

Breast Hyperplasias, Risk Signature, and Breast Cancer

Indira Poola¹, Qingqi Yue¹, John W. Gillespie¹, Peggy S. Sullivan², Josephine Aguilar-Jakthong², JianYu Rao², Abeer M. Shaaban³, Edward R. Sauter⁴, and Andrew J. Ricci⁵



Abstract

We address the dilemma faced by oncologists in administering preventative measures to "at risk" patients diagnosed with atypical and nonatypical hyperplasias due to lack of any molecular means of risk stratification and identifying high-risk subjects. Our study purpose is to investigate a four marker risk signature, MMP-1, CEACAM6, HYAL1, and HEC1, using 440 hyperplastic tissues for identifying high-risk subjects who will benefit from preventative therapies. We assayed the markers by IHC and combined their expression levels to obtain a composite value from 0–10, which we called a "Cancer Risk Score." We demonstrate that the four marker-based risk scores predict subsequent cancer development with an accuracy of 91% and 86% for atypical and nonatypical subjects, respectively. We have established a correlation between risk scores

and cancer rates by stratifying the samples into low risk (score ≤ 0.5); intermediate risk (score ≤ 5.4), and high risk (score >5.4) groups using Kaplan–Meier survival analysis. We have evaluated cancer rates at 5, 10, and 15 years. Our results show that the average cancer rates in the first 5 years among low- and intermediate-risk groups were 2% and 15%, respectively. Among high-risk group, the average cancer rates at 5 years were 73% and 34% for atypical and nonatypical subjects, respectively. The molecular risk stratification described here assesses a patient's tumor biology-based risk level as low, intermediate, or high and for making informed treatment decisions. The outcomes of our study in conjunction with the available prophylactic measures could prevent approximately 20%–25% of sporadic breast cancers.

Introduction

Breast cancer continues to be the most diagnosed cancer striking approximately 225,000 every year and the second leading cause of cancer-related deaths for women in the United States (1) in spite of major strides in development of numerous therapies. The high death rates are attributed to the inability to contain cancer once it develops. A significant reduction in the cancer incidence and the deaths can be envisioned if high-risk subjects can be identified at the precancerous stage and treated to stop in the first place.

Currently, prevention of about 5%–10% of breast cancers has become a reality by identifying high-risk BRCA1 and BRCA2 gene mutation carriers among "at risk" women with family history. However, such tools are not available for identifying high-risk subjects among the largest established "at risk group of subjects" diagnosed with proliferative hyperplastic precancerous lesions to prevent sporadic breast cancers.

Surveys show that, in addition to cancer cases, an estimated 600,000–800,000 women are diagnosed with some form of benign breast lesions every year in the United States of which about 50% are proliferative growths that include atypical hyperplasias, usual ductal hyperplasias (UDH), intraductal papilloma, and sclerosing adenosis (2, 3). Numerous clinical studies by Wellings (4), Dupont (5, 6), London (7), Page (8), and others (9–15) have established that diagnosis of atypical and nonatypical proliferative growths increases the cancer risk by 5-fold and 2-fold, respectively. Three studies surveyed the incidence of subsequent cancer among proliferative benign group and reported that approximately 20% among atypical and approximately 10% among nonatypical cohorts subsequently developed cancer (16–18). On the basis of the above reports, about 40, 000–50,000 high-risk women who are diagnosed with proliferative growths every year

¹Silbiotech, Inc., Gaithersburg, Maryland. ²Pathology Division, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California. ³Pathology Division, Leeds Hospitals, Leeds, United Kingdom. ⁴Department of Surgery, Hartford Hospital, Hartford, Connecticut. ⁵Department of Pathology, Hartford Hospital, Hartford, Connecticut.

Current address for A.M. Shaaban: Queen Elizabeth Hospital Birmingham, Birmingham, United Kingdom; current address for E.R. Sauter, NCI, NIH, Rockville, Maryland.

Corresponding Author: Indira Poola, Silbiotech, 7858 Beechcraft Ave, Gaithersburg, MD 20879. Phone: 301-787-1216; Fax: 301-987-9724; E-mail: ipoola@silbiotech.com

Cancer Prev Res 2019;12:471–80

doi: 10.1158/1940-6207.CAPR-19-0051

©2019 American Association for Cancer Research.

subsequently develop cancer and are the most suitable candidates for preventative therapies similar to BRCA1 and BRCA2 gene mutation carriers. However, currently there are no tools available to identify a molecular subclass of high-risk true precancerous lesions among the proliferative growths.

Because of very high incidence of cancer among atypical cohort, prophylactic endocrine treatments, tamoxifen, raloxifene, aromatase inhibitors (AI), and/or mastectomies are approved by FDA to prevent cancer similar to BRCA gene mutation carriers. However, both patients and their health care providers are faced with a decisional dilemma of undertaking/administering the preventative measures because of the lack of any molecular tools that can precisely stratify the approximately 20% of high-risk subjects who will go on to develop cancer. The uncertainty surrounding the risk puts enormous emotional burden on patients as well as physicians. Currently, treatments are administered without any molecular basis to willing patients not knowing their actual risk. As a result, patients who have low risk are overtreated and unnecessarily subjected to side effects of the above drugs such as pulmonary embolism, deep vein thrombosis, stroke, endometrial cancers, cataracts and vasomotor instability (tamoxifen; refs. 19–22), or musculoskeletal pain and bone loss (AIs), or undergoing unnecessary mastectomies. On the other hand, patients who have the real risk but choose not to receive above therapies to avoid side effects are not getting the benefit of prophylactic treatments. Because of lack of any tests to distinguish a subclass of true precancerous atypical lesions, high level of anxiety persists among women diagnosed with these types of growths. Molecular tests that can be applied to screen atypical lesions are urgently needed to precisely identify women who are highly likely to develop cancer so that they can be treated and prevented from developing cancer. In addition, currently there are no preventative treatment guidelines for subjects diagnosed with nonatypical proliferative lesions, although a majority of proliferative lesions diagnosed are the nonatypical type and large majority of cancers from proliferative group are from the nonatypical cohort.

To understand which proliferative lesions are true precancerous growths and associated with future cancer development, we have previously performed gene expression studies of atypical tissues and identified several cancer molecules that potentially indicate a proliferative lesion to be a true precursor (23). In this study, our goal was to investigate a panel of risk signature molecules that can be

applied for identifying high-risk subjects who will most likely develop cancers so that personalized treatment decisions can be made based on the patient's tumor biology. For this purpose, we have investigated a panel of selected risk signature, from the above-identified cancer molecules, in a total of 440 proliferative tissues with clinical follow-up information on cancer development status. We demonstrate here that the selected risk signature stratifies the high-risk group that subsequently developed cancer from the low-risk cancer-free group among atypical as well as nonatypical cohorts with 91% and 86% accuracy, respectively. We also report risk stratification into low-, intermediate-, and high-risk groups and cancer rates in each group at 5 years, 10 years, and 15 years after hyperplasia diagnosis. Finally, we report significant differences in terms of risk signature between atypical and nonatypical lesions that reflect in the cancer rates at various time periods.

