The International Journal of Biological Markers, Vol. 17 no. 1, pp. 5-23

REVIEW ARTICLE

A low-density DNA microarray for analysis of markers in breast cancer

M. Lacroix[1], N. Zammatteo[2], J. Remacle[2], G. Leclercq[1]

[1] Laboratoire Jean-Claude Heuson de Cancérologie Mammaire, Institut Jules Bordet, Université Libre de Bruxelles
[2] Laboratoire de Biochimie et de Biologie Cellulaire, Facultés Universitaires Notre Dame de la Paix, Namur - Belgium

ABSTRACT: Breast cancer remains a major cause of death in women from Western countries. In the near future, advances in both nucleic acids technology and tumor biology should be widely exploited to improve the diagnosis, prognosis, and outcome prediction of this disease. The DNA microarray, also called biochip, is a promising tool for performing massive, simultaneous, fast, and standardized analyses of multiple molecular markers in tumor samples. However, most currently available microarrays are expensive, which is mainly due to the amount (several thousands) of different DNA capture sequences that they carry. While these high-density microarrays are best suited for basic studies, their introduction into the clinical routine remains hypothetical. We describe here the principles of a low-density microarray, carrying only a few hundreds of capture sequences specific to markers whose importance in breast cancer is generally recognized or suggested by the current medical literature. We provide a list of about 250 of these markers. We also examine some potential difficulties (homologies between marker and/or variant sequences, size of sequences, etc.) associated with the production of such a low-cost microarray. (Int J Biol Markers, 2002; 17: 5-23)

Key words: Breast cancer, Molecular markers, Diagnosis, Prognosis, Prediction, Microarray, Chip, Low-density, mRNA, DNA

INTRODUCTION

In Western countries about one of 11 women will develop a breast carcinoma in their lifetime. Despite considerable progress in tumor detection as well as in radio-, chemo-, and hormone therapy, more than one third of patients still succumb to the disease. In most cases death results from the dissemination of cancer cells and their proliferation at secondary sites.

To ensure a better characterization and treatment of the extensively heterogeneous breast tumors, new approaches are needed to complement the classical clinicopathological analysis. In particular, tools that exploit the most recent molecular biology knowledge and technological advances would be required. They should mainly help the clinician to predict, at the highest level of precision, the evolution of tumors, including their sites of metastasis, as well as the response of these tumors to different therapies. Moreover, they could accelerate the development of novel therapeutic agents, since this process increasingly requires the understanding of the molecular bases of breast cell transformation and tumor development.

One major objection raised against the generalized use of new technologies in the oncological routine is their cost. In this paper we examine how this problem might be solved in the peculiar case of the DNA microarray. We suggest the design of a breast cancer-oriented microarray carrying a limited number of oligonucleotide sequences specific to recognized or potential markers in breast oncology. By screening the medical literature, we have identified and listed about 250 of these markers.

MOLECULAR MARKERS

Current tumor characterization procedures include clinical examination, imaging, microscopic analysis (cytology, study of tumor margins) – all providing mainly qualitative data – and the measurement of a few specific molecules (molecular markers) by immunohistochemical, biochemical, and molecular biology techniques, the latter providing quantitative or semiquantitative data.

There is a growing need for additional reliable molecular markers, since the perfect marker for breast cancer may not even exist. As a matter of fact, the ideal marker

0393-6155/005-19$09.50/0

should be produced solely by cancer cells or in their immediate vicinity; it should be specific and sensitive, and easily measurable in a reproducible way through simple, fast, and inexpensive techniques; it should allow estimation of the tumor volume and assessment of the efficacy of therapy and might itself constitute a highly tumor-specific therapeutic target. In reality, the highly heterogeneous nature of breast tumors makes their exhaustive description based on the expression levels of only a few genes impossible. This is further hampered by the diversity of processes (proliferation, adhesion, proteolysis, chemoresistance, hormone sensitivity) that characterize tumor behavior. Accounting for the complexity of tumors undoubtedly requires recourse to a panel of selected indicators.

Studies in the last 10 years have unveiled the great molecular complexity of breast carcinomas. Thousands of human genes have been cloned and characterized from various sources. For more than 250 of them, listed in Table I, it has been shown that their corresponding mRNA and/or protein level may vary in breast tumors or cell lines. These variations can, or potentially could

- help to identify and characterize tumors (diagnostic markers, or indicators);
- allow to foresee the evolution and the complications - notably metastasis - of tumors (prognostic markers);
- provide an estimation of the patient's responsiveness to specific therapy (predictive markers).

Besides their possible value as clinical indicators, some of these markers could also be used as therapeutic targets.

For a given molecular marker to earn the label of "reliable clinical marker" it must undergo extensive, strictly controlled and reproducible expression studies, often covering years. To date this process has been completed for a few candidates only. For instance, the longest established breast cancer molecular indicator, the estrogen receptor-alpha (ER-alpha, gene *ESR1*), has been quantified for more than 30 years in tumor samples. This has led to its definitive acceptance both as a prognostic indicator (its expression is associated with longer survival) and a predictor of patient responsiveness to antiestrogens, particularly in node-negative patients (reviewed in (1)). c-erbB-2 (*ERBB2*) is another marker whose prognostic relevance has been demonstrated by numerous studies. More recently, it has been suggested that its overexpression, observed in 10% to 40% of breast tumors, might predict patient responsiveness to chemotherapeutic agents such as doxorubicin and paclitaxel (reviewed in (2)). However, confirmation of the latter property requires further investigation. Thus, the c-erbB-2 expression level may at present be considered both a *recognized* (for prognosis) and a *potential* (for prediction) reliable indicator. Moreover, c-erbB-2 could also serve as a target for antibody-based therapeutic strategies, which emphasizes the need to evaluate its expression level in all tumor samples. Urokinase-type plasminogen activator (uPA, gene *PLAU*) and plasminogen activator inhibitors-1 and 2 (*SERPINE1* and *SERPINB2*) are three additional markers that have recently been introduced into the clinical routine after extensive investigation of their expression levels in breast tumors. They all possess prognostic and predictive properties. Moreover, uPA might be the target of therapeutic strategies based on the use of antiproteases (reviewed in (3)).

The clinical importance of several other markers has also been repeatedly suggested by studies examining variations in their mRNA and/or protein amounts in tumors. Among these are bcl-2 (gene *BCL2*), cathepsin D (*CTSD*), cyclin D1 (*CCND1*), epidermal growth factor receptor (*EGFR*), p53 (*TP53*), progesterone receptor (*PgR*), pS2 (*TFF1*), and urokinase receptor (*PLAUR*) (3-5).

Apart from the few markers that are generally recognized as reliable clinical indicators, a vast majority still require extensive investigation before acquiring (or not) this status. In view of the great amount of work involved in such studies, we clearly have to consider the opportunity of using recently developed techniques for massive, parallel, rapid, and standardized determination of gene expression in biological samples. These techniques might be helpful not only to clinicians but also to basic researchers wishing to gain access to the expression of a large number of tumor molecules in order to improve their knowledge of pathways underlying the occurrence and evolution of breast carcinomas.

