Membrane-Associated and Secreted Genes in Breast Cancer

Nathan O. Stitziel,1 Brenton G. Mar,2 Jie Liang,1 and Carol A. Westbrook3

Departments of 1Bioengineering and 2Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, Illinois; and 3Department of Medicine, Boston University, Boston, Massachusetts

ABSTRACT

The identification of membrane-associated and secreted genes that are differentially expressed is a useful step in defining new targets for the diagnosis and treatment of cancer. Extracting information on the subcellular localization of genes represented on DNA microarrays is difficult and is limited by the incomplete sequence and annotation that is available in existing databases. Here we combine a biochemical and bioinformatic approach to identify membrane-associated and secreted genes expressed in the MCF-7 breast cancer cell line. Our approach is based on the analysis of differential hybridization levels of RNAs that have been physically separated by virtue of their association with polysomes on the endoplasmic reticulum. This approach is specifically applicable to oligonucleotide microarrays such as Affymetrix, which use single-color hybridization instead of dual-color competitive hybridizations. Assignment to membrane-associated and secreted class membership is based on both the differential hybridization levels and an expression threshold, which are calculated empirically from data collected on a reference set of known cytoplasmic and membrane proteins. This method enabled the identification of 755 membrane-associated and secreted probe sets expressed in MCF-7 cells for which this annotation did not previously exist. The data were used to filter a previously reported expression dataset to identify membrane-associated and secreted genes which are associated with poor prognosis in breast cancer and represent potential targets for diagnosis and treatment. The approach reported here should provide a useful tool for the analysis of gene expression patterns, identifying membrane-associated or secreted genes with biological relevance that have the potential for clinical applications in diagnosis or treatment.

INTRODUCTION

With the advent of high-throughput global genomic strategies, the potential exists for the identification of many novel genes that have a specific association with cancer, and the gene product of which has diagnostic or therapeutic implications. The task remains to determine which of these have the most immediate potential for clinical translation. Among the most useful proteins in the clinical setting are those that are associated with the cancer cell membrane, including those that are membrane-bound and those that are secreted extracellularly (referred to as membrane-associated and secreted or membrane-associated and secreted genes). Membrane-bound proteins include surface antigen targets for diagnosis or treatment, receptors for external factors that regulate cell growth, and proteins that regulate cell adhesion and metastases. Secreted proteins and peptides can be used as circulating tumor markers for diagnosis and monitoring.

The characterization of a novel gene as one that encodes a membrane-associated or secreted protein can be difficult. Although computational methods exist for predicting whether a protein is membrane-bound or secreted (1, 2), these methods cannot be applied to incomplete or poorly annotated gene sequence and are inexact even in the best setting.

An alternative method to identify membrane-associated and secreted genes experimentally based on differential hybridization to glass slide cDNA arrays was recently shown (3). This method takes advantage of the fact that proteins function at the membrane surface or are immediately secreted are preferentially translated from ribosomes at the endoplasmic reticulum to which they are directed by their signal peptide. Because their association with the endoplasmic reticulum membrane makes them less dense, these membrane-bound polysomes can be separated from their heavier cytosolic counterparts by sucrose gradient centrifugation (4). RNA prepared from these two cellular subfractions is used for differential cDNA hybridization to identify those that are most highly associated with the membrane-bound polysomes.

Here we report the application of this method to the global analysis of the genes expressed in a breast cancer cell line, MCF-7, modifying it for the widely used Affymetrix chips. We develop a statistical approach to determine the membrane association of each Affymetrix probe set, as expressed in the cell line, by comparing the ratio on two chips to a reference set of known cytoplasmic and membrane proteins. The results of this study were then used to analyze the data from a previously reported differential expression study in breast cancer (5), to identify membrane-associated and secreted genes that are associated with poor prognosis, demonstrating the utility of this approach to identify potential targets for diagnosis and treatment from differential hybridization studies.

