CALL FOR PAPERS | Comparative Genomics

Bovine mammary gene expression profiling using a cDNA microarray enhanced for mammary-specific transcripts

Steven P. Suchyta,1,2 Sue Sipkovsky,2 Robert G. Halgren,3 Rachael Kruska,1,2
Michael Elftman,1,2 Miriam Weber-Nielsen,1 Michael J. Vandehaar,1
Lan Xiao,1,2 Robert J. Tempelman,1,2 and Paul M. Coussens1,2

1Department of Animal Science, 2Center for Animal Functional Genomics, and 3Genomics Technology Support Facility, Michigan State University, East Lansing, Michigan 48824

Submitted 5 March 2003; accepted in final form 7 October 2003

MAMMARY GLAND DEVELOPMENT involves a series of complex physiological processes. There are six defined stages involved in development of the mammary gland including embryonic, prepubertal, pubertal, pregnancy, lactation, and involution. Each of these stages involves a complex interaction of hormones, growth factors, and signal transduction pathways, ultimately leading to expression of developmentally regulated genes. Although some factors involved in mammary development, such as the hormones prolactin, estrogen, and progesterone, have been known for some time, it has only been in the last 10 years that individual genes involved in mammogenesis have been identified. Gene targeting models have been used to uncover several important signaling pathways and genes that are directly involved in mammary development (for review see Ref. 22). For example, a functional Jak2/Stat5 pathway is absolutely required for mammary cell differentiation (35, 36, 44, 57). Additionally, the differentiation factor RANKL, which interacts with the Jak2/Stat5 pathway, is obligatory for alveolar development (14, 24). The cyclin D1/p27 genes are also crucial for mammary development during pregnancy (13, 15, 18, 30, 45, 46 67).

Although these discoveries have greatly improved our understanding of the pathways and genes involved in mammary gland physiology, our understanding of this complex process is still far from complete. For example, target genes for the transcription factors involved in mammary gland development, such as Stat3, are still unknown. Furthermore, studies of cyclin D1/p27 have shown that there are additional, as yet unknown, cell cycle pathways involved (18, 21, 51). There is also very little understanding of how these different components interact with one another to coordinate the many physiological changes observed in developing mammary tissue. Clearly, a better understanding of the target genes and pathways involved in normal mammary development is needed before we can take a more focused approach to treatment and prevention of mammary gland diseases.

A good example of this is the relationship between lactation and breast cancer risk. It is known that breast-feeding greatly reduces the risk of breast cancer (5, 31, 47, 53, 69). The reason this occurs is not fully understood. Studies of apoptosis as it relates to normal lactation cycles and onset of breast cancer have led to some possible connections. Abnormal regulation of apoptosis has been demonstrated to be directly involved in human breast cancer (8, 20, 27, 34). Apoptosis also occurs during the processes of lactation and involution, utilizing many of the same apoptotic pathways that are involved in breast cancer development (for review see Ref. 20). Many of the genes that play a role in breast cancer apoptotic pathways have been identified (i.e., EGFR, IGF-I, bcl-2) using breast cancer cell lines. Although these studies have led to elucidation of some critical pathways in programmed cell death (PCD) related to breast cancer, the complex interaction of PCD pathways in vivo as well as their function in normal lactation cycles are unknown. A better understanding of the PCD pathways involved in normal lactation cycles and mammary develop-
ment in vivo will be necessary before we can understand the link between protection from breast cancer and breast-feeding.

One of the models to study apoptosis during normal lactation cycles is the bovine. Because of their large size, cows provide a large source of lactating mammary tissue, which could be a limiting factor in other species. The cellular structure of the human mammary fat pad, parenchymal growth, and epithelial-stromal interactions are histologically similar to ruminants, but distinct from the mouse (23, 25). Despite differences in fatty acid synthesis in mammary cells of ruminants and nonruminants, the cellular similarities between ruminants and humans suggest that the bovine may prove to be a good physiological model along with rodents for human mammary physiology.

Bovine cell lines (including a bovine mammary epithelial cell line, MAC-T; Ref. 26) have been used successfully as model systems in a number of studies focused on host immune response (9, 32) and breast cancer (19, 48, 68). However, these studies were limited because, until recently, there has been a lack of attention in development and use of functional genomic resources in livestock animals. Functional genomics resources, such as a high-density cDNA microarrays would allow for global analysis of gene expression from fresh tissues, providing a critical link to in vitro studies.

This paper describes development of a >1,000 gene cDNA microarray enhanced for bovine mammary transcripts (BMAM). To identify pathways and mechanisms specifically related to bovine mammary gland growth and development, gene expression patterns from eight different tissues including developing mammary tissue were compared using the BMAM microarray. We have identified a set of 16 genes that we believe to be preferentially expressed in the developing mammary gland. These genes may thus represent a developing bovine mammary gland gene expression “signature,” providing good targets for future studies aimed at gaining a better understanding of mammary physiology. One of our long-term goals is to elucidate the genes and pathways associated with bovine lactation and involution and to use this as a model for human mammary gland development as it relates to human breast cancer risks. To begin this process, we compared gene expression profiles of lactating bovine mammary tissue against nonlactating bovine mammary tissue on the BMAM microarray. Our results suggest that many genes of interest exhibit differential expression in lactating mammary tissue relative to nonlactating tissue. In addition to the expected milk protein genes, genes involved in induction/repression of apoptosis, cell cycle control, and energy metabolism were also differentially expressed.

