

Prediction of lymph node involvement in breast cancer from primary tumor tissue using gene expression profiling and miRNAs

A. Smeets · A. Daemen · I. Vanden Bempt · O. Gevaert · B. Claes · H. Wildiers · R. Drijkonigen · P. Van Hummelen · D. Lambrechts · B. De Moor · P. Neven · C. Sotiriou · T. Vandorpe · R. Paridaens · M. R. Christiaens

Received: 21 October 2010 / Accepted: 12 November 2010 / Published online: 30 November 2010
© Springer Science+Business Media, LLC. 2010

Abstract The aim of this study was to investigate whether lymph node involvement in breast cancer is influenced by gene or miRNA expression of the primary tumor. For this purpose, we selected a very homogeneous patient population to minimize heterogeneity in other tumor and patient characteristics. First, we compared gene expression profiles of primary tumor tissue from a group of 96 breast cancer patients balanced for lymph node involvement using Affymetrix Human U133 Plus 2.0 microarray chip. A model was built by weighted Least-Squares Support Vector Machines and validated on an internal and external dataset. Next, miRNA profiling was performed on a subset of 82 tumors using Human MiRNA-microarray chips (Illumina). Finally, for each miRNA the number of significant inverse correlated targets was determined and compared with 1000 sets of randomly chosen targets. A model based on 241

genes was built (AUC 0.66). The AUC for the internal dataset was 0.646 and 0.651 for the external datasets. The model includes multiple kinases, apoptosis-related, and zinc ion-binding genes. Integration of the microarray and miRNA data reveals ten miRNAs suppressing lymph node invasion and one miRNA promoting lymph node invasion. Our results provide evidence that measurable differences in gene and miRNA expression exist between node negative and node positive patients and thus that lymph node involvement is not a genetically random process. Moreover, our data suggest a general deregulation of the miRNA machinery that is potentially responsible for lymph node invasion.

Keywords Breast cancer · Microarrays · miRNA · Lymph node · Prediction

Electronic supplementary material The online version of this article (doi:10.1007/s10549-010-1265-5) contains supplementary material, which is available to authorized users.

A. Smeets (✉) · H. Wildiers · P. Neven · T. Vandorpe · R. Paridaens · M. R. Christiaens
Multidisciplinary Breast Centre University Hospital,
Herestraat 49, B-3000 Leuven, Belgium
e-mail: ann.smeets@uzleuven.be

A. Daemen · O. Gevaert · B. De Moor
Department of Electrical Engineering, University of Leuven,
Leuven, Belgium

I. Vanden Bempt
Department of Microbiology, University Antwerp,
Antwerp, Belgium

O. Gevaert
Department of Radiology, Stanford University
School of Medicine, Stanford, USA

B. Claes · D. Lambrechts
The Vesalius Research Centre, VIB and University of Leuven,
Leuven, Belgium

R. Drijkonigen
Department of Pathology, Jessa Hospital, Hasselt, Belgium

P. Van Hummelen
Center for Cancer Genome Discovery (CCGD), Dana Farber
Cancer Institute, 44 Binney St., Boston, USA

C. Sotiriou
Department of Medical Oncology, Jules Bordet Institute,
Brussels, Belgium

Introduction

Lymph node involvement is the most important prognostic factor in breast [1]. However, 20–30% of node positive patients remain free of distant metastases whereas 20–30% of lymph node negative patients will develop metastasis [2]. Given this limited correlation it remains unclear whether metastasis to distant sites proceeds sequentially from lymph node metastasis or in parallel by a hematogenous route. Moreover, it is unclear whether lymph node metastasis reflects the chronologic age or the biology of the tumor and whether there is an influence of the host on the process of metastasis to the lymph nodes [3–6]. Known factors associated with axillary lymph node metastasis include increasing tumor size, presence of lymphovascular invasion, poor histologic grade, and age [7–9]. However, even in the subgroup of patients with all favorable factors, still 13% had involved lymph nodes [8].

The genetic signature of a primary tumor holds significant prognostic value [10]. However, when looking into the literature it is unclear whether lymph node involvement can be predicted from primary tumor tissue. Huang et al. [11] identified a gene expression pattern associated with the breast tumor's likelihood of having lymph node metastasis at diagnosis. In contrast Weigelt et al. [12] did not find an expression signature that could predict the lymph node status. Therefore, the question remains as to whether or not it is possible to predict lymph node metastases from patients' primary tumor based on gene expression data.

MicroRNAs (miRNAs) are a class of small non-coding RNAs able to negatively regulate gene expression at the post-transcriptional level [13]. With over 700 human miRNAs reported and hundreds of target genes per miRNA, these RNAs represent one of the largest classes of gene regulators. The influence of miRNAs on potentially every cellular pathway makes it likely that deregulated miRNA expression is implicated in cancer progression. Growing evidence suggests that miRNAs can function as oncogenes or tumor [13]. Emerging evidence reveals that the pattern of miRNA expression correlates well with clinicopathological characteristics and disease outcome [14, 15].

We sought to determine whether the presence of metastasis in the regional lymph node could be predicted from the primary tumor. To answer this question, we first compared gene expression profiles of primary tumor tissue. In contrast with previous studies, we selected a group of breast tumors with very homogeneous histological characteristics, balanced for lymph node involvement. Second, in a subgroup of these breast tumors, miRNA expression profiling was performed. Third, the results of both analyses were integrated.

