In Silico and In Vitro Analysis of Small Breast Epithelial Mucin as a Marker for Bone Marrow Micrometastasis in Breast Cancer

Manuel Valladares Ayerbes, Silvia Díaz-Prado, Daniel Ayude, Rosario G. Campelo, Pilar Iglesias, Mar Haz, Vanessa Medina, Isabel Gallegos, Maria Quindós, and Luis Antón Aparicio

Summary Molecular signatures associated with malignant phenotype would be useful for detection of micrometastatic carcinoma cells. The small breast epithelial mucin (SBEM) gene is predicted to code for a low molecular weight glycoprotein. To evaluate its potential role as a marker for bone marrow (BM) micrometastasis in breast cancer (BC) patients, we have studied in silico and in vitro expression profiles of SBEM gene. Digital SBEM expression in libraries obtained from normal and neoplastic tissues and cell -lines (CL) were displayed and counted on the SAGE Anatomic Viewer. Profiles for cytokeratin-19 and mammaglobin (hMAM), commonly targets used for detection of disseminated BC cells were obtained and compared with SBEM data. Human breast and haematopoietic cancer CL and normal BM were examined by RT-PCR for SBEM and hMAM. Bioinformatics tools were used to gain further insights about the biological role of SBEM in normal breast and BC. Genes with expression patterns in breast libraries correlating with SBEM were identified using two-dimensional display. SBEM tag was detected in 40 libraries (21 BC; 8 non-cancerous breast tissues). Intermediate to high expression was found on 15/21 BC libraries and 7/8 non-tumor breast tissue. SBEM tag count was correlated with ERBB2 (0.662), hMAM (0.409), and RRM2 (−0.379). A model system based on RT-PCR for SBEM mRNA was highly sensitive and specific in order to detect isolated tumor cells. Our results demonstrate that SBEM mRNA may be an imp ortant marker for targeting BC micrometastasis.

Introduction

Although relative survival from breast cancer (BC) in women improved steadily in all European countries in the latest years, age-standardized 5-year relative survival remains in 60–82.6% (1). Metastatic hematogenous spreading is one of the most important factors affecting the prognosis of carcinoma patients, including BC. Circulating tumor cells and occult metastasis (micrometastasis) are considered early events in the progression of BC. Detection of carcinoma cells in the blood or minimal deposits in distant organs as bone marrow could be important to identify patients at high risk of relapse or disease progression (2). PCR amplification of

J.J. Li et al. (eds.), *Hormonal Carcinogenesis V.*

© Springer 2008

tissue or tumor selective mRNA is the most powerful tool for detection of this circulating or micrometastatic cells. Cytokeratins and mammaglobin are among the most frequents mRNA markers used in different reverse-transcriptase polymerase-chain reaction (RT-PCR) assays in BC patients. However down-regulation of mRNA marker in tumor cells (3) or low-level transcription of selected target in the hematopoietic compartment (4) could compromise both sensitivity and specificity of molecular methods. Selection of novel breast-specific transcripts and development of multimarker RT-PCR assays are clearly outstanding research questions. In this context we have evaluated the potential role for small breast epithelial mucin (SBEM) as a marker for bone marrow (BM) micrometastasis in BC. The SBEM gene [Genbank (#AF414087)] was identified by Miksicek, et al. (5) using the cDNA xProfiler tool. SBEM is similar to proteins B511s (6) and BS106 (7). SBEM gene is predicted to code for a low molecular weight glycoprotein with a specific patterns of expression, limited to breast and salivary glands.

Materials and Methods

In Silico Expression Profiles: Serial Analysis of Gene Expression (SAGE). We used an in silico analysis approach to examine SBEM gene expression in normal and cancerous tissues and cell lines. All available published SAGE data were used for analysis of SBEM gene expression. We obtained a mapping of UniGene cluster (Hs.348419) to NIaIII tags from the SAGE tag to Gene Mapping (SAGEmap) search tool (8) available at the NCBI Web site http://www.ncbi.nlm.nih.gov/SAGE/. Expression levels are displayed as blots with different densities and corrected as tag/million (tpm) to facilitate evaluation. (9) Digital SBEM gene expression profiles were analyzed using SAGE Genie tools (http://cgap.nci.nih.gov/). SAGE Genie automatically identifies SAGE tags and provides a link between gene names and SAGE transcript levels (counts). SBEM transcript expression in different libraries obtained from normal and tumor tissues and cell lines were displayed and counted on the SAGE Anatomic Viewer. Libraries were constructed by using *Nla*III as the anchoring enzyme and *Bsm*FI as the tagging enzyme. In addition, in silico expression profiles for cytokeratin-19 and mammaglobin 1 (hMAM, secretoglobin, family 2A, member 2) two commonly targets mRNA used for detection of disseminated BC cells were obtained. These results were compared with SBEM transcript expression. In order to gain further insights about the biological role of SBEM in normal breast tissue and BC, bioinformatics tools were used.

