

## Research Article

See related article by Caan et al., p. 725

# Intrinsic Subtypes from PAM50 Gene Expression Assay in a Population-Based Breast Cancer Cohort: Differences by Age, Race, and Tumor Characteristics

Carol Sweeney<sup>1,2</sup>, Philip S. Bernard<sup>2,3</sup>, Rachel E. Factor<sup>2,3</sup>, Marilyn L. Kwan<sup>4</sup>, Laurel A. Habel<sup>4</sup>, Charles P. Quesenberry Jr<sup>4</sup>, Kaylynn Shakespear<sup>1</sup>, Erin K. Weltzien<sup>4</sup>, Inge J. Stijleman<sup>2</sup>, Carole A. Davis<sup>2</sup>, Mark T.W. Ebbert<sup>3</sup>, Adrienne Castillo<sup>4</sup>, Lawrence H. Kushi<sup>4</sup>, and Bette J. Caan<sup>4</sup>

## Abstract

**Background:** Data are lacking to describe gene expression–based breast cancer intrinsic subtype patterns for population-based patient groups.

**Methods:** We studied a diverse cohort of women with breast cancer from the Life After Cancer Epidemiology and Pathways studies. RNA was extracted from 1 mm punches from fixed tumor tissue. Quantitative reverse-transcriptase PCR was conducted for the 50 genes that comprise the PAM50 intrinsic subtype classifier.

**Results:** In a subcohort of 1,319 women, the overall subtype distribution based on PAM50 was 53.1% luminal A, 20.5% luminal B, 13.0% HER2-enriched, 9.8% basal-like, and 3.6% normal-like. Among low-risk endocrine-positive tumors (i.e., estrogen and progesterone receptor positive by immunohistochemistry, HER2 negative, and low histologic grade), only 76.5% were categorized as luminal A by PAM50. Continuous-scale luminal A, luminal B, HER2-enriched, and normal-like scores from PAM50 were mutually positively correlated. Basal-like score was inversely correlated with other subtypes. The proportion with non-luminal A subtype decreased with older age at diagnosis,  $P_{\text{Trend}} < 0.0001$ . Compared with non-Hispanic Whites, African American women were more likely to have basal-like tumors, age-adjusted OR = 4.4 [95% confidence intervals (CI), 2.3–8.4], whereas Asian and Pacific Islander women had reduced odds of basal-like subtype, OR = 0.5 (95% CI, 0.3–0.9).

**Conclusions:** Our data indicate that over 50% of breast cancers treated in the community have luminal A subtype. Gene expression–based classification shifted some tumors categorized as low risk by surrogate clinicopathologic criteria to higher-risk subtypes.

**Impact:** Subtyping in a population-based cohort revealed distinct profiles by age and race. *Cancer Epidemiol Biomarkers Prev*; 23(5); 714–24. ©2014 AACR.

## Introduction

Gene expression profiling has revealed intrinsic subtypes of breast cancer that improve prognostication (1–8) and prediction of response to therapy (7, 9, 10) compared with categories defined by clinicopathologic characteristics. The luminal A subtype has best prognosis and is, in most populations examined, the most frequent subtype. Defining subtypes of breast tumors for participants in breast cancer epidemiologic studies is of interest for several reasons: the distribution of subtypes by host characteristics, or associations between sub-

types and risk factors, may shed light on etiologic pathways; survival differences for subtype groups should be defined in population-based studies; the influence of modifiable risk factors on recurrence and survival may vary by subtype. Much of the existing data on gene expression–based breast cancer intrinsic subtypes have been derived from clinical trial populations or other selected populations, for example estrogen receptor positive (ER<sup>+</sup>) cases only (8, 9), cases diagnosed at ages younger than 55 years (3), or patients with node-negative or low histologic grade disease (10, 11). It is not known how well the subtype distributions estimated from these studies describe the population across all ages, across a range of clinical characteristics, and across racial and ethnic groups.

Microarray gene expression assay is the gold standard for intrinsic subtyping, but because fresh-frozen tissue is required, this technology is usually not feasible for large research study populations. Instead, strategies for assigning subtypes based on clinicopathologic variables, that is ER, progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), and proliferation

**Authors' Affiliations:** <sup>1</sup>Division of Epidemiology, Department of Internal Medicine; <sup>2</sup>Huntsman Cancer Institute, University of Utah; <sup>3</sup>The Associated Regional and University Pathologist Institute for Clinical and Experimental Pathology, Salt Lake City, Utah; and <sup>4</sup>Division of Research, Kaiser Permanente Northern California, Oakland, California

**Corresponding Author:** Carol Sweeney, Division of Epidemiology, University of Utah, 295 Chipeta Way, Salt Lake City, UT 84108. Phone: 801-581-5865; Fax: 801-581-3623; E-mail: carol.sweeney@hsc.utah.edu

doi: 10.1158/1055-9965.EPI-13-1023

©2014 American Association for Cancer Research.

markers or tumor grade, have been applied in clinical and epidemiologic studies (12–22). Limitations of the clinicopathologic subtyping approach are that staining and scoring of immunohistochemical (IHC) markers is subject to variability and that subtypes classified using clinicopathologic variables may not align with intrinsic subtypes classified by gene expression–based assays (10, 23). Subtype classifiers based on quantitative reverse-transcriptase PCR (qRT-PCR) of gene products from fixed tissue are a third strategy for intrinsic subtyping. qRT-PCR classifiers are more feasible for large studies than microarray techniques and more quantitative than IHC (6, 24). Although the clinical utility of qRT-PCR classifiers is an active area of research (25–27), examples of research applying these classifiers in epidemiology are very limited (28).

In this study, we applied the PAM50 assay, a well-characterized qRT-PCR intrinsic subtyping classifier that measures expression of 50 genes selected as characteristic of 5 breast cancer intrinsic subtypes (6, 10, 11, 29), to archived primary tumor tissue from participants in the Life After Cancer Epidemiology (LACE) and Pathways breast cancer cohorts. We describe the overall distribution of subtypes luminal A, luminal B, HER2-enriched, basal-like, and normal-like, and subtype variation in relation to clinicopathologic categories and patient characteristics.