Materials and methods

Selection of patients

Tumor samples were selected from the multidisciplinary breast center database. Cases were chosen from women with primary breast cancer in whom axillary lymph node status was known, and the cohort was balanced for nodal status. We only selected postmenopausal patients with a poorly differentiated, estrogen receptor positive, her2-neu negative invasive ductal cancer. Node positive patients were those with at least one node containing a tumor deposit of >2 mm. Node negative patients were those with pathologically negative nodes. Within 30 min after surgical extirpation, tissues were deep frozen and stored at -80°C . Frozen tumor blocks were thin-sectioned and stained with hematoxylin/eosin and only those that were judged to contain at least 70% viable tumor by area were carried on for RNA extraction. The final collection of tumors for the microarray study consisted of 48 lymph node negative and 48 lymph node positive tumors (training set). 82 out of these tumors were carried on for miRNA profiling. In all cases, ER, PR, and HER-2 status were determined for diagnostic purposes. Nuclear ER and PR immunostaining (Allred score 3–8) were considered ER-positive and PR-positive, respectively. Lack of membranous HER-2 immunostaining (score 0 or 1) was considered HER-2 negative. Table 1 gives an overview of the clinical characteristics of the patients.

Microarrays

RNA extraction

Total RNA was extracted from eight 10–20 μm slides using Trizol reagents and further purified on column (RNAeasy Minikit, Qiagen, Valencia, CA, USA). RNA concentration and purity were determined spectrophotometrically using the Nanodrop ND-1000 (Nanodrop Technologies), and RNA integrity was assessed using a Bioanalyser 2100 (Agilent).

Microarray expression profiling

The analyses detailed here comply with the MIAME (minimal information about a microarray experiment) guidelines established by the Microarray gene expression data society (www.mged.org). cRNA target preparation, hybridization to Affymetrix U133 Plus 2.0 arrays, and washing and array signal acquisition were performed at the Microarray Facility of the Flanders Interuniversity Institute for Biotechnology (VIB) in Belgium.

Table 1 Clinical characteristics of training set

Characteristic	All patients (<i>n</i> = 96)	Lymph node negative (<i>n</i> = 48)	Lymph node positive (<i>n</i> = 48)
Age (y)			
50–60	26	16	10
61–70	30	14	16
>70	40	18	22
Pathological tumor stage (mm)			
T1	26	18	8
T2	55	24	31
T3/T4	15	6	9
Number of positive lymph nodes			
0	48	48	–
1–3	39	–	39
4–9	9	–	9
Progesteron receptor status			
Positive	83	38	45
Negative	13	10	3

Per sample, an amount of 2 µg of total RNA spiked with bacterial RNA transcript positive controls (Affymetrix) was converted to double-stranded cDNA in a reverse transcription reaction. Subsequently, the sample was converted and amplified to antisense cRNA and labeled with biotin in an in vitro transcription reaction. All the steps were carried out according to the manufacturer's protocol (Affymetrix).

All amplification and labeling reactions were performed on a Biomek 3000 ArrayPlex Workstation (Beckman Coulter). A mixture of purified and fragmented biotinylated cRNA and hybridisation controls (Affymetrix) was hybridised on Affymetrix HG U133 Plus 2.0 arrays followed by staining and washing in a GeneChip® fluidics station 450 (Affymetrix) according to the manufacturer's procedures. To assess the raw probe signal intensities, chips were scanned using a GeneChip® scanner 3000 (Affymetrix). Microarray data are available at the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo>), with accession code GSE23177.

Microarray data analysis

The microarray dataset was preprocessed with MAS 5.0, the GeneChip Microarray Analysis Suite 5.0 software (Affymetrix). We used an alternative annotation for the conversion of probes to genes provided by Dai et al. We also took the low signal-to-noise ratio of microarray data into account by unsupervised filtering out genes with low variation across all samples. The 5000 most varying genes were included. Finally, our dataset was standardized per sample across all genes. To be able to include all 96 patients independent of number of positive nodes, we

defined the lymph node ratio (LNratio) as the number of lymph nodes found to be positive divided by the total number of examined lymph nodes [16]. The LN ratio reflects the severity, and the Spearman correlation coefficient was used to identify genes that gradually increase or decrease with changing LN ratio. Models for the prediction of the lymph node status were built by weighted Least-Squares Support Vector Machines [17, 18] on random splits of 96 patients in 10-folds, and this 10-fold cross validation was repeated 100 times for robustness [19]. Genes that were selected in at least half of the cross-validation iterations at a significance level of 0.05 were considered.

