Breast Cancer Res Treat (2011) 128:483–493
DOI 10.1007/s10549-011-1347-z

EPIDEMIOLOGY

Breast cancer as heterogeneous disease: contributing factors and carcinogenesis mechanisms

Julia Kravchenko · Igor Akushevich · Victoria L. Seewaldt · Amy P. Abernethy · H. Kim Lyerly

Received: 6 November 2010 / Accepted: 4 January 2011 / Published online: 12 January 2011

Abstract The observed bimodal patterns of breast cancer incidence in the U.S. suggested that breast cancer may be viewed as more than one biological entity. We studied the factors potentially contributing to this phenomenon, specifically focusing on how disease heterogeneity could be linked to breast carcinogenesis mechanisms. Using empirical analyses and population-based biologically motivated modeling, age-specific patterns of incidence of ductal and lobular breast carcinomas from the SEER registry (1990–2003) were analyzed for heterogeneity and characteristics of carcinogenesis, stratified by race, stage, grade, and estrogen (ER)/progesterone (PR) receptor status. The heterogeneity of breast carcinoma age patterns decreased after stratification by grade, especially for grade I and III tumors. Stratification by ER/PR status further reduced the heterogeneity, especially for ER(+)/PR(−) and ER(−)/(−) tumors; however, the residual heterogeneity was still observed. The number of rate-limiting events of carcinogenesis and the latency of ductal and lobular carcinomas differed, decreasing from grade I to III, with poorly differentiated tumors associated with the least number of carcinogenesis stages and the shortest latency. Tumor grades play important role in bimodal incidence of breast carcinoma and have distinct mechanisms of carcinogenesis. Race and cancer subtype could play modifying role. ER/PR status contributes to the observed heterogeneity, but is subdominant to tumor grade. Further studies on sources of "remaining" heterogeneity of population with breast cancer (such as genetic/epigenetic characteristics) are necessary. The results of this study could suggest stratification rather than unification of breast cancer prevention strategies, risk assessment, and treatment.

Keywords Breast lobular carcinoma · Breast ductal carcinoma · Age patterns · Grade · Estrogen receptor · Progesterone receptor · Carcinogenesis

Introduction

Multiple clinical, pathological, and molecular analyses support the theory that breast cancer is a heterogeneous disease [1–3]. In 1957, Armitage and Doll [4] demonstrated that the age-specific patterns of incidence of some solid cancers had linear slopes on a logarithmic scale. Later, two peaks were identified in the slopes of several solid cancers, including breast cancer, suggesting that at least a two-disease model was required to describe each component [3, 5–7]. The bimodal age pattern of breast cancer incidence with the Clemmesen's hook corresponding to the dip between the bimodal peaks [9] suggested the existence of two different rate curves—specifically of early and late forms of breast cancer [8–11]. Recently, two peaks of breast cancer incidence were observed in patients at the ages of 50 and 70 years in the categories of histopathological subtypes,

Electronic supplementary material The online version of this article (doi:10.1007/s10549-011-1347-z) contains supplementary material, which is available to authorized users.

J. Kravchenko (✉) · A. P. Abernethy · H. K. Lyerly
Duke Cancer Institute, Duke University Medical Center (DUMC), 2424 Erwin Road, Box#2732, Hock Plaza, Suite G05, Durham, NC 27705, USA
e-mail: julia.krauchanka@duke.edu

I. Akushevich
Center for Population Health and Aging, Duke University, Durham, NC, USA

V. L. Seewaldt · A. P. Abernethy
Division of Medical Oncology, Department of Medicine, Duke University Medical Center, Durham, NC, USA

hormone receptor status, tumor characteristics, and molecular signatures, suggesting a link between breast cancer etiology and outcome [3, 5, 12–14].

It is still not well understood which factors contribute to the Clemmesen's hook and how the bimodality of breast carcinoma patterns could be linked to specific characteristics of breast carcinogenesis in humans. As the results obtained from animal and in vitro experiments cannot fully represent the processes occurring in humans, population-based modeling of breast carcinogenesis might be a useful option. We hypothesized that the observed bimodality in the age patterns of breast cancer incidence is a multi-component phenomenon (i.e., multiple factors can contribute to it, and the contribution of each factor can be more or less pronounced), and these factors could also be linked to breast carcinogenesis mechanisms. Thus, it is possible to build a "bridge" between the heterogeneity of a population of patients with breast cancer (e.g., by age, race, risk factors, or genetic/epigenetic characteristics) and tumor heterogeneity due to the different mechanisms of carcinogenesis.

Data and methods

Data

The patterns of incidence in breast carcinomas as they relate to patient age were analyzed using SEER registry data. This program has collected data on ER and PR statuses since 1990; therefore, we performed analyses on 1990–2003 data. The codes 850 and 852 (ICD-O-2) were used for ductal and lobular breast carcinomas, respectively. Grade I (well differentiated), grade II (intermediate differentiation), and grade III (poorly differentiated) tumors were analyzed. Hormone receptor status was coded as estrogen receptor-positive [ER(+)], progesterone receptor-positive [PR(+)], receptor negative [ER(−), PR(−)], missing, borderline, or unknown. For this study, the missing, borderline, or unknown data were combined into one "Unknown" group to control for data completeness.

Statistical methods and carcinogenesis modeling

The incidence rates were calculated per 100,000 person-years. The age structure of the SEER registry population in 2000 was chosen as a standard population. We considered one-year age-specific incidence rates for ductal and lobular carcinomas for Caucasian and African-American females.

The age-specific incidence rates were analyzed using the Armitage-Doll model with a random frailty, which takes into account individual predisposition to cancer. This model demonstrated a better fit than others (such as the two-stage clonal expansion model and models with hidden frailty); for a detailed description of the model selection procedure, see Section 7.3 in Manton et al. [7]. To evaluate the model fit for age-specific patterns of incidence, the residuals were analyzed; the fit was considered good when all residuals fluctuated randomly around zero, without abnormally large values (i.e., staying between −2 and +2 on the Y scale), age periods with regular (i.e., non-stochastic) behavior, and large periods with residuals of the same sign (i.e., plus/minus).