Materials and Methods

Tissue samples

Archival formalin-fixed paraffin-embedded (FFPE) breast hyperplastic tissue samples for which clinical follow-up information available were included in this study. A total of 440 tissues of which 149 atypical and 291 nonatypical hyperplasias from subjects who had not received any preventative measures were included. Of the 440 tissues, 70 were from the previous studies (23) and 370 were the new independent set. The atypical group included was ductal (ADH) as well as lobular (ALH) types. Most of the atypical specimens also had one or more of the following nonatypical hyperplasias: UDH, papilloma, or sclerosing adenosis. The degree of proliferative ducts in the tissues ranged from a minimum of two to more than 10 ducts. Of the 440 specimens included, 162 (atypical $n = 74$ and nonatypical, $n = 88$) were from subjects who subsequently developed Ductal Carcinoma *In Situ* (DCIS, $n = 54$) or invasive ($n = 108$) cancer in either ipsilateral ($n = 124$) or contralateral ($n = 38$) breast after a minimum of 1 year and a maximum of 13 years (cancer group). The other 278 (atypical, $n = 75$ and nonatypical, $n = 203$) were from subjects who did not have prior breast cancer and did not develop for a minimum of 5 years and a maximum of 19 years (cancer-free group; details of specimens in the study are summarized in Table 1). The mean time period for developing cancer among atypical subjects was 4.3 years (SD 4.2) and for nonatypical subjects it was 7.1 years (SD 5.4). The ages of the subjects at diagnosis of

Table 1. Description of the proliferative tissues included in the study

Proliferative tissue samples	#	Benign histology		Time to cancer or cancer-free surveillance	Laterality of cancer		Cancer-developed	
		Atypical	Nonatypical		Ipsilateral	Contralateral	Invasive	DCIS
Cancer (developed) group	162	74	88	1 yr–13 yrs	124	38	108	54
Cancer-free group	278	75	203	Min. 5 yrs, Max. 19 yrs	NA	NA	NA	NA
Total	440	149	291	440	124	38	108	54

hyperplasia in the cancer group ranged from 28 years to 80 years (mean, 53.7, SD 12.1) and in the cancer-free group it ranged from 19 years to 72 years (mean, 46.7 and SD 12.4). The tissue samples with clinical surveillance information were obtained from the pathology divisions of the following institutions: UCLA Medical School (Los Angeles, CA), Leeds Hospitals (Leeds, United Kingdom), and Hartford Hospital (Hartford, CT). Institutional Review Boards of the respective institutions approved the archival tissue sample retrieval procedures for the study. To ascertain that all the sections cut from each block had the hyperplasia, the first and the last section cut from each paraffin block were stained with haematoxylin and eosin and examined for histology and only those containing the desired tissues were used.

Assay of risk signature molecules in FFPE tissues

A panel of four cancer markers, MMP-1 (Matrix Metallo Proteinase 1), CEACAM6 (Carcino Embryonic Antigen Cell Adhesion Molecule 6), HYAL1 (Hyaluronoglucosaminidase 1), and HEC1 (Highly Expressed in Cancer Rich in Leucine Heptad Repeats 1) that were previously identified by gene expression studies (23) and an established breast cancer marker, ESR1 (Estrogen Receptor 1; ref. 24) were selected for this study. We assayed the markers by IHC method, which visually ascertains the presence of relatively limited number of proliferating ducts that could vary from block to block and section to section of a specimen. We performed IHC assays of markers as described previously (23, 25–28) blinded to the knowledge on cancer development status of the specimen. The positive controls were breast cancer tissues, which express the above markers. The negative controls were breast hyperplastic tissues stained without primary or secondary antibody. All the slides were graded by board-certified and licensed pathologists blinded to the knowledge on the development or nondevelopment of cancer. The slides were graded on the basis of the stain intensity and the proportion of proliferating epithelial cells staining. The tissue was considered positive if greater than 10% of proliferative cells showed staining. The staining intensity was graded in comparison with the negative control (score, 0) and positive control (cancer; grade, 4). The final grading values ranged from 0.0 to 4.0. To ascertain that the above marker proteins were not deteriorated with storage, we evaluated two markers in proliferative tissues in the cancer group that were stored for 12 years and found no deterioration of the markers.

Statistical methods

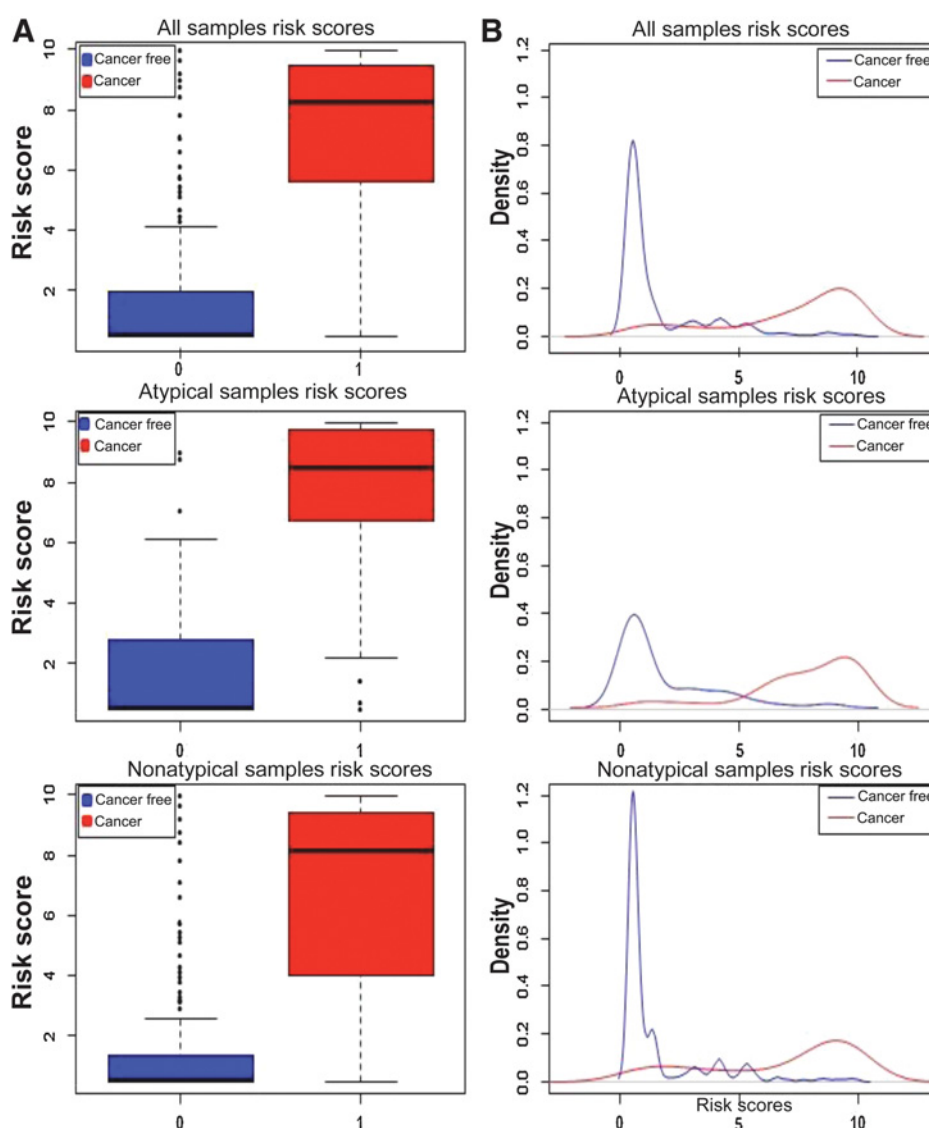
Our goal, in this study, was to stratify the high-risk subjects who developed cancer (cancer group) from the low-risk (cancer-free) group based on the composite levels of multiple cancer markers expressed in the proliferative tissues. The IHC grades of the multiple markers assayed indicate their expression levels. To determine an optimum method to combine the marker levels and obtain a com-

