Discovery and preliminary confirmation of novel early detection biomarkers for triple-negative breast cancer using preclinical plasma samples from the Women's Health Initiative observational study

Christopher I. Li[1], **Justin E. Mirus**[1], **Yuzheng Zhang**[1], **Arturo B. Ramirez**[1], **Jon J. Ladd**[1], **Ross L. Prentice**[1], **Martin McIntosh**[1], **Samir M. Hanash**[1], and **Paul D. Lampe**[1]

[1]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA

Abstract

Triple-negative breast cancer is a particularly aggressive and lethal breast cancer subtype that is more likely to be interval-detected rather than screen-detected. The purpose of this study is to discover and initially validate novel early detection biomarkers for triple-negative breast cancer using preclinical samples. Plasma samples collected up to 17 months prior to diagnosis from 28 triple-negative cases and 28 matched controls from the Women's Health Initiative Observational Study were equally divided into a training set and a test set and interrogated using a customized antibody array. Data were available on 889 antibodies, and in the training set statistically significant differences in case vs. control signals were observed for 93 (10.5%) antibodies at p<0.05. Of these 93 candidates, 29 were confirmed in the test set at p<0.05. Areas under the curve for these candidates ranged from 0.58 to 0.79. With specificity set at 98%, sensitivity ranged from 4% to 68% with 20 candidates having a sensitivity 20% and 6 having a sensitivity 40%. In an analysis of KEGG gene sets, the pyrimidine metabolism gene set was upregulated in cases compared to controls (p=0.004 in the testing set) and the JAK/Stat signaling pathway gene set was downregulated (p=0.003 in the testing set). Numerous potential early detection biomarkers specific to triple-negative breast cancer in multiple pathways were identified. Further research is required to follow-up on promising candidates in larger sample sizes and to better understand their potential biological importance as our understanding of the etiology of triple-negative breast cancer continues to grow.

Keywords

Breast cancer; triple-negative; biomarkers; early detection

Introduction

Annual or biennial mammography is effective at detecting breast cancer early and has been shown in multiple randomized trials to reduce mortality rates.[18] However, its effectiveness varies by breast cancer subtype. With respect to hormone receptor status, it has been shown that interval-detected cancers are 1.8 to 2.6-fold more likely to be estrogen receptor (ER) negative compared to screen-detected tumors.[6, 26] Improving the early detection of ER-

Corresponding author: Christopher I. Li, MD, PhD, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., M4-C308, P.O. Box 19024, Seattle, WA 98109-1024, Telephone: 206.667.7444, Fax: 206.667.5948, cili@fhcrc.org.

Conflicts of interest
The authors declare that they have no conflict of interest.

cancers is of great clinical importance because these tumors are more likely to present at an advanced stage, and a higherstage carries a higher risk of breast cancer mortality.[7]

One approach to developing new tools for detecting cancer early is through the identification and validation of blood based cancer specific biomarkers. In applying this approach to breast cancer, one potential challenge is its considerable heterogeneity. The characterization of distinct molecular subtypes of breast cancer based on patterns of gene expression has shifted how we approach this complex disease.[20, 32] The unique molecular signatures of the different subtypes suggest that they likely have unique etiologies, and a growing number of studies indicate that several well established breast cancer risk factors differ markedly in their associations with the various molecular subtypes.[8, 16, 21–25] The most common subtypes are ER+ (comprising the luminal A and luminal B subtypes), while one of the most aggressive and difficult to treat subtypes is triple-negative (TN) breast cancer. These tumors lack ER, progesterone receptor (PR), and HER2-neu (HER2) expression and the majority of them have the so called basal-like phenotype.[5, 12] Beyond their molecular differences, this subtyping is also of considerable clinical relevance given the differences in survival rates of luminal A and TN cancers: while luminal A tumors have a ~90% 5-year survival rate, the reported 5-year survival rate for TN breast cancers ranges from 35–80%.[3, 5, 11, 15] Thus, given the molecular, clinical and epidemiological differences from ER+ cancers, one might reasonably hypothesize that there may be unique early detection biomarkers specific to TN breast cancer, and that biomarkers for this subtype may be more readily discovered given the highly aggressive nature of these tumors. One challenge to the discovery of useful biomarkers for TN disease is the procurement of sufficient samples collected prior to disease diagnosis. Large cohort studies that have collected biospecimens and have good follow-up are excellent potential sources.

