

## Detection of Elevated Plasma Levels of Epidermal Growth Factor Receptor Before Breast Cancer Diagnosis among Hormone Therapy Users

Sharon J. Pitteri, Lynn M. Amon, Tina Busald Buson, Yuzheng Zhang, Melissa M. Johnson, Alice Chin, Jacob Kennedy, Chee-Hong Wong, Qing Zhang, Hong Wang, Paul D. Lampe, Ross L. Prentice, Martin W. McIntosh, Samir M. Hanash, and Christopher I. Li

### Abstract

Applying advanced proteomic technologies to prospectively collected specimens from large studies is one means of identifying preclinical changes in plasma proteins that are potentially relevant to the early detection of diseases such as breast cancer. We conducted 14 independent quantitative proteomics experiments comparing pooled plasma samples collected from 420 estrogen receptor-positive (ER<sup>+</sup>) breast cancer patients  $\leq 17$  months before their diagnosis and matched controls. Based on the more than 3.4 million tandem mass spectra collected in the discovery set, 503 proteins were quantified, of which 57 differentiated cases from controls with a  $P$  value of  $< 0.1$ . Seven of these proteins, for which quantitative ELISA assays were available, were assessed in an independent validation set. Of these candidates, epidermal growth factor receptor (EGFR) was validated as a predictor of breast cancer risk in an independent set of preclinical plasma samples for women overall [odds ratio (OR), 1.44;  $P = 0.0008$ ] and particularly for current users of estrogen plus progestin (E + P) menopausal hormone therapy (OR, 2.49;  $P = 0.0001$ ). Among current E + P users, the EGFR sensitivity for breast cancer risk was 31% with 90% specificity. Whereas the sensitivity and specificity of EGFR are insufficient for a clinically useful early detection biomarker, this study suggests that proteins that are elevated preclinically in women who go on to develop breast cancer can be discovered and validated using current proteomic technologies. Further studies are warranted to examine the role of EGFR and to discover and validate other proteins that could potentially be used for early detection of breast cancer. *Cancer Res*; 70(21); 8598–606. ©2010 AACR.

### Introduction

Breast cancer is a disease of considerable public health importance given that it is the most commonly diagnosed cancer in women worldwide. At present, the best available tool for early detection of breast cancer is mammography, as it has been shown to reduce mortality in randomized trials (1). Despite improvements in technology and the widespread use of mammography, breast cancer still remains the second leading cause of cancer mortality in women in the United States (2) and the leading cause of cancer mortality in women worldwide (3).

Considerable effort has been invested in trying to interrogate changes related to breast cancer in biological specimens. Numerous challenges have been faced including use

of discovery platforms with poor reproducibility and interrogation of specimens collected either at or after a breast cancer diagnosis (when the specimens of greatest clinical interest are those ascertained before diagnosis). We used advanced technologies for proteome profiling to systematically examine changes in the plasma proteome of preclinical samples from a case-control study nested in the Women's Health Initiative Observational Study (WHI OS) prospective cohort. The purpose of this study was to identify proteins that may be elevated preclinically in estrogen receptor-positive (ER<sup>+</sup>) breast cancer cases and to validate promising candidates in an independent sample set.

### Materials and Methods

We conducted a nested case-control study within the WHI OS prospective cohort that was approved by the Fred Hutchinson Cancer Research Center Institutional Review Board. The WHI OS is a prospective cohort study, and details on its scientific rationale, eligibility criteria, and design have been published previously (4, 5). Briefly, 93,676 postmenopausal women 50 to 79 years of age were enrolled between October 1, 1993 and December 31, 1998 through 40 clinical centers in the United States. Epidemiologic data and biological specimens were collected from participants

**Authors' Affiliation:** Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

**Corresponding Author:** Christopher I. Li, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M4-C308, P.O. Box 19024, Seattle, WA 98109-1024. Phone: 206-667-7444; Fax: 206-667-5948; E-mail: [cili@fhcrc.org](mailto:cili@fhcrc.org).

doi: 10.1158/0008-5472.CAN-10-1676

©2010 American Association for Cancer Research.

uniformly according to standardized Institutional Review Board–approved procedures and protocols by trained study staff. Blood specimens were collected at two time points, at enrollment (baseline) and at year 3 of follow-up. All participants provided written informed consent for this research study at the time of enrollment.

### Case and control identification

Cohort members completed self-administered demographic and risk factor questionnaires annually, and the medical records of all women reporting a breast cancer diagnosis were reviewed by a study adjudicator to verify the diagnosis. The medical records of verified cases were then forwarded to the WHI coordinating center for coding of breast cancer stage, ER status, progesterone receptor (PR) status, and histology. A total of 943 women clinically diagnosed with invasive breast cancer within 17 months of either their baseline or year 3 blood draw without a prior history of breast cancer were identified. Given the heterogeneity of breast cancer and the likelihood that biomarkers for early detection of breast cancer may be specific to certain breast cancer types, the work described herein focused exclusively on ER<sup>+</sup> breast cancer cases.

Potential controls were selected from a pool of women never diagnosed with any type of cancer through September 15, 2005. Controls were individually matched 1:1 to cases on age at enrollment ( $\pm 1$  year), race/ethnicity (white, black, Hispanic, Asian/Pacific Islander, or other), blood draw date ( $\pm 1$  year), and clinical center of enrollment. Matching was done in a time-forward manner to ensure that each control had at least as much follow-up time following her blood draw as the time from blood draw to breast cancer diagnosis of the case to which she was matched.