posite value, we performed the following: we first tested classical linear logistic regression method by evaluating its performance using typical measures such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and the area under the ROC curve (AUC), as well as log-likelihood ratio. To evaluate possible interactions and nonlinearity of the marker levels, we also compared pair-wise interaction methods and nonlinear methods. We found piece-wise linear logistic regression to be the optimum method for combining the multiple marker levels and obtaining the composite value between 0–10, which we call a "cancer risk score."

The risk scores of 440 samples were displayed as box plots and density plots to examine their distributions in the cancer group and the cancer-free groups of samples. The Odds Ratio (OR) was calculated on the basis of the Odd at the mean risk score and the Odd at mean risk score plus one SD. We used the calculated cancer risk scores and the corresponding cancer status of the samples to obtain the ROC statistics such as sensitivity, specificity, PPV, NPV, and the accuracy of cancer prediction (29). The ROC curves were drawn to evaluate the performance of the risk signature for cancer prediction. The commonly used 10-fold cross-validation method was applied for verifying the cancer prediction values, sensitivity, specificity, PPV, NPV, and accuracy. The R package caret is used for the stratified cross-validation, where each of the 10 subsets has about the same sample size of 10% of the total data sample size and has about the same case-control ratio. For each of the 10 subsets, say S_i , combining the rest of the 9 subsets as the training dataset to determine the accuracy of cancer prediction values. We have investigated a correlation between cancer risk scores and cancer rates at 5 years, 10 years, and 15 years, and beyond using Kaplan–Meier survival analysis (30). All analyses were performed for UDH and ADH samples separately, and for all tissues combined.

Results

We have developed here a clinically applicable risk-scoring method that assesses cancer risk in terms of a score from 0 to 10 by taking into account the expression levels of four markers. We present results to show that the cancer group of samples can be stratified from the cancer-free group based on the risk scores. We demonstrate that the risk scores of samples from patients who did not develop cancer after a minimum of 5 years and maximum of 19 years ranged from 0 to 0.5 and for samples from patients who subsequently developed cancer after a minimum of 1 year, the scores ranged from 5.5 to 9.0. In addition, we present results to show that molecular signature-based risk scores predict cancer development with very high sensitivity, specificity, and accuracy. Finally, data are presented on correlation between risk scores and cancer rates at 5 years, 10 years, and 15 years for atypical and nonatypical types of samples. The details of the results are presented below.

**Figure 1.**

Box plots and density plots of composite risk scores obtained from the expression levels of MMP-1, CEACAM6, HYAL1, and HEC1 in proliferative hyperplastic tissues are shown. **A**, Box plots of risk scores (0–10) for the cancer-free (0, blue) and cancer group of samples (1, red). The plots show clear separation of cancer group from the cancer-free group in atypical (middle) and nonatypical samples (bottom) as well as all samples together (top). **B**, Density plots of risk scores (0–10) for atypical, nonatypical, and the combination of all samples are shown. The peak risk scores of cancer-free group (blue) and cancer group (red) were 1 and 9.5 for atypical (middle), 0.5 and 8.0 for nonatypical (bottom), and 0.5 and 9.0 for all samples combined (top), respectively.

The molecular signature–based risk scores stratify the cancer group from cancer-free group

After assaying multiple markers in 440 samples, we applied piece-wise linear logistic regression method to combine the expression levels of all the markers assayed to obtain composite risk scores. We found that the four markers, MMP-1, CEACAM6, HYAL1, and HEC1 were highly associated with cancer prediction by linear as well as piece-wise logistic regression methods. However, ESR1, did not significantly contribute to the predictive values by any method tested, therefore, was not included in the final risk score calculations.

After obtaining the risk scores of all 440 samples, we examined the score distributions in the cancer group and the cancer-free group of samples by displaying in box plots and density plots. The box plots (Fig. 1A) show that the cancer group of samples (red) is clearly segregated from the cancer-free group (blue). The segregation could

be observed in atypical cohort, nonatypical cohort, or all the samples combined (Fig. 1A, middle, bottom, and top plots, respectively). The density plots (Fig. 1B) show that in the cancer-free group (blue) the risk scores peaked around 0.5, while in the cancer group (red) the scores peaked around 9.0 (Fig. 1B, top plot). When we separately plotted the atypical and nonatypical samples, the peak scores were 9.5 and 8.0, respectively, among cancer groups but minor differences in the cancer-free groups (Fig. 1B, middle, and bottom plots, respectively). These results clearly establish that risk signature segregates the cancer group from the cancer-free group of tissues. When we analyzed the risk scores between atypical and nonatypical tissues for any differences using *t* test, we found significantly higher values in the cancer group of atypical tissues ($P = 0.02$) than the nonatypical tissues but did not find any significant differences in the cancer-free groups.

