the most promising proteins for breast cancer detection.

Although studies have shown correlations between serum proteins and breast cancer, it is often unclear how these correlations translate into clinical applicability for diagnosis. To quantify the diagnostic performance of a set of proposed biomarkers, we have implemented many statistical and machine-learning models. We have focused on Bayesian models, which provide full posterior predictive distributions. This detailed information would help a physician to judge how much emphasis to place on the classifier's diagnostic prediction.

The goal of this study was to quantify the ability of serum proteins to detect breast cancer. For measuring predictive performance on this noisy data set, we used many statistical classification models. To better understand the cancer-specificity of the screening test, we ran the classifiers on proteins from malignant lesions, benign lesions, and normal breast tissue. Our data indicated that some serum proteins can detect moderately well the presence of a breast lesion, but they could not distinguish benign from malignant lesions.

Methods

Enrolling subjects

This study enrolled 97 subjects undergoing diagnostic biopsy for breast cancer and 68 normal controls at Duke University Medical Center from June 1999 to October 2005. Women donating blood to this study were either undergoing image guided biopsy to diagnose a primary breast lesion (benign and cancer) or were undergoing routine screening mammography and had no evidence of breast abnormalities (normal). All subjects were enrolled only after obtaining written informed consent for this Duke IRB approved study (P.I., JRM, current registry number, 9204-07-11E1ER, Early Detection of Breast Cancer Using Circulating Markers). All subjects were premenopausal women. Table 1 shows the demographics of the study population. Additional file 1 shows subjects' ages and histology findings.

Measuring serum protein levels with ELISA

Blood sera were collected under the HIPAA-compliant protocol "Blood and Tissue Bank for the Discovery and Validation of Circulating Breast Cancer Markers." Blood

Table 1: Subject demographics

	Normal	Benign	Malignant	Total
Number of subjects	68 (41%)	48 (29%)	49 (30%)	165 (100%)
Mean age (years)	36 ± 8	38 ± 9	42 ± 4	38 ± 8
Race: Black	23 (41%)	19 (34%)	14 (25%)	56 (34%)
Race: White	45 (41%)	29 (27%)	35 (32%)	109 (66%)

was collected from subjects prior to surgical resection. All specimens were collected in red-stoppered tubes and processed generally within 4 hours (but not greater than 12 hours) after collection and stored at -80°C. Sera were assayed using the Enzyme-Linked ImmunoSorbent Assay (ELISA, Luminex platform) and reagents in the Luminex Core Facility of University of Pittsburgh Cancer Institute. The Luminex protocol was optimized for analytical performance as described by Gorelik *et al.* [12]. One replicate per patient sample was performed with reactions from 100 beads measured and averaged. All samples were analyzed on the same day using the same lot of reagents. Complete information about characteristics of individual assays including inter- and intra-assay coefficients of variation (CVs) is available from the manufacturers of assays [see Additional file 2] and from the Luminex Core website <http://www.upci.upmc.edu/luminex/sources.cfm>. Based on our analysis of assays performed monthly within one month interval for 3 months using the same lot of reagents, 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).

Biomarkers were selected based on the known literature reports about their association with breast cancer. The 98 assayed proteins are shown in Table 2, with further details in Additional file 2. In addition to the protein levels, patient age and race were also recorded.

Regression with variable selection

In order to incorporate these proteins into a breast cancer screening tool, we built statistical models linking the protein levels to the probability of malignancy. We used the following three common regression models: linear regression $Y_i = X_i\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$, logistic regression $\Pr(Y_i = 1 | \beta) = \frac{e^{X_i\beta}}{1 + e^{X_i\beta}}$, and probit regression $\Pr(Y_i = 1 | \beta) = \Phi(X_i, \beta)$, where Y is the response vector (breast cancer diagnosis), X is the matrix of observed data (protein levels), β is the vector of coefficients, ε is additive noise, and $\Phi(\cdot)$ is the cumulative distribution function of the normal distribution. These classical models become unstable

and predict poorly if there are relatively few observations (curse of dataset sparsity [40]) and many features (curse of dimensionality [41]). It is better to choose a subset of useful features. We used stepwise feature selection [42-46] to choose the set of proteins that optimized model fit.

However, choosing only one feature subset for prediction comes with an inherent risk. When multiple possible statistical models fit the observed data similarly well, it is risky to make inferences and predictions based only on a single model [47]. In this case predictive performance suffers, because standard statistical inference typically ignores model uncertainty.

Accounting for model uncertainty with Bayesian model averaging

We accounted for model-selection ambiguity by using a Bayesian approach to average over the possible models. We considered a set of models M_1, \dots, M_B , where each model M_k consists of a family of distributions $p(D | \theta_k, M_k)$ indexed by the parameter vector θ_k , where $D = (X, Y)$ is the observed data. Y is the response vector (breast cancer diagnosis), and X is the matrix of observed data (protein levels). Using a Bayesian method to average over the set of considered models [47-52], we first assigned a prior probability distribution $p(D | M_k)$ to the parameters of each model M_k . This formulation allows a conditional factorization of the joint distribution,

$$p(D, \theta_k, M_k) = p(D | \theta_k, M_k)p(\theta_k | M_k)p(M_k). \quad (1)$$

Splitting the joint distribution in this way allowed us to implicitly embed the various models inside one large hierarchical mixture model. This form allowed us to fit these models using the computational machinery of Bayesian model averaging (BMA).

BMA accounts for model uncertainty by averaging over the posterior distributions of multiple models, allowing for more robust predictive performance. If we are interested in predicting a future observation D_f from the same process that generated the observed data D , then we can represent the predictive posterior distribution $p(D_f | D)$ as an average over the models, weighted by their posterior probabilities [47,53,54]:

$$p(D_f | D) = \sum_k p(D_f | D, M_k)p(M_k | D) \quad (2)$$

where the sum's first term $p(D_f | D, M_k)$ is a posterior weighted mixture of conditional predictive distributions

$$p(D_f | D, M_k) = \int p(D_f | \theta_k, M_k)p(\theta_k | D, M_k)d\theta_k \quad (3)$$

Table 2: List of the 98 serum proteins measured by ELISA assay (Luminex platform)

