

## COMMENTARY

# How Many Etiological Subtypes of Breast Cancer: Two, Three, Four, Or More?

William F. Anderson, Philip S. Rosenberg, Aleix Prat, Charles M. Perou, Mark E. Sherman

Manuscript received August 15, 2013; revised May 1, 2014; accepted May 13, 2014.

**Correspondence to:** William F. Anderson, MD, MPH, National Cancer Institute, Division of Cancer Epidemiology and Genetics, Biostatistics Branch, 9609 Medical Center Dr, Rm 7E144, Bethesda, MD 20892–9780 (e-mail: [wanderso@mail.nih.gov](mailto:wanderso@mail.nih.gov)).

Breast cancer is a heterogeneous disease, divisible into a variable number of clinical subtypes. A fundamental question is how many etiological classes underlie the clinical spectrum of breast cancer? An etiological subtype reflects a grouping with a common set of causes, whereas a clinical subtype represents a grouping with similar prognosis and/or prediction. Herein, we review the evidence for breast cancer etiological heterogeneity. We then evaluate the etiological evidence with mRNA profiling data. A bimodal age distribution at diagnosis with peak frequencies near ages 50 and 70 years is a fundamental characteristic of breast cancer for important tumor features, clinical characteristics, risk factor profiles, and molecular subtypes. The bimodal peak frequencies at diagnosis divide breast cancer overall into a “mixture” of two main components in varying proportions in different cancer populations. The first breast cancer tends to arise early in life with modal age-at-diagnosis near 50 years and generally behaves aggressively. The second breast cancer occurs later in life with modal age near 70 years and usually portends a more indolent clinical course. These epidemiological and molecular data are consistent with a two-component mixture model and compatible with a hierarchal view of breast cancers arising from two main cell types of origin. Notwithstanding the potential added value of more detailed categorizations for personalized breast cancer treatment, we suggest that the development of better criteria to identify the two proposed etiologic classes would advance breast cancer research and prevention.

JNCI J Natl Cancer Inst (2014) 106(8): dju165

Clinically, breast cancer is widely recognized as a heterogeneous disease. In this commentary, we focus instead upon the etiological perspective of breast cancer heterogeneity. By etiological heterogeneity, we mean breast cancer subtypes (components, classes, or groupings) that share common sets of causes. This is distinct from a clinical subtype, which refers to tumors with common prognostic characteristics and/or predictive features (response to targeted-treatment) (1).

As efforts proceed to improve the taxonomy of clinical breast cancer, a fundamental question persists; how many etiological subtypes actually exist? Clinical taxonomic systems have defined multiple classes in an effort to optimize therapeutic management. At its extreme, this approach translates into “precision” or “personalized” medicine, with a view that each person’s tumor is unique.

We propose a more parsimonious view for breast cancer etiology, which we believe is consistent with a hierarchal view of breast cancer derived from two main cell types of origin (2–9). We show that any given breast cancer molecular or clinical category demonstrates a mixture of two stereotypical age-specific incidence patterns with bimodal peak frequencies near ages 50 years and 70 years. We hypothesize that the consistency of this pattern supports a two-component mixture model, where different molecular and/or clinical categorizations represent variable combinations of two etiological subtypes. In this model, it is the difference in the relative distributions of the two putative subtypes that endows any

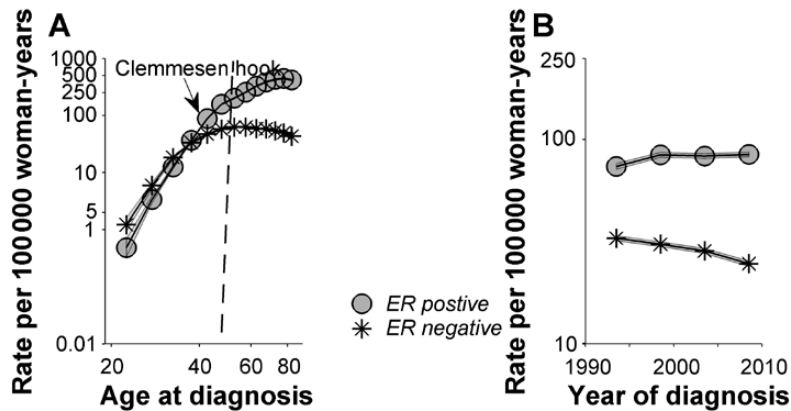
given breast cancer categorization with its distinguishable biological features.

## Historical Developments in the Understanding of Breast Carcinogenesis and Pathogenesis From an Epidemiological Perspective

### Multistage (Log-Linear) Cancer model

More than 50 years ago, Armitage and Doll noted that cancer rates rise exponentially with advancing age for a number of epithelial malignancies (10–12), thus providing the theoretical foundation for multistage tumor initiation, promotion, and progression (13–15). An epidemiological prediction of the multistage cancer model is a log-linear (or log-additive) relationship between cancer incidence and chronological age with a linear (or steady) rise in the logarithm of cancer rates as a function of the logarithm of age at diagnosis.

However, breast cancer incidence does not demonstrate an exponential increase with advancing age. Incidence rates slow before age 50 years (Figure 1A). The change point in incidence has been termed “Clemmesen’s hook” after its discoverer, Johannes Clemmesen (16,17). Clemmesen’s hook is characteristic of female breast cancers worldwide (18–20), coincides with the female climacteric, and is not found among male breast cancers (21).



**Figure 1.** Breast cancer case and population data were obtained from the National Cancer Institute’s Surveillance, Epidemiology, and End Results 9 Registries Database from 1990 through 2010 among women with invasive estrogen receptor (ER)-positive and ER-negative breast cancer. The dataset included thirteen 5-year age groups (ages 20–24, 25–29, ..., 80–84 years) and four 5-year time periods (1991–1995, 1996–2000, 2001–2005, 2006–2010), spanning 16 partially overlapping 10-year birth cohorts, referred to by mid-year of birth (1911, 1916, ..., 1986). We

To account for the distinctive incidence rate pattern for breast cancer, Pike et al. introduced the concept of “breast tissue age” as a better marker of risk than chronological age (22). In his model, key reproductive events affect the shape of the age-specific incidence rate curve. Specifically, risk factors accelerate and protective factors attenuate breast tissue aging, with Clemmesen’s hook representing the net effect. Others have refined the Pike model (23–26); however, a key feature of all of these models is that all breast cancers share a common pathogenesis, reflected in a single, age-specific incidence rate curve.

### Two-Component Cancer Model

Lilienfeld (27) and de Waard (28–30) pioneered the concept that breast cancers develop by two distinct pathways (rather than one), each with a different age-specific incidence rate curve. The first pathway results in mainly premenopausal tumors with peak occurrence early in life, similar to estrogen receptor (ER)-negative cancers in the general US population (Figure 1A). The second pathway results in predominantly postmenopausal cancers with peak incidence later in life, similar to late-onset ER-positive cancers (Figure 1A). In this model, Clemmesen’s hook can be seen as the confluence or superimposition of the age-specific incidence rate curves for the early-onset and late-onset subtypes of breast cancer (31,32).

## Descriptive Epidemiology

### Age-Specific Incidence Rates

The menopause transition was predicted to affect incidence rates of ER-positive more than ER-negative breast cancers, given the presumed greater role of sex-steroid hormones in the pathogenesis of hormone sensitive cancers (33,34). Nonetheless, and somewhat paradoxically, menopause [or rather its surrogate, age 50 years (35,36)] is associated with greater impact upon ER-negative than ER-positive cancers (31,33,34) (Figure 1A). ER-negative rates rise rapidly early in life then flatten or fall soon after menopause (37),

used the age–period–cohort framework to obtain the fitted or longitudinal age-specific incidence rate curve for the mid-cohort, adjusted for period effects (A) and the fitted temporal trends (B). A) Age-specific incidence rates for ER-negative cancers rise rapidly early in reproductive life and then flatten or fall. Rates for ER positive cancers rise rapidly early in life and then continue to rise at a slower pace. B) ER-positive and ER-negative temporal trends have diverged over time. ER-positive rates have risen, whereas ER-negative rates have fallen.

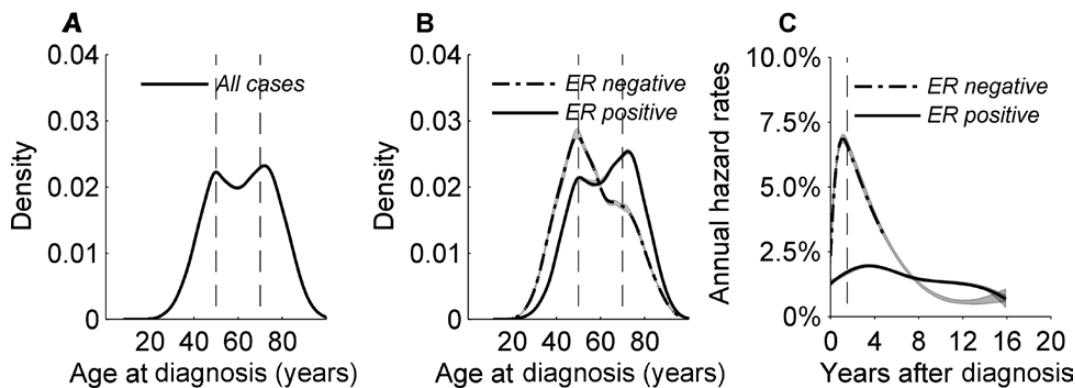
whereas ER-positive rates rise continuously irrespective of menopause, albeit more slowly after age 50 years.