**Materials and Methods**

Selection of expressed sequence tag (EST) clones. The United States Department of Agriculture (USDA), Agricultural Research Service, Beltsville Agricultural Research Center (ARS), previously constructed a cDNA library from pooled mRNA isolated from mammary tissues at eight physiological, developmental, and disease states (55). Additionally, there are currently available four pooled-tissue normalized bovine cDNA libraries constructed by the USDA Meat Animal Research Center (MARC) encompassing a wide variety of bovine tissue types, excluding mammary tissue (Table 1) (54). All ESTs in the MARC and BARC libraries are represented in The Institute for Genomic Research (TIGR) Cattle Gene Index (BgI), and are ordered according to sequence overlap in clusters (TIGR Clusters, TC).

EST clones derived from mammary tissue were selected from available TCs using an “electronic subtraction” scheme. Based on the concordance of TC number and EST composition (TIGR file BTGI. TC_EST.0101), the library name (MARC or BARC) was determined for ESTs associated with a given TC through examination of the GenBank record for each EST. The TIGR BgI (Release 5.0) contained a total of 20,209 TCs, 18,263 of which contained at least one MARC or BARC clone. All TCs containing a MARC clone were excluded, since these would have been derived from tissue other than mammary gland. When the 18,263 TCs were analyzed, 1,174 did not contain a MARC clone. If more than one BARC clone was available for a given TC, then the clone with the longest sequence read length was selected to represent each cluster through examination of the nucleotide sequence given in each clone’s GenBank entry. Several Perl (v. 5.6.1) scripts were written to facilitate clone selection, and these are available upon request. Selected BARC clones (1,174) were used to form a bovine EST clone set (BMAM) that is enriched for transcripts derived from mammary tissue.

**DNA sequencing and sequence analysis.** To make the BMAM EST clone set from the selected BARC clones, 1 μl of liquid culture from the selected BARC clones was transferred into individual wells of 96 well plates that contained 150 μl of Luria broth (LB) plus 8% wt/vol glycerol and 1,000 μg/ml ampicillin using a Tecan Genosys 150 liquid handling robot (Tecan US, Research Triangle Park, NC). Plates were covered with foil and grown overnight at 37°C with constant shaking at 225 rpm. Replicates of the BMAM library were made from the overnight cultures and stored at −80°C. A replicate of the entire BMAM EST clone set was processed for commercial high-throughput sequence verification (Genome Therapeutics). Genome Therapeutics uses the Amersham TempliPhi φ29 DNA Polymerase-based Rolling Circle Amplification kit (Amersham Biosciences, Piscataway, NJ) to amplify the plasmid prior to sequencing. Vector and quality trimmed sequence data from all of the BMAM clones were compared with the BARC EST clone set via BLAST. Of the 1,172 EST clones sent for sequencing, 1,114 matched (95%) with the original BARC clone sequence. This subset of 1,114 EST clones was selected to create the BMAM cDNA microarray.

**Production of cDNA microarrays.** Clone inserts from the BMAM library were amplified by adding 1 μl of a 1:10 dilution [in double-

---

**Table 1. Bovine tissue distribution of cDNA libraries used in selection of EST clones for the BMAM EST collection**

<table>
<thead>
<tr>
<th>cDNA Library</th>
<th>Tissues</th>
<th>No. of Clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARC 1</td>
<td>Lymph node, ovary, fat hypothalamus, and pituitary gland</td>
<td>32,000</td>
</tr>
<tr>
<td>MARC 2</td>
<td>Testis, thymus, semitendinosus muscle, longissimus muscle, pancreas, adrenal, and endometrium</td>
<td>34,000</td>
</tr>
<tr>
<td>MARC 3</td>
<td>Marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle</td>
<td>39,000</td>
</tr>
<tr>
<td>MARC 4</td>
<td>Day 20 and day 40 embryos</td>
<td>29,000</td>
</tr>
<tr>
<td>BARC 5</td>
<td>Mammary tissue at eight physiological, developmental, and disease states</td>
<td>23,000</td>
</tr>
</tbody>
</table>

Total number of clones is 157,000. MARC, United States Dept. of Agriculture (USDA). Meat Animal Research Center, Clay Center, NE; BARC, USDA Beltsville Animal Research Center, Beltsville, MD. EST, expressed sequence tag.
distilled H$_2$O (ddH$_2$O$_2$) of the TempliPhi d29 Polymerase (Amersham Biosciences) reaction used for sequencing directly into PCR master mix [20 μM each dNTP, 0.5 μM forward primer, 0.5 μM reverse primer, 200 nM Tris-HCl (pH 8.4), 500 mM KCl, and 2.0 U of Taq DNA polymerase] in 96-well plates. The forward primer (5′ ATCTCCGGGATACCTGCAG 3′) and reverse primer (5′ CCTCAGGAGGATACTCTAGACG 3′) are immediately adjacent to the insertion site for the cloning vector pSPORT6 used for construction of clones. Following a preheat step of 94°C for 2 min, PCR was performed in a thermocycler (model PE 9700; PerkinElmer, Palo Alto, CA) using the following conditions: 94°C for 30 s; 60°C for 1 min, and 72°C for 3 min; for 30 cycles. Based on agarose gel electrophoresis of purified amplicons, our success rate allowed triple spotting of each clone within a patch and an overall efficiency of 48% (Genomic Solutions, Ann Arbor, MI) equipped with a 48-pin head, of liver, spleen, thymus, adrenal, adipose, ileum, and lymph tissue from a primiparous cow that was slaughtered at 181 days of lactation. A reference RNA for use in microarray hybridization was prepared from an immortalized bovine macrophage cell line (56). We felt that a macrophage cell line would be a good source of cDNA that would not be expected to have a high level of gene expression differences between a panel of genes expressed for mammary gland transcripts. Additionally, a stable cell line should be a reliable source of cDNA that could facilitate a connection between this study and future studies. For further experiments, we used the reference cDNA as a base to compare gene expression levels from various bovine tissue samples. The reference cell line was grown in 75-mm² flasks containing RPMI-1640 with 10% fetal bovine serum, 2 mM l-glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin in a humidified atmosphere of 5% CO$_2$ and 95% air at 39°C. RNA was extracted from all tissues and cells using TRIzol reagent (Invitrogen) as described previously (59). Quantity and quality of total RNA extracted from all tissues and cells was estimated by UV spectrophotometry and electrophoresis on 1.2% native agarose gels. For all experiments, total RNA (8 μg) was used as template in reverse transcription reactions, using the Atlas Glass PowerScript Fluorescent labeling system (BD Biosciences, Alameda, CA) with oligo(dT)$_1$$_$_4$ as primer. Synthesis of cDNA in this system incorporates an amino-modified dUTP into the cDNA. Following first strand synthesis, cDNAs were labeled using n-hydroxysuccinate (NHS)-derivatized Cy3 and Cy5 dyes (Amersham Biosciences). Labeled cDNAs were purified to remove unincorporated dyes, combined and concentrated to ~10 μl using Microcon 30 spin concentrators (Millipore). Pre-warmed (70°C) SlideHyb no. 3 (Ambion), 100 μl was added to probe mixtures, and the final probe solutions were applied to precoated microarray slides via the port of a GeneTAC Hydridization Station microarray hybridization chamber (Genomic Solutions). Hybridizations were conducted for 18 h using step-down temperatures from 65°C to 42°C in a sealed microarray hybridization station (Genomic Solutions). Following hybridizations, the station applies three washes, two with medium stringency buffer and one with high-stringency buffer (Genomic Solutions). Slides were finally rinsed briefly at room temperature in 2× saline sodium citrate and once in ddH$_2$O. Washed microarrays were dried by centrifugation at 1,000 rpm in a cushioned 50-ml conical centrifugation tube.