Datasets

Besides the training set of 96 patients, an independent internal dataset of 20 patients balanced for lymph node involvement was selected from the multidisciplinary breast center database in the same way as described for the training set. The same approach for RNA extraction and microarray expression profiling was used as well. Also six publicly available datasets on breast cancer for which the lymph node status was provided were considered in the microarray study. In the independent dataset and the datasets for external validation only those patients with the defined characteristics were included (Table 2).

miRNA (82 tumors)

miRNA extraction

RNA from each tumor specimen was extracted from five 10 µm slides using a mirVana MiRNA isolation kit

Table 2 Microarray datasets used in this study

Identifier	Institution	No of samples	No of N0/N+	Microarray platform	AUC	% 241 genes measured
KUL1 (training)	Gasthuisberg	96	48/48	U133plus2	0.659	100
KUL2 (internal validation)	Gasthuisberg	20	9/11	U133plus2	0.646	100
TAM_A	Uppsala, John Radcliffe	13	6/7	U133A	0.786	63.5
TAM_plus2	Guys' Hospital	29	15/14	U133plus2	0.567	97.5
UPP_AB	Uppsala	12	6/6	U133A+B	0.889	93
VDX, UNT, MAINZ ^a	Erasmus, John Radcliffe, Uppsala, Mainz	52	52/0	U133A	0.672 ^a	63.5
Complete external validation		106	79/27		0.651	63.5

^a Because these datasets contain only LN negative patients, they could not be validated separately and were validated together with the patients in TAM_A for which the same platform was used

(Applied biosystems). RNA quality was checked with an Agilent BioAnalyzer lab on chip.

miRNA expression profiling

Biotinylated first strand cDNA was prepared from 5 µg of total RNA and hybridized onto Human MiRNA assay version 1 microarray chips (Illumina) according to the manufacturer's instructions. This assay detects 470 miRNAs described in the miRBase database v9.1 and 265 potential miRNAs identified in a RAKE analysis study [20]. Images of the arrays were acquired using Illumina Beadstudio software, and visual inspection of microarray images revealed no visual artifacts. The signal intensity was calculated for each spot while adjusting for local background. Data were exported to Genemaths XT microarray analysis software (Applied Maths) for further analysis. Spots with signal-to-noise ratio below 200 were marked as missing values, and probes that had 50% or more missing values were excluded, hereby retaining 235 high quality probes (195 validated and 40 potential miRNAs). Signal intensities were \log_2 transformed after setting values smaller than 10^{-6} to 10^{-6} . A median centering normalization algorithm was applied to all remaining probes (i.e., all arrays were scaled with respect to the global median).

The resulting dataset was used for statistical analysis.

To evaluate the miRNAs that are significantly differentially expressed according to lymph node status, significance analysis was performed using the Wilcoxon rank sum test at significance level 0.05 without correction for multiple testing.

Paired miRNA-microarray analysis

The correlation between a miRNA and its computationally predicted targets was evaluated in our dataset of 82 tumors.

Although the targets of many miRNAs are not yet known, databases exist consisting of computationally predicted targets based on sequence complementarity between the miRNA and its target site and on evolutionary target site conservation. We used the microRNA.org database of computationally predicted targets [21] to assign genes to their corresponding miRNA. The spearman correlation coefficient was calculated between each miRNA and its computationally predicted targets. The analysis was done starting from all miRNAs and mRNAs passing preprocessing without supervised selection of only the differentially expressed miRNAs or mRNAs. A one-sided hypothesis test was used to determine the significance of inverse correlation between a miRNA and its target, with significance threshold of 0.05. For each miRNA, the number of significant inverse correlated targets was determined and compared with 1000 sets of randomly chosen targets within the node positive and node negative patients separately, to assess whether miRNA expression is more correlated with its computationally predicted targets than expected with random sets of targets. This allowed identifying miRNAs with lymph node-specific inverse correlation.

Results

Differential gene expression

A 10-fold cross-validation strategy applied to the microarray training set of 96 patients has led to a model based on 241 genes (Table 1 in the Supplementary Appendix) with an average 10-fold area under the ROC curve (AUC) of 0.66. Figure 1 shows the heatmap for the 241 genes and 96 patients.

Next, to validate the classifier, an additional independent set of primary tumors from 20 patients was selected

(9 lymph node negative and 11 lymph node positive patients). The lymph node status was predicted with the classifier and resulted in an AUC of 0.646 (Fig. 2).

Finally, the model was validated on six publicly available datasets, all gathered with the Affymetrix technology. Again, only samples with the defined characteristics were selected. Starting from 4800 patients in six publicly available breast cancer datasets, data from only 79 lymph node negative and 27 lymph node positive patients could be used, due to the requirement for homogeneous characteristics (Table 2). On external datasets, there was a large variance in results, with an AUC ranging from 0.57 to 0.89 (Table 2 and Fig. 2).

To gain insight into the biological meaning of the genes in our expression signature, we used the Gene Ontology Tree Machine [22] and the Molecular Signatures Database [23]. We observed an important increase in apoptosis-related genes and zinc ion-binding genes (Table 3). The model includes 11 gene-mediating BAF57 cell deaths (S100A8, CITED2, SMURF2, KIAA0895, SORBS2, SOD2, PRR6, MYH10, KYNU, PALMD, CDF15).

A subset of genes upregulated in the node positive group are cytokines and genes related to interactions between the tumor and the host (VEGFB, CXCL13, TNFRSF12A, TNFAIP-8, IL1RAP, S100A8).

Differential miRNA expression

Eight miRNAs (miR-195, miR-191, miR-132, miR-203, miR-431, miR-16, miR-30c, miR-30a) were significantly differentially expressed according to lymph node status (negative or positive) with *P*-values varying from 0.014 to 0.05. Only miR-431 was upregulated in the lymph node positive patients while the seven other miRNAs were lower expressed in these patients.