The purpose of this study was to discover and initially validate novel biomarkers for the early detection of TN breast cancer using a novel high-density antibody array and plasma samples collected prior to diagnosis among women enrolled in the Women's Health Initiative (WHI) observational study. The antibody microarray contains approximately 1000 antibodies to many important signaling proteins important in inflammatory, immune response, proliferation, and insulin signaling pathways. Content includes many cytokines, adipokines and other growth factors, and is enriched for antibodies to secreted and/or membrane proteins. This includes proteins in pathways known to be deregulated in breast cancer including those involved in apoptosis, angiogenesis, T-cell activation/infiltration, inflammation/prostaglandins, insulin, and insulin resistance signaling. Since antibodies are nature's best affinity capture reagents, they are perfectly suited for characterizing complex proteomes such as human plasma due to their high affinity and specificity and when used in a high dimensional format can give a rather comprehensive view of the plasma proteome. We have previously shown that this approach has excellent concordance with ELISA assays for specific proteins, and has yielded new biomarkers of ovarian cancer that have been confirmed by alternate methods.[14, 28–30]

Methods

Study design

We conducted a nested case-control study of breast cancer within the Women's Health Initiative (WHI) Observational Study (OS), a prospective cohort of 93,676 post-menopausal women enrolled from 1993 to 1998 in the United States. Detailed descriptions of the design and methods of the WHI OS have been previously published.[9, 33] Our nested case-control study included 28 ER-/PR-/HER2- breast cancer cases and 28 controls without a prior history of any type of cancer, individually matched 1:1 to cases on age at enrollment (±1 year), race/ethnicity (white, black, Hispanic, Asian/Pacific Islander, or other), blood draw

date (±1 year), and clinical center of enrollment. Cases were included in this study if they had an available study blood specimen drawn within 17 months prior to their breast cancer diagnosis. Information on ER, PR, and HER2 status was obtained from medical records and centrally adjudicated by WHI staff. The 28 matched sets were divided equally and randomly into a training set, used for discovery, and an independent testing set, used for confirmation.

Laboratory methods

These preclinical samples were evaluated on a customized antibody array populated with 977 full length antibodies to many secreted, integral membrane, cytoplasmic and nuclear proteins involved in a diverse array of signaling pathways. Detailed descriptions of our protocols for array fabrication, plasma treatment, plasma labeling, incubation of plasma with arrays, array scanning and statistical analyses have been previously reported.[14, 29] Triplicate features of each antibody were printed on Nexterion Slide H hydrogel-coated glass slides (Schott, Elmsford, NY) using a Genetix Q-array 2 microarray printer (San Jose, CA) and blocked with 0.3% ethanolamine, 0.05 M sodium borate pH 8.0. Albumin and IgG were depleted from 100 ul plasma using a ProteoPrep Immunoaffinity Albumin and IgG Depletion Kit (Sigma Chemical, St. Louis, MO) per the manufacturer's directions. The depleted plasma was concentrated to its original volume using Amicon Ultra 10k MWCO centrifugal filters (Millipore, Billerica, MA), measured for total protein concentration by BCA assay (Pierce Biotechnology, Rockford, IL), and labeled with the amine reactive dyes Cy3- and Cy5-maleimide (GE Amersham, Piscataway, NJ) according to the manufacturer's instructions. Unincorporated dye was removed using Amicon Ultra 10k MWCO centrifugal filters. For this study, 500 µg case and control plasma were labeled with Cy5, and separately incubated for 90 minutes with Cy3-labeled reference plasma (a common pool of plasma comprised of samples collected from 7 women aged 45–72 years was used as a reference for all samples) in approximately 100 µl total volume (kept from drying using LifterSlips, Fisher Scientific, Pittsburgh, PA). After washing slides were scanned in a GenePix 4000B microarray scanner and data extracted using GenePix Pro 6.0 software (Molecular Devices, Sunnyvale, CA).

Statistical analysis

For each antibody, fold change of signal (red channel) compared to reference (green channel), the M value, was calculated as $\log 2(R_c/G_c)$; where R_c is red corrected and G_c is green corrected (using the normexp background correction method developed by Smyth). [31] Technical sources of variation were normalized using loess procedures developed for microarrays, including within-array print-tip loess and between-arrays quartile normalization. Following normalization, triplicate features were summarized using their median. M values were further normalized using linear regression to remove the systematic bias due to experimental factors such as printing and hybridization day. After this normalization data were available on a total of 889 antibodies. All statistical analyses were conducted on M values, and analyses of the training set data and testing set data were performed independently.