ACTH	Adiponectin	AFP	Angiostatin
Apolipoprotein A1	Apolipoprotein Apo A2	Apolipoprotein Apo B	Apolipoprotein Apo C2
Apolipoprotein Apo C3	Apolipoprotein Apo E	CA 15-3	CA 19-9
CA-125	CA72-4	CD40L (TRAP)	CEA
Cytokeratin 19	DR5	EGF	EGFR
EOTAXIN	ErbB2	FGF-b	Fibrinogen
Fractalkine	FSH	G-CSF	GH
GM-CSF	GROa	Haptoglobin	HGF
IFN-a	IFN-g	IGFBP-1	IL-10
IL-12p40	IL-13	IL-15	IL-17
IL-1a	IL-1b	IL-1Ra	IL-2
IL-2R	IL-4	IL-5	IL-6
IL-7	IL-8	IP-10	Kallikrein 10
Kallikrein 8	Leptin	LH	MCP-1
MCP-2	MCP-3	Mesothelin(IgY)	MICA
MIF	MIG	MIP-1a	MIP-1b
MMP-1	MMP-12	MMP-13	MMP-2
MMP-3	MMP-7	MMP-8	MMP-9
MPO	NGF	PAI-1(active)	PROLACTIN
RANTES	Resistin	S-100	SAA
SCC	sE-Selectin	sFas	sFasL
sICAM-1	sIL-6R	sVCAM-1	TGFa
TNF-a	TNF-RI	TNF-RII	tPAI-1
TSH	TTR	ULBP-1	ULBP-2
ULBP-3	VEGF		

and the sum's second term $p(M_k|D)$ is a model's posterior distribution

$$p(M_k | D) = \frac{p(D|M_k)p(M_k)}{\sum_k p(D|M_k)p(M_k)}, \quad (4)$$

which incorporates the model's marginal likelihood

$$p(D | M_k) = \int p(D | \theta_k, M_k)p(\theta_k | M_k)d\theta_k. \quad (5)$$

Promoting computational efficiency with iterated BMA

BMA allows us to average over all possible models, containing all possible subsets of features. However, considering many models would require extensive computations, especially when computing the posterior predictive distributions. Such computations would be prohibitively long for a quick screening tool that is intended not to impede clinicians' workflow during routine breast cancer screening. Because it was computationally infeasible to consider all possible $2^{100} \approx 1.26 * 10^{30}$ models, we first chose a set of the best fitting models. For computational efficiency in model selection, this study followed Yeung *et al.* [55] and used a deterministic search based on an Occam's window approach [54] and the "leaps and bounds" algorithm [56] to identify models with higher posterior probabilities.

In addition to choosing the best models, we also chose the best proteins. We applied an iterative adaptation of BMA [55]. This method initially ranks each feature separately by the ratio of between-group to within-group sum of squares (BSS/WSS) [57]. For protein j the ratio is

$$\frac{BSS(j)}{WSS(j)} = \frac{\sum_i \left[I(Y_i=0)(\bar{X}_{0j} - \bar{X}_j)^2 + I(Y_i=1)(\bar{X}_{1j} - \bar{X}_j)^2 \right]}{\sum_i \left[I(Y_i=0)(X_{ij} - \bar{X}_{0j})^2 + I(Y_i=1)(X_{ij} - \bar{X}_{1j})^2 \right]} \quad (6)$$

where $I(\cdot)$ is an indicator function, X_{ij} is the level of protein j in sample i , \bar{X}_{0j} and \bar{X}_{1j} are respectively the average levels of protein j in the normal and cancer groups, and \bar{X}_j is the average level of protein j over all samples.

Ordered by this BSS/WSS ranking, the features were iteratively fit into BMA models, which generated posterior probability distributions for the proteins' coefficients. We then discarded proteins that had low posterior probabilities of relevance, $\Pr(b_j \neq 0 | D) < 1\%$, where

$$\Pr(b_j \neq 0 | D) = \sum_{M_k \in \Gamma} \Pr(M_k | D) \quad (7)$$

where $\Pr(b_j \neq 0 | D)$ is the posterior probability that protein j 's coefficient is nonzero, Γ is the subset of the considered models M_1, \dots, M_B that include protein j . By discarding proteins that have small influence on classification, this iterative procedure keeps only the most relevant proteins.

Other models for high-dimensional data

To compare iterated BMA's classification and generalization performance, we also classified the data using two other dimensionality-reducing methods: a support vector machine (SVM) [58] with recursive feature selection [59,60] and least-angle regression (LAR, a development of LASSO) [61].

All modeling was performed using the R statistical software (version 2.6.2), and specifically the BMA package (version 3.0.3) for iterated BMA, the packages e1071 (version 1.5–16) and R-SVM [31] for the SVM with recursive feature selection, the lars package (version 0.9–5) for least angle regression, and the ROCR package (version 1.0–2) for ROC analysis. We extended the BMA package to compute the full predictive distributions (See Equations 2–6) within cross-validation using an MCMC approach. Additional file 3 contains the R code and Additional file 4 contains the data in comma-delimited format.

Evaluating classification performance

The classifiers' performances were analyzed and compared using receiver operating characteristic (ROC) analysis [62]. ROC curve metrics were compared statistically using a nonparametric bootstrap method [63]. To estimate generalization performance on future cases, we first defined independent train and test sets by randomly choosing 70% of samples for training and optimizing the classifiers, and then we tested the classifiers on the remaining 30% of the samples. To compare with the train/test split, we also performed leave-one-out cross-validation (LOOCV). Feature selection was performed within each fold of the cross-validation.

Table 3: Cross-validation classification errors, normal versus cancer

Model	FN	FP
BMA of linear models	19	8
BMA of logistic models	15	12
BMA of probit models	18	7
SVM with RFS	18	12
LAR	24	5