Virtual Northern. Monochromatic SAGE/cDNA virtual northern for SBEM, mammaglobin and CK-19 were accessed from the gene info tool of the CGAP site. It provides an output indicating the relative abundance of each expressed sequence tag (EST) and SAGE sequence. Spot images represent expression level of the gene. For each combination of tissue and histology (normal vs. cancer), expression is computed by dividing the number of ESTs or SAGE tags representing the gene divided by the total number of ESTs or SAGE tags in all libraries with the given

tissue/histology. This ratio is then multiplied by 200,000, giving the number of ESTs or SAGE tags per 200,000. To measure the significance of differences in transcript expression, the method had been described (10).

Cell Lines. BC-derived cell lines (BCCL) MCF-7, MDA-MB468, T47D, BT-549, and PM1 and hematopoietic cell lines (HCL) Jurkat, KG1 and K562, were grown in RPMI-1640 supplemented with 10% (vol/vol) fetal calf serum, L-glutamine, penicillin, streptomycin, and amphotericin at 37°C in 5% CO_2. Cells from adherent cultures were recovered with trypsin-EDTA or nonenzymatic cell dissociating reagent (SIGMA). From each cell line at 50–70% confluence 10^6 cells were obtained for RNA isolation.

RNA Extraction and RT-PCR. Purification of RNA from cell cultures was performed with High Pure RNA Isolation Kit (Roche) as suggested by the manufacturer. Total RNA was treated with DNase I and it was quantified at a wavelength of 260 nm using a spectrophotometer. The reverse transcription was performed using the SuperScript II First-Strand Synthesis System for RT-PCR (Invitrogen) using random hexamers as manufacturer's instructions. PCR amplifications were carried out with AmpliTaq Gold PCR Master Mix (Applied Biosystems). First round PCR amplification of SBEM mRNA was performed using specific primers (SBEM-U-O 5′CTT TGA AGC ATT TTT GTC TGT G3′ and SBEM-L-O 5′AAG GTA AGT AGT TGG ATG AAA T3′) and described by Miksicek (5). In the nested PCR new primers pair (SBEM-S-I 5′TGA TCT TCA GGT CAC CAC CA3′ and SBEM-A-I 5′TGG ATA CGT GTC AGC TGG AG3′) designed using software available on Internet was used (11). First round PCR was performed in 50 μL of reaction mixture containing 2 μL of template cDNA, deionized water, outer SBEM primers U and L, 1 μL 20 μM of each, and PCR Master Mix (2X) 25 μL. In SBEM nested reaction 1 μL of first round PCR template and 0.5 μL at 20 μM of each inner SBEM primers pair was used. For first round SBEM PCR amplification, an initial activation at 95°C for 5 min was used, followed by 35 cycles of 95°C 30 s, 54°C 1 min, and 72°C 1 min. Finally, last extension was at 72°C for 7 min. For SBEM nested reaction, an initial activation at 95°C for 2 min was used, followed by 20 cycles of 95°C 40 s, 62°C 15 s and 72°C 20 s. Last extension was at 72°C for 7 min. First round and nested-PCR amplification of hMAM mRNA was performed using specific primers described previously (12). hMAM first round and nested PCR were performed as described for SBEM with the exception that for nested PCR 1 μL of 1/100 dilution of first round PCR template was used. For first round hMAM PCR amplification, an initial activation at 95°C for 5 min was used, followed by 35 cycles of 95°C 30 s, 57°C 1 min, and 72°C 1 min. Finally, last extension was at 72°C for 7 min. For hMAM nested reaction, an initial activation at 95°C for 5 min was used, followed by 15 cycles of 95°C 30 s, 60°C 1 min, and 72°C 1 min. Last extension was at 72°C for 7 min. PCR products were electrophoresed through agarose gel and stained with 5% ethidium bromide. β-2 microglobulin serves as a positive control target. Negative controls were included in each experiment.