Values were standardized such that the mean value and standard deviation of the cancer free control group were set to zero and one, respectively. Multivariate linear regression was used to compute log2 odds ratios (OR), p-values, and 95% confidence intervals (CI) for case versus control comparisons. All ORs were adjusted for age, race/ethnicity, body mass index, menopausal hormone therapy use, and array hybridization day. We also calculated the area under the curve (AUC) and the sensitivity at 98% specificity as metrics of the extent to which individual markers could discriminate between cases and controls. AUC estimates were two-sided, such that an AUC>0.5 indicates that the marker is higher in cases compared to controls and an AUC<0.5 indicates that it is lower among cases compared to controls.

We conducted Gene Set Enrichment Analyses (GSEA) based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) gene sets that are available from the Molecular Signatures Database (MSigDB) (http://www.broadinstitute.org/gsea/msigdb/index.jsp). The 889 antibodies available for analysis correspond to 732 unique genes. Of the 186 KEGG gene sets available from MSigDB, our arrays included at least 5 proteins corresponding to gene members in 91 of these gene sets. With respect to the GO gene sets, of the 1454 available GO gene sets, our arrays included at least 5 gene members in 535 of these gene sets. We then tested if the proteins corresponding to groups of genes in a given gene set had a higher statistical ranking than the proteins not in this gene set based on Wilcoxon testing in the training set. Gene sets that were statistically significantly different in cases compared to controls were then evaluated in the testing set using the same approach.

Results

The cases and controls in our training and test sets were generally balanced with respect to age and race/ethnicity (Table 1). Somewhat higher proportions of cases compared to controls were overweight or obese (body mass index $\geq$ 25.0 kg/m^2) in both the training and testing sets. The same proportion of controls and cases in both training and testing sets were not currently using menopausal hormone therapy, though there was some variation in the proportions using unopposed estrogen versus combined estrogen and progestin regimens. In the training set, of the 889 antibodies assessed, 93 (10.5%) showed statistically significant differences in signal between cases and controls at p<0.05; in the testing set, 146 (16.4%) were statistically different. Of the 93 candidates identified in the training set, 29 were validated in the test set at p<0.05 (Table 2). 28 of these 29 candidates were higher in cases compared to controls and the AUCs for these individual candidates ranged from 0.58 to 0.79. With sensitivity set at 98%, across these 29 candidates specificity ranged from 4% to 68%. As a comparison, a prostate specific antigen (PSA) value of $\geq$ 4.1 has an estimated 20.5% sensitivity and 93.8% specificity for detecting prostate cancer.[34] Here, 20 of the 29 candidates had a sensitivity of $\geq$ 20% and 6 had a sensitivity of $\geq$ 40% including NADH dehydrogenase 1a subcomplex 10 (NDUFA10, 68%), protein-tyrosine phosphatase, mitochondrial 1 (PTPMT1, 52%), integrin beta-1 (ITGB1, 48%), mast/stem cell growth factor receptor (KIT, 46%), DNA-directed RNA polymerase (POLR2L, 43%), and ephrin-A5 (EFNA5, 41%), again with specificity set at 98%. We also applied a more stringent multiple testing correction procedure to our analysis. Ten candidates from the training set had a p-value of <0.01 and independently five of these ten discriminated TN cases from controls at a nominal p<0.005 in the testing set (DUSP9, EED, EFNA5, ITGB1, and PTPMT1), so that these five have a Bonferroni-corrected p-value of <0.05.

An additional eleven candidates had a p-value of <0.1 in the training set and a p-value of <0.05 in the test set, and 12 had a p-value of <0.05 in the training set and a p-value of <0.1 in the test set (Table 3). Among these twenty-three candidates, 2 had sensitivities of $\geq$ 40% at 98% specificity, toll-like receptor 2 (TLR2, 52%) and anoctamin 1 (ANO1, 43%).

For our gene set analysis, a total of 91 KEGG gene sets met our inclusion criteria of $\geq$ 40% sensitivity at 98% specificity. Of these gene sets, seven had a p-value <0.05 in our training set, of which 2 had p-values <0.05 in our testing set (Table 4). These were pyrimidine metabolism, which was upregulated in cases compared to controls (p-value=0.0004 in the training set and p-value=0.004 in the testing set) and the JAK/Stat signaling pathway, which was downregulated in cases compared to controls (p-value=0.03 in the training set and p-value=0.003 in the testing set). A total of 535 GO gene sets met our inclusion criteria, 62 of which had a p-value <0.05 in our training set. Of these sixty-two, 17 had a p-value <0.05 in the testing set (Table 4). These included several gene sets with overlapping membership

related to cellular and RNA metabolism, regulation of DNA transcription, and interleukin activity.