MARKER ANALYSIS - DNA MICROARRAY

Gene expression may be measured at both transcriptional (mRNA) and translational (protein) levels. Because they ultimately support cell functions and tumor properties such as growth, angiogenesis, and dissemination, proteins and enzyme activities are the target of choice for investigations in tumor samples. Indeed, their systematic and large-scale measurement should be performed in the future by techniques including *antibody-based arrays*, allowing the simultaneous measurement of multiple proteins in a sample (6); *tissue arrays*, to measure one protein in multiple tumor samples (7, 8); and, more generally, *proteomics*. The latter term refers to the description of proteome (protein expression level and post-translational modifications) by the association of 2D electrophoresis, laser, mass spectrometry, and computerized image and data analysis (9, 10). Although promising, such methods are cumbersome and time-consuming and require further technical development before being applicable on a large scale and at a reasonable cost. At present it is easier to assay for gene expression in a biological sample by qualitative and quantitative analysis of its mRNA population (transcriptome). This may be performed through the use of the so-called DNA microarrays or DNA biochips

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS

Gene name	Gene product name (s)	Cell function or tumor property	Locus	PubMed identifier
ABCB1	ATP-binding cassette, subfamily B, member 1; P-glycoprotein; Multidrug resistance protein 1 (MDR1)	Chemoresistance	7q21	10092063
ABCC1	ATP-binding cassette, subfamily C, member 1; Multidrug resistance-associated protein (MRP, MRP-1)	Chemoresistance	16p13.1	9815925
ABCG2	ATP-binding cassette, subfamily G, member 2; Breast cancer resistance protein (BCRP); Placenta-specific ATP-binding cassette transporter (ABCP)	Chemoresistance	4q22	10070941
AIB3	Thyroid hormone receptor binding protein (TRBP); Cancer-amplified transcriptional coactivator ASC-2; Nuclear receptor coactivator RAP250; Peroxisome proliferator-activated receptor interacting protein (PRIP); KIAA0181 protein	Hormone sensitivity	20q11	10567404
ANGPT1	Angiopoietin 1	Angiogenesis	8q22	11250735
AP1S2	Adaptor-related protein complex 1, sigma 2 subunit; Sigma1B subunit of AP-1 clathrin adaptor complex		19q13.2-19q13.3	11034073
APC	Adenomatous polyposis of the colon		5q21-5q22	8318422
APPBP2	Amyloid beta precursor protein (cytoplasmic tail)-binding protein 2; Kinesin light chain-related protein; Protein interacting with APP tail 1 (PAT1); KIAA0228 protein		17q22-17q23.1	11034067
AR	Androgen receptor; Dihydrotestosterone receptor (DHTR)	Hormone sensitivity	Xq11-Xq12	2249895
ARHC	Ras homolog gene family, member C (rhoC)		1p21-1p13	10499627
ATM	Ataxia telangiectasia mutated		11q22-11q23	9766563
BACE2	Beta-site APP-cleaving enzyme 2; Aspartic-like protease 56kDa (ALP56)	Protein cleavage	21q22	10838186
BAG1	BCL-2-associated athanogene	Cell cycle regulation	9p12	10430086
BAK1	BCL-2-antagonist/killer 1	Cell cycle regulation	6p21.3-6p21.2	9500190
BAX	BCL2-associated X protein	Cell cycle regulation	19q13.3-19q13.4	9495351
BCAR1	Breast cancer anti-estrogen resistance 1; p130Cas adaptor protein	Hormone sensitivity	16q22-16q23	10639512
BCAR3	Breast cancer anti-estrogen resistance 3; Novel SH2-containing protein (NSP2)	Hormone sensitivity		9582273
BCAS1	Breast carcinoma amplified sequence 1; Amplified in breast carcinoma 1 (AIBC1); Novel amplified in breast cancer 1 (NABC1)		20q13.2-20q13.3	9671742
BCAS2	Breast carcinoma amplified sequence 2; Putative spliceosome-associated protein; DAM1-encoded protein		1p21-1p13.3	10403562
BCL2	B-cell CLL/lymphoma 2	Cell cycle regulation	18q21.33	7896458

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS (contd.)

Gene name	Gene product name (s)	Cell function or tumor property	Locus	PubMed identifier
BCL2L1	Bcl-2-like 1; Bcl-X	Cell cycle regulation	20	10430086
BECN1	Bcl-2-interacting protein beclin-1	Cell cycle regulation	17q21	10395800
BIN1	Bridging integrator 1; Amphiphysin II		2q14	10652430
BRCA1	Breast cancer 1, early onset; Breast cancer type I susceptibility protein		17q21	10371343
BRCA2	Breast cancer 2, early onset Breast cancer type II susceptibility protein		13q12-13q13	10498873
BRMS1	Breast cancer metastasis suppressor 1			10850410
BSG	Basigin; Extracellular matrix metalloproteinase inducer (EMMPRIN) ; Tumor collagenase stimulatory factor (TCSF); CD147	Proteolysis	19p13.3	9154157
BZRP	Benzodiazepin receptor, peripheral		22q13.31-22qter	11062691
CALCR	Calcitonin receptor		7q21.3	9399657
CALML3	Calmodulin-like 3		10pter-10p13	10935476
CAV1	Caveolin-1; Caveolae protein, 22kDa		7q31.1	9717814
CCND1	Cyclin D1; Parathyroid adenomatosis 1 (PRAD1); Bcl-1	Cell cycle regulation	11q31.1	10706127
CD24	CD24 antigen		6q21	10037815
CD44	CD44 antigen; Hyaluronate receptor; Hermes antigen gp90 homing receptor	Adhesion	11p13	9131272
CD9	CD9 antigen; p24 antigen		12p13	9736046
CDC25A	Cell division cycle 25A	Cell cycle regulation	3p21.3-3p21.2	11099949
CDH1	Cadherin-1; E-cadherin (epithelial); Uvomorulin	Adhesion	16q22.1	8604681
CDH11	Cadherin-11; OB-cadherin (osteoblast)	Adhesion	16q22.1	10029089
CDH13	Cadherin-13; H-cadherin (heart)	Adhesion	16q24.2-16q24.3	8673923
CDH2	Cadherin 2; N-cadherin (neuronal)	Adhesion	18q12.1	10545506
CDKN1A	Cyclin-dependent kinase inhibitor 1A; p21/waf1/cip1	Cell cycle regulation	6p21.2	7591270
CDKN1B	Cyclin-dependent kinase inhibitor 1B; p27/kip1	Cell cycle regulation	12p13.1-12p12	10936889

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS (contd.)

Gene name	Gene product name (s)	Cell function or tumor property	Locus	PubMed identifier
CDKN1C	Cyclin-dependent kinase inhibitor 1C; p57/waf2	Cell cycle regulation	11p15.5	10537284
CDKN2A	Cyclin-dependent kinase inhibitor 2A; p16/ink4/mts1	Cell cycle regulation	9p21	7591270
CEACAM5	Carcinoembryonic antigen-related cell adhesion molecule 5; CD66e	Adhesion	19q13.1-19q13.2	10592099
CLDN1	Claudin 1; Senescence-associated membrane protein 1 (SEMP1)	Adhesion	3q28-3q29	11071387
CSE1L	Chromosome segregation 1-like; Cellular apoptosis susceptibility protein (Cas)	Cell cycle regulation	20q13	10969801
CSF1	Colony stimulating factor 1; Macrophage-colony stimulating factor (M-CSF, M-CSF1)		1p21-1p13	1334964
CSF1R	Colony stimulating factor 1 receptor; c-fms-encoded protein		5q33-5q35	1334964
CSPG2	Chondroitin sulfate proteoglycan 2; Versican		5q12-q14	10353737
CST6	Cystatin E/M	Proteolysis	11q13	8995380
CSTA	Cystatin A; Stefin A	Proteolysis	3q21	1515089
CTGF	Connective tissue growth factor; Insulin-like growth factor-binding protein 8 (IGFBP8)		6q23.1	9076950
CTNNB1	Catenin (cadherin-associated protein), beta 1 (88kDa)		3p21	10759547
CTSB	Cathepsin B	Proteolysis	8p22	10738903
CTSD	Cathepsin D	Proteolysis	11p15.5	9888472
CTSL	Cathepsin L	Proteolysis	9q21-9q22	10738903
CYP19	Cytochrome P450, subfamily XIX; Aromatase; Estrogen synthetase	Oxidative metabolism	15q21.1	9797013
CYP1A1	Cytochrome P450, subfamily I (aromatic compound-inducible), polypeptide 1; Cytochrome P4501A1	Oxidative metabolism	15q22-15q24	10357772
CYP1B1	Cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1; Cytochrome P4501B1	Oxidative metabolism	2p21	10357772
CYR61	Cysteine-rich, angiogenesis inducer, 61	Angiogenesis	1p31-1p32	11059746
DCC	Deleted in colorectal carcinoma		18q21.3	8318422
DCN	Decorin		12q23	11054714
DDIT3	DNA damage-inducible transcript 3; Growth arrest- and DNA damage-inducible gene 153 (GADD153); C/EBP-homologous protein (CHOP)		12q13.1-12q13.2	

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS (contd.)