## Materials and Methods

### Study population

The study population consisted of breast cancer survivors from the LACE and Pathways cohorts. LACE participants were 18 to 79 years old when diagnosed with early-stage breast cancer (stage I with tumor size 1 cm or greater, stage II, or stage IIIA) from 1996 to 2000. Additional eligibility criteria at the time of LACE recruitment in 2000 to 2002 were that the woman was within 39 months of diagnosis (mean time from diagnosis to enrollment = 23 months, 61% between 12 and 24 months), had completed any chemotherapy or radiation therapy, was free of breast cancer recurrence, and had no other cancer diagnosis within 5 years. LACE study methods and baseline characteristics of participants have been described (30). For the intrinsic subtyping study, we included LACE participants recruited from the Kaiser Permanente Northern California (KPNC) Cancer Registry and the Utah Cancer Registry. The Pathways Study enrolled women diagnosed with invasive breast cancer from 2005 to 2013 in KPNC with no previous diagnosis of other invasive cancer, and at least 21 years of age. Most women were approached for enrollment within 2 months of diagnosis (mean time from diagnosis to enrollment = 1.8 months, maximum 7.2 months). Details of the Pathways study methods have been described (31). For the PAM50 intrinsic subtyping study, we included Pathways women diagnosed in 2006 to 2008. Participants provided informed consent under protocols approved by the Kaiser Permanente Northern California Institutional Review Board (IRB) and the University of Utah IRB.

Self-reported race and ethnicity were obtained at study enrollment using mailed (LACE) or in-person (Pathways) questionnaires. Characteristics at the time of diagnosis, including age, disease stage (American Joint Commission on Cancer, AJCC), tumor size, node status, histologic grade, ER status, PR status, and HER2 overexpression or amplification in the primary tumor, were abstracted from tumor registry data and medical records review.

A stratified random sample of women meeting eligibility criteria was selected for intrinsic subtyping. This sample forms a subcohort that will be used for future case-cohort analysis of outcomes of breast cancer recurrence and death. ER, PR, and HER2 status based on IHC (and/or FISH for HER2) defined the strata used for sampling, with an 18% random sample selected among cases of the common breast cancer phenotype that is positive for ER or PR expression and negative for HER2 (and has low risk of recurrence) and a 100% sample of tumors that were ER<sup>-</sup> and PR<sup>-</sup> or HER2<sup>+</sup>.

### Tissue samples

For cohort members selected for the subcohort, we contacted the hospital where surgery for resection of the primary tumor was performed, or the institution's pathology storage facility, to obtain formalin-fixed, paraffin-embedded (FFPE) tissue blocks from that procedure and corresponding slides and pathology reports. Slides were reviewed by one pathologist (R.E. Factor). Final eligibility for the PAM50 subtyping study was determined based on the review of slides and pathology reports. If the appearance of the primary tumor tissue in the slides indicated to the pathologist that neoadjuvant therapy had been used before resection, or if the area of invasive tumor was observed to be smaller than 0.5 cm in diameter, the case was classified as ineligible. If the pathology report indicated bilateral disease, the case was classified as ineligible. For eligible cases, the pathologist marked an area of representative tumor tissue on a slide. Tissue punches 1 mm in diameter were obtained from the area of the FFPE tissue block corresponding to the marked slide. Two punches per case, or one punch if the primary tumor was less than 0.7 cm in diameter, were placed in plastic tubes labeled with a sample identifying number.

### Clinical tissue markers

For about 6% of cohort members, ER, PR, and/or HER2 clinical results were missing, precluding selection for the PAM50 assay according to the stratified random sampling scheme. We attempted to obtain tissue blocks for determination of these markers. For many of these cases with missing data, no blocks could be obtained. If blocks were obtained, sections were cut from the block and slides were made to assess expression of these markers by IHC staining. Equivocal HER2 results (IHC 2+) were subjected to FISH for determination of HER2 amplification; those with a HER2/CEP17 ratio greater than 2.2 were classified as HER2<sup>+</sup>. For  $n = 11$  cases determined by this process to be ER and PR negative or HER2<sup>+</sup>, placing the case in a

category with 100% sampling frequency, the case was selected for PAM50.

### Gene expression assay

The tissue punch was deparaffinized and tissue was digested for RNA extraction as described previously (11). qRT-PCR was conducted for the 50 target genes that comprise the PAM50 (6). Details of qRT-PCR methods have been provided elsewhere (11, 32). Quality control included a negative control (no template) and a positive control (reference DNA template) for each gene in each plate, and qRT-PCR of 5 housekeeping genes from each tissue sample. Laboratory personnel were blinded to clinical information and received only a sample identification number to track the sample. Each batch of tissue punches sent to the laboratory for PAM50 assay included a mix of clinicopathologic types.

### Subtype classification variables

We defined 2 subtype classification variables for use in data analysis. The first was clinicopathologic subtype, based on clinical characteristics, and the second was subtype based on the PAM50 gene expression assay. The characteristics used to define clinicopathologic subtype were IHC results for ER and PR, IHC and/or FISH results for HER2, and tumor grade. Following St. Gallen conference recommendations (33) with a recent modification (34), we defined a low-risk, endocrine positive, that is surrogate luminal A, clinicopathologic subtype category as PR<sup>+</sup>, HER2<sup>-</sup>, and well or moderately differentiated. A higher-risk endocrine positive or clinicopathologic surrogate luminal B category included tumors with ER<sup>+</sup> or PR<sup>+</sup> and any of PR<sup>-</sup>, HER2<sup>+</sup>, or tumor grade of poorly or undifferentiated. A third clinicopathologic category included HER2<sup>+</sup>, endocrine-negative tumors, that is ER<sup>-</sup>, PR<sup>-</sup>, and HER2<sup>+</sup>, and a fourth group was "triple negative" or ER<sup>-</sup>, PR<sup>-</sup>, and HER2<sup>-</sup>. To classify intrinsic subtypes from the gene expression data, we applied centroid-based algorithms to the calibrated log-expression ratio for the 50 genes (6). This process generates, for each sample, 5 continuous-scale normalized subtype scores representing degree of correlation of gene expression with that of archetypal luminal A, luminal B, HER2-enriched, basal-like, and normal-like breast tumors. Tumors were classified as the subtype with the highest normalized subtype score.

### Data analysis

All analyses incorporated the original sampling weights and the stratified sampling design for unbiased estimation of population parameters and valid estimates of standard errors (with "svy" commands in Stata software, StataCorp.). Frequency distributions of intrinsic subtypes were estimated for the overall cohort and within subgroups defined by clinicopathologic categories and by race and ethnicity. We calculated the  $\kappa$  statistic to describe agreement between the clinicopathologic and intrinsic subtypes. Tumors categorized by PAM50 as normal-like,

a subtype with no corresponding clinicopathologic category, were excluded from the  $\kappa$  calculation. Sensitivity and specificity were calculated for subtype classification, dichotomized as yes or no, treating clinicopathologic classification from IHC as the "test" and PAM50 subtype as the gold standard.