MATERIALS AND METHODS

Cell Line Preparation

MCF-7 cells were purchased from the American Type Culture Collection (Manassas, VA) and were cultured in Eagle’s MEM supplemented with 0.01 mg/mL bovine insulin, 10% fetal bovine serum in 5% CO2 at 37°C.

Polysome Fractionation

Polysomes were fractionated by sucrose density gradient centrifugation with a modification of the method described by Mechler (4). After treatment with cyclohexamide (10 μg/mL) for 10 minutes at 37°C, 3 × 106 MCF-7 cells in log growth were collected by scraping the dishes into cold PBS. The cells were then resuspended at a concentration of 2.5 × 10⁷ cells/mL in a hypotonic lysis buffer [10 mMol/L HCl, 1.5 mMol/L MgCl2, and 10 mMol/L Tris-Cl (pH 7.4)] and were allowed to rest on ice for 10 minutes. After lysing cells with a Dounce homogenizer, nuclear and cell debris were removed by centrifugation at 2,000 × g (4°C) for 2.5 minutes. The supernatant was loaded on a discontinuous-step sucrose gradient (2.5 mol/L, 2.1 mol/L, 1.95 mol/L, and 1.3 mol/L sucrose) and centrifuged at 26,000 × g for 5 hours. After centrifugation, Successive 1.5 mL fractions were collected from the bottom of the centrifugation tube, and the A260 nm was measured to estimate the RNA content.

RNA Preparation

Total RNA was isolated from the pooled sucrose gradient fractions by mixing with TRizol LS reagent (Invitrogen, Carlsbad, CA) at a 3:1 ratio
Real-time Quantitative Reverse-transcriptase PCR

First-Strand cDNA Synthesis. For generation of first-strand cDNA, 1 μg of RNA was reverse-transcribed with Superscript II Reverse Transcriptase Kit (Invitrogen) in the presence of oligodeoxynucleotidic acid (12–18) in a final 20-μL reaction volume with reverse transcriptase per manufacturer’s recommended protocol followed by RNase H treatment.

Real-time Reverse Transcriptase PCR Setup. Real-time PCR reactions were done with DNase-free cDNA templates generated above and SYBR Green PCR Core Reagents (Applied Biosystems, Branchburg, NJ) following manufacturer’s protocol with the following modifications: a 25 μL reaction was used, which contained 1× SYBR Green PCR mix, 3 mM/L MgCl2, 1× deoxyoligonucleotidic triphosphate blend (0.2 mM/L of dATP, dCTP, dGTP, and dTTP), 0.625 units of AmpliTaq Gold, 0.125 units of AmpliF-erase UNG, 50 mM/L each of forward and reverse primer, and 1 μL of cDNA template. Default PCR amplification cycles were used as specified by the ABI Prism 7700 Sequence Detection System (Applied Biosystems): 50°C for 2 minutes, 95°C for 10 minutes, 40 cycles of 15 seconds at 95°C, and 60°C for 1 minute. PCR amplification was followed by melting curve analysis with the following 3 hold cycles: 95°C for 15 seconds, 60°C for 20 seconds, and 95°C for 15 seconds, with the ramping time at maximum value 19:59 minutes set at the last hold cycle. PCR amplification analysis was done on Sequence Detector v.1.7a, and melting curves were analyzed on Dissociation Curves v.1 according to Applied Biosystems guidelines.

MCF-7 cDNA template was used to generate a relative standard curve for either the endogenous control or the target gene. MCF-7 cDNA was serially diluted at 1:10 dilution factors starting with the highest arbitrary concentration of 50 ng/μL (taken from one twentieth of a reverse transcriptase reaction of 1 μg of starting total RNA). The sample dilutions were diluted at 1:5. All samples were done in triplicate. The sequence of the primers used in real time reverse transcriptase-PCR is as follows: endogenous control 18S rRNA: F:5'-GTAACCCGCTGAAACCCATT-3', R:5'-CCATCCATCGGTAGTCCG-3' (6) with the expected size of 150 bp; and junctional adhesion molecule (JAM1): F:5'-CCCTCTTGCTGTATGGTCCG-3', R:5'-TGACCTTGACTGATGGCTTGATTTTGC-3' (JAM1): F:5'-GTAACCCGCTGAAACCCATT-3', R:5'-CCATCCATCGGTAGTCCG-3' (6) with the expected size of 150 bp. The glyceraldehyde-3-phosphate dehydrogenase (GAPDH) primers were obtained from Applied Biosystems (Foster City, CA). The quantity of target RNA (either JAM1 or GAPDH) was calculated by interpolating from the standard curve generated for that specific target. Both JAM1 and GAPDH quantities were normalized to 18S rRNA to calculate the relative quantity.