The age patterns of tumor incidence were analyzed for homogeneity. Homogeneous patterns (total and histotype-, grade-, and hormone receptor status-specific) were defined as patterns that could be well-described by the model we used for analysis with a good fit. In an "ideal" case, a homogeneous population could be described as a population of individuals with exactly the same cancer risk predisposition/susceptibility, as was done in the classic Armitage-Doll model or its extensions, the Moolgavkar-Venzon-Knudson-type (MVK) models. However, these models failed to describe the age patterns of cancer incidence in patients of advanced age (80+ years), where a decline in incidence is observed. This decline can be reproduced considering that individuals in this population are distributed in accordance with their individual predisposition/susceptibility to cancer risk. Therefore, when observed age patterns could not be described by the Armitage-Doll model with a good quality of fit, the population is considered heterogeneous. This indicates that the population could consist of two or more subpopulations that differ in cancer risk, resulting in the bimodal distribution of cancer incidence. This points toward the existence of distinct subpopulations with specific genetic or epigenetic characteristics or patients that differed in their exposure to a specific risk factor—this group gives rise to a smaller peak in a bimodal pattern.

In this study, ductal carcinoma was chosen for detailed analysis because it provided a number of cases large enough for stratification by each studied factor. A step-by-step analysis of the age-specific patterns of incidence was performed. Step 1 included an analysis of stage-specific patterns of cancer incidence and how stage contributes to the observed heterogeneity of age patterns. For step 2, analyses were performed for race-stage-specific patterns to evaluate the contribution of race. For step 3, analyses were stratified by grade, race, and stage to study the role of tumor grade. Finally, step 4 included analyses of ER/PR status, grade, race, and stage to estimate the role that ER/PR status plays in the bimodality phenomenon. At each step, decreases in heterogeneity were evaluated by analyzing the residuals and by estimating the significance of the results using the chi-square test for non-linear least squares, i.e., by evaluating χ^2 per degree of freedom (DOF).

To investigate how heterogeneity in age-related patterns of breast cancer incidence could be linked to the mechanisms of breast carcinogenesis, a five-parameter version of the frailty model with a Weibull baseline was used to obtain the parameters characterizing carcinogenesis mechanisms:

$$I(x) = \frac{(x-l)^{m-1}}{c^m(1+n\sigma^2 c^{-m}m^{-1}(x-l)^m)^{1/n}} \quad (1)$$

In this equation, m was the number corresponding to carcinogenesis stage ("m"—from "*m*alignant") or the number of rate-limiting events from cell initiation by exposure to carcinogen and to the moment when the malignant cell appeared; l (in years) represents the *lag* period (i.e., the period between the occurrence of first malignant cell and the date of cancer onset); c (in years) indicates the scale parameter related to the maximum age in cancer incidence age pattern; σ^2 stands for the frailty distribution reflecting individual susceptibility to cancer risk; and n represents the parameter describing the shape of frailty distribution ($n = 1$, 2, and 0 correspond to gamma-distribution, inverse Gaussian distribution, and the distribution suggested in Manton et al. [7], respectively; for $n \leq 1$, the model has a maximum with the age equaling $l + c(m(m-1)\times(n+m-mn)^{-1}\sigma^{-2})^{1/m}$. The initial age for modeling was identified as the minimal age of cancer incidence based on the results of empirical analysis.

In addition to the base model (1), two simplified models were also used. The first was a model with n fixed to one ($n = 1$), which corresponded to gamma-distributed frailty. The second was a model where *lag* was fixed at 20 years with the $n = 1$ condition ($lag = 20$, $n = 1$). Thus, to control for overparameterization, three scenarios were used: (1) all parameters were "free" or non-fixed; (2) $n = 1$; and (3) $n = 1$ and $lag = 20$. The results for the carcinogenesis parameters obtained from the analysis for all three scenarios were evaluated for their significance using the chi-square test.

Results

Grade- and hormone status-specific frequencies of ductal and lobular breast carcinomas are presented in Table 1. For both types of carcinomas, among ER(+)/PR(+) tumors, grades I and II were diagnosed more often, and among ER(−)/PR(−) tumors, grade III was more commonly diagnosed. With increasing age (from 25 to 85 years), the frequencies of grade I and grade II tumors increased by 6–15%, and for ER(+) tumors, the frequency increased by 12–26%, whereas grade III carcinomas decreased by 12–25% and ER(−) tumors decreased by 22–25% (percents varied by race and cancer subtype). Poorly differentiated tumors were more commonly diagnosed in younger (<45 years) African-American females (in 44–54% of ductal cases and in 18–33% of lobular carcinomas; percents varied by age). This population group also displayed ER(−) ductal tumors more frequently than Caucasian females (21–30 vs. 11–19%, respectively).

The frequencies of the different stages at diagnosis in 1990–2003 are presented in Table 2. For both subtypes, grade I tumors were more often diagnosed at earlier stages, whereas grade III tumors were most commonly diagnosed at advanced stages. From 1990 to 2003, the frequencies of in situ tumors of grade II and III slightly increased, especially for lobular carcinomas, whereas the distribution of other stages remained almost unchanged.

Analyses of heterogeneity in carcinoma incidence patterns

Age patterns of incidence of ductal and lobular invasive carcinomas were analyzed grade-specifically (Fig. 1). In patients with ductal carcinomas, grade III prevailed over grade I, especially at younger ages, which suggests an important role for the formation of the first component of the bimodal pattern. In patients with lobular carcinomas, the slight predominance of grade III tumors over grade I was observed in patients younger than 50. Grade II tumors were observed at a higher rate than grade I or III in women older than 50 years for ductal, and at all ages for lobular carcinomas.

Age patterns of ductal carcinoma incidence were analyzed for all stages together and separately for invasive carcinomas using the five-parameter model (1). A one-disease model was applied to the incidence curve to emphasize the presence of the second component; a substantial increase in residuals was noted in patients aged 30–50, suggesting the existence of an early-onset component (Fig. 2). In invasive-alone cases, the Clemmesen's hook became slightly less pronounced than for the all-stage analysis; however, the decrease in the first component (i.e., "early-onset" cancer) was not significant (Fig. 2) (note that the "early-onset" component was excluded from the optimization procedure while estimating the chi-square value to describe the fit of the one-disease model).

Next, the age patterns of invasive ductal carcinoma were analyzed according to race using a model with fixed $n = 1$ and $lag = 20$ for Caucasians and African-American combined and for Caucasian females alone. For Caucasian females, the residuals had a tendency to become less pronounced, which suggests that race could contribute to breast cancer bimodality; however, the chi-square values were still high (Fig. 3).