Dried microarrays were scanned immediately using a GeneTAC LS IV microarray scanner and GeneTAC LS software (Genomic Solutions). GeneTAC Integrator 4.0 software was then used to process array images, align spots, integrate robot-spotting files with the microarray image, and to export reports of spot intensity data. The final report was retrieved as raw spot intensities in comma separated value files.

**Microarray data analysis.** Microarray data was normalized using the scatter plot smoother LOWESS (10) from the statistical package SAS (“PROC LOESS”) according to the method of Yang et al. (64). Briefly, the normalization process is as follows, M-A plots are constructed for each slide, where log-intensity ratios M = $\log(Cy3/Cy5)$ = $[\log(Cy3) - \log(Cy5)]$ are plotted against the mean log-intensity A = $[\log(Cy3) + \log(Cy5)]/2$ as described by Yang et al. (64). Normalization is then performed considering a robust local regression technique (11), using the LOWESS procedure of SAS (SAS Institute, Cary, NC). Efficiency of LOWESS normalization is assessed by monitoring M-A plots before and after normalization. The normalized data is then back transformed prior to further statistical analysis using the formula

$$\log Cy3^{\#} = A + M^{*}/2 \quad \text{and} \quad \log Cy5^{\#} = A - M^{*}/2,$$

where $\log Cy3^{\#}$ and $\log Cy5^{\#}$ are the normalized log intensities. Here, $M^{\#} = M - M'$ and represents each of the normalized M values, with M being the LOWESS predicted value for each spot.

LOWESS normalized array data was imported into Microsoft Excel for further analysis. After back-transforming normalized intensity values as above, the median blank intensity on a microarray for each dye was subtracted from the respective normalized spot intensity values. Residual intensity values were again log transformed prior to further analysis. Log-transformed and blank subtracted values were used to calculate a mean log expression difference for each gene for each tissue compared with reference cell cDNA. Since each gene is technically replicated three times on an array, a P value (t-test) can be determined for evidence of differential expression on each gene. This P value strictly refers only to the particular sample vs. reference rather than a biologically replicated mammary vs. reference comparison for...
any given gene on the array. The antilog of the mean log expression difference for an individual gene on the array yields the approximate fold change in expression in this gene between the tissue of interest and the reference cDNA. For a general mammary vs. reference fold change in expression in this gene between the tissue of interest and the reference RNA corrected intensity value and the reference RNA corrected intensity value was used.

By performing experiments that hybridized the same sample labeled with both Cy3 and Cy5 on the same slide, we were able to empirically determine the rate of false-positive signals expected on the BMAM cDNA microarray. By determining the false-positive rate we have a measurement of the rate of truly null features being called significant. Additionally, a conservative implementation of the false discovery rate (FDR) was calculated on individual microarray experiments (6). The FDR gives the rate that significant features are truly null. For a gene to be considered as significantly different, the FDR P value was required to be \( \leq 0.1 \). This liberal cutoff was used to avoid declaring too many false negatives, which may be as serious as false-positive declarations in genomic studies.

Microarray experiments. A reference microarray experimental design (28, 29, 65) was used to compare prepubertal mammary gland (136-kg Holstein heifer) gene expression against a panel of seven other tissues (experiment 1). This was accomplished by first starting with the 44 genes that had significantly greater expression in developing mammary tissue compared with reference cells (fold change, >1.5; FDR: \( P \leq 0.1 \)). The seven other tissues’ gene expression levels for these 44 genes relative to reference (log fluorescence intensity of the tissue’s cDNA minus the fluorescence intensity of the reference cells within the corresponding arrays) were subtracted from the developing mammary gland’s log fluorescence intensity (also expressed as a deviation from the fluorescence intensity of the reference cells within the corresponding arrays). Following the subtraction, the antilog of the difference was used to calculate the fold difference between the tissue and mammary gland. This comparison was done for each of the seven tissue types against the prepubertal mammary gland.