Analysis of mRNA Markers in Normal Human Bone Marrow. Total RNA acids isolated from human normal BMs (BM, $n = 23$) were purchased from BD

Biosciences-Clontech and were examined by RT-PCR for the expression of SBEM and hMAM transcripts. cDNA synthesis was carried out as previously described using different amounts of RNA (up to 2 μg). PCR amplifications were carried out with AmpliTaq Gold PCR Master Mix (Applied Biosystems) as indicated.

Assay Detection Sensitivity and Specificity. Serial RNA dilution analysis was used to assess detection performance of the assay. Total RNA was isolated and purified from the cancer cell line MDA-MB468 and then serially diluted in molecular biology grade water. To additionally assess the detection sensitivity and specificity of the assay, an in vitro model was set up by serially diluting RNA from MDA-MB468 tumor cells in pooled normal human BM derived RNA. One-round and nested RT-PCR were performed for each marker on serially diluted RNA.

DNA Sequencing. PCR products were used as template DNA. Products were purified by enzymatic method (ExoSAP-It, Amersham USB). DNA sequencing was performed in a reference facility on ABI 3700 (Applied Biosystems) using Big Dye Terminators. Sense SBEM and hMAM-specific primers were used.

Results

In Silico Expression Profiles: SAGE. We obtained a mapping of UniGene cluster (Hs.348419) to NIaIII tags from the SAGEmap. Tag CTTCCTGTGA (Ref seq nm-058173.1) was selected. The output file shown expression levels are displayed as blots with different densities and corrected as tpm to facilitate evaluation. Tag CTTCCTGTGA was found in 21 mRNA-source sequences. Of these 20 clustered in UniGene Hs.348419 (LOC118430 small breast epithelial mucin). SBEM sequence tag was detected in 40 libraries, including 21 BC libraries, eight libraries obtained from noncancerous breast tissues, and 11 libraries from other sources. Intermediate or high value (13) for gene expression (cut-off value ≥37 tpm) was found on 15/21 BC libraries and 7/8 nontumor breast tissue. Moderate SBEM expression (≥37 tpm) was only present in four libraries from nonmammary tissues. Interestingly it is included a library developed from white blood cells obtained from a breast carcinoma patient. Digital SBEM gene expression profiles were analyzed using SAGE Genie tools. SBEM transcript expression in different libraries obtained from normal and tumor tissues and cell lines were displayed and counted on the SAGE Anatomic Viewer. SBEM was found in 48 libraries; 31 of them were obtained from mammary tissues. SBEM was expressed in 89 and 67% of libraries derived from normal breast and BC, respectively. Moderate to high expression was found in 41% of the libraries from BC. In addition, in silico expression profiles for cytokeratin-19 and mammaglobin 1 (hMAM, secretoglobin, family 2A, member 2) two commonly targets mRNA used for detection of disseminated BC cells were obtained. These results (Table 1) were compared with SBEM transcript expression.

Table 1 In silico expression profiles results for cytokeratin-19, mammaglobin 1 and SBEM

Gene	SAGE tag	Libraries	Positivity (%)	Low expression (%)	Moderate-to-High expression (%)	Mean Density-TAGS per 200,000
SBEM	CTTCCTGTGA	Normal breast	8/9(89)	2/9(22)	6/9(67)	210,33
		Breast Cancer	18/27(67)	7/27(26)	11/27(41)	211,3
KRT19	GACATCAAGT	Normal breast	3/9(78)	0/9	7/9(78)	183,609
		Breast Cancer	24/27(89)	2/27(7.5)	22/23(81.5)	136,48
SCGB2A2	TTTATTTTAA	Normal breast	7/9(78)	2/9(22)	5/9(56)	10,259
		Breast Cancer	16/27(59)	6/27(22)	10/27(37)	54,15