Further analyses were performed grade-specifically. Heterogeneity decreased for all three grades compared to

Table 1 Frequencies of the grades and hormone receptor status for ductal and lobular breast carcinoma, for both races, SEER registry, 1990–2003

Characteristics	Ductal carcinoma		Lobular carcinoma	
	All stages (%)	Invasive only (%)	All stages (%)	Invasive only (%)
Grade				
Grade I	13.75	14.81	13.20	14.98
Grade II	34.71	37.16	31.32	36.08
Grade III	31.46	34.64	14.33	16.54
Not determined/not stated/n/a	16.27	11.12	39.63	30.93
ER status				
ER(+)	52.82	61.19	63.93	75.75
ER(−)	17.34	20.12	5.85	6.81
None done/not in chart/unknown/ no information	29.40	18.18	29.94	17.10
PR status				
PR(+)	44.40	51.46	51.41	60.85
PR(−)	23.81	27.66	15.42	18.23
None done/not in chart/unknown/ no information	31.13	20.10	32.53	20.17
ER(+)				
Grade I	72.08	79.10	77.77	81.95
Grade II	66.30	73.08	78.20	81.29
Grade III	43.12	46.11	67.95	70.30
ER(−)				
Grade I	4.09	4.35	3.43	3.58
Grade II	9.73	10.73	4.98	5.15
Grade III	34.65	37.44	13.16	13.58
PR(+)				
Grade I	60.96	66.85	61.04	64.26
Grade II	55.99	61.75	63.89	66.40
Grade III	35.71	38.24	54.97	56.91
PR(−)				
Grade I	12.40	13.54	15.24	16.05
Grade II	17.74	19.58	15.69	16.33
Grade III	40.52	43.73	23.55	24.32

non-grade-specific analysis (Fig. 4 compared to Fig. 3): the most pronounced decrease was observed for grade III (χ^2 1.89), followed by grade I and grade II tumors.

Finally, the analysis was stratified by ER/PR status. The results from the most common combinations, such as ER(+)/PR(+), ER(+)/PR(−), and ER(−)/PR(−), are shown in Fig. 5: the most "improved" homogeneity (compared to non-ER/PR-specific analysis) was registered for ER(+)/PR(−) tumors (χ^2 1.01, 1.42, and 1.23 for grades I, II, and III, respectively) and for ER(−)/PR(−) tumors (χ^2 0.79, 1.09, and 1.79, respectively). There was still heterogeneity that remained among the ER(+)/PR(+) tumors, especially for those of grade II.

Therefore, heterogeneity decreased substantially (see the residuals and chi-square test) when the analysis was specified by grade; additionally, ER/PR status contributed to "decreasing" heterogeneity, but to a less extent than grade. The contributions of stage and race were not statistically significant, but they played roles in reducing the observed bimodality of breast cancer incidence. Note that the remaining minimal heterogeneity observed for ER/PR-specific patterns suggested the effects of other factors beyond the scope of this study.

Carcinogenesis modeling: assessments of characteristics

To investigate whether factors contributing to breast cancer bimodality could be linked to specific mechanisms of carcinogenesis, analyses of breast carcinogenesis mechanisms associated with empirically observed age-specific patterns were performed, and these analyses took into

Table 2 Frequencies of the stages at breast ductal and lobular carcinomas diagnoses, SEER registry, 1990–2003

Cancer	Grade	Year	In situ (%)	Localized (%)	Regional (%)	Distant (%)	Unstaged (%)
Ductal carcinoma	All grades	1990	12.8	54.5	27.2	4.0	1.5
		2003	18.9	50.5	26.2	3.4	1.0
	Grade I	1990	8.9	72.4	17.2	1.5	0.1
		2003	11.7	70.7	16.0	0.8	0.7
	Grade II	1990	1.2	65.2	29.3	3.1	1.3
		2003	15.2	54.7	26.8	2.5	0.8
	Grade III	1990	0.9	50.4	40.6	6.6	1.6
		2003	12.5	45.9	35.6	5.2	0.9
Lobular carcinoma	All grades	1990	20.8	48.6	25.8	3.4	1.3
		2003	18	47.6	29.4	4.2	0.8
	Grade I	1990	6.3	73.0	20.6	–	–
		2003	6.6	64.5	25.7	2.8	0.3
	Grade II	1990	0.5	65.4	28.1	5.4	0.5
		2003	5.7	54.4	36	3.4	0.6
	Grade III	1990	–	47.9	42.7	8.9	0.5
		2003	6.8	44.1	41.8	6.7	0.5

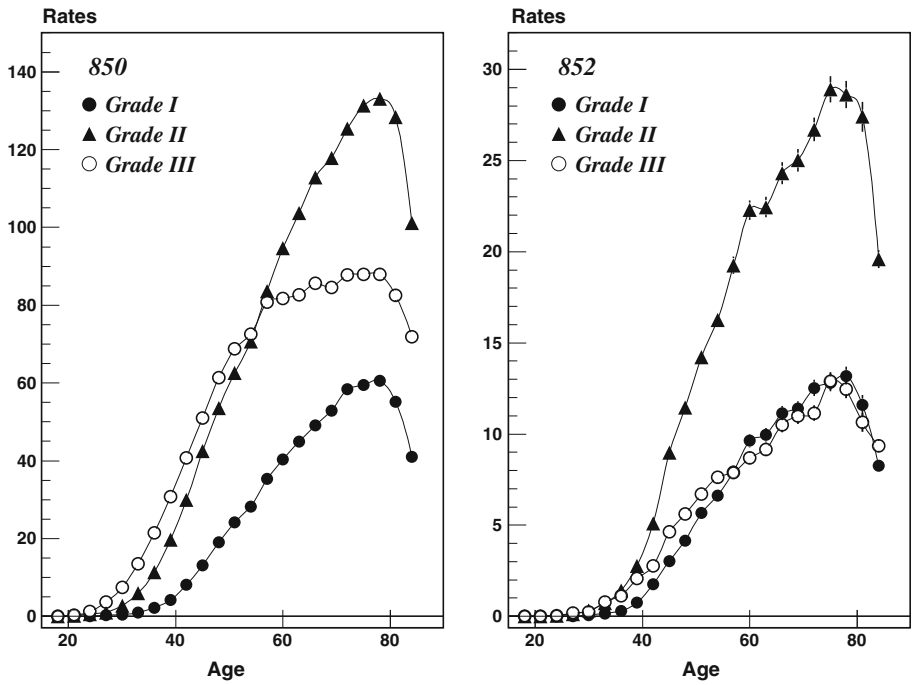

Fig. 1 Grade-specific age patterns of invasive breast ductal (*left*) and lobular (*right*) carcinoma incidence for both races, SEER registry data, 1990–2003

