



HAL
open science

Effects of infiltrating lymphocytes and estrogen receptor on gene expression and prognosis in breast cancer

Alberto Calabrò, Tim Beissbarth, Ruprecht Kuner, Michael Stojanov, Axel Benner, Martin Asslaber, Ferdinand Ploner, Kurt Zatloukal, Hellmut Samonigg, Annemarie Poustka, et al.

► **To cite this version:**

Alberto Calabrò, Tim Beissbarth, Ruprecht Kuner, Michael Stojanov, Axel Benner, et al.. Effects of infiltrating lymphocytes and estrogen receptor on gene expression and prognosis in breast cancer. *Breast Cancer Research and Treatment*, 2008, 116 (1), pp.69-77. <10.1007/s10549-008-0105-3>. <hal-00478245>

HAL Id: hal-00478245

<https://hal.science/hal-00478245v1>

Submitted on 30 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Effects of infiltrating lymphocytes and estrogen receptor on gene expression and prognosis in breast cancer

Alberto Calabrò · Tim Beissbarth · Ruprecht Kuner · Michael Stojanov · Axel Benner · Martin Asslaber · Ferdinand Ploner · Kurt Zatloukal · Hellmut Samonigg · Annemarie Poustka · Holger Sültmann

Received: 19 March 2008 / Accepted: 12 June 2008 / Published online: 1 July 2008
© Springer Science+Business Media, LLC. 2008

Abstract The involvement of the immune system for the course of breast cancer, as evidenced by varying degrees of lymphocyte infiltration (LI) into the tumor is still poorly understood. The aim of this study was to evaluate the prognostic value of LI in breast cancer samples using microarray-based screening for LI-associated genes. Starting from the observation that most published ER gene signatures are heavily influenced by the LI effect, we developed and applied a novel approach to dissect molecular signatures. Further, a meta-analysis encompassing 1,044 hybridizations showed that LI alone is not sufficient to highlight breast cancer patients with different prognosis. However, for ER positive patients, high LI was associated

with shorter survival times, whereas for ER negative patients, high LI is significantly associated with longer survival. Annotation of LI, in addition to ER status, is important for breast cancer patient prognosis and may have implications for the future treatment of breast cancer.

Keywords Breast cancer · Computational microdissection · Prognosis · Lymphocyte infiltration · Estrogen receptor

Introduction

Breast cancer is the most frequent cancer in women in western countries [14]. The most common breast cancer subtype, the invasive ductal carcinoma (IDC), represents more than 75% of all cases and is not further sub-classified with currently established methods. Many reports on the molecular classification of breast cancer entities are available. Some of these studies attempted to predict patient survival [36, 40]. Others found different molecular subclasses to be associated with clinical parameters like lymph node status or grading [36, 42]. However, only one of the identified prognosis signatures has so far entered the clinical practice [13]. Effective methods to stratify patients for different therapeutic regimens, and consecutively to estimate individual outcome, are still urgently needed. Malignant human tumors are often accompanied by the infiltration of immune cells into the region of tumor cell proliferation. Various publications [6, 21, 27, 31, 33] reported the effects of lymphocyte infiltration (LI) in human solid tumors. However, the prognostic significance of LI in cancer remains controversial. LI was described to be beneficial for patient outcome in certain publications [32], and detrimental in others [15, 17, 18, 23]. In breast

Alberto Calabrò and Tim Beissbarth contributed equally to this manuscript.

Electronic supplementary material The online version of this article (doi:10.1007/s10549-008-0105-3) contains supplementary material, which is available to authorized users.

A. Calabrò · T. Beissbarth · R. Kuner · M. Stojanov · A. Poustka · H. Sültmann (✉)
Division of Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany
e-mail: h.sueltmann@dkfz.de

A. Benner
Division of Biostatistic, German Cancer Research Center, 69120 Heidelberg, Germany

M. Asslaber · K. Zatloukal
Institutes of Pathology, Medical University of Graz, 8036 Graz, Austria

F. Ploner · H. Samonigg
Clinical Oncology, Medical University of Graz, 8036 Graz, Austria

cancer the use of this parameter as a prognostic factor remains a matter of debate [2, 7, 22, 29]. The main reason for this can be attributed to the intrinsic difficulty in separating confounding factors in the analysis: LI is more pronounced in ER negative than in ER positive tumors [38, 43, 44]. Consequently, every breast cancer screening study focusing on LI and patient survival will see its outcome greatly affected by the well-known role of the ER. A second reason for the difficulties in assigning a role to LI for patient survival may be due to the fact that, in contrast to the ER status, the occurrence of LI is not routinely assessed in histopathological reports and consequently, data for comparative studies are often lacking. To overcome these limitations, we developed a microarray-based approach to estimate the presence of LI. We used this estimator for LI to computationally microdissect the gene signatures that distinguish the ER positive and ER negative tumors, and we applied it to a novel microarray dataset encompassing 155 breast cancer samples. We suggest that these signatures reflect more accurately the biological processes which play a role in breast cancer progression. Furthermore, in an individual patient data IPD meta-analysis with altogether 1,044 patient samples from five publicly available breast cancer microarray datasets [10, 30, 36, 37, 41] as well as our own dataset, we found that LI has contrasting effects on the survival of patients suffering from breast cancer, depending on whether ER is expressed or not.