	Tissue	EST Data		SAGE Data		EST Data			SAGE Data		
		Normal	Cancer	Normal	Cancer	Normal	Cancer	P	Normal	Cancer	P
SBEM	ALL TISSUES					13/2014309	12/1970906	0.45	455/4270902	1707/8895803	0.00
	bone marrow				–	0/14453	0/20958	–	0/204563	–	–
	mammary gland					4/39573	11/79977	0.32	450/538122	1693/1418238	0.00
	white blood cells	–	–		–	–	–	–	0/79858	–	–
hMAM	ALL TISSUES					1/2271660	20/2063711	0.00	111/7080391	424/11023868	0.00
	bone marrow					0/14831	0/21185	–	0/204563	0/16808	–
	mammary gland					0/44822	18/81550	0.00	22/445680	324/1679255	0.00
	white blood cells	–	–		–	–	–	–	1/645820	–	–
CK19	ALL TISSUES					452/2271660	1405/2063711	0.00	1154/7080391	3171/11023868	0.00
	bone marrow					0/14831	0/21185	–	0/204563	0/16808	–
	mammary gland					62/44822	138/81550	0.09	146/445680	1177/1679255	0.00
	white blood cells	–	–		–	–	–	–	11/645820	–	–

Fig. 1 Monochromatic SAGE/cDNA Virtual Northern for SBEM, hMAM, and CK19 in different tissues, including mammary gland and hematopoietic tissues (BM and WBC)

Virtual Northern. Monochromatic SAGE/cDNA Virtual Northern for SBEM, hMAM, and CK19 were accessed from the Gene Info tool of the CGAP site. We analyzed the relative abundance of each EST and SAGE sequence in different tissues, including mammary gland and hematopoietic tissues (BM and WBC). We observed high expression of SBEM, CK19, and hMAM in mammary gland. These reflect the lack of SBEM sequence in BM and WBC compartments (Fig. 1).

Biological Role for Small Breast Epithelial Mucin. In order to gain further insights about the biological role of SBEM in normal breast tissue and BC, bioinformatics tools were used. First, the expression of SBEM in a series of human breast carcinomas SAGE libraries ($n = 27$) was quantified and correlated with the

tags numbers of different molecular markers associated with BC progression. Expression of SBEM was correlated (Spearman's rho) to ERBB2 (0.662: $p=0.000$), hMAM (0.409; $p=0.034$) and RRM2 (−0.379; $p = 0.051$). No significant correlations with SBEM expression were found for estrogen receptor α (ESR1), CK19, STAT1, EGFR, FLT1, HIF1A, FGF 18, GSTM3, TP53, PTTG1, and EpCAM.

In addition, tags from normal breast tissue and breast cancer SAGE data (available from CGAP SAGE Genie) correlating SBEM sequence tag expression were identified and displayed in array format. The color spots are based on normalized values. Correlation coefficients, means, and standard deviations that we display were based on the set of unnormalized values. Results are shown in Fig. 2.

Analysis of mRNA Markers in Cell Lines and Bone Marrow. BC-derived cell lines (MCF-7, MDA-MB468, T47D, BT-549, and PM1) and hematopoietic cell lines