In addition to analyses that treated each intrinsic subtype as a discrete category, we examined the distributions of the 5 continuous-scale normalized subtype scores from the PAM50 assay. Kernel-smoothed probability density functions were constructed, showing the distribution of the scores for each intrinsic subtype within the groups of subjects defined by subtype categories. Pairwise associations between subtype scores were assessed with Pearson correlation coefficients.

We described associations of participant and tumor characteristics with intrinsic subtypes by fitting a multinomial logistic regression model. This is similar to the case-case analysis approach widely used for dichotomous tumor characteristics (35), extended to the 5 subtype categories via the multinomial model. Treating the most prevalent subtype, luminal A, as the base outcome, we estimated age-adjusted OR and 95% confidence intervals (CI) for each of the non-luminal A subtypes. Association between continuous-scale subtype scores and participant characteristics were also assessed in linear regression models.

### Results

The combined LACE and Pathways cohorts include women diagnosed with breast cancer at ages 25 to 91 years, with the largest proportion diagnosed between 50 and 59 years of age (Table 1). In addition to the non-Hispanic White majority of the participants, the cohort has representation of African American, Hispanic, Asian or Pacific Islander, and American Indian or Alaska Native women. The cohort includes all stages at diagnosis, with over 90% diagnosed at AJCC stages I or II. Of 1,622 subjects selected for the intrinsic subtyping by PAM50, 103 were determined to be ineligible on pathology review because of insufficient area of invasive tumor in available blocks, 14 because of neoadjuvant therapy, and 6 because of bilateral disease. Fifty-six study participants did not consent for the tumor tissue assay, and blocks could not be obtained for another 105 participants. Laboratory assay failure prevented PAM50 subtyping for only 19 cases. After accounting for the stratified sampling scheme, the subcohort of 1,319 participants with PAM50 results was similar to the parent cohort with respect to age, race and ethnicity, and clinical characteristics (Table 1), with a mean age at diagnosis of 59 years.

The distribution of PAM50-based intrinsic subtypes in the cohort was estimated to be 53.1% luminal A, 20.5% luminal B, 13.0% HER2-enriched, 9.8% basal-like, and 3.6% normal-like (Table 2). In an analysis restricted to the 697 cases who were enrolled in the study within 1 year after diagnosis, the distribution was 52.8% luminal A, 21.6% luminal B, 12.7% HER2-enriched, and 8.5% basal-like and 4.3% normal-like ( $P = 0.41$  for comparison to those enrolled

**Table 1.** Characteristics at time of diagnosis for LACE and Pathways breast cancer cohort participants and a subcohort subtyped by PAM50 gene expression–based classifier

	LACE study <i>n</i> = 2,135	Pathways study <i>n</i> = 2,172	Combined cohort <i>n</i> = 4,307	Subcohort <sup>a</sup> <i>n</i> = 1,319	
	%	%	%	Raw %	Weighted <sup>b</sup> %
Age at diagnosis (years)					
<40	4.6	4.9	4.7	6.9	5.0
40–49	18.6	17.5	18.0	20.2	17.0
50–59	30.4	29.9	30.1	31.5	30.8
60–69	28.4	28.0	28.2	25.3	27.6
70–79	18.0	15.0	16.5	13.9	16.6
80+	0	4.7	2.4	2.2	3.0
Race and ethnicity					
White, non-Hispanic	79.6	67.6	73.7	69.2	72.6
Black or African American	5.4	6.9	6.2	8.7	6.9
Hispanic	6.5	11.5	9.0	9.3	8.8
Asian or Pacific Islander	6.5	10.7	8.6	9.9	9.2
American Indian or Alaska Native	1.0	2.1	1.6	1.7	1.6
Other or missing	1.0	0.9	1.0	1.1	0.8
AJCC stage					
I	46.6	50.7	48.7	44.3	49.8
II	50.6	33.9	42.2	47.6	43.6
III	2.9	9.3	6.1	7.2	6.2
IV	0.0	1.7	0.9	0.9	0.5
Missing	0	4.3	2.2	0	0
Tumor size					
≤2 cm	65.0	64.5	64.8	59.7	66.1
>2 cm	35.0	31.5	33.2	40.3	33.9
Tumor markers from IHC/FISH <sup>c</sup>					
ERPR <sup>+</sup> HER2 <sup>-</sup>	67.5	72.7	70.1	33.0	74.8
ERPR <sup>+</sup> HER2 <sup>+</sup>	11.1	9.3	10.2	26.0	9.8
ERPR <sup>-</sup> HER2 <sup>-</sup>	10.9	12.5	11.7	30.7	11.5
ERPR <sup>-</sup> HER2 <sup>+</sup>	3.7	4.5	4.1	10.3	3.9
Missing	6.7	0.9	3.8	0	0

<sup>a</sup>Subjects with PAM50 subtyping results. The subcohort sample was chosen with the following selection probabilities by clinical subtype: ERPR<sup>+</sup> HER2<sup>-</sup>, 18%; ERPR<sup>+</sup> HER2<sup>+</sup>, 100%; triple negative, 100%; ERPR<sup>-</sup> HER2<sup>+</sup>, 100%.

<sup>b</sup>Estimated distribution of characteristics in the subcohort after accounting for stratified sampling.

<sup>c</sup>Categories based on tumor ER and PR expression, as determined by clinical IHC, and for HER2 overexpression by IHC and/or FISH. ERPR<sup>+</sup> includes subjects with ER<sup>+</sup> and/or PR<sup>+</sup>; ERPR<sup>-</sup> includes subjects with both ER<sup>-</sup> and PR<sup>-</sup>.

one or more years after diagnosis). Among the tumors classified as low-risk endocrine positive or surrogate luminal A by clinicopathologic factors (Table 2), 76.5% were classified as luminal A by PAM50, and a minimal percentage were classified as basal-like, 0.3%. Within the group classified as endocrine positive, higher-risk or surrogate luminal B by clinicopathologic factors, only 36.2% were classified as luminal B by PAM50, with appreciable proportions classified as luminal A or HER2-enriched. For the cohort members with clinicopathologic HER2 positive, endocrine negative tumors, 75.7% were HER2-enriched by PAM50. Among participants with triple negative tumors, 69.9% were basal-like by intrinsic subtype. The

$\kappa$  statistic for agreement between clinicopathologic and intrinsic subtype categories across 4 groups was 0.49. The sum of sensitivity and specificity for the clinicopathologic categories was highest for triple-negative as a surrogate classifier of basal-like subtype (Table 2). When we considered the distribution of intrinsic subtypes within racial and ethnic groups (Fig. 1), the proportions with basal-like subtype among African American women and Asian women, 30.4% (95% CI, 21.0–41.7) and 5.0% (95% CI, 3.0–8.2), respectively, had CIs that excluded the proportion classified as basal-like among non-Hispanic Whites.