Gene name	Gene product name (s)	Cell function or tumor property	Locus	PubMed identifier
DFNA5	Deafness, autosomal dominant 5; Inversely correlated to estrogen receptor expression-1 (ICERE-1)		7p15	9523727
DRIM	Downregulated in metastasis		12	9673349
DSP	Desmoplakin (DP1, DP2)	Adhesion	6p24	
ECGF1	Endothelial cell growth factor 1, platelet-derived (PD-ECGF); Thymidine phosphorylase (TP)	Angiogenesis	22q13.33	9041202
EDN1	Endothelin-1	Angiogenesis	6p24.1	10672062
EDN3	Endothelin-3	Angiogenesis	20q13.2-20q13.3	10672062
EDNRB	Endothelin receptor, type B (ETB, ETRB)	Angiogenesis	13q22	11250735
EGF	Epidermal growth factor		4q25	9191980
EGFR	Epidermal growth factor receptor		7p12	10326794
EIF4E	Eukaryotic translation initiation factor 4E (EIF4-E); Cap-binding protein	Protein synthesis	4q21-4q25	10216944
EMP1	Epithelial membrane protein-1 (EMP-1); Tumor-associated membrane protein homolog (TMP); Progression Associated Protein (PAP); B4B; CL-20		12p12.3	9066625
EMS1	Cortactin; Amplaxin	Adhesion	11q13	10706127
ENG	Endoglin; CD105	Angiogenesis	9q34.1	11250735
ERBB2	c-erbB-2; Herstatin (HER-2); Neu		17q11.2-17q12	10706127
ERBB3	c-erbB-3		12q13	10918187
ERBB4	c-erbB-4		2q33.3-2q34	10918187
ESR1	Estrogen receptor 1; Estrogen receptor-alpha	Hormone sensitivity	6q25.1	9816251
ESR2	Estrogen receptor 2; Estrogen receptor-beta	Hormone sensitivity	14q	10554009
ETS1	c-ets-1		11q23.3	9247254
ETV4	Ets translocation variant 4; E1A enhancer binding protein (E1AF); Pea3 transcription factor		17q21	9380403
F2R	Coagulation factor II receptor; Thrombin receptor (TR); Protease-activated receptor 1 (PAR1)	Angiogenesis	5q13	11250735
FABP3	Fatty acid binding protein 3; Mammary-derived growth inhibitor (MDGI)	Fatty acid metabolism Cell cycle regulation	1p33-1p32	8895735

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS (contd.)

Gene name	Gene product name (s)	Cell function or tumor property	Locus	PubMed identifier
FABP7	Fatty acid binding protein 7; Mammary-derived growth inhibitor related (MRG)	Fatty acid metabolism Cell cycle regulation	6q22-6q23	9242429
FGF2	Fibroblast growth factor 2; Fibroblast growth factor, basic (bFGF)	Angiogenesis	4q26-4q27	9041202
FGF8	Fibroblast growth factor 8 (androgen-induced)	Angiogenesis	10q24	10023681
FGFR1	FGF receptor-1; Fms-related tyrosine kinase 2	Angiogenesis	8p11.2-8p11.1	10706127
FGFR2	FGF receptor-2	Angiogenesis	10q26	7705943
FHIT	Fragile histidine triad; Bis(5′-adenosyl) triphosphatase; Diadenosine triphosphate (Ap3A) hydrolase	Cell cycle regulation	3p14.2	9824201
FIGF	c-fos-induced growth factor; Vascular endothelial growth factor D (VEGFD)	Angiogenesis	Xp22.31	10551327
FLT1	Fms-related tyrosine kinase 1; VEGF receptor 1	Angiogenesis	13q12	7821921
FLT4	Fms-related tyrosine kinase 4; VEGF receptor 3	Angiogenesis	5q34-5q35	10329591
FN1	Fibronectin 1	Adhesion	2q34	9306263
GATA3	GATA-binding protein 3		10p15	10037815
GJA1	Gap junction protein, alpha 1, 43kDa; Connexin 43 (Cx43)	Adhesion	6q21-6q23.2	10463615
GJB1	Gap junction protein, beta 1, 32kDa; Connexin 32 (Cx32)	Adhesion	Xq13.1	11280719
GJB2	Gap junction protein, beta 2, 26kDa; Connexin 26 (Cx26)	Adhesion	13q11-13q12	1324944
GRN	Granulin		17q21.32	11134521
GSN	Gelsolin		9q33	8895730
GSTM3	Glutathione-S-transferase M(u)3	Chemoresistance	1p13.3	10037815
GSTP1	Glutathione-S-transferase P(i)1	Chemoresistance	11q13	2466554
HGF	Hepatocyte growth factor; Scatter factor (SF); Hepapoietin A	Angiogenesis	7q21.1	9024713
HIF1A	Hypoxia-inducible factor 1, alpha subunit; Basic helix-loop-helix transcription factor	Angiogenesis	14q21-14q24	10582706
HMG1	High mobility group protein 1		13q12	10944450
HPSE	Heparanase	Angiogenesis	4q21.3	10395325
HSPA5	Heat-shock protein A5; 78 kDa glucose-regulated stress protein (GRP78)	Protein stabilization	9q33-9q34.1	10752676
HSPB1	Heat shock 27kDa protein 1	Protein stabilization	7q	1988702

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS (contd.)

Gene name	Gene product name (s)	Cell function or tumor property	Locus	PubMed identifier
HSPCA	Heat shock 90kDa protein 1, alpha	Protein stabilization	1q21.2-1q22	8878452
HSPCB	Heat shock 90kDa protein 1, beta	Protein stabilization	6p12	8878452
HXB	Hexabrachion; Tenascin(-C) (Tn, TN-C); Cytotactin		9q33	9306263
IBSP	Integrin-binding sialoprotein; Bone sialoprotein (BSP)		4q21-4q25	10418827
ICAM1	Intercellular adhesion molecule-1; Rhinovirus receptor; CD54 antigen	Adhesion	19p13.3-19p13.2	9892213
ID1	Inhibitor of DNA binding 1; Dominant negative helix-loop-helix protein		20q11	10728695
IL11	Interleukin 11; Adipogenesis inhibitory factor (ADIF)	Inflammatory response	19q13.3-19q13.4	9619855
IL1A	Interleukin 1, alpha	Inflammatory response	2q14	10049521
IL1B	Interleukin 1, beta	Inflammatory response	2q14	9241076
IL6	Interleukin 6; Interferon, beta 2	Inflammatory response	7p21	10813718
IL6R	Interleukin-6 receptor; CD126 antigen	Inflammatory response	1q21	10813718
IL6ST	Interleukin 6 signal transducer; Oncostatin M receptor; gp130; CD130 antigen	Inflammatory response	5q11	10813718
IL8	Interleukin-8; Monocyte-derived neutrophil-activating protein (MONAP); Monocyte-derived neutrophil chemotactic factor (MDNCF)	Angiogenesis	4q13-4q21	9378554
IMP-1	IGF-II mRNA-binding protein 1; Coding region determinant-binding protein (CRDBP)		17	10850408
ING1	Inhibitor of growth 1 family, member 1; p33ING1 protein	Cell cycle regulation	13q34	10754201
ITGA6	Integrin, alpha 6	Adhesion	2	10545019
ITGAV	Integrin, alpha V; Vitronectin receptor alpha polypeptide; CD51 antigen	Adhesion	2q31-2q32	10651305
ITGB1	Integrin, beta 1; Fibronectin receptor, beta polypeptide; CD29 antigen	Adhesion	10p11.2	8388172
ITGB3	Integrin, beta 3; Platelet glycoprotein IIIa; CD61 antigen	Adhesion	17q21.32	8727100
JUP	Junction plakoglobin; Gamma catenin	Adhesion	17q21	10640987

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS (contd.)