account cancer subtype, race, grade, and ER/PR status (see the Electronic supplemental Table for the detailed results). To avoid the possible effects of a correlation between *m*-stages and *lag*, the scenario of $n = 1$ and $lag = 20$ was used. For both cancer subtypes in both races, decreases were observed for *m*-stages with a gradient from grade I (the highest number of *m*-stages) to grade III (the least number of *m*-stages) tumors. No statistically significant differences in the number of *m*-stages were observed among all seven combinations of ER/PR status for the cancer subtypes. Due to the absence of a modifying effect on carcinogenesis parameters caused by hormone receptor status, the subtype-, race-, and grade-specific characteristics of carcinogenesis were selected for further detailed analysis.

The option with $n = 1$ was used for latency period analyses (see Table 3). A decrease in *m*-stages and shorter latency were observed from grade I to grade III for both

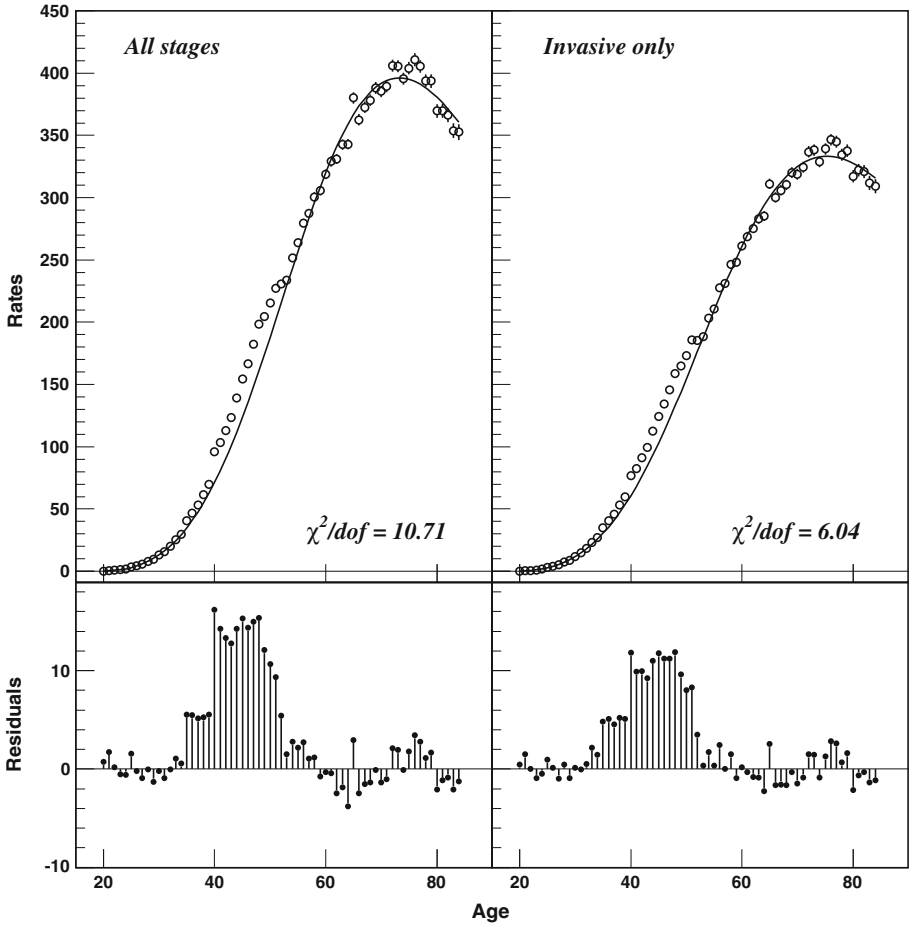

Fig. 2 Age patterns of breast ductal carcinoma incidence and model fit (for residuals) for all stages (*left*) and for invasive only (*right*), Caucasian females, SEER registry data, 1990–2003

cancer subtypes in both races. Well-differentiated tumors required more *m*-stages for their development and had a longer latency than intermediate grade tumors, and, in particular, poorly differentiated tumors. Depending on subtype and race, the difference between grade I and grade III tumors varied from 1.21 to 1.92 for *m*-stages (grade I—from 4.47 ± 0.14 to 5.94 ± 0.40 vs. grade III—from 3.26 ± 0.07 to 4.02 ± 0.23). Poorly differentiated tumors had the shortest latency. Depending on histotype and race, grade III tumors had a *lag* of about 5–10 years shorter than grade I tumors (i.e., for grade III ductal carcinomas—from 14.9 ± 1.8 to 19.2 ± 0.6 years, vs. grade I—24.3 ± 1.5 and 24.9 ± 2.8 years).

Breast lobular carcinomas tended to have more *m*-stages and longer latency than ductal tumors (see Table 3). However, for *m*-stages, the differences in most cases were less than one stage. With regard to latency, subtype-specific differences were observed predominantly in African-American females. In these patients, the *lag* period was about 5.3–13.4 years longer for lobular than ductal carcinoma, and it varied by grade.

Ductal carcinomas in African-American females tended to have fewer *m*-stages and shorter latencies than those in Caucasians, whereas the opposite association was observed for patients with lobular carcinomas (see Table 3). The race-specific differences for *m*-stages were less than one stage for both subtypes. Caucasian females had a longer latency period for ductal carcinomas by about 4.3 years than African-Americans and a shorter latency for lobular carcinomas by about 6.1–7.2 years.

Therefore, the differences between the grades of differentiation suggested that each grade could be described as having its own number of *m*-stages and latency. Race and cancer subtype may also differ in their characteristics of carcinogenesis, particularly by latency period, suggesting that these factors can modulate the mechanisms of carcinogenesis; however, these factors were less important than tumor grade.