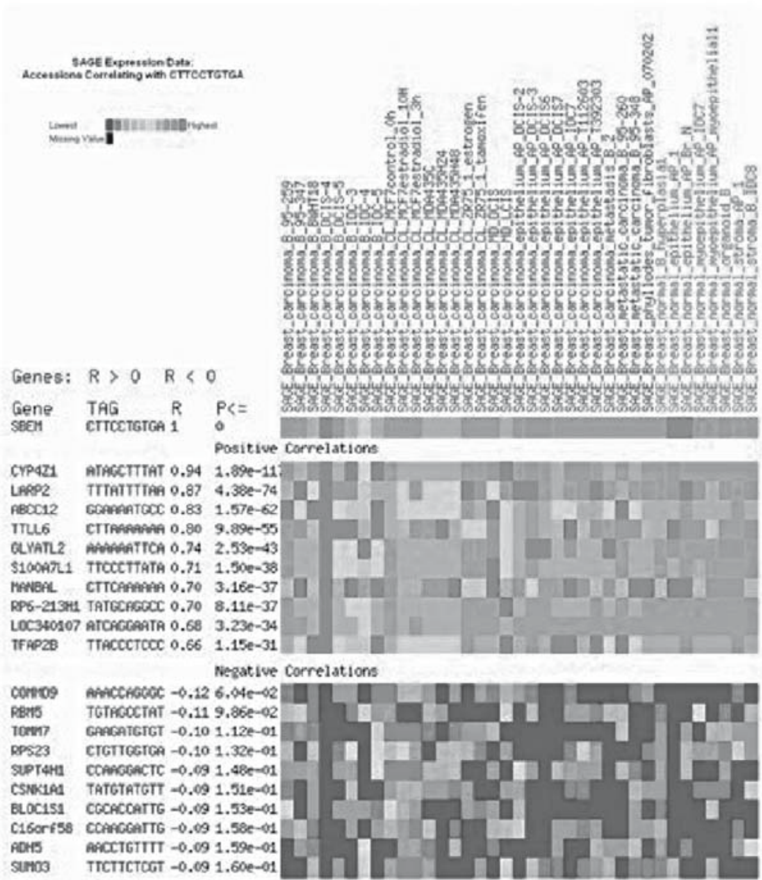

Fig. 2 Tags from normal breast tissue and breast cancer SAGE data correlating SBEM sequence tag expression were identified and displayed in array format

(Jurkat, KG1, and K562) were assessed for hMAM and SBEM mRNA expression by first round and nested RT-PCR. Specific transcript for hMAM was detected in 3/5 BCCL both in first and nested RT-PCR. SBEM mRNA was detected in 2/5 BCCL in first round PCR (Fig. 3) and in 5/5 using nested approach. In both cases, hMAM and SBEM mRNA was not detected in all hematopoietic cell lines assessed. We analyzed SBEM and hMAM mRNA expression in three different pools of normal human bone marrow ($n = 23$). We did not obtain neither SBEM nor hMAM mRNA expression by RT-PCR approach in all pools tested. Low expression was detected on nested RT-PCR in one BM pool ($n = 7$) for both markers.

Sensitivity and Specificity of Nested RT-PCR for SBEM and hMAM. RNA from the cancer cell line MDA-MB468 was serially diluted in molecular biology grade water. One-round and nested RT-PCR were performed for each marker on serially diluted RNA. Using SBEM nested approach we could detect the presence of two cancer cells (Fig. 4) whereas in hMAM nested approach we detected up to one cancer cell.

To additionally asses the detection sensitivity and specificity of the assay, we performed an in vitro model by serially diluting RNA from MDA-MB468 tumor

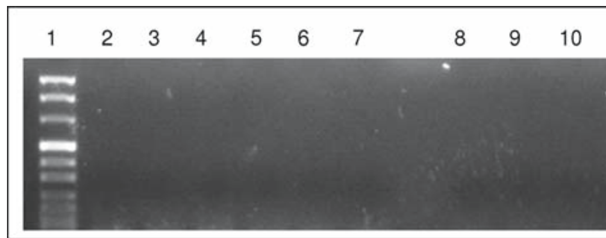

Fig. 3 PCR results for SBEM mRNA expression in different BC-derived cell lines. 1 MW marker, 2–6 (–) controls of BT-549, T47D, PM1, MDA-MB468, and MCF cancer cell lines, respectively, without SuperScript, 7(–) control without RNA, 8–10, BT-549, T47D, and PMI cancer cell lines, respectively

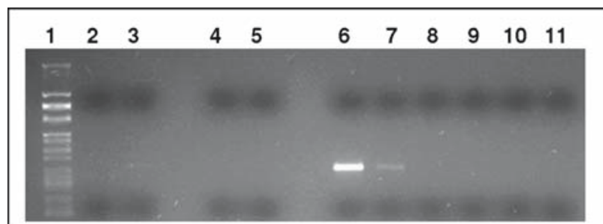

Fig. 4 RT-PCR for SBEM mRNA detection. 1 MW marker, 2 (–) control of normal human BM without SuperScript, 3 normal human BM, 4 MDA-MB468 (–) control without SuperScript, 5 (–) control without RNA, 6 MDA-MB468 not diluted, 7, 8, 9, 10, and 11 MDA-MB468 dilutions at 1:100, 1:1000, 1:10.000, 1:100.000, and 1:1.000.000, respectively