Table 2. Risk score–based ROC statistics

Tissue samples	Sensitivity	Specificity	PPV	NPV	Accuracy
Atypical	89% ± 7%	93% ± 6%	93% ± 6%	90% ± 7%	91% ± 5%
Nonatypical	68% ± 10%	94% ± 3%	83% ± 9%	87% ± 4%	86% ± 4%
Atypical + Nonatypical	78% ± 6%	93% ± 3%	87% ± 5%	88% ± 4%	88% ± 3%

The four marker-based risk scores predict cancer with very high accuracy—ROC statistics

We used the calculated cancer risk scores of 440 samples and the corresponding cancer status of the samples to obtain the ROC statistics such as sensitivity, specificity, PPV, NPV, and the accuracy of cancer prediction. The results are presented in Table 2. As seen in the Table, overall, the four marker signature predicts cancer with very high sensitivity, specificity, PPV, NPV, and accuracy (78% ± 6%, 93% ± 3%, 87% ± 5%, 88% ± 4%, and 88% ± 3%, respectively). When we compared the above values obtained from the 70 samples of the previous studies (23) with the 370 samples of the new independent set, we got similar results. Cross-verification using 10-fold cross-validation method gave similar values for sensitivity, specificity, PPV, NPV, and accuracy (81% ± 5%, 92% ± 3%, 86% ± 5%, 89% ± 3%, and 88% ± 3%, respectively). When the atypical samples were separately evaluated, the accuracy was higher, 91% ± 5%, than the nonatypical group, 86% ± 4% (Table 2). Although these values are slightly lower for nonatypical group of tissues, they are highly significant in predicting cancer development. We have further analyzed the risk score data by displaying the ROC curves. The results displayed (Fig. 2) show that the performance of the four marker risk signature is very high in predicting cancer. The AUCs for atypical and nonatypical types were 0.95 and 0.9, respectively (Fig. 2, middle and bottom curves) and for all samples combined the AUC was 0.92 (Fig. 2, top curve). The values of above 0.90 AUC show very high predictive value of risk signature molecules tested.

Correlation between risk scores and cancer rates at 5 years, 10 years, and 15 years—Kaplan–Meier survival analysis

We have investigated a correlation between the risk scores and cancer rates at various time durations after hyperplasia diagnosis using Kaplan–Meier survival analysis. To draw the survival curves, we have divided the samples into three groups based on the risk scores: low risk (Score ≤ 0.5), intermediate-risk (Score ≤ 5.4), and high-risk (Score > 5.4) groups. We chose score 0.5 as low-risk group because the samples that had a score of 0.5 did not express any of the four molecules. The samples that had risk score above 5.4 were grouped as high risk because 5.4 was the cutoff for the optimal performance measurements such as sensitivity, specificity, PPV, and NPV as in Table 2. The samples that had scores of more than 0.5 and less or equal to 5.4 were grouped as intermediate-risk group. The survival plots drawn for

the above three risk groups are shown in Fig. 3. The top panel shows the survival curves for all samples (atypical and nonatypical samples combined). In the first 5 years, 53% of subjects in the high-risk group, 12% in the intermediate-risk group, and 1% in the low-risk group developed cancer. When we divided the data into

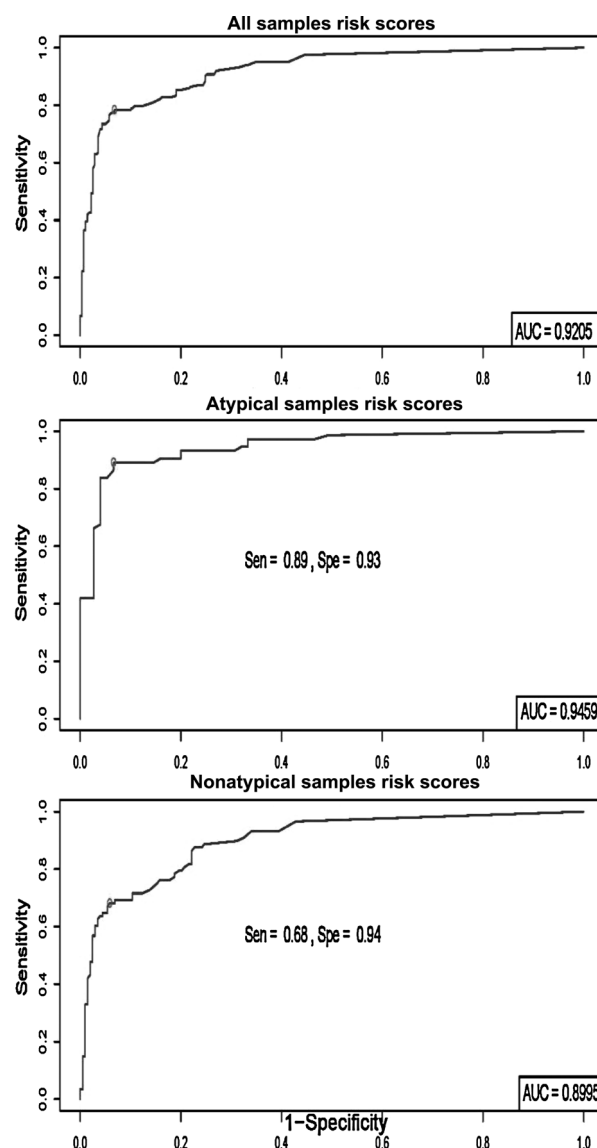


Figure 2. ROC curves for atypical, nonatypical, and combination of all samples are shown. The AUCs for atypical (middle), nonatypical (bottom), and combination of all samples (top) were 0.95, 0.9, and 0.92, respectively.