There was substantial overlap in the distribution of continuous-scale luminal A scores between tumors

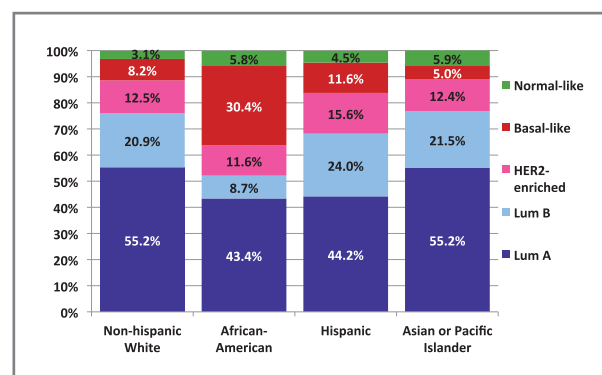


**Table 2.** Distribution of breast cancer intrinsic subtypes from PAM50 gene expression assay in the population-based LACE and Pathways studies, in a subcohort overall and by clinicopathologic categories

PAM50 subtype	Clinicopathologic category <sup>a</sup>															
	Subcohort n = 1,319			Low-risk endocrine + n = 298			High-risk endocrine + n = 480			HER2 <sup>+</sup> , endocrine -- n = 138			Triple negative n = 405			
	%	95% CI	Sens/Spec <sup>b</sup>	%	95% CI	Sens/Spec <sup>b</sup>	%	95% CI	Sens/Spec <sup>b</sup>	%	95% CI	Sens/Spec <sup>b</sup>	%	95% CI	Sens/Spec <sup>b</sup>	
Luminal A	53.1	49.8–56.5	76.5	71.4–81.0	0.74/0.74	40.5	34.5–46.7	2.2	0.7–6.7	3.0	1.7–5.2	3.0	1.7–5.2	3.0	1.7–5.2	
Luminal B	20.5	17.7–23.6	14.8	11.2–19.3	0.59/0.73	36.2	30.6–42.3	7.4	4.0–13.2	4.7	3.0–7.2	4.7	3.0–7.2	4.7	3.0–7.2	
HER2E	13.0	11.3–15.0	3.7	2.1–6.6	0.23/0.99	17.5	13.7–22.0	75.7	67.8–82.2	0.23/0.99	20.3	16.6–24.5	20.3	16.6–24.5	20.3	16.6–24.5
Basal-like	9.8	8.8–10.8	0.3	0.0–2.0	14.7	1.3–6.2	14.7	9.7–21.7	0.0	0.0	69.9	65.2–74.2	69.9	65.2–74.2	69.9	65.2–74.2
Normal-like	3.6	2.5–5.4	4.7	2.8–7.8	0.0	1.4–5.9	0.0	0.0	0.0	0.0	2.2	1.2–4.2	2.2	1.2–4.2	2.2	1.2–4.2

<sup>a</sup>Clinicopathologic categories (columns), defined in methods, are based on ER and PR expression, as determined by IHC, HER2 overexpression by IHC, and/or FISH, and tumor grade.

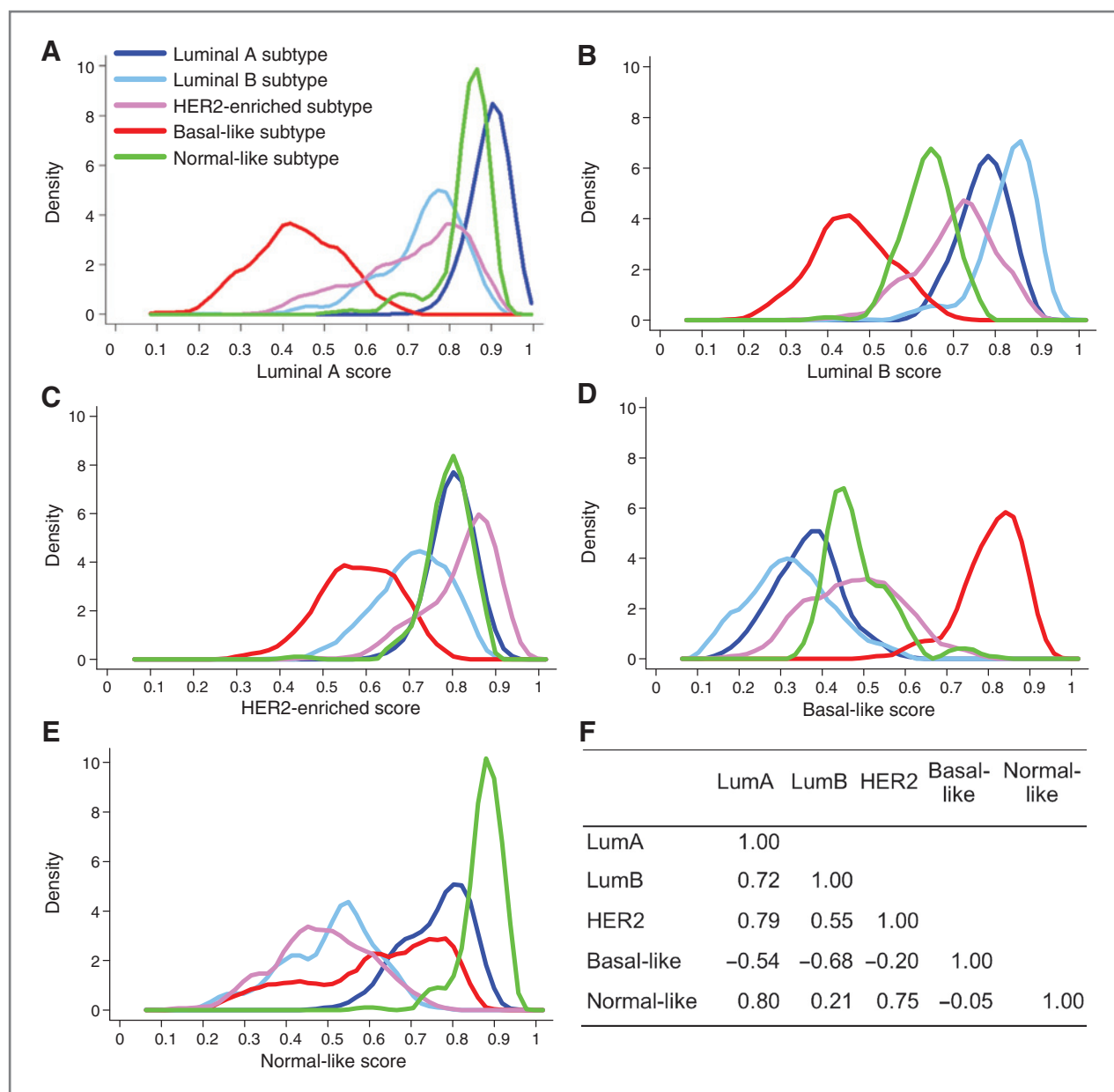
<sup>b</sup>Sensitivity and specificity, treating clinicopathologic category as test and the most similar PAM50 subtype classification as the gold standard.