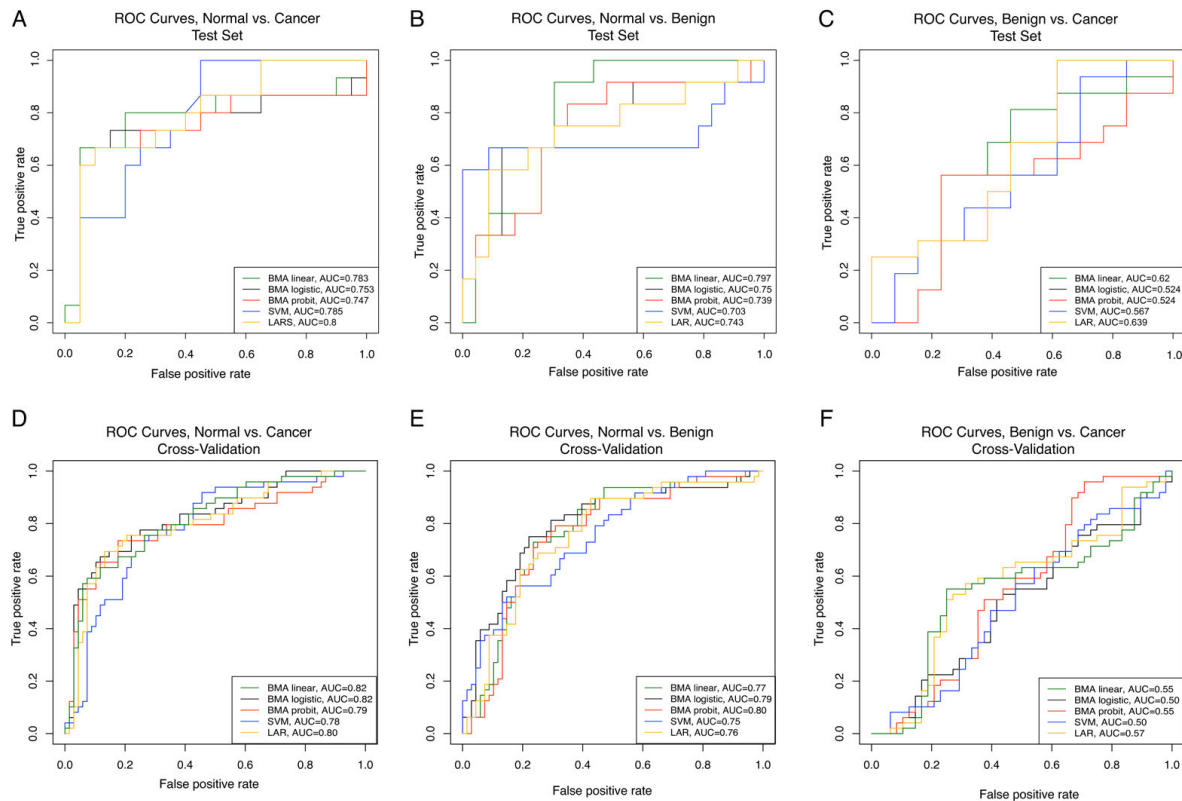


Figure 1

ROC curves showing the classification performance of statistical models using the serum protein levels. The models were run with a 70% train and 30% test split of the data set (A-C) and also with leave-one-out cross-validation (LOOCV) (D-F). The classifiers performed similarly, with moderate classification results for normal vs. malignant or benign lesions (A, B, D, E) and poor classification results for malignant vs. benign lesions (C, F).

Evaluating feature-Selection methods

We compared models with feature-selection techniques using two methods: feature concentration and classifier performance. For feature concentration, we performed feature selection within each fold of a LOOCV. We recorded how many times each feature was chosen. This method distinguished the feature selection methods that chose few versus many features. Using classifier performance, we investigated the effect of feature selection and sampling. We compared linear models run on the data with these four techniques: 1) no feature selection (using all the proteins in the model), 2) preselecting the features (using all the data to choose the best features, and then running the model using only those preselected features in LOOCV), 3) stepwise feature selection, and 4) iterated BMA. For each case, we ran the classifier in LOOCV.

Results

Classifier performance

The classifiers achieved moderate classification performance for both normal vs. malignant ($AUC = 0.77 \pm 0.02$ on

the test set, and $AUC = 0.80 \pm 0.02$ for LOOCV) and normal vs. benign ($AUC = 0.75 \pm 0.03$ on the test set, and $AUC = 0.77 \pm 0.02$ for LOOCV), but very poor performance for malignant vs. benign tumors ($AUC = 0.57 \pm 0.05$ on the test set, and $AUC = 0.53 \pm 0.03$ for LOOCV). The classification performance is shown as ROC curves in Figure 1. Whereas the ROC curves show the classification performance over the entire range of prediction thresholds, we also considered the threshold of $\Pr(Y_i = 1 | \beta) = 0.5$ in particular. For this threshold, Table 3 shows the classification error. The models performed similarly, with approximately 20 false negatives and 10 false positives. All classifiers were run with leave-one-out cross-validation (LOOCV). The classifiers chose the best subsets of proteins for classification.

Figure 2 plots the full posterior predictive distributions for BMA of probit models run with LOOCV. In general, the predictive distributions were more "decided" (concentrated further from the 0.5 probability line) for the tasks of normal vs. cancer and normal vs. benign, but they were

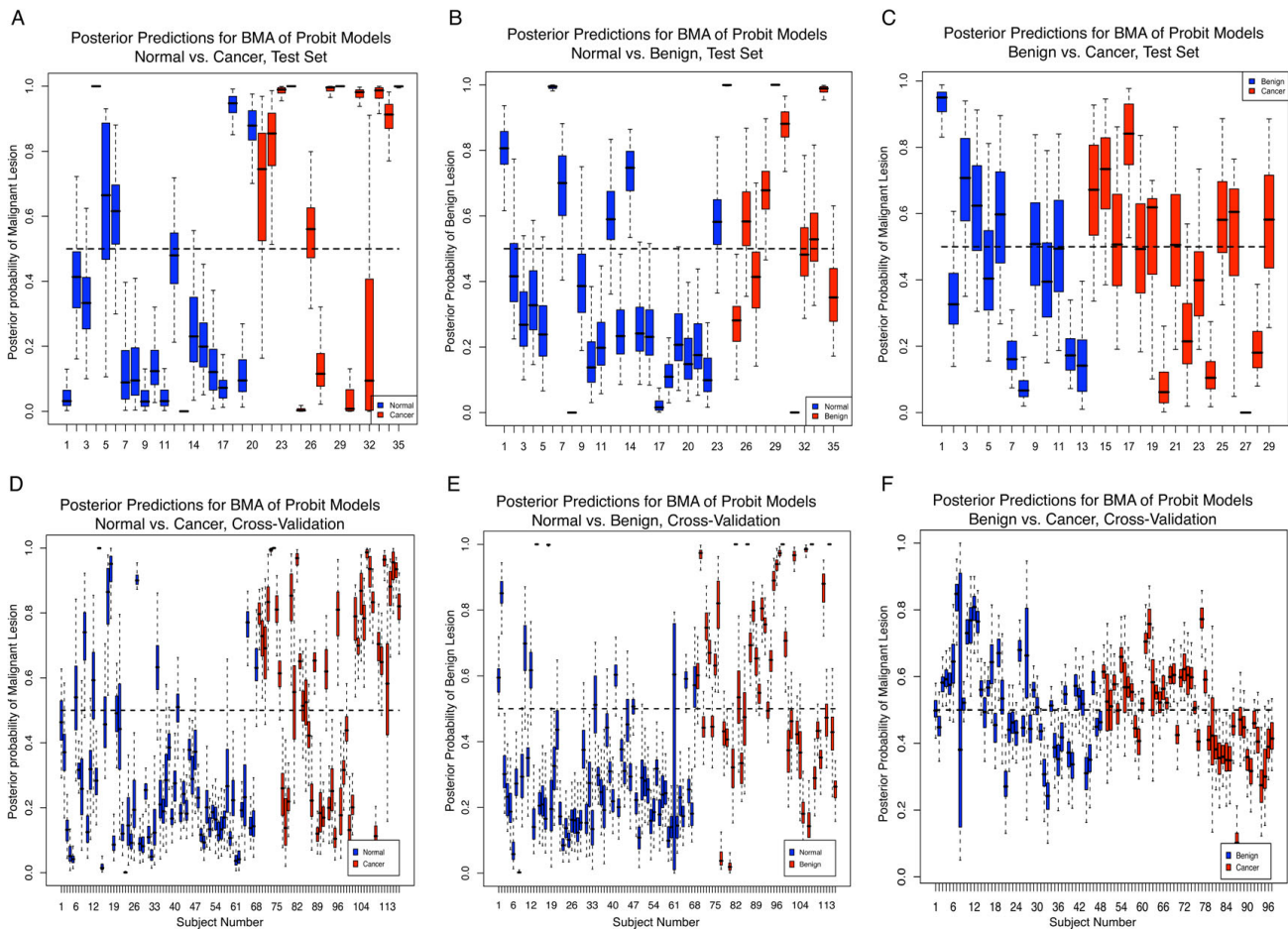


Figure 2
Posterior predictions of Bayesian model averaging (BMA) of probit models, run with a 70% train and 30% test split of the data set (A-C) and also with leave-one-out cross-validation (LOOCV) (D-F). The classifiers achieved moderate classification results for normal vs. malignant or benign lesions (A, B, D, E) and poor classification results for malignant vs. benign lesions (C, F).

less "decided" (concentrated closer to the 0.5 probability line) for benign vs. cancer. This trend indicated that the serum protein levels were very similar for malignant and benign lesions.