Discussion

In this biomarker discovery and preliminary confirmation study utilizing preclinical plasma samples, we identified several novel putative early detection biomarkers for triple-negative breast cancer. Interpreting the results of a study such as this is challenging given that our understanding of the molecular characteristics of TN breast cancer is evolving.[13] Numerous functional groups of genes and proteins have been identified to be overexpressed in TN breast cancer including activators of signaling pathways that are deregulated in cancer, cell growth related genes involved in proliferation and cell cycle control, tyrosine kinase receptors that participate in transcription and signal transduction, as well as extracellular matrix receptors and other genes involved in the structure and anchoring of basal epithelial cells.[27] Consequently there are multiple individual and groups of plasma proteins that could feasibly differentiate TN breast cancer cases and cancer-free controls preclinically. As might be expected for biomarkers released into the plasma, we observed that a higher percentage of significantly changed proteins had at least some portion exposed extracellularly compared to the list of antibodies as a whole (46% vs. 37%). Of the 29 proteins with $p<0.05$ in both the training and test sets, seven are membrane proteins with extracellular protein domains, four are secreted, and three are involved in protein export to the plasma membrane (assignments based on Ingenuity Pathway Analysis).

In order to evaluate sets or families of proteins that were changed in cases compared to controls, we performed gene set enrichment analyses (GSEA). Based on this approach, cytokine signaling, and specifically the JAK-STAT signaling pathway, were of interest. With respect to individual cytokines, CCL27 and CCL28 were present at higher levels among cases compared to controls with reasonable AUC values (0.76 and 0.67, respectively) and sensitivities at the designated 98% specificity cut-off (25% and 29%, respectively). Among the other individual biomarkers identified, ephrin A5 (41% sensitivity) has previously been observed in serum[10] and has been described as a potential cancer therapeutic target.[4] SRP54 (41% sensitivity) is also found in plasma[19] and has been shown to be upregulated in breast cancer (http://www.itb.cnr.it/breastcancer/). FAS was observed with two different antibodies (one performing with 36% sensitivity and the other with 25% sensitivity), and FAS/FAS-ligand expression are significant predictors of skeletal spread in primary breast cancers.[1, 2] Cleaved extracellular domains of integral membrane proteins could also be compelling biomarkers. Integrin β1 and integrin β-like 1 proteins (48% and 29% sensitivity at 98% specificity) are interesting candidates due to the role of the β-1 integrins in cell growth control and breast tumor induction.[35] Finally, several statistically significant nuclear and cytoplasmic proteins were observed in the training and test sets. Most of these are known to be expressed in breast tissue and several have been shown to be overexpressed in cancer including DUSP9, TSPO, BRCA1, HEXIM1, POLR2L, UPP1, XBP1, and RNF113A. Many previous biomarker studies have also observed cytoplasmic and nuclear proteins in blood[19] and suggest higher proliferation/ apoptosis/necrosis of tumor tissue may be the cause. Not yet mentioned are some of the top ranked candidates based on sensitivity at 98% specificity, including NDUFA10 (68% sensitivity), PTPMT1 (52%), and KIT (46%). Of these three, KIT has the highest direct relevance to cancer as it regulates cell survival and proliferation. NDUFA10 has both dehydrogenase and oxidoreductase activity and is involved in electron transfer from NADPH to the respiratory chain. PTPM1 is a protein phosphatase that is important for ATP production.

A limitation of the laboratory approach used is that we were only able to evaluate biomarkers for which there were antibodies included on the array. Thus, we did not perform a fully comprehensive assessment of the plasma proteome so the potential of biomarkers not included on the array could not be assessed. This in particular limited our gene set analyses as the gene sets we could assess were limited by the candidates included on the array. The array itself also has certain limitations. Although the array yielded data on 889 putative proteins, it uses only a single antibody to capture the antigen. Consequently, there is neither enzymic amplification nor the specificity inherent in sandwich ELISA assays which require antigens to bind to two different antibodies at different epitopes. However, we have optimized dye labeling and plasma processing methods to concentrate and label the less abundant plasma proteins to levels several-fold higher than in native plasma thereby increasing sensitivity to the point that we could readily detect pg increases in IL1β.[14] Furthermore, this methodology is inherently "discovery" in nature and will require further follow-up and validation of promising biomarkers in independent sample sets. As mentioned previously, the interpretation of our results is also hampered by the limited, though evolving, knowledge regarding the biological underpinnings of triple-negative breast cancer as there is also emerging evidence that there are multiple subtypes of TN disease.[13, 17]

This study suggests that there may be unique early detection biomarkers specific to triple-negative breast cancer. These candidates warrant additional follow-up in larger studies to further characterize their potential clinical utility.