Gene name	Gene product name (s)	Cell function or tumor property	Locus	PubMed identifier
KAI1	Kangai-1; Suppression of tumorigenicity 6 (ST6); CD82 antigen		11p11.2	9736046
KDR	Kinase insert domain receptor; VEGF receptor 2; Flk-1 protein	Angiogenesis	4q11-4q12	10371349
KISS1	Kiss-1 metastasis suppressor		1q32	9192814
KLK10	Kallikrein 10; Normal epithelial cell-specific protease 1 (NES1)	Proteolysis	19q13.3-19q13.4	9809976
KLK3	Kallikrein 3; Prostate-specific antigen (PSA)	Proteolysis	19q13.3-19q13.4	7527295
KPNA2	Karyopherin alpha 2; RAG cohort 1 (RC1); Importin alpha 1	Protein transport	17q23.1-17q23.3	10969801
KRT5	Keratin 5		12q13	1690428
KRT7	Keratin 7		12q12-12q14	1690428
KRT14	Keratin 14		17q12-17q21	1690428
KRT18	Keratin 18		12q13	9701399
KRT19	Keratin 19		17q21	9701399
KRT20	Keratin 20		17	9704713
KRT8	Keratin 8		12q13	9701399
LAMA3	Laminin, alpha 3	Adhesion	18q11.2	9848077
LAMB3	Laminin, beta 3	Adhesion	1q32	9848077
LAMC2	Laminin, gamma 2	Adhesion	9q31-9q34	9848077
LAP18	Leukemia-associated phosphoprotein p18; Oncoprotein 18 (OP18); Stathmin	Cell cycle regulation	1p36.1-1p35	10638981
LASP1	LIM and SH3 protein 1; Malignant 50 (MLN50)		17q11-17q21.3	9848085
LOX	Lysyl oxidase		5q23.3-5q31.2	9484712
LOXL1	Lysyl oxidase-like 1		15q22	9484712
LUM	Lumican		12q21.3-12q22	11054714
MAD2L1	MAD2 (mitotic arrest deficient, yeast, homolog)-like 1	Cell cycle regulation	4q27	11066082
MAP3K8	Mitogen-activated protein kinase-kinase-kinase 8; Cancer Osaka thyroid oncogene (COT); Tumor progression locus 2 (TPL-2)		10p11.2	10490831
MAWD	MAWD			10646843
MCAM	Melanoma cell adhesion molecule (MCAM); MUC18 glycoprotein; CD166 antigen	Adhesion	11q23.3	9284823

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS (contd.)

Gene name	Gene product name (s)	Cell function or tumor property	Locus	PubMed identifier
MDM2	Mouse double minute 2, human homolog of; p53 binding protein	Cell cycle regulation	12q14.3-12q15	10706127
MET	Met-protooncogene product; Hepatocyte growth factor receptor	Angiogenesis	7q31	9221809
MGB1	Mammaglobin 1		11q13	10397237
MGB2	Mammaglobin 2; Mammaglobin B; Lacryglobin (LGB)		11q13	10767368
MKI67	Ki-67 antigen; Mib-1 antigen	Cell cycle regulation	10q25-10qter	10445426
MLN64	Malignant 64; Steroidogenic acute regulatory protein related		17q11-17q12	9139840
MMP1	Matrix metalloproteinase 1; Interstitial collagenase	Proteolysis	11q22.3	10628798
MMP11	Matrix metalloproteinase 11; Stromelysin 3	Proteolysis	22q11.23	8645587
MMP13	Matrix metalloproteinase 13; Collagenase 3	Proteolysis	11q22.3	8207000
MMP14	Matrix metalloproteinase 14 (membrane-inserted); Membrane-type matrix metalloproteinase 1 (MT1-MMP)	Proteolysis	14q11-14q12	9158005
MMP15	Matrix metalloproteinase 15 (membrane-inserted); Membrane-type matrix metalloproteinase 2 (MT2-MMP)	Proteolysis	16q13-16q21	9158005
MMP2	Matrix metalloproteinase 2; Gelatinase A; 72kDa gelatinase	Proteolysis	16q13-16q21	8686751
MMP3	Matrix metalloproteinase 3; Stromelysin 1	Proteolysis	11q22.3	10208466
MMP7	Matrix metalloproteinase 7; Matrilysin	Proteolysis	11q21-11q22	9932604
MMP9	Matrix metalloproteinase 9; Gelatinase B; 92kDa gelatinase	Proteolysis	20q11.2-20q13.1	8645587
MSN	Moesin		Xq11.2-Xq12	9706140
MST1R	Macrophage stimulating 1 receptor; RON protein tyrosine kinase (RON)		3p21.3	9671413
MT1E	Metallothionein 1E		16q13	10917545
MTA1	Metastasis-associated gene	Hormone sensitivity	14q31.2	7607577
MUC1	Mucin-1, transmembrane; CA15-3 antigen; Episialin; Polymorphic epithelial mucin (PEM); Epithelial membrane antigen (EMA);	Adhesion	1q21	9309122
MVP	Major vault protein; Lung resistance protein (LVP)	Chemoresistance	16p13.1-16p11.2	10697509

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS (contd.)

Gene name	Gene product name (S)	Cell function or tumor property	Locus	Pubmed identifier
MYB	V-myb avian myeloblastosis viral oncogene homolog	Cell cycle regulation	6q22-6q23	11034064
MYBL2	V-myb avian myeloblastosis viral oncogene homolog-like 2	Cell cycle regulation	20q13.1	10969801
MYC	V-myc avian myelocytomatosis viral oncogene homolog	Cell cycle regulation	8q24.12-8q24.13	8432027
NCOA1	Nuclear receptor coactivator 1; Steroid receptor coactivator 1 (SRC-1)	Hormone sensitivity	2p23	9541193
NCOA2	Nuclear receptor coactivator 2; Steroid receptor coactivator 2 (SRC-2); Transcriptional intermediary factor 2 (TIF2); Glucocorticoid receptor interacting protein 1 (GRIP1)	Hormone sensitivity	8	10612426
NCOA3	Nuclear receptor coactivator 3; Steroid receptor coactivator 3; Amplified in breast cancer (AIB1); Thyroid hormone receptor activator molecule (TRAM-1); Receptor-associated coactivator 3 (RAC3)	Hormone sensitivity	20q12	9252329
NCOR1	Nuclear receptor corepressor 1; KIAA1047 protein	Hormone sensitivity	17p11.2	10690532
NCOR2	Nuclear receptor corepressor 2; Silencing mediator of retinoid and thyroid hormone action (SMRT)	Hormone sensitivity	12q24	10589759
NF2	Neurofibromatosis, type II; Neurofibromin; Merlin; Schwannomin		22q12.2	8162073
NLVCF	Nuclear localization signal deleted in velocardiofacial syndrome; Upregulated in metastasis (URIM) protein		22q11.21	10226592
NME1	Non-metastatic cells 1, protein expressed in; Nm23-h1, nm23A; Nucleoside diphosphate kinase A (NDKA)	Cell cycle regulation	17q21.3	7892043
NME2	Non-metastatic cells 2, protein expressed in; Nm23-h2, nm23B; Nucleoside diphosphate kinase B (NDKB)	Cell cycle regulation	17q21.3	1988104
NR3C1	Nuclear receptor subfamily 3, group C, member 1; Glucocorticoid receptor (GR)	Hormone sensitivity	5q31	2249895
ODC1	Ornithine decarboxylase 1	Cell cycle regulation	2p25	9472098
P8	p8 protein; Candidate of metastasis 1 protein		16p11.2	11034106
PCNA	Proliferating cell nuclear antigen	Cell cycle regulation	20pter-20p12	10445426
PDZK1	PDZ domain-containing 1		1q21	11103799
PECAM1	Platelet-endothelial cell adhesion molecule 1; CD31	Angiogenesis	17q13	11250735
PGR	Progesterone receptor	Hormone sensitivity	11q22-11q23	7927940

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS (contd.)