Selected serum proteins

The iterated BMA algorithm chose the best-fitting probit models. The chosen models and their proteins are shown in Figure 3. The best proteins for each classification task are listed in Table 4. The top protein for both normal vs. cancer and normal vs. benign was macrophage migration inhibitory factor (MIF), a known inflammatory agent [64-66]. Other selected proteins also play roles in inflammation and immune response, such as MMP-9 [67,68], MPO [69], sVCAM-1 [70], ACTH [71], MICA [72], IL-5 [73], IL-12 p40 [74-76], MCP-1 [77], and IFN α [78-80]. For benign vs. cancer, the top protein was CA-125, which is

used as a biomarker for ovarian cancer [12,81-83]. However, the greater presence of CA-125 in cancer tissue was still too subtle to allow the classifiers to achieve good classification performance.

Complementary to the models' matrix plots for feature strength are the coefficients' marginal posterior probability distribution functions (PDFs), which the BMA technique calculates by including information from all considered models. Figure 4 shows the marginal posterior PDFs for the top coefficients for the BMA models. The coefficients' distributions are mixture models of a normal distribution and a point mass at zero. This point mass is much larger for the benign vs. cancer models than for normal vs. cancer and normal vs. benign models. The higher weight at zero indicates that the proteins are less suitable

Table 4: Proteins chosen by BMA of linear models

Protein	Biological Role	Higher Prevalence In
<i>Normal vs. Cancer</i>		
MIF (macrophage migration inhibitory factor)	Inflammation, regulates macrophage function in host defense through the suppression of anti-inflammatory effects of glucocorticoids [64-66]	Cancer
MMP-9 (matrix metalloproteinase)	Breakdown and remodeling of extracellular matrix [67], regulates growth plate angiogenesis and apoptosis of hypertrophic chondrocytes [68]	Cancer
MPO (myeloperoxidase)	Inflammation, produces HOCl, modulates vascular signaling and vasodilatory functions of nitric oxide (NO) during acute inflammation [69]	Normal
sVCAM-1 (soluble vascular cell adhesion molecule 1)	Mediates leukocyte-endothelial cell adhesion and signal transduction, membrane-bound adhesion molecules and the process of vascular inflammation of the vessel wall [70]	Normal
ACTH (adrenocorticotrophic hormone)	Stimulates secretion of adrenal corticosteroids [71]	Cancer
<i>Normal vs. Benign</i>		
MIF (macrophage migration inhibitory factor)	Inflammation, regulates macrophage function in host defense through the suppression of anti-inflammatory effects of glucocorticoids [64-66]	Benign
MICA (MHC class I polypeptide-related sequence A)	Stress-induced antigen that is broadly recognized by intestinal epithelial gamma delta T cells, ligands for natural killer cells [72]	Benign
IL-5 (Interleukin 5)	Stimulates B cell growth, increases immunoglobulin secretion, mediates eosinophil activation [73]	Normal
IL-12 p40 (Interleukin 12, p40 chain)	Differentiation of naive T cells into Th1 cells, stimulates the growth and function of T cells, stimulates the production of interferon-gamma (IFN- γ) and [74,75](TNF- α) from T and natural killer cells, and reduces IL-4 mediated suppression of IFN- γ [76]	Normal
MCP-1 (Monocyte chemotactic protein-1)	Induces recruitment of monocytes, T lymphocytes, eosinophils, and basophils and is responsible for many inflammatory reactions to disease [77]	Benign
<i>Benign vs. Cancer</i>		
CA-125 (cancer antigen 125)	Marker for ovarian cancer [12,81,82]	Cancer
IFN α (Interferon type I)	Secreted by leukocytes, fibroblasts, or lymphoblasts in response to viruses [78] or interferon inducers; implicated in autoimmune diseases [79,80]	Benign
MICA (MHC class I polypeptide-related sequence A)	Stress-induced antigen that is broadly recognized by intestinal epithelial gamma delta T cells, ligands for natural killer cells [72]	Benign

for distinguishing benign from malignant lesions than they are for distinguishing lesions from normal tissue.

Feature selection and Bayesian model averaging

While doing feature selection for normal vs. cancer within LOOCV, we recorded the counts of how many times each

protein was selected. The selection frequencies are shown as a heatmap in Figure 5. Iterated BMA and least-angle regression selected the fewest proteins, whereas stepwise feature selection chose many more proteins. The strongest proteins were chosen consistently across all feature selection techniques, as shown by the horizontal lines in the figure.

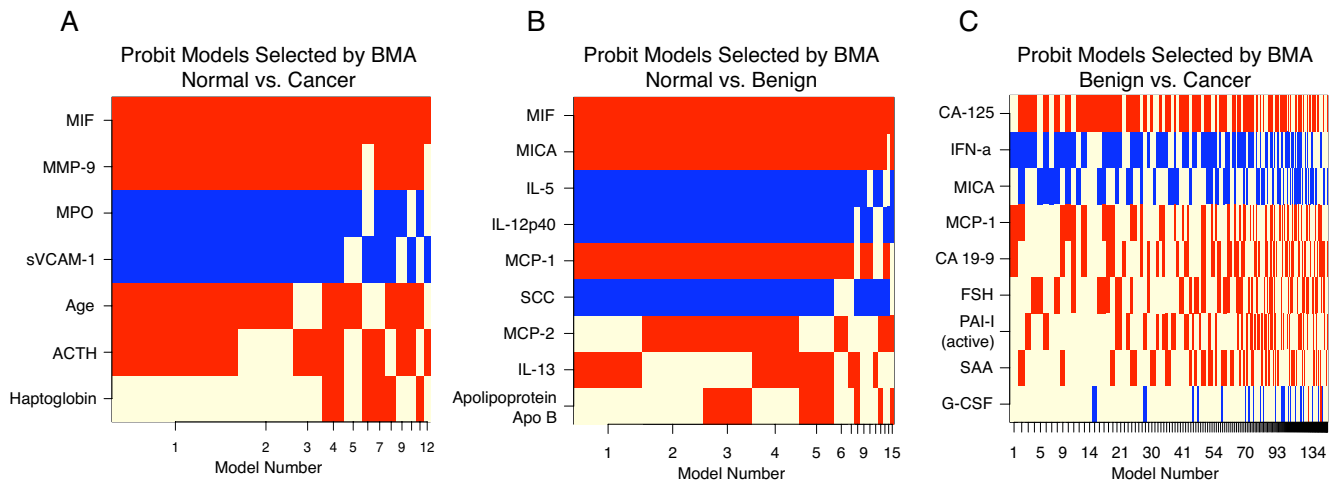


Figure 3
Models selected by BMA of linear models. Features are plotted in decreasing posterior probability of being nonzero. Models are ordered by selection frequency, with the best, most frequently selected models on the left and the weakest, rarest chosen on the right. Coefficients with positive values are shown in red and negative values in blue. Strong, frequently selected features appear as solid horizontal stripes. A beige value indicates that the protein was not selected in a particular model.