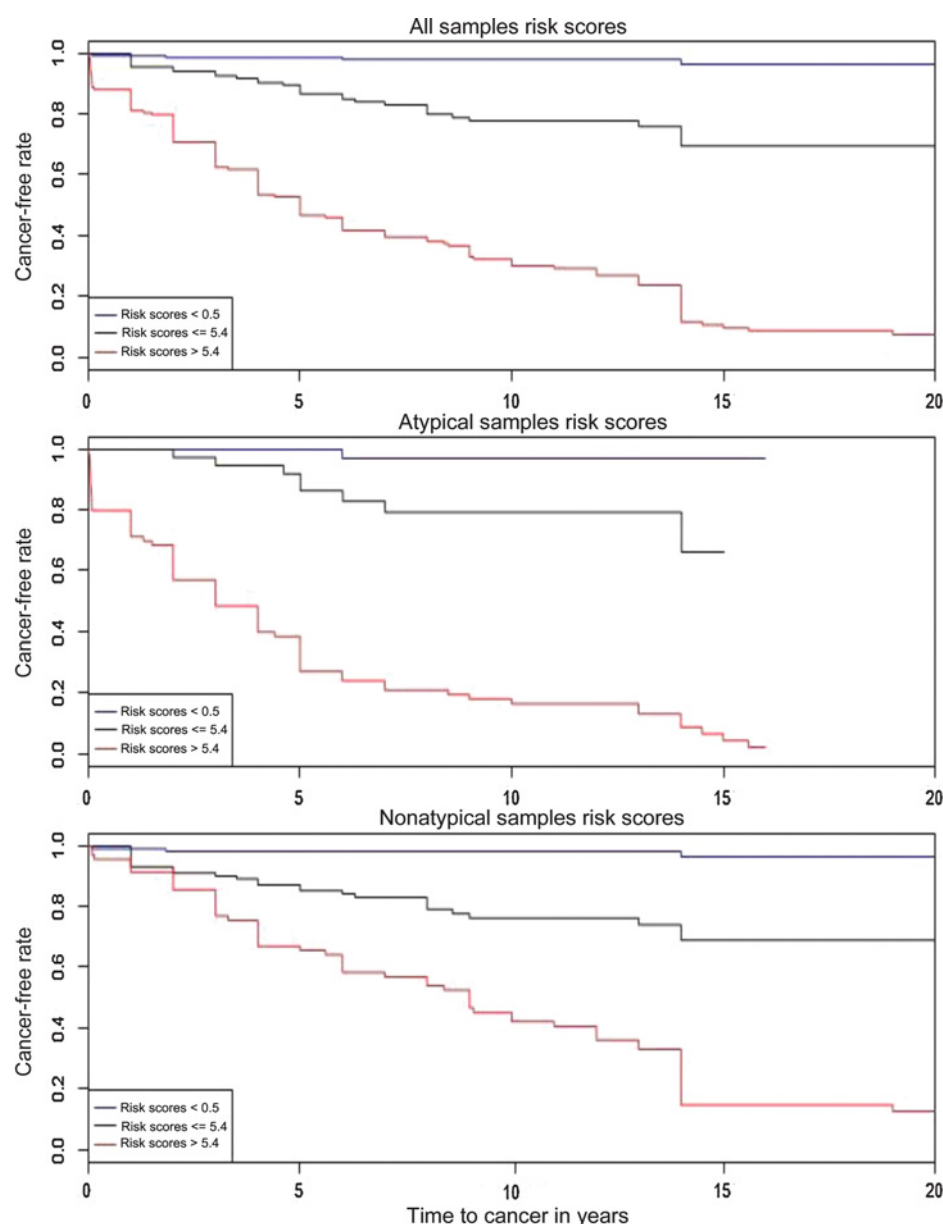


Figure 3. Kaplan-Meier survival curves for the low-risk, intermediate-risk, and high-risk groups. Survival curves for atypical (middle), nonatypical (bottom), and combination of all samples (top) are shown. The survival curves are distinct for the low-risk (score ≤ 0.5 ; blue), intermediate-risk (Score ≤ 5.4 ; black), and high-risk (score > 5.4 ; red) groups. The cancer rates for the three risk categories are shown in Table 3.

atypical and nonatypical groups, 73% of atypical subjects in high-risk group developed cancer in the first 5 years, while in the nonatypical group the cancer rates were significantly lower (34%; Fig. 3, middle and bottom panels, respectively). However, in the low- and intermediate-risk groups, there were no significant differences in the cancer rates between atypical and nonatypical types. We have summarized the data on cancer rates in the first 5 years, 10 years, 15 years, and beyond 15 years in Table 3. As seen in Fig. 3 and Table 3, there is a direct correlation between risk scores and cancer rates. Overall, among low-risk group with scores from 0.0 to 0.5, the cancer-free rates were at least 97% for up to 15 years and beyond.

MMP-1, CEACAM6, HYAL1, and HEC1 are colocalized along with ESR1 in proliferating ducts

We have examined individual proliferating ducts for the localization patterns of above proteins to get insights into their role in initiating carcinogenesis in the milk ducts where over 95% of breast cancers arise. Examples of colocalization patterns of the above proteins in ADH, UIDH, and papilloma types of proliferative ducts are shown in Fig. 4, panels 1–3, respectively. As seen in Fig. 4, we observed colocalization of all the expressed molecules in the proliferating cells of the same ducts. Interestingly, the above four cancer promoters were colocalized along with ESR1 as shown in Fig. 4, panels 1–3, although its expression levels were not significantly different in the cancer

Table 3. Cancer rates among low, intermediate-, and high-risk groups at various time periods

Sample details and risk levels			Percent cancer rates in 5 years, 10 years, 15 years, and beyond			
Sample types and risk score	Sample size	Risk level	Up to 5 years	Up to 10 years	Up to 15 years	Beyond 15 years
Atypical and risk scores < 0.5	39	Low	0	3	3	3
Atypical and risk scores < 5.4	40	Intermediate	12	18	20	20
Atypical and risk scores > 5.4	70	High	73	83	91	93
Nonatypical and risk scores < 0.5	119	Low	2	2	3	3
Nonatypical and risk scores < 5.4	102	Intermediate	15	22	25	25
Nonatypical and risk scores > 5.4	70	High	34	57	81	83
Atypical + Nonatypical and risk scores < 0.5	163	Low	1	2	3	3
Atypical + Nonatypical and risk scores < 5.4	137	Intermediate	12	20	23	23
Atypical + Nonatypical and risk scores > 5.4	140	High	53	69	85	86