Experiment 2 compared animal 1 vs. nonlactating mammary tissue (136-kg Holstein heifer) and experiment 3 compared animal 2 vs. nonlactating mammary tissue (227-kg Holstein heifer). Mean log expression differences from animal 1 vs. nonlactating (136-kg Holstein heifer) (experiment 2) and mean log expression differences from animal 2 vs. nonlactating (227-kg Holstein heifer) (experiment 3) were used as biological replicates. An overall average log difference between lactating and nonlactating tissue was calculated using the mean log expression difference for each gene for animal 1 and animal 2. A P value (\( t \)-test) for differential expression was based on these two averages. Data for the 13 microarray experiments used in this report can be found in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO), series accession number GSE279.

Real-time quantitative PCR. To confirm gene expression differences observed from microarray results, quantitative real-time PCR was performed on six genes with overexpression in lactating mammary tissue relative to nonlactating tissue (based on experiments 2 and 3). These six genes were chosen because they had a known function that has not been previously associated with bovine lactation. To better account for biological variation among the samples, a total of seven nonlactating mammary samples (two prepubertal samples used in the microarray experiments and five additional prepubertal nonpregnant mammary samples) and seven lactating mammary samples (two lactating samples used in the microarray experiments and five additional midlactation mammary samples) were used in real-time quantitative PCR. The six genes were BMAM_BE589961 (Caspase8), BMAM_BE476665 (CD47 antigen/integrin-associated protein), BMAM_BF68293 (LIM and SH3 protein 1), BMAM_BE588590 (PBK1), BMAM_BF239097 (Vav 3 oncogene), and BMAM_BE481095 (c-myc proto-oncogene).

Real-time quantitative PCR primers were designed using Primer Express ver. 2.0 (Applied Biosystems, Foster City, CA) using the sequence of the respective EST as the source. Primers for the six genes used were BMAM_BE589961 (forward, 5’ AATGGAGCTGCT-GTTAATCCTATTGAT 3’; reverse, 5’ GCCCTGGCGGTATCTGG 3’), BMAM_BE476665 (forward, 5’ TGTACCCTTATATGTAATGCTAGCTGCTCTTGAT 3’; reverse, 5’ AGGGAGCTGCTCCAAAGGACA 3’), BMAM_BF868293 (forward, 5’ GAGGAAAGGGAAGATGGTGCTA 3’; reverse, 5’ AATCAGGCAAGGTTGTGCTTAG 3’), BMAM_BE588590 (forward, 5’ CTGAGAGGCTGCAAAGG 3’; reverse, 5’ GGAACACCCCACGGITTTGG 3’), BMAM_BE239097 (forward, 5’ CGCGTCCATCCCCATCTGTT 3’; reverse, 5’ GGGTAGGTTAGAGGAAAGCCA 3’), and BMAM_BE481095 (forward, 5’ GGCAGAAGGATGCTGTCCTT 3’; reverse, 5’ GCCGCTCTTCCTCCGGAGTGT 3’).

There was not publicly available a full-length cDNA sequence for bovine \( \beta \)-actin; therefore, the BMAM clone BMAM_BG689033 (596 bp) that has significant BLASTN similarity to human and mouse \( \beta \)-actin (\( E \) value \( < 10^{-50} \)) was sequenced in its entirety. The entire length of clone BMAM BG689033 was 1.8 kb and contained the full-length cDNA sequence for bovine \( \beta \)-actin (GenBank accession no. AY141970). Real-time quantitative PCR primers were designed using Primer Express ver. 2.0 (Applied Biosystems) based on the bovine \( \beta \)-actin sequence (forward, 5’ CGCCATGGATGATGATATCAG 3’; reverse, 5’ AAGGCCGCTCCGACAT 3’).

Quantitative PCR was performed in duplicate for each sample on an ABI 7000 Sequence detection system (Applied Biosystems) using the default two-step amplification procedure and 2X SYBR Green Master Mix (Applied Biosystems) in 50 \( \mu \)l reaction volume with 300 nM of each primer and 50 ng of cDNA.

Relative quantitation of mRNA abundance for genes of interest were calculated using the \( 2^{-\Delta\Delta CT} \) method of Livak and Schmittgen (37). Amplification of \( \beta \)-actin from each sample was used for normalizing quantitative PCR data (37). Relative quantitation of genes from individual lactating mammary gland samples was compared with mean data from nonlactating tissue samples.

An overall average between lactating mammary samples and nonlactating mammary samples was calculated by taking the average of the relative concentration of each lactating mammary sample compared with the overall nonlactating mammary samples. A \( P \) value (\( t \)-test) for the overall average was established using these seven averages. Thus these \( P \) values appropriately take into account true biological replication.

RESULTS

BMAM EST characterization and database. A web-accessible resource was established (http://nbfgc.msu.edu) to provide information on all EST clones contained in the BMAM microarray. The web database can be searched by cDNA clone name, ontology, or keyword (keyword based on BLAST information from the cDNA sequence). Clones in the BMAM EST collection were functionally classified using GeneSpring ver. 4.1.5 (Silicon Genetics, Redwood City, CA) Build Simplified Ontology function, which is based on the Gene Ontology Consortium classifications (2). Ontology classification assignment for each clone is based on the description line of the top BLASTX hit if available; otherwise, it is based on the top BLASTX hit corresponding to a functional gene ontology. Ontology classification for each EST is web-linked to AmiGO!, an HTML-based browser for The Gene Ontology from the Gene Ontology Consortium (http://www.geneontology.org). Of the original 1,174 BMAM EST clones, 733 (62%) matched to known genes (based on BLASTN or BLASTX matches with \( E \) value \( < 10^{-15} \)). Of these 733 clones, 116 matched to characterized bovine
genes; the remaining 617 ESTs matched to known genes from other species. In total, 246 of the 733 clones had matches to hypothetical genes or genes with no known function in either bovine (18), human (225), or mouse (5). We assigned 383 of the 773 genes with putative gene identifications to one of the 3 major categories of ontology (biological process, cellular component, and molecular function) using GeneSpring ver. 4.1.5 (Silicon Genetics).