Materials and methods

Sample preparation

155 cryo-preserved human primary breast tumor samples which had been surgically resected between the years 1990 and 1992 were retrieved from the *biobank* of the Medical University of Graz [5]. Before enrollment into the microarray experiments, the tissue samples underwent a careful

re-analysis of the histopathology by two independent pathologists. The sample annotation (Table 1) included patients age at time of surgery (mean = 59 years), estrogen receptor status (negative, n = 61; positive, n = 94), lymphocyte infiltration (negative, n = 18; positive, n = 27) and overall survival time. The study has been approved by the Ethical Committee of the Medical University of Graz. Total cellular RNA was isolated from slices of tissue stored in *RNAlater* (Qiagen, Hilden, Germany) at -80°C using an RNeasy Mini kit (Qiagen) after homogenization with a Mikro-Dismembrator S (Braun Biotech, Melsungen, Germany). The quality of RNA was verified with the Agilent 2100 bioanalyzer (Agilent Technologies, Waldbronn, Germany). Only high-quality RNA samples (28S:18S ribosomal RNA ratio > 1.8) were selected for oligonucleotide microarray hybridization. Amplification, cDNA synthesis and labeling were performed using the TacKle protocol [34].

Microarray processing

The microarrays carried the Human oligonucleotide set V4.0 (Operon technologies, Cologne, Germany), which consists of 35,035 oligonucleotides (average length: 70 bases) representing 33,791 transcripts of the Ensembl human build NCBI-35c, and 28,902 of Refseq. The oligonucleotides were spotted using the VersArray ChipWriter Pro (Bio-Rad, Munich, Germany) and SMP3 pins (Telechem, Sunnyvale, CA) onto epoxysilane-coated glass slides (Nexterion slide E, Schott, Mainz, Germany). Afterwards, microarrays were rehydrated, and the DNA was denatured with boiling water prior to washing with 0.2% sodium dodecyl sulfate, water, ethanol, and isopropanol. The arrays were dried with air pressure.

Microarray hybridization and data analysis

Amplified tumor-derived RNA was labeled with Cy5, and amplified common reference RNA (Stratagene, La Jolla,

Table 1 Studies included in the analysis

Reference	Number of patients	Mean age (years)	ER+/ER-	Follow up (years)	Mapped genes
van't Veer et al. [3]	117	44.2	78/39	–	16
Calabrò et al.	155	58.9	94/61	7.3	18
Bild et al. [36]	158	–	110/48	4.8	13
Miller et al. [37]	247	62.1	213/34	8.2	12
Sorlie et al. [2]	109	58.5	81/28	2.7	13
Sotiriou et al. [38]	98	57.4	67/31	6	11
Van de Vijver et al. [31]	295	43.9	226/69	7.9	16
Total	1062	54.7	791/271	6.7	18

This table enumerates only the patients that in the original study with annotation for the ER status and overall survival. The study from van't Veer et al. is included as part of the validation process for the LI marker genes

CA) was labeled with Cy3. Cy3- and Cy5- labeled samples were purified on Microcon YM-30 columns (Millipore, Bedford, MA). Labeled DNA samples were pooled, purified and resuspended in 50 μ l of 1 \times DIG-Easy hybridization buffer (Roche Diagnostics) containing 10 \times Denhardt's solution and 2 ng/ μ l of Cot1-DNA (Invitrogen, Karlsruhe, Germany). The samples were incubated on the microarray slide for 17 h at 39°C. After removing of unspecific signals, the arrays were scanned with the GenePix 4000B microarray scanner (Axon Instruments, Union City, CA) and analyzed using GenePix Pro 4.1 software (Axon Instruments). Spot intensities were calibrated and transformed by the variance stabilized normalization method using the *arrayMagic* (version 1.16) software tool [12]. The *limma* (version 2.12) software package [35] was used to identify differentially expressed genes. All data analyses were performed using the R (version 2.6) statistical computing environment [1]. The entire dataset is available at GEO [20] under the ID: GSE10510.

Computational microdissection based on quantitative markers

A linear model was applied to test for significant effects of ER status and LI on gene expression, when analyzing microarray data obtained from patient material. First, the microarray data were transformed to log₂ values and quantile-normalized. We used an indicator variable [0,1] to distinguish ER negative and ER positive patients and continuous variables based on the gene expression of marker genes to quantify the presence of LI. Next, we fitted a linear model for each gene, which modeled gene expression measurement according to ER status, LI effect and their potential interaction. The *P*-value for each explanatory factor was computed by using moderated *t*-statistics, including empirical Bayes estimation of the residual standard deviation [35]. *P*-values were adjusted for multiple testing controlling the *false-discovery-rate* (FDR) as defined by Benjamini and Hochberg [24]. All calculations were performed using the R *limma* package. The marker genes for LI were annotated in Suppl. Table A according to the REMARK criteria [28].