### Discovery studies

Fourteen independent large-scale quantitative proteomics experiments were performed using a previously described method (6) that used immunodepletion, isotopic labeling with acrylamide (7), extensive fractionation (8), and high-resolution tandem mass spectrometry (MS/MS). Each experiment compared a pool of 300  $\mu$ L of plasma from 35 prediagnostic breast cancer cases (equal volume) with a pool of 300  $\mu$ L from 35 matched controls (equal volume). Pools of case and control plasma were immunodepleted of the top six most abundant proteins (albumin, IgG, IgA, transferrin, haptoglobin, and antitrypsin) using a Hu-6 column (4.6  $\times$  250 mm, Agilent). Columns were equilibrated with buffer A at 0.5 mL/min for 13 minutes, and pooled plasma was injected after filtration through a 0.22- $\mu$ m syringe filter. Flow-through fractions were collected for 10 minutes at a flow rate of 0.5 mL/min with buffer A and stored at  $-80^{\circ}\text{C}$  until further use. Immunodepleted samples were concentrated using a Centricon YM-3 device (Millipore) and rediluted in 8 mol/L urea, 30 mmol/L Tris (pH 8.5), 0.5% octyl- $\beta$ -D-glucopyranoside (Roche Diagnostics). Samples were reduced with DTT, and isotopic labeling of cysteine residues of intact proteins was done. Pools received either the light acrylamide isotope (C12; Sigma-

Aldrich) or the heavy 1,2,3-<sup>13</sup>C-acrylamide isotope (C13; Cambridge Isotope Laboratories). Alkylation with acrylamide was performed for 1 hour at room temperature. The pools were then mixed together for further analysis.

The combined isotopically labeled samples were separated by an automated online two-dimensional HPLC system (Shimadzu). The combined labeled plasma samples were separated in the first dimension by anion exchange chromatography (Poros HQ/10, 10 mm i.d.  $\times$  100 mm l; Applied Biosystems) using an eight-step elution (0–1,000 mmol/L NaCl) at 0.8 mL/min. Fractions from each of the eight anion-exchange separation steps were automatically transferred onto a reversed-phase column (PorosR2/10, 4.6 mm i.d.  $\times$  100 mm l; Applied Biosystems) for a second dimension of separation. A 25-minute gradient elution (5–95% mobile phase B) was used at 2.4 mL/min. Mobile phase A for anion-exchange chromatography consisted of 20 mmol/L Tris (Sigma-Aldrich), 6% isopropanol (Fisher Scientific), and 4 mol/L urea (pH 8.5); mobile phase B was the same composition and pH as mobile phase A with 1 mol/L NaCl (Fisher Scientific) added. Mobile phase A for reversed-phase chromatography consisted of 95% water, 5% acetonitrile, and 0.1% trifluoroacetyl or trifluoroacetic acid (TFA; Supelco); mobile phase B consisted of 90% acetonitrile, 10% water, and 0.1% TFA.

In-solution digestion was performed with lyophilized aliquots from the reversed-phase (second dimension) fractionation step. Proteins were resuspended in 0.25 mol/L urea containing 50 mmol/L ammonium bicarbonate and 4% acetonitrile and then digested overnight at 37°C with modified trypsin (Promega). The digestion was stopped by addition of 10% formic acid. Aliquots were subjected to MS shotgun analysis. Ninety-six fractions were analyzed for each experiment using a LTQ-Orbitrap (Thermo) mass spectrometer coupled with a NanoLC-1D (Eksigent). Liquid chromatography separation was performed on a 25-cm column [Pico frit 75  $\mu$ m i.d. (New Objectives), packed in-house with Magic C18 resin] with a 90-minute linear gradient from 5% to 40% acetonitrile in 0.1% formic acid at 300 nL/min for shotgun analysis. Spectra were acquired in a data-dependent mode over the  $m/z$  range 400 to 1,800 and included selection of the five most abundant doubly or triply charged ions of each MS spectrum for MS/MS analysis. Mass spectrometer parameters included capillary voltage of 2.0 kV, capillary temperature of 200°C, resolution of 60,000, and target value of 1,000,000.

Each experiment compared a set of pooled plasma samples collected from 35 breast cancer cases to a set of pooled plasma samples from their 35 matched controls; thus, a total of 490 cases and 490 controls contributed to these discovery experiments. The characteristics of our 14 experiments are summarized in Table 1. Eight experiments were performed comparing plasma from ER<sup>+</sup>/PR<sup>+</sup> ductal breast cancer patients to controls, with four experiments focusing on cases whose blood was drawn 0 to 38 weeks before their breast cancer diagnosis and four experiments on cases whose blood was drawn between 38 weeks and 17 months before diagnosis. Two experiments compared ER<sup>+</sup>/PR<sup>-</sup> ductal breast cancer

patients to controls, one containing specimens collected 0 to 38 weeks before diagnosis and the other containing cases collected 38 weeks to 17 months before diagnosis. Two experiments compared ER<sup>+</sup>/PR<sup>+</sup> lobular breast cancer patients to controls, one containing cases collected 0 to 38 weeks before diagnosis and the other containing cases collected 38 weeks to 17 months before diagnosis. Lastly, two experiments compared ER<sup>-</sup>/PR<sup>-</sup> ductal breast cancer patients to controls, with one experiment comparing cases whose blood was drawn 0 to 38 weeks before their breast cancer diagnosis and one experiment comparing cases whose blood was drawn between 38 weeks and 17 months before diagnosis. Seven of the experiments were run with the C13 (heavy) acrylamide on the cases and with C12 (light) on the controls, and the other seven experiments had the labels reversed to account for potential artifacts of this labeling.

This pooling strategy and sample size was selected because it is effectively equivalent to sample sizes of 255 individual cases and 255 individual controls for our overall analysis and 145 individual cases and 145 individual controls for the analyses restricted to ER<sup>+</sup> cases. With these sample sizes, we had a >80% specificity to discover markers with an area under the curve (AUC) of 0.6 and a sensitivity of only 20%.