A wide range of functional ontology categories are represented in the BMAM EST collection (Fig. 1). A large proportion of signal transduction, transport, and mitochondrial related genes were identified, consistent with the many diverse functions in growth, differentiation, milk synthesis, immunity, and involution characteristic of the mammary gland. The BMAM EST clone collection also contains 11 casine and casnine kinase related genes, as well as many plasma membrane associated genes that may be highly expressed in development of mammary parenchymal tissue, basement membrane, and extracellular matrix (ECM). Since the BARC library is a pooled library from prepubertal mammary tissue (136-kg Holstein heifer) was compared with cDNA from the reference RNA cells on the BMAM microarray. There were 44 BMAM EST clones that had a significantly greater expression (fold change, >1.5; FDR, P \(\leq 0.1\)) in prepubertal mammary tissue and 1 EST had a significantly greater expression (fold change, >1.5; FDR, P \(\leq 0.1\)) in reference cells relative to prepubertal mammary gland. This result demonstrates our ability to detect gene expression in developing mammary tissue and that, as expected, gene expression in the reference cells was low compared with mammary tissue.

Based in part on our preliminary analysis, we hypothesized that the BMAM cDNA microarray could be used to establish a gene expression profile in developing mammary gland. Specifically, we were interested in identifying genes that would have a high probability of being preferentially expressed in the prepubertal mammary gland. To accomplish this, an experiment was performed to compare gene expression levels of several tissue types (adipose, liver, adrenal, lymph, spleen, thymus, and ileum) to developing mammary tissue by utilizing a reference microarray experimental design (28, 29, 65) In this case, cDNA from the various tissue types was directly compared with cDNA from reference cells, using separate microarrays for each tissue. Of the 44 genes with a >1.5-fold expression in prepubertal mammary vs. reference, 39 had >1.5-fold higher expression in prepubertal mammary relative to adipose. Of the remaining 39 genes, 35 were expressed more than 1.5-fold greater in prepubertal mammary than in the 7 other tissues examined. These 16 genes did not show any significant difference in expression when RNA
from prepubertal and postpubertal mammary tissue were directly compared on the BMAM microarray (overall mean expression ratio 0.95) and when RNA from prepubertal (136-kg Holstein heifer) and lactating mammary were directly compared with lactating mammary tissue (overall mean expression ratio 1.05).

**Defining a set of genes differentially expressed in lactating mammary tissue.** Our next objective was to find genes that are differentially expressed in lactating mammary tissue relative to nonlactating tissue. To accomplish this, we performed microarray experiments directly comparing RNA from nonlactating mammary tissue (136-kg Holstein heifer and 227-kg Holstein heifer) to RNA from mammary tissue of two midlactation cows (animal 1, animal 2) (experiments 2 and 3). Using a cutoff of $\geq 1.5$ fold difference in gene expression at FDR $P \leq 0.1$ in lactating tissue compared with nonlactating tissue, there were a total of 75 genes upregulated in lactating mammary tissue from animal 1 relative to nonlactating tissue (136-kg heifer), and 112 genes upregulated in lactating animal 2 relative to nonlactating tissue (227-kg Holstein heifer). As expected, genes encoding $\alpha$S1-casein and $\beta$-casein were among those with enhanced expression in lactating tissue from both cows relative to nonlactating tissue. A total of 77 genes for animal 1 and 141 genes for animal 2 were expressed significantly lower (less than or equal to a $-1.5$-fold difference, FDR $P \leq 0.1$) in lactating mammary tissue relative to nonlactating tissue.