Microarray Hybridization and Data Processing

To minimize the effects of technical variability, membrane-associated and secreted cytoplasmic RNA pools from one fractionation were processed in triplicate as follows. In three parallel reactions, 10 μg of total RNA from each pool was labeled, hybridized to the Affymetrix U133A microarray, processed, and scanned according to standard Affymetrix protocols. The six resulting CEL files (three membrane-associated and secreted and three cytoplasmic) were processed with the Bioconductor software suite (a set of libraries for R; ref. 7). The robust multiarray average algorithm (8–10) was used for normal-

MEMBRANE-ASSOCIATED AND SECRETED GENES IN BREAST CANCER

Downloaded from cancerres.aacrjournals.org on July 13, 2017. © 2004 American Association for Cancer Research.

4Web address: www.affymetrix.com/analysis/
RESULTS

RNA Fractionation and Verification. The fractionation of poly-
somes by sucrose density gradient centrifugation was first described
by Mechler (4), and it was based on the observation that genes
encoding proteins that are membrane-associated or secreted are trans-
lated by ribosomes bound to the endoplasmic reticulum (membrane-
bound polysomes), whereas genes encoding proteins that are cytosolic
are translated by ribosomes free in the cytosol (free polysomes). Membrane-bound polysomes, being less dense, rise to the top of the
gradient, whereas free polysomes remain near the bottom of the
gradient.

Here, the method was used to separate intact polysomes prepared
from the MCF-7 breast cancer cell line. In a typical fractionation, the
separation results in two distinct peaks of A_{260} nm, with the lower peak
representing free polysomes and the upper peak containing the less
dense membrane-bound polysomes. To prepare sufficient RNA for
Affymetrix hybridizations, it was necessary to fractionate polysomes
from 3 x 10^6 cells; the results of this procedure are shown in Fig. 1.
Fractions from each peak were pooled, and the fractions in the peak
nearest the bottom (2 to 15) of the gradient are designated cyto-
plasmic, whereas fractions in the peak nearest the top of the gradient (20
through 26) are designated membrane and secreted. Sucrose fractions
with near-baseline absorption (16 through 19) in between the cyto-
plasmic and membrane-associated and secreted pools were saved as
negative controls. The peak at the surface of the gradient (the top 1.5
mL fraction) was discarded.

To confirm that the membrane-associated and secreted and cyto-
plasmic pools were enriched for membrane-associated and secreted-
associated and cytoplasmic-associated mRNA, respectively, real-time
quantitative reverse-transcriptase PCR was done with two primer
pairs expected to amplify coding sequences specific for each popula-
tion. We reverse transcribed 1 µg of total RNA from the mem-
brane-associated and secreted and cytoplasmic pools and labeled
the resulting cDNA in three separate reactions for each pool. JAM1 is
primarily cell-surface associated, whereas GAPDH is a protein found
free in the cytoplasm. Because of their different biological sequestering,
we expected JAM1 to be more highly represented in the mem-
brane-associated RNA, whereas the opposite will be true for
GAPDH. To confirm the physical separation of these two RNAs, the
membrane-associated and secreted/cytoplasmic expression ratio was
calculated (see Table 1) by taking the ratios of the averages from the
triplicates. As seen in Table 1, the membrane-associated and secreted/
cytoplasmic expression ratio is about 1,000-fold greater for JAM1
than for GAPDH, demonstrating a marked enrichment in the mem-
brane-associated and secreted pool for JAM1.