Acknowledgments

References

1. Bebenek M, Dus D, Kozlak J. Fas and Fas ligand as prognostic factors in human breast carcinoma. Med Sci Monit. 2006; 12:CR457–CR461. [PubMed: 17072269]

2. Bebenek M, Dus D, Kozlak J. Fas/Fas-ligand expressions in primary breast cancer are significant predictors of its skeletal spread. Anticancer Res. 2007; 27:215–218. [PubMed: 17352235]

3. Brown M, Tsodikov A, Bauer KR, Parise CA, Caggiano V. The role of human epidermal growth factor receptor 2 in the survival of women with estrogen and progesterone receptor-negative, invasive breast cancer: the California Cancer Registry, 1999–2004. Cancer. 2008; 112:737–747. [PubMed: 18189290]

4. Campbell TN, Attwell S, Arcellana-Panlilio M, Robbins SM. Ephrin A5 expression promotes invasion and transformation of murine fibroblasts. Biochem Biophys Res Commun. 2006; 350:623–628. [PubMed: 17022942]

5. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006; 295:2492–2502. [PubMed: 16757721]

6. Collett K, Stefansson IM, Eide J, Braaten A, Wang H, Eide GE, Thoresen SO, Foulkes WD, Akslen LA. A basal epithelial phenotype is more frequent in interval breast cancers compared with screen detected tumors. Cancer Epidemiol Biomarkers Prev. 2005; 14:1108–1112. [PubMed: 15894660]

7. Dunnwald LK, Rossing MA, Li CI. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. Breast Cancer Res. 2007; 9:R6. [PubMed: 17239243]

8. Gaudet MM, Press MF, Haile RW, Lynch CF, Glaser SL, Schildkraut J, Gammon MD, Douglas TW, Bernstein JL. Risk factors by molecular subtypes of breast cancer across a population-based study of women 56 years or younger. Breast Cancer Res Treat. 2011; 130:587–597. [PubMed: 21667121]

9. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, Rossouw JE. The Women's Health Initiative recruitment methods and results. Ann Epidemiol. 2003; 13:S18–S77. [PubMed: 14575939]

10. Kalin M, Cima I, Schiess R, Fankhauser N, Powles T, Wild P, Templeton A, Cerny T, Aebersold R, Krek W, Gillessen S. Novel prognostic markers in the serum of patients with castration-resistant prostate cancer derived from quantitative analysis of the pten conditional knockout mouse proteome. Eur Urol. 2011; 60:1235–1243. [PubMed: 21741162]

11. Kaplan HG, Malmgren JA. Impact of triple negative phenotype on breast cancer prognosis. Breast J. 2008; 14:456–463. [PubMed: 18657139]

12. Kim MJ, Ro JY, Ahn SH, Kim HH, Kim SB, Gong G. Clinicopathologic significance of the basal-like subtype of breast cancer: a comparison with hormone receptor and Her2/neu-overexpressing phenotypes. Hum Pathol. 2006; 37:1217–1226. [PubMed: 16938528]

13. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011; 121:2750–2767. [PubMed: 21633166]

14. Loch CP, Ramirez AB, Liu Y, Sather CL, Delrow JJ, Scholler N, Garvik GM, Urban ND, McIntosh MW, Lampe PD. Use of High Density Antibody Arrays to Validate and Discover Cancer Serum Biomarkers. Molecular Oncology. 2007; 1:313–320. [PubMed: 19383305]

15. Lund MJ, Trivers KF, Porter PL, Coates RJ, Leyland-Jones B, Brawley OW, Flagg EW, O'Regan RM, Gabram SG, Eley JW. Race and triple negative threats to breast cancer survival: a population-based study in Atlanta, GA. Breast Cancer Res Treat. 2009; 113:357–370. [PubMed: 18324472]

16. Ma H, Wang Y, Sullivan-Halley J, Weiss L, Marchbanks PA, Spirtas R, Ursin G, Burkman RT, Simon MS, Malone KE, Strom BL, McDonald JA, Press MF, Bernstein L. Use of four biomarkers to evaluate the risk of breast cancer subtypes in the women's contraceptive and reproductive experiences study. Cancer Res. 2010; 70:575–587. [PubMed: 20068186]

17. Metzger-Filho O, Tutt A, de AE, Saini KS, Viale G, Loi S, Bradbury I, Bliss JM, Azim HA Jr, Ellis P, Di LA, Baselga J, Sotiriou C, Piccart-Gebhart M. Dissecting the Heterogeneity of Triple-Negative Breast Cancer. J Clin Oncol. 2012 In press.

18. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. Ann Intern Med. 2009; 151:727–742. [PubMed: 19920273]

19. Omenn GS. The Human Proteome Organization Plasma Proteome Project pilot phase: reference specimens, technology platform comparisons, and standardized data submissions and analyses. Proteomics. 2004; 4:1235–1240. [PubMed: 15188391]

20. Perou CM, Sorlie T, Eisen MB, van de RM, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. Nature. 2000; 406:747–752. [PubMed: 10963602]

21. Phipps AI, Buist DS, Malone KE, Barlow WE, Porter PL, Kerlikowske K, Li CI. First-degree family history of breast cancer and triple-negative breast cancer risk. Breast Cancer Res Treat. 2011; 136:671–678. [PubMed: 20814817]

22. Phipps AI, Chlebowski R, Prentice R, McTiernan A, Wactawski-Wende J, Kuller LH, Adams-Campbell LL, Lane D, Stefanick ML, Vitolins M, Kabat GC, Rohan TE, Li CI. Reproductive history and oral contraceptive use in relation to risk of triple-negative breast cancer. J Natl Cancer Inst. 2010 in press.

23. Phipps AI, Chlebowski RT, Prentice R, McTiernan A, Stefanick ML, Wactawski-Wende J, Kuller LH, Adams-Campbell LL, Lane D, Vitolins M, Kabat GC, Rohan TE, Li CI. Body size, physical activity, and risk of triple-negative and estrogen receptor-positive breast cancer. Cancer Epidemiol Biomarkers Prev. 2011; 20:454–463. [PubMed: 21364029]

24. Phipps AI, Malone KE, Porter PL, Daling JR, Li CI. Body size and risk of luminal, HER2-overexpressing, and triple-negative breast cancer in postmenopausal women. Cancer Epidemiol Biomarkers Prev. 2008; 17:2078–2086. [PubMed: 18664548]

25. Phipps AI, Malone KE, Porter PL, Daling JR, Li CI. Reproductive and hormonal risk factors for postmenopausal luminal, HER-2-overexpressing, and triple-negative breast cancer. Cancer. 2008; 113:1521–1526. [PubMed: 18726992]

26. Porter PL, El-Bastawissi AY, Mandelson MT, Lin MG, Khalid N, Watney EA, Cousens L, White D, Taplin S, White E. Breast tumor characteristics as predictors of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst. 1999; 91:2020–2028. [PubMed: 10580027]

27. Rakha EA, Reis-Filho JS, Ellis IO. Basal-like breast cancer: a critical review. J Clin Oncol. 2008; 26:2568–2581. [PubMed: 18487574]

28. Ramirez AB, Lampe PD. Discovery and validation of ovarian cancer biomarkers utilizing high density antibody microarrays. Cancer Biomark. 2010; 8:293–307. [PubMed: 22045360]

29. Ramirez AB, Loch CM, Zhang Y, Liu Y, Wang X, Wayner EA, Sargent JE, Sibani S, Hainsworth E, Mendoza EA, Eugene R, Labaer J, Urban ND, McIntosh MW, Lampe PD. Use of a single-chain antibody library for ovarian cancer biomarker discovery. Mol Cell Proteomics. 2010; 9:1449–1460. [PubMed: 20467042]

30. Scholler N, Gross JA, Garvik B, Wells L, Liu Y, Loch CM, Ramirez AB, McIntosh MW, Lampe PD, Urban N. Use of cancer-specific yeast-secreted in vivo biotinylated recombinant antibodies for serum biomarker discovery. J Transl Med. 2008; 6:41. [PubMed: 18652693]

31. Smyth GK, Speed T. Normalization of cDNA microarray data. Methods. 2003; 31:265–273. [PubMed: 14597310]

32. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003; 100:8418–8423. [PubMed: 12829800]

33. 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. [PubMed: 9492970]

34. Thompson IM, Ankerst DP, Chi C, Lucia MS, Goodman PJ, Crowley JJ, Parnes HL, Coltman CA Jr. 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. [PubMed: 15998892]

35. White DE, Kurpios NA, Zuo D, Hassell JA, Blaess S, Mueller U, Muller WJ. Targeted disruption of beta1-integrin in a transgenic mouse model of human breast cancer reveals an essential role in mammary tumor induction. Cancer Cell. 2004; 6:159–170. [PubMed: 15324699]

Table 1

Descriptive characteristics of triple-negative cases and cancer free controls in the training and testing sets