The different incidence rate patterns by ER expression represent an age interaction or effect modification. Under the null hypothesis of no interaction, the age-specific incidence rates for ER-positive and ER-negative cancers would be parallel on the log scale (38–40), yielding a constant incidence rate ratio (IRR) irrespective of age at diagnosis ( $IRR_{ERneg\ to\ ERpos} = \text{constant for all ages}$ ). We also make a distinction between quantitative (noncrossover) and qualitative (crossover or reversing) age interactions (37–40). A quantitative age interaction varies in magnitude but not direction, whereas a qualitative interaction differs in both magnitude and direction. At the extreme ages in Figure 1A, the incidence rate ratio of ER-negative to ER-positive cancer is 2.3 during ages 20 to 24 years ( $IRR_{ERneg\ to\ ERpos} > 1.0$ ) and the incidence rate ratio of ER-negative to ER-positive cancer is 0.10 during the ages 80 to 84 years ( $IRR_{ERneg\ to\ ERpos} < 1.0$ ). True qualitative or reversing interactions are considered rare (40–42) but, when found, can be statistical surrogates for age-dependent etiological heterogeneity. In the context of a qualitative age interaction, Clemmesen’s hook can be further seen as the crossover in falling ER-negative rates and rising ER-positive rates.

### Bimodal Age Distributions at Diagnosis

If breast cancer followed the log-linear incidence curve described by Armitage and Doll (10–12), breast cancer cases in the general population would be predicted to show a unimodal age distribution at diagnosis. In contrast, breast cancer overall demonstrates a bimodal pattern, with the modal ages near 50 and 70 years representing the central tendencies for the early-onset and late-onset breast cancers (Figure 2A). Density plots are constructed with 1-year increments using a “smoothing” method for the corresponding age distributions at diagnosis (31,43), where the area under the curve includes all of the breast cancer cases in a given population.

Crossing ER-positive and ER-negative age-specific rates (Figure 1A) also shows bimodal age distributions at diagnosis



**Figure 2.** Invasive female breast cancer case data were obtained from the National Cancer Institute's Surveillance, Epidemiology, and End Results 9 Registries Database from 1990 through 2010 database overall and for estrogen receptor (ER)-positive and ER-negative cancers. Bimodal breast cancer populations have fluctuated over time likely because of complex interactions between age-related biologic, risk factor, and screening phenomena, as previously described (186). For illustration, this figure has been restricted to the 1995 to 1998 period, during which a bimodal female breast cancer population was evenly distributed between early-onset and late-onset subtypes. Age distributions at diagnosis (or density plots) with 95% confidence intervals were constructed in 1-year age increments using a kernel density estimator applied to the corresponding age-at-diagnosis frequency histogram. The area under the curve represents 100% of the cancer records. The vertical axis shows the smoothed distribution (or proportion) with the frequency value  $\times 100 =$  percentage distribution. **A)** Density plot for breast cancer overall demonstrates a bimodal age distribution at diagnosis with the modal ages near 50 and 70 years representing the central tendencies for early-onset and late-onset breast cancers. **B)** Density plot for ER-negative tumors also shows a bimodal age distribution at diagnosis with a dominant early-onset mode near

age 50 years and a minor mode around age 70 years. Density plot for ER-positive tumors shows bimodal age distributions at diagnosis with a dominant late-onset mode near age 70 years and a minor mode around age 50 years. **C)** The risk for breast cancer-specific death can be expressed as an annual hazard rate, which describes the instantaneous rate of dying from breast cancer in a specified time interval (ie, percentage dying per year) after diagnosis among women who are alive at the beginning of that time interval. Nonparametric hazard function estimators were applied that modeled the hazard profile of ER-positive and ER-negative cancers, allowing both the shape and magnitude to be estimated free of ad hoc mathematical assumptions. Specifically, the hazard rate curves were generated using cubic splines with joinpoints selected by Akaike's information criteria and 95% confidence intervals applied with bootstrap resampling (187-189). Bimodal age distributions at diagnosis among women **(B)** are associated with two very different cancer-specific outcomes. ER-negative hazards for breast cancer death peak near 7.5% per year approximately 2 years after initial diagnosis and then decline rapidly. ER-positive hazards lack a sharp peak but are relatively constant at 1% to 2% per year. Falling ER-negative and constant ER-positive hazards cross over approximately 8 years after breast cancer diagnosis.

(Figure 2B) (31,43), as do many other important breast cancer features such as tumor size, lymph nodal status, and histological grade (44,45). Although the modal ages of 50 and 70 years have been robust irrespective of breast cancer characteristic, the relative distributions (or mixtures) for the early-onset and late-onset breast cancer subtypes can vary by specific stratifying factor, providing that class with distinguishable features. In general, high-risk tumors such as ER-negative cancers have bimodal breast cancer patterns with a predominant early-onset peak (Figure 2B). Lower-risk tumors, such as ER-positive cancers, have bimodal breast cancer patterns with a more dominant late-onset peak (Figure 2B).

### Age-Adjusted Secular Trends

Recent studies in the United States show an unexpected divergence of ER-positive and ER-negative breast cancer trends (46-48); ER-positive cancers have risen over the long term, whereas ER-negative breast cancers have declined (Figure 1B). Theoretically, divergent ER trends in the United States could have resulted from statistical anomalies (49), changes in assay methodology or application of lower thresholds for classifying ER tests as positive (50), and/or the implementation of organized screening mammography (51,52). However, secular trends in other countries with better control of these potentially confounding factors (eg, Denmark) showed similar trends as in the United States (53), consistent with the hypothesis that divergent ER-positive and ER-negative trends might be due to changes in different risk factor profiles by ER subtype over time (47,48,54).

### Biostatistical Models

Our group has leveraged the application of biostatistical models to complement descriptive epidemiology. First, we used two-component mixture models to determine whether bimodal age distributions at diagnosis fitted the data better than a single density (55) (Figure 2, A and B). Then, to confirm qualitative age interactions (eg, Figure 1A), we used age-period-cohort (APC) models to evaluate age-specific effects independent of calendar-period effects (that relate to screening, changing diagnostic and/or practice patterns) and/or birth-cohort effects (generational and/or exposure factors) (56,57).

### Two-Component Mixture Models

We initially identified three distinct age-specific incidence rate patterns that were closely associated with seven histopathological breast cancer subtypes (58). Incidence rates of infiltrating duct, tubular, and lobular carcinomas rose rapidly until age 50 years, then increased more slowly, similar to rates for breast cancer overall. Rates for medullary and inflammatory breast carcinomas increased rapidly until age 50 years, then flattened or fell, similar to ER-negative rates (Figure 1A). Finally, rates for papillary and mucinous carcinomas increased steadily with age, similar to ER-positive rates (Figure 1A) and much like cancers at many other organ sites such as colorectal cancer (12).

Notwithstanding the three distinct incidence rate patterns, two-component mixture models demonstrated that six of the histopathological subtypes had bimodal age distributions at diagnosis with early-onset and/or late-onset peak frequencies around the

stereotypical ages of 50 and 70 years (45). The one notable exception was medullary carcinoma, which showed a unimodal age distribution with mode close to age 50 years. Medullary carcinomas are linked to the loss of *BRCA1* function (59–61), which we propose may represent the closest known approximation to an etiologically pure early-onset subtype of breast cancer.

Similar to ER-positive and ER-negative breast cancers (Figure 2B), the bimodal peaks for different histopathological categories do not sharply divide cancers into pure groups, but rather reflect central tendencies of what we propose are two fundamental etiological classes. Breast cancers that develop at extreme ages are likely to be highly enriched for one etiological class, but both of these classes span the entire lifespan, with substantial mixing during the middle years, when many cancers occur. The mixing fraction or proportion of each etiological group varies within a class of breast cancers depending on its definition; however, the peak ages remain near ages 50 and 70 years.