**Figure 1.** Distribution of breast cancer intrinsic subtypes from PAM50 assay in a population-based cohort by race and ethnicity, LACE and Pathways studies.

classified as luminal A and normal-like (Fig. 2A). For luminal B and HER2-enriched scores (Fig. 2B and C), overlap was observed between tumors classified in different subtypes. The distribution of the basal-like score among tumors classified by PAM50 as basal-like was distinct, exhibiting almost no overlap with the distributions for other subtypes (Fig. 2D). Tumors with subtypes other than normal-like exhibited a broad range of normal-like scores (Fig. 2E). The continuous-scale PAM50 subtype scores for luminal A, luminal B, HER2-enriched, and normal-like were positively correlated with each other, whereas the basal-like score was inversely correlated with each of the other subtype scores (Fig. 2F).

Age at diagnosis was strongly related to intrinsic subtype. Using luminal A as the comparison group, there were significant trends of decreasing odds of having a breast tumor of luminal B ( $P = 0.001$ ), HER2-enriched ( $P = 0.05$ ), or basal-like ( $P < 0.0001$ ) subtype with older age at diagnosis (Table 3). In age-adjusted models, African American women with breast cancer had significantly higher odds of having a basal-like tumor versus a luminal A tumor, OR = 4.39 (95% CI, 2.29–8.29), compared with non-Hispanic Whites. Hispanic women were somewhat more likely to have any of the non-luminal A subtypes, that is luminal B, HER2-enriched, or basal-like, in univariate analysis, but the differences were attenuated after age adjustment and did not reach statistical significance. Asian and Pacific Islander women with breast cancer had significantly reduced relative odds of having basal-like versus luminal A tumors, OR = 0.48 (95% CI, 0.25–0.92). Results for American Indian and Alaska Native women were somewhat suggestive of higher prevalence of non-luminal A subtypes, but confidence intervals were large because of small numbers in this racial group in the cohort. In linear regression models, luminal A subtype score was positively associated with age at diagnosis ( $P = 0.01$ ) and inversely associated with African American race ( $P < 0.001$ ) among women with non-LumA subtype. Basal-like subtype score was inversely associated with age at diagnosis ( $P = 0.005$ ) among women with non-basal-like subtype.



**Figure 2.** Distributions of normalized, continuous-scale scores for breast cancer intrinsic subtypes estimated from PAM50 assay results, LACE and Pathways cohort. A–E, each panel shows, for one continuous-scale intrinsic subtype score obtained from PAM50, the kernel-smoothed probability density function of that score within PAM50 intrinsic subtype category groups of tumors, indicated by color. F, correlation coefficients for the subtype scores.

Several clinical characteristics at diagnosis were associated with intrinsic subtype (Table 3). Higher histologic tumor grade was associated with increased prevalence of luminal B, HER2-enriched, and basal-like subtypes ( $P_{\text{Trend}} < 0.0001$  for each). Luminal B, HER2-enriched, and basal-like subtypes were relatively more prevalent among tumors with higher stage at diagnosis ( $P_{\text{Trend}} < 0.0001$ ,  $< 0.0001$ , and  $0.06$ , respectively). When we considered 2 of the component characteristics of stage at diagnosis, tumor size, and positive lymph nodes, tumors greater than 2 cm in diameter were more likely than smaller tumors to have

luminal B, HER2-enriched, and basal-like subtypes. Increasing number of positive nodes was associated with luminal B ( $P_{\text{Trend}} = 0.01$ ) and HER2-enriched subtypes ( $P_{\text{Trend}} = 0.007$ ), but not with basal-like subtype. The odds of having the normal-like subtype was inversely associated with stage at diagnosis ( $P = 0.04$ ) but was not significantly associated with age, race, or any other clinical characteristic (data not shown).

We conducted a subanalysis among cohort members whose clinicopathologic tumor features placed them in the low-risk endocrine positive or surrogate luminal A

**Table 3.** Demographic and clinical characteristic associations with breast cancer intrinsic subtype from PAM50 assay, LACE and Pathways cohort: ORs for non-luminal A subtypes versus luminal A

