Identification of gene sets and pathways associated with lactation performance in mice

Jerry Wei, Palaniappan Ramanathan, Ian C. Martin, Christopher Moran, Rosanne M. Taylor, and Peter Williamson

Faculty of Veterinary Science, The University of Sydney, New South Wales, Australia

Submitted 20 September 2011; accepted in final form 1 January 2013

Wei J, Ramanathan P, Martin IC, Moran C, Taylor RM, Williamson P. Identification of gene sets and pathways associated with lactation performance in mice. Physiol Genomics 45: 171–181, 2013. First published January 2, 2013; doi:10.1152/physiolgenomics.00139.2011.—Mammary transcriptome analyses across the lactation cycle and transgenic animal studies have identified candidate genes for mammmogenesis, lactogenesis and involution; however, there is a lack of information on pathways that contribute to lactation performance. Previously we have shown significant differences in lactation performance, mammary gland histology, and gene expression profiles during lactation [lactation day 9 (L9)] between CBA/CaH (CBA) and the superior performing QS15 strains of mice. In the present study, we compared these strains at midpregnancy [pregnancy day 12 (P12)] and utilized these data along with data from a 14th generation of intercross (AIL) to develop an integrative analysis of lactation performance. Additional analysis by quantitative reverse transcription PCR examined the correlation between expression profiles of lactation candidate genes and lactation performance across six inbred strains of mice. The analysis demonstrated that the mammary epithelial content per unit area was similar between CBA and QS15 mice at P12, while differential expression was detected in 354 mammary genes (false discovery rate < 0.1). Gene ontology and functional annotation analyses showed that functional terms associated with cell division and proliferation were the most enriched in the differentially expressed genes between these two strains at P12. Further analysis revealed that genes associated with neuroreceptor ligand-receptor interaction and calcium signaling pathways were significantly upregulated and positively correlated with lactation performance, while genes associated with cell cycle and DNA replication pathways were downregulated and positively correlated with lactation performance. There was also a significant negative correlation between Grb10 expression and lactation performance. In summary, using an integrative genomic approach we have identified key genes and pathways associated with lactation performance.

gene sets; lactation performance; mammary transcriptome; microarray; qRT-PCR

LACTATION IS A COMPLEX PROCESS involving biosynthesis, secretory activation, and metabolic regulation. However, compared with the physiological events of mammary (alveolar) development, secretory differentiation during pregnancy, secretory activation at parturition, and involution (48, 56), genetically determined variation in lactation performance can be relatively subtle, and thus genetic variation in expression of the key genes for regulating milk production can be difficult to detect. A strategy employing inbred mouse strains to identify candidate genes and pathways associated with lactation performance has been developed in our laboratory (52). Comparison of gene expression profiles of lactating mammary glands between the high-performance QS15 strain and the relatively low-performance CBA/CaH strain suggested that Wnt, MAPK, and tight junction pathways contribute to the superior lactation phenotype in the QS15 mice. It has also been proposed that two imprinted genes, namely insulin-like growth factor 2 receptor (Igf2r) and growth factor receptor-bound protein 10 (Grb10) are negatively associated with variation in lactation phenotype (52). Grb10 functions as a negative regulator in insulin signaling (14, 40, 53) and the insulin-like growth factor (IGF) (21) pathways, and Igf2r is directly involved in binding IGF via the ligand internalization of M6P/IGF2R, which leads to the IGF2 degradation in lysosome (15, 16). The IGF signaling pathway is well known for its crucial roles in postnatal growth and metabolism. Transgenic mouse models have suggested that IGF1 not only modulates mammary development but also alters the lactation phenotype.

Multiple genes and pathways play key roles in postnatal mammary development, secretory activation, and mammary cell apoptosis, as indicated by integrative analysis of gene ontology (GO) and gene expression profiles of mammary tissue across the lactation cycle. For example, it has been suggested that the PI3K-AKT pathway is a significant regulator during pregnancy and secretory activation; the integrin pathway is also enriched during lactation, and the protein ubiquitination pathway is the most enriched pathway downregulated in both lactation and involution (37). The PI3K-AKT pathway plays a key role in cell cycle proliferation, protein synthesis, and glucose metabolism (72, 79). Integrins are cell surface receptors that interact with extracellular matrix and mediate intracellular signal transduction. The integrin pathways are involved in maintaining cell shape, cell migration, and proliferation (30). Ubiquitination is a posttranslational process that attaches ubiquitin to proteins marking them for destruction. It is a universal process for a broad range of proteins involved in cell proliferation, apoptosis, and other cell-specific metabolic pathways (9). Mechanisms underlying cell apoptosis are reflected in gene expression profiles during mammary gland involution and there is an indication of a significant role for immunological pathways (10, 11, 64, 65). However, most studies have focused on the function of genes at each stage of mammary development, and very few focused on the variation in lactation performance phenotypes.