**Table 2. BMAM ESTs that have preferential expression in prepubertal mammary gland**

<table>
<thead>
<tr>
<th>Clone Name</th>
<th>Description</th>
<th>Ontology</th>
<th>Mean Fold Overexpression</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMAM_BG690952</td>
<td>Fibrinogen gamma-B chain</td>
<td>Cell adhesion molecule, blood coagulation, blood pressure regulation, positive control of cell proliferation</td>
<td>2.100</td>
</tr>
<tr>
<td>BMAM_BG91636</td>
<td>Putative membrane protein</td>
<td>Unknown</td>
<td>2.188</td>
</tr>
<tr>
<td>BMAM_BG687920</td>
<td>Sirtuin (silent mating type information regulation 2 S. cerevisiae homolog)</td>
<td>Unknown</td>
<td>2.079</td>
</tr>
<tr>
<td>BMAM_BE486110</td>
<td>Stomatin peptide</td>
<td>Cytoskeleton, integral plasma membrane protein, glucose transport</td>
<td>2.422</td>
</tr>
<tr>
<td>BMAM_BG91557</td>
<td>Isopentenyl-diphosphate Delta-isomerase (IDI1)</td>
<td>Peroxisome, isoprenoid biosynthesis, related to hypertension</td>
<td>2.368</td>
</tr>
<tr>
<td>BMAM_BE485618</td>
<td>GTP-binding protein, alpha-subunit</td>
<td>Peripheral plasma membrane protein, heterotrimeric G-protein GTPase, alpha-subunit, signal transducer</td>
<td>1.849</td>
</tr>
<tr>
<td>BMAM_BE486352</td>
<td>21615 ADH, oxidoreductase</td>
<td>Oxidoreductase</td>
<td>1.855</td>
</tr>
<tr>
<td>BMAM_BE486522</td>
<td>Unknown EST</td>
<td>Unknown</td>
<td>2.423</td>
</tr>
<tr>
<td>BMAM_BG687733</td>
<td>p53-binding protein 2</td>
<td>Cell cycle control, SH3/SH2 adaptor protein, cytoplasm, signal transduction</td>
<td>2.207</td>
</tr>
<tr>
<td>BMAM_BG691672</td>
<td>Similar to lipase A lysosomal acid cholesterol esterase (Wolman disease)</td>
<td>Lipoprotein lipase, lysosome, N-linked glycosylation</td>
<td>2.265</td>
</tr>
<tr>
<td>BMAM_BG690604</td>
<td>Organic cation transporter OCTN1</td>
<td>Unknown</td>
<td>2.343</td>
</tr>
<tr>
<td>BMAM_BE486444</td>
<td>Unknown EST</td>
<td>Nucleotide binding, integral plasma membrane protein, fluid secretion, organic cation porterorganic cation transport</td>
<td>2.414</td>
</tr>
<tr>
<td>BMAM_BE477209</td>
<td>Unknown EST</td>
<td>Unknown</td>
<td>2.046</td>
</tr>
<tr>
<td>BMAM_BE478125</td>
<td>Developmentally regulated RNA-binding protein 1</td>
<td>Unknown</td>
<td>1.939</td>
</tr>
<tr>
<td>BMAM_BE478128</td>
<td>Unknown EST</td>
<td>Receptor, integral plasma membrane protein</td>
<td>1.893</td>
</tr>
<tr>
<td>BMAM_BG691504</td>
<td>Butyrophilin, a plasma membrane protein involved in synthesis/release of milk fat globules</td>
<td>Unknown</td>
<td>2.101</td>
</tr>
</tbody>
</table>

"Mean Fold Overexpression" is the average fold overexpression of the EST in prepubertal mammary gland compared with adipose, liver, adrenal, lymph, spleen, thymus, and ileum.
To establish the biological significance of differential gene expression observed in nonlactating vs. lactating microarray comparisons, expression data from the two lactating cow samples were combined and analyzed. Genes that had a $\geq 3$-fold overall overexpression in lactating mammary gland relative to nonlactating are included in Table 3. Genes that had $\leq 3$-fold overall repression in lactating mammary tissue relative to nonlactating are included in Table 4. There was large variability in some genes because the magnitude of expression differed between cows, even though the direction of expression was the same. Due to this large variability, real-time quantitative PCR was performed on selected genes to confirm microarray results.

Quantitative real-time PCR. Figure 3 depicts the average relative amount of mammary mRNA for six genes with overexpression in lactating mammary tissue relative to nonlactating tissue (based on experiments 2 and 3) from seven lactating animals compared with seven nonlactating mammary samples using quantitative real-time PCR (36). Five genes were highly significant ($P \leq 0.05$) and had $\geq 2.5$-fold expression in lactating mammary tissue relative to nonlactating (C-myc, Caspase 8, VAV3, LASP1, and PBK1, Fig. 3). CD47 had $3.5$-fold

Table 3. BMAM clones tending to be upregulated in lactating mammary tissue compared with nonlactating mammary tissue

<table>
<thead>
<tr>
<th>Clone Name</th>
<th>Description</th>
<th>Lactating animal 1</th>
<th>Lactating animal 2</th>
<th>Avg</th>
<th>P Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMAM_BG689327</td>
<td>Unknown EST</td>
<td>17.07</td>
<td>34.97</td>
<td>24.42</td>
<td>0.0712</td>
</tr>
<tr>
<td>BMAM_BE589961</td>
<td>Caspase 8</td>
<td>16.33</td>
<td>36.38</td>
<td>24.38</td>
<td>0.0794</td>
</tr>
<tr>
<td>BMAM_BE476665</td>
<td>CD74 antigen/integrin-associated protein (CD74/ITGA6)</td>
<td>10.76</td>
<td>34.6</td>
<td>19.29</td>
<td>0.1466</td>
</tr>
<tr>
<td>BMAM_BE868293</td>
<td>LIM and SH3 protein 1 (LASP1)</td>
<td>21.17</td>
<td>16.32</td>
<td>18.58</td>
<td>0.0283</td>
</tr>
<tr>
<td>BMAM_BE588590</td>
<td>PBK1</td>
<td>27.74</td>
<td>11.14</td>
<td>17.57</td>
<td>0.1005</td>
</tr>
<tr>
<td>BMAM_BE490907</td>
<td>v-Abl oncogene (VAV3)</td>
<td>11.31</td>
<td>24.48</td>
<td>16.64</td>
<td>0.0868</td>
</tr>
<tr>
<td>BMAM_BE466198</td>
<td>Similar to yeast suppressor protein SRP40</td>
<td>17.44</td>
<td>13.64</td>
<td>15.43</td>
<td>0.0286</td>
</tr>
<tr>
<td>BMAM_BE481095</td>
<td>e-myc proto-oncogene</td>
<td>11.56</td>
<td>16.44</td>
<td>13.79</td>
<td>0.0427</td>
</tr>
<tr>
<td>BMAM_BE483939</td>
<td>Unknown EST</td>
<td>12.31</td>
<td>12.7</td>
<td>12.5</td>
<td>0.0039</td>
</tr>
<tr>
<td>BMAM_BE484690</td>
<td>Acetyl-CoA synthetase</td>
<td>10.68</td>
<td>5.955</td>
<td>10.1</td>
<td>0.0153</td>
</tr>
<tr>
<td>BMAM_BE481675</td>
<td>Myelination transcription factor 1</td>
<td>7.997</td>
<td>9.034</td>
<td>8.497</td>
<td>0.0176</td>
</tr>
<tr>
<td>BMAM_BE487635</td>
<td>Molybdenum cofactor synthesis 2 (MOCS2)</td>
<td>3.612</td>
<td>3.225</td>
<td>3.413</td>
<td>0.0293</td>
</tr>
<tr>
<td>BMAM_BE482108</td>
<td>Complementation group 3 (xeroderma pigmentosum group B complementing) (ERCC3)</td>
<td>3.129</td>
<td>3.929</td>
<td>3.507</td>
<td>0.0576</td>
</tr>
<tr>
<td>BMAM_BE480086</td>
<td>Chaperonin containing TCP1, subunit 5 (epsilon) (CCT5)</td>
<td>4.470</td>
<td>4.341</td>
<td>4.405</td>
<td>0.0063</td>
</tr>
<tr>
<td>BMAM_BE481948</td>
<td>PAS-4</td>
<td>3.548</td>
<td>1.528</td>
<td>2.329</td>
<td>0.2943</td>
</tr>
<tr>
<td>BMAM_BE476665</td>
<td>CD74 antigen/integrin-associated protein (CD74/ITGA6)</td>
<td>4.553</td>
<td>11.727</td>
<td>7.307</td>
<td>0.1487</td>
</tr>
<tr>
<td>BMAM_BE479994</td>
<td>Unknown EST</td>
<td>7.561</td>
<td>7.888</td>
<td>7.723</td>
<td>0.0066</td>
</tr>
</tbody>
</table>