Gene name	Gene product name (s)	Cell function or tumor property	Locus	PubMed identifier
PI3	Proteinase inhibitor 3; Elafin; Elastase-specific inhibitor (ESI); Skin-derived antileukoproteinase (SKALP)	Proteolysis	20q12-20q13	7780965
PIP	Prolactin-induced protein; Gross cystic disease fluid protein 15 (GCDFP-15)		7q34	10576657
PLAT	Plasminogen activator, tissue-type (tPA)	Proteolysis	8p12	2112958
PLAU	Plasminogen activator, urokinase (uPA)	Proteolysis	10q24	10930085
PLAUR	Plasminogen activator, urokinase receptor (uPAR); CD87 antigen	Proteolysis	19q13	10930085
PLD1	Phospholipase D1		3q26	11090971
PLU-1	Putative DNA/chromatin binding motif; Retinoblastoma binding protein 2 homolog 1 (RBBP2H1)		1q32.1	10336460
PPARBP	Peroxisome proliferator-activated receptor binding protein; Vitamin D receptor-interacting protein complex component (DRIP205); Thyroid hormone receptor-associated protein complex component (TRAP220); Thyroid hormone receptor interactor protein 2 (TRIP2)	Hormone sensitivity	17q12	10485914
PTGS2	Prostaglandin-endoperoxide synthase 2; Prostaglandin G/H synthase; Cyclooxygenase-2 (COX-2)	Inflammatory response	1q25.2-1q25.3	9521170
PTHLH	Parathyroid hormone-like hormone; Parathyroid hormone-related protein		12p12.1-12p11.2	10517339
PXN	Paxillin		12q24	10383144
RAD51C	RAD51 (S. cerevisiae) homolog C		17	11034067
RB1	Retinoblastoma-1	Cell cycle regulation	13q14.2	7588321
RBP1	Retinol-binding protein, cellular	Hormone sensitivity	3q23	10716965
REA	Repressor of estrogen receptor activity; B-cell associated protein (BAP); D-prohibitin	Hormone sensitivity	12p13	10850416
RPS6KB1	Ribosomal protein S6 kinase, 70kDa, polypeptide 1	Protein synthesis	17	11034067
S100A2	S100 calcium-binding protein A2		1q21	1372446
S100A4	S100 calcium-binding protein A4; Metastasin; Placental calcium-binding protein (CAPL)		1q21	9570150
S100A6	S100 calcium-binding protein A6; Calcyclin (CACY); Prolactin-receptor-associated protein (PRA)		1q21	2448309
S100A7	S100 calcium-binding protein A7; Psoriasin 1		1q21	10595935

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS (contd.)

Gene name	Gene product name (s)	Cell function or tumor property	Locus	PubMed identifier
SAFB	Scaffold attachment factor B; Hsp27 ERE-TATA-binding protein (HET); HnRNP A1 associated protein (HAP)	Hormone sensitivity	19p13	9679963
SERPINB2	Serine (or cysteine) proteinase inhibitor, clade B, member 2; Plasminogen activator inhibitor, type II (PAI-2)	Proteolysis	18q21.3	9496254
SERPINB5	Serine (or cysteine) proteinase inhibitor, clade B, member 5; Protease inhibitor 5 (PI5); Maspin	Proteolysis	18q21.3	9635871
SERPINE1	Serine (or cysteine) proteinase inhibitor, clade E, member 1; Plasminogen activator inhibitor, type I (PAI-1)	Proteolysis	7q21.3-7q22	10930085
SFN	Stratifin; 14-3-3sigma	Cell cycle regulation	1p	10811911
SLC9A3R1	Solute carrier family 9, isoform A3, regulatory factor 1; Sodium/hydrogen exchanger, regulatory factor of (NHERF); ERM-binding phosphoprotein, 50-kDa (EBP50)		17	11141479
SNAI1	Snail 1, Drosophila, homolog of		20q13.1	10655586
SNCG	Synuclein gamma; Breast cancer specific protein 1 (BCSG1); Persyn		10q23.2-10q23.3	11016747
SOD2	Superoxide dismutase-2, mitochondrial; Manganese-containing superoxide dismutase (MnSOD)	Angiogenesis	6q25.3	7905787
SPARC	Secreted protein, acidic, cysteine-rich; Osteonectin		5q31.1-5q32	9389930
SPP1	Secreted phosphoprotein 1; Osteopontin		4q21-4q25	10418827
SPRR1A	Small proline-rich protein 1A		1q21-1q22	10501656
SRA1	Steroid receptor RNA activator (1)	Hormone sensitivity	5q	10485452
STK15	Serine/threonine protein kinase 15; Aurora-related kinase 1 (ARK1); Breast tumor-amplified kinase (BTAK)		20q13	11291073
TBX2	T-box 2		17q23	11034067
TEK	Tyrosine kinase, endothelial; Protein receptor tyrosine kinase, epithelial-specific Tie-2	Vasculogenesis, Angiogenesis	9p21	11250735
TERC	Telomerase, RNA component		3q21-3q28	9516976
TERT	Telomerase reverse transcriptase; Telomerase, catalytic component		5p15.33	9620778
TFF1	Trefoil factor 1; pS2; Breast cancer, estrogen-inducible sequence expressed in (BCEI)	Hormone sensitivity	21q22.3	10573116
TGFA	Transforming growth factor, alpha		2p13	9191980
TGFB1	Transforming growth factor, beta 1		19q13.2	8198961
TGFB2	Transforming growth factor, beta 2		1q41	8198961

TABLE I - A LIST OF MARKERS WHOSE SPECIFIC mRNA LEVEL HAS BEEN SHOWN TO VARY IN BREAST CANCER TISSUES OR CELLS (contd.)