The aim of the present study was to identify genes that may correlate with lactation performance. The investigation involved a comparative analysis of genome-wide expression profiles from the mammary glands of CBA/CaH and QS15 mice at day 12 of pregnancy (P12). Further analysis was also completed comparing the mammary transcriptome profiles of these strains and an intercross line derived from them. Addi-
tionally, quantitative RT-PCR profiling of selected lactation candidate genes across an extended genetic background of six inbred strains of mice was performed.

**MATERIALS AND METHODS**

**Animals.** Two inbred mouse strain models, namely the QS5 and CBA/CaH (CBA), and an advanced intercross line (AIL; CBA × QS5) were housed at the animal housing facility, Faculty of Veterinary Science, The University of Sydney, and were maintained as per protocol approved by the Animal Ethics Committee, The University of Sydney. The room temperature was kept at 21°C, and the mice were exposed to a daily photoperiod of 12 h (0600–1800). A pelleted diet (rat and mouse cubes; Specialty Feeds, Glen Forrest, WA, Australia) and water were available ad libitum. For mammary samples during pregnancy, five nulliparous female mice of QS5 and CBA strains were each mated with their male counterparts respectively. The appearance of a vaginal plug was recorded as day 0 of pregnancy, and the males were removed from this stage. Inguinal mammary glands were harvested at P12. Mammary tissue was immediately frozen in liquid nitrogen upon removal and stored at −80°C. For mammary samples collected during lactation, the AIL was produced by crossing the parental CBA and QS5 strains by a breeding protocol where each generation was sequentially produced by an intercrossing of the previous generation as described by Darvasi and Soller (12) and maintained for 14 generations (F14). The AIL mice from the F13/F14 were used in this study. The experimental protocols of assessing lactation performance and collecting mammary gland samples from the AIL dams at lactation day 9 (L9) were described previously (52). In brief, mice during the second lactation were used for the lactation study. The day of parturition was defined as lactation day 1 (L1). Litter size was standardized to six for the AIL dams to assess lactation performance of each animal, and the average pup weight gain for the first 8 days of lactation was used as the primary indicator (52). Inguinal mammary glands were collected on L9 and snap-frozen for further gene expression analyses.

**Histomorphometric analysis of mammary glands.** The left inguinal mammary glands were collected from five CBA mice and six QS5 mice at P12, respectively. Immediately the mammary gland was weighed and fixed in 10% neutral buffered formalin solution. Subsequently the formalin fixed samples were paraffin-embedded, sectioned, and mounted on the glass slides. Standard hematoxylin and eosin staining was performed to differentiate the mammary epithelial and interstitial area. The paraffin wax was removed from the sections, and the sections were rehydrated in water. The sections were stained in Whitlock’s hematoxylin for 3 min and washed in water to remove any excess stain. The sections were then incubated in Scott’s bluing agent for 2 min, washed, and rinsed in 70% ethanol. Subsequently the sections were counterstain in 1% eosin for 28 s and rapidly dehydrated with absolute ethanol. After dehydration, the sections were cleared and coverslipped. Two images were randomly captured for each sample. Percentage of mammary epithelial area (dark fields) was analyzed using the ImageJ image analysis tool (1, 54) (http://imagej.nih.gov/ij/) for the analysis between P12 and L9 with the leading-edge subset from the previous study (52). The average pup weight gain served as the primary indicator (52). Inguinal mammary glands were collected on L9 and snap-frozen for further gene expression analyses.

**RNA isolation and transcriptome profiling.** Total RNA isolation from the mammary tissue was performed with the RNaseasy Mini Kit (Qiagen, Victoria, Australia) as per manufacturer’s instructions. RNA quantity and integrity were assured by using the Agilent 2100 Bioanalyzer (Agilent Technologies) prior to hybridization. The biotin-labeled amplified RNA (aRNA) was generated and hybridized to an individual GeneChip Mouse Genome 430 2.0 Array, according to the manufacturer’s recommended procedures (Affymetrix; and AGRF, VIC, Australia). The minimum value of the RNA integrity number (RIN) (58) for the samples from the AIL mice was 9.2. Samples measured from lactation samples were also >7, but the RNA samples from the pregnant CBA and QS5 mice were more variable (mean 5.8 ± 1.3). It has been reported that the RIN can be variable in mammary tissue during periods of proliferative growth (25, 73). We therefore proceeded with samples on the basis of 28s/18s rRNA (ratios were >1.8; mean 2.1 ± 0.6).