### APC Models

A useful APC function is the fitted (or longitudinal) age-specific incidence rate curve (37,62) (Figure 1A). The fitted curve stitches together the age-specific incidence rates from a collection of birth cohorts, each one observed over a limited and variable age span (ie, younger cohorts are observed at younger ages and older cohorts at older ages). The resulting curve estimates the age-specific rates of the middle or reference cohort over the entire age range. In contrast with the typical cross-sectional, age-specific incidence rate curve that may be confounded by period and cohort effects (63,64), the fitted curve is conditioned upon cohort and adjusted for period changes.

Clemmesen's menopausal hook for breast cancer overall was once dismissed as a birth-cohort artifact (65), where the progressive increase in breast cancer risk from one generation to the next gave the appearance of falling incidence rates among older persons (64). This view is refuted by APC models (37), which demonstrate that the Clemmesen's phenomenon is a true age-related event that persists in the fitted age-specific incidence rate curve (Figure 1A). Similar modeling approaches confirm that the qualitative age interaction remains for the fitted curves for ER-positive and ER-negative breast cancers, respectively (37).

### Analytic Epidemiology

Consistent with descriptive epidemiology and biostatistical models, analytic epidemiology also supports a two-component breast cancer mixture model based upon the identification of major risk factors and genetic susceptibility.

### Risk Factor Epidemiology

Despite the well-established protective effect of full-term pregnancy for breast cancer overall (22,66,67), it has been suggested, albeit without total agreement (68,69), that parity is associated with an early increase in risk, followed by long-term protection (70–86). Alternatively, the early risk and late protection for parity could reflect another qualitative (crossover or reversing) age interaction for breast cancer (40), where parity increases risk of early-onset subtypes and reduces risk of late-onset ones (37). Indeed, parity and multiple live

births are associated with reduced risk for ER-positive and luminal A intrinsic breast cancer subtypes (Table 1) and increased risk of ER-negative tumors (specifically, ER-negative basal-like or triple-negative intrinsic breast cancer subtypes) (87–91). The combined effect of early age at first birth and lack of breast feeding appears to impart an especially high risk for basal-like (triple-negative and *BRCA1*) cancers (88,89,92–98), particularly among some specific ethnic groups.

Obesity is another risk factor with dual effects by age at diagnosis and ER status (90,99,100), which are possibly mediated through the cholesterol metabolite 27-hydroxycholesterol (101–103). Body weight has a direct association with postmenopausal breast cancer and an inverse relationship with premenopausal cancer (104–109). Obesity also has a stronger positive association with hormone receptor-positive than hormone receptor-negative cancers (101,102,110), although obesity may also increase the risk of basal-like, triple-negative, and inflammatory breast cancers (88,90,111,112). Rising ER-positive cancers among older women and falling ER-negative tumors among younger women (Figure 1B) are consistent with rising obesity and declining parity in the United States (113) and Denmark (114), as well as many other parts of the world (115). Other studies have also found dual risk factor associations by ER status [reviewed in (54)].

### Genetic Susceptibility

Many genetic loci are known to contribute to the risk of familial breast cancer, including highly penetrant deleterious germline mutations in the *BRCA1* and *BRCA2* tumor suppressor genes. Of note, *BRCA1* and *BRCA2* breast cancers are distinct in their expression of hormone receptors (116–118). Roughly 75% of *BRCA1* breast cancers are ER negative and 25% are ER positive. On the other hand, 75% of *BRCA2* breast cancers are ER positive and 25% are ER negative, similar to breast cancer in the general population. Recent genome-wide association studies also have identified more than 75 low-penetrant susceptibility loci with evidence for specificity by tumor subtype (119–129). For 13 loci (119), there is greater relative risk for ER-positive than ER-negative cancers. For two notable exceptions (single nucleotide polymorphisms rs6828523 and rs7072776), the relative risk is opposite for ER-positive and ER-negative tumors (119), consistent with risk factor differences by ER expression. Lastly, there are a number of susceptibility loci only for ER-negative but not ER-positive tumors (eg, the hTERT loci) (130, 131).

### Molecular Class Discovery And Prognosis

Analyses of gene-expression profiling data reveal that ER-positive and ER-negative tumors are fundamentally distinct diseases in molecular terms (132). There are two predominantly ER-positive intrinsic molecular subtypes (ie, luminal A and luminal B) and two predominantly ER-negative intrinsic subtypes (ie, HER2-enriched and basal-like) (Table 1). The intrinsic molecular subtypes are largely distinguished by the expression of genes involved in luminal epithelial differentiation (eg, ER and PR genes), proliferation (eg, Ki67 gene), human epidermal growth factor receptor 2 pathway (eg, HER2 gene), and basal differentiation (133–136).

The intrinsic molecular signatures are robust across multiple genomic platforms (137,138), apply to both carcinoma in situ and

**Table 1.** Immunohistochemical staining for the intrinsic breast cancer molecular subtypes\*

IHC subtype (%)†	ER and/or PR	HER2	Ki67	Intrinsic subtype
Luminal A (73%)	Positive	Negative	Low	Luminal A
Luminal B (10%)	Positive	Negative or positive	Low or high	Luminal B
HER2 positive, nonluminal (5%)	Negative	Positive	Not needed	HER2-enriched
Triple-negative (12%)	Negative	Negative	Not needed	Basal-like

\* There are two predominantly hormone-positive (estrogen receptor [ER] and/or progesterone receptor [PR]) intrinsic molecular subtypes (luminal A and luminal B) and two predominantly hormone-negative intrinsic subtypes (human epidermal growth factor receptor 2 [HER2] enriched and basal-like) (133–136). Additionally, there are two predominantly HER2-positive intrinsic molecular subtypes (luminal B and HER2 enriched) and two predominantly HER2-negative intrinsic subtypes (luminal A and Basal-like). Adapted from Goldhirsch et al. (157). IHC = immunohistochemical.

† Estimated percentage distribution for IHC-derived subtypes among women with breast cancer and known ER, PR, and HER2 expression in the general population of the United States in 2010, provided by the National Cancer Institute's SEER database (178).

invasive breast cancers (139–141), and are identifiable within different racial groups (142–144). Although the intrinsic gene set was originally developed through agnostic profiling, the power of this approach reflects its ability to define molecular subtypes that vary with respect to prognosis and treatment (134,136,137,145–156). The intrinsic subtypes have been variously approximated with immunohistochemical staining algorithms (Table 1); however, there can be considerable discordance between the gene-based and immunohistochemical-based expression profiles for the intrinsic subtypes (157–159).

Luminal breast cancers are the most heterogeneous intrinsic subtypes, with the luminal A tumors distinguished by the high expression of luminal epithelial genes, low expression of the Ki67, and the best prognosis (160,161). The difference between the luminal A and luminal B gene patterns is less distinct than the difference between the luminal A and basal-like subtypes, which appear to be anticorrelated (88,146,162–165). Among the intrinsic subtypes, the basal-like tumors have the most unique and distinctive genomic profile (132,166–169); they are in many ways more similar to squamous cell carcinomas of the lung and high-grade serous ovarian carcinomas than to all other subtypes of breast cancer (9,170). Basal-like tumors are enriched with *BRCAl*-mutated and triple-negative breast cancers but also include some special histopathological subtypes such as medullary and adenoid cystic tumors (138,157,158,171). Finally, the HER2-enriched (HER2E) subtype shows a global gene signature that lies more closely to the luminal than basal-like cancers (172), with HER2 cell surface expression possibly playing an important role in regulating the luminal cancer stem cell population (5–7).

If every molecular subtype was a unique biological entity, one might anticipate that each would demonstrate a distinct age-specific incidence rate curve and unimodal age frequency distribution at diagnosis (45,58). Unfortunately, incidence rate data are not readily available for the intrinsic molecular signatures because gene-expression analyses have mostly been limited to case series and/or small observational studies from convenience and/or hospital-based samples. Nonetheless, in an early study that used the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Residual Tissue Repository (164,173), we estimated the age-specific incidence rates for the molecular subtypes with immunohistochemical staining of breast cancer tissue microarrays and imputed population data. Luminal A (defined as ER positive and HER2 negative) incidence rates rose continuously, although more slowly after age 50 years (similar to ER-positive