#### Analysis of discovery data

Mass spectra were searched using Mascot against the human International Protein Index (IPI) database (v. 3.13). Quantitative information was extracted from acrylamide labeled peptides using an in-house script (Q3) for peptides with minimum PeptideProphet = 0.75, expect score <0.10, and maximum fractional delta mass = 20 ppm (7). Also, because the label orientation was reversed in some experiments, we additionally required that any quantified peptide had to be observed in both isotopic states at least once in at least one experiment. This criterion removes spurious ratios that arise due to misidentification. Proteins with ProteinProphet scores >0.90 were aligned by their protein group number to identify master groups of indistinguishable proteins. The master group ratio for each experiment was set to the geometric mean of the corresponding peptide ratios, logarithmically transformed, and median centered. Proteins that

had become defunct in IPI or were immunodepleted were removed from the analysis. Mean log<sub>2</sub> ratios and moderated *P* values for groups of experiments were computed using the LIMMA package from bioconductor.org and weighted using the number of quantitated events (peptides; ref. 9).

The validation work described here focuses on ER<sup>+</sup> breast cancer, and thus we selected candidate proteins from the following subsets of experiments: all cases (*n* = 14 experiments), ER<sup>+</sup> (*n* = 12 experiments), ER<sup>+</sup>/PR<sup>+</sup> (*n* = 10 experiments), and ER<sup>+</sup>, with blood samples collected 0 to 38 weeks before diagnosis (*n* = 6 experiments). Candidate proteins that were associated with a fold change ≥1.15 and a *P* value of <0.10 and were quantified in at least two experiments in any one of the subsets were chosen for validation. Of the 503 total proteins quantified in our discovery experiments, 57 met these criteria and are listed in Supplementary Table S1. More stringent statistical criteria were not applied to our discovery data to make our candidate lists more inclusive, given our plan to use a rigorous validation to identify the false positives while not discarding potential true positives that could be missed by use of more stringent criteria.

#### Validation assays

Of the 57 top ranked candidates, we selected 7 for which an ELISA assay was readily available for validation. Assays were performed according to directions provided by the various manufacturers. ELISA reagents for EGFR, IGFBP1, and NOV were obtained from R&D Systems. ELISA kits for FN1, LTF, and VWF were purchased from Bendermed Systems, Calbiochem, and Assaypro, respectively. TFF3 assays were performed using commercial reagents based on a previously described procedure (10). All samples were assayed in duplicate blinded to case/control status. We conducted two rounds of ELISA validation assays. The first round included preclinical plasma samples from 105 women who were later diagnosed with invasive ER<sup>+</sup> breast cancer and their matched controls not used in our discovery experiments and involved assessment of all seven candidates. The second round included preclinical plasma samples from 93 women who were later diagnosed with invasive ER<sup>+</sup> breast cancer and their matched controls not

**Table 1.** Composition of case and control pools interrogated in our discovery experiments

ER/PR status	Histology	Timing of specimen collection with respect to breast cancer diagnosis	Median (min, max) no. of weeks specimens were collected before breast cancer diagnosis	No. of case/control sets per pool	No. of pools	Total no. of cases/controls
ER <sup>+</sup> /PR <sup>+</sup>	Ductal	0–38 wk before diagnosis	21.4 (0.1, 37.4)	35	4	140/140
	Ductal	38 wk–17 mo before diagnosis	55.5 (38.6, 73.4)	35	4	140/140
ER <sup>+</sup> /PR <sup>+</sup>	Lobular	0–38 wk before diagnosis	22.7 (2.7, 37.0)	35	1	35/35
	Lobular	38 wk–17 mo before diagnosis	58.3 (38.9, 73.7)	35	1	35/35
ER <sup>+</sup> /PR <sup>-</sup>	Ductal	0–38 wk before diagnosis	19.7 (2.0, 38.0)	35	1	35/35
	Ductal	38 wk–17 mo before diagnosis	56.0 (38.3, 72.7)	35	1	35/35
ER <sup>-</sup> /PR <sup>-</sup>	Ductal	0–38 wk before diagnosis	24.1 (0.9, 36.7)	35	1	35/35
	Ductal	38 wk–17 mo before diagnosis	57.0 (38.1, 73.0)	35	1	35/35

**Table 2.** Characteristics of breast cancer cases and controls used for biomarker discovery and validation

Characteristic	Discovery		Validation	
	Cases (n = 490)	Controls (n = 490)	Cases (n = 198)	Controls (n = 198)
	n (%)	n (%)	n (%)	n (%)
Age at blood draw, y				
50–59	127 (25.9)	127 (25.9)	41 (20.7)	41 (20.7)
60–69	237 (48.4)	237 (48.4)	91 (46.0)	91 (46.0)
70–79	126 (25.7)	126 (25.7)	66 (33.3)	66 (33.3)
Race/ethnicity				
Non-Hispanic white	435 (88.8)	435 (88.8)	179 (90.4)	179 (90.4)
African American	22 (4.5)	22 (4.5)	8 (4.0)	8 (4.0)
Hispanic white	14 (2.9)	14 (2.9)	3 (1.5)	3 (1.5)
Asian/Pacific Islander	14 (2.9)	14 (2.9)	6 (3.0)	6 (3.0)
Other	5 (1.0)	5 (1.0)	2 (1.0)	2 (1.0)
Use of menopausal hormone therapy at blood draw				
Never user	141 (28.8)	173 (35.3)	51 (25.8)	69 (34.8)
Former user	60 (12.2)	78 (15.9)	23 (11.6)	20 (10.1)
Current unopposed estrogen user	131 (26.7)	136 (27.8)	49 (24.7)	53 (26.8)
Current estrogen and progestin user	158 (32.2)	103 (21.0)	75 (37.9)	56 (28.3)
Time to breast cancer diagnosis				
0–38 wk	245 (50.0)	—	71 (35.9)	—
38–74 wk	245 (50.0)	—	127 (64.1)	—
First-degree family history of breast cancer				
No	378 (77.5)	393 (80.2)	157 (79.3)	151 (76.6)
Yes	110 (22.5)	97 (19.8)	41 (20.7)	46 (23.4)
Missing	2 (0.4)	0 (0)	0 (0)	1 (0.5)
BMI, kg/m <sup>2</sup>				
<25.0	199 (41.2)	224 (46.3)	80 (40.4)	92 (46.7)
25.0–29.9	166 (34.4)	156 (32.2)	74 (37.4)	62 (31.5)
>30.0	118 (24.4)	104 (21.5)	44 (22.2)	43 (21.8)
Missing	7 (1.4)	6 (1.2)	0 (0)	1 (0.5)
Hysterectomy				
No	309 (63.1)	291 (59.5)	132 (67.0)	121 (61.1)
Yes	181 (36.9)	198 (40.5)	65 (33.0)	77 (38.9)
Missing	0 (0.0)	1 (0.2)	1 (0.5)	—