The Affymetrix GeneChip Mouse Genome 430 2 array data (cel files) of mammary glands from lactating (L9) CBA and QS5 mice, respectively, derived from the previous study [available on the National Center for Biotechnology (NCBI) Gene Expression Omnibus (GEO) data repository, accession no. GSE9668] (52) and the expression data from mammary glands of the CBA and QS5 mice at P12 and of the AIL mice at L9 from the current study (available on the NCBI GEO data repository, accession no. GSE32020) were analyzed with affyليمGUI software package (77). Differentially expressed (DE) genes were identified using the limma package (61) incorporated in affyليمGUI. The data were background corrected, quantile normalized, and expression levels calculated using the robust multichip average algorithm (29). All packages used were available in the Bioconductor website (http://www.bioconductor.org/) (20) and implemented in the R software environment (http://www.r-project.org/).

**Functional annotation and gene set enrichment analyses.** Genes DE between the CBA and QS5 mice at P12 were defined using the linear model approach (61). We applied the Benjamini and Hochberg adjusted P value [false discovery rate (FDR)] cutoff of 0.1 to determine statistical significance (5). The DE genes were then analyzed for functional annotation enrichment using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov/) (13, 27, 28). The annotation categories for the DAVID analysis were set as default, the classification stringency was medium and the background was Affymetrix Mouse Genome 430 2 array. The EASE score representing the geometric mean of all the enrichment P values in each annotation term is used to determine the significance of results from the functional annotation clustering and gene functional classification analyses (28). An enrichment score (ES, minus log transformation of the EASE score) cut-off of 1.3 (EASE = 0.05) was applied in the current study.

We used the Gene Set Enrichment Analysis (GSEA) software (43, 67) for identification of the enriched gene sets DE between pregnancy and lactation regardless of strain differences. Microarray gene expression data from the five CBA and five QS5 mice at P12 were grouped as the query gene expression dataset, and data from five CBA and five QS5 at L9 (52) were grouped as the lactation set for the GSEA using the Molecular Signatures Database (MSigDB v2.5) (67). Gene sets from the four major collections, namely curated, motif, computational, and GO gene sets in MSigDB were used in the analysis. To statistically rank the ES, the permutation type algorithm was “gene_set” and the number of permutation was 1,000 with the “Signal2Noise” metric. Applying a FDR cut-off of 0.05, we further analyzed the top-ranked upregulated and downregulated gene sets at L9 using “leading-edge analysis” for the leading-edge gene subset.

We used the GSEA software with MSigDB v2.5 to analyze mammary gene expression and lactation performance phenotypic data of the 10 AIL mice at L9 from the current study, incorporating the microarray and phenotypic data of the CBA and QS5 at L9 derived from the previous study (52). The average pup weight gain served as the primary indicator of lactation performance, microarray data were together uploaded to the GSEA program, and the Pearson metric was used for the analysis. A total of 3,906 gene sets from the four major collections, namely, “curated,” “motif,” “computational,” and “GO” datasets in MSigDB, were used in the analysis. To assess the statistical significance of the ES, the permutation type algorithm was gene_set, the number of permutation was 1,000, and FDR was set at 0.05. Top-ranked gene sets were analyzed by the leading-edge analysis function for identifying the leading-edge gene subset associated with lactation performance.

We then compared the leading-edge subset identified from the analysis between P12 and L9 with the leading-edge subset from the correlation analysis with lactation performance at L9. The leading-
edge genes were categorized into four groups according to their expression profiles that satisfied either one of two in each criterion from both analyses: 1) up- or downregulated during lactation compared with pregnancy, and 2) positively or negatively correlated with lactation performance. Genes that were upregulated during lactation and positively correlated with lactation performance were categorized as group 1 (G1); genes that were downregulated during lactation and negatively correlated with lactation performance were categorized as group 2 (G2); genes that were upregulated during lactation and negatively correlated with lactation performance were categorized as group 3 (G3); and genes that were downregulated during lactation and negatively correlated with lactation performance were categorized as group 4 (G4). The leading-edge genes in each group were analyzed for functional annotation enrichment by DAVID with the aforementioned parameters.