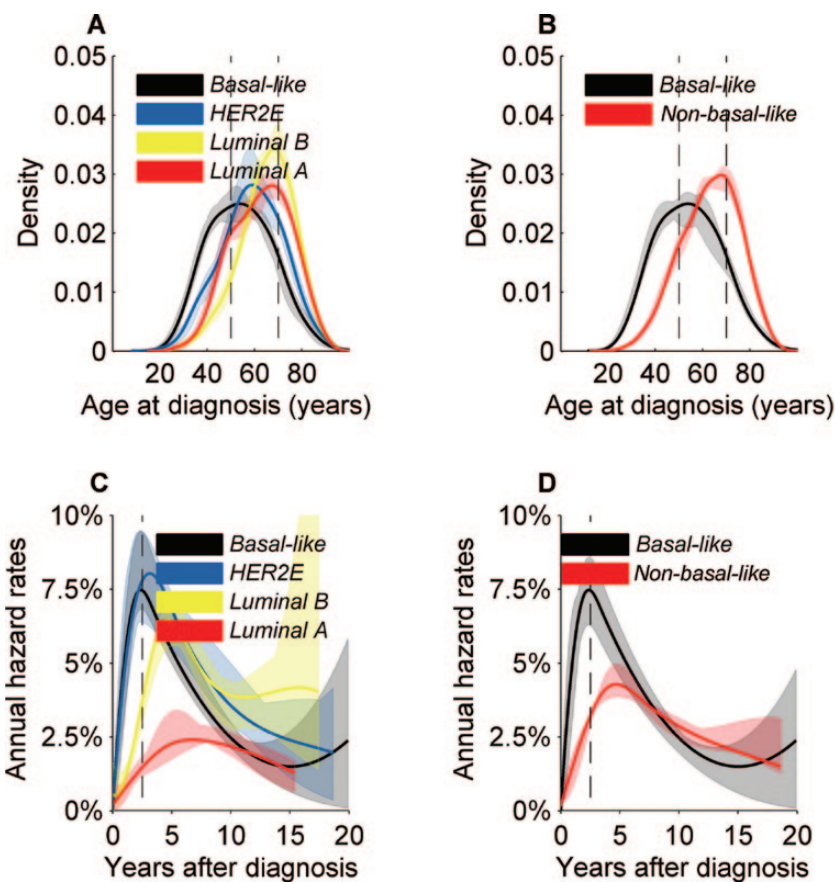
cancers) (Figure 1A). Basal-like rates (defined as ER negative and HER2 negative) increased rapidly early in life then flattened or fell (similar to ER negative cancer) (Figure 1A). These patterns are consistent with subsequent studies from the California Cancer Registry (174–177) and emerging data from SEER's large-scale population-based database (178).

Given the limited availability of population-based incidence rate data, we applied the intrinsic gene set classification algorithm (147) to approximately 2000 breast cancer cases reported by METABRIC Group (179,180). Although METABRIC described 10 distinct groups based primarily upon copy number data, the age distribution patterns by gene expression–defined intrinsic molecular subtypes are reminiscent of the bimodal patterns defined by ER protein expression (Figure 2B). Luminal A and luminal B cases had bimodal age distributions with predominant late modes near age 70 years and minor modes near age 50 years (Figure 3A), similar to ER-positive cancers in SEER (Figure 2B). Basal-like cancers had an early-onset mode around age 50 years (Figure 3, A and B), similar to ER-negative cancers in SEER (Figure 2B). The HER2-enriched age distribution lay midway between the luminal and basal-like cancers (Figure 3A). Combining the HER2-enriched, luminal A, and luminal B cases into a single non-basal-like group did not appreciably alter the bimodal shape of the molecular subtypes (Figure 3B).

In METABRIC, hazard rates for breast cancer–specific death for basal-like, HER2-enriched, and luminal B tumors peaked near 7.5% per year approximately 2 years after initial breast cancer diagnosis then declined (Figure 3C), similar to ER-negative cases in SEER (Figure 2C). Luminal A hazard rates lacked a sharp peak and were relatively constant at 2% to 2.5% per year (Figure 3C), similar to ER-positive cases in SEER (Figure 2C). Overlaying the hazard rate curves for non-basal-like and basal-like cancers demonstrated crossing hazard rates approximately 8 years after diagnosis (Figure 3D), similar to ER-positive and ER-negative cancers in SEER (Figure 2C).

## Directions for Future Research

We present the hypothesis that breast cancer comprises two fundamental etiological components, classes or subtypes; which, as of yet, are not specifically defined but which induce bimodal age distributions at diagnosis irrespective of the classification applied. The two putative main etiological subtypes are characterized by sharply contrasting tendencies related to age-specific incidence



**Figure 3.** MEATBRIC case data for approximately 2000 breast tumors (179). Global gene expression was assessed with the intrinsic gene set. **A)** Density plots for age at diagnosis of basal-like cancers had an early-onset mode near age 50 years. Luminal A and luminal B cases showed bimodal age distributions with predominant late-onset modes near age 70 years and minor modes near age 50 years. The human epidermal growth factor receptor 2-enriched (HER2E) age distribution at diagnosis lies in between the density plots for the luminal and basal-like cancers. **B)** Combining HER2E and luminal cases into a single non-basal-like group did not appreciably alter the shape of the bimodal age distribution plots shown for molecular subtypes. **C)** Hazard rates of breast cancer-specific

death for basal-like, HER2E, and luminal B cancers peaked near 7.5% per year approximately 2 years after initial breast cancer diagnosis then declined, similar to estrogen receptor (ER)-negative cases in the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database (Figure 2C). Luminal A hazard rates lacked a sharp peak and were relatively constant at 2% to 2.5% per year, similar to ER-positive cases in SEER (Figure 2C). **D)** Combining the HER2E and luminal cases into a single non-basal-like group resulted in a hazard plot that was intermediate between the basal-like and luminal A hazard rates. Hazard rates crossed over approximately 8 years after breast cancer diagnosis, similar to ER-positive and ER-negative cancers in SEER (Figure 2C).

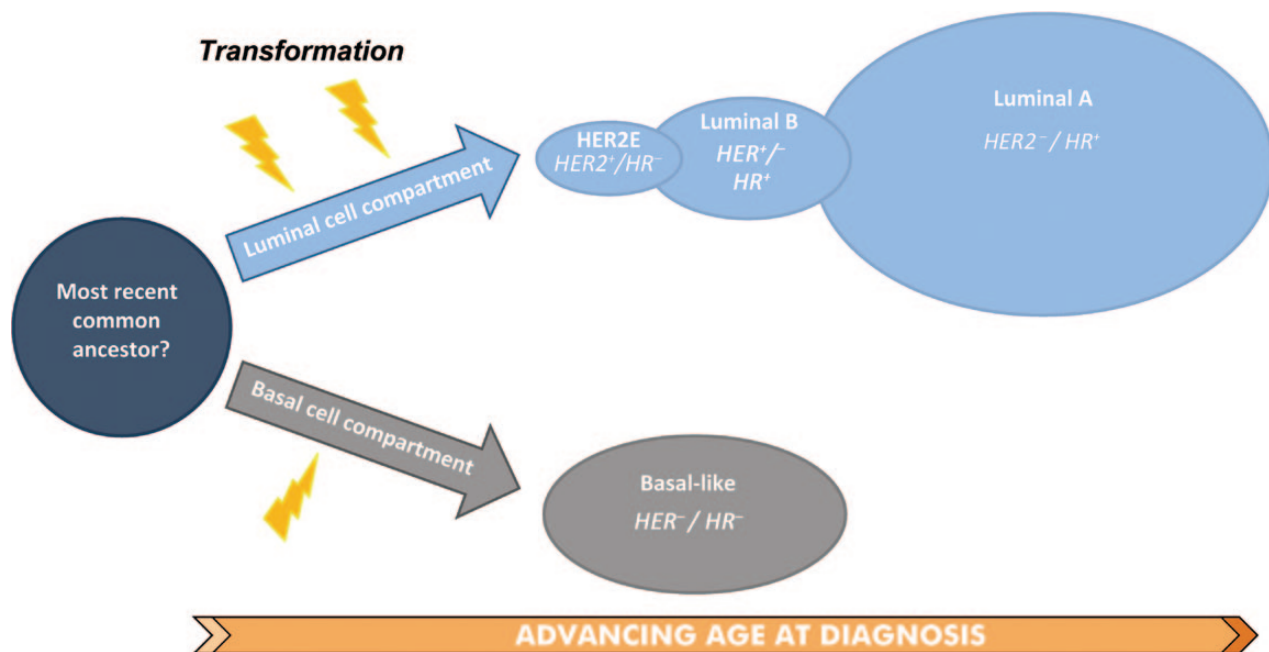
rates, clinical prognosis, risk factor profiles, and somatic gene and protein expression. ER expression by age at diagnosis provides an epidemiologically useful correlate for our two-component breast cancer mixture model, but it is not a perfect surrogate.

From the molecular perspective, multianalyte genomic profiling has revealed many novel breast cancer subtypes with distinguishable, if not distinctive, clinical tendencies. We speculate that these categorizations also may reflect varying mixtures of two main etiological components, as demonstrated by a ubiquitous bimodal age distribution at diagnosis. Presently, the luminal A and basal-like molecular signatures are the most clearly distinguishable subtypes (Figure 3A) and appear to be polar opposites at most every level (3,6,163,181).