used in either our discovery experiments or the first-round validation. This second-round validation focused only on EGFR and TFF3 based on the promising data generated from the first-round validation as an effort to strengthen our validation conclusions in an additional independent set. A sample size of ~100 cases and ~100 controls was selected for each round of validation because it provided 70% power to characterize a marker with an AUC of 0.60 and a sensitivity of 20%, and 99% power to characterize a marker with an AUC of 0.65 and a sensitivity of 30%.

### Statistical methods

ELISA measurements above and below the detection limit for assays were imputed by the maximum and minimum computable values for the assay. All statistical analyses were done on a log 2 scale. Values were Winsorized by 99%/1%

percentile to compensate for the influence of extreme values and were standardized such that the mean value and SD of the control group of each set of assays were set to 0 and 1, respectively. Multivariate logistic regression was used to compute odds ratios (OR), *P* values, and 95% confidence intervals (95% CI) comparing cases to controls. All ORs were adjusted for the following potential confounders: age, time between blood draw and cancer diagnosis for cases/matched reference date for controls, race/ethnicity, body mass index (BMI) at baseline, and first-degree family history of breast cancer. Analyses stratified by recency of menopausal hormone therapy use at the time of blood draw [never/former user, current unopposed estrogen (E) user, current estrogen plus progestin (E + P) user] were conducted. Of these covariates, BMI was measured and family history of breast cancer and use of menopausal hormone therapy were self-reported.

**Table 3.** Results from the first round of ELISA-based validations on 105 case/control pairs

Protein name	Abbreviation	Discovery results			Validation results	
		Fold change* (95% CI)	P	Subset with lowest P value	OR <sup>†</sup> (95% CI)	P
Epidermal growth factor receptor	EGFR	1.17 (0.98–1.40)	0.070	ER <sup>+</sup> /PR <sup>+</sup>	1.64 (1.20–2.24)	0.002
Fibronectin 1	FN1	0.77 (0.64–0.93)	0.020	All experiments	1.24 (0.95–1.61)	0.11
Protein NOV homologue	NOV	1.54 (1.11–2.14)	0.029	ER <sup>+</sup> /PR <sup>+</sup>	1.21 (0.89–1.64)	0.22
Insulin-like growth factor binding protein 1	IGFBP1	0.85 (0.77–0.93)	0.003	ER <sup>+</sup> , 0–33 wk	1.14 (0.86–1.51)	0.35
von Willebrand factor	VWF	0.68 (0.46–1.00)	0.048	All experiments	1.11 (0.83–1.49)	0.49
Trefoil factor 3	TFF3	1.28 (1.08–1.51)	0.010	All experiments	1.11 (0.83–1.50)	0.47
Lactoferrin	LTF	1.80 (1.19–2.73)	0.025	All experiments	0.98 (0.75–1.28)	0.88

\*Fold change refers to geometric mean of case-to-control ratio across samples in the subset.

†ORs were computed using logistic regression and were adjusted for age, time between blood draw and breast cancer diagnosis, race/ethnicity, BMI, and first-degree family history of breast cancer.

In analyses where samples were paired, the time to diagnosis for a healthy control was set to that of her matched case.

## Results

In both the discovery and validation sets, cases and controls had similar frequencies of first-degree family history of breast cancer, but cases were somewhat more likely to be current E + P users and to be overweight/obese (BMI  $\geq$  25.0 kg/m<sup>2</sup>) and less likely to have had a hysterectomy compared with controls (Table 2).

The discovery experiments acquired 3,412,733 tandem mass spectra from which 3,154 proteins were identified and assigned to 1,491 groups of indistinguishable proteins. Acrylamide labels cysteine-containing peptides, and thus only 603 of the protein groups had measurable ratios. After removing proteins that were either immunodepleted or had defunct IPI entries, 503 protein groups remained for statistical analysis.

Although the ORs and P values for EGFR, FN1, IGFBP1, LTF, NOV, TFF3, and VWF did not necessarily place them near the top of our overall candidate list, they were selected for validation based on the fact that these were the only candidates with commercially available ELISA assays.