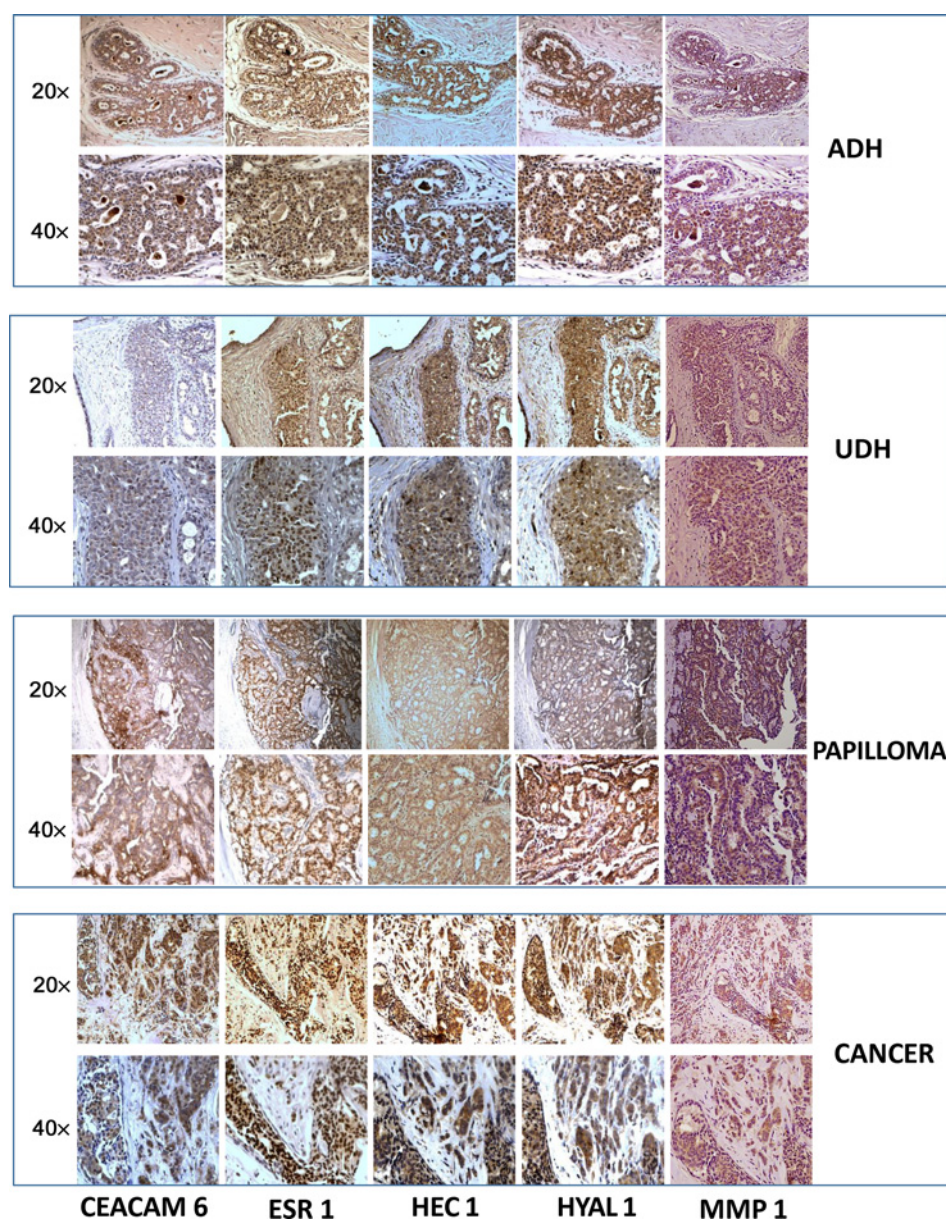
group of tissues compared with cancer-free group by IHC, consistent with our previous gene expression data (23). For comparative purposes, expression levels of above markers in breast cancer tissues are also shown in Fig. 4, panel 4.

Discussion

In this study, we have addressed the dilemma faced by the oncologists in treating the subjects who are diagnosed

Figure 4.

Colocalization of ESR1, MMP-1, CEACAM6, HYAL1, and HEC1 in proliferative hyperplastic breast tissues by IHC. Examples of colocalization of the five oncoproteins is shown in atypical ductal hyperplasia (ADH), UDH, and papilloma types of proliferating breast tissues (top three panels, respectively). For comparative purpose, colocalization of these molecules in breast cancer tissues is also shown (bottom). The micrographs shown are at 20× and 40× magnification.



with breast atypical, as well as nonatypical hyperplasias due to lack of any molecular means of risk stratification. We have recognized the need for identifying the high-risk subjects who would benefit from preventative endocrine treatments at the same time spare the low-risk subjects from unnecessary treatments and their side effects or mastectomies. We have presented results here to demonstrate that a four marker risk signature, MMP1, CEACAM6, HYAL1, and HEC1, could be applied to stratify the high-risk atypical as well as nonatypical subjects who will benefit from the prophylactic measures.

The risk signature molecules, MMP-1, CEACAM6, HYAL1, and HEC1, applied here were discovered in our previous gene expression studies and were highly elevated (35.5, 37, 13.3, and 8 times, respectively) in cancer group of hyperplastic tissues compared with cancer-free group (23). The selected molecules were also shown to promote the progression of several types of cancers by various mechanisms (31–47). In addition, three of the individual markers (MMP-1, CEACAM6, and HYAL1) were shown to predict cancer with varying accuracies in pilot studies (23, 26, 27). In this study, we have developed a method to combine the levels of all the expressed markers by piece-wise linear logistic regression for obtaining a composite value between 0 and 10, which we call a "cancer risk score." The results presented in Fig. 1 show that the risk scores of cancer group of samples were distinct from the cancer-free group demonstrating differences in the biology of the two types of tissues. The density plots (Fig. 1B) also show that the risk scores varied widely in cancer group of samples suggesting heterogeneity in marker levels and biology.

The ROC statistics data (Table 2) and ROC curves data (Fig. 2) established that the four marker signature-based risk scores predict cancer development with very high sensitivity, specificity, PPV, NPV, and accuracy. The accuracy is slightly higher for atypical tissues (91%) compared with nonatypical tissues (86%). The OR for our risk signature is 8.4. On the basis of the OR of 8.4, the sample size of 102 has the power of more than 99% with significance level of 0.0001. Therefore, the sample size of 440 applied in this study was sufficient for stratifying the cancer group from the cancer-free group of samples. We compared risk scores of samples from subjects who developed cancer in ipsilateral ($n = 124$) versus contralateral ($n = 38$) breast by ANOVA and found no significant differences ($P = 0.09453$). Similarly, no significant association was found between risk scores and age at hyperplasia diagnosis, menopausal status, number of atypical foci, or any characteristics of the subsequent cancer developed such as Estrogen Receptor (ER) status, nodes, grade, stage, DCIS, or invasive cancer. These observations demonstrate that the risk signature-based cancer prediction is independent of laterality, age at hyperplasia diagnosis, menopausal status, or any of the characteristics of the future cancer developed.

The survival curves that were drawn by categorizing the subjects based on risk scores into low-Risk (score ≤ 0.5), intermediate-risk (score ≤ 5.4), and high-risk (score > 5.4) groups are distinct for the three groups (Fig. 3) demonstrating that the risk scoring could be applied for identifying high-risk subjects who will subsequently develop cancer. The data presented in Fig. 3 and Table 3 also demonstrate a direct correlation between risk scores and cancer rates. Overall, atypical samples in the high-risk group had significantly higher risk scores ($P = 0.02$ by t test) than nonatypical samples, which reflected in higher cancer rates and shorter cancer-free survival. The lower cancer rates in the first 5 years among the nonatypical high-risk subjects suggest that these are relatively slow-growing preatypical stage lesions. Kaplan–Meier survival data presented above establish that the cut-off risk scores we have set here for risk stratification into low, intermediate, and high risk could be applied to assess a patient's risk level and treatment decisions.