Furthermore, emerging molecular evidence shows that across different types of cancer, breast cancer is one of the few cancers with two major divisions. One division consists of basal-like tumors and the other is composed of luminal A, luminal B, and HER2-enriched cancers (9,170), which we have conceptualized with a cartoon in Figure 4. These data, together with the epidemiological

findings reported herein, suggest that breast cancer overall may be viewed as a hierarchal disease derived from two main cell types of origin (ie, basal/myoepithelial vs luminal cellular compartment) (2–9). Basal-like breast cancers arise from the basal/myoepithelial cell compartment, whereas non-basal-like cancers (ie, HER2-enriched, luminal B, and luminal A) emerge from a more luminal-like cell compartment.

Data in this commentary also suggest themes for future etiological and clinical research. Large-scale and population-based epidemiological studies could stratify etiological analyses by molecular subtypes such as basal-like vs non-basal-like or luminal. Our current thinking is that there are two main etiological subtypes for sporadic breast cancers, but there could be a few more, potentially including rare subtypes such as inflammatory breast cancer (112). Given the challenges in molecular subtyping in large-scale population-based studies, it would be helpful to further develop and validate parsimonious robust marker panels for studies that merge individual-level molecular and clinical data in populations within well-defined catchment areas (173,182,183).



**Figure 4.** Results and data herein suggest that breast cancer overall may be viewed as a hierarchical disease, consisting of a two-component mixture of two main cell types of origin (ie, luminal vs basal/myoepithelial). As hypothesized in this cartoon, human epidermal growth factor

receptor 2-enriched (HER2E) and luminal intrinsic molecular subtypes are initiated (**lightning bolts**) from the luminal cell compartment. On the other hand, in this model, basal-like breast cancers (HER<sup>-</sup>/HR<sup>-</sup>) are initiated within the basal/myoepithelial cell compartment.

Understanding the etiological heterogeneity of breast cancer, whether it is fundamentally bimodal as proposed herein or more complex, would have critical implications for breast cancer prevention. For example, selective ER modulators are effective in reducing breast cancer incidence among women at higher risk overall (184), but the means for identifying women specifically at elevated risk for hormonally driven cancers are lacking. Furthermore, ER-positive status is only a surrogate of hormone dependence; selective ER modulators do not prevent all ER-positive cancers nor do all ER-positive tumors respond to adjuvant endocrine therapy (4,185). If the hormone-responsive phenotype is linkable to a risk factor profile (eg, the etiological class of breast cancers enriched, but not exclusively within the late onset peak), then identifying this subtype more specifically would represent an advance with translational potential for both prevention and treatment. Similarly, understanding the basic biology of tumors enriched within the early-onset age distribution at diagnosis might offer analogous opportunities for *BRCA1*-related, basal-like, and/or triple-negative breast cancers.

As a final thought, although a two-component mixture model may seem too simplistic for breast cancer clinical heterogeneity, it is not too simple for etiology. Much of the clinical heterogeneity for breast cancer may result from tumor promotion and progression, which is very likely far downstream of tumor initiation (31). Therefore, breast cancer etiology (tumor initiation) may be less complex than subsequent tumor promotion and progression. Additionally, although parsimonious, the complexity of mixture models should not be underappreciated. The potential for an infinite number of mixtures of just two main components is complicated enough to account for much of the observed breast cancer heterogeneity.

## References

- Henderson IC, Patek AJ. The relationship between prognostic and predictive factors in the management of breast cancer. *Breast Cancer Res Treat.* 1998;52(1-3):261-288.
- Anderson WF, Matsuno RK. Breast cancer heterogeneity: a mixture of at least two main types? *J Natl Cancer Inst.* 2006;98(14):948-951.
- Li JJ, Li SA, Mohla S, Rochefort H, Maudelonde T. Hormonal Carcinogenesis V. *Advances in Experimental Medicine and Biology.* Vol. 617. New York: Springer; 2008. ISBN 978-0-387-69078-0.
- Ellis MJ, Ding L, Shen D, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature.* 2012;486(7403):353-360.
- Ithimakin S, Day KC, Malik F, et al. HER2 drives luminal breast cancer stem cells in the absence of HER2 amplification: implications for efficacy of adjuvant trastuzumab. *Cancer Res.* 2013;73(5):1635-1646.
- Korkaya H, Wicha MS. HER2 and breast cancer stem cells: more than meets the eye. *Cancer Res.* 2013;73(12):3489-3493.
- Korkaya H, Wicha MS. Breast cancer stem cells: we've got them surrounded. *Clin Cancer Res.* 2013;19(3):511-513.
- Prat A, Perou CM. Mammary development meets cancer genomics. *Nat Med.* 2009;15(8):842-844.
- Prat A, Adamo B, Fan C, et al. Gene expression analyses across six cancer types identify basal-like breast cancer as a unique molecular entity. *Scientific Rep.* 2013;3:1-12.
- Armitage P, Doll R. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br J Cancer.* 1957;11(2):161-169.
- Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer.* 1954;8(1):1-12.
- Knudson AG. Two genetic hits (more or less) to cancer. *Nat Rev Cancer.* 2001;1(2):157-162.
- Doll R. Commentary: the age distribution of cancer and a multistage theory of carcinogenesis. *Int J Epidemiol.* 2004;33(6):1183-1184.
- Frank SA. Commentary: mathematical models of cancer progression and epidemiology in the age of high throughput genomics. *Int J Epidemiol.* 2004;33(6):1179-1181.
- Moolgavkar SH. Commentary: fifty years of the multistage model: remarks on a landmark paper. *Int J Epidemiol.* 2004;33(6):1182-1183.