The overall results of our first round of ELISA validations on 105 cases and 105 controls are shown in Table 3. Overall, levels of EGFR significantly differentiated cases from controls (OR, 1.64; P = 0.002), whereas levels of FN1, IGFBP1, LTF, TFF3, NOV, and VWF did not. Although the multivariate adjusted ORs estimated from our ELISA validation are not directly comparable to the fold changes calculated from our discovery data using pooled samples, the EGFR discovery and validation data are consistent in showing that EGFR levels are on average higher in cases compared with controls. Menopausal hormone therapy has been shown to affect a significant portion of the serum proteome (6). For this reason, we stratified our validation

**Table 4.** First- and second-round validation results for EGFR

ELISA round	Menopausal hormone therapy use											
	Overall			Never/former users			Current E users			Current E + P users		
	n*	OR <sup>†</sup> (95% CI)	P	n*	OR <sup>†</sup> (95% CI)	P	n*	OR <sup>†</sup> (95% CI)	P	n*	OR <sup>†</sup> (95% CI)	P
Round 1	105/105	1.64 (1.20–2.24)	0.002	42/51	1.08 (0.71–1.66)	0.71	41/24	1.72 (0.86–3.42)	0.12	22/30	4.54 (1.71–12.05)	0.002
Round 2	93/93	1.31 (0.96–1.81)	0.09	32/38	0.96 (0.58–1.61)	0.89	34/32	1.17 (0.57–2.41)	0.67	27/23	2.13 (1.13–4.03)	0.02
Rounds 1 and 2	198/198	1.44 (1.16–1.79)	0.0008	74/89	1.05 (0.76–1.43)	0.78	75/56	1.43 (0.90–2.27)	0.13	49/53	2.49 (1.56–3.99)	0.0001

\*Number of cases/controls.

†ORs were computed using logistic regression and were adjusted for age, time between blood draw and breast cancer diagnosis, race/ethnicity, BMI, and first-degree family history of breast cancer.

**Table 5.** Quartile distributions of EGFR validation results among all case/control sets and among current estrogen and progestin users

EGFR quartile*	All cases and controls			
	Controls n (%)	Cases n (%)	OR <sup>†</sup> (95% CI)	P
(-2.5, -0.63)	50 (25.3)	29 (14.6)	1.00 (reference)	
(-0.63, 0.02)	49 (24.7)	50 (25.3)	1.79 (0.98–3.32)	0.06
(0.02, 0.66)	49 (24.7)	41 (20.7)	1.47 (0.78–2.76)	0.23
(0.66, 2.48)	50 (25.3)	78 (39.4)	2.90 (1.60–5.32)	0.0005

EGFR quartile*	Current estrogen and progestin users			
	Controls n (%)	Cases n (%)	OR <sup>†</sup> (95% CI)	P
(-1.87, -0.63)	14 (25.0)	6 (8.0)	1.00 (reference)	
(-0.63, -0.08)	14 (25.0)	12 (16.0)	2.37 (0.68–8.93)	0.19
(-0.08, 0.46)	14 (25.0)	15 (20.0)	2.95 (0.85–11.25)	0.10
(0.46, 1.85)	14 (25.0)	42 (56.0)	9.04 (2.78–33.21)	0.0004

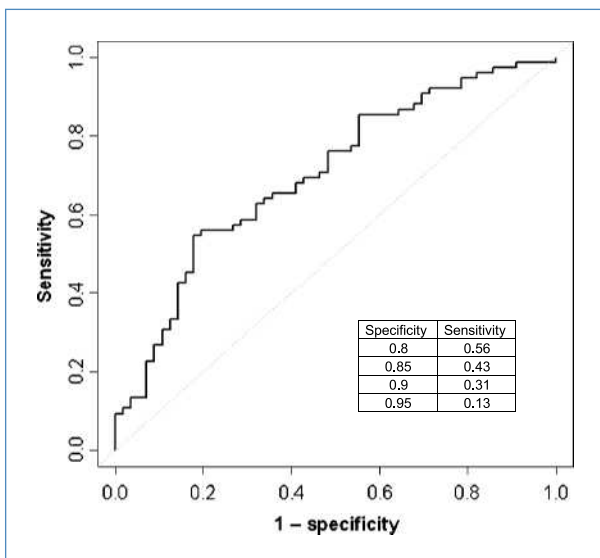
\*Units were standardized to give controls a mean of 0 and a SD of 1. Quartiles were computed from the control distribution.

<sup>†</sup>ORs were computed using logistic regression and were adjusted for age, time, BMI, race, and first-degree family history of breast cancer.

results according to use of menopausal hormone therapy (never/former users, current E users, and current E + P users). The first-round validation data suggested that EGFR levels were elevated in both current E users and current E + P users (Table 4). In addition, the data suggested that whereas TFF3 levels did not distinguish cases from controls, TFF3 levels were higher in current menopausal hor-

none therapy users compared with never/former users regardless of case/control status. E users had an OR of 2.15 ( $P < 0.001$ ) compared with never/former users, whereas E + P users had an OR of 1.90 ( $P < 0.001$ ) compared with never/former users. Associations with the other proteins assessed in our first round of ELISA validations were not influenced by use of menopausal hormones.

To further validate these findings, a second-round validation was conducted where the finding that EGFR levels were elevated in cases compared with controls who were E + P users was replicated (Table 4). Combining data from both validation rounds, EGFR levels were elevated in cases versus controls among all women (OR, 1.44; 95% CI, 1.16–1.79), particularly among E + P users (OR, 2.49; 95% CI, 1.56–3.99). We used logistic regression modeling to test whether differences in the ORs across menopausal hormone therapy use subgroups were statistically different. The  $P$  value comparing the ORs among current E + P users to never/former users was 0.0019, and the  $P$  value comparing current E + P users to current E users was 0.12. We also evaluated risk of breast cancer by quartile of EGFR levels. Overall, women in the highest EGFR quartile had a 2.90-fold (95% CI, 1.60–5.32) increased risk of developing breast cancer compared with those in the lowest quartile, but among current E + P users, those in the highest EGFR quartile had a 9.04-fold (95% CI, 2.78–33.21) increased risk (Table 5). The performance characteristics of EGFR in women taking E + P hormone therapy are shown in Fig. 1. The receiver operating characteristic curve has an AUC of 0.7. At 80% specificity, the sensitivity of EGFR as a single marker is 56%, and at 90% specificity, its sensitivity