We also investigated the effect of feature selection upon classifier generalization. Figure 6 shows the ROC and accuracy curves for linear models with various feature selection strategies. Preselecting the features generated a very optimistically biased classification performance. When the same feature selection technique (stepwise) was applied within each fold of LOOCV, the classification performance fell dramatically. Using no feature selection (using all proteins) had extremely poor classification performance – no better than guessing. The poor performance demonstrated that the linear model needs feature selection for good classification performance when the

number of features is roughly the same as the number of samples in noisy data. Iterated BMA of linear models significantly outperformed the stepwise, single-model method. This performance increase demonstrated the better predictive ability of the BMA models; averaging a set of the most promising models was better than using only the single best model.

Discussion

The serum proteins allowed the classifiers to detect lesions (malignant or benign) from normal tissue moderately well, but they were very poor for distinguishing benign

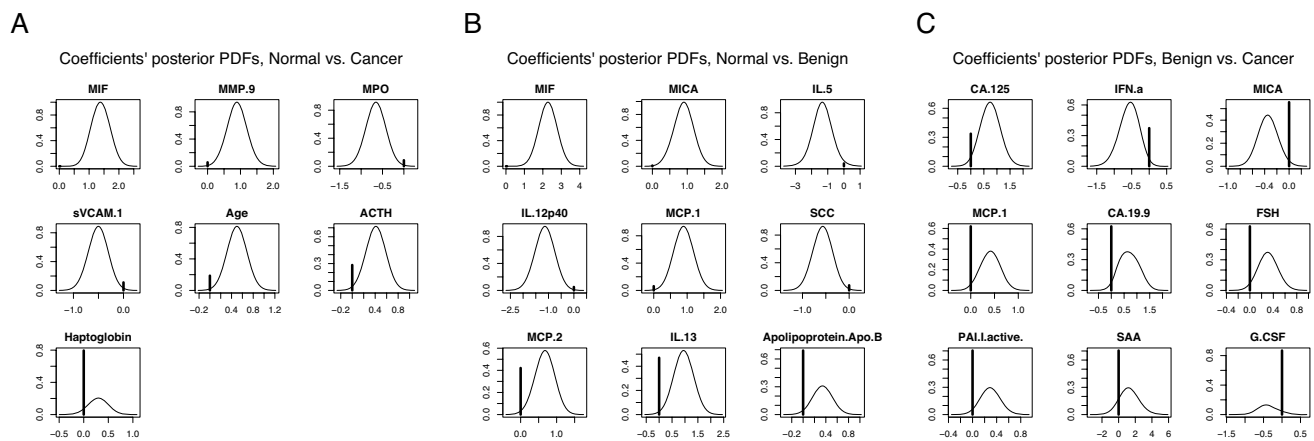


Figure 4
Posterior distributions of the model coefficients for the proteins. The distributions are mixtures of a point mass at zero and a normal distribution. The height of the solid line at zero represents the posterior probability that the coefficient is zero. The nonzero part of the distribution is scaled so that the maximum height is equal to the probability that the coefficient is nonzero.

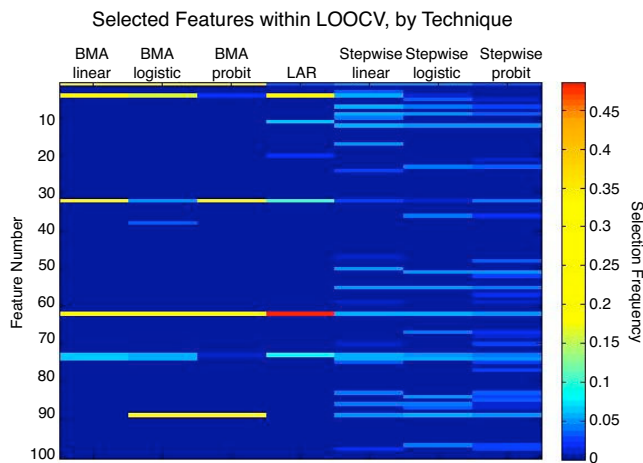


Figure 5
Heatmap of normalized frequencies of selected features, normal vs. cancer. The feature selection frequencies were averaged over all folds of the LOOCV. For comparison across techniques, the frequencies in each column were scaled to sum to one. Less-frequently selected features appear as cooler dark blue colors, whereas more frequently selected features appear as hotter, brighter colors. Models that used fewer features appear as dark columns with a few bright bands, whereas models that used more features appear as denser smears of darker bands.

from malignant lesions. These classification results were consistent in both the test set and the leave-one-out cross-validation. This consistency implies that the classification

results are not highly dependent on the sampling scheme but rather highlight consistent trends in our data set. The classification results show that the proteins were not specific for cancer and suggest that they may indicate states of biological or immunological stress. A good candidate for the dominating biological effect is inflammation, since the top protein selected for both normal vs. cancer and normal vs. benign was macrophage migration inhibitory factor (MIF), which is active in inflammation [64-66]. The best protein for distinguishing malignant from benign tumors was cancer antigen 125 (CA-125), which is a prominent serum marker for ovarian cancer [12,81,83]. However, CA-125 levels are influenced by other factors, such as age, hormone-replacement therapy, smoking habits, and use of oral contraceptives [84]. In general it is very difficult to ascertain whether biomarker proteins are generated by the cancer itself or by secondary effects, such as immune response. Once potential biomarker proteins are identified in initial studies, follow-up studies can focus on those proteins to discover their origin and role in the disease process. Helpful experimental study designs would control for known secondary causes of biomarker activity and would collect enough samples to average over unintended secondary causes. Longitudinal studies would also lessen the effect of transient secondary causes.

To quantify and compare classification performances, we used ROC analysis, which fairly compares classifiers that may be operating at different sensitivities due to arbitrary decision thresholds applied to the classifiers' output val-