Paired miRNA-microarray analysis

48 miRNAs had a higher number of inversely correlated targets for the node negative group than expected by chance while not having a significant number of inversely correlated targets in the node positive group. For the node positive group, 21 miRNAs had a higher number of inversely correlated targets than expected by chance while not having a significant number of inversely correlated targets in the node negative group. Next, there were also miRNAs which had a lower number of inversely correlated targets than expected by chance. These miRNAs can be hypothesized as not binding to their targets. There were 21 such miRNAs in the lymph node positive group. 17 miRNAs are inactive in the lymph node negative group

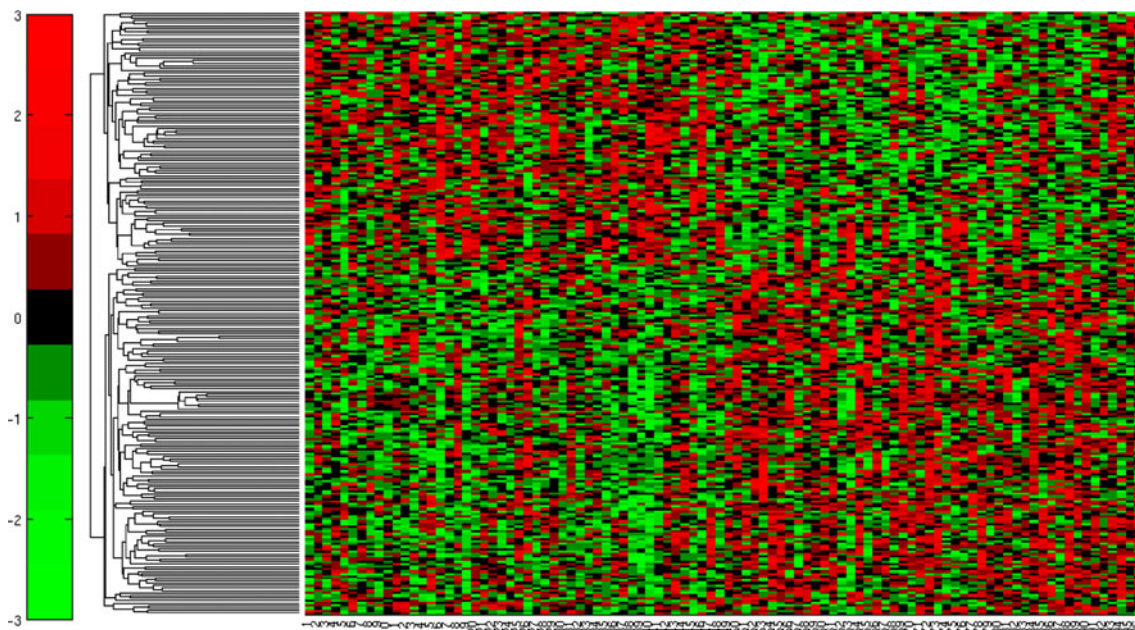


Fig. 1 Heatmap of the 241 genes used to determine lymph node involvement in a series of 96 breast cancer patients. Each column represents the profile of the 241 marker genes for one tumor, and each row represents the relative level of expression of each gene. The tumors are numbered from 1 to 96 on the x axis. The first 48 patients

are lymph node positive, the last lymph node negative. A red color indicates a high level of expression of messenger RNA (mRNA) in the tumor, as compared with the reference level of mRNA, and a green color indicates a low level of expression

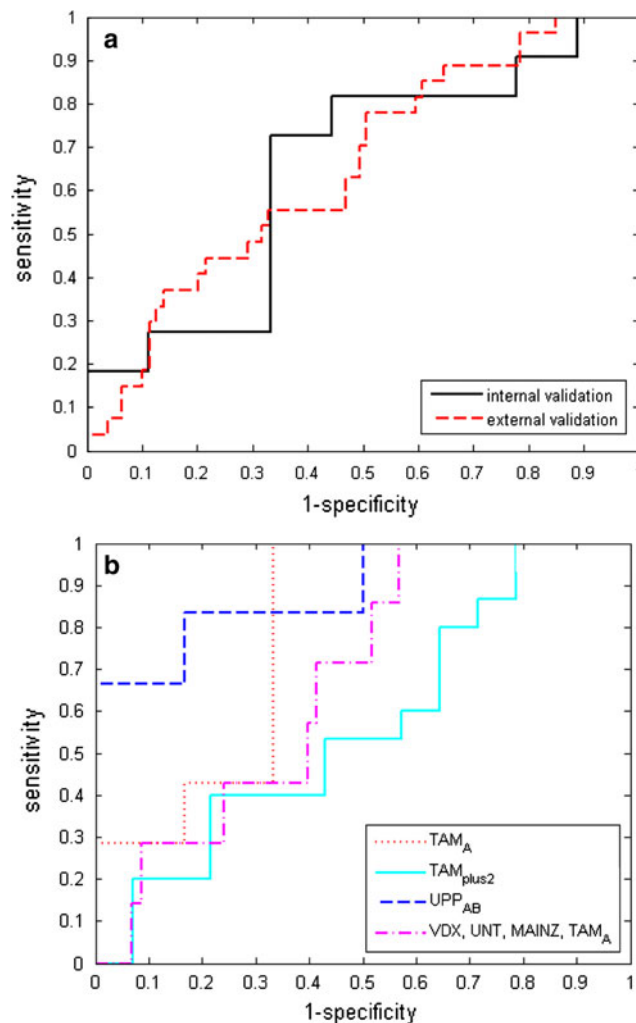


Fig. 2 Area under the receiver operating characteristic curves (AUC) for (a) the internal and external validation and (b) the separate external validation datasets. Because the datasets VDX, UNT and MAINZ only contain lymph node negative patients, they could not be validated separately. We therefore validated them together with the subset of patients in TAM for which the same Affymetrix platform U133A was used

(Table 2 in the Supplementary Appendix). In general, there are more miRNAs significantly inactive and less miRNAs significantly active in the lymph node positive group compared to the lymph node negative group. These results point to general deregulation of the miRNA machinery potentially responsible for lymph node invasion.