Discussion

The bimodality of age-specific patterns of breast carcinoma incidence observed in our study was in agreement with other studies describing breast cancer as a disease with "early-" and "late-onset" components [1–3]. Our study

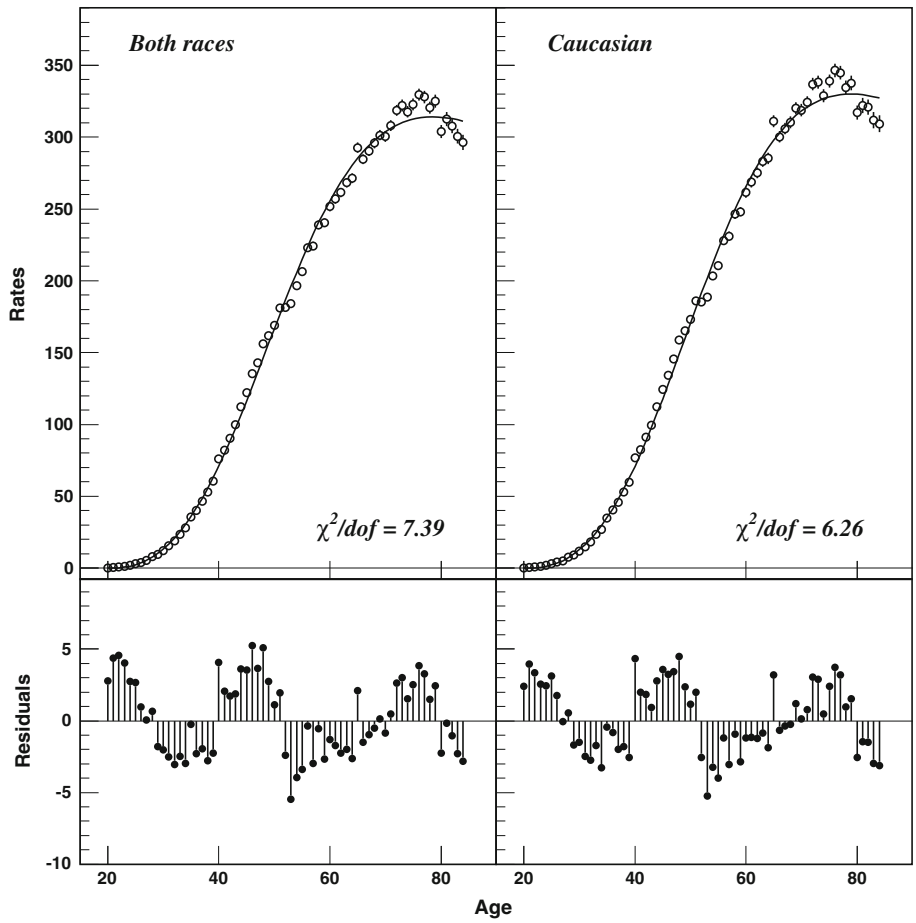

Fig. 3 All-grade/all-ER/all-PR age patterns of breast ductal carcinoma incidence and model fit (i.e., residuals) for both races (*left*) and for Caucasian females only (*right*), SEER registry data, 1990–2003

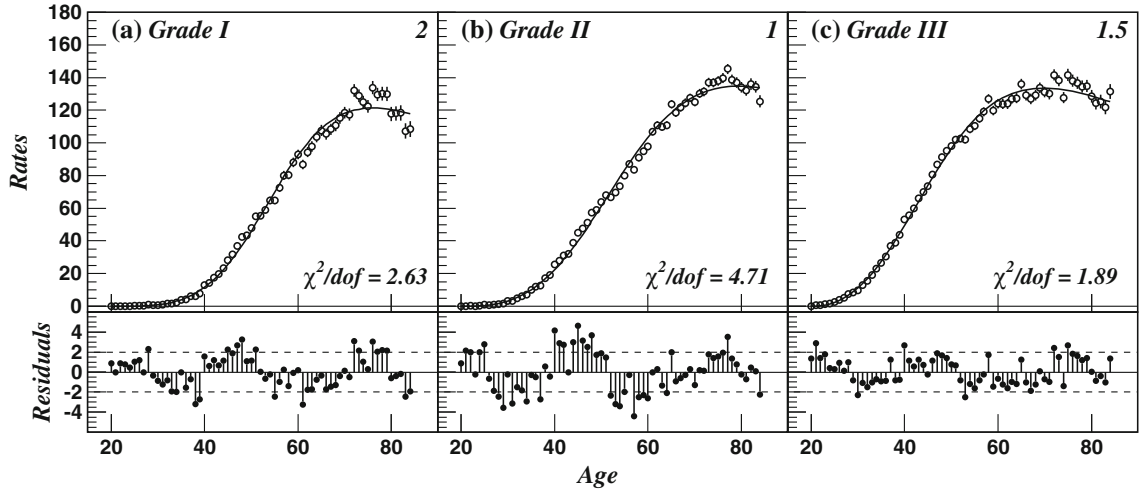

Fig. 4 Grade-specific age patterns of breast ductal carcinoma incidence and model fit (i.e., residuals) for Caucasian females for all-ER/all-PR cases for grade I (**a**), grade II (**b**), and grade III (**c**) tumors, SEER registry data, 1990–2003. The number in the upper right corner of each plot is a rescaling factor: incidence rate should be obtained by dividing the values at the plot by the rescaling factor

allowed for deeper analysis of the sources and mechanisms causing this heterogeneity. We used a biologically motivated model that was specified by cancer subtype (ductal and lobular), race, stage, grade, and hormone receptor status, thus extending existing breast carcinogenesis models [6, 15–19]. The results demonstrated that grade plays the primary role in breast cancer heterogeneity, with ER/PR status "strengthening" the grade effect.