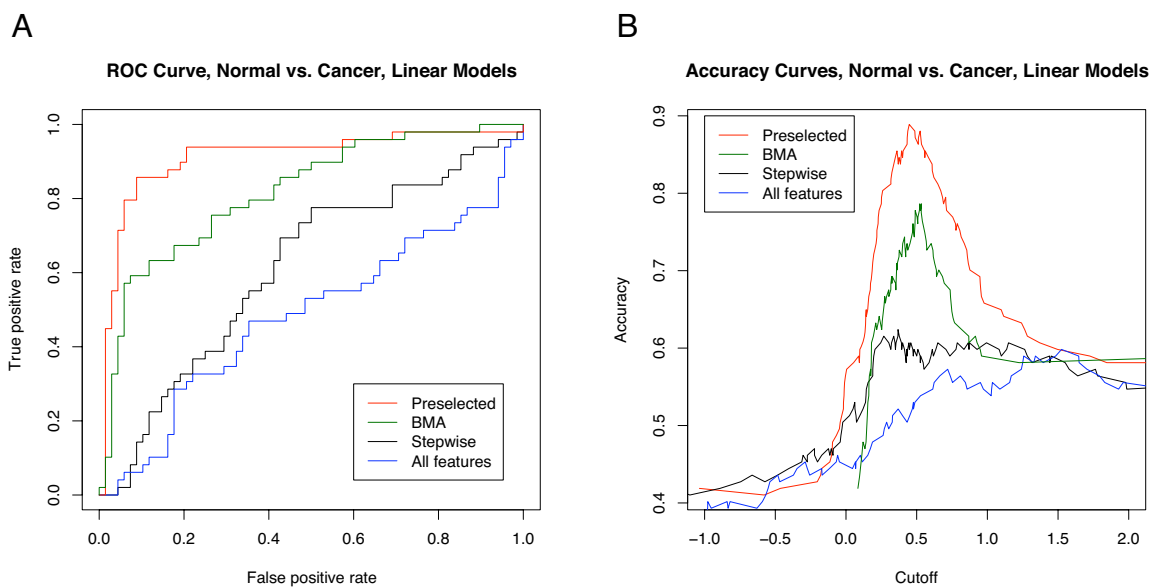


Figure 6
ROC and accuracy curves for linear models with four feature selection techniques. 1) Preselected: the features (using all the data to choose the best features, and then running the model using only those preselected features in LOOCV), 2) BMA: iterated Bayesian model averaging, 3) Stepwise feature selection, and 4) All features: using all the proteins in the model, no feature selection.

ues. Although our data set comprised three classes (normal, benign, and cancer), current ROC methods required us to split an inherently three-class classification problem into three different two-class tasks: normal vs. benign, normal vs. cancer, and benign vs. cancer. The field of ROC analysis is still in development for the three-class problem; no consensus has yet been reached about how to quantitatively score the resulting six-dimensional ROC hypersurface [85-88]. However, for other methods of classifier comparison, such as the generalized Brier score or discrete counts of classification errors, full three-class models could have been used, albeit with decision thresholds.

This study investigated a group of 98 serum proteins (Table 2), which is relatively small sample of all detectable serum proteins. Future studies may identify other proteins with stronger relationships to breast cancer. Rather than relying on only a few proteins, upcoming protein technologies will allow the screening of large populations with protein-based tests that require a larger set of proteins. Microfluidics chips would simplify the process of automating blood tests in a high-throughput fashion. However, with current assay technology and cost-benefit analysis of screening programs, the fixed cost per protein assayed essentially limits the number of proteins that can be used for screening. To lower screening costs, we chose small subsets of the features via feature-selection methods. Iterated BMA and least-angle regression were able to classify well using a far smaller set of features than those chosen by stepwise feature selection.

High feature correlation impedes many feature-selection techniques. For stochastic feature-selection methods, two highly correlated features are each likely to be chosen in alternation. Similarly, a cluster of highly correlated features causes the feature selection technique to spread the feature selection rate among each feature in the cluster, essentially diluting each feature's selection rate. Severe dilution of selection rates can cause none of the cluster's features to be selected. Future work will entail adding cluster-based methods to the iterated BMA algorithm.

The currently proposed serum biomarkers for breast cancer are not sensitive or specific enough for breast cancer screening. However, better biomarkers may be identified with newer protein assay technology and larger data sets. A protein's subtle diagnostic ability may be enhanced by the assimilation of other medical information, such as gene expression and medical imaging. The proteins will boost diagnostic performance only if they provide complementary and non-redundant information with the clinical practice of mammograms, sonograms, and physical examination. The relationship of medical imaging and protein screening remains to be determined in future work.

Conclusion

We have performed feature-selection and classification techniques to identify blood serum proteins that are indicative of breast cancer in premenopausal women. The best features to detect breast cancer were MIF, MMP-9, and MPO. While the proteins could distinguish normal tissue from cancer and normal tissue from benign lesions, they could not distinguish benign from malignant lesions. Since the same protein (MIF) was chosen for both normal vs. cancer and normal vs. benign lesions, it is likely that this protein plays a role in the inflammatory response to a lesion, whether benign or malignant, rather than in a role specific for cancer. While the current set of proteins show moderate ability for detecting breast cancer, their true usefulness in a screening program remains to be seen in their integration with imaging-based screening practices.

List of abbreviations used

This article used the following abbreviations: AUC: for Area under the ROC curve; BMA: for Bayesian model averaging; CV: for coefficient of variation; ELISA: for enzyme-linked immunosorbent assay; HIPAA: for Health Insurance Portability and Accountability Act; LOOCV: for leave-one-out cross-validation; MICA: for human major histocompatibility complex class I chain-related A; MIF: for macrophage migration inhibitory factor; MMP-9: for matrix metalloproteinase 9; PDF: for probability density function; and ROC: for receiver operating characteristic.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JRM coordinated the sample collection and study design. AEL and ZY conducted the ELISA assays. JLJ performed the computations and statistical analysis. SM and MC guided the use of Bayesian Model Averaging in the study. JLJ and JYL contributed to the drafting of the manuscript, which was approved by all authors.

Additional material

Additional File 1

Subject diagnoses and protein levels. These data represent this study's selected serum proteins and their antibodies.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-9-164-S1.xls>]

Additional File 2

Proteins and antibodies. These data show sample information with the serum protein levels measured by ELISA.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-9-164-S2.xls>]

Additional File 3

Classification code in R. This file is a compressed folder containing this study's source code for the statistical program R.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-9-164-S3.zip>]

Additional File 4

Serum protein data. This file is a comma-delimited file of blood serum ELISA data intended for computational analysis in the R scripts of Additional file 1.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-9-164-S4.csv>]

Acknowledgements

The work was supported in part by the NIH (NIH CA 84955 and R01 CA-112437-01) and the U.S. Army Breast Cancer Research Program (Grant No. W81XWH-05-1-0292).