The RNA pools were then labeled and hybridized to Affymetrix
U133A microarrays. Membrane-associated and secreted expression
and cytoplasmic expression are the values returned by the robust
multiarray average calculation of expression measured on the Af-
nymetrix array hybridized to membrane-associated and secreted and
cytoplasmic RNA, respectively, and were calculated for each microar-ray element by averaging the expression value across the appropriate
triplicate (supplementary data). The membrane-associated and secreted/
cytoplasmic ratio for GAPDH (AFFX-HUMGAPDH/M33197_M_at) and
JAM1 (221664_s_at) was then calculated, as shown in Table 1. As expected, the membrane-associated and secreted pool shows an
enrichment for JAM1 as compared with GAPDH, whereas the cyto-
plasmic pool shows an enrichment for GAPDH. All microarray data
are available at the Gene Expression Omnibus (14) as accession
number GSE1400.

Reference Set Construction. Because the distribution of mem-
brane-associated and secreted/cytoplasmic ratios for either class is not
known a priori, it was necessary to train a classifier with a reference
set of genes with known subcellular localization. Of all 22,283 ele-
ments on the Affymetrix U133A array, subcellular location annota-
tion, as described in Materials and Methods, was available for 9,851
elements. Unambiguous membrane-associated and secreted annota-
tion was found for 3,188 of these, whereas unambiguous cytoplasmic
annotation was found for 798 elements. These elements with unam-
biguous annotation represent the reference set.

Expression Threshold Calculation. It is likely that only a subset
of the elements on the U133A microarray will be expressed in MCF-7
cells at a level great enough for meaningful measurement. To deter-
mine that level, we evaluated our ability to distinguish known mem-
brane-associated and secreted genes from known cytosolic genes in
the reference set at various total expression (E_T) levels, where E_T
= membrane-associated and secreted expression + cytoplasmic expres-
sion, where membrane-associated and secreted and cytoplasmic ex-
pression are the average exponentiated expression values for the
membrane-associated and secreted and cytoplasmic microarrays, re-
spectively. A 10-fold cross validation was done at increasing thresh-
old levels of E_T, including only training set members with an E_T
value ≥ threshold. Briefly, for each E_T level, the data were randomly
partitioned into 10 groups, 9 of which were used as a “training” set,
and the remaining group was designated as a “testing” set. At each E_T
level, the membrane-associated and secreted/cytoplasmic ratio thresh-
old was calculated (as described in Materials and Methods) for the
training set at that E_T level. The positive predictive value, sensitivity,
and specificity were calculated by examining the performance of
predicting the testing set for that E_T level, and averages over the 10
groups were recorded. The results of these calculations are shown in
Fig. 2. As shown in Fig. 2A, the performance of prediction for E_T
thresholds ranging from 22,283 (100% of the microarray elements) to
1,106 (4.9% of the microarray elements) was examined. The E_T level
that corresponded to the highest sensitivity without a significant drop
in positive predictive value or specificity was 738. At this level only
24.6% of probe sets with the highest E_T are included, resulting in a

<table>
<thead>
<tr>
<th>Target</th>
<th>MS/CYT ratio, as measured by reverse transcriptase-PCR</th>
<th>MS/CYT ratio, as measured by Affymetrix U133A microarray</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>0.00064</td>
<td>0.986</td>
</tr>
<tr>
<td>JAM1</td>
<td>0.387</td>
<td>1.173</td>
</tr>
</tbody>
</table>

Abbreviation: MS/CYT, membrane-associated and secreted/cytoplasmic.
final dataset of 5,483 probe sets that pass this threshold filtering. At this ET level, our membrane-associated and secreted prior probability estimate is 7.2%, with a corresponding cytoplasmic prior probability of 92.8%. Of those probe sets above this ET, 538 have unambiguous membrane-associated and secreted annotation and 305 have unambiguous cytoplasmic annotation. Additionally, at this level, our 10-fold cross-validation yields a 97.5% positive predictive value with 80.7% sensitivity and 96.9% specificity.

