Improved classification of breast cancer by analysis of genetic alterations and gene expression profiling

Horlings, H.M.

Publication date
2011

6

Refinement of breast cancer classification by molecular characterization of histological special types

6

Refinement of breast cancer classification by molecular characterization of histological special types†

B Weigelt,[1*‡§] HM Horlings,[1§] B Kreike,[1] MM Hayes,[2] M Hauptmann,[3] LFA Wessels,[3] D de Jong,[4] MJ Van de Vijver,[4,5] LJ Van't Veer[1,4*] and JL Peterse[4||]

[1]Division of Experimental Therapy, The Netherlands Cancer Institute, Amsterdam, The Netherlands
[2]Department of Pathology, British Columbia Cancer Agency and Department of Pathology & University of British Columbia, Vancouver, Canada
[3]Division of Molecular Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
[4]Division of Pathology, The Netherlands Cancer Institute, The Netherlands
[5]Department of Pathology, Academic Medical Center, Amsterdam, The Netherlands

†*This article is dedicated to the memory of Dr Hans Peterse.*

**Correspondence to:*
B Weigelt, Ernest Orlando Lawrence Berkeley National Laboratory, Life Sciences Division, 1 Cyclotron Road, MS-977-225A, Berkeley, CA 94720, USA.
E-mail: bweigelt@lbl.gov

LJ Van't Veer, The Netherlands Cancer Institute, Department of Pathology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.
E-mail: l.vt.veer@nki.nl

‡*Current address: Life Sciences Division, Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, California, USA.*

§*These authors contributed equally to this work.*

||*Deceased.*

Conflicts of interest: LJ Van't Veer is an employee of, and holds shares in, Agendia. BV.

Received: 20 April 2008
Revised: 1 July 2008
Accepted: 2 July 2008

Abstract

Most invasive breast cancers are classified as invasive ductal carcinoma not otherwise specified (IDC NOS), whereas about 25% are defined as histological 'special types'. These special-type breast cancers are categorized into at least 17 discrete pathological entities; however, whether these also constitute discrete molecular entities remains to be determined. Current therapy decision-making is increasingly governed by the molecular classification of breast cancer (luminal, basal-like, HER2+). The molecular classification is derived from mainly IDC NOS and it is unknown whether this classification applies to all histological subtypes. We aimed to refine the breast cancer classification systems by analysing a series of 11 histological special types [invasive lobular carcinoma (ILC), tubular, mucinous A, mucinous B, neuroendocrine, apocrine, IDC with osteoclastic giant cells, micropapillary, adenoid cystic, metaplastic, and medullary carcinoma] using immunohistochemistry and genome-wide gene expression profiling. Hierarchical clustering analysis confirmed that some histological special types constitute discrete entities, such as micropapillary carcinoma, but also revealed that others, including tubular and lobular carcinoma, are very similar at the transcriptome level. When classified by expression profiling, IDC NOS and ILC contain all molecular breast cancer types (ie luminal, basal-like, HER2+), whereas histological special-type cancers, apart from apocrine carcinoma, are homogeneous and only belong to one molecular subtype. Our analysis also revealed that some special types associated with a good prognosis, such as medullary and adenoid cystic carcinomas, display a poor prognosis basal-like transcriptome, providing strong circumstantial evidence that basal-like cancers constitute a heterogeneous group. Taken together, our results imply that the correct classification of breast cancers of special histological type will allow a more accurate prognostication of breast cancer patients and facilitate the identification of optimal therapeutic strategies.

Keywords: breast cancer; expression profiling; histological classification; molecular subtypes

Introduction

Invasive breast cancers are a heterogeneous group of tumours that show a wide variation with regard to their clinical presentation, behaviour, and morphological spectrum. At least 18 different histological breast cancer types (ie pathological entities) are described by the World Health Organization (WHO) [1]. Invasive ductal carcinoma not otherwise specified (IDC NOS) accounts for the large majority of breast cancers (ie 50–80%). IDC NOS is a diagnosis by default, being defined by the WHO as a tumour that fails to exhibit sufficient morphological characteristics to be classified into one of the histological special types [1]. Approximately 25% of invasive breast cancers are recognized as 'special types', and characterized by distinctive growth patterns and cytological features [1–3] (Table 1 and Figure 1). However, carcinomas of special type are often not recognized as such at pathological examination and are lumped together with IDC NOS.

Table 1. Frequency and outcome of histological types of invasive breast carcinoma [1,2]

Histopathological type of invasive breast carcinoma	Frequency	10-year overall survival rate
Invasive ductal carcinoma not otherwise specified (IDC NOS)	50–80%	35–50%
Invasive lobular carcinoma (ILC)	5–15%	35–50%
Adenoid cystic carcinoma	0.1%	90–100%
Apocrine carcinoma	0.3–4%	Like IDC NOS
IDC with osteoclastic giant cells	Unknown	Like IDC NOS
Medullary carcinoma	1–7%	50–90%
Metaplastic carcinoma	<5%	Unknown
Micropapillary carcinoma	<3%	Unknown
Mucinous carcinoma	<5%	80–100%

Recently, gene expression profiling studies established a widely applied molecular classification of breast cancers distinguishing three major subtypes, luminal, basal-like, and HER2+ breast cancers, which are characterized by distinct transcriptomic features and, most importantly, patient outcomes [4,5]. This molecular subtyping, however, has been developed based on the gene expression profiles of largely IDC NOS and a few ILCs only [4]. It is unknown whether the molecular classification system also applies to the other histological special types. Likewise, it is unknown whether prognostic gene sets, including the 70-gene prognosis profile [6,7] and 21-gene recurrence score [8], have similar prognostic power in the special types of breast cancer.

Here we describe a comprehensive characterization of a series of 11 different histological special-type breast carcinomas by immunohistochemistry and gene expression profiling in an attempt to refine breast cancer classification and improve patient stratification.