16. Clemmesen J. Carcinoma of the breast. *Br J Radiol.* 1948;21(252):583–590.
17. Clemmesen J. On the etiology of some human cancers. *J Natl Cancer Inst.* 1951;12(1):1–21.
18. Muguti GI. Experience with breast cancer in Zimbabwe. *J R Coll Surg Edinb.* 1993;38(2):75–78.
19. Chie WC, Chen CF, Lee WC, Chen CJ, Lin RS. Age-period-cohort analysis of breast cancer mortality. *Anticancer Res.* 1995;15(2):511–515.
20. Sant M, Gatta G, Micheli A, et al. Survival and age at diagnosis of breast cancer in a population-based cancer registry. *Eur J Cancer.* 1991;27(8):981–984.
21. Thomas DB. Breast cancer in men. *Epidemiol Rev.* 1993;15(1):220–231.
22. Pike MC, Krailo MD, Henderson BE, Casagrande JT, Hoel DG. “Hormonal” risk factors, “breast tissue age: and the age-incidence of breast cancer. *Nature.* 1983;303(5920):767–770.
23. Pathak DR, Whittemore AS. Combined effects of body size, parity, and menstrual events on breast cancer incidence in seven countries. *Am J Epidemiol.* 1992;135(2):153–168.
24. Rosner B, Colditz GA. Nurses’ Health Study: log-incidence mathematical model of breast cancer incidence. *J Natl Cancer Inst.* 1996;88(6):359–364.
25. Rosner B, Colditz GA, Willett WC. Reproductive risk factors in a prospective study of breast cancer: the Nurses’ Health Study. *Am J Epidemiol.* 1994;139(8):819–835.
26. Moolgavkar SH, Day NE, Stevens RG. Two-stage model for carcinogenesis: epidemiology of breast cancer in females. *J Natl Cancer Inst.* 1980;65(3):559–569.
27. Lilienfeld AM, Johnson EA. The age distribution in female breast and genital cancers. *Cancer.* 1955;8:875–882.
28. De Waard F. Premenopausal and postmenopausal breast cancer: one disease or two? *J Natl Cancer Inst.* 1979;63(3):549–552.
29. De Waard F, Baanders-van Halewijn EA, Huizinga J. The bimodal age distribution of patients with mammary carcinoma; evidence for the existence of 2 types of human breast cancer. *Cancer.* 1964;17(2):141–151.
30. De Waard F, De Laive JWJ, Baanders-van Halewijn EA. On bimodal age distribution of mammary carcinoma. *Br J Cancer.* 1960;14(3):437–448.
31. Anderson WF, Chatterjee N, Ershler WB, Brawley OW. Estrogen receptor breast cancer phenotypes in the Surveillance, Epidemiology, and End results database. *Breast Cancer Res Treat.* 2002;76(1):27–36.
32. Benz CC. Impact of aging on the biology of breast cancer. *Crit Rev Oncol Hematol.* 2008;66(1):65–74.
33. Yasui Y, Potter JD. The shape of age-incidence curves of female breast cancer by hormone-receptor status. *Cancer Causes Control.* 1999;10(5):431–437.
34. Tarone RE, Chu KC. The greater impact of menopause on ER- than ER+ breast cancer incidence: a possible explanation (United States). *Cancer Causes Control.* 2002;13(1):7–14.
35. Morabia A, Costanza MC. International variability in ages at menarche, first livebirth, and menopause. World Health Organization Collaborative Study of Neoplasia and Steroid Contraceptives. *Am J Epidemiol.* 1998;148(12):1195–1205.
36. Morabia A, Flandre P. Misclassification bias related to definition of menopausal status in case-control studies of breast cancer. *Int J Epidemiol.* 1992;21(2):222–228.
37. Anderson WF, Rosenberg PS, Menashe I, Mitani A, Pfeiffer RM. Age-related crossover in breast cancer incidence rates between black and white ethnic groups. *J Natl Cancer Inst.* 2008;100(24):1804–1814.
38. Anderson WF, Matsuno RK, Sherman ME, et al. Estimating age-specific breast cancer risks: a descriptive tool to identify age interactions. *Cancer Causes Control.* 2007;18(4):439–447.
39. Anderson WF, Chen BE, Brinton LA, Devesa SS. Qualitative age interactions (or effect modification) suggest different cancer pathways for early-onset and late-onset breast cancers. *Cancer Causes Control.* 2007;18(10):1187–1198.
40. Anderson WF, Jatoi I, Sherman ME. Qualitative age interactions in breast cancer studies: mind the gap. *J Clin Oncol.* 2009;27(32):5308–5311.
41. Cochrane AL. *Effectiveness and Efficiency: Random Reflections on Health Services.* London: Nuffield Provincial Hospitals Trust; 1972.
42. Peto R. *Statistical Aspects of Cancer Trials.* London: Chapman and Hall; 1982.
43. Anderson WF, Chu KC, Chatterjee N, Brawley OW, Brinton LA. Tumor variants by hormone receptor expression in white patients with node-negative breast cancer from the Surveillance, Epidemiology, and End Results database. *J Clin Oncol.* 2001;19(1):18–27.
44. Anderson WF, Jatoi I, Devesa SS. Distinct breast cancer incidence and prognostic patterns in the NCI’s SEER program: suggesting a possible link between etiology and outcome. *Breast Cancer Res Treat.* 2005;90(2):127–137.
45. Anderson WF, Pfeiffer RM, Dores GM, Sherman ME. Comparison of age frequency distribution patterns for different histopathologic types of breast carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2006;15(10):1899–1905.
46. Desantis C, Howlader N, Cronin KA, Jemal A. Breast cancer incidence rates in US women are no longer declining. *Cancer Epidemiol Biomarkers Prev.* 2011;20(5):733–739.
47. Anderson WF, Katki HA, Rosenberg PS. Breast cancer incidence in the United States: current and future trends. *J Natl Cancer Inst.* 2011;103(18):1397–1402.
48. Desantis C, Ma J, Bryan L, Jemal A. Breast cancer statistics, 2013. *CA Cancer J Clin.* 2014;64(1):52–62.
49. Berry DA, Ravdin PM. Breast cancer trends: a marriage between clinical trial evidence and epidemiology. *J Natl Cancer Inst.* 2007;99(15):1139–1141.
50. Hammond ME, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol.* 2010;28(16):2784–2795.
51. Porter PL, El-Bastawissi AY, Mandelson MT, et al. Breast tumor characteristics as predictors of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst.* 1999;91(23):2020–2028.
52. Narod SA, Dube MP. Biologic characteristics of interval and screen-detected breast cancers. *J Natl Cancer Inst.* 2001;93(2):151–152.
53. Anderson WF, Rosenberg PS, Petito L, et al. Divergent estrogen receptor-positive and -negative breast cancer trends and etiologic heterogeneity in Denmark. *Int J Cancer.* 2013;133(9):2201–2206.
54. Toriola AT, Colditz GA. Trends in breast cancer incidence and mortality in the United States: implications for prevention. *Breast Cancer Res Treat.* 2013;138(3):665–673.
55. Pfeiffer RM, Carroll RJ, Wheeler W, Whitby D, Mbulaiteye S. Combining assays for estimating prevalence of human herpesvirus 8 infection using multivariate mixture models. *Biostatistics.* 2008;9(1):137–151.
56. Rosenberg PS, Anderson WF. Proportional hazard models and age-period-cohort analysis of cancer rates. *Stat Med.* 2010;29(11):1228–1238.
57. Holford TR. The estimation of age, period and cohort effects for vital rates. *Biometrics.* 1983;39(2):311–324.
58. Anderson WF, Chu KC, Chang S, Sherman ME. Comparison of age-specific incidence rate patterns for different histopathologic types of breast carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2004;13(7):1128–1135.
59. Marcus JN, Page DL, Watson P, Narod SA, Lenoir GM, Lynch HT. BRCA1 and BRCA2 hereditary breast carcinoma phenotypes. *Cancer.* 1997;80(S3):543–556.
60. Lakhani SR, Gusterson BA, Jacquemier J, et al. The pathology of familial breast cancer: histological features of cancers in families not attributable to mutations in BRCA1 or BRCA2. *Clin Cancer Res.* 2000;6(3):782–789.
61. Armes JE, Venter DJ. The pathology of inherited breast cancer. *Pathology (Phila).* 2002;34(4):309–314.
62. Rosenberg PS, Anderson WF. Age-period-cohort models in cancer surveillance research: ready for prime time? *Cancer Epidemiol Biomarkers Prev.* 2011;20(7):1263–1268.
63. Moolgavkar SH, Stevens RG, Lee JA. Effect of age on incidence of breast cancer in females. *J Natl Cancer Inst.* 1979;62(3):493–501.
64. Parkin DM, Bray FI, Devesa SS. Cancer burden in the year 2000. The global picture. *Eur J Cancer.* 2001;37(Suppl 8):S4–66.
65. MacMahon B. Breast cancer at menopausal ages: an explanation of observed incidence changes. *Cancer.* 1957;10(5):1037–1044.
66. Lane-Clayton JE. A further report on cancer of the breast with special reference to its associated antecedent conditions. *Rep Public Health Med Subj (Lond).* 1926;32.
67. MacMahon B, Cole P, Lin TM, et al. Age at first birth and breast cancer risk. *Bull World Health Organ.* 1970;43(2):209–221.