Gene name	Gene product name (s)	Cell function or tumor property	Locus	PubMed identifier
TGFB3	Transforming growth factor, beta 3		14q24	8198961
TGFBR2	Transforming growth factor-beta receptor, type II; Hereditary nonpolyposis colorectal cancer, type 6 (HNPCC6)		3p22	10375614
TGFBR3	Transforming growth factor-beta receptor, type III; Beta-glycan		1p33-1p32	10375614
THBS1	Thrombospondin-1	Angiogenesis	15q15	9012463
THBS2	Thrombospondin-2	Angiogenesis	6q27	9012463
TIMP1	Tissue inhibitor of metalloproteinase-1; Erythroid potentiating activity (EPA)	Proteolysis	Xp11.3-Xp11.23	9815852
TIMP2	Tissue inhibitor of metalloproteinase-2	Proteolysis	17q25	9815852
TJP1	Tight junction protein-1; Zonula occludens 1 protein (ZO-1)	Adhesion	15q13	9846967
TJP2	Tight junction protein-2; Zonula occludens 2 protein (ZO-2)	Adhesion	9	11018256
TMSB10	Thymosin, beta 10	Cell cycle regulation	2	10487837
TNF	Tumor necrosis factor alpha	Angiogenesis	6p21.3	8119765
TOP2A	Topoisomerase (DNA) II alpha (170kDa)	Chemoresistance	17q21-17q22	8664122
TP53	Tumor protein p53	Cell cycle regulation	17p13.1	1537617
TRAF4	TNF receptor-associated factor 4; Malignant 62 (MLN62); Cysteine-rich domain associated with RING and TRAF domains (CART1)		17q11-17q12	8752152
TSP50	Testes-specific protease 50	Proteolysis	3p14-3p12	10397268
TYMS	Thymidylate synthetase; Thymidylate synthase (TS, TMS)	Cell cycle regulation	18p11.32	1704999
VDR	Vitamin D receptor; 1,25-dihydroxyvitamin D3 receptor	Hormone sensitivity	12q12-12q14	2249895
VEGF	Vascular endothelial growth factor; Vascular permeability factor (VPF)	Angiogenesis	6p12	9828838
VEGFB	Vascular endothelial growth factor B	Angiogenesis	11q13	9665470
VEGFC	Vascular endothelial growth factor C	Angiogenesis	4	9665470
VIM	Vimentin		10p13	7538357
WISP2	WNT1 inducible signaling pathway protein 2		20q12-20q13.1	10944450
WISP3	WNT1 inducible signaling pathway protein 3; Lost in inflammatory breast cancer (LIBC)		6q22-6q23	10499627
WT1	Wilms tumor 1	Cell cycle regulation	11p13	9223327
ZNF217	Zinc finger protein 217; Breast cancer putative transcription factor (ZABC1)		20q13.2	9671742

The list has been composed by extensive screening of the medical literature covering the last 15 years. For each marker the following items are mentioned: the main gene name as recommended by the Human Genome Organisation (HUGO, http://www.gene.ucl.ac.uk/hugo/); one or more denomination(s) of the gene product, as found in the breast cancer literature; one tumor property or cell function in which the marker is involved, if this is known; the gene locus; the PubMed identificator (PMID) of one of the most recent references reporting a variation of the marker mRNA in breast tumors or breast cancer cell lines. PubMed URL: http://www.ncbi.nlm.nih.gov/PubMed/

(reviewed in (11)).

DNA microarrays are solid surfaces bearing multiple cDNA or oligonucleotide spots, which play the role of capture probes. These capture probes, which represent genes or parts of genes, are either chemically synthesized *in situ* on the surface (oligonucleotide-based arrays) or laid down through the use of a special device, the "arrayer" (cDNA-based arrays). Glass and plastic slides are progressively replacing nylon membranes as supports.

cDNA-based microarrays are ideal for extensive screening of gene expression and expression profiling of unknown or poorly characterized genes. They may be very useful to examine cell or tissue response to various agents, to identify new therapeutic targets, or to classify biological materials (such as tumors). In these arrays the capture sequences are obtained from cDNA libraries or PCR products, and are thus generally heterogeneous in size. This may make any direct quantitative comparison between two different genes represented on the array impossible. Moreover, in some cases the capture sequences may be unable to discriminate between different genes exhibiting common or closely related regions, or between variants of the same gene. This would require the design and use of gene-specific hybridization probes. Oligonucleotide-based microarrays are best suitable for expression studies of one or a few well characterized genes. Usually the capture sequences supported by such arrays have the same size. This allows quantitative comparisons between two genes or two variants of the same gene represented on the array. Capture sequences may be chosen to fit to highly specific gene regions and multiple sequences corresponding to different regions may be used for each gene. However, this operation may be tedious and time-consuming.

In most studies involving microarrays, labeled target cDNAs obtained by reverse transcription from the population of tumoral mRNAs are incubated with the array, and the amount of material hybridized to the specific capture probes is determined by various techniques (radioactivity, colorimetry, fluorescence and others). Microarrays have the inherent advantage of detecting the expression of genes in parallel, with a direct readout of the hybridization results. Most of the commercially available DNA microarrays carry several thousands of capture probes; due to this vast amount of sequences to be synthesized, purified, quantified, and fixed on the solid support, they are expensive. They also entail rather complicated data analysis. Moreover, they may carry many capture probes devoid of real interest from a perspective of routine breast cancer evaluation, because these are specific to genes that are unexpressed or invariable, or whose expression level has never been explored in this kind of cancer. Thus, although these "high-density" DNA microarrays may provide the basic researcher with a means to identify possible novel mRNAs transcripted in tumors, they do not provide a data/price ratio high enough to satisfy clinicians requiring a tool applicable to routine analysis in their everyday clinical activities.

A LOW-DENSITY BREAST CANCER-RELATED MICROARRAY

The concept of the low-density microarray is an attempt to reconcile the need for large amounts of breast cancer-relevant data with the need for low cost. A low-density microarray should support no more than a few hundreds of oligonucleotide capture sequences. However, these should be specific to markers that are recognized as reliable clinical indicators or serious candidates for this status on the basis of previous work. Since potential or established clinical markers continue to be the subject of fundamental studies, the low-density microarray could also be useful for basic researchers.

Due to their reduced number of capture sequences, low-density microarrays appear to be more suited for "customization" than high-density ones. For instance, it is conceivable to replace one capture probe by another on such arrays if, in the course of studies, the original sequence turns out to be not as interesting as initially suggested. Moreover, the possibility to select the probes one by one should allow the design of arrays specially dedicated to the study of one specific tumor characteristic such as angiogenesis, proliferation, or proteolysis.

The estrogen receptors (ER) alpha and beta and their functionally related molecules provide a good example of how a low-density DNA chip could be exploited by both clinicians and basic researchers working on breast cancer. In addition to the reliable clinical indicator ER-alpha and the more recently identified ER-beta (gene *ESR2*), it is recommended to analyze tumors for the expression of several genes whose transcription is known to be induced by ER. For instance, levels of mRNA encoded by *TFF1*, *CTSD*, *CCND1*, *PGR*, and *MYC* are relevant to the transcriptional functionality of the ER. The possibility of using the expression of these genes as predictors of patient response to endocrine therapy is thus open and has to be validated.

Other mRNAs have been suggested to be under estrogenic regulation in breast cancer cells. Among them are keratin 19 (gene *KRT19*), parathyroid hormone-related peptide (*PTHLH*), interleukin-6 (*IL6*), vascular endothelial growth factor (*VEGF*) and bcl-2 (*BCL2*) (12-16). Before they can be confirmed as reliable clinical indicators, their expression in tumors should be thoroughly examined.

Recent findings indicate that the activity of ER in response to estrogens/antiestrogens depends on the expression of several coactivator and corepressor molecules (17, 18). These molecules are currently under extensive investigation. The understanding of their mechanisms of

action may benefit from systematic studies of their expression in tumors and the resulting data could probably modulate the endocrine therapy given to patients.

The number of genes whose expression is more or less related to the presence and function of ER in tumors is indeed so high that it could almost justify the production of an ER chip, which would constitute a specialized form of a breast cancer-specific chip.