Materials and methods

Selection of tumours

Specimens ($n = 113$) of 11 histological pure variants of invasive breast cancer were selected from the frozen tissue bank of The Netherlands Cancer Institute/Antoni van Leeuwenhoek hospital (NKI/AVL). Before and after cutting tissue sections for RNA isolation, a representative section was stained with haematoxylin and eosin and semi-quantitatively assessed for the percentage of tumour areas over the total sample area by two of the authors (BW and JLP). Only samples containing ≥50% tumour cells [median 80% (range 60–95%)] were selected for downstream analysis (for detailed information on tumour cell content of samples see Supporting information, Supplementary Table 1). Tumours were classified based on the WHO criteria as ILC ($n = 22$; $n = 18$ classic, $n = 4$ pleomorphic, $n = 0$ tubulo-lobular), tubular ($n = 9$), mucinous ($n = 19$), neuroendocrine ($n = 10$), apocrine ($n = 6$), IDC with osteoclastic giant cells ($n = 5$), micropapillary ($n = 8$), adenoid cystic ($n = 4$), metaplastic ($n = 20$), and typical medullary carcinoma ($n = 10$) [1]. Mucinous tumours were subdivided into hypocellular mucinous (mucinous A) ($n = 10$) and cellular mucinous (mucinous B) ($n = 9$) based on the criteria of Capella *et al* [9]. The selection was carried out by independent review of the tumour sections by three pathologists (MMH, MvdV, and JLP) and only cases that fulfilled the diagnostic criteria for pure special types according to all observers were included. In addition, 45 IDCs NOS [1] composed of more than 85% of areas morphologically only classifiable as ductal NOS patterns and containing ≥50% tumour areas [median 70% (range 50–90%)] were selected (clinicopathological and immunohistochemical characteristics are summarized in the Supporting information, Supplementary Table 2). This study was approved by the Medical Ethical Committee of the NKI/AVL.

Tissue microarrays and immunohistochemistry

A tissue microarray of 112 of the 113 breast carcinomas (the paraffin block of one neuroendocrine tumour was unavailable) was constructed using a manual tissue arrayer (Beecher Instruments, Silver Spring, MD, USA) as previously described [10]. 600 μm tissue cores were taken from each paraffin-embedded tumour donor block and arrayed in triplicate into a new recipient paraffin block.

Serial sections of 3 μm were cut from the tissue microarray blocks, deparaffinized in xylene, and hydrated in a graded series of alcohol. Detailed information on the antibodies, staining, and scoring methods is available in the supporting information, Supplementary Table 3. When the staining score differed among the three cores analysed, the highest score was recorded. In the very few cases where the staining result could not be evaluated on the TMA, staining was repeated on whole paraffin sections.

Statistical analysis of immunohistochemistry

We compared the distribution of immunohistochemical markers across the histological special types using the Kruskal–Wallis test for singly ordered $R \times C$ contingency tables, where $R = 11$ histological subtypes and

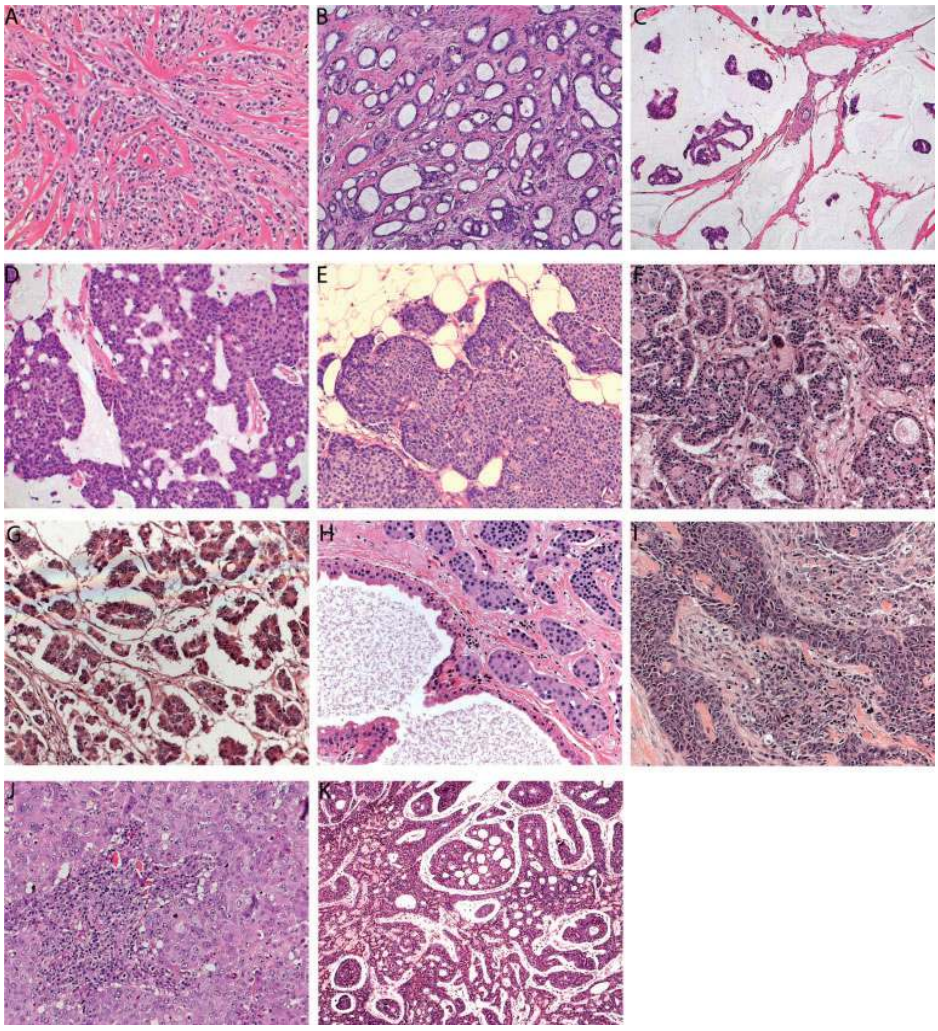

Figure 1. Histology of invasive breast carcinomas. Representative micrographs of special type breast cancers: (A) invasive lobular carcinoma, (B) tubular, (C) mucinous A, (D) mucinous B, (E) neuroendocrine, (F) IDC with osteoclastic giant cells, (G) micropapillary, (H) apocrine, (I) metaplastic, (J) medullary, and (K) adenoid cystic carcinoma

C represents up to four ordered categories of staining intensity [11]. Because of the large sparse tables, we used 100 000 Monte Carlo samples to approximate exact *p* values.