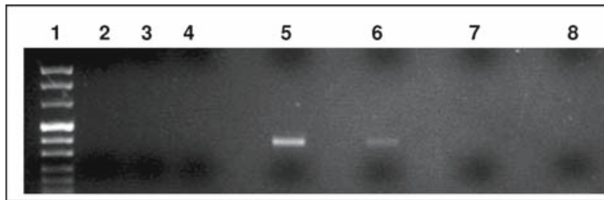

Fig. 5 In vitro model by serially diluting RNA from MDA-MB468 tumor cells in pooled normal human BM derived RNA 1 MW marker, 2, and 3 (–) controls of normal human BM, and MDA-MB468 cancer cell line, respectively, without SuperScript, 4 (–) control without RNA, 5, 6, 7, and 8 pool normal human BM + 1:10, 1:100, 1:1.000, 1:10.000 MDA-MB468 dilutions, respectively. 9 normal human BM

cells in pooled normal human BM derived RNA. One-round and nested RT-PCR were performed for each marker on serially diluted RNA. For SBEM nested approach we could detect up to one cancer cell among 1 μg/μL of normal BM RNA (Fig. 5). In case of hMAM we could also detect up to one cancer cell among 1 μg/μL of normal BM RNA.

Conclusions

Bioinformatics approach based on SAGE and EST data confirms the selective and high expression of SBEM both in normal and BC tissues. Moreover, SBEM was over expressed in BC comparing normal mammary gland based on SAGE counts ($p < 0.005$). In addition moderate to high expression was found in 41% of BC libraries comparing with 37% for hMAM. Lack of SBEM expression in hematopoietic-derived libraries was confirmed by our in silico analysis. Nested RT-PCR for SBEM mRNA was highly sensitive and specific in order to detect isolated tumor cells in a model system. Our results probe that SBEM mRNA could serve as a marker for targeting BC micrometastasis.

Acknowledgments This work was supported by a Grant of the Universidade da Coruña 2000-5435256011.

References

1. Coleman MP, Gatta G, Verdecchia A, et al. (2003) EUROCARE-3 summary: cancer survival in Europe at the end of the 20th century. Ann Oncol 14(Suppl. 5):v128–v149.
2. Braun S, Naume B (2005) Circulating and disseminated tumor cells. review article. J Clin Oncol 23:1623–1626.
3. Woelfle U, Cloos J, Sauter G, et al. (2003) Molecular signature associated with bone marrow micrometastasis in human breast cancer. Cancer Res 63:5679–5684.

4. Kruger WH, Jung R, Detlefsen B, et al. (2001) Interference of cytokeratin-20 and mammaglobin-reverse-transcriptase polymerase chain assays designed for the detection of disseminated cancer cells. Med Oncol 18:33–38.
5. Miksicek RJ, Myal Y, Watson PH, et al. (2002) Identification of a novel breast- and salivary gland-specific, mucin-like gene strongly expressed in normal and tumor human mammary epithelium. Cancer Res 62:2736–2740.
6. Houghton RL, Dillon DC, Molesh DA, et al. (2001) Transcriptional complementarity in breast cancer: application to detection of circulating tumor cells. Mol Diagn 6:79–91.
7. Roberts L, Russell J, Friedman P, et al. (2002) Identification and immunohistochemical characterization of a mucin-like glycoprotein expressed in early stage breast carcinoma. Tumour Biol. 23:263–278.
8. Lash AE, Tolstoshev CM, Wagner L, et al. (2000). SAGEmap: a public gene expression resource. Genome Res 10:1051–1060.
9. Boon K, Elison C, Greenhut SF, et al. (2002) An anatomy of normal and malignant gene expression. Proc Natl Acad Sci 99:11287–11292.
10. Lal A, Lash AE, Altschul SF, et al. (1999) A public database for gene expression in human cancers. Cancer Res 59: 5403–5407.
11. Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, NJ, pp. 365–386.
12. Grunewald K, Haun M, Urbanek M, et al. (2000) Mammaglobin gene expression: a superior marker of breast cancer cells in peripheral blood in comparison to epidermal-growth-factor receptor and cytokeratin-19. Lab Invest 80:1071–1077.
13. Lercher MJ, Urrutia AO, Hurst LD (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet 31:180–183.