Analysis of significant biological function represented in a gene list

Functional gene categories were identified with the assistance of the Ingenuity pathway analysis (IPA) version 5.5.1 (Ingenuity Systems, Mountain View, CA. <https://analysis.ingenuity.com>). IPA's functional analysis compares the data across different biological functions and produces a scored list. The classes defined as “Immune and Lymphatic

System Development and Function”, “Immunological Disease” and “Immune Response” were used to deplete the ER gene list from the genes related to LI. Enrichment of gene ontology (GO) classes was computed based on contingency tables from either of the three gene lists of interest and from the complete array and tested using Fisher's exact tests [8].

Prediction of patient survival

Survival analysis was performed using merged data from the different platforms. ER status was based on the pathologists' annotation, which was available for all 1,044 patients. Presence and intensity of LI were predicted based on the gene expression signatures of marker genes from the microarray studies. The primary analysis was done to test the effects and significance of LI and ER status on patient survival by fitting a Cox proportional hazards regression model including an interaction factor to test for possible interactive effects of ER and LI [16]. A stratification factor for each platform was included in the model to account for the different data sources. To provide quantitative information on the relevance of results, 95% confidence intervals of hazard ratios (HR) were computed. For LI hazard ratio estimates were computed for a change from lower to upper quartile of computed LI intensities.

Stratified Cox models were fitted using ER only, the LI only and the LI in the ER negative and ER positive patients separately, as well as in the subset defined by IDC. The method of Kaplan and Meier was used to estimate survival time distributions. For illustration purposes, the continuous LI variable was dichotomized to build two groups at a ratio of 1:2 (reflecting the ER−/ER+ ratio in the population). Kaplan-Meier plots were drawn for the subgroups defined by ER and dichotomized LI (Fig. 2). All analyses were performed using the R packages *survival* (version 2.32) and *design* (version 2.1). The IDC patient subset was generated by selecting the samples annotated as IDC in Sorlie et al. [36] and in our own platforms, the only two datasets which included such information.

Publicly available datasets for an IPD meta-analysis

Public datasets were obtained from the GEO database [20]. The criteria for the selection of the publicly available dataset were: the presence of annotation for overall survival and the presence of a record for ER status. Hybridizations present in more than one study were counted only once in the survival analysis. Only samples annotated as IDC in the original paper were considered for the IDC only analysis. The information related to the publicly available datasets used for the IPD meta-analysis is summarized in Table 1. The van't Veer dataset (3) was

not included in the IPD meta-analysis as the samples are largely overlapping with the dataset of van de Vijver [41].