*P*-distribution, $n = 2.
expression in lactating mammary gland relative to nonlactating mammary gland ($P = 0.09$, Fig. 3).

**DISCUSSION**

A cDNA microarray resource for studies of bovine mammary gland physiology has been developed and tested for its utility in detection of differential gene expression during physiologically relevant points in mammary gland development. The BMAM EST clone set and microarray contain a directed set of clones enhanced for transcripts found in bovine mammary gland and is annotated in a web-accessible database (http://nbfcg.msu.edu) designed to assist in interpretation of microarray data. BMAM clone data in our database allows retrieval of the top BLAST results for each clone against sequences in GenBank nonredundant (nr) database. Additional information, including clone sequence data, the TIGR TC number, and gene ontology information are also displayed. Links to public databases, such as GenBank, TIGR BtG1, and the Gene Ontology Consortium’s AmiGO! allow rapid retrieval of associated information for each clone. The database is searchable with regards to clone name, keywords derived from BLAST hits, BLAST searching within the database, and ontology term. Search results are of primary importance when analyzing data from microarray experiments, allowing for easy integration of gene expression data and putative gene identification and ontology information of the transcript represented by each clone.

We assessed the ability of BMAM microarrays to detect gene expression in developing mammary gland. In total, 44 genes displayed greater expression in prepubertal mammary gland relative to reference cells. Expression of RNAs represented on the BMAM cDNA microarray were further evaluated in a group of seven tissues (adipose, liver, adrenal, lymph, spleen, thymus, and ileum) using a reference design, allowing for cross-microarray comparison. After comparing gene expression data from all 7 tissues with prepubertal mammary gland, 16 genes were identified as having a high probability of being preferentially expressed in developing mammary tissue. For example, expression of butyrophilin (Table 2), a plasma membrane protein involved in discharge of milk fat globules (4), would be expected particularly during lactation, although a role may also exist in the developing mammary gland. Preferential expression of another 10 of these genes in prepubertal mammary gland is reasonable given their indicated roles as integral membrane proteins in cell proliferation and in other functions likely occurring in tissue that is growing and differentiating. The fact that 5 genes of the 16 genes (BMAM_BG691636, BMAM_BE486522, BMAM_BE486043, BMAM_BE477209, and BMAM_BE478128, Table 2) in this group have an unknown function provides an indication that there is much to learn despite significant prior research on control of mammary gland development and function in humans, mice, and cattle.

The fact that only 16 genes were detected with expression highly specific to developing mammary gland may, in part, be due to the relatively recent evolutionary acquisition of mammary glands in mammals. It has been suggested that many, if not all genetic pathways involved in development of the mammary gland are used in tissues that appear earlier in evolution (21). Genes regulating mammary gland development have been gained from a wide variety of tissue types, as illustrated by genes known to be involved in mammogenesis, such as Jak2/Stat5, RANKL, and the cyclins. These genes are

![Table 4. BMAM clones tending to be repressed in lactating mammary tissue compared to nonlactating mammary tissue](image_url)
also known to regulate bone morphogenesis, reproduction, retinal development, and brain and other tissue developmental programs. Despite this, the 16 genes identified in the current report appear to be highly specific to developing mammary gland and as such are potentially very good candidates for playing an important role in regulating mammogenesis.

Similarly, in a comparison of gene expression profiles in cDNA derived from lactating and nonlactating mammary tissue RNA on the BMAM microarray, we were able to identify numerous interesting genes that are affected by lactation in either a positive or negative manner (Tables 3 and 4). Many of the genes identified have not been previously associated with lactation. For example, LASP1 has 18-fold higher expression in lactating mammary tissue relative to nonlactating tissue as determined by microarray analysis (Table 3). LASP1 was first identified from a breast cancer-derived metastatic lymph node cDNA library (58). LASP1 has subsequently been mapped to the long arm of human chromosome 17 (17) and mouse chromosome 11 (49), which is almost completely syntenic with bovine chromosome 19 (3). Chromosome 17 in human is also where the breast cancer related genes c-erbB-2 and BRCA1 (43) are located. It has been suggested that c-erbB-2 and LASP1 belong to the same amplicon due to the co-amplification and co-overexpression of both genes in 12% of mammary tumors (7). LASP1 may be involved in tumorigenesis and/or in tumor progression (7). Schreiber et al. (50) showed that LASP1 is expressed in various degrees in a wide variety of adult mouse tissue, including mammary gland (some of the tissues where there was the strongest expression, such as placenta and colon were not included in the MARC cDNA libraries, which may explain why LASP1 was included in the BMAM EST collection). The strong evolutionary conservation, actin binding domain, and ubiquituous expression of LASP1, suggests that this gene plays an essential role in basic cellular functions (50). Lin et al. (33) have demonstrated that LASP1 is a downstream cytoskeletal target of c-ABL that functions to prevent cell migration on the ECM. Marti et al. (39) showed that there was a weak induction in actin during involution compared with lactating tissue. The fact that LASP1 is overexpressed in lactating mammary gland may be related to its function in regulating cell migration via its interaction with the actin cytoskeleton.