Finally, both analyses were intersected, to investigate whether there are miRNAs that are significantly active in one group and significantly inactive in the other group. This holds for ten miRNAs being significantly active in the node negative group and significantly inactive in the node positive group (Table 4). Figure 3 shows the network for these ten miRNAs and represents the node negative patients because these miRNAs repress their targets for the

node negative patients while not performing this task in the node positive patients. This network can thus be hypothesized as suppressing lymph node invasion. Moreover, the network shows that part of the targets are shared (TGFB1, CNIH, SLC5A3, ...). These shared targets can be a good starting point for further investigating whether these genes are related.

Conversely, there is only one miRNA that is significantly active in the lymph node positive group and inactive in the lymph node negative group, hsa-miR-361-5p (Fig. 3). This miRNA and its targets can be hypothesized as promoting lymph node invasion.

Discussion

It is unclear whether lymph node involvement reflects tumor chronology, tumor genetics or whether it serves as a marker of the host response to the tumor. In a meta-analysis of published correlations between various prognostic factors in breast cancer by Mitra and MacRae [4] is concluded that the axillary lymph node status is simply a reflection of the chronologic age of the tumor. On the other hand, several studies show that nodal metastasis is not only a marker of diagnosis at a later point in the natural history of breast cancer but also a marker of an aggressive phenotype [3, 8]. Moreover, one might speculate that besides chronology and phenotype, a weakened host response results in early metastasis to the axillary lymph nodes [9].

Our results demonstrate that measurable differences in gene expression exist between node negative and node positive patients and thus that lymph node involvement is not a genetically random process. However, an area under the ROC curve of only 0.65 indicates that, besides tumor genetics, other factors such as tumor chronology influence lymph node involvement. Moreover, it is possible that we miss important information to define lymph node involvement by studying mainly the tumor and ignoring the continuous interplay between the tumor, the host, and his environment. It has long been recognized that the prognosis of cancer patients is determined not only by the intrinsic properties of the tumor but also by the tumor environment [24] and that interactions with immune and stromal cells play a role in tumor aggressiveness [25, 26]. Tumor-infiltrating lymphocytes are considered a manifestation of the host immune response to tumor and are strongly correlated with lymph node metastasis [27, 28].

Another possibility is that mRNA expression not always correlates with the protein expression and possible functional roles of the genes of interest.

There is an overlap of five genes between our model and the Van't Veer profile (IGFBP5; KRT18; BTG2; PEX12;

Table 3 Functional analysis

Biological process/ molecular function	Overexpressed N+ group	Overexpressed N0 group
Kinases	CXCL13,IGFBP5, MAPK13, MAPK9, PIP4K2C, PIP5K2A, PPP4C, S100A8,S100P, VEGFB	RIPK3
Apoptosis	ACTC1, BCL6, BCL7C, CDKN1A, CEBPB, DRAM1, KRT18P19, MAPK9, MSX1, RPS27L, TNFAIP8, TNFRSF12A, TRAF3IP2, SQSTM1, TSTA3, XRCC4,	BTG2, CITED2, LITAF, MDM4, MYC, NMNAT1, PHB, RIPK3, SOCS2, SOD2, TP5313
Zinc ion binding	ADAM15, ADAM8, AGAP1/CENTG2, BBS1/DPP3, BCL6, CDKN1A, IL1RAP, MID2, RPS27L, SEPX1, SLC39A11, SQSTM1, ZNF385A, ZNF750, ZNF768	C21orf57, ERI2/EXOD1, MAN2A2, MYRIP, NR5A2, PDE5A, PEX12, RNF43, RPAIN, SORBS2, THAP2, TRIM45, TTC3, ZNF383, ZNF397, ZNF420, ZNF514, ZNF568, ZNF569, ZNF570, ZNF571, ZNF585A, ZSWIM5

Table 4 The inverse correlation of miRNAs and the computationally predicted targets

miRNAs active N0 and Inactive N+	miRNAs active N+ and inactive N0
hsa-let-7i	hsa-miR-361-5p
hsa-miR-143	
hsa-miR-16	
hsa-miR-196a	
hsa-miR-26a	
hsa-miR-27a	
hsa-miR-375	
hsa-miR-503	
hsa-miR-519a	
hsa-miR-519b-3p	

CIRBP). As this is far more than expected by chance it might be, in contrast to previous publications [12], an indication of deregulation of the same pathways in lymphatic and hamatogenous metastasis.

Our study is unique because a very homogeneous group of tumors was selected, reducing the problem of multiple testing. Most microarray studies are conducted using a convenience sample of patients for whom tissue is available, but the cohort is often far too heterogeneous to support relevant conclusions [29, 30].