Colocalization studies of risk signature molecules revealed that all the expressed markers in a tissue were colocalized in every positive duct as shown in Fig. 4. Interestingly, three of the four molecules, MMP-1, CEACAM6, and HYAL1, are known to be involved in disrupting cell–cell and/or cell–ECM interactions. MMP-1 and HYAL1 are matrix-cleaving enzymes and are known to promote cancer progression by opening up spaces in the matrix and facilitating the migration of cancer cells (32, 48). CEACAM6 was shown to regulate cell adhesion and over-expression is known to disrupt cell–cell interactions and tissue architecture (34). It is possible that the above three cancer-promoting molecules along with another established cancer promoter, HEC1, may contribute to carcinogenesis.

In summary, data are presented in establishing MMP-1-, CEACAM6-, HYAL1-, and HEC1-based risk scoring and risk stratification strategy to classify a hyperplasia subject as low risk, intermediate risk, or high risk for developing future breast cancer. In addition, data on cancer rates for a given risk group at 5 years, 10 years, and beyond are presented for atypical as well as nonatypical patients. The risk scoring model described here offers for the first time a tumor biology-based personalized risk scoring method for making informed treatment decisions. Prophylactic treatment decisions can now be made based on the cancer risk score and consultation between a patient and her health care professional. For example, a standard screening for women with low risk, increased surveillance for subjects of intermediate risk, and risk-reducing endocrine therapies and/or prophylactic mastectomy for high-risk group similar to BRCA gene mutation carriers. Currently, the above treatment guidelines are approved for only atypical subjects but not for nonatypical subjects. Our data presented here on nonatypical samples warrants the need for developing treatment guidelines similar to subjects with atypical

growths. Finally, molecular risk scoring of hyperplastic tissues in conjunction with the already available prophylactic endocrine therapies and/or prophylactic mastectomy measures could prevent about 20% to 25% of sporadic breast cancers.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: I. Poola

Development of methodology: I. Poola

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): I. Poola, Q. Yue, J.W. Gillespie, P.S. Sullivan, J. Rao, A.M. Shaaban, E.R. Sauter, A.J. Ricci, J. Aguilar-Jakthong

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): I. Poola, Q. Yue

Writing, review, and/or revision of the manuscript: I. Poola

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): I. Poola, Q. Yue

Study supervision: I. Poola

Others (obtained funding from NCI, NIH, and NSF; principal investigator on the grants that funded the project for the above manuscript.): I. Poola

Acknowledgments

This work was supported, in part, by grants from the NCI, NIH (CA173919 and CA206774) and the National Science Foundation (IIP-1314287) awarded to I. Poola.

Received January 25, 2019; revised February 27, 2019; accepted May 13, 2019; published first June 25, 2019.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics 2019. *CA Cancer J Clin* 2019;69:7–34.
2. MedTech Insight report: US breast disease detection and diagnostic technologies (report no. A400). Newport Beach, CA: MedTech Insight; 2002.
3. Tavassoli FA, Norris HJ. A comparison of the results of long term follow up for atypical intraductal hyperplasia and intraductal hyperplasia of the breast. *Cancer* 1990;65:518–29.
4. Wellings SR, Jenson HM, Marcum RG. An atlas of sub-gross pathology of the human breast with special reference to possible precancerous lesions. *J Natl Cancer Inst* 1975;55:231–273.
5. Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast cancer. *N Engl J Med* 1985;312:146–51.
6. Dupont WD, Parl FF, Hartman WH. Breast cancer risk associated with proliferative breast disease and atypical hyperplasia. *Cancer* 1993;71:1258–65.
7. London SJ, Conolly JL, Schnitt SJ, Colditz GA. A prospective study of benign breast disease and the risk of breast cancer. *JAMA* 1992; 267:941–4.
8. Page DL, Dupont WD. Association indicators (histologic and cytologic) of increased breast cancer risk. *Breast Cancer Res Treat* 1993;28:157–166.
9. Foote FW, Stewart FW. Comparative studies of cancerous versus noncancerous breasts. *Ann Surg* 1945;121:197–222.
10. Ryan JA, Cody CV. Intraductal epithelial proliferation in the human breast- a comparative study. *Cancer J Surg* 1962;5:2–8.
11. Black MM, Barclay TH, Cutler SV, Hankey BF, Asire AJ. Association of atypical characteristics of benign breast lesions with subsequent risk of breast cancer. *Cancer* 1972;29:338–43.
12. Karpus CM, Leis HP, Oppenheim A, Mersheimer WL. Relationship of fibrocystic disease to carcinoma of the breast. *Ann Surg* 1995;162:1–8.
13. Allerd DC, Mohsin SK, Fuqua SAW. Histological and biological evolution of human premalignant breast disease. *Endocr Relat Cancer* 2001;8:47–61.
14. Krishnamurthy S, Sneige N. Molecular and biological markers of pre-malignant lesions of human breast. *Adv Anat Pathol* 2002;9: 185–97.
15. Guray M, Sahin AA. Benign breast diseases: classification, diagnosis, and management. *Oncologist* 2006;11:435–49.
16. Hartman L, Sellers TA, Frost MH, Lingle WL, Frhmom SC, Ghosh K, et al. Benign breast disease and the risk of breast cancer. *N Engl J Med* 2005;353:229–37.
17. Coopey SB, Mazolla E, Buckley JM, John S. The role of chemoprevention in modifying the risk of breast cancer in women with atypical breast lesions. *Breast Cancer Res Treat* 2012;136: 627–33.
18. Worsham MJ, Abrams J, Raju U, Kapke A, Lu M, Cheng J, et al. Breast cancer incidence in a cohort of women with benign breast disease from a multiethnic, primary health care population. *Breast J* 2007;13:116–21.
19. Tan-Chiu E, Wang J, Costantino JP, Paik S, Butch C, Wickerham L, et al. Effects of tamoxifen on benign breast disease in women at high risk for breast cancer. *J Natl Cancer Inst* 2003;95:302–7.
20. Veronesi U, Maisonneuve P, Rotmensz N, Costa A, Sacchini V, Travaglini R, et al. Italian Randomized trial among women with hysterectomy: tamoxifen and hormone-dependent breast cancer in high risk women. *J Natl Cancer Inst* 2003;95:160–5.
21. Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for prevention of cancer: report of the NSABP. *J Natl Cancer Inst* 1998;90:1371–88.
22. Cuzick J, Forbes J, Edwards R, Baum M, Cawthorn S, Coates A, et al. First results from the international breast cancer intervention study (IBIS-I): a randomized prevention trial. *Lancet* 2002;360: 817–24.
23. Poola I, DeWitty RL, Marshallack JJ, Bhatnagar R, Abraham J, et al. Identification of MMP-1 as a putative breast cancer predictive molecular marker by global gene expression analysis. *Nat Med* 2005;11:481–3.
24. Summer S, Fuqua SAW. Estrogen receptor and breast cancer. *Semin Cancer Biol* 2001;11:39–352.
25. Poola I, Fuqua SAW, DeWitty RL, Abraham J, Marshallack JJ, Liu A. ER negative breast cancer tissues express significant levels of estrogen receptor beta- 5: potential molecular targets for chemoprevention of ER -negative breast cancer. *Clin Cancer Res* 2005; 11:7579–85.
26. Poola I, Shokrani B, Bhatnagar R, DeWitty R, Yue Q, Bonney G. Expression of carcinoembryonic antigen cell adhesion molecule 6 (CEACAM6) in atypical ductal hyperplastic tissues is associated with development of invasive breast cancer. *Clin Cancer Res* 2006;12:4773–83.
27. Poola I, Abraham J, Marshallack JJ, Yue Q, Lokeshwar V, Bonney G, et al. Molecular risk assessment for breast cancer development in patients with ductal hyperplasias. *Clin Cancer Res* 2008;14: 1274–80.