It is interesting to note that there is increased expression during lactation of one of the initiators of apoptosis (Caspase 8, 24-fold greater expression relative to nonlactating mammary tissue, Table 3), and two oncogenes, VAV3 and C-myc, (>10-fold higher in lactating tissue relative to nonlactating tissue, Table 3). Both oncogenes are regulators of cell division. Although the occurrence of apoptosis is well established during involution of mammary gland in cattle (63) and mice (38, 39, 40, 41, 59), few studies have addressed the significance of apoptosis in normal lactation. One study using immunohistochemistry to analyze expression of apoptosis-related genes in lactating goat mammary tissue (62), showed an increase in expression of many apoptosis-related genes over the course of lactation. Activated genes included TGF-B1, TGFBIIR, Bcl-2, and caspase 3. Since VAV3 is regulated by TGFβ (61), our results provide preliminary evidence that PCD may also be involved during lactation in cattle.

In addition to known genes discussed previously, we also identified five genes that have a greater than fourfold higher expression in lactating mammary tissue relative to nonlactating tissue that do not match any known genes (Table 3). For example, BMAM_BG689327 had >24-fold higher expression in lactating mammary tissue compared with nonlactating mammary tissue (Table 3). Further characterization of these genes and their protein products, as well as their role in mammary physiology is clearly warranted.

Glyoxalase I (GLO1) was significantly repressed (greater than 3-fold, Table 4) in lactating mammary tissue relative to nonlactating mammary tissue. GLO1 is a detoxifying enzyme, which has elevated expression in human breast cancer tissue compared with normal tissue (49). GLO1 is also abundantly expressed in bladder and prostate cancer tissue (12, 42) and has been implicated in multidrug resistance in tumor cells (10). There has been no previous association of GLO1 with lactation physiology.

The milk proteins comprise a majority of the proteins expressed in lactating mammary gland (for review see Refs. 1 and 16), so clearly transcripts encoding the caseins would have the highest absolute expression in lactating mammary tissue. The caseins (αS1 and β) and the mammary cell-specific proteins of the whey fraction (lactalbumin and lactoglobulin) had greater relative expression in lactating mammary tissue compared with nonlactating tissue (Table 3). Because of the difference in fold change between animal 1 and animal 2 relative to nonlactating tissue (4.5 and 11.7, respectively, Table 3), αS1-casein had a P value for the combined data of 0.14 (Table 3). We acknowledge a limitation of the present experiment was biological variation from the limited sample size and heterogeneity in time of collection during lactation. However, for the individual microarray experiments, αS1-casein has the highest significance level (animal 1, FDR $P = 4.3 \times 10^{-105}$, animal 2, $P = 5.6 \times 10^{-90}$) for the genes that have increased expression in lactating tissue relative to nonlactating tissue. These expected results lend credence to the overall gene expression patterns observed in this microarray study.

To confirm gene expression data observed in lactating and nonlactating mammary tissues on the BMAM microarray, expression of six selected genes was tested with quantitative real-time PCR. All six had similar gene expression profiles in quantitative real-time PCR assays compared with microarray studies, although the level of expression varied between samples (Fig. 3). There was also variability in level of expression between lactating samples, although the trend was in agreement with microarray results (Fig. 3). This variability could be due to heterogeneity in cell type (since whole tissue samples were used), variation between animals (since samples were not strictly age matched), or to the outbred nature of animals used in this preliminary study. Regardless of the source of variability, it is clear that the trend is upregulation of CASP8, CD47, c-MYC, LASP1, PBK1, and VAV3 in lactating mammary gland vs. nonlactating mammary gland. Future work is needed to further elucidate gene expression patterns during the full course of mammary development and lactation cycles.

In summary, we have designed an EST and microarray resource enhanced for transcripts derived from bovine mammary tissue (BMAM). This resource should prove valuable in functional genomics studies of differential gene expression in the bovine mammary gland. We have preliminary gene expression patterns associated with mammary gland development and have identified 16 genes comprising a potential developing
mammary gland gene expression “signature.” These genes may prove to be useful targets in studies of hormonal and metabolic regulation of mammary development. Analysis of differential gene expression between nonlactating and lactating mammary tissue revealed the identities of several previously unknown genes that may play important roles in the bovine mammary gland during lactation. Further characterization of these genes and their protein products should improve our understanding of mechanisms important in mammary physiology. In addition, the BMAM microarray, combined with other bovine-specific microarrays developed by our group (66) should help advance in vivo studies aimed at shedding light on mechanisms that underlie diseases of the mammary gland, such as mastitis and cancer.

ACKNOWLEDGMENTS

We gratefully acknowledge the excellent assistance of Guilherme Rosa, Department of Animal Science, Michigan State University, in statistical analysis of microarray data. We also acknowledge the excellent support of Peter Saama, Department of Animal Science, Michigan State University, in maintaining the Center for Animal Functional Genomics website and database.

REFERENCES