The lymph node ratio was defined as the number of lymph nodes found to be positive divided by the total number of examined lymph nodes and has been shown superior to the number of involved lymph nodes as a prognostic factor [16]. It was used to be able to include all 96 patients independent of the number of positive nodes. The Spearman correlation coefficient was used for the identification of genes that gradually increase or decrease with changing LN ratio.

Our model includes an important increase in kinases, apoptosis-related genes, and zinc ion-binding genes

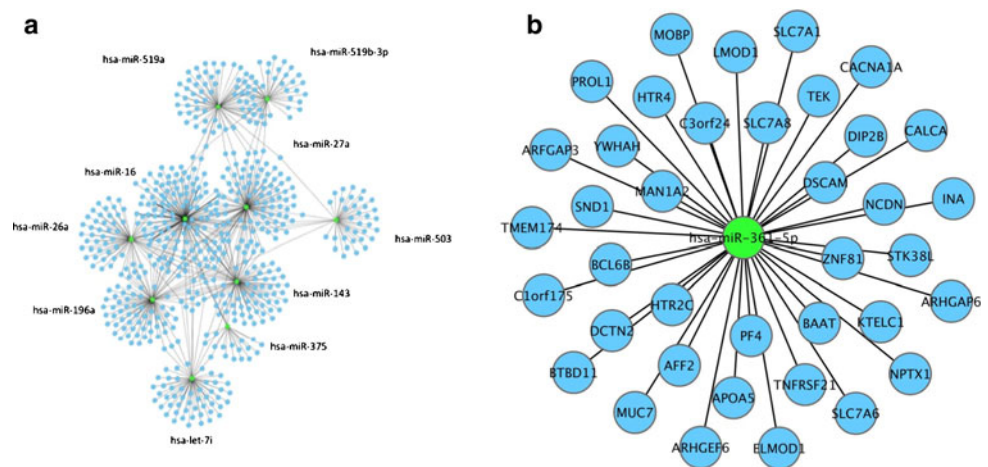
(Table 3). Zinc is essential for cell survival, and zinc chelation leads to cell death by involvement of the NFkappaB pathway and MAP kinase signaling [31]. The model includes 11 gene-mediating BAF57 cell deaths. BAF57-induced cell death involves downregulation of gene products critical for inhibition of apoptosis [32]. A subset of genes upregulated in the node positive group are cytokines and genes related to interactions between the tumor and the host.

IL1RAP and TNFAIP8 are potent pro-inflammatory cytokines, the so called alarm cytokines that are secreted by macrophages and initiate inflammation [33].

TNFAIP8 is an antiapoptotic molecule induced by the activation of the transcription factor NF-kappaB and has been shown to play a role in tumor progression [34]. Moreover, activation of NF-kappaB was shown to be an essential link between inflammation and tumorigenesis [33] and a key event in neoplastic progression [35]. The mechanism by which IL-1 promotes tumor growth remains unclear, though the protein is believed to act via stimulation of adjacent cells to produce VEGF and TNFAIP8.

VEGF-B expression on tumor cells is associated with a poor prognosis [36]. In several reports on malignant tumors, high levels of VEGF-B expression were associated with lymph node metastasis in colorectal cancer and enhanced migration in pancreatic cells through activation of the MAPK pathway [37, 39, 40]. VEGF-B binds on VEGFR-1 and it has been shown that VEGFR-1 positive cells may regulate the homing of tumor cells [40]. IGFBP5, Insulin growth factor binding protein 5, is one of the five genes of our model overlapping with the Van't Veer profile [41]. In several independent studies it was associated with metastasis by protecting cells from apoptosis. Hao et al. [42] found that IGFBP5 was more frequently overexpressed in lymph node metastasis compared with the matched primary cancer tissues. In addition, IGFBP5 was more frequently overexpressed in T1 breast carcinoma that has

Fig. 3 The miRNA-mRNA network of (a) the 10 miRNAs that are active in the node negative group and inactive in the node positive group and (b) the miRNA that is active in the node positive group and inactive in the node negative group



lymph node metastasis compared with T1 carcinoma with no lymph node metastasis [43]. Li et al. [44] recently found that the mRNA level of IGFBP5 positively correlates with the invasion of axillary lymph nodes.

S100A8 is a strong inflammatory kinase, induced by TNFAIP8 and VEGF. It is produced by the primary tumor and attracts monocytes and neutrophils in the premetastatic site [35, 40]. This process is called immune prepping. In this way, the primary tumor influences the environment in premetastatic sites before metastasis. Our findings might reflect the dynamic changes in tumor cell interactions with the microenvironment and suggest that some of the differences between lymph node negative and positive tumors relate to the stromal component rather than to the tumor itself.

The analysis of the inverse correlation between a miRNA and its computationally predicted target genes shows that overall miRNAs were more active and less inactive in the lymph node negative group.

Finally, ten miRNAs were significantly active in the lymph node negative group and at the same time significantly inactive in the lymph node positive group, strengthening the evidence that these miRNAs are suppressing lymph node invasion. When looking into the literature, six of these miRNAs (let-7i, miR-143, miR-16, miR-R-26a, miR-375, and miR-519) have been associated with decreased migration and proliferation of tumor cells and with a better survival [45–50]. It has been proposed that miR-375 is a tumor suppressor miRNA in different types of tumors by inducing apoptosis. There was only one miRNA (has-miR-361-5p) promoting lymph node invasion. These results point to general deregulation of the miRNA machinery potentially responsible for lymph node invasion. However, only an exploratory analysis was performed and in a next step, these results will be used to build predictive models for lymph node involvement.