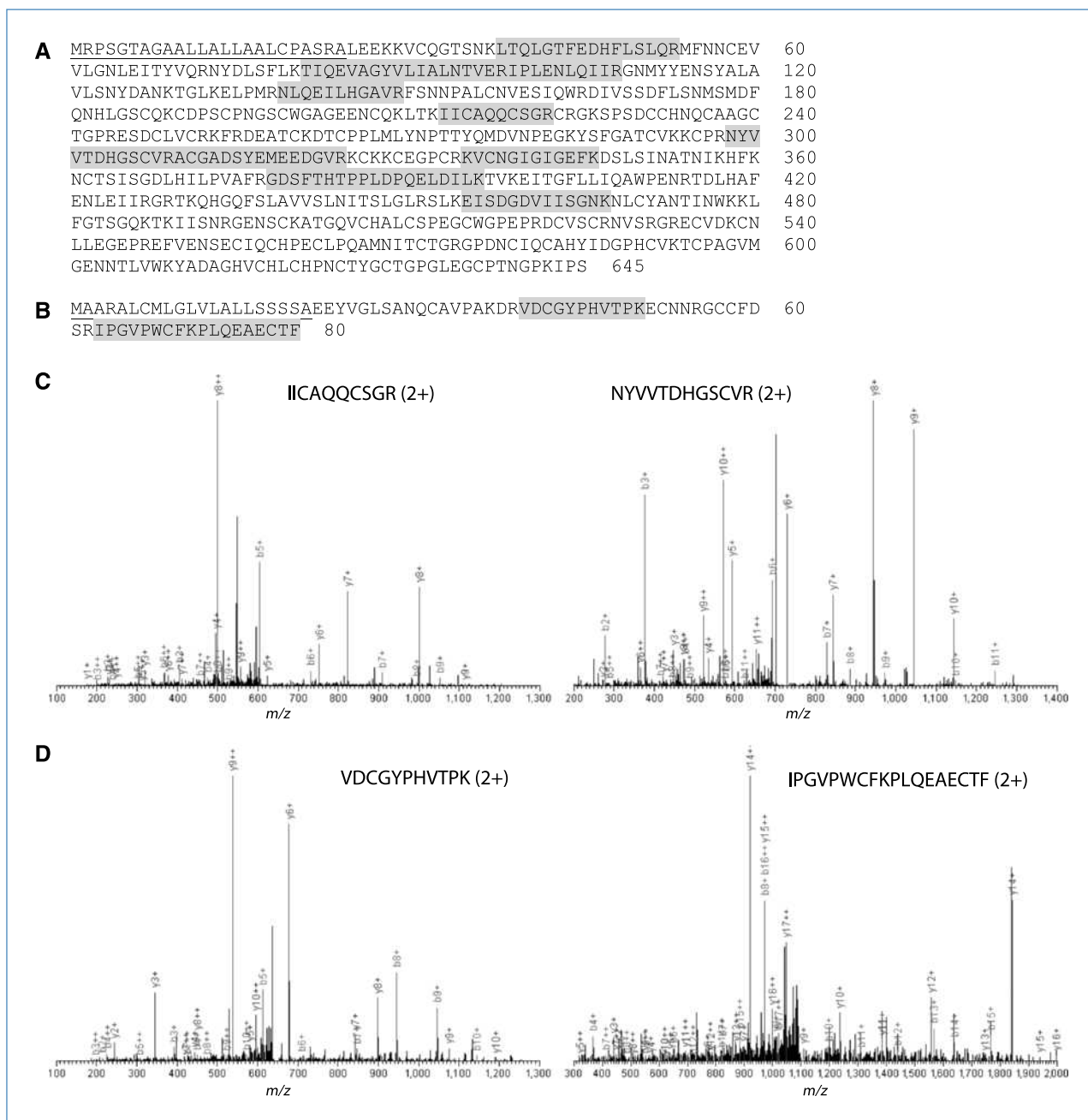


**Figure 1.** EGFR receiver operating characteristic curve for cases versus controls among all current E + P users.

is 31%. Of note, these risk estimates did not vary by breast cancer stage. Combining data from the two rounds of ELISA assays, EGFR was elevated overall among women who went on to be diagnosed with both localized (OR, 1.39; 95% CI, 1.09–1.78) and regional/distant disease (OR, 1.75; 95% CI, 1.04–2.92) and among current E + P users who went on to be diagnosed with both localized (OR, 2.52; 95% CI, 1.37–4.64) and regional/distant disease (OR, 2.84; 95% CI, 1.03–7.80).

The peptide sequence coverage for EGFR in the discovery experiments is shown in Fig. 2A. EGFR is a transmembrane protein, and the peptides identified by MS are all located on the extracellular region. The TFF3 peptides observed in the discovery work are shown in Fig. 2B. TFF3 is a small secreted protein, and the sequence coverage suggests that we are observing the intact form of the secreted protein.

In our second-round validation, we also replicated our TFF3 results. When we combined data from both validation



**Figure 2.** Sequence coverage for (A) EGFR (extracellular domain) and (B) TFF3. The signal peptide is underlined, and identified peptides are highlighted. Representative MS/MS spectra for quantified peptides for (C) TFF3 and (D) EGFR.

rounds, TFF3 levels were elevated among both unopposed estrogen users (OR, 2.17;  $P < 0.001$ ) and E + P users (OR, 1.93;  $P < 0.001$ ) compared with never/former users.