References

- Ferrini R, et al: **Screening mammography for breast cancer: American College of Preventive Medicine practice policy statement.** *Am J Prev Med* 1996, **12(5)**:340-1.
- Meyer JE, et al: **Occult breast abnormalities: percutaneous preoperative needle localization.** *Radiology* 1984, **150**:335-337.
- Rosenberg AL, et al: **Clinically occult breast lesions: localization and significance.** *Radiology* 1987, **162**:167-170.
- Yankaskas BC, et al: **Needle localization biopsy of occult lesions of the breast.** *Radiology* 1988, **23**:729-733.
- Kreunin P, et al: **Proteomic profiling identifies breast tumor metastasis-associated factors in an isogenic model.** *Proteomics* 2007, **7(2)**:299-312.
- Isaacs C, Stearns V, Hayes DF: **New prognostic factors for breast cancer recurrence.** *Semin Oncol* 2001, **28(1)**:53-67.
- Duffy MJ: **Urokinase plasminogen activator and its inhibitor, PAI-1, as prognostic markers in breast cancer: from pilot to level I evidence studies.** *Clin Chem* 2002, **48(8)**:1194-7.
- Malkas LH, et al: **A cancer-associated PCNA expressed in breast cancer has implications as a potential biomarker.** *Proc Natl Acad Sci USA* 2006, **103(51)**:19472-7.
- Hye A, et al: **Proteome-based plasma biomarkers for Alzheimer's disease.** *Brain* 2006, **129(Pt 11)**:3042-50.
- Wang TJ, et al: **Multiple biomarkers for the prediction of first major cardiovascular events and death.** *N Engl J Med* 2006, **355(25)**:2631-9.
- Polascik TJ, Oesterling JE, Partin AW: **Prostate specific antigen: a decade of discovery—what we have learned and where we are going.** *J Urol* 1999, **162(2)**:293-306.
- Gorelik E, et al: **Multiplexed immunobead-based cytokine profiling for early detection of ovarian cancer.** *Cancer Epidemiol Biomarkers Prev* 2005, **14(4)**:981-7.
- Duffy MJ: **Serum tumor markers in breast cancer: are they of clinical value?** *Clin Chem* 2006, **52(3)**:345-51.
- Duffy MJ, et al: **High preoperative CA 15-3 concentrations predict adverse outcome in node-negative and node-positive breast cancer: study of 600 patients with histologically confirmed breast cancer.** *Clin Chem* 2004, **50(3)**:559-63.
- Cheung KL, Graves CR, Robertson JF: **Tumour marker measurements in the diagnosis and monitoring of breast cancer.** *Cancer Treat Rev* 2000, **26(2)**:91-102.
- Li J, et al: **Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer.** *Clin Chem* 2002, **48(8)**:1296-304.
- Mathelin C, et al: **Serum biomarkers for detection of breast cancers: A prospective study.** *Breast Cancer Res Treat* 2006, **96(1)**:83-90.
- Molina R, et al: **Tumor markers in breast cancer- European Group on Tumor Markers recommendations.** *Tumour Biol* 2005, **26(6)**:281-93.
- Lumachi F, et al: **Relationship between tumor markers CEA and CA 15-3, TNM staging, estrogen receptor rate and MIB-1 index in patients with pT1-2 breast cancer.** *Anticancer Res* 2004, **24(5B)**:3221-4.
- Skates SJ, et al: **Pooling of case specimens to create standard serum sets for screening cancer biomarkers.** *Cancer Epidemiol Biomarkers Prev* 2007, **16(2)**:334-41.
- Kiang DT, Greenberg LJ, Kennedy BJ: **Tumor marker kinetics in the monitoring of breast cancer.** *Cancer* 1990, **65(2)**:193-9.
- Yasasever V, et al: **Utility of CA 15-3 and CEA in monitoring breast cancer patients with bone metastases: special emphasis on "spiking" phenomena.** *Clin Biochem* 1997, **30(1)**:53-6.
- Pentheroudakis G, et al: **The neutrophil, not the tumor: serum CA 15-3 elevation as a result of granulocyte-colony-stimulating factor-induced neutrophil MUC overexpression and neutrophilia in patients with breast carcinoma receiving adjuvant chemotherapy.** *Cancer* 2004, **101(8)**:1767-75.
- Colomer R, et al: **Circulating CA 15-3 levels in the postsurgical follow-up of breast cancer patients and in non-malignant diseases.** *Breast Cancer Res Treat* 1989, **13(2)**:123-33.
- Hashimoto T, Matsubara F: **Changes in the tumor marker concentration in female patients with hyper-, eu-, and hypothyroidism.** *Endocrinol Jpn* 1989, **36(6)**:873-9.
- Symeonidis A, et al: **Increased serum CA-15.3 levels in patients with megaloblastic anemia due to vitamin B12 deficiency.** *Oncology* 2004, **67(5-6)**:359-67.
- Symeonidis A, et al: **Increased CA-15.3 levels in the serum of patients with homozygous beta-thalassaemia and sickle cell/ beta-thalassaemia.** *Br J Haematol* 2006, **133(6)**:692-4.
- Sahab ZJ, Semaan SM, Sang QXAS: **Methodology and Applications of Disease Biomarker Identification in Human Serum.** *Biomarker Insights* 2007, **2**:21-43.
- Zissimopoulos A, et al: **[Procollagen-I, collagen telopeptide I, CEA, CA 15-3 as compared to bone scintigraphy in patients with breast cancer].** *Hell J Nud Med* 2006, **9(1)**:60-4.
- Belluco C, et al: **Serum proteomic analysis identifies a highly sensitive and specific discriminatory pattern in stage I breast cancer.** *Ann Surg Oncol* 2007, **14(9)**:2470-6.
- Bouchal P, et al: **Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (ITRAQ-2DLC-MS/MS) based quantitative proteomic analysis.** *J Proteome Res* 2009, **8(1)**:362-73.
- Callesen AK, et al: **Combined experimental and statistical strategy for mass spectrometry based serum protein profiling for diagnosis of breast cancer: a case-control study.** *J Proteome Res* 2008, **7(4)**:1419-26.
- Laronga C, Drake RR: **Proteomic approach to breast cancer.** *Cancer Control* 2007, **14(4)**:360-8.
- Rui Z, et al: **Use of serological proteomic methods to find biomarkers associated with breast cancer.** *Proteomics* 2003, **3(4)**:433-9.
- Wulfkuhle JD, et al: **New approaches to proteomic analysis of breast cancer.** *Proteomics* 2001, **1(10)**:1205-15.
- Wulfkuhle JD, et al: **Proteomics of human breast ductal carcinoma in situ.** *Cancer Res* 2002, **62(22)**:6740-9.
- Callesen AK, et al: **Reproducibility of mass spectrometry based protein profiles for diagnosis of breast cancer across clinical studies: a systematic review.** *J Proteome Res* 2008, **7(4)**:1395-402.
- Gast MC, et al: **SELDI-TOF MS serum protein profiles in breast cancer: assessment of robustness and validity.** *Cancer Biomark* 2006, **2(6)**:235-48.
- Wulfkuhle JD, et al: **Multiplexed cell signaling analysis of human breast cancer applications for personalized therapy.** *J Proteome Res* 2008, **7(4)**:1508-17.
- Somorjai RL, Dolenko B, Baumgartner R: **Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions.** *Bioinformatics* 2003, **19(12)**:1484-91.
- Bellman RE: **Adaptive control processes: a guided tour.** Princeton, N.J.: Princeton University Press; 1961:255.
- Sahiner B, et al: **Stepwise linear discriminant analysis in computer-aided diagnosis: the effect of finite sample size.** In *Medical Imaging 1999: Image Processing* San Diego, CA, USA: SPIE; 1999.