Characteristic	Training set				Testing set			
	Cases		Controls		Cases		Controls	
	n	%	n	%	n	%	n	%
Age								
50–59	5	35.7	5	35.7	5	35.7	5	35.7
60–69	7	50.0	7	50.0	7	50.0	7	50.0
70–79	2	14.3	2	14.3	2	14.3	2	14.3
Race/ethnicity								
Non-Hispanic white	11	78.6	11	78.6	12	85.7	12	85.7
African American	2	14.3	2	14.3	2	14.3	2	14.3
Hispanic white	1	7.1	1	7.1				
Body mass index, kg/m²								
<25.0 (normal)	4	28.6	7	50.0	5	35.7	7	50.0
25.0–29.9 (overweight)	4	28.6	5	35.7	4	28.6	4	28.6
30.0 (obese)	6	42.9	2	14.3	5	35.7	3	21.4
Current use of menopausal hormone therapy								
Non user	8	57.1	8	57.1	8	57.1	8	57.1
Current unopposed estrogen user	3	21.4	5	35.7	4	28.6	3	21.4
Current estrogen+progestin user	3	21.4	1	7.1	2	14.3	3	21.4

Table 2

Log2 ratios and p-values of top ranked antibodies with p-values <0.05 in both the training and test sets

Antibody name	Gene name	Training set		Test set		AUC	Sensitivity at 98% specificity
		log2 ratio	p-value	log2 ratio	p-value		
P-value <0.05 in both the training set and the test set							
NADH dehydrogenase 1a subcomplex, 10	NDUFA10	1.78	0.005	1.43	0.018	0.79	68%
Protein-tyrosine phosphatase, mitochondrial 1	PTPMT1	1.85	0.000	1.62	0.001	0.72	52%
Integrin beta-1	ITGB1	1.33	0.007	2.22	0.001	0.76	48%
Mast/stem cell growth factor receptor	KIT	0.65	0.049	2.24	0.001	0.75	46%
DNA-directed RNA polymerase L, 7.6 kDa	POLR2L	1.16	0.044	1.00	0.049	0.69	43%
Ephrin-A5	EFNA5	1.09	0.009	1.46	0.002	0.70	41%
Regulator of G-protein signaling 5	RGS5	0.80	0.019	1.34	0.007	0.64	39%
Stomatin-like protein 2	STOML2	2.07	0.018	1.96	0.009	0.66	36%
TNFR superfamily member 6 (Ab1)	FAS	1.69	0.005	1.78	0.006	0.62	36%
Signal transducer and activator of transcription 6	STAT6	2.05	0.013	1.69	0.003	0.69	33%
Cysteine-rich protein 2	CSRP2	1.28	0.010	1.76	0.020	0.60	31%
Single-stranded DNA binding protein 1	SSBP1	0.97	0.049	1.63	0.005	0.62	31%
Embryonic ectoderm development protein	EED	1.96	0.002	1.59	0.007	0.71	31%
Small inducible cytokine A28	CCL28	1.31	0.006	1.27	0.012	0.67	29%
Signal recognition particle 54 kDa protein	SRP54	1.11	0.024	1.19	0.017	0.55	29%
p16	CDKN2A	1.26	0.034	2.00	0.026	0.58	27%
Small inducible cytokine A27	CCL27	1.35	0.020	1.63	0.015	0.76	25%
Antigen 85-A	fbpA	1.30	0.029	1.41	0.008	0.70	25%
Mitogen-activated protein kinase kinase 1	MAP2K1	1.38	0.023	1.49	0.009	0.63	22%
Breast and ovarian cancer susceptibility protein 1	BRCA1	1.63	0.005	1.30	0.013	0.67	20%
Endoglin	ENG	1.79	0.006	1.34	0.007	0.69	19%
Exportin-T	XPOT	1.66	0.049	1.27	0.047	0.67	19%
DiGeorge syndrome critical region gene 6	DGCR6	0.98	0.032	1.70	0.008	0.66	19%
Propionyl-CoA carboxylase alpha chain m	PCCA	1.58	0.048	1.35	0.013	0.60	19%
Adaptor-related protein complex 3, beta 2 subunit	AP3B2	1.98	0.012	1.12	0.011	0.68	19%
Dual specificity protein phosphatase homolog 9	DUSP9	1.45	0.003	2.57	0.005	0.63	18%
Peripheral-type benzodiazepine receptor	TSPO	1.19	0.047	1.46	0.004	0.53	8%

Antibody name	Gene name	Training set		Test set			
		log2 ratio	p-value	log2 ratio	p-value	AUC	Sensitivity at 98% specificity
TNF receptor-associated factor 4	TRAF4	−1.97	0.016	−2.49	0.036	0.41	4%

Table 3

Log2 ratios and p-values of top ranked antibodies with p-values <0.1 in the training and test sets