## Discussion

There is paucity of breast cancer proteomic studies that have used prediagnostic blood samples. From a list of candidates identified by a high-dimensional MS-based discovery approach, we attempted to validate a handful of proteins that were not particularly highly ranked but met particular statistical criteria and had an available ELISA assay that could readily be used for validation. Of the seven proteins assessed, EGFR may have some utility as an indicator of breast cancer risk before diagnosis, particularly among current E + P users, based on its high statistical significance in our validation set. However, its performance with respect to its sensitivity and specificity is insufficient for it to serve as a single early detection marker.

We found that EGFR had an AUC of 0.70; at 80% specificity, its sensitivity is 56%, and at 90% specificity, its sensitivity is 31%. In comparison, prostate-specific antigen, which is clinically used to screen men for prostate cancer, has an AUC of 0.68; at 81.1% specificity, its sensitivity is 40.5%, and at 93.8% specificity, its sensitivity is 20.5% (11). Whereas EGFR is insufficient as a stand-alone early detection marker, its utility as a risk factor in conjunction with mammography has yet to be determined. For example, more frequent mammography among E + P users with elevated EGFR levels could potentially be useful for detecting breast cancer earlier among these women, but this needs to be formally evaluated.

EGFR is a cell surface tyrosine kinase receptor and is a member of the *ERBB* proto-oncogene family, which also includes *HER2*. The identification of extracellular peptides and no transmembrane or intracellular peptides by MS in our discovery IPAS experiments suggests shedding of the EGFR extracellular domain by cells. EGFR is involved in numerous cancer-relevant pathways involving cell proliferation, survival, differentiation, and migration, and binding of EGFR by various ligands can result in increased uncontrolled proliferation of cancer cells (12, 13). Further, EGFR is overexpressed in 20% to 81% of breast tumors (14–16). Several studies have measured blood levels of EGFR in relation to breast cancer, although overall, the results are quite inconsistent. It is difficult to directly compare the results of these studies to ours because none involved measurements of EGFR in the preclinical period before a breast cancer patient's diagnosis. Among women with hormone receptor–positive disease, serum EGFR levels decreased significantly after 1 and 3 months of letrozole therapy versus pretreatment conditions (17). Reduction in serum levels of EGFR in postoperative breast cancer cases compared with preoperative cases has been shown to correlate with disease-free survival (18). Whereas EGFR is involved in hormonal pathways relevant to breast cancer, there is no clear explanation at this point for why EGFR may be useful

for the early detection of breast cancer primarily among E + P users, but not among either E users or never/former users. Thus, further investigations of the potential utility of EGFR as an indication of increased risk of breast cancer are warranted.

TFF3 is a small stable secreted protein that is predominantly expressed in the gastrointestinal tract as well as in a variety of other normal tissues and tumors. TFF3 has been shown to be estrogen responsive at both the gene and protein levels. The gene encoding TFF3 contains a palindromic estrogen response element that is conserved between human and mouse (19). We have recently shown an association between proteins containing estrogen response elements in the upstream region of their gene sequence, such as TFF3, and increased serum levels when comparing women taking oral estrogen treatment after 1 year compared with baseline (6). Our observation that TFF3 levels are elevated in current E and current E + P users compared with never/former users is consistent with this work. Although TFF3 was upregulated in our discovery work, this was likely due to the imbalance in hormone therapy use in cases compared with controls (58.9% versus 48.8% were current users) in our discovery experiments.

Of the seven proteins we attempted to validate, EGFR was the only one that showed statistically significant differences in overall levels for cases compared with controls. The likely primary reason why we failed to validate the other candidates is that all of them had a relatively high false discovery rate, as we used fairly broad statistical criteria when including candidates on our list of those warranting further follow-up. The primary limitation of this study is the lack of readily available means to validate the numerous other candidates we discovered, most of which were much more compelling candidates based on their discovery ORs,  $P$  values, and false discovery rates.

Whereas EGFR alone cannot be viewed as clinically useful for early detection of breast cancer, our validation of increased EGFR levels as an indication of increased risk for ER<sup>+</sup> breast cancer specific to E + P users is important in two respects. First, no prior studies have validated even a single biomarker for early detection of breast cancer in preclinical specimens to the degree we have done in this study, validating increased EGFR levels in two completely independent validation sets. Second, consideration of other exposures in biomarker discovery studies is likely critical given that although overall EGFR was not found to distinguish cases from controls among E + P users, it is highly statistically significant. Thus, incorporation of factors such as use of hormone therapy, which has been shown to have a major effect on the plasma proteome (6), is critical in this type of work. Future work aimed at discovering and validating preclinical changes related to breast cancer is needed, and our EGFR finding warrants further replication and study.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.



## Acknowledgments

**WHI Program Office:** Jacques Rossouw, Shari Ludlam, Joan McGowan, Leslie Ford, and Nancy Geller (National Heart, Lung, and Blood Institute, Bethesda, Maryland).

**WHI Clinical Coordinating Center:** Ross Prentice, Garnet Anderson, Andrea LaCroix, Charles L. Kooperberg (Fred Hutchinson Cancer Research Center, Seattle, WA); Evan Stein (Medical Research Labs, Highland Heights, KY); Steven Cummings (University of California at San Francisco, San Francisco, CA).