CHOICE OF MARKERS

mRNAs that could be candidates (see Table I) for quantification by the low-density microarray are expected to fulfill four criteria:

- Their levels should vary in breast cancer tissues or cells compared to their normal counterparts, or from one tumor to another, or at least from one cell line to another (e.g. MCF-7 versus MDA-MB-231);
- These variations should be recognized as being, or at least suspected to be, of diagnostic, prognostic, or predictive significance;
- They should be expressed in tumors at levels detectable by the DNA chip. This requirement might become a limiting factor due to the decreasing size of samples obtained at surgery. The total RNA amount routinely used to label radioactive or fluorescent target cDNA for microarray hybridization may be as low as 1 or 50, respectively. With such amounts, however, some specific mRNAs whose basal level is low (as is notably the case for cytokines) could be undetectable. Amplification of the mRNA sequences could solve this problem but introduces an additional level of manipulation and very often negatively affects the quantification since not all mRNA sequences are amplified with the same efficiency.
- The variations in their levels should parallel the variations in the corresponding protein and gene levels. Basically, the amount of mRNA reflects an equilibrium between transcription and degradation, which is influenced by many factors. This amount may also be determined by gene deletions and amplifications (frequently observed for genes like *ERBB2*, *MYC*, *CCND1*, *EMS1*, *FGFR1*, and *MDM2* (19)). Proteic structures and activities are the ultimate support of cell functions and tumor properties. A good correlation between mRNA and protein levels may be crucial when it concerns therapeutic target proteins, such as proteinases, vascular endothelial growth factor (*VEGF*), c-erbB-2 (*ERBB2*), or resistance markers such as breast cancer resistance protein (*ABCG2*), multidrug resistance protein-1 (*ABCB1*) or the lung resistance protein (*MVP*). Regarding ER-alpha (ESR1) we have recently shown that a simple linear relationship exists between its 6.7-kb mRNA measured by Northern blot and the receptor level evaluated by ligand-binding assay (LBA) (20). However, similar studies remain to be done for a number of candidate markers. This means that, besides results obtained at the mRNA level, it will be mandatory to measure the corresponding protein in tumor samples. The use of DNA chips should lead to a first selection of markers and to the elimination of candidates for which a protein study will not be necessary.

CHOICE OF CAPTURE PROBES

Size, cost of production, time needed for efficient hybridization with the target cDNA, and specificity are determinant factors in the choice of a capture probe. In some cases the constraints defined by these factors may be difficult to be reconciled.

In theory, longer probes lead to more efficient hybridization but they also require longer hybridization times and are more expensive. Furthermore, it may happen that a capture probe must absolutely be short because its target cDNA itself is short. For instance, *FGF2*, *PIP* (encoding prolactin-induced protein), *MGB1* (mammaglobin 1), *TFF1* (pS2), and *SPRR1A* (small proline-rich protein 1A) are transcribed into mRNAs that do not exceed 500 bases.

Retrotranscription of mRNAs begins at their 3'-poly(A+) region and is not always complete, which gives rise to a population of more or less complete cDNAs. To improve the efficiency of hybridization it may thus be preferable to design capture sequences specific to mRNA regions close to their 3' end. On the other hand, care must be taken to avoid regions of high (>50%) homology between genes, able to lead to cross-hybridization between capture probes and target cDNAs. Homologies between mRNA species are often found in regions encoding highly conserved protein domains, such as transmembrane or catalytic portions. This is exemplified by the estrogen receptors alpha and beta, whose DNA-binding domains show up to 84% homology at the nucleotide level, while these receptors are largely unrelated by their N-terminus (21).

VARIANTS

It may happen that multiple mRNAs are transcribed from the same gene. These variants often exhibit more or less overlapping sequences. When selecting one or more capture probes specific to such a marker, one should wonder whether all variants have the same potential importance as diagnostic, prognostic, or predictive indicators, or whether it is recommended to detect only some specific ones.

For instance, two integrin alpha 6 (*ITGA6*) mRNA variants have been identified in breast tumors. They encode proteins differing by their C-terminal cytoplasmic

domain (22). Increased integrin alpha 6 expression could be associated with the metastatic phenotype of breast cancer cells (23), but this has not been specifically ascribed to any of the variants. Similarly, the potential prognostic value of bcl-2 (*BCL2*) in breast cancer (5) has not been clearly associated with either of its two transcripts, which differ by their C-terminus. For *ITGA6* as well as *BCL2* it thus appears appropriate to design a capture probe recognizing both variant mRNAs.

At least four variants (A to D) have been found for the integrin beta 1 (*ITGB1*), which mediates interactions between cells and the extracellular matrix. The relative amount of these forms is likely to vary in breast tumors. However, it is clearly recommendable to detect the C variant, which has been shown to inhibit cell proliferation *in vitro* and is downregulated in carcinomas (24). Two capture probes should therefore be designed, one specific to the C form, the other detecting all *ITGB1* variants.

A number of more or less truncated ER-alpha mRNA variants have been described in breast tumors, mainly on the basis of RT-PCR studies. There is no evidence that these rare (25) small-sized forms specify any functional protein. On the other hand, very little is known about their potential as diagnostic, prognostic or predictive factors (26). While the design of a capture probe recognizing the major full-size ER-alpha is an absolute requirement for a breast cancer-specific chip, it might thus be of interest to define additional probes specific to the different variants. Such probes could ideally be incorporated into an even more specialized estrogen receptor-related array.

CD44 glycoproteins are involved in cell-cell and cell-matrix interactions. The *CD44* gene contains 20 exons. Exons 1-5 and 16-20 are spliced together to form a transcript that encodes the ubiquitously expressed standard isoform (known as CD44s). The ten variable exons 6-15 (also named v1-v10) can be alternatively spliced and included within the standard exons at an insertion site between exons 5 and 16, giving rise to a large number of so-called CD44v variants (27). Preliminary studies suggest that not all variants have the same interest as indicators in breast cancer. According to these studies it is recommended to at least define a capture probe recognizing the CD44v6 variant, which might be a marker to identify node-negative patients with a relatively favorable prognosis (28). It has also been suggested that the CD44v7-v8 variant could direct breast tumor cells to lymph nodes and lymphatic vessels (29) and it should be measured by a specific probe.

The case of *MUC1* is one of the most complicated ones. This gene encodes at least three proteins, Muc1/y, Muc1/sec, and Muc1/rep. The latter two contain a large, highly glycosylated core region made up of 30-100 tandem repeats of a 20-amino-acid sequence (30). All Muc1 proteins appear to play a role in reducing cell-cell and integrin-mediated cell-matrix interactions, and probably in the metastatic spread of cancer cells from the primary tumor site. More generally, *MUC1* expression might enhance tumor initiation and progression. Muc1 proteins could also serve as targets for immunotherapy using antibodies directed against the glycosylated variable core region. Molecular studies have shown that the non-secreted Muc1/rep and the secreted Muc1/sec are able to bind to Muc1/y serving as a signaling receptor. This suggests that the impact of *MUC1* on tumor properties could depend on the relative levels of its three protein products. While the design of a unique capture probe recognizing all *MUC1* mRNAs is practically possible, it would be of considerable interest to specifically detect the messengers encoding the three proteins Muc1/y, Muc1/sec and Muc1/rep.

CONCLUSION

Besides the traditional clinicopathological analysis, reliable molecular markers will be increasingly used, alone or in combination (index), for diagnosis, prognosis, and response prediction in breast oncology. We need to increase the number of available markers and thus to validate new, potentially interesting ones. To achieve this aim the availability of simultaneous, fast, and standardized tumor sample analysis techniques that are routinely applicable appears greatly helpful. While proteome analysis ("proteomics") is a major challenge for the future, a low-density breast cancer-specific DNA microarray appears currently to best combine the above-mentioned imperatives with the necessity of an inexpensive and easy-to-use tool. On the other hand, the design of pertinent capture probes, discriminating between closely related markers and/or variant forms, may be a fastidious step in the development of such a tool. Nevertheless, specialized low-density microarrays are now under development and are expected to become, with the aid of computer-based data interpretation, a mandatory complement to future proteomics.

ACKNOWLEDGEMENT

This work and the "TUMOCHIP" project are supported by grants from the "Recherche d'initiatives" program (Région Wallonne - Walloon Region), from the "Fonds Médic" and from the "Fonds Jean-Claude Heuson".