Results

Determination of LI through gene expression studies based on marker genes

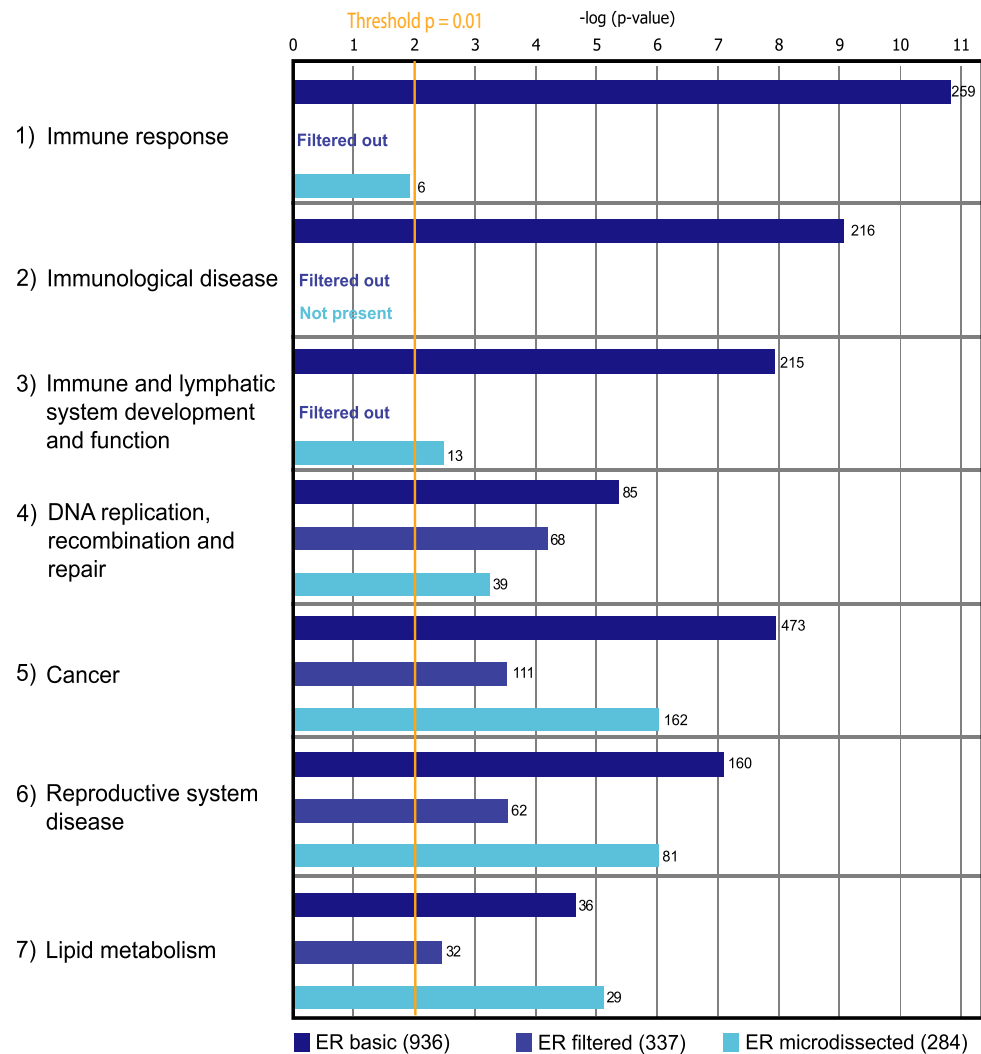
In order to compute an intensity score for lymphocyte infiltration (LI), we selected known lymphocyte-specific marker genes from the literature, including the genes coding for cell surface proteins, immunoglobulin genes, and others. These genes were evaluated for their transcriptional activity using tissue specific EST expression databases (SAGEmap [26]; S-O.U.R.C.E. [19]). The eighteen following specific lymphocyte marker genes (*CCL5*, *CD19*, *CD37*, *CD3D*, *CD3E*, *CD3G*, *CD3Z*, *CD79A*, *CD79B*, *CD8A*, *CD8B1*, *IGHG3*, *IGJ*, *IGLC1*, *CD14*, *LCK*, *LTB*, *MS4A1*; Suppl. Table A, Suppl. Fig. D) were then tested on the expression data of our own microarray platform. In order to use these as quantitative markers in the linear model analysis of cell mixtures and in the survival analysis, the gene expression profiles of each of the features corresponding to these genes were standardized to have mean zero and unit variance. To remove low quality features, spots with fold changes smaller than 1.5 were excluded. Subsequently, for each patient the mean of all remaining standardized features was used as a score for the presence of LI. To validate the performance of this method, we mapped the LI marker genes onto data from an independent array platform and compared the outcome to the LI information. To this end, we selected the dataset by van't Veer [40] and Bertucci [9], which are among the few breast cancer microarray datasets in which the LI annotation based on histopathological characterization of the tissues is provided. We observed significantly positive correlation coefficients between the pathological annotation data and the continuous parameter that we computed based on the molecular markers in both these platforms. The correlation reached a positive correlation coefficient of 0.65 (CI = 0.43–0.75) in our own dataset, a positive correlation coefficient of 0.36 (CI = 0.19–0.50) in the van't Veer dataset and a positive correlation coefficient of 0.47 (CI = 0.26–0.65) in the Bertucci dataset.

Computational microdissection of ER and LI effects

In order to understand the effects of ER expression in breast cancer cells, it is common practice to perform gene expression studies and compute the gene signatures that distinguish ER positive from ER negative patients. However, the published signatures are, however, heavily

influenced by the LI effect and are of limited use for functional interpretations, as they include a large fraction of the genes that are expressed in immune cells [38]. In the van't Veer dataset, 538 out of 2,556 features are annotated as lymphocyte related (20.8%). Similarly in our platform these numbers are 599 out of 2,160 (27.7%). In order to distinguish the different effects, we developed a method to computationally microdissect the gene expression signatures from the various cell types in these complex tissues and applied it to our dataset of 155 breast cancer samples. The linear regression analysis comparing patients with different ER status according to histopathology resulted in 2,160 differentially expressed features (FDR < 0.05), 936 of which were functionally annotated (Suppl. Table C). The analysis of this gene list (“ER basic”) revealed a large portion of genes involved in “immune response” and “activation of leukocytes” that were significantly associated with ER status (Fig. 1). Therefore, we applied our computational microdissection method in order to eliminate the transcriptional variation generated from the infiltrating lymphocytes. The resulting “ER microdissected” gene list analysis revealed only 629 genes associated with ER expression. Of these, 284 were fully annotated with functional categories “DNA replication and repair”, “cancer” and “reproductive system disease” being the most significant ones. In order to evaluate the result of our new method, we performed a third analysis in which we deprived the “ER basic” of all genes with a GO annotation related to expression in lymphocytes. This gene list that we identify as “ER filtered” represents an alternative approach to remove lymphocyte related genes present in the ER positive versus ER negative comparison. The “ER filtered” list consisted of 337 genes, which were sorted according to known gene functions. The genes filtered in this process belonged to functions mainly characteristic for immune cells but among them we also found genes which might be expressed in breast epithelial cell. This is not surprising as the GO annotation may represent several different functions of a gene and therefore has the intrinsic limitation in assigning a gene univocally to a biological function in a specific condition. For example, the ER gene itself that was filtered as related with LI. Figure 1 compares these three different methods at the GO level: the “ER basic”, the “ER filtered” and the “ER microdissected” gene lists. While the most significant terms in the “ER gene list” were associated with immune processes, the analysis of the “ER microdissected” list did not show any significant lymphocyte elements. In summary, the microdissection method helped to remove the GO classes related with lymphocytes and the resulting gene list became more focused on the biological processes relevant to tumor cells. We performed similar analyses with the other publicly available datasets (Table 1). The genes in each analysis