	Total n	Intrinsic subtype <sup>a</sup> from PAM50											
		LumA		LumB		LumB vs. LumA		HER2E		HER2E vs. LumA		Basal vs. LumA	
		% <sup>b</sup>	% <sup>b</sup>	OR <sup>c</sup>	95% CI	% <sup>b</sup>	% <sup>b</sup>	OR <sup>c</sup>	95% CI	% <sup>b</sup>	% <sup>b</sup>	OR <sup>c</sup>	95% CI
Age at diagnosis													
<40	91	2.9	8.1	2.48	0.98–6.29	5.5	1.72	0.68–4.34	10.8	3.05	1.40–6.65		
40–49	266	14.2	20.7	1.27	0.72–2.27	16.3	1.03	0.59–1.79	26.5	1.51	0.93–2.46		
50–59	415	28.3	32.4	1	Reference	31.6	1	Reference	35.0	1	Reference		
60–69	334	31.9	20.8	0.57	0.33–0.97	28.8	0.81	0.49–1.32	17.5	0.44	0.27–0.72		
70–79	184	19.1	15.5	0.71	0.39–1.30	15.2	0.71	0.39–1.29	9.5	0.41	0.22–0.76		
80+	29	3.6	2.4	0.57	0.17–1.95	2.6	0.65	0.19–2.30	0.6	0.13	0.028–0.61		
<i>P</i> <sub>Trend</sub>					0.001			0.05			<0.0001		
Race/ethnicity													
White	913	75.5	74.3	1	Reference	70.1	1	Reference	60.9	1	Reference		
Black or African American	115	5.6	2.9	0.51	0.19–1.33	6.1	1.15	0.55–2.38	21.3	4.38	2.29–8.39		
Hispanic	123	7.3	10.4	1.29	0.63–2.62	10.6	1.46	0.71–3.01	10.5	1.47	0.82–2.63		
Asian or Pacific Islander	131	9.5	9.6	0.89	0.46–1.73	9.0	0.93	0.55–1.54	4.7	0.48	0.25–0.92		
American Indian or Alaska Native	22	1.0	2.5	2.9	0.45–10.3	2.9	2.74	0.61–12.4	2.0	1.95	0.52–7.33		
Other	37	1.0	0.3	0.30	0.05–1.76	1.3	1.47	0.37–5.87	0.6	0.77	0.13–4.80		
Grade													
Well-differentiated	162	33.2	9.9	1	Reference	8.8	1	Reference	1.2	1	Reference		
Moderately differentiated	453	46.9	47.7	3.19	1.65–6.15	39.2	3.06	1.40–6.69	11.7	6.31	2.11–18.8		
Poor/undifferentiated	630	9.4	38.6	12.80	6.14–26.5	47.4	18.50	8.30–41.3	82.5	217.40	74.6–633.8		
Unknown	72	10.5	3.8	1.23	0.41–3.72	4.6	1.68	0.53–5.30	4.7	13.20	3.44–50.9		
<i>P</i> trend					<0.0001			<0.0001			<0.0001		
Stage													
I	562	57.2	36.1	1	Reference	38.5	1	Reference	44.0	1	Reference		
II	607	37.8	52.0	2.04	1.34–3.12	55.0	2.09	1.40–3.11	49.6	1.47	1.01–2.12		
III	90	5.0	10.8	3.05	1.45–6.40	5.7	1.60	0.81–3.17	5.2	1.05	0.51–2.15		
IV	12	0.1	1.1	24.50	2.23–269.6	0.9	20.70	2.24–192.2	1.2	14.90	1.59–139.1		
<i>P</i> <sub>Trend</sub>					<0.0001			<0.0001			0.06		
Tumor size													
≤2 cm	797	74.9	54.3	1	Reference	55.8	1	Reference	56.0	1	Reference		
>2 cm	522	25.1	45.7	2.36	1.56–3.56	44.2	2.29	1.53–3.41	44.0	2.06	1.43–2.96		
Positive lymph nodes													
0	832	70.2	59.0	1	Reference	58.0	1	Reference	72.9	1	Reference		
1–3	321	22.8	25.3	1.18	0.73–1.90	29.9	1.50	0.97–2.32	19.8	0.68	0.43–1.07		
4+	139	6.9	15.6	2.48	1.30–4.71	12.2	2.05	1.11–3.76	7.4	0.87	0.48–1.60		
<i>P</i> <sub>Trend</sub>					0.01			0.007			0.19		

<sup>a</sup>LumA, luminal A; LumB, luminal B; HER2E, HER2-enriched; Basal, basal-like. The normal-like subtype accounted for fewer than 5% of subtype results and is not shown.

<sup>b</sup>Estimated distribution of characteristics within the subtype category after correction for sampling weights.

<sup>c</sup>Prevalence ORs for cross-sectional association between subject or clinical characteristic at diagnosis and non-luminal A subtype.

category, to evaluate predictors of being classified as worse-prognosis non-luminal A (i.e., luminal B, HER2-enriched, or basal-like) subtypes by the PAM50 assay (data not shown). Among tumors that were low-risk, endocrine-positive from clinicopathologic data, younger age at diagnosis ( $P_{\text{Trend}} = 0.007$ ) showed a significant trend of being classified by the PAM50 assay as a worse-prognosis subtype.

## Discussion

Existence of the breast cancer intrinsic subtypes luminal A, luminal B, HER2-enriched, and basal-like has been validated across multiple laboratory methods (36, 37) and multiple study populations, including confirmation that the subtypes initially characterized in European and North American populations are also present in breast cancers in Asian women (38). However, this study is the first, to our knowledge, to describe intrinsic subtypes based on a gene expression assay in a U.S. population-based study, and in a study population that includes representation of multiple racial and ethnic groups.

Our estimated prevalence of intrinsic subtype luminal A, 53.1% of all breast cancers, is higher than the fraction estimated by most prior studies, whereas the fraction with normal-like subtype in our cohort, 3.6%, is lower. Others have suggested that categorization as normal-like may be an artifact of contamination of tumor RNA with RNA from normal breast cells (11, 39, 40). The low percentage of normal-like results in this study suggests that the strategy of obtaining a tissue punch from the FFPE tumor block, guided by the pathologist's marked slide, was effective in minimizing normal tissue contamination. The inverse association between normal-like subtype and tumor stage in this study may be an indication of contamination of the tumor punch with normal tissue when smaller tumors are sampled.

Population selection likely contributes to the high prevalence of luminal A in this study compared with other datasets. Specifically, selection of clinical trial populations with restrictions such as lymph-node positive cases (10, 11) or young cases (3) for intrinsic subtype studies, and the general tendency for clinical trial participants to be younger than the overall cancer population (41, 42), would result in skewing study populations toward more aggressive, non-luminal A subtypes. A study population more similar to ours was based on a population-based cancer registry in Sweden with no age or stage restrictions and a mean age at diagnosis of 58 years (23). In that study, if normal-like results were excluded, the distribution of subtypes (from microarray analysis of fresh tissue) would be 44% luminal A, 19% luminal B, 16% HER2-enriched, and 21% basal-like.

Although this study used population-based recruitment, cohort exclusions and enrollment processes may have resulted in a sample set that deviates somewhat from the true population distribution. Specifically, exclusions of tumors with less than 0.5 cm of invasive tumor or neoadjuvant chemotherapy, exclusion of late-stage cases

and of cases older than 79 years from the LACE cohort, and missed enrollment for the LACE cohort of potentially eligible cases who died within about 2 years after diagnosis, introduce a limitation for describing the population distribution of subtypes. The small tumors and older women excluded would be more likely to be luminal A subtype, whereas advanced-stage cases and those who died within a short time after diagnosis probably include more poor prognosis, non-luminal A subtypes, so that there is no uniform direction of bias expected from these exclusions. In analysis of women enrolled within the first year after diagnosis, the distribution of subtypes did not differ significantly, and the proportion of luminal A remained 53%.

Although the literature on intrinsic subtyping initially emphasized classification and distinguishing between subtypes, and discrepancies of subtype classification for the same tumor in cross-platform validation studies are seen as problematic for clinical application (39), within-subtype molecular heterogeneity is increasingly recognized (32, 36, 43). The overlapping probability density functions for luminal A, luminal B, and HER2-enriched from this large, population-based study support the interpretation that, within these subtypes, a tumor of one subtype can have varying degrees of expression of genes characteristic of other subtypes (32). In a previous study, use of subtype scores resulted in better-fitting prognostic models compared with subtype category (6). Researchers should further consider applying not just subtype classification but also subtype scores, and individual gene expression data obtained from assays such as PAM50, when investigating tumor characteristics in relation to prognostic and etiologic characteristics.