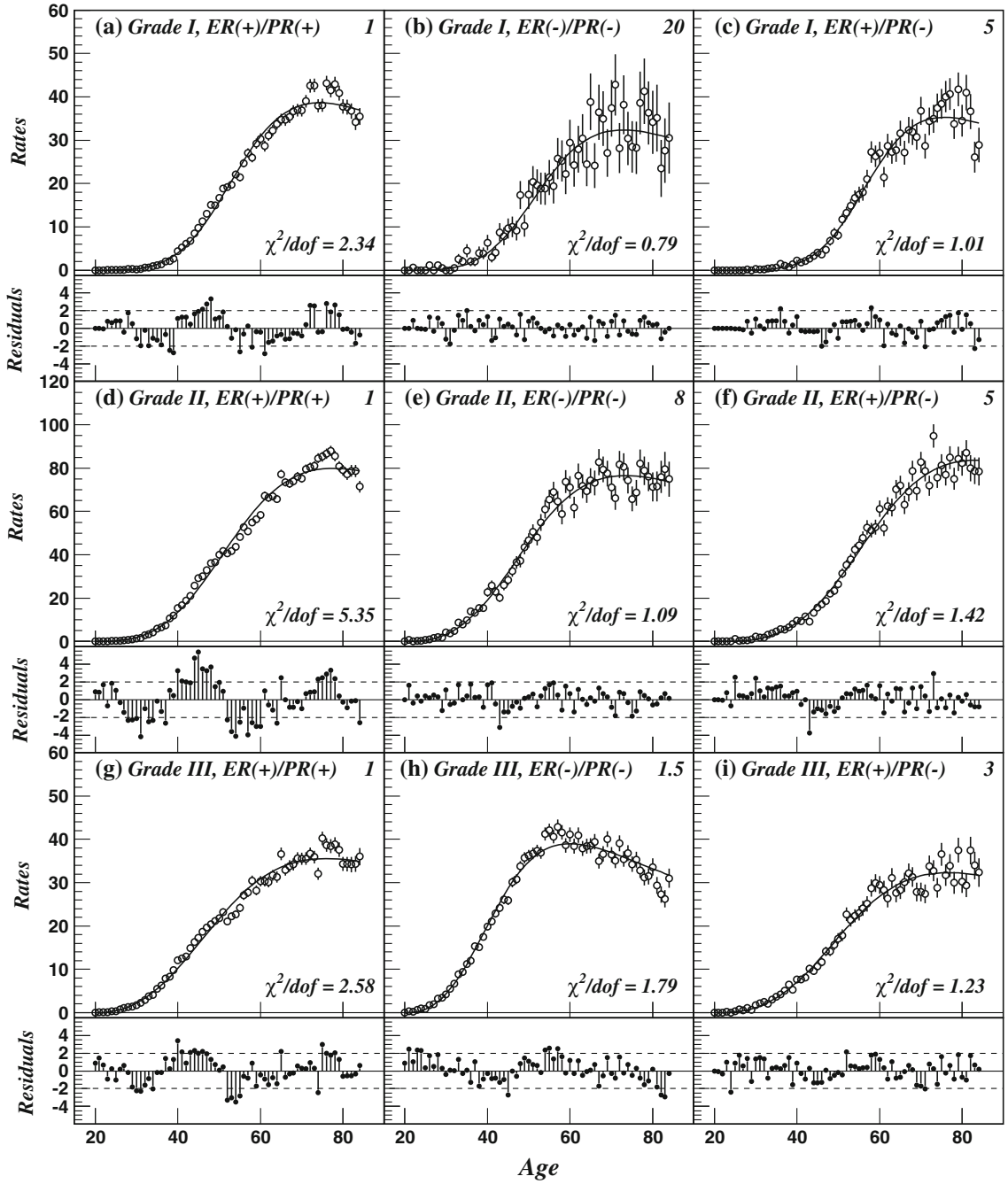

Fig. 5 ER/PR status-specific age patterns of breast ductal carcinoma incidence and model fit (i.e., residuals) for Caucasian females for grade I (*upper row*), grade II (*center row*), and grade III (*lower row*) tumors with ER(+)/PR(+) (*left column*), ER(−)/PR(−) (*center column*), and ER(+)/PR(−) (*right column*) status, SEER registry data, 1990–2003. The number in the upper right corner of each plot is a rescaling factor: incidence rate should be obtained by dividing the values at the plot by the rescaling factor

Using relatively short period of time in this study (from 1990 to 2003) allowed us to minimize the period and cohort effects. In our earlier SEER-based studies of breast carcinoma the *m*-stages parameter (which was the main focus of our model in this study) was stable in respect of time period effect (see Table 7.3 in [7]). Based on Anderson et al. [5], it could be suggested that for time period analyzed in this study the period-cohort effects on breast cancer bimodality and Clemmesen's hook are minor: the age-related effects on breast cancer incidence have been shown to be more important (about 10-fold greater) than calendar period or birth cohort effects [5].

A steady increase in lobular carcinoma incidence suggests the importance of studying its mechanism [20].

Table 3 Grade-specific characteristics of carcinogenesis for breast lobular and ductal carcinomas in Caucasian and African-American females (M ± SE), for ER(all)/PR(all) status

Grade	Ductal carcinoma				Lobular carcinoma			
	Caucasians		African-Americans		Caucasians		African-Americans	
	m	*lag*	*m*	*lag*	*m*	*lag*	*m*	*lag*
G1	4.76 ± 0.06	24.3 ± 1.5	4.47 ± 0.14[a]	24.9 ± 2.8	5.31 ± 0.12[b]	28.1 ± 2.4	5.94 ± 0.40[a,b]	35.3 ± 1.2[a,b]
G2	3.85 ± 0.04[c]	23.2 ± 0.9	3.52 ± 0.07[a,c]	19.0 ± 1.5[a,c]	4.64 ± 0.09[b,c]	23.9 ± 2.2	4.47 ± 0.216[b,c]	24.3 ± 3.5[b,c]
G3	3.28 ± 0.03[c]	19.2 ± 0.6[c]	3.26 ± 0.07[c]	14.9 ± 1.8[a,c]	3.90 ± 0.08[b,c]	22.2 ± 1.5[b,c]	4.02 ± 0.23[b,c]	28.3 ± 1.2[a,b,c]

[a] Difference between Caucasians and AAs for respective parameters

[b] Difference between ductal and lobular breast carcinomas for respective parameters

[c] Difference for parameter decrease from grade I to grade III (i.e., G1 > G2 > G3)

A shorter latency period of ductal compared to lobular carcinoma obtained from our analysis could be due (at least in part) to higher proliferation and apoptosis [21]. Our results are also in agreement with other molecular studies on cancer proliferation, biology, and genetics, which demonstrated that lobular tumors exhibited similarities to low-grade ductal tumors [22–24].

Several studies have revealed a striking similarity between in situ and invasive breast cancers, where all markers correlated with grade rather than with tumor invasiveness [22, 25–27], and there was a good concordance between grade in primary and metastatic tumors [28–30]. Recent molecular studies have demonstrated that qualitative and quantitative differences exist between ER(+) and ER(−) tumors and between low- and high-grade cancers, but not among different stages [22, 31–34]. After taking into account that grade III tumors in our study were more often diagnosed at advanced stages, and grade I tumors were more likely to be diagnosed at earlier stages, the differences obtained between the grades for a number of *m*-stages and *lag* could be even more pronounced. Further detailed analyses of the effects of stage at diagnosis on carcinogenesis parameters are required with the model generalization describing cancer cases registered at certain stages (as registered in the SEER registry), such as the "time-to-next-stage" model (an approach similar to that obtained from cancer survival modeling).