Fig. 1 Analysis of biological function
Comparison among the functional classes in the three cases of ER related gene lists. Sectors 1 to 3 represent the most significant biological function of the “ER basic”. Sectors 4 to 6 show the most significant functions for the “ER filtered” list. Sectors 5 to 7 represent the most relevant classes in the “ER microdissected” list. Numbers show the number of genes accounted in every gene list per class. The “ER filtered” list has been assigned the “Filtered out” value when the biological classes used for the depletion step are considered. “Not present” means no value has been assigned to that biological function



show a high degree of consistency with the list based on our dataset. Consistency was higher in the “ER microdissected” than in the “ER basic” (Suppl. Table B).

Survival analysis based on ER and LI marker genes

To quantitatively evaluate the effect of LI as a parameter indicative for the prognosis of breast cancer patients, we used the previously selected marker genes to evaluate quantitatively the presence of LI in the original tissue samples. The molecular markers indicated their general suitability for the analysis of complex tissues microarray data. We mapped the marker genes to published datasets, which were selected depending on the availability of data on ER and overall survival (OS) of patients. For this purpose, we included data from six microarray studies (Table 1). These six datasets included 1,044 hybridizations with primary breast cancer samples. In the following, we used the immunohistochemistry based on the pathologist’s annotations for ER status and our quantitative estimate for

the LI for survival analysis. The prognostic value of ER expression and LI was evaluated in nested analyses. The presence or absence of LI alone did not reveal a significant impact on OS (stratified Cox regression $P = 0.12$; Fig. 2a). However, as expected, patients with ER positive tumors had a significantly better prognosis when compared to patients with ER negative tumors (stratified Cox regression: P -value < 0.001 ; Fig. 2b). LI was not significantly associated with therapy in the two datasets that were amenable for this analysis (Sotiriou et al. [37] and our own; data not shown).

In an attempt to test for interactive effects of LI and ER, we used a stratified Cox model with interaction factor. We identified a statistically significant interaction between ER and LI (Table 2; Fig. 2c). In this analysis, patients with an increased LI level showed a slightly worse prognosis in the ER positive patients. In the ER positive patients the estimated hazard ratio (HR), when comparing patients at the 75% percentile and at the 25% percentiles of LI, was 1.15 with a 95% confidence interval $CI = 0.94–1.40$. In

Fig. 2 Patient stratification
These Kaplan-Meier plots show the survival of different patient subclasses deduced from all of the 1,044 patients analyzed in the study. The first panel (a), shows the trend of the LI positive and negative patients. Panel b reports the effect of ER on survival. In Panel c the patients are stratified by ER status as well LI, combining the two factors. Panel d shows only the patients annotated as IDC. In order to draw the Kaplan-Meier plots the continuous LI values were converted to categories “+” and “-” describing high and low LI levels

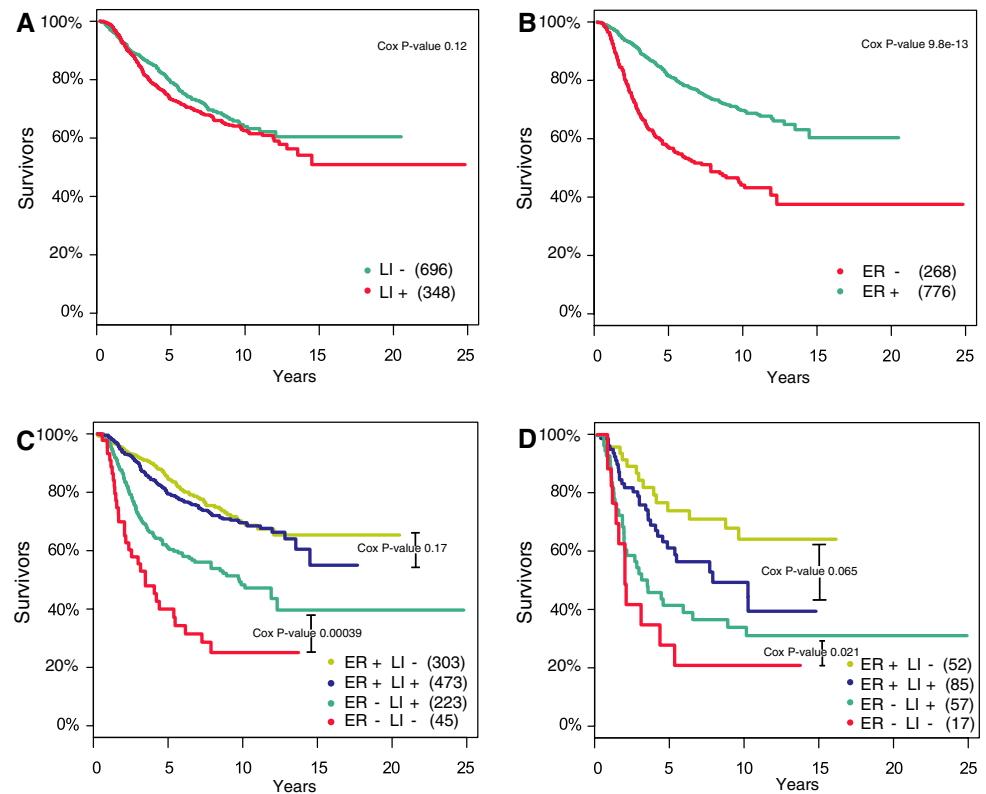


Table 2 Cox-regression analysis to model ER and LI effects on survival *P*-values for LI are based on the Wald statistics for testing the LI effect (LI + LI by ER interaction)