Conclusion

We can conclude that measurable differences in gene and miRNA expression exist between N0 and N+ patients and thus that lymph node involvement is not a genetically random process. Our model includes an important increase in kinases, apoptosis-related genes, and zinc ion-binding genes. Our findings might reflect the dynamic changes in tumor cell interactions with the microenvironment and suggest that some of the differences between lymph node negative and positive tumors relate to the stromal component rather than to the tumor itself. Moreover, our data suggest a general deregulation of the miRNA machinery potentially responsible for lymph node invasion. For an accurate prediction of lymph node involvement, we hypothesize that, besides the tumor, also the host and its environment have to be taken into account.

Acknowledgment AS is supported by a grant from Research Foundation Flanders (FWO) and Pfizer Oncology.

References

1. Shek LL, Godolphin W (1988) Model for breast cancer survival: relative prognostic roles of axillary nodal status, TNM stage, estrogen receptor concentration, and tumor necrosis. *Cancer Res* 48:5565–5569
2. Weigelt B, Peterse JL, van't Veer LJ (2005) Breast cancer metastasis: markers and models. *Nat Rev Cancer* 5(8):591–602
3. Jatoi I, Hilsenbeck SG, Clark GM, Osborne CK (1999) Significance of axillary lymph node metastasis in primary breast cancer. *J Clin Oncol* 17(8):2334–2340
4. Mitra I, MacRae KD (1991) A meta-analysis of reported correlations between prognostic factors in breast cancer: does axillary lymph node metastasis represent biology of chronology? *Eur J Cancer* 27(12):1574–1583
5. Fisher B, Jeong JH, Anderson S, Bryant J, Fisher ER, Wolmark N (2002) Twenty-five-year follow-up of a randomized trial