The majority of grading systems (e.g., the Scarff-Bloom-Richardson method) combine histological assessment of nuclear pleomorphism, mitotic activity, and tubule formation [35]. However, our results suggest that tumor grade can differ not only by histological characteristics, but also by developmental mechanisms. The results showing fewer *m*-stages and shorter latency periods associated with poorly differentiated tumors obtained from our analysis were concordant with molecular studies of proliferation/apoptosis-related markers showing that mitotic/apoptotic activity was higher in poorly differentiated tumors [34]. More recent molecular studies have suggested that breast cancer grade may be associated with distinct molecular subtypes with unique origins and pathogenesis, thus requiring distinct molecular targets for treatment [36]. In our study, grade II tumors demonstrated the highest heterogeneity, and they had carcinogenesis characteristics intermediate between grades I and III. This is in agreement with other studies that showing that grade I and III breast tumors have distinct gene expression profiles, whereas the gene expression profiles of grade II tumors were intermediate or a heterogeneous mixture of grades I and III profiles, with two components of different molecular and tumorigenic characteristics [36–38].

The results we obtained were in agreement with other studies that reported an association between grade I and ER-positive tumor characteristics and between grade III and ER-negative tumor characteristics [39, 40]. The ER/PR status showed less contribution to the bimodality phenomenon than grade. A recent breast cancer gene expression study has demonstrated that grade is more strongly associated with breast cancer clinical outcome than ER status [38]. In our study, ER(+)/PR(−) tumors were the most homogeneous for all grades. These tumors have recently been described to be growth factor-dependent, constituting a unique subgroup of ER+ patients [41]. Our results also demonstrated that factors beyond the scope of this study likely contributed to the remaining heterogeneity of ER(+)/PR(+) tumors and less to ER(−)/PR(−) grade III tumors. The importance of these factors have recently been shown in several molecular studies where gene expression patterns have been used to classify the tumors into clinically relevant subgroups (such as by expression of luminal epithelial specific genes) associated with different clinical prognoses and age at disease onset [42, 43]. Further analysis is required to determine whether luminal A and B subtypes of ER(+) tumors, and HER-2(+), basal-like, and normal breast-like subtypes of ER(−) tumors could explain this remaining heterogeneity [44, 45] as well as how

variations in the population of breast cancer patients by other genetic and epigenetic characteristics contribute to the remaining heterogeneity.

Study limitations

As the estimates of carcinogenesis characteristics were obtained within our model which assumed that carcinogenesis is a multistage process, model dependence is one of the limitations of our approach. This limitation is common in carcinogenesis modeling approaches. However, we reviewed and tested a spectrum of different carcinogenesis models based on different assumptions, and we concluded that the model we chose describes age patterns of incidence rates most adequately. Underlying cancer heterogeneity could not be currently captured by specific groupings, and the inclusion of further tumor classifications, such as molecular pathway-specific analysis, could help further decrease tumor heterogeneity. Other assumptions include the stage at diagnosis, which was not explicitly incorporated into our model; however, we took other assumptions into account when analyzing their influence on breast carcinoma bimodality and their possible effects on grade-specific characteristics of carcinogenesis. Future studies involving modeling of stage-to-stage progression and the transition time between stages could be applied to data that are stratified by stages.

Conclusions

Among the studied factors, grade plays the primary role in the heterogeneity of breast cancer age patterns, and hormone receptor status has an important biological impact due to its association with the "remaining" heterogeneity. In this study, the approach of biologically motivated mathematical modeling was applied to the specific analysis of breast carcinogenesis with various factors taken into account. The grade-specific differences in carcinogenesis permitted speculation regarding tumor grades being associated with certain carcinogenesis characteristics, such as the number of *m*-stages and latency. A common biological mechanism may be unique to grade-specific breast carcinomas. The different mechanisms could determine the different molecular pathways associated with each cancer subtype and subsequent progression and outcome. For example, breast carcinomas in young females are characterized by more aggressive phenotypes; however, the biological underpinnings driving this phenomenon are largely unknown. Recently, analyses of the patterns of gene expression have provided a genetic explanation explaining the highly aggressive phenotype of grade III breast tumors in young females [46]. Further analyses of breast cancer heterogeneity in female populations could refine our understanding of breast carcinogenesis and have important implications for the stratification, rather than unification, of breast cancer prevention strategies, risk assessment, and treatment. The developed model could be used for estimation of ages at highest risk of potential carcinogens exposure (e.g., obesity, aryl hydrocarbon or xenoestrogens exposure, ionizing radiation, alcohol) for grade-ER/PR-specific breast cancers of "early" and "later" onsets to make prevention strategies focused on tumor-initiating events. Incorporation of genetic and epigenetic characteristics of patients into the model will allow developing of more individualized approach to cancer prevention and screening. By incorporating the molecular data, the model could be further developed to understand the normal and pathological breast cells growth [45, 47, 48], as well as for studying the effects of therapy targeting cancer-initiating progenitor cells.

Acknowledgments The authors thank Fred and Alice Stanback for supporting this study with a philanthropic donation to the Duke Comprehensive Cancer Center.

References

1. Lilienfeld AM, Johnson EA (1955) The age distribution in female breast and genital cancers. Cancer 8:875–882
2. Althuis MD, Fergenbaum JH, Garcia-Closas M et al (2004) Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. Cancer Epidemiol Biomarkers Prev 13:1558–1568
3. Anderson WF, Matsuno R (2006) Breast cancer heterogeneity: a mixture of at least two main types? JNCI 98:948–951
4. Armitage P, Doll R (1957) A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. Br J Cancer 11:161–169
5. Anderson WF, Reiner AS, Matsuno RK, Pfeiffer RM (2007) Shifting breast cancer trends in the United States. J Clin Oncol 25:3923–3929
6. Manton KG, Stallard E (1980) A two-disease model of female breast cancer: mortality in 1969 among white females in the United States. J NCI 64:9–16
7. Manton KG, Akushevich I, Kravchenko J (2009). Cancer mortality and morbidity patterns in the U.S. population: an interdisciplinary approach. Springer, 455 pp
8. Clemmesen J (1948) Carcinoma of the breast. Br J Radiol 21: 583–590
9. Anderson E (1950) Possible relationship between menopause and age at onset of breast cancer. Cancer 3:410–411
10. De Waard F (1979) Premenopausal and postmenopausal breast cancer: one disease or two? J Natl Cancer Inst 63:549–552
11. Anderson WF, Chatterjee N, Ershler WB (2002) Estrogen receptor breast cancer phenotypes in the Surveillance, Epidemiology, and End Results database. Breast Cancer Res Treat 76:27–36
12. Anderson WF, Pfeiffer RM, Dores GM (2006) Comparison of age frequency distribution patterns for different histopathologic types of breast carcinoma. Cancer Epidemiol Biomarkers Prev 15: 1899–1905