Variable	All 1,044 patients		211 IDC patients	
	Hazard ratio for death (95% CI)	<i>P</i> -value	Hazard ratio for death (95% CI)	<i>P</i> -value
LI (IQR)		0.002		0.01
ER+	1.15 (0.94, 1.40)		1.48 (0.92, 2.38)	
ER–	0.67 (0.53, 0.86)		0.60 (0.41, 0.88)	
ER (ER+ vs. ER–)		<0.001		<0.001
LI (25% percentile)	0.29 (0.21, 0.40)		0.22 (0.13, 0.37)	
LI (50% percentile)	0.36 (0.28, 0.46)		0.31 (0.20, 0.48)	
LI (75% percentile)	0.50 (0.39, 0.65)		0.55 (0.33, 0.94)	

Hazard ratios and confidence intervals (CIs) for LI are computed for an increment corresponding to the interquartile range (IQR) of LI intensities in reference to ER. Likewise the *P*-values for ER are based on the Wald statistics for testing the ER effect (ER + LI by ER interaction). Hazard ratios and confidence intervals (CIs) for ER are computed for the ER positive patients vs. the ER negative patients and adjusted for LI intensity

contrast, ER negative patients with high LI levels show a dramatically better prognosis than ER negative patients with low LI levels (HR = 0.67, CI = 0.53–0.86). To further investigate the difference in survival, we reduced the influence of possible confounding factors due to histopathological heterogeneity. To account for this, we restricted the analysis to the invasive ductal breast carcinomas (IDC) only and performed a similar analysis as described before. Only 211 out of 1,044 samples were annotated as IDC in the two out of six different platforms, since no annotation was available for most of the patients. Figure 2d illustrates

the result of the analysis limited to the IDC subset. Despite of the drastic reduction of the sample numbers, the results remained comparable to those obtained computing all samples (Table 2).

Discussion

Breast cancer is highly heterogeneous with respect to clinical and histopathological appearance as well as to patient survival. The involvement of hormonal and growth

factors in breast cancer progression has been known for decades. Consequently, highly effective therapies are available which target the major promoters of breast cancer, i.e. the estrogen and the epidermal growth factor receptors (ESR1 and ERBB2). However, the prediction of breast cancer therapy success and patient benefit is still poorly developed. Improved patient stratification tools are urgently required before novel therapeutic targets can be identified and specific therapies can be devised. Therefore, we developed a method for the *in silico* microdissection of complex molecular signatures from microarray experiments. It relies on the definition, validation and application of specific genes as representatives for the expression of a much larger number of genes behaving similarly in certain cell types and tissues. We call this approach “computational microdissection” of complex gene expression patterns. It allows the a posteriori separation of different cell types in microarray experiments performed on complex tissue samples (e.g. the tumor cells and their microenvironment). Theoretically, the method can be applied to any tumor entity or array platform. The computational microdissection has proved its usefulness as an approach applicable to circumstances where the physical separation of different cell types before molecular analysis is not possible (as for retrospective microarray studies). In addition, it might be a useful knowledge-based tool to distinguish the different contributions of tumor and stromal cells to cancer development and progression.

Furthermore, we approached the highly debated relevance of LI as a prognostic marker. Our finding is consistent with recent reports [25, 38, 39] and suggest opposite role of LI in ER positive and negative patients. These studies do in fact highlight that a better prognosis of ER negative patients is associated with lymphocyte related genes in the tumor microenvironment. We aimed to estimate the level of LI by using a small set of lymphocyte specific marker genes that enables us to study the interaction between the important prognostic factors ER status and LI, in independent studies even without LI annotation data. Due to the larger sample size in our analysis, it was possible to show a statistically significant effect of LI on survival in ER negative patients. Our study, based on molecular data of more than 1,000 breast cancer samples from multiple centers, suggests that LI when considered in relation to ER status is significantly associated with patient survival in breast tumors. A major contributing factor to overall survival remains ER expression. Nonetheless we identified significant adverse effects of LI on the overall survival of breast cancer patients with or without ER expression: LI is beneficial for ER negative patients but probably unfavorable for ER positive patients. This is particularly true for the patients belonging to the IDC subset. However, since we used a limited number of B- and

T-cell marker genes, it is likely that further gene signatures associated to distinct immune response or tumor-intrinsic characteristics are also present and prognostically relevant.