- comparing radical mastectomy, total mastectomy, and total mastectomy followed by irradiation. *N Engl J Med* 347(8):567–575
6. Koscielny S, Le MG, Tubiana M (1989) The natural history of human breast cancer. The relationship between involvement of axillary lymph nodes and the initiation of distant metastases. *Br J Cancer* 59:775–782
 7. Yip CH, Taib NA, Tan GH, Ng KL, Yoong BK, Choo WY (2009) Predictors of axillary lymph node metastases in breast cancer: is there a role for minimal axillary surgery? *World J Surg* 33:54–57
 8. Rivadeneira DE, Simmons RM, Christos PJ, Hanna K, Daly JM, Osborne MP (2000) Predictive factors associated with axillary lymph node metastases in T1a en T1b breast carcinomas: analysis in more than 900 patients. *J Am Coll Surg* 191(1):1–6
 9. Wildiers H, Van Calster B, van de Poll-Franse LV et al (2009) Relationship between age and axillary lymph node involvement in women with breast cancer. *J Clin Oncol* 27(18):2931–2937
 10. Sotiriou C, Pusztai L (2009) Gene-expression signatures in breast cancer. *N Engl J Med* 360:790–800
 11. Huang E, Cheng SH, Dressman H et al (2003) Gene expression predictors of breast cancer outcomes. *Lancet* 361(9369):1590–1596
 12. Weigelt B, Wessels LFA, Bosma AJ et al (2005) No common denominator for breast cancer lymph node metastasis. *Br J Cancer* 93(8):924–932
 13. He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 5(7):522–531
 14. Ng EK, Wong CLP, Ma ES, Kwong A (2009) MicroRNAs as new players for diagnosis, prognosis and therapeutic targets in breast cancer. *J Oncol* 2009:305420
 15. Heneghan HM, Miller N, Kerin MJ (2010) MiRNAs as biomarkers and therapeutic targets in cancer. *Curr Opin Pharmacol* 10:1–8
 16. Martinez-Ramos D, Escrig-Sos J, Alcalde-Sanchez M, Torrelar-Ramos A, Salvador-Sanchis JL (2009) Disease-free survival and prognostic significance of metastatic lymph node ratio in T1–T2 N positive breast cancer patients. A population registry-based study in a European country. *World J Surg* 33:1659–1664
 17. Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J (2002) Least squares support vector machines. World Scientific, Singapore
 18. Suykens JAK, Vandewalle J (1999) Least squares support vector machine classifiers. *Neural Process Lett* 9:293–300
 19. Cawley GC (2006) Leave-one-out cross-validation based model selection criteria for weighted LS-SVMs. *Proc Int Joint Conf Neural Netw* 1661–1668
 20. Berezikov E, vanTetering G, Verheul M et al (2006) Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. *Genome Res* 16:1289–1298
 21. Betel D, Wilson M, Gabow A, Marks DS, Sander C (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res* 36:D149–D153
 22. Zhang B, Schmoyer D, Kirov S, Snoddy J (2004) GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies. *BMC Bioinformatics* 5:16
 23. Subramanian A, Tamayo P, Mootha VK et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102(43):15545–15550
 24. Hu M, Polyak K (2008) Microenvironmental regulation of cancer development. *Curr Opin Genet Dev* 18(1):27–34
 25. van Kempen LCLT, Ruiters DJ, van Muijen GNP, Coussens LM (2003) The tumor microenvironment: a critical determinant of neoplastic evolution. *Eur J Cell Biol* 82(11):539–558
 26. Bogenrieder T, Herlyn M (2003) Axils of evil: molecular mechanisms of cancer metastasis. *Oncogene* 22(42):6524–6536
 27. Sheu BC, Kuo WH, Chen RJ, Huang SC, Chang KJ, Chow SN (2008) Clinical significance of tumor-infiltrating lymphocytes in neoplastic progression and lymph node metastasis of human breast cancer. *Breast* 17(6):604–610
 28. Macchetti AH, Marana HRC, Silva JS, de Andrade JM, Ribeiro-Silva A, Bighetti S (2006) Tumor-infiltrating CD4+T lymphocytes in early breast cancer reflect lymph node involvement. *Clinics* 61(3):203–208
 29. Simon R (2005) Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 23(29):7332–7341
 30. Michiels S, Koscielny S, Hill C (2007) Interpretation of microarray data in cancer. *Br J Cancer* 96(8):1155–1158
 31. Ding WQ, Yu HJ, Lind SE (2008) Zinc-binding compounds induce cancer cell death via distinct modes of action. *Cancer Lett* 271(2):251–259
 32. Wang L, Baiocchi RA, Pal S, Mosialos G, Caligiuri M, Sif S (2005) The BRG1- and hBRM-associated factor BAF57 induces apoptosis by stimulating expression of the cyclinomatosis tumor suppressor gene. *Mol Cell Biol* 25(18):7953–7965
 33. Apte RN, Dotan S, Elkabets M et al (2006) The involvement of IL-1 in tumorigenesis, tumor invasiveness, metastasis and tumor-host interactions. *Cancer Metastasis Rev* 25(3):387–408
 34. Zhang C, Chakravarty D, Sakabe I et al (2006) Role of SCC-S2 in experimental metastasis and modulation of VEGFR-2, MMP-1, and MMP-9 expression. *Mol Ther* 13(5):947–955
 35. Németh J, Stein H, Haag D et al (2009) S100A8 and S100A9 are novel nuclear factor kappa B target genes during malignant progression of murine and human liver carcinogenesis. *Hepatology* 50(4):1251–1262
 36. Kanda M, Nomoto S, Nishikawa Y et al (2008) Correlations of the expression of vascular endothelial growth factor B and its isoforms in hepatocellular carcinoma with clinico-pathological parameters. *J Surg Oncol* 98(3):190–196
 37. Fan F, Wey JS, McCarty MF et al (2005) Expression and foundation of vascular endothelial growth factor receptor-1 on human colorectal cancer cells. *Oncogene* 24(16):2647–2653
 38. Wey JS, Fan F, Gray MJ et al (2005) Vascular endothelial growth factor receptor-1 promotes migration and invasion in pancreatic carcinoma cell lines. *Cancer* 104(2):427–438
 39. Gunningham SP, Currie MJ, Han C et al (2001) VEGF-B expression in human primary breast cancers is associated with lymph node metastasis but not angiogenesis. *J Pathol* 193(3):325–332
 40. Hiratsuka S, Watanabe A, Aburatani H, Maru Y (2006) Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis. *Nat Cell Biol* 8(12):1369–1375
 41. Van't Veer L, Dai H, van de Vijver M et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(6871):530–536
 42. Hao X, Sun B, Hu L et al (2004) Differential gene and protein expression in primary breast malignancies and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. *Cancer* 100(6):1110–1122
 43. Nishidate T, Katagiri T, Lin ML et al (2004) Genome-wide gene-expression profiles of breast-cancer cells purified with laser microbeam microdissection: identification of genes associated with progression and metastasis. *Int J Oncol* 25(4):797–819
 44. Li X, Cao X, Li X, Zhang W, Feng Y (2007) Expression level of insulin-like growth factor binding protein 5 mRNA is a prognostic factor for breast cancer. *Cancer Sci* 98(10):1592–1596
 45. Abdelmohsen K, Kim MM, Srikantan S, et al. (2010) miR-519 suppresses tumor growth by reducing HuR levels. *Cell Cycle* 9(7):1538–4101

46. Yang N, Kaur S, Volinia S et al (2008) MicroRNA microarray identifies Let-7i as a novel biomarker and therapeutic target in human epithelial ovarian cancer. *Cancer Res* 68(24):10307–10314
47. Bhattacharya R, Nicoloso M, Arvizo R et al (2009) MiR-15a and MiR-16 control Bmi-1 expression in ovarian cancer. *Cancer Res* 69(23):9090–9095
48. Tsukamoto Y, Nakada C, Noguchi T et al (2010) Micro-RNA-375 is downregulated in gastric carcinomas and regulates cell survival by targeting PDK1 and 14-3-3. *Cancer Res* 70(6):2339–2349
49. Arndt GM, Dossey L, Cullen LM et al (2009) Characterization of global microRNA expression reveals oncogenic potential of miR-145 in metastatic colorectal cancer. *BMC Cancer* 9:374
50. Villanueva A, Hoshida Y, Toffanin S et al (2010) New strategies in Hepatocellular carcinoma: genomic prognostic markers. *Clin Cancer Res* 16(19):4688–4994