68. Cummings P, Weiss NS, McKnight B, Stanford JL. Estimating the risk of breast cancer in relation to the interval since last term pregnancy. *Epidemiology*. 1997;8(5):488–494.
69. Thompson WD. Age at and time since: modeling temporal aspects of exposure. *Epidemiology*. 1997;8(5):471–473.
70. Woods KL, Smith SR, Morrison JM. Parity and breast cancer: evidence of a dual effect. *Br Med J*. 1980;281(6237):419–421.
71. Lubin JH, Burns PE, Blot WJ, et al. Risk factors for breast cancer in women in northern Alberta, Canada, as related to age at diagnosis. *J Natl Cancer Inst*. 1982;68(2):211–217.
72. Janerich DT, Hoff MB. Evidence for a crossover in breast cancer risk factors. *Am J Epidemiol*. 1982;116(5):737–742.
73. Bruzzi P, Negri E, La Vecchia C, et al. Short term increase in risk of breast cancer after full term pregnancy. *BMJ*. 1988;297(6656):1096–1098.
74. Williams EM, Jones L, Vessey MP, McPherson K. Short term increase in risk of breast cancer associated with full term pregnancy. *BMJ*. 1990;300(6724):578–579.
75. Hsieh C, Pavia M, Lambe M, et al. Dual effect of parity on breast cancer risk. *Eur J Cancer*. 1994;30A(7):969–973.
76. Lambe M, Hsieh C, Trichopoulos D, Ekblom A, Pavia M, Adami HO. Transient increase in the risk of breast cancer after giving birth. *N Engl J Med*. 1994;331(1):5–9.
77. Leon DA, Carpenter LM, Broeders MJ, Gunnarskog J, Murphy MF. Breast cancer in Swedish women before age 50: evidence of a dual effect of completed pregnancy. *Cancer Causes Control*. 1995;6(4):283–291.
78. Albrektsen G, Heuch I, Kvale G. The short-term and long-term effect of a pregnancy on breast cancer risk: a prospective study of 802,457 parous Norwegian women. *Br J Cancer*. 1995;72(2):480–484.
79. Lambe M, Hsieh CC, Tsaih SW, Ekblom A, Trichopoulos D, Adami HO. Parity, age at first birth and the risk of carcinoma in situ of the breast. *Int J Cancer*. 1998;77(3):330–332.
80. Wohlfahrt J, Olsen JH, Melby M. Breast cancer risk after childbirth in young women with family history (Denmark). *Cancer Causes Control*. 2002;13(2):169–174.
81. Pathak DR. Dual effect of first full term pregnancy on breast cancer risk: empirical evidence and postulated underlying biology. *Cancer Causes Control*. 2002;13(4):295–298.
82. Melbye M, Wohlfahrt J, Olsen JH, et al. Induced abortion and the risk of breast cancer. *N Engl J Med*. 1997;336(2):81v85.
83. Schedin P. Pregnancy-associated breast cancer and metastasis. *Nat Rev Cancer*. 2006;6(4):281–291.
84. Lyons TR, Schedin PJ, Borges VF. Pregnancy and breast cancer: when they collide. *J Mammary Gland Biol Neoplasia*. 2009;14(2):87–98.
85. Tiede B, Kang Y. From milk to malignancy: the role of mammary stem cells in development, pregnancy and breast cancer. *Cell Res*. 2011;21(2):245–257.
86. Kobayashi S, Sugiura H, Ando Y, et al. Reproductive history and breast cancer risk. *Breast Cancer*. 2012;19(4):302–308.
87. Colditz GA, Rosner BA, Chen WY, Holmes MD, Hankinson SE. Risk factors for breast cancer according to estrogen and progesterone receptor status. *J Natl Cancer Inst*. 2004;96(3):218–228.
88. Millikan RC, Newman B, Tse CK, et al. Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat*. 2008;123(1):123–139.
89. Shinde SS, Forman MR, Kuerer HM, et al. Higher parity and shorter breastfeeding duration: association with triple-negative phenotype of breast cancer. *Cancer*. 2010;116(21):4933–4943.
90. Yang XR, Chang-Claude J, Goode EL, et al. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the breast cancer association consortium studies. *J Natl Cancer Inst*. 2011;103(3):250–263.
91. Phipps AI, Chlebowski RT, Prentice R, et al. Reproductive history and oral contraceptive use in relation to risk of triple-negative breast cancer. *J Natl Cancer Inst*. 2011;103(6):470–477.
92. Clavel-Chapelon F, Gerber M. Reproductive factors and breast cancer risk. Do they differ according to age at diagnosis? *Breast Cancer Res Treat*. 2002;72(2):107–115.
93. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. *Lancet*. 2002;360(9328):187–195.
94. Ma H, Bernstein L, Pike MC, Ursin G. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. *Breast Cancer Res*. 2006;8(4):R43.
95. 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(7):1521–1526.
96. Gaudet MM, Press MF, Haile RW, et al. Risk factors by molecular subtypes of breast cancer across a population-based study of women 56 years or younger. *Breast Cancer Res Treat*. 2011;130(2):587–597.
97. Kotsopoulos J, Lubinski J, Salmena L, et al. Breastfeeding and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. *Breast Cancer Res*. 2012;14(2):R42.
98. Li CI, Beaber EF, Tang MT, Porter PL, Daling JR, Malone KE. Reproductive factors and risk of estrogen receptor positive, triple-negative, and HER2-neu overexpressing breast cancer among women 20–44 years of age. *Breast Cancer Res Treat*. 2013;137(2):579–587.
99. 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(8):2078–2086.
100. Cleary MP, Maihle NJ. The role of body mass index in the relative risk of developing premenopausal versus postmenopausal breast cancer. *Proc Soc Exp Biol Med*. 1997;216(1):28–43.
101. Wu Q, Ishikawa T, Sirianni R, et al. 27-Hydroxycholesterol promotes cell-autonomous, er-positive breast cancer growth. *Cell Rep*. 2013;5(3):637–645.
102. Nelson ER, Wardell SE, Jasper JS, et al. 27-Hydroxycholesterol links hypercholesterolemia and breast cancer pathophysiology. *Science*. 2013;342(6162):1094–1098.
103. Warner M, Gustafsson JA. On estrogen, cholesterol metabolism, and breast cancer. *N Engl J Med*. 2014;370(6):572–573.
104. Huang Z, Hankinson SE, Colditz GA, et al. Dual effects of weight and weight gain on breast cancer risk. *JAMA*. 1997;278(17):1407–1411.
105. Lahmann PH, Hoffmann K, Allen N, et al. Body size and breast cancer risk: findings from the European Prospective Investigation into Cancer And Nutrition (EPIC). *Int J Cancer*. 2004;111(5):762–771.
106. Carmichael AR, Bates T. Obesity and breast cancer: a review of the literature. *Breast (Edinburgh, Scotland)*. 2004;13(2):85–92.
107. Harvie M, Howell A, Vierkant RA, et al. Association of gain and loss of weight before and after menopause with risk of postmenopausal breast cancer in the Iowa women's health study. *Cancer Epidemiol Biomarkers Prev*. 2005;14(3):656–661.
108. Eliassen AH, Colditz GA, Rosner B, Willett WC, Hankinson SE. Adult weight change and risk of postmenopausal breast cancer. *JAMA*. 2006;296(2):193–201.
109. Teras LR, Goodman M, Patel AV, Diver WR, Flanders WD, Feigelson HS. Weight loss and postmenopausal breast cancer in a prospective cohort of overweight and obese US women. *Cancer Causes Control*. 2011;22(4):573–579.
110. Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME. Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. *Cancer Epidemiol Biomarkers Prev*. 2004;13(10):1558–1568.
111. Chang S, Buzdar AU, Hursting SD. Inflammatory breast cancer and body mass index. *J Clin Oncol*. 1998;16(12):3731–3735.
112. Schairer C, Li Y, Frawley P, et al. Risk factors for inflammatory breast cancer and other invasive breast cancer. *J Natl Cancer Inst*. 2013;105(18):1373–1384.
113. Tarone RE. Breast cancer trends among young women in the United States. *Epidemiology*. 2006;17(5):588–590.
114. Lidegarrd O, Kroman N. The epidemiology of breast cancer. *Eur Clinics Obstet Gynaecol*. 2005;1(1):24–28.
115. Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the Human Development Index (2008–2030): a population-based study. *Lancet Oncol*. 2012;13(8):790–801.
116. Karp SE, Tonin PN, Begin LR, et al. Influence of BRCA1 mutations on nuclear grade and estrogen receptor status of breast carcinoma in Ashkenazi Jewish women. *Cancer*. 1997;80(3):435–441.