Our results might reflect intrinsic differences in the biology of breast-tumor subtypes, leading to a difference in tumor immune surveillance depending on the estrogen receptor status. LI occurs as a reaction of the organism to the growing tumor mass and it is known to play a role in generating a signaling microenvironment for the tumor. This stroma might become a source of endocrine factors fostering tumor growth [17]. Recent publications have shown that regulation of the immune system by ER is possible [11] and that the tumor is acting on the signaling microenvironment in order to promote immune tolerance [4]. However, further cellular and molecular analysis is required to unravel the mechanism underlying this hypothesis. Despite these open questions, our results suggest that the acquisition of multiple clinical, histopathological and molecular parameters, combined with IPD meta-analyses of microarray datasets can considerably contribute to breast cancer patient stratification to predict disease outcome. Our results indicate that LI, when combined with ER status, is a relevant prognostic factor for breast cancer. This confirms similar observations of a recent study, reporting the association of LI with HER2-positive breast cancer [3]. We suggest that existing as well as novel specific targets aiming at the treatment of breast cancer patient subgroups should be evaluated in the light of these data.

Acknowledgements We thank Sabrina Balaguer-Puig for excellent technical assistance, Andreas Bunes for retrieving the external datasets and Dirk Ledwinka for IT support. The study was supported by a grant of the German Federal Ministry for Education and Research (NGFN grant 01GR0418; NGFN grant 01GR0450) and the Austrian Genome Research Program (GEN-AU).

Authors contributions MS, RK and TB had the initial ideas for the study. AC collected all the data and performed the experiments. TB did the statistical analysis with the help of AC and AB. MA, FP, KZ and HSa collected and reevaluated the patient samples creating the patient samples and annotation for our own dataset. AC, TB, RK, AP and HSü interpreted the results and wrote the manuscript. AC, RK, TB, AP and HSü contributed in discussions. All authors read and approved the final manuscript.

References

1. The R core project team (2007) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
2. Aaltomaa S, Lipponen P, Eskelinen M, Kosma VM, Marin S, Alhava E et al (1992) Lymphocyte infiltrates as a prognostic variable in female breast cancer. *Eur J Cancer* 28A:859–864. doi: [10.1016/0959-8049\(92\)90134-N](https://doi.org/10.1016/0959-8049(92)90134-N)
3. Alexe G, Dalgin GS, Scandfeld D, Tamayo P, Mesirov JP, DeLisi C et al (2007) High expression of lymphocyte-associated genes in