13. Anderson WF, Jatoi I, Devesa SS (2005) 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 90:127–137
14. Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. Nat Med 10:789–799
15. Pike MC, Krailo MD, Henderson BE et al (1983) "Hormonal" risk factors, "breast tissue age" and the age-incidence of breast cancer. Nature 303:767–770
16. Colditz GA, Baer HJ, Tamimi RM (2006) Breast cancer. In: Schottenfeld D, Fraumeni JF Jr (eds) Cancer epidemiology and prevention, 3rd edn. Oxford University Press, New York
17. Nunney L (1999) Lineage selection and the evolution of multistage carcinogenesis. Proc Biol Sci 266:493–498
18. Parmigiani G, Berry D, Aquilar O (1998) Determining carrier probabilities for breast cancer susceptibility genes BRCA1 and BRCA2. Am J Hum Genet 62:145–158
19. Colditz GA, Rosner BA (2000) Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. Am J Epidemiol 152:950–964
20. Li CI, Anderson BO, Daling JR et al (2003) Trends in incidence rates of invasive lobular and ductal breast carcinoma. J Am Med Assoc 289:1421–1424
21. Kruger S, Fahrenkrog T, Muller H (1999) Proliferative and apoptotic activity in lobular breast carcinoma. Int J Mol Med 4:171–174
22. Lacroix M, Toillon R-A, Leclercq G (2004) Stable "portrait" of breast tumors during progression: data from biology, pathology and genetics. Endocr Relat Cancer 11:497–522
23. Ferlicot S, Vincent-Salomon A, Medioni J (2004) Wide metastatic spreading in infiltrating lobular carcinoma of the breast. Eur J Cancer 40:336–341
24. Zhao H, Langerod A, Ji Y (2004) Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. Mol Biol Cell 15:2523–2536
25. Warnberg F, Nordgren H, Bergkvist L (2001) Tumor markers in breast carcinoma correlate with grade rather than with invasiveness. Br J Cancer 85:869–874
26. Iglehart JD, Kerns BJ, Huper G (1995) Maintenance of DNA content and erbB-2 alterations in intraductal and invasive phases of mammary cancer. Breast Cancer Res Treat 34:253–263
27. Mommers EC, Leonhart AM, Falix F (2001) Similarity in expression of cell cycle proteins between in situ and invasive ductal breast lesions of same differentiation grade. J Pathol 194: 327–333
28. Millis RR, Barnes DM, Lampejo OT (1998) Tumor grade does not change between primary and recurrent mammary carcinoma. Eur J Cancer 34:548–553
29. Bijker N, Peterse JL, Duchateau L (2001) Histological type and marker expression of the primary tumor compared with its local recurrence after breast-conserving therapy for ductal carcinoma in situ. Br J Cancer 84:539–544
30. Cserni G (2002) Tumor histological grade may progress between primary and recurrent invasive mammary carcinoma. J Clin Pathol 55:293–297
31. Visscher D, Jimenez RE, Grayson M III (2000) Histopathologic analysis of chromosome aneuploidy in ductal carcinoma in situ. Hum Pathol 31:201–207
32. Boeker W, Buerger H, Schmitz K (2001) Ductal epithelial proliferations of the breast: a biological continuum? Comparative genomic hybridization and high-molecular-weight cytokeratin expression patterns. J Pathol 195:415–421
33. Richard F, Pacyna-Gengelbach M, Schluns K (2000) Patterns of chromosomal imbalances in invasive breast cancer. Int J Cancer 89:305–310
34. Buerger H, Mommers EC, Littmann R (2000) Correlation of morphologic and cytogenetic parameters of genetic instability with chromosomal alterations in in situ carcinomas of the breast. Am J Clin Pathol 114:854–859
35. Elston CW, Ellis IO (2002) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histophatology 41:154–161
36. Roylance R, Gorman P, Harris W (1999) Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer. Cancer Res 59:1433–1436
37. Ivshina A, George J, Senko O (2006) Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res 66:10292–10301
38. Sotiriou C, Wirapati P, Loi S (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 98:262–272
39. Zafrani B, Aubriot MH, Mouret E (2000) High sensitivity and specificity of immunohistochemistry for the detection of hormone receptors in breast carcinoma: comparison with biochemical determination in a prospective study of 793 cases. Histopathology 37:536–545
40. Ringberg A, Anagnostaki L, Anderson H (2001) Cell biological factors in ductal carcinoma in situ (DCIS) of the breast-relationship to ipsilateral local recurrence and histopathological characteristics. Eur J Cancer 37:1514–1522
41. Finn RS, Dering J, Ginther C (2006) ER+ PR− breast cancer defines a unique subtype of breast cancer that is driven by growth factor signaling and may be more likely to respond to EGFR targeted therapies. J Clin Oncol 24(18S):514
42. Sorlie T, Tibshirani R, Parker J (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 100:8418–8423
43. Sorlie T, Wang Y, Xiao C (2006) Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analysis across three different platforms. BMC Genomics 7:127
44. Nguyen PL, Taghian AG, Katz MS (2008) Breast cancer subtype approximated by estrogen receptor, progesterone receptor, and HER-2 is associated with local and distant recurrence after breast-conserving therapy. J Clin Oncol 26:2373–2378
45. Perou CM, Sorlie T, Eisen MB et al (2000) Molecular portraits of human breast tumours. Nature 406:747–752
46. Anders CK, Hsu DS, Broadwater G (2008) Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. J Clin Oncol 26:3324–3330
47. Boecker W, Buerger H (2003) 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 36(1 Suppl):73–84
48. Jones C, Nonni AV, Fulford L et al (2001) CGH analysis of ductal carcinoma of the breast with basaloid/myoepithelial cell differentiation. Br J Cancer 85:422–427