**WHI Clinical Centers:** Sylvia Wassertheil-Smoller (Albert Einstein College of Medicine, Bronx, NY); Haleh Sangi-Haghepeykar (Baylor College of Medicine, Houston, TX); JoAnn E. Manson (Brigham and Women's Hospital, Harvard Medical School, Boston, MA); Charles B. Eaton (Brown University, Providence, RI); Lawrence S. Phillips (Emory University, Atlanta, GA); Shirley Beresford (Fred Hutchinson Cancer Research Center, Seattle, WA); Lisa Martin (George Washington University Medical Center, Washington, DC); Rowan Chlebowski (Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA); Erin LeBlanc (Kaiser Permanente Center for Health Research, Portland, OR); Bette Caan (Kaiser Permanente Division of Research, Oakland, CA); Jane Morley Kotchen (Medical College of Wisconsin, Milwaukee, WI); Barbara V. Howard (MedStar Research Institute/Howard University, Washington, DC); Linda Van Horn (Northwestern University, Chicago/Evanston, IL); Henry Black (Rush Medical Center, Chicago, IL); Marcia L. Stefanick (Stanford Prevention Research Center, Stanford, CA); Dorothy Lane (State University of New York at Stony Brook, Stony Brook, NY); Rebecca Jackson (The Ohio State University, Columbus, OH); Cora E. Lewis (University of Alabama at Birmingham, Birmingham, AL); Cynthia A. Thomson (University of Arizona, Tucson/Phoenix, AZ); Jean Wactawski-Wende (University at Buffalo, Buffalo, NY); John Robbins (University of California at Davis, Sacramento, CA); F. Allan Hubbell (University of

California at Irvine, CA); Lauren Nathan (University of California at Los Angeles, Los Angeles, CA); Robert D. Langer (University of California at San Diego, LaJolla/Chula Vista, CA); Margery Gass (University of Cincinnati, Cincinnati, OH); Marian Limacher (University of Florida, Gainesville/Jacksonville, FL); J. David Curb (University of Hawaii, Honolulu, HI); Robert Wallace (University of Iowa, Iowa City/Davenport, IA); Judith Ockene (University of Massachusetts/Fallon Clinic, Worcester, MA); Norman Lasser (University of Medicine and Dentistry of New Jersey, Newark, NJ); Mary Jo O'Sullivan (University of Miami, Miami, FL); Karen Margolis (University of Minnesota, Minneapolis, MN); Robert Brunner (University of Nevada, Reno, NV); Gerardo Heiss (University of North Carolina, Chapel Hill, NC); Lewis Kuller (University of Pittsburgh, Pittsburgh, PA); Karen C. Johnson (University of Tennessee Health Science Center, Memphis, TN); Robert Brzyski (University of Texas Health Science Center, San Antonio, TX); Gloria E. Sarto (University of Wisconsin, Madison, WI); Mara Vitolins (Wake Forest University School of Medicine, Winston-Salem, NC); Michael S. Simon (Wayne State University School of Medicine/Hutzel Hospital, Detroit, MI).

## Grant Support

The WHI program is funded by the National Heart, Lung, and Blood Institute, NIH, U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 05/11/2010; revised 07/25/2010; accepted 08/20/2010; published OnlineFirst 10/19/2010.

## References

- Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. *JAMA* 1995; 273:149-54.
- Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J Clin* 2009;59:225-49.
- Hortobagyi GN, de la Garza Salazar J, Pritchard K, et al. The global breast cancer burden: variations in epidemiology and survival. *Clin Breast Cancer* 2005;6:391-401.
- Hays J, Hunt JR, Hubbell FA, et al. The Women's Health Initiative recruitment methods and results. *Ann Epidemiol* 2003;13:S18-77.
- The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Control Clin Trials* 1998;19:61-109.
- Katayama H, Paczesny S, Prentice R, et al. Application of serum proteomics to the Women's Health Initiative conjugated equine estrogens trial reveals a multitude of effects relevant to clinical findings. *Genome Med* 2009;1:47.
- Faca V, Coram M, Phanstiel D, et al. Quantitative analysis of acrylamide labeled serum proteins by LC-MS/MS. *J Proteome Res* 2006; 5:2009-18.
- Faca V, Pitteri SJ, Newcomb L, et al. Contribution of protein fractionation to depth of analysis of the serum and plasma proteomes. *J Proteome Res* 2007;6:3558-65.
- Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;3:Article3.
- Bignotti E, Ravaggi A, Tassi RA, et al. Trefoil factor 3: a novel serum marker identified by gene expression profiling in high-grade endometrial carcinomas. *Br J Cancer* 2008;99:768-73.
- Thompson IM, Ankerst DP, Chi C, et al. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. *JAMA* 2005;294:66-70.
- Wieduwilt MJ, Moasser MM. The epidermal growth factor receptor family: biology driving targeted therapeutics. *Cell Mol Life Sci* 2008;65:1566-84.
- Mitsudomi T, Yatabe Y. Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. *FEBS J* 2010;277:301-8.
- Abd El-Rehim DM, Pinder SE, Paish CE, et al. Expression and co-expression of the members of the epidermal growth factor receptor (EGFR) family in invasive breast carcinoma. *Br J Cancer* 2004;91: 1532-42.
- Hudelist G, Singer CF, Manavi M, Pischinger K, Kubista E, Czerwenka K. Co-expression of ErbB-family members in human breast cancer: Her-2/neu is the preferred dimerization candidate in nodal-positive tumors. *Breast Cancer Res Treat* 2003;80:353-61.
- Tsutsui S, Kataoka A, Ohno S, Murakami S, Kinoshita J, Hachitanda Y. Prognostic and predictive value of epidermal growth factor receptor in recurrent breast cancer. *Clin Cancer Res* 2002;8:3454-60.
- Lafky JM, Baron AT, Cora EM, et al. Serum soluble epidermal growth factor receptor concentrations decrease in postmenopausal metastatic breast cancer patients treated with letrozole. *Cancer Res* 2005;65:3059-62.
- Rocca A, Cancellato G, Bagnardi V, et al. Perioperative serum VEGF and extracellular domains of EGFR and HER2 in early breast cancer. *Anticancer Res* 2009;29:5111-9.
- Bourdeau V, Deschenes J, Metivier R, et al. Genome-wide identification of high-affinity estrogen response elements in human and mouse. *Mol Endocrinol* 2004;18:1411-27.