Results from this study confirm the Swedish population-based study's report of a significant association of older age at diagnosis with luminal A subtype (23). A recent report described changes in gene expression patterns in normal breast tissue with age (44). These results taken together suggest that future research should consider the links between normal tissue context and tumor subtype.

Our observation of a significant, 4-fold higher prevalence of PAM50 basal-like subtype among African Americans is consistent with the high proportion of clinicopathologic basal-like phenotypes among African Americans first described in a population-based study in North Carolina (12). We observed a reduced odds of having a basal-like tumor, versus luminal A, among Asian and Pacific Islander women. A high proportion of clinicopathologic luminal A subtypes and/or low prevalence of triple negative breast cancers has been reported among Asian populations in the United States (45) and in Asia (22, 46).

Among participants whose clinical tumor markers placed them in the clinicopathologic endocrine positive, low risk or surrogate luminal A category, about 1 in 5 were shifted to higher-risk luminal B, HER2-enriched, or basal-like subtypes based on subtyping by PAM50 assay. This shift in subtype is clinically significant because a St. Gallen



conference (33) recommended that subtypes be used for early breast cancers to distinguish tumors for which adjuvant chemotherapy is indicated. In our analysis of cases with clinicopathologic endocrine positive, low-risk surrogate luminal A tumors that shifted to other subtypes based on PAM50 results, women with younger age at diagnosis were more likely to shift to the higher-risk subtypes.

Study resources were not sufficient to support uniform IHC assessment of ER, PR, and HER2 across the full cohort. Instead, we relied on clinical determination from the time of diagnosis. Changes in scoring of ER, PR, or HER2 over time are a limitation in the data, as is our reliance on only "positive" or "negative" results, given that details on percent of cells stained or HER2 score were not consistently found in the medical records review.

Studies of qRT-PCR-based intrinsic classifiers indicate promise for prognosis, but reviewers have suggested that more evidence is needed to support application of these classifiers in clinical decision making (25). The potential application of PCR-based classifiers for epidemiologic studies is intriguing. To our knowledge, this is the first study to apply such a tool in a large, population-based study. We were successful in obtaining PAM50 subtype results for a high proportion of cases in a population-based study using FFPE tumor blocks, many of which had been stored 10 to 12 years. Thus, PAM50 assay is a feasible classifier for epidemiologic studies. Subtype differences by age and by race were distinct. Epidemiologic studies have reported that the influence of breast cancer risk factors varies by clinicopathologic subtype (13, 15, 16, 21, 22, 47–50). PCR-based intrinsic classifiers may prove a more sensitive tool for describing etiologic heterogeneity.

In conclusion, the PAM50 qRT-PCR assay proved feasible in an epidemiologic study. Luminal A was the intrinsic subtype for the majority of a population-based study cohort, but intrinsic subtyping shifted some tumors categorized as low risk by clinicopathologic criteria to higher-risk subtypes. Gene expression-based subtyping

in a population-based cohort supported the concept of molecular heterogeneity within subtypes and revealed distinct intrinsic subtype profiles by age and race.

#### Disclosure of Potential Conflicts of Interest

P.S. Bernard is a partner in University Genomics. P.S. Bernard has ownership interest (including patents) in University Genomics and Bio-classifier LLC. L.A. Habel has a commercial research grant from bioTheragnostics. No potential conflicts of interest were disclosed by the other authors.

#### Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

#### Authors' Contributions

**Conception and design:** C. Sweeney, L.A. Habel, L.H. Kushi, B.J. Caan  
**Development of methodology:** C. Sweeney, P.S. Bernard, L.A. Habel, C.P. Quesenberry Jr, E.K. Weltzien, B.J. Caan

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** C. Sweeney, P.S. Bernard, R.E. Factor, K. Shakespear, E.K. Weltzien, I.J. Stijleman, A. Castillo, L.H. Kushi, B.J. Caan

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** C. Sweeney, P.S. Bernard, M.L. Kwan, L.A. Habel, C.P. Quesenberry Jr, E.K. Weltzien, M.T.W. Ebbert, L.H. Kushi, B.J. Caan

**Writing, review, and/or revision of the manuscript:** C. Sweeney, P.S. Bernard, R.E. Factor, M.L. Kwan, L.A. Habel, C.P. Quesenberry Jr, E.K. Weltzien, M.T.W. Ebbert, A. Castillo, L.H. Kushi, B.J. Caan

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** K. Shakespear, E.K. Weltzien, C.A. Davis, A. Castillo

**Study supervision:** C. Sweeney

#### Grant Support

This work is supported by U.S. National Institutes of Health awards R01CA129059 (B.J. Caan, PI) and R01CA105274 (L.H. Kushi, PI). Additional support from Bioinformatics and Biostatistics core resources of the Huntsman Cancer Institute, P30CA042014. The Utah Cancer Registry is funded by Contract No. HHSN261201000026C from the National Cancer Institute's SEER Program with additional support from the Utah State Department of Health and the University of Utah.

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 October 7, 2013; revised December 23, 2013; accepted December 24, 2013; published OnlineFirst February 12, 2014.

#### References

- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–52.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
- van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6.
- Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003;100:8418–23.
- Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006;7:96.
- Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;27:1160–7.
- EGAPP Working Group. Recommendations from the EGAPP Working Group: can tumor gene expression profiling improve outcomes in patients with breast cancer? *Genet Med* 2009;11:66–73.
- Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res* 2010;16:5222–32.
- Dunbier AK, Anderson H, Ghazoui Z, Salter J, Parker JS, Perou CM, et al. Association between breast cancer subtypes and response to neoadjuvant anastrozole. *Steroids* 2011;76:736–40.
- Chia SK, Bramwell-Wesley V, Tu D, Shepherd L, Jiang S, Vickery T, et al. A 50 gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clin Cancer Res* 2012;18:4465–72.