RNA isolation and microarray expression profiling

Detailed protocols for RNA isolation, amplification, labelling, and hybridization can be found at http://microarrays.nki.nl/download/protocols.html.
RNA quality was assessed by measurement of the OD 260/280 ratio using the NanoDrop 1000 (Fisher Scientific, Pittsburgh, USA) and only samples with a ratio $\geq$1.95 were included. RNA integrity was assessed by gel electrophoresis. Samples were co-hybridized with a standard reference of pooled and amplified RNA from 100 breast tumours; each sample was hybridized using reverse colour labelling (ie 'dye swaps'). Oligo microarrays with a complexity of 34 580 probes representing 24 650 genes were prepared at the Central Microarray Facility (CMF) of the NKI/AVL (http://microarrays.nki.nl). Fluorescent images of the microarrays were obtained using the Agilent DNA microarray scanner (Agilent Technologies, Palo Alto, USA). Fluorescent intensities were quantified using ImaGene 5 (Biodiscovery, Marina

Table 2. Results of the Kruskal–Wallis test for singly ordered contingency tables of 22 immunohistochemical markers and 11 histological breast cancer subtypes (112 tumours). Cells are colour-coded with respect to the corresponding mean ranks (shown in each cell). High values indicate, on average, a higher amount of staining. (Asymp = asymptotic)

	Histological Subtype											Kruskal-Wallis Test (a, b)			
Antibody	Apocrine	Mucinous A	Mucinous B	Neuroendocrine	IDC Osteo	Micropap	Adenoid	Medullary	Metaplastic	ILC	Tubular	Chi-Square	df	Asymp. p-value	exact p-value
ER	42	71	81	81	81	81	23	23	23	67	68	79.83	10	7.95E-10	<1e-5
E-cadherin	59	63	68	68	68	68	68	68	55	17	68	72.99	10	7.46E-10	<1e-5
CK 19	64	74	67	69	64	74	27	33	24	70	69	61.86	10	2.08E-09	<1e-5
CD117	47	47	47	47	47	47	103	81	72	47	47	59.29	10	5.41E-09	<1e-5
AR	87	65	81	67	47	68	28	28	28	68	67	54.08	10	4.73E-08	<1e-5
EMA	60	43	61	47	82	104	12	62	31	62	63	53.58	10	5.84E-08	<1e-5
CK 8/18	66	60	66	66	66	66	38	33	35	66	66	47.94	10	6.36E-07	<1e-5
PR	34	72	62	84	96	64	34	34	34	63	63	47.71	10	7.00E-07	<1e-5
Vimentin	45	55	45	56	45	45	85	74	78	45	45	45.01	10	2.17E-06	<1e-5
S100	36	54	36	36	36	36	70	88	74	62	51	44.07	10	3.19E-06	<1e-5
Synaptophysin	51	62	88	69	51	51	51	51	51	53	51	43.09	10	4.78E-06	4.00E-05
GCDFP-15	68	34	66	53	34	40	34	34	34	39	52	36.03	10	8.33E-05	1.83E-03
CK 14	57	53	48	48	48	48	76	64	76	50	48	31.74	10	0.0004431	0.00057
CK 5/6	52	48	43	43	54	64	85	82	65	53	43	29.96	10	0.0008682	0.0005
P63	51	56	51	51	51	51	81	62	68	51	51	25.47	10	0.004528	0.00655
Chromogranin	55	55	67	61	55	55	55	55	55	55	55	18.07	10	0.05386	0.0538
CEA	66	63	53	66	70	67	47	47	49	56	53	13.55	10	0.1945	0.181
CD56	47	52	53	59	47	47	47	63	66	57	59	10.85	10	0.3693	0.357
P53	70	45	40	53	71	68	58	47	56	65	53	10.42	10	0.4043	0.411
EGFR	64	55	55	55	55	55	55	60	60	55	55	9.32	10	0.5022	0.459
CD10	61	52	58	52	52	52	52	57	60	62	52	8.07	10	0.6219	0.619
HER2	62	58	59	53	53	59	53	53	58	58	53	4.68	10	0.9115	0.952

Kruskal-Wallis Test, a = Based on 100,000 Monte Carlo sampled; b = Grouping Variable = Histological subtypes. Numbers represent median ranks of Kruskal-Wallis Test
Abbreviations: Adenoid = Adenoid cystic carcinoma; ILC = Invasive lobular carcinoma; IDC Osteo = Invasive ductal carcinoma with osteoclastic giant cells; Micropap = Micropapillary

Ranks

Lowest — Highest

Del Rey, USA), normalized, and corrected for a variety of biases that affect the intensity measurements [12]. Weighted averages and confidence levels were computed according to the Rosetta error model [13]. Microarray data of the 113 special types are available at Array Express (http://www.ebi.ac.uk/arrayexpress/), experiment number E-NCMF-3.

Unsupervised hierarchical clustering

In order to remove those genes of the 34 580 probes on the array with low expression variation across tumours, we only retained genes that were significantly regulated ($p < 0.01$) in at least 14 of the 113 samples with missing data in three samples or less, resulting in a set of 8513 genes. The p value was derived based on the Rosetta error model [13]. We performed unsupervised hierarchical clustering on these 8513 genes using centred Pearson correlation as the similarity metric and complete linkage clustering. Cluster 3.0 software was used for clustering [14] and the results were visualized using Java Treeview (http://jtreeview.sourceforge.net/).

Molecular subtype, 70-gene prognosis profile, and 21-gene recurrence score classification

For molecular subtype classification, hierarchical clustering analysis of the updated 'Intrinsic/UNC' gene list comprising 1300 unique genes, of which 1098 were identified on our microarray platform, was employed [15]. The molecular subtypes of the samples were determined by the branch of the dendrogram that was associated with characteristic gene expression patterns. In addition, correlations to the class centroids were calculated using the 'Intrinsic/UNC' centroids comprising 306 unique genes [15], of which 293 could be identified.

For 70-gene prognosis profile classification, the correlation coefficient of the expression level of the 70 genes, of which 60 could be identified on our microarray platform, with an average good prognosis profile was calculated as reported previously [6,7,16]. To classify tumours according to the recurrence score predictor, microarray data for all 21 recurrence score genes were used and the normalization, recurrence score computation and assignment to low-, intermediate-, and high-risk categories, was performed as described previously [8,16]. Both tests are microarray readings of the gene sets of the two published diagnostic tests, with adapted calculations to derive the prognostic indices.