117. Loman N, Johannsson O, Bendahl PO, Borg A, Ferno M, Olsson H. Steroid receptors in hereditary breast carcinomas associated with BRCA1 or BRCA2 mutations or unknown susceptibility genes. *Cancer*. 1998;83(2):310–319.
118. Foulkes WD, Metcalfe K, Sun P, et al. Estrogen receptor status in BRCA1- and BRCA2-related breast cancer: the influence of age, grade, and histological type. *Clin Cancer Res*. 2004;10(6):2029–2034.
119. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. 2013;45(4):353–361.
120. Garcia-Closas M, Hall P, Nevanlinna H, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet*. 2008;4(4):e1000054.
121. Garcia-Closas M, Chanock S. Genetic susceptibility loci for breast cancer by estrogen receptor status. *Clin Cancer Res*. 2008;14(24):8000–8009.
122. Li J, Humphreys K, Darabi H, et al. A genome-wide association scan on estrogen receptor-negative breast cancer. *Breast Cancer Res*. 2010;12(6):R93.
123. Figueroa JD, Garcia-Closas M, Humphreys M, et al. Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. *Hum Mol Genet*. 2011;20(23):4693–4706.
124. Broeks A, Schmidt MK, Sherman ME, et al. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum Mol Genet*. 2011;20(16):3289–3303.
125. Haiman CA, Chen GK, Vachon CM, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat Genet*. 2011;43(12):1210–1214.
126. Hein R, Maranian M, Hopper JL, et al. Comparison of 6q25 breast cancer hits from Asian and European genome wide association studies in the Breast Cancer Association Consortium (BCAC). *PLoS One*. 2012;7(8):e42380.
127. Lambrechts D, Truong T, Justenhoven C, et al. 11q13 is a susceptibility locus for hormone receptor positive breast cancer. *Hum Mutat*. 2012;33(7):1123–1132.
128. Stevens KN, Fredericksen Z, Vachon CM, et al. 19p13.1 is a triple-negative-specific breast cancer susceptibility locus. *Cancer Res*. 2012;72(7):1795–1803.
129. Garcia-Closas M, Couch FJ, Lindstrom S, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet*. 2013;45(4):392–398.
130. Bojesen SE, Pooley KA, Johnatty SE, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet*. 2013;45(4):371–384.
131. Stevens KN, Vachon CM, Couch FJ. Genetic susceptibility to triple-negative breast cancer. *Cancer Res*. 2013;73(7):2025–2030.
132. Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet*. 2011;378(9805):1812–1823.
133. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–752.
134. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869–10874.
135. Chung CH, Bernard PS, Perou CM. Molecular portraits and the family tree of cancer. *Nat Genet*. 2002;32(Suppl):533–540.
136. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*. 2003;100(14):8418–8423.
137. Wirapati P, Sotiriou C, Kunkel S, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res*. 2008;10(4):R65.
138. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61–70.
139. Livasy CA, Perou CM, Karaca G, et al. Identification of a basal-like subtype of breast ductal carcinoma in situ. *Hum Pathol*. 2007;38(2):197–204.
140. Livasy CA, Karaca G, Nanda R, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol*. 2006;19(2):264–271.
141. Tamimi RM, Baer HJ, Marotti J, et al. Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer. *Breast Cancer Res*. 2008;10(4):R67.
142. 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(16):5508–5517.
143. Calza S, Hall P, Auer G, et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. *Breast Cancer Res*. 2006;8(4):R34.
144. Ihemelandu CU, Leffall LD Jr, Dewitty RL, et al. Molecular breast cancer subtypes in premenopausal and postmenopausal African-American women: age-specific prevalence and survival. *J Surg Res*. 2007;143(1):109–118.
145. Winer EP, Carey LA, Dowsett M, Tripathy D. Beyond anatomic staging: are we ready to take the leap to molecular classification? In: Perry MC, ed. *American Society of Clinical Oncology Education Book*. Alexandria, VA: American Society of Clinical Oncology; 2005:46–59.
146. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med*. 2006;355(6):560–569.
147. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160–1167.
148. Jatoi I, Anderson WF, Jeong JH, Redmond CK. Breast cancer adjuvant therapy: time to consider its time-dependent effects. *J Clin Oncol*. 2011;29(17):2301–2304.
149. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295(21):2492–2502.
150. Kim MJ, Ro JY, Ahn SH, Kim HH, Kim SB, Gong G. Clinicopathologic significance of the basal-like subtype of breast cancer: a comparison with hormone receptor and Her2/neu-overexpressing phenotypes. *Hum Pathol*. 2006;37(9):1217–1226.
151. Carey LA, Dees EC, Sawyer L, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res*. 2007;13(8):2329–2334.
152. Rouzier R, Perou CM, Symmans WF, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res*. 2005;11(16):5678–5685.
153. Wo JY, Taghian AG, Nguyen PL, et al. The association between biological subtype and isolated regional nodal failure after breast-conserving therapy. *Int J Radiat Oncol Biol Phys*. 2010;77(1):188–196.
154. Prat A, Ellis MJ, Perou CM. Practical implications of gene-expression-based assays for breast oncologists. *Nat Rev Clin Oncol*. 2012;9(1):48–57.
155. Haque R, Ahmed SA, Inzhakova G, et al. Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades. *Cancer Epidemiol Biomarkers Prev*. 2012;21(10):1848–1855.
156. Ellis MJ, Perou CM. The genomic landscape of breast cancer as a therapeutic roadmap. *Cancer Disc*. 2013;3(1):27–34.
157. Goldhirsch A, Wood WC, Coates AS, et al. 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(8):1736–1747.
158. Prat A, Adamo B, Cheang MC, Anders CK, Carey LA, Perou CM. Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. *Oncologist*. 2013;18(2):123–133.
159. Prat A, Cheang MC, Martin M, et al. Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal a breast cancer. *J Clin Oncol*. 2013;31(2):203–209.
160. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61–70.
161. Ciriello G, Sinha R, Hoadley KA, et al. The molecular diversity of luminal A breast tumors. *Breast Cancer Res Treat*. 2013;141(3):409–420.
162. Sorlie T, Wang Y, Xiao C, et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. *BMC Genomics*. 2006;7(1):127.
163. Sorlie T. Introducing molecular subtyping of breast cancer into the clinic? *J Clin Oncol*. 2009;27(8):1153–1154.
164. Gierach GL, Burke A, Anderson WF. Epidemiology of triple negative breast cancers. *Breast Dis*. 2010;32(1–2):5–24.
165. Perou CM. Molecular stratification of triple-negative breast cancers. *Oncologist*. 2010;15(Suppl 5):39–48.
166. Lusa L, McShane LM, Reid JF, et al. Challenges in projecting clustering results across gene expression-profiling datasets. *J Natl Cancer Inst*. 2007;99(22):1715–1723.

167. Weigelt B, Mackay A, A'Hern R, et al. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *Lancet Oncol*. Apr 2010;11(4):339–349.
168. Perou CM, Parker JS, Prat A, Ellis MJ, Bernard PS. Clinical implementation of the intrinsic subtypes of breast cancer. *Lancet Oncol*. 2010;11(8):718–719; author reply 720–721.
169. Sorlie T, Borgan E, Myhre S, et al. The importance of gene-centring microarray data. *Lancet Oncol*. 2010;11(8):719–720; author reply 720–721.
170. Hoadley KA, Yau C, Wolf DM, et al. Multi-platform integration of 12 cancer types reveals cell-of-origin with distinct molecular signatures. *Cell*. 2014; in press.
171. Ghabach B, Anderson WF, Curtis RE, Huycke MM, Lavigne JA, Dores GM. Adenoid cystic carcinoma of the breast in the United States (1977–2006): a population-based cohort study. *Breast Cancer Res*. 2012;12(4):R54.
172. Prat A, Parker JS, Fan C, Perou CM. PAM50 assay and the three-gene model for identifying the major and clinically relevant molecular subtypes of breast cancer. *Breast Cancer Res Treat*. 2012;135(1):301–306.
173. Anderson WF, Luo S, Chatterjee N, et al. Human epidermal growth factor receptor-2 and estrogen receptor expression, a demonstration project using the residual tissue repository of the Surveillance, Epidemiology, and End Results (SEER) program. *Breast Cancer Res Treat*. 2009;113(1):189–196.
174. Telli ML, Chang ET, Kurian AW, et al. Asian ethnicity and breast cancer subtypes: a study from the California Cancer Registry. *Breast Cancer Res Treat*. 2011;127(2):471–478.
175. Kurian AW, Fish K, Shema SJ, Clarke CA. Lifetime risks of specific breast cancer subtypes among women in four racial/ethnic groups. *Breast Cancer Res*. 2010;12(6):R99.
176. Keegan TH, DeRouen MC, Press DJ, Kurian AW, Clarke CA. Occurrence of breast cancer subtypes in adolescent and young adult women. *Breast Cancer Res*. 2012;14(2):R55.
177. Clarke CA, Keegan TH, Yang J, et al. Age-specific incidence of breast cancer subtypes: understanding the black-white crossover. *J Natl Cancer Inst*. 2012;104(14):1094–1101.
178. SEER-18. Surveillance, Epidemiology, and End Results (SEER) program ([www.seer.cancer.gov](http://www.seer.cancer.gov)) SEER\*Stat Database: Incidence - SEER 18 Regs Custom Data (includes additional CS SSF data for 2010 cases), Nov 2012 Sub (2000–2010) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969–2011 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released May 2013, based on November 2012 submission. 2012. [www.seer.cancer.gov](http://www.seer.cancer.gov). Accessed May 1, 2014.
179. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346–352.
180. Dawson SJ, Rueda OM, Aparicio S, Caldas C. A new genome-driven integrated classification of breast cancer and its implications. *EMBO J*. 2013;32(5):617–628.
181. Boecker W, Buerger H. Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept. *Cell Prolif*. 2003;36(Suppl 1):73–84.
182. Goodman MT, Hernandez BY, Hewitt S, et al. Tissues from population-based cancer registries: a novel approach to increasing research potential. *Hum Pathol*. Jul 2005;36(7):812–820.
183. Chaturvedi AK, Engels EA, Pfeiffer RM, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol*. 2011;29(32):4294–4301.
184. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast And Bowel Project P-1 Study. *J Natl Cancer Inst*. 1998;90(18):1371–1388.
185. Osborne CK, Schiff R. Estrogen-receptor biology: continuing progress and therapeutic implications. *J Clin Oncol*. 2005;23(8):1616–1622.
186. Anderson WF, Reiner AS, Matsuno RK, Pfeiffer RM. Shifting breast cancer trends in the United States. *J Clin Oncol*. 2007;25(25):3923–3929.
187. Rosenberg PS. Hazard function estimation using B-splines. *Biometrics*. 1995;51(3):874–887.
188. Akaike H. Information theory and an extension of the maximum likelihood principle. Paper presented at 2nd International Symposium on Information Theory; 1973; Budapest, Hungary.
189. Efron B, Tibshirani R. *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability 57*. New York: Chapman & Hall; 1993.

### Funding

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Cancer Institute (NCI), and by extramural funds to CMP (NCI Breast SPORE program; P50-CA58223-09A1).

### Notes

The authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of reported findings. CMP is an equity stock holder and board of director member of BioClassifier LLC and University Genomics. CMP is also listed an inventor on a patent application on the PAM50 molecular assay.

We wish to thank Dr Hyuna Sung and Mr David Check, Department of Health and Human Services/NIH/NCI/Division of Cancer Epidemiology and Genetics for advice and enhancements for Figure 4.

**Affiliations of authors:** Division of Cancer Epidemiology and Genetics Biostatistics Branch (WFA, PSR), and Division of Cancer Prevention (MES), National Cancer Institute, National Institutes of Health, Bethesda, MD; Translational Genomics Group, Vall d'Hebron Institute of Oncology, Barcelona, Spain (AP); Department of Genetics and Pathology & Laboratory Medicine, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC (CMP).