Membrane-Associated and Secreted/Cytoplasmic Ratio Threshold Calculation. All of the 843 probe sets in the reference set (with an ET above the threshold of 723) were used to determine the membrane-associated and secreted/cytoplasmic ratio that corresponds to the maximum posterior probability of belonging to the membrane-associated and secreted class. The distribution of membrane-associated and secreted/cytoplasmic ratios for genes with known localizations was examined (Fig. 3). It is interesting to note that the cytoplasmic genes show a discrete peak, whereas the membrane-associated and secreted genes show a bimodal distribution with a smaller peak that associates with the cytoplasmic genes. The membrane-associated and secreted/cytoplasmic ratio of 1.08 was calculated as giving the maximum posterior probability. Note that above this level, the majority of known cytoplasmic genes are excluded (only 3.2% are above this level), and a sizeable fraction of the known membrane-associated and secreted genes (22%) show a lower membrane-associated and secreted/cytoplasmic ratio. Thus, genes with a ratio below 1.08 cannot be designated with certainty as either membrane-associated and secreted or cytoplasmic.

The distribution of membrane-associated and secreted/cytoplasmic ratios for the remaining probe sets (genes of unknown cellular localization) is plotted in Fig. 4. Of these, 755 probe sets fall above the expression threshold and above the membrane-associated and secreted/cytoplasmic ratio of 1.08. These 755 probe sets are labeled as “predicted membrane-associated and secreted.” The remaining 3,885 probe sets found above the expression threshold and below the membrane-associated and secreted/cytoplasmic ratio of 1.08 are labeled as “indeterminate,” because we expect a mixture of cytoplasmic and membrane-associated and secreted genes in this range of membrane-associated and secreted/cytoplasmic ratios. Of the predicted membrane-associated and secreted probe sets, 323 were found to have a tentative subcellular annotation but did not meet the criteria previously established for the reference set. The remaining 432 probe sets have no subcellular annotation. A similar percentage of indeterminate probe sets were found to have some tentative subcellular annotation (1516 of 3885).

The Swiss-Prot annotations were searched for terms that might

Fig. 2. Identifying the optimal ET threshold for predicting membrane-associated and secreted genes. A. The number of probes with total expression level ET above specific threshold values of ET. B. The percentage of correctly labeled membrane-associated and secreted and cytoplasmic genes at differing ET thresholds. C. The number of known membrane-associated and secreted genes that are correctly predicted at differing ET thresholds. D. The number of known cytoplasmic genes that are correctly predicted at differing ET thresholds. Averages of the 10-fold cross validation results are plotted in B–D.

Fig. 3. Distribution of membrane-associated and secreted/cytoplasmic ratios for all of the genes in the reference set expressing above the ET. The midpoints of bins from frequency histograms are plotted (for visual clarity, bins are 0.05 units wide). The vertical line indicates a membrane-associated and secreted/cytoplasmic ratio of 1.08. The distribution of membrane-associated and secreted genes is plotted with dashed lines, whereas the solid line indicates the distribution of cytoplasmic genes. (MS/CYT, membrane-associated and secreted/cytoplasmic)