Ingenuity Pathway Analysis

The Ingenuity Pathway Analysis program (http://www.ingenuity.com) was used to analyse pathways and networks that were significantly regulated in the gene expression data of the different histological subtypes. Details of the significance, symbols, and annotations used by Ingenuity Pathway Analysis can be found in the supporting information.

Results

Immunohistochemical and gene expression analysis of histological special types of breast carcinoma

To explore whether the 11 histological subtypes selected for this study also constitute distinct entities at the molecular level, we analysed their protein expression pattern by immunohistochemical staining on tissue microarrays with a panel of 22 antibodies representing markers specific for cell type and

differentiation (see supporting information, Supplementary Table 3). We observed significant heterogeneity in the staining intensity for ER, E-cadherin, CK19, CD117, AR, EMA, CK8/18, PR, vimentin, S100, synaptophysin, GCDFP-15, CK14, and CK5/6 ($p < 0.05/22 = 0.0023$ including Bonferroni adjustment for 22 tests performed), but not for p63, chromogranin, CEA, CD56, p53, EGFR, CD10, and HER2 (Table 2). The ER varied mostly across the 11 special-type classes studied and discriminated the ER-positive from the ER-negative subtypes (adenoid cystic, medullary, and metaplastic, $p < 0.00001$). Furthermore, the histological special types could be distinguished in luminal keratin-positive (eg mucinous, ILC, tubular carcinoma; CK8/18 $p < 0.000001$) versus basal keratin-positive-derived subtypes (adenoid cystic, medullary, and metaplastic; CK14 $p < 0.00057$ and CK5/6 $p < 0.0005$, respectively) (Table 2).

Except for E-cadherin, which was significantly down-regulated in ILCs ($p < 0.00001$), the overall staining pattern of ILCs showed great similarities to those of tubular carcinomas (Table 2). As expected, all eight micropapillary carcinomas studied showed 'inside-out' staining for the epithelial membrane antigen (EMA) [17] ($p < 0.00001$) (Supporting information, Supplementary Table 4). In addition, the micropapillary tumours were characterized by decreased expression of S100. IDCs with osteoclastic giant cells shared some characteristics with micropapillary carcinomas, including increased CEA and p53, decreased S100, and 'inside-out' EMA staining (Table 2 and Supporting information, Supplementary Table 4).

The neuroendocrine, mucinous A, and mucinous B tumours stained positive for the endocrine markers synaptophysin and chromogranin [2] (Table 2). Also, the adenoid cystic, medullary, and metaplastic carcinomas showed a similar overall immunohistochemical staining pattern, which was characterized by low levels of CK19, AR, CK8/18, and PR expression, and elevated levels of CD117, vimentin, S100, CK14, and CK5/6 expression, compared with the other subtypes (Table 2).

In summary, immunohistochemical staining revealed that a number of histological special types have similar protein expression patterns (eg ILC and tubular; mucinous and neuroendocrine; adenoid cystic, medullary and metaplastic carcinoma) which may suggest a common aetiological background and/or the involvement of common genetic/epigenetic pathways during tumourigenesis.

In addition, we performed gene expression profiling for the 113 breast carcinomas. Unsupervised hierarchical cluster analysis using 8518 significantly regulated genes divided the tumours into two groups based on their ER expression [6,18] (Figure 2). Within the ER-negative group, apocrine tumours and pleomorphic ILCs, which also exhibited apocrine differentiation, formed a separate cluster. The adenoid cystic carcinomas clustered in one branch within the metaplastic and medullary carcinomas, all of which showed relatively similar gene expression patterns, paralleling the immunohistochemistry results (Table 2).

Within the ER-positive tumours, seven of the eight micropapillary carcinomas clustered together in one distinct branch (Figure 2). Mucinous B tumours clustered together with neuroendocrine and mucinous A tumours, supporting the results of the immunohistochemical analysis. Of note, a number of mucinous A cancers formed a separate cluster, which was characterized by increased expression of proliferation and cell cycle genes compared with the other mucinous A tumours (data not shown). IDCs with osteoclastic giant cells were most similar in gene expression to mucinous A and micropapillary tumours. Tubular carcinomas, however, showed remarkable similarities at the transcriptome level to and intermingled with ILCs (Figure 2). Collectively, hierarchical clustering analysis confirmed the identity of special types, such as micropapillary carcinoma. The similarities seen between tubular and lobular, mucinous and neuroendocrine, and medullary, metaplastic, and adenoid cystic carcinoma on the protein level were further corroborated and expanded by gene expression profiling.

Identification of molecular subtypes in special-type breast cancers

To test whether the molecular subtypes described for IDC NOS and ILC also exist in the special-type breast cancers, clustering analysis was performed on 45 IDCs NOS and the 113 special-type cancers. Hierarchical clustering using the 'Intrinsic/UNC' gene set subdivided IDCs NOS and the special types into luminal, basal-like, and HER2+ tumours (Figure 3) [4,15]. In addition, a recently described 'molecular apocrine' group of breast cancers could be identified [19], which included androgen receptor (AR)-positive and ER-negative apocrine and pleomorphic ILCs. Remarkably, the IDCs NOS and ILCs consist of different molecular subtypes, whereas the histological special types, with the exception of apocrine carcinomas, are very homogeneous and each belongs to only one molecular subtype (Figure 3 and Supporting information, Supplementary Table 5). Of note, the medullary and adenoid cystic carcinomas, which are known to be associated with a favourable outcome (Table 1), cluster as poor prognosis basal-like tumours based on their intrinsic gene expression profiles.