- node-negative HER2+ breast cancers correlates with lower recurrence rates. *Cancer Res* 67:10669–10676. doi:10.1158/0008-5472.CAN-07-0539
4. Aspord C, Pedroza-Gonzalez A, Gallegos M, Tindle S, Burton EC, Su D et al (2007) Breast cancer instructs dendritic cells to prime interleukin 13-secreting CD4+ T cells that facilitate tumor development. *J Exp Med* 204:1037–1047. doi:10.1084/jem.20061120
 5. Asslaber M, Zatloukal K (2007) Biobanks: transnational, European and global networks. *Brief Funct Genomic Proteomic* 6:193–201
 6. Balkwill F, Mantovani A (2001) Inflammation and cancer: back to Virchow? *Lancet* 357:539–545. doi:10.1016/S0140-6736(00)04046-0
 7. Bates GJ, Fox SB, Han C, Leek RD, Garcia JF, Harris AL et al (2006) Quantification of regulatory T cells enables the identification of high-risk breast cancer patients and those at risk of late relapse. *J Clin Oncol* 24:5373–5380. doi:10.1200/JCO.2006.05.9584
 8. Beissbarth T, Speed TP (2004) Gostat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* 20:1464–1465. doi:10.1093/bioinformatics/bth088
 9. Bertucci F, Finetti P, Cervera N, Charafe-Jauffret E, Mamessier E, Adelaide J et al (2006) Gene expression profiling shows medullary breast cancer is a subgroup of basal breast cancers. *Cancer Res* 66:4636–4644. doi:10.1158/0008-5472.CAN-06-0031
 10. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D et al (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439:353–357. doi:10.1038/nature04296
 11. Biswas DK, Singh S, Shi Q, Pardee AB, Iglehart JD (2005) Crossroads of estrogen receptor and NF-kappaB signaling. *Sci STKE* 2005:pe27. doi:10.1126/stke.2882005pe27
 12. Buness A, Huber W, Steiner K, Sultmann H, Poustka A (2005) arrayMagic: two-colour cDNA microarray quality control and preprocessing. *Bioinformatics* 21:554–556. doi:10.1093/bioinformatics/bti052
 13. Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ (2008) Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol* 26:729–735. doi:10.1200/JCO.2007.14.3222
 14. Clamp A, Danson S, Clemons M (2002) Hormonal risk factors for breast cancer: identification, chemoprevention, and other intervention strategies. *Lancet Oncol* 3:611–619. doi:10.1016/S1470-2045(02)00875-6
 15. Coussens LM, Werb Z (2002) Inflammation and cancer. *Nature* 420(6917):860–867. doi:10.1038/nature01322
 16. Cox DR (1972) Regression models and life tables. *J R Stat Soc [Ser A]*, 187–220
 17. de Visser KE, Eichten A, Coussens LM (2006) Paradoxical roles of the immune system during cancer development. *Nat Rev Cancer* 6:24–37. doi:10.1038/nrc1782
 18. DeNardo DG, Coussens LM (2007) Inflammation and breast cancer. Balancing immune response: crosstalk between adaptive and innate immune cells during breast cancer progression. *Breast Cancer Res* 9:212. doi:10.1186/bcr1746
 19. Diehn M, Sherlock G, Binkley G, Jin H, Matese JC, Hernandez-Boussard T et al (2003) SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. *Nucleic Acids Res* 31:219–223. doi:10.1093/nar/gkg014
 20. Edgar R, Domrachev M, Lash AE (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30:207–210. doi:10.1093/nar/30.1.207
 21. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Page C et al (2006) Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 313:1960–1964. doi:10.1126/science.1129139
 22. Griffith CD, Ellis IO, Bell J, Burns K, Blamey RW (1990) Density of lymphocytic infiltration of primary breast cancer does not affect short-term disease-free interval or survival. *J R Coll Surg Edinb* 35:289–292
 23. Hayes DF (2005) Prognostic and predictive factors revisited. *Breast* 14:493–499. doi:10.1016/j.breast.2005.08.023
 24. Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. *Stat Med* 9:811–818. doi:10.1002/sim.4780090710
 25. Kreike B, van Kouwenhove M, Horlings H, Weigelt B, Peterse H, Bartelink H et al (2007) Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Res* 9:R65. doi:10.1186/bcr1771
 26. Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ et al (2000) SAGEmap: a public gene expression resource. *Genome Res* 10:1051–1060. doi:10.1101/gr.10.7.1051
 27. Marques LA, Franco EL, Torloni H, Brentani MM, da Silva-Neto JB, Brentani RR (1990) Independent prognostic value of laminin receptor expression in breast cancer survival. *Cancer Res* 50:1479–1483
 28. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM (2005) Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 97:1180–1184
 29. Menard S, Tomasic G, Casalini P, Balsari A, Pilotti S, Cascinelli N et al (1997) Lymphoid infiltration as a prognostic variable for early-onset breast carcinomas. *Clin Cancer Res* 3:817–819
 30. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A et al (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA* 102:13550–13555. doi:10.1073/pnas.0506230102
 31. Nixon AJ, Neuberg D, Hayes DF, Gelman R, Connolly JL, Schnitt S et al (1994) Relationship of patient age to pathologic features of the tumor and prognosis for patients with stage I or II breast cancer. *J Clin Oncol* 12:888–894
 32. Oldford SA, Robb JD, Codner D, Gadag V, Watson PH, Drover S (2006) Tumor cell expression of HLA-DM associates with a Th1 profile and predicts improved survival in breast carcinoma patients. *Int Immunol* 18:1591–1602. doi:10.1093/intimm/dx1092
 33. Rilke F, Colnaghi MI, Cascinelli N, Andreola S, Baldini MT, Bufalino R et al (1991) Prognostic significance of HER-2/neu expression in breast cancer and its relationship to other prognostic factors. *Int J Cancer* 49:44–49. doi:10.1002/ijc.2910490109
 34. Schlingemann J, Thuerigen O, Itrich C, Toedt G, Kramer H, Hahn M et al (2005) Effective transcriptome amplification for expression profiling on sense-oriented oligonucleotide microarrays. *Nucleic Acids Res* 33:e29. doi:10.1093/nar/gni029
 35. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments *Stat Appl Genet Mol Biol* 3:Article3
 36. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H et al (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98:10869–10874. doi:10.1073/pnas.191367098
 37. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A et al (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci USA* 100:10393–10398. doi:10.1073/pnas.1732912100
 38. Teschendorff AE, Journee M, Absil PA, Sepulchre R, Caldas C (2007) Elucidating the altered transcriptional programs in breast cancer using independent component analysis. *PLOS Comput Biol* 3:e161. doi:10.1371/journal.pcbi.0030161
 39. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C (2007) An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol* 8:R157. doi:10.1186/gb-2007-8-8-r157

40. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530–536. doi:[10.1038/415530a](https://doi.org/10.1038/415530a)
41. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW et al (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347:1999–2009. doi:[10.1056/NEJMoa021967](https://doi.org/10.1056/NEJMoa021967)
42. West RB, Nuyten DS, Subramanian S, Nielsen TO, Corless CL, Rubin BP et al (2005) Determination of stromal signatures in breast carcinoma. *PLoS Biol* 3:e187. doi:[10.1371/journal.pbio.0030187](https://doi.org/10.1371/journal.pbio.0030187)
43. Whitford P, Mallon EA, George WD, Campbell AM (1990) Flow cytometric analysis of tumour infiltrating lymphocytes in breast cancer. *Br J Cancer* 62:971–975
44. Yao C, Lin Y, Ye CS, Bi J, Zhu YF, Wang SM (2007) Role of interleukin-8 in the progression of estrogen receptor-negative breast cancer. *Chin Med J (Engl)* 120:1766–1772