ADDITIONAL DATA ON BREAST CANCER MARKERS AVAILABLE ON THE WEB

Useful information on breast cancer markers (location, expression level, clinical value, etc.) may be obtained at the following URLs:

http://bioinfo.amc.uva.nl/HTM-bin/index.cgi
The Human Transcriptome Map (HTM). It relates about 2.5 million tags contained in public SAGE (Serial Analysis of Gene Expression) libraries (National Center for Biotechnology Information SAGEmap database, reference 31) to the Unigene clusters mapped in GeneMap1999 (32).

http://www.cancerindex.org/geneweb/X0401.htm
Cancer Genetics Web – Various data on gene expression and abnormalities in breast cancer.

http://www.med.uni-muenchen.de/egtm/detail/2.htm
Tumor markers in breast cancer, data on clinical significance and recommendations by the European Group on Tumor Markers (EGTM).

http://www.geocities.com/HotSprings/Spa/3430/marqueurs/lismark.htm
Various data on breast cancer markers and their expression in tumor and cell lines.

http://www.mdanderson.org/leukemia/methylation/cgi.html
A list of genes whose expression is affected by promoter CpG island methylation in various cancers (including breast cancer).

http://cc.ucsf.edu/people/waldman/GENES/completegenes.html
A list of cancer genes and loci.

Address for correspondence:
Prof. Marc Lacroix
Laboratoire Jean-Claude Heuson de Cancérologie Mammaire
Institut Jules Bordet, Université Libre de Bruxelles
127, Boulevard de Waterloo
B-1000 Bruxelles, Belgium
e-mail: labo.cancerologie.mammaire@bordet.be,
lcanmamm@ulb.ac.be

REFERENCES

1. Valavaara R. Reliability of estrogen receptors in predicting response to antiestrogens. Oncology (Huntingt) 1997; 11: 14-8.
2. Menard S, Tagliabue E, Campiglio M, Pupa SM. Role of HER2 gene overexpression in breast carcinoma. J Cell Physiol 2000; 182: 150-62.
3. Look MP, Foekens JA. Clinical relevance of the urokinase plasminogen activator system in breast cancer. APMIS 1999; 107: 150-9.
4. Lapidus RG, Nass SJ, Davidson NE. The loss of estrogen and progesterone receptor gene expression in human breast cancer. J Mammary Gland Biol Neoplasia 1998; 3: 85-94.
5. Fitzgibbons PL, Page DL, Weaver D, et al. Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med 2000; 124: 966-78.
6. Davies H, Lomas L, Austen B: Profiling of amyloid beta peptide variants using SELDI Protein Chip arrays. Biotechniques 1999; 27: 1258-61.
7. Schraml P, Kononen J, Bubendorf L, et al. Tissue microarrays for gene amplification surveys in many different tumor types. Clin Cancer Res 1999; 5: 1966-75.
8. Camp RL, Charrette LA, Rimm DL. Validation of tissue microarray technology in breast carcinoma. Lab Invest 2000; 80: 1943-9.
9. Pandey A, Mann M. Proteomics to study genes and genomes. Nature 2000; 405: 837-46.
10. Harry JL, Wilkins MR, Herbert BR, Packer NH, Gooley AA, Williams KL. Proteomics: capacity versus utility. Electrophoresis 2000; 21: 1071-81.
11. De Benedetti VM, Biglia N, Sismondi P, De Bortoli M. DNA chips: the future of biomarkers. Int J Biol Markers 2000; 15: 1-9.
12. Choi I, Gudas LJ, Katzenellenbogen BS. Regulation of keratin 19 gene expression by estrogen in human breast cancer cells and identification of the estrogen responsive gene region. Mol Cell Endocrinol 2000; 164: 225-37.
13. Funk JL, Wei H. Regulation of parathyroid hormone-related protein expression in MCF-7 breast carcinoma cells by estrogen and antiestrogens. Biochem Biophys Res Commun 1998; 251: 849-54.
14. Galien R, Garcia T. Estrogen receptor impairs interleukin-6 expression by preventing protein binding on the NF-kappaB site. Nucleic Acids Res 1997; 25: 2424-9.
15. Hyder SM, Nawaz Z, Chiapetta C, Stancel GM. Identification of functional estrogen response elements in the gene coding for the potent angiogenic factor vascular endothelial growth factor. Cancer Res 2000; 60: 3183-90.
16. Perillo B, Sasso A, Abbondanza C, Palumbo G. 17beta-estradiol inhibits apoptosis in MCF-7 cells, inducing bcl-2 expression via two estrogen-responsive elements present in the coding sequence. Mol Cell Biol 2000; 20: 2890-901.
17. Klinge CM. Estrogen receptor interaction with co-activators and co-repressors. Steroids 2000; 65: 227-51.
18. Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. Cell 2000; 103: 843-52.
19. Cuny M, Kramar A, Courjal F, et al. Relating genotype and phenotype in breast cancer: an analysis of the prognostic significance of amplification at eight different genes or loci and of p53 mutations. Cancer Res 2000; 60: 1077-83.
20. Lacroix M, Querton G, Hennebert P, Larsimont D, Leclercq

G. Estrogen receptor analysis in primary breast tumors by ligand-binding assay, immunocytochemical assay, and northern blot: a comparison. Breast Cancer Res Treat 2001; 67: 263-71.
21. Mosselman S, Polman J, Dijkema R. ER beta: identification and characterization of a novel human estrogen receptor. FEBS Lett 1996; 392: 49-53.
22. Patriarca C, Ivanyi D, Fles D, et al. Distribution of extracellular and cytoplasmic domains of the alpha 3 and alpha 6 integrin subunits in solid tumors. Int J Cancer 1995; 63: 182-9.
23. Mukhopadhyay R, Theriault RL, Price JE. Increased levels of alpha6 integrins are associated with the metastatic phenotype of human breast cancer cells. Clin Exp Metastasis 1999; 17: 325-32.
24. Manzotti M, Dell'Orto P, Maisonneuve P, Fornaro M, Languino LR, Viale G. Down-regulation of beta(1C) integrin in breast carcinomas correlates with high proliferative fraction, high histological grade, and larger size. Am J Pathol 2000; 156: 169-74.
25. van Dijk MA, Hart AA, van 't Veer LJ. Differences in estrogen receptor alpha variant messenger RNAs between normal human breast tissue and primary breast carcinomas. Cancer Res 2000; 60: 530-3.
26. Chappell SA, Johnson SM, Shaw JA, Walker RA. Expression of oestrogen receptor alpha variants in non-malignant breast and early invasive breast carcinomas. J Pathol 2000; 192: 159-65.
27. Goodison S, Urquidi V, Tarin D. CD44 cell adhesion molecules. J Clin Pathol: Mol Pathol 1999; 52: 189-96.
28. Foekens JA, Dall P, Klijn JG, et al. Prognostic value of CD44 variant expression in primary breast cancer. Int J Cancer 1999; 84: 209-15.
29. Kinoshita J, Haga S, Shimizu T, Imamura H, Watanabe O, Kajiwara T. The expression of variant exon v7-v8 CD44 antigen in relation to lymphatic metastasis of human breast cancer. Breast Cancer Res Treat 1999; 53: 177-83.
30. Baruch A, Hartmann M, Yoeli M, et al. The breast cancer-associated MUC1 gene generates both a receptor and its cognate binding protein. Cancer Res 1999; 59: 1552-61.
31. Lal A, Lash AE, Altschul SF, et al. A public database for gene expression in human cancers. Cancer Res 1999; 59: 5403-7.
32. Deloukas P, Schuler GD, Gyapay G, et al. A physical map of 30,000 human genes. Science 1998; 282: 744-6.

Received: June 15, 2001
Accepted: September 12, 2001