Similar results were obtained using the 'Intrinsic/UNC' class centroids for molecular subtype assignment (data not shown) [15]. As no centroids are available for the molecular apocrine subtype, the by clustering 'molecular apocrine' pleomorphic ILCs and apocrine carcinomas were classified based on the correlation coefficient to either the luminal or the basal-like subtype. One apocrine tumour did not show a sufficient correlation with any molecular subtype. In addition, four ILCs and three tubular carcinomas switched from

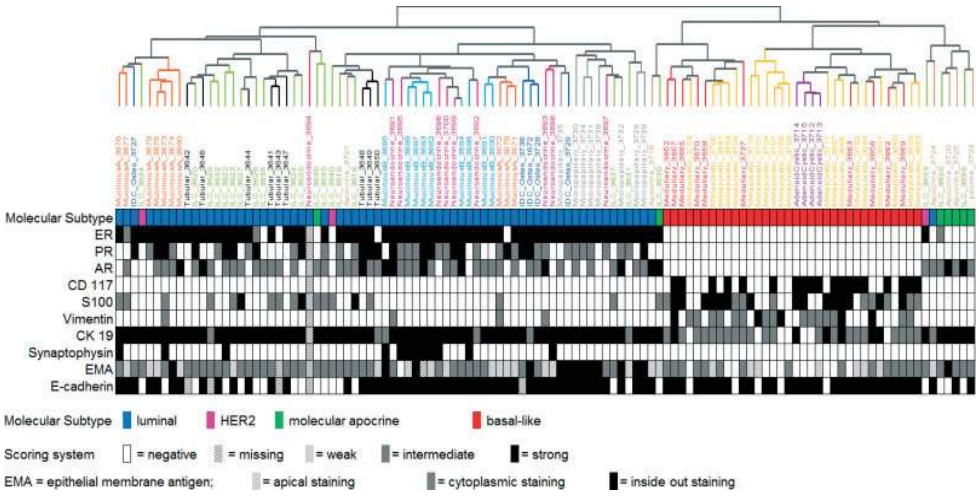

Figure 2. Unsupervised hierarchical clustering of histological special types. Hierarchical clustering of 113 breast carcinomas of 11 histological types measured over 8518 genes whose expression varied most across samples. Immunohistochemical staining results of selected markers are included

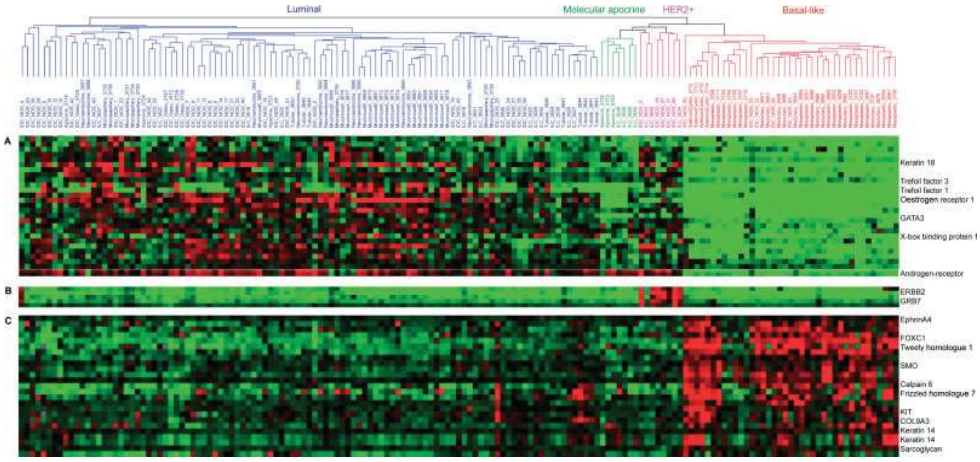

Figure 3. Molecular subtype identification. Hierarchical clustering of IDC NOS and 11 breast cancer special types using the 'Intrinsic/UNC' gene set [15]. (A) Luminal/ER-positive, molecular apocrine AR-positive gene cluster. (B) HER2 and GRB7-containing expression cluster. (C) Basal-like cluster

luminal type by clustering to the normal breast-like subtype, which could not be identified by clustering, as did two basal-like adenoid cystic carcinomas. The gene expression patterns of these ILCs and tubular carcinomas had very high correlation coefficients to the luminal centroid, and the two adenoid cystic carcinomas to the basal-like centroid, but the correlation coefficient to the normal breast-like centroids was in all cases slightly higher (data not shown). The basal-like nature of adenoid cystic carcinoma, however, is supported by CK5/6 and CD117 expression and lack of ER, PR, and HER2 expression [20] (Supporting information, Supplementary Table 4). In addition, the special-type breast carcinomas have been classified according to the 70-gene prognosis profile [6,7] and the 21-gene recurrence score [8] by microarray-derived readings of the gene sets of the two diagnostic tests [16] (Supporting information, Supplementary Table 6).

Pathway analysis

Ingenuity Pathway Analysis was applied to identify specific regulatory networks of genes operating in

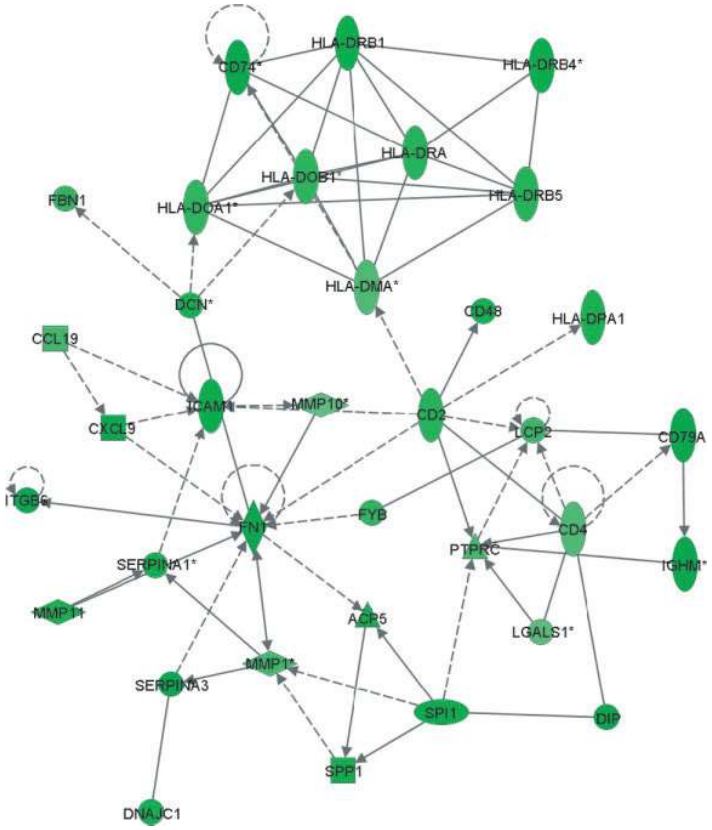

Figure 4. Top-scoring network identified by Ingenuity Pathway Analysis in adenoid cystic carcinomas. Network of genes associated with migration, proliferation, and immune response (score 54). The intensity of the node colour indicates the degree of up-regulation (red) or down-regulation (green)

the histological subtypes of breast cancer. In mucinous B carcinomas, which have a favourable outcome [1,2], one network involving migration, invasion, and proliferation genes was significantly down-regulated (score 63) (Supporting information, Supplementary Figure 1). Also, in the molecularly similar neuroendocrine carcinomas, one major down-regulated network of genes involved in migration, invasion, and proliferation was identified (score 67) (Supporting information, Supplementary Figure 2).