11. Bastien RR, Rodriguez-Lescure A, Ebbert MT, Prat A, Munariz B, Rowe L, et al. PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. *BMC Med Genomics* 2012;5:44.
12. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 2006;295:2492–502.
13. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. *Cancer Epidemiol Biomarkers Prev* 2007;16:439–43.
14. Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y, McCue P, et al. Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. *Cancer* 2007;110:876–84.
15. 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;113:1521–6.
16. Kwan ML, Kushi LH, Weltzien E, Maring B, Kutner SE, Fulton RS, et al. Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. *Breast Cancer Res* 2009;11:R31.
17. Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med* 2010;7:e1000279.
18. O'Brien KM, Cole SR, Tse CK, Perou CM, Carey LA, Foulkes WD, et al. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res* 2010;16:6100–10.
19. Dawood S, Hu R, Homes MD, Collins LC, Schnitt SJ, Connolly J, et al. Defining breast cancer prognosis based on molecular phenotypes: results from a large cohort study. *Breast Cancer Res Treat* 2011;126:185–92.
20. Phipps AI, Buist DS, Malone KE, Barlow WE, Porter PL, Kerlikowske K, et al. Reproductive history and risk of three breast cancer subtypes defined by three biomarkers. *Cancer Causes Control* 2011;22:399–405.
21. Tamimi RM, Colditz GA, Hazra A, Baer HJ, Hankinson SE, Rosner B, et al. Traditional breast cancer risk factors in relation to molecular subtypes of breast cancer. *Breast Cancer Res Treat* 2012;131:159–67.
22. Sueta A, Ito H, Islam T, Hosono S, Watanabe M, Hirose K, et al. Differential impact of body mass index and its change on the risk of breast cancer by molecular subtype: a case-control study in Japanese women. *SpringerPlus* 2012;1:39.
23. Calza S, Hall P, Auer G, Bjohle J, Klaar S, Kronenwett U, et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. *Breast Cancer Res* 2006;8:R34.
24. Kelly CM, Bernard PS, Krishnamurthy S, Wang B, Ebbert MT, Bastien RR, et al. Agreement in risk prediction between the 21-gene recurrence score assay (Oncotype DX<sup>®</sup>) and the PAM50 breast cancer intrinsic classifier<sup>™</sup> in early-stage estrogen receptor-positive breast cancer. *Oncologist* 2012;17:492–8.
25. Azim HA Jr., Michiels S, Zagouri F, Delaloge S, Filipits M, Namer M, et al. Utility of prognostic genomic tests in breast cancer practice: the IMPAKT 2012 Working Group Consensus Statement. *Ann Oncol* 2013;24:647–54.
26. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* 2007;25:5287–312.
27. Colombo PE, Milanezi F, Weigelt B, Reis-Filho JS. Microarrays in the 2010s: the contribution of microarray-based gene expression profiling to breast cancer classification, prognostication and prediction. *Breast Cancer Res* 2011;13:212.
28. Habel LA, Shak S, Jacobs MK, Capra A, Alexander C, Pho M, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* 2006;8:R25.
29. Prat A, Parker JS, Fan C, Cheang MC, Miller LD, Bergh J, et al. Concordance among gene expression-based predictors for ER-positive breast cancer treated with adjuvant tamoxifen. *Ann Oncol* 2012;23:2866–73.
30. Caan B, Sternfeld B, Gunderson E, Coates A, Quesenberry C, Slattery ML. Life After Cancer Epidemiology (LACE) Study: a cohort of early stage breast cancer survivors (United States). *Cancer Causes Control* 2005;16:545–56.
31. Kwan ML, Ambrosone CB, Lee MM, Barlow J, Krathwohl SE, Ergas IJ, et al. The Pathways Study: a prospective study of breast cancer survivorship within Kaiser Permanente Northern California. *Cancer Causes Control* 2008;19:1065–76.
32. Ebbert MT, Bastien RR, Boucher KM, Martin M, Carrasco E, Caballero R, et al. Characterization of uncertainty in the classification of multivariate assays: application to PAM50 centroid-based genomic predictors for breast cancer treatment plans. *J Clin Bioinform* 2011;1:37.
33. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* 2011;22:1736–47.
34. Prat A, Cheang MC, Martin M, Parker JS, Carrasco E, Caballero R, et al. Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal a breast cancer. *J Clin Oncol* 2012;31:203–9.
35. Begg CB, Zhang ZF. Statistical analysis of molecular epidemiology studies employing case-series. *Cancer Epidemiol Biomarkers Prev* 1994;3:173–5.
36. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490:61–70.
37. Prat A, Ellis MJ, Perou CM. Practical implications of gene-expression-based assays for breast oncologists. *Nat Rev Clin Oncol* 2012;9:48–57.
38. Yu K, Lee CH, Tan PH, Tan P. Conservation of breast cancer molecular subtypes and transcriptional patterns of tumor progression across distinct ethnic populations. *Clin Cancer Res* 2004;10:5508–17.
39. Weigelt B, Mackay A, A'Hern R, Natrajan R, Tan DS, Dowsett M, et al. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *Lancet Oncol* 2010;11:339–49.
40. Weigelt B, Reis-Filho JS. Molecular profiling currently offers no more than tumour morphology and basic immunohistochemistry. *Breast Cancer Res* 2010;12 Suppl 4:S5.
41. Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr., Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med* 1999;341:2061–7.
42. Scher KS, Hurria A. Under-representation of older adults in cancer registration trials: known problem, little progress. *J Clin Oncol* 2012;30:2036–8.
43. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. *Mol Oncol* 2011;5:5–23.
44. Pirone JR, D'Arcy M, Stewart DA, Hines WC, Johnson M, Gould MN, et al. Age-associated gene expression in normal breast tissue mirrors qualitative age-at-incidence patterns for breast cancer. *Cancer Epidemiol Biomarkers Prev* 2012;21:1735–44.
45. Chuang E, Paul C, Flam A, McCarville K, Forst M, Shin S, et al. Breast cancer subtypes in Asian-Americans differ according to Asian ethnic group. *J Immigr Minor Health* 2012;14:754–8.
46. Kurebayashi J, Moriya T, Ishida T, Hirakawa H, Kurosumi M, Akiyama F, et al. The prevalence of intrinsic subtypes and prognosis in breast cancer patients of different races. *Breast* 2007;16 Suppl 2: S72–7.
47. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, et al. Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat* 2008;109:123–39.

48. 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; 17:2078–86.
49. Phipps AI, Chlebowski RT, Prentice R, McTiernan A, Wactawski-Wende J, Kuller LH, et al. Reproductive history and oral contraceptive use in relation to risk of triple-negative breast cancer. *J Natl Cancer Inst* 2011;103:470–7.
50. Phipps AI, Chlebowski RT, Prentice R, McTiernan A, Stefanick ML, Wactawski-Wende J, et al. Body size, physical activity, and risk of triple-negative and estrogen receptor-positive breast cancer. *Cancer Epidemiol Biomarkers Prev* 2011;20:454–63.