Poola et al.

28. Poola I, Abraham J, Marshallack J, Yue Q, Fu S, Viswanath L, et al. Molecular constitution of breast but not other reproductive tissues is rich in growth promoting molecules: a possible link to highest incidence of tumor growths. *FEBS Lett* 2009;83: 3069–75.
29. Tobias S, Oliver S, Niko B, Thomas L. ROCR: visualizing classifier performance in R. *Bioinformatics* 2005;21:3940–1.
30. Therneau T. A package for survival analysis in S. version 2.38, (2015). Available from: <https://CRAN.R-project.org/package=survival>. Survival analysis.
31. Magita T, Sato E, Saito K, Mizoi T, Shiiba K, Matsuno S, et al. Differing expression of MMPs-1 and 9 and urokinase receptor between diffuse and intestinal-type gastric carcinoma. *Int J Cancer* 1999;84:74–9.
32. Balduyck M, Zerimech F, Lemaire R, Lemaire R, Hemon B, Grard G, et al. Specific expression of matrix metalloproteinases 1, 3, 9 and 13 associated with invasiveness of breast cancer cells *in vitro*. *Clin Exp Metastasis* 2000;18:171–8.
33. Hojilla CV, Mohammed FF, Kokha R. Matrix metalloproteinases and their tissue inhibitors direct cell fate during cancer development. *British J Cancer* 2003;89:1817–21.
34. Ilantzis C, DeMarte L, Sreaton RA, Stanners CP. Deregulated expression of human tumor marker CEA family member. CEACAM6 disrupts tissue architecture and blocks colonocyte differentiation. *Neoplasia* 2002;4:151–63.
35. Jantschkeff P, Terracciano L, Lowy A, Glatz-Krieger K, Grunert F, Micheel B, et al. Expression of CEACAM6 in resectable colorectal cancer: a factor of independent prognostic significance. *J Clin Oncol* 2003;21:3638–46.
36. Dexbury MS, Ito H, Zinner MJ, Ashley SW, Whang EE. CEACAM6 gene silencing impairs resistance and *in vivo* metastatic ability of pancreatic adenocarcinoma cells. *Oncogene* 2004;23: 465–73.
37. Blumenthal RD, Hansen HJ, Goldenberg DM. Inhibition of adhesion, invasion, and metastasis by antibodies targeting CEACAM6 (NCA 90) and CEACAM5. *Cancer Res* 2005;65: 8809–17.
38. Lokeshwar VB, Cerwinka W, Lokeshwar BL. HYAL1 hyaluronidase: a molecular determinant of bladder tumor growth and invasion. *Cancer Res* 2005;65:2243–50.
39. Lokeshwar VB, Cerwinka WH, Itoyama T, Lokeshwar BL. HYAL1 hyaluronidase in prostate cancer: tumor promoter and suppressor. *Cancer Res* 2005;65:7782–9.
40. Posey JT, Soloway MS, Ekici S, Sofer M, Civantos F, Soloway MS, et al. Evaluation of the prognostic potential of hyaluronic acid and hyaluronidase (HYAL1) for prostate cancer. *Cancer Res* 2003; 63:2638–44.
41. Bertrand P, Girard N, Duval C, Anjou JD, Chauzy C, Menard JP, et al. Increased hyaluronidase levels in breast tumor metastases. *Int J Cancer* 1996;73:327–31.
42. Franzmann EJ, Schroeder GL, Goodwin WJ, Weed DT, Fisher P, Lokeshwar VB. Expression of tumor markers hyaluronic acid and hyaluronidase (HYAL1) in head and neck tumors. *Int J Cancer* 2003;106:438–45.
43. Liu D, Pearlman E, Diacnou E, Guo K, Mori H, Haqqi T, et al. Expression of hyaluronidase by tumor cells induces angiogenesis *in vivo*. *Proc Natl Acad Sci U S A* 1996;93:7832–7.
44. Huang LY, Chang CC, Lee YS, Chang JM, Huang JJ, Chuang SH, et al. Activity of a novel Hec1-targeted anticancer compound against breast cancer cell lines *in vitro* and *in vivo*. *Mol Cancer Ther* 2014;13:1419–30.
45. Tang NH, Toda T. Mapping the Ndc80 loop in cancer: a possible link between Ndc80/Hec1 overproduction and cancer formation. *Bioessays* 2015;37:248–56.
46. Qu Y, Li J, Cai Q, Liu B. Hec1/Ndc80 is overexpressed in human gastric cancer and regulates cell growth. *J Gastroenterol* 2014;49: 408–18.
47. Huang LY, Chang CC, Lee YS, Huang JJ, Chuang SH, Chang JM, et al. Inhibition of Hec1 as a novel approach for treatment of primary liver cancer. *Cancer Chemother Pharmacol* 2014;74: 511–20.
48. Knudson W. Tumor associated hyaluronan providing an extracellular matrix that facilitates invasion. *Am J Pathol* 1996;148: 1721–6.

Cancer Prevention Research

Breast Hyperplasias, Risk Signature, and Breast Cancer

Indira Poola, Qingqi Yue, John W. Gillespie, et al.

Cancer Prev Res 2019;12:471-480. Published OnlineFirst June 25, 2019.

Updated version Access the most recent version of this article at:
doi:[10.1158/1940-6207.CAPR-19-0051](https://doi.org/10.1158/1940-6207.CAPR-19-0051)

Cited articles This article cites 46 articles, 12 of which you can access for free at:
<http://cancerpreventionresearch.aacrjournals.org/content/12/7/471.full#ref-list-1>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://cancerpreventionresearch.aacrjournals.org/content/12/7/471>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.