For adenoid cystic carcinomas, a tumour type associated with an excellent prognosis [1,21], Ingenuity Pathway Analysis determined two major networks containing genes associated with migration, proliferation, and immune response (score 54), which were down-regulated (Figure 4 and Supporting information, Supplementary Figure 3). Remarkably, almost the entire antigen presentation pathway is down-regulated in this tumour type (Supporting information, Supplementary Figure 3).

Discussion

The correct classification of the histological special types of breast cancer is not just an academic exercise, as it has both prognostic and predictive implications. For instance, patients with pure tubular or adenoid cystic carcinomas have overall survival rates similar to those of the general population, and ILCs have been shown to have a distinct metastatic pattern and poor response to neo-adjuvant chemotherapy [1,3,21,22].

The current system of histological classification has been proven to be subjective and not to reflect accurately the biological complexity of breast cancers. With the exception of a few examples (eg loss of E-cadherin expression in lobular carcinomas), there is a paucity of molecular markers to resolve the histological classification of equivocal cases. Although transcriptome analyses of breast cancers using high-throughput methods have been performed, these have been largely restricted to IDCs NOS, a few ILCs, and metaplastic breast cancers [4,6,7,15,23].

We demonstrate not only that expression profiling confirmed that some special types of breast cancer are specific entities, but also that a number of histological subtypes do not constitute distinct entities. Several special types have been shown to be remarkably similar at the transcriptome level, whereas others have rather heterogeneous transcriptome profiles. It should be noted that in our analysis both neoplastic and stromal cells were included, given that both together form and characterize each breast cancer special type. As the special types are heterogeneous with regard to their stromal composition, the results of the hierarchical clustering may be based on the transcriptome of stroma and tumour cells, rather than solely on the characteristics of the cancer cells.

Here we demonstrate that pure micropapillary carcinomas have a characteristic immunoprofile and constitute a distinct group of ER-positive cancers in hierarchical clustering analysis (Table 2 and Figure 2). In addition, micropapillary tumours have recently been reported to have distinct molecular genetic profiles from IDCs NOS [24], confirming that micropapillary carcinomas of the breast constitute a specific pathological entity.

On the other hand, we provide strong circumstantial evidence to suggest the existence of two large subgroups of ER-positive special types of breast cancer: one characterized by neuroendocrine differentiation and the other composed of special types with indolent clinical behaviour. In this study, ER-positive neuroendocrine, mucinous A, and mucinous B tumours, tumours classified as distinct breast cancer special types based on the histological WHO criteria [1], pertain to a single molecular subgroup. These three subtypes stained positive for the neuroendocrine markers synaptophysin and chromogranin (Table 2 and Supporting information, Supplementary Table 4), showed high similarity in overall gene expression (Figure 2), and were of luminal molecular subtype (Figure 3). This is not surprising, given that these special types of breast cancer are reported to have a similar age distribution, occasionally show overlapping morphological features, and have similar clinical behaviour [1]. In addition, we identified gene networks of invasion and proliferation to be down-regulated in both mucinous B and neuroendocrine carcinomas (Supporting information, Supplementary Figures 1 and 2), which may explain the low incidence of metastasis in patients with mucinous carcinoma (Table 1) [1,2].

ER-positive tumours with an indolent clinical behaviour form a distinct group within the luminal subtype (Figure 2). Classic ILCs and tubular carcinomas show remarkably similar transcriptomic and immunohistochemical profiles. However, ILC can be differentiated from tubular carcinoma based on the expression levels of E-cadherin (Table 2 and Supporting information, Supplementary Table 4) [25,26]. Our findings provide molecular support for the hypothesis that classic ILCs and tubular carcinomas, both members of low-grade breast neoplasia, might originate from the same family of low-grade precursors [26]. Based on an *in silico* analysis of our microarray data, 38% of the classic ILCs and tubular carcinomas studied here have a low or intermediate risk 21-gene recurrence score and 69% a good 70-gene prognosis signature [6,8] (see the Materials and methods section and Supporting information, Supplementary Table 6).

The four pleomorphic ILCs clustered together with apocrine tumours in the hierarchical clustering (Figure 2). These pleomorphic ILCs, unlike classic ILC, were not classified as luminal but as either HER2+ or molecular apocrine subtypes (Figure 3). These findings provide molecular support for the definition of pleomorphic ILCs based on the presence of apocrine features in conjunction with nuclear pleomorphism, as initially proposed by Eusebi *et al* [27]. Although classic and pleomorphic ILCs may co-exist [27], have similar genetic aberrations [28,29], and the latter may progress from classic ILC [29,30], the significant differences in the molecular profiles of classic and pleomorphic ILCs, together with the reported aggressive clinical behaviour of pleomorphic ILC [27,31], suggest that pleomorphic ILC should merit a status distinct from classic ILC. Notably, *in silico* analysis employing microarray-derived readings of two prognostic gene sets indicates that the apocrine carcinomas and pleomorphic ILCs of molecular apocrine subtype may be associated with a poor outcome. In fact, all seven 'molecular apocrine' tumours have a high-risk recurrence score and six of seven a poor 70-gene prognosis signature [6–8] (see Supporting information, Supplementary Table 6).

The immunohistochemical staining patterns and gene expression profiles of the ER-negative adenoid cystic, medullary, and metaplastic carcinomas were highly similar. However, adenoid cystic carcinomas do not intermingle with medullary and metaplastic tumours in the hierarchical clustering, but form a separate group (Figures 2 and 3). The favourable prognosis of adenoid cystic carcinomas, despite the fact that they do not express ER and they harbour a poor signature, may be explained not only by their low histological grade, but also by the low expression of genes associated with immune response and inflammation (Figure 4 and Supporting information, Supplementary Figure 3). Chronic activation of various cell types of the immune system has been suggested to promote tumour development by releasing proteolytic enzymes and angiogenic factors [32].