Fig. 4. Distribution of membrane-associated and secreted/cytoplasmic ratios for genes that are not in the reference set expressing above the ET. The midpoints of bins from frequency histograms are plotted (for visual clarity, bins are 0.05 units wide). The vertical line indicates a membrane-associated and secreted/cytoplasmic ratio of 1.08. (MS/CYT, membrane-associated and secreted/cytoplasmic)
indicate a tentative assignment to a cellular fraction (e.g., “membrane by similarity” or “nuclear”). Table 2 summarizes the tentative localization annotations for the predicted membrane-associated and secreted and indeterminate groups. Seventy percent (214 of 323) of the predicted membrane-associated and secreted probe sets with tentative annotations indicate a membrane-associated and secreted subcellular location. The binomial probability (with the prior probability of membrane-associated and secreted class membership as calculated in Statistical Calculations) of obtaining this number of membrane-associated and secreted probe sets by chance is very low \((P < 0.005)\), indicating that the probe set population with membrane-associated and secreted/cytoplasmic ratios \(\geq 1.08\) is significantly enriched for membrane-associated and secreted genes. Less than 2\% (6 of 323) of the predicted membrane-associated and secreted probe sets with tentative annotation are thought to be cytoplasmic. The remaining probe sets have either conflicting annotation or are thought to be localized to the nucleus, the endoplasmic reticulum, mitochondria, or other intracellular locations. Biochemical process annotation was available for 224 of these 323 probe sets in Gene Ontology (15). Over half of these seem to be involved in metabolism, whereas one third are involved in cell growth. Almost 25\% of the predicted membrane-associated and secreted class are involved in cell communication. (Although these annotations seem to comprise a greater number than the actual number of annotated probe sets, Gene Ontology is organized in a way such that multiple annotations can correspond to a single probe set.)

In contrast, 7.5\% (113 of 1516) of the indeterminate probe sets with tentative annotation indicate a membrane-associated and secreted localization. Although only 12.5\% (189 of 1516) of these are thought to be cytoplasmic, a significant fraction of the probe sets with conflicting annotation indicate a possible cytoplasmic localization. Interestingly, >60\% of the indeterminate probe sets contain nuclear or mitochondrial annotation.

**Analysis of a Gene Expression Study for Membrane-Associated and Secreted Gene Content.** The MCF-7 membrane-associated and secreted gene dataset was used to filter a differential gene expression study in breast cancer, which compared tumors with good versus poor 5-year outcome (5). We asked whether the membrane-associated and secreted localization provided by our study might give additional insight into the interpretation of the results and facilitate the selection of target genes for additional evaluation.

In the van’t Veer et al. (5) study, RNA from 98 primary breast tumors was hybridized to cDNA microarrays, and the resultant analysis led to a 231-gene expression profile associated with poor prognosis. The original study was performed on cDNA glass slide microarrays; therefore, we needed to find which elements of the Affymetrix U133A microarray corresponded to the 231 genes from the original study. It was possible to map 166 of these 231 genes to 269 probe sets on the Affymetrix microarray. Of these 269 probe sets, 20 were found in our predicted membrane-associated and secreted database representing 15 unique genes (see Table 3); an additional 52 were found in our training set of previously known membrane-associated and secreted genes. Of the genes not in the training set, almost half (7 of 15) had no subcellular location annotation in Gene Ontology or Swiss-Prot, although one had a published characteriza-

tion. Of the 9 genes with functional annotation, 5 are involved in metabolism, along with one each involved in signal transduction, cell-cycle regulation, proteolysis, and calcium binding. It is interesting to note that of the genes without functional annotation, HCCR1 is a putative proto-oncogene, fucosyltransferase 8 is thought to contribute to malignancy, “G protein-coupled receptor 126” contains a “protein tyrosine phosphatase-like protein” domain, and “hypothetical protein FLJ22341” contains a rhomboid domain, thought to regulate epidermal growth factor receptor expression. Any of these proteins, the up-regulation of which is associated with poor prognosis in breast cancer, merit additional investigation as potential treatment targets.

**DISCUSSION**

We describe here a novel set of membrane-associated and secreted genes expressed in MCF-7 cells. We are able to annotate 755 probe sets as membrane-associated or secreted, 432 of which had no previous subcellular location annotation. Two levels of validation strengthen our location predictions. First, we did 10-fold cross validation on the set of genes with annotated localization, which is a robust method for estimating performance on future datasets with similar characteristics. On the basis of the results of the 10-fold cross validation, it is likely that a great number of the predicted membrane-associated and secreted genes will have membrane-associated and secreted localization. This is reflected by the average 97\% positive predictive value observed in the 10-fold cross validation. Second, we examined the tentative annotations of genes in the set that were not used in the cross validation test and for which we predicted subcellular localization. Many of these have some tentative annotation, which we do not consider definitive. Nevertheless, our membrane-associated and secreted predictions coincide with these tentative annotations 70\% of the time.