Antibody name	Gene name	Training set		Test set			
		log2 ratio	p-value	log2 ratio	p-value	AUC	Sensitivity at 98% specificity
P-value <0.1 in the training set and and <0.05 in the test set							
Uridine phosphorylase 1	UPP1	1.32	0.075	1.33	0.045	0.67	33%
YEATS domain-containing protein 4	YEATS4	1.60	0.065	0.66	0.050	0.68	32%
Tyrosine-protein phosphatase, receptor type, E	PTPRE	1.19	0.069	1.59	0.009	0.74	31%
TNFR superfamily member 6 (Ab2)	FAS	1.25	0.053	1.43	0.009	0.66	25%
Hexamethylene bis-acetamide inducible 1	HEXIM1	0.91	0.050	1.20	0.044	0.67	22%
Guanine nucleotide-binding protein beta	GNB4	1.33	0.074	1.99	0.031	0.62	21%
Src-like-adaptor	SLA	1.22	0.063	1.17	0.043	0.67	20%
Apoptotic protease-activating factor 1	APAF1	1.17	0.062	1.74	0.014	0.63	14%
Vacuolar ATP synthase subunit G 1	ATP6V1G1	1.75	0.086	0.88	0.012	0.65	11%
X box-binding protein 1	XBP1	1.16	0.075	1.20	0.021	0.71	8%
PTP, non-receptor type 11	PTPN11	−0.62	0.066	−1.39	0.048	0.30	4%
P-value <0.05 in the training set and and <0.1 in the test set							
Toll-like receptor 2	TLR2	2.10	0.001	1.07	0.087	0.71	52%
Anoctamin 1	ANO1	1.49	0.006	1.52	0.062	0.69	43%
Hydroxysteroid (17-beta) dehydrogenase 10	HSD17B10	1.67	0.013	1.27	0.082	0.60	30%
Integrin, beta-like 1	ITGBL1	1.91	0.004	1.40	0.059	0.60	29%
G patch domain and KOW motifs protein	GPKOW	1.66	0.001	1.06	0.071	0.74	22%
Myosin Va	MYO5A	2.46	0.023	1.41	0.075	0.76	22%
RING finger protein 113A	RNF113A	0.78	0.023	1.22	0.076	0.64	18%
nNOS	NOS1	−0.72	0.006	−1.98	0.063	0.41	8%
p300	EP300	−1.62	0.005	−1.41	0.097	0.36	7%
CD45	PTPRC	−1.82	0.019	−1.41	0.090	0.38	4%
Patched	PTCHD1	−1.35	0.020	−2.13	0.064	0.40	4%
c-Myc	MYC	−1.45	0.011	−1.88	0.063	0.46	0%

Table 4

KEGG and GO gene sets with p-values <0.05 in both the training and testing sets

Gene set name	Number of genes in set	Number of unique genes observed	Training set		Testing set	
			AUC	p-value	AUC	p-value
KEGG sets						
Pyrimidine metabolism	98	6	0.89	0.0004	0.82	0.0040
JAK/Stat signaling pathway	155	48	0.42	0.0308	0.39	0.0033
GO sets						
Behavior	149	19	0.62	0.0496	0.69	0.0016
RNA biosynthetic process	626	66	0.44	0.0484	0.40	0.0017
Transcription DNA dependent	624	66	0.44	0.0484	0.40	0.0017
RNA metabolic process	811	69	0.43	0.0353	0.40	0.0029
Regulation of RNA metabolic process	458	48	0.42	0.0321	0.40	0.0090
Regulation of transcription DNA dependent	453	48	0.42	0.0321	0.40	0.0090
Stress activated protein kinase signaling pathway	49	14	0.64	0.0433	0.66	0.0159
Hematopoietin interferon class D 200 domain cytokine receptor activity	32	7	0.20	0.0070	0.24	0.0163
Cell proliferation	501	76	0.44	0.0336	0.43	0.0180
Regulation of nucleobase, nucleoside, nucleotide, and nucleic acid metabolic process	602	68	0.43	0.0278	0.43	0.0251
Chemoking receptor binding	43	7	0.76	0.0160	0.74	0.0281
G protein coupled receptor binding	54	7	0.76	0.0160	0.74	0.0281
Transmembrane receptor activity	411	45	0.40	0.0087	0.42	0.0315
Regulation of cellular metabolic process	768	98	0.43	0.0196	0.44	0.0327
Interleukin binding	24	5	0.23	0.0386	0.23	0.0343
Interleukin receptor activity	19	5	0.23	0.0386	0.23	0.0343
Regulation of metabolic process	780	101	0.44	0.0206	0.44	0.0343