The down-regulation of genes involved in cellular growth and proliferation (data not shown), an effective host immune response, enhanced tumour cell apoptosis, and elevated levels of metastasis-inhibiting and low levels of metastasis-promoting factors, as reported by others [33,34], may account for the good prognosis of medullary carcinomas.

Although apocrine carcinomas displayed high levels of AR and GCDFP-15 protein expression, our results demonstrate that despite the limited sample size

($n = 6$), these tumours are unlikely to constitute a distinct entity. Apocrine carcinomas were shown to have heterogeneous gene expression profiles and to pertain to multiple molecular subtypes (Figures 2 and 3). As breast carcinomas of any type and grade may display features of apocrine differentiation [1], our data suggest that it might be more clinically and biologically relevant to identify the group of 'molecular apocrine' tumours, which show not only features of apocrine differentiation at the histological level, but also increased androgen signalling [19]. In a way similar to the success of targeting AR signalling in hormone-dependent prostate cancers [35], drugs inhibiting AR activity may constitute a novel therapeutic strategy for the management of patients with 'molecular apocrine' breast cancers.

We studied the existence of molecular subtypes as identified in IDC NOS and ILC in the rare phenotypes of breast cancer. Special-type breast cancers subdivide into the different molecular subtypes and admix with the IDCs NOS and ILCs (Figure 3). However, all histological special types of breast cancer but apocrine carcinomas were shown to be less heterogeneous than IDC NOS and ILC and to belong almost exclusively to one intrinsic subtype. Analysis of the composition of each molecular subtype in terms of the distribution of breast cancer special types revealed that basal-like breast cancers, which are generally associated with a poor clinical outcome [5], constitute a heterogeneous group of tumours. Our findings provide molecular support for previous studies demonstrating that this subgroup encompasses tumours with variable histology, clinical features, and response to chemotherapy [36–40].

Apart from grade III IDC NOS, basal-like breast cancers were shown to encompass all metaplastic [41], and the good outcome medullary [34,42,43] and adenoid cystic carcinomas [21]. The high rate of concordance between the 'intrinsic gene list' molecular subtypes and other prognostic gene signatures for patients with breast cancer [16] suggest that basal-like medullary and adenoid cystic carcinomas should be classified as aggressive tumour types by those outcome predictors. In fact, *in silico* analysis [16] of the 70-gene prognosis profile [6] and 21-gene recurrence score [8] using our gene expression data revealed that these two special types of breast carcinoma should also be assigned to the poor outcome 70-gene poor prognosis profile and 21-gene high-risk recurrence score (Supporting information, Supplementary Table 6), despite their reported favourable prognosis. Our findings emphasize that it is critical to develop new approaches to identify subgroups of patients with basal-like breast cancer that have a good outcome or a high likelihood of response to chemotherapy. In addition, deeper insight into the molecular heterogeneity of basal-like cancers may also contribute to the identification of novel therapeutic targets for this molecular tumour type.

Owing to the rarity of some of the entities analysed here (eg adenoid cystic carcinoma, IDC NOS with osteoclastic giant cells), our results on some of the special types should be interpreted as hypothesis-generating. Notwithstanding the limitation in sample size due to the nature of our study, our data prompt a re-evaluation of the existing histological classification system of breast tumours and suggest that the panel of 11 breast cancer subtypes selected following WHO criteria might be reduced to a smaller set based on their molecular profiles. The analysis of additional breast cancer special-type samples will be required to validate our findings, to determine the biological and clinical relevance of the novel 'molecular entities' of special-type cancers described here, and to identify molecular markers for their detection. Furthermore, we have shown that the molecular classification system of breast cancer using the 'intrinsic' genes and most likely other prognostic gene sets as well may be improved by a thorough and systematic analysis of special types of breast cancer. Taken together, our results represent a step forward towards a taxonomy that not only best reflects the biology of breast cancers, but also paves the way for a refinement in the prognostication of breast cancer patients and the identification of novel tailored therapies.

Acknowledgements

This article is dedicated to the memory of Dr Hans Peterse — we thank him for his inspiration, his critical mind, his mentorship and friendship. We thank A Broeks and LM Braaf for providing IDC NOS samples; R de Groot for making the tissue microarrays; and J Aantjes, J Houtgraaf, and DM Majoor for immunohistochemistry (NKI/AVL). This work was supported by the Dutch Cancer Society (KWF 02-2575) and the Cancer Genomics Center (Netherlands Genomics Initiative).

Supporting information

Supporting information may be found in the online version of this article.

References

1. Tavassoli FA, Devilee P. *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Breast and Female Genital Organs*. IARC Press: Lyon, 2003.
2. Schnitt SJ, Guidi AJ. Pathology of invasive breast cancer. In *Diseases of the Breast* (2nd edn), Harris JR, Lippman ME, Morrow M, Osborne CK (eds). Lippincott Williams & Wilkins: Philadelphia, 2004; 541–584.
3. Page DL. Special types of invasive breast cancer, with clinical implications. *Am J Surg Pathol* 2003;**27:**832–835.
4. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, *et al*. Molecular portraits of human breast tumours. *Nature* 2000;**406:**747–752.
5. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, *et al*. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;**98:**10869–10874.