Here we describe a general method of applying density gradient fractionation of RNA to the Affymetrix platform, including a robust statistical analysis. Furthermore, we have described an approach that can easily be modified for other tissues or states for comparative studies.

To minimize technical variability, hybridization data were collected in triplicate, with three independent labeling experiments on RNA collected from one fractionation experiment. It was not possible to compare the results obtained from multiple fractionations of different cell cultures because of the prohibitive cost of processing these large volumes of cells and of Affymetrix hybridization. Thus, the results shown here represent a “snapshot” of a cell line at a single point in time; it is possible that the representation of some genes, and even their membrane-associated and secreted/cytoplasmic distribution, will vary with different culture conditions. Indeed, this approach might be used to investigate global changes in subcellular distribution of proteins under various biological conditions, which to our knowledge has not been addressed previously.

Our Bayesian analysis may be over- or underestimating membrane-associated and secreted localization because of some violations of the equation assumptions. The localization of different genes are not entirely independent observations. For instance, there are clearly genes that colocalize because of genetic interactions. In addition, we

**Table 2 Tentative subcellular annotation for probe sets with predicted localization**

<table>
<thead>
<tr>
<th>Predicted location</th>
<th>Total probe sets</th>
<th>Probe sets with tentative subcellular annotation</th>
<th>Tentative MS annotation</th>
<th>Tentative cytoplasmic annotation</th>
<th>Tentative nuclear or mitochondrial annotation</th>
<th>Conflicting annotation</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>755</td>
<td>323</td>
<td>214 (69.8%)</td>
<td>6 (1.4%)</td>
<td>56 (15.9%)</td>
<td>15 (3.6%)</td>
<td>32 (9.2%)</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>3885</td>
<td>1516</td>
<td>113 (7.5%)</td>
<td>189 (12.5%)</td>
<td>961 (63.4%)</td>
<td>219 (14.4%)</td>
<td>34 (2.2%)</td>
</tr>
</tbody>
</table>

Abbreviation: MS, membrane-associated and secreted.
make the assumption that these two classes are mutually exclusive, which may not be true for a small fraction of genes. The robust multiaarray average algorithm might be a different source of underestimation for membrane-associated and secreted genes. It is possible that alternative microarray processing algorithms may yield additional predicted membrane-associated and secreted genes. Despite these drawbacks, we believe this will be a useful tool for investigators wishing to filter existing or future breast cancer Affymetrix datasets to look for membrane-associated and secreted genes. Alternative statistical methods may be useful for additional analysis and confirmation of our results.

There are a significant number of genes with unambiguous membrane-associated and secreted annotation that fall below our membrane-associated and secreted/cytoplasmic threshold. It is unclear if this is because of a real biological process (some of those membrane-associated and secreted genes are not membrane-associated and secreted localized in MCF-7 cells, for instance) or a processing artifact. Additional experimental analysis is needed to elucidate the mechanism in action. Additional study is also needed to determine whether the protein localization we discovered for MCF-7 cells holds true when analyzing other breast cancer cell lines.

ACKNOWLEDGMENTS

The authors wish to acknowledge the helpful equipment support of Dr. Alexander Mankin from University of Illinois at Chicago Pharmaceutical Biotechnology and the technical assistance of Mr. Doron Galili.

REFERENCES

Membrane-Associated and Secreted Genes in Breast Cancer

Nathan O. Stitziel, Brenton G. Mar, Jie Liang, et al.

*Cancer Res* 2004;64:8682-8687.

Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/64/23/8682

This article cites 13 articles, 7 of which you can access for free at:
http://cancerres.aacrjournals.org/content/64/23/8682.full#ref-list-1

This article has been cited by 4 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/64/23/8682.full#related-urls

Sign up to receive free email-alerts related to this article or journal.

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.