43. Sahiner B, et al.: **Feature selection and classifier performance in computer-aided diagnosis: the effect of finite sample size.** *Medical Physics* 2000, **27(7)**:1509-22.
44. Lachenbruch PA: **Discriminant analysis.** Volume ix. New York,: Hafner Press; 1975:128.
45. Tatsuoaka MM: **Multivariate analysis; techniques foreducational and psychological research.** Volume xiii. New York,:Wiley; 1971:310.
46. Draper NR, Smith H: **Applied regression analysis.** In *Wiley series in probability and statistics. Texts and references section Volume xvii.* 3rd edition. New York: Wiley; 1998:706.
47. Hoeting JA, et al.: **Bayesian model Averaging: A Tutorial.** *Statistical Science* 1999, **14(4)**:382-417.
48. Hodges JS: **Uncertainty, Policy Analysis and Statistics.** *Statistical Science* 1987, **2(3)**:259-275.
49. Draper D: **Assessment and Propagation of Model Uncertainty.** *Journal of the Royal Statistical Society. Series B (Methodological)* 1995, **57(1)**:45-97.
50. Clyde M, George EI: **Model Uncertainty.** *Statistical Science* 2004, **19(1)**:81-94.
51. Berger J, Pericchi L: **Objective Bayesian methods for model section.** In *Model Section* Edited by: Lahiri P. IMS: Beachwood, OH; 2001:135-207.
52. Chipman H, George EI, McCulloch R: **The Practical Implementation of Bayesian Model Selection.** In *Model Selection* Edited by: Lahiri P. IMS: Beachwood, OH; 2001:65-134.
53. Raftery AE: **Bayesian Model Selection in Social Research.** *Sociological Methodology* 1995, **25**:111-163.
54. Madigan D, Raftery AE: **Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window.** *Journal of the American Statistical Association* 1994, **89(428)**:1535-1546.
55. Yeung KY, Bumgarner RE, Raftery AE: **Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data.** *Bioinformatics* 2005, **21(10)**:2394-2402.
56. Furnival GM, Wilson RW Jr: **Regressions by Leaps and Bounds.** *Technometrics* 1974, **16(4)**:499-511.
57. Dudoit S, Fridlyand J, Speed TP: **Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data.** *Journal of the American Statistical Association* 2002, **97(457)**:77-87.
58. Vapnik NV: **The Nature of Statistical Learning Theory.** In *Statistics for Engineering and Information Science* 2nd edition. New York, NY: Springer-Verlag New York, Inc; 1995:188.
59. Guyon I, et al.: **Gene Selection for Cancer Classification using Support Vector Machines.** *Machine Learning* 2002, **46(1)**:389-422.
60. Zhang X, et al.: **Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data.** *BMC Bioinformatics* 2006, **7**:197.
61. Efron B, et al.: **Least Angle Regression.** *The Annals of Statistics* 2004, **32(2)**:407-451.
62. Obuchowski NA: **Receiver operating characteristic curves and their use in radiology.** *Radiology* 2003, **229(1)**:3-8.
63. Efron B, Tibshirani RJ: **An Introduction to the Bootstrap.** In *Monographs on Statistics and Applied Probability* Edited by: Cox DR, et al. New York, NY: Chapman & Hall; 1993:436.
64. Lolis E, Bucala R: **Macrophage migration inhibitory factor.** *Expert Opin Ther Targets* 2003, **7(2)**:153-64.
65. Morand EF: **New therapeutic target in inflammatory disease: macrophage migration inhibitory factor.** *Intern Med J* 2005, **35(7)**:419-26.
66. Bucala R: **MIF rediscovered: cytokine, pituitary hormone, and glucocorticoid-induced regulator of the immune response.** *Faseb J* 1996, **10(14)**:1607-13.
67. Poitevin S, et al.: **Type I collagen induces tissue factor expression and matrix metalloproteinase 9 production in human primary monocytes through a redox-sensitive pathway.** *J Thromb Haemost* 2008, **6(9)**:1586-94.
68. Vu TH, et al.: **MMP-9/gelatinase B is a key regulator of growth plate angiogenesis and apoptosis of hypertrophic chondrocytes.** *Cell* 1998, **93(3)**:411-22.
69. Eiserich JP, et al.: **Myeloperoxidase, a leukocyte-derived vascular NO oxidase.** *Science* 2002, **296(5577)**:2391-4.
70. Gearing AJ, Newman W: **Circulating adhesion molecules in disease.** *Immunol Today* 1993, **14(10)**:506-12.
71. Gwynne JT, et al.: **Adrenal cholesterol uptake from plasma lipoproteins: regulation by corticotropin.** *Proc Natl Acad Sci USA* 1976, **73(12)**:4329-33.
72. Bauer S, et al.: **Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA.** *Science* 1999, **285(5428)**:727-9.
73. Clutterbuck EJ, Hirst EM, Sanderson CJ: **Human interleukin-5 (IL-5) regulates the production of eosinophils in human bone marrow cultures: comparison and interaction with IL-1, IL-3, IL-6, and GM-CSF.** *Blood* 1989, **73(6)**:1504-12.
74. Jana M, et al.: **Induction of tumor necrosis factor-alpha (TNF-alpha) by interleukin-12 p40 monomer and homodimer in microglia and macrophages.** *J Neurochem* 2003, **86(2)**:519-28.
75. Mattner F, et al.: **Treatment with homodimeric interleukin-12 (IL-12) p40 protects mice from IL-12-dependent shock but not from tumor necrosis factor alpha-dependent shock.** *Infect Immun* 1997, **65(11)**:4734-7.
76. Ma X, et al.: **The interleukin 12 p40 gene promoter is primed by interferon gamma in monocytic cells.** *J Exp Med* 1996, **183(1)**:147-57.
77. Taub DD, et al.: **Monocyte chemotactic protein-1 (MCP-1), -2, and -3 are chemotactic for human T lymphocytes.** *J Clin Invest* 1995, **95(3)**:1370-6.
78. Braganca J, Civas A: **Type I interferon gene expression: differential expression of IFN-A genes induced by viruses and double-stranded RNA.** *Biochimie* 1998, **80(8-9)**:673-87.
79. Prummel MF, Laurberg P: **Interferon-alpha and autoimmune thyroid disease.** *Thyroid* 2003, **13(6)**:547-51.
80. Imagawa A, et al.: **Autoimmune endocrine disease induced by recombinant interferon-alpha therapy for chronic active type C hepatitis.** *J Clin Endocrinol Metab* 1995, **80(3)**:922-6.
81. Nossou 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* 2008, **199(3)**:215-23.
82. Mano A, et al.: **CA-125 AUC as a predictor for epithelial ovarian cancer relapse.** *Cancer Biomark* 2008, **4(2)**:73-81.
83. Mano A, et al.: **CA-125 AUC as a new prognostic factor for patients with ovarian cancer.** *Gynecol Oncol* 2005, **97(2)**:529-34.
84. Dehaghani AS, et al.: **Factors influencing serum concentration of CA125 and CA15-3 in Iranian healthy postmenopausal women.** *Pathol Oncol Res* 2007, **13(4)**:360-4.
85. Dreiseitl S, Ohno-Machado L, Binder M: **Comparing three-class diagnostic tests by three-way ROC analysis.** *Med Decis Making* 2000, **20(3)**:323-31.
86. Xin H, et al.: **Three-class ROC analysis-a decision theoretic approach under the ideal observer framework.** *Medical Imaging, IEEE Transactions on* 2006, **25(5)**:571-581.
87. Edwards DC, et al.: **Estimating three-class ideal observer decision variables for computerized detection and classification of mammographic mass lesions.** *Med Phys* 2004, **31(1)**:81-90.
88. Chan HP, et al.: **Design of three-class classifiers in computer-aided diagnosis: Monte Carlo simulation study.** In *Medical Imaging 2003: Image Processing* San Diego, CA, USA: SPIE; 2003.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2407/9/164/prepub>