6. Van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, *et al*. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;**415:**530–536.
7. Van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, *et al*. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;**347:**1999–2009.
8. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, *et al*. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;**351:**2817–2826.
9. Capella C, Eusebi V, Mann B, Azzopardi JG. Endocrine differentiation in mucoid carcinoma of the breast. *Histopathology* 1980;**4:**613–630.
10. Kononen J, Bubendorf L, Kallioniemeni A, Bärlund M, Schraml P, Leighton S, *et al*. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature Med* 1998;**4:**844–847.
11. Siegel S, Castellan NJ Jr. In *Nonparametric Statistics for the Behavioral Sciences* (2nd edn). McGraw-Hill: London, 1988; 206–216.
12. Yang Y, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, *et al*. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 2002;**30:**e15.
13. Weng L, Dai H, Zhan Y, He Y, Stepaniants SB, Bassett DE. Rosetta error model for gene expression analysis. *Bioinformatics* 2006;**22:**1111–1121.
14. Eisen M, Spellman P, Brown P, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;**95:**14863–14868.
15. Hu Z, Fan C, Oh D, Marron JS, He X, Qaqish BF, *et al*. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006;**7:**96.
16. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, *et al*. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;**355:**560–569.
17. Cheuk W, Chan JK. Subcellular localization of immunohistochemical signals: knowledge of the ultrastructural or biologic features of the antigens helps predict the signal localization and proper interpretation of immunostains. *Int J Surg Pathol* 2004;**12:**185–206.
18. Gruvberger S, Ringner M, Chen Y, Panavally S, Saal LH, Borg A, *et al*. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* 2001;**61:**5979–5984.
19. Farmer P, Bonnefoi H, Becette V, Tubiana-Hulin M, Fumoleau P, Larsimont D, *et al*. Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* 2005;**24:**4660–4671.
20. Azoulay S, Lae M, Freneaux P, Merle S, Al Ghuzlan A, Chnecker C, *et al*. KIT is highly expressed in adenoid cystic carcinoma of the breast, a basal-like carcinoma associated with a favourable outcome. *Mod Pathol* 2005;**18:**1623–1631.
21. Arpino G, Clark GM, Mohsin S, Bardou VJ, Elledge RM. Adenoid cystic carcinoma of the breast: molecular markers, treatment, and clinical outcome. *Cancer* 2002;**94:**2119–2127.
22. Mathieu MC, Rouzier R, Llombart-Cussac A, Sideris L, Koscielny S, Travagli JP, *et al*. The poor responsiveness of infiltrating lobular breast carcinomas to neoadjuvant chemotherapy can be explained by their biological profile. *Eur J Cancer* 2004;**40:**342–351.
23. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, *et al*. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;**365:**671–679.
24. Marchio C, Iravani M, Natrajan R, Lambros M, Savage K, Tamber N, *et al*. Genomic and immunophenotypical characterization of pure micropapillary carcinomas of the breast. *J Pathol* 2008;**215:**398–410.
25. Esposito NN, Chivukula M, Dabbs DJ. The ductal phenotypic expression of the E-cadherin/catenin complex in tubulolobular carcinoma of the breast: an immunohistochemical and clinicopathologic study. *Mod Pathol* 2007;**20:**130–138.
26. Abdel-Fatah TM, Powe DG, Hodi Z, Lee AH, Reis-Filho JS, Ellis IO. High frequency of coexistence of columnar cell lesions, lobular neoplasia, and low grade ductal carcinoma *in situ* with invasive tubular carcinoma and invasive lobular carcinoma. *Am J Surg Pathol* 2007;**31:**417–426.
27. Eusebi V, Magalhaes F, Azzopardi JG. Pleomorphic lobular carcinoma of the breast: an aggressive tumor showing apocrine differentiation. *Hum Pathol* 1992;**23:**655–662.
28. Nishizaki T, Chew K, Chu L, Isola J, Kallioniemi A, Weidner N, *et al*. Genetic alterations in lobular breast cancer by comparative genomic hybridization. *Int J Cancer* 1997;**74:**513–517.
29. Simpson PT, Reis-Filho JS, Lambros MBK, Jones C, Steele D, Mackay A, *et al*. Molecular profiling pleomorphic lobular carcinomas of the breast: evidence for a common molecular genetic pathway with classic lobular carcinomas. *J Pathol* 2008;**215:**231–244.
30. Reis-Filho JS, Simpson PT, Jones C, Steele D, Mackay A, Iravani M, *et al*. Pleomorphic lobular carcinoma of the breast: role of comprehensive molecular pathology in characterization of an entity. *J Pathol* 2005;**207:**1–13.
31. Weidner N, Semple JP. Pleomorphic variant of lobular carcinoma of the breast. *Hum Pathol* 1992;**23:**1167–1171.
32. de Visser K, Eichten A, Coussens L. Paradoxical roles of the immune system during cancer development. *Nature Rev Cancer* 2006;**6:**24–37.
33. Yakirevich E, Izhak OB, Rennert G, Kovacs ZG, Resnick MB. Cytotoxic phenotype of tumor infiltrating lymphocytes in medullary carcinoma of the breast. *Mod Pathol* 1999;**12:**1050–1056.
34. Bertucci F, Finetti P, Cervera N, Charafe-Jauffret E, Mamessier E, Adelaide J, *et al*. Gene expression profiling shows medullary breast cancer is a subgroup of basal breast cancers. *Cancer Res* 2006;**66:**4636–4644.
35. Taplin ME. Drug insight: role of the androgen receptor in the development and progression of prostate cancer. *Nat Clin Pract Oncol* 2007;**4:**236–244.
36. Kreike B, van Kouwenhove M, Horlings H, Weigelt B, Bartelink H, van de Vijver MJ. Gene expression profiling and histopathological characterization of triple negative/basal-like breast carcinomas. *Breast Cancer Res* 2007;**9:**R65.
37. Fadare O, Tavassoli FA. The phenotypic spectrum of basal-like breast cancers: a critical appraisal. *Adv Anat Pathol* 2007;**14:**358–373.
38. Fulford LG, Reis-Filho JS, Ryder K, Jones C, Gillett CE, Hanby A, *et al*. Basal-like grade III invasive ductal carcinoma of the breast: patterns of metastasis and long-term survival. *Breast Cancer Res* 2007;**9:**R4.
39. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, *et al*. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res* 2007;**13:**2329–2334.
40. Liedtke C, Mazouni C, Hess KR, Andre F, Tordai A, Mejia JA, *et al*. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol* 2008;**26:**1275–1281.
41. Reis-Filho JS, Milanezi F, Steele D, Savage K, Simpson PT, Nesland JM, *et al*. Metaplastic breast carcinomas are basal-like tumours. *Histopathology* 2006;**49:**10–21.
42. Jacquemier J, Padovani L, Rabayrol L, Lakhani SR, Penault-Llorca F, Denoux Y, *et al*. Typical medullary breast carcinomas have a basal/myoepithelial phenotype. *J Path* 2005;**207:**260–268.
43. Vincent-Salomon A, Gruel N, Lucchesi C, MacGrogan G, Dendale R, Sigal-Zafrani B, *et al*. Identification of typical medullary breast carcinoma as a genomic sub-group of basal-like carcinomas, a heterogeneous new molecular entity. *Breast Cancer Res* 2007;**9:**R24.