Quantitative reverse transcriptase PCR. Primer sequences for the mouse Neo1 (ID: 6679036a1), Igf2r (ID: 2395605a1), Grb10 (ID: 16359305a2), and Ppia (ID: 6679439a1) genes were obtained from Primer Bank (http://pga.mgh.harvard.edu/primerbank/index.html) (62, 75) and manufactured by Sigma Genosys (NSW, Australia). Details of animal housing and experimental design were described previously (76). In brief, the inguinal mammary gland samples were collected from the six inbred strains of mice, namely QSi5, FVB/N (FVB), AKR, C57BL/6J (C57), CBA, and DBA/1J, at L9. RNA isolation followed by first-strand cDNA synthesis were completed using the RNeasy Mini Kit (Qiagen, Victoria, Australia) and MMLV reverse transcriptase (Invitrogen, NSW, Australia), respectively, following the instructions provided by the manufacturers. All RNA samples were checked for purity with a 260 nm/280 nm ratio >1.8. Five cDNA samples from each strain were analyzed. Quantitative reverse transcription PCR (qRT-PCR) was performed under the following conditions: 10 μl reaction contained 10 ng cDNA template, 3 mM MgCl2, 0.2 mM dNTP, 20 ng of each primer, Syto9 2 μM (Invitrogen), Platinum Taq DNA polymerase 0.2 U with 1× PCR buffer (Invitrogen) using a Rotor Gene 6000 system (Corbett Research, NSW, Australia). CDNA samples were denatured at 95°C for 5 min, followed by a 40 cycle-PCR of 95°C for 20 s, 65–67°C (touchdown by 1°C each cycle for the first 8 cycles) for 20 s and 72°C for 30 s. Upon completion of 40 cycles the samples were held at 72°C for 5 min and followed by melting-curve analysis. Samples of five animals from each strain were examined. Several commonly used housekeeping genes including Actb, Gaddph, Ppia, Rn18S, and Sdha for qPCR were evaluated, and the peptidylprolyl isomerase A (Ppia; cyclophilin A) were selected as Actb, Gapdh, Ppia, Rn18S, and Sdha for qPCR were evaluated, and the peptidylprolyl isomerase A (Ppia; cyclophilin A) were selected as candidates for qPCR analysis. The melting-curve analysis of representative histological sections estimated the adipose and epithelial cell content at P12. Quantitative image analysis of representative histological sections estimated the adipose and epithelial cell content at P12. Quan-

titative image analysis of representative histological sections from mammary gland tissue showed that >90% of cells in the mammary gland were adipocytes in both the CBA (Fig. 1, A1 and A2) and the QSi5 (Fig. 1, B1 and B2) mice at P12. This is consistent with previous observations that adipocytes constitute the majority of cells in the mammary gland during pregnancy (2, 76). The average proportion of mammary gland epithelial content in CBA mice was 9.97 ± 0.97% and 8.15 ± 0.87% in QSi5 mice (P = 0.198) (Fig. 1C). These results also confirm that the mammary tissue samples at P12 used for microarray analysis were consistent with regard to cellular content.

Genes DE between strains at P12. In a previous study to identify genes associated with the superior lactation performance of QSi5 mice, we compared gene expression profiles in the lactating mammary gland between CBA and QSi5 mice and tentatively identified several candidate genes and pathways (52). To characterize whether there are differences in mammary gene expression during pregnancy between CBA and QSi5, we performed a microarray analysis of mammary gland gene expression at P12. Using the linear model approach (61) and an FDR cut-off of 0.1 (P = 0.000985), we identified a total of 444 probe IDs (representing 354 annotated genes) as differentially expressed between these two strains. This comprised a total of 274 probe IDs that were expressed at a relatively higher level in QSi5 mice, and 170 probe IDs that were expressed at a relatively higher level in the CBA mice (Supplementary Data S1).1 These DE probe IDs were then analyzed for functional annotation clustering and gene function classification using DAVID and an ES cut-off of 1.3 (28). Results of the functional annotation clustering analysis showed that, within the genes expressed in significant higher levels in the QSi5 mammary gland vs. in the CBA, four functional annotation clusters had scores of >1.3. Two of these are associated with cell cycle and mitosis (ES 1.76) and cell cycle and meiosis (ES 1.58), and the other two are associated with immune response, and antigen processing and presentation (ES 1.66 and 1.60) (Supplementary Data S2). In contrast, only one enriched functional annotation cluster in association with immunoglobulin and major histocompatibility complex was identified (ES 1.75) (Supplementary Data S3) within the DE genes expressed at lower levels in QSi5 mice vs. in CBA.

Results of the gene function classification analysis identified two groups (ES >1.3) within the genes expressed at higher levels in QSi5 vs. in CBA (Table 1, Supplementary Data S4), and this included a total of eight genes. The eight functionally significant genes are all associated with cell cycle. No groups within the genes expressed in higher levels in CBA vs. QSi5 were identified to have an ES of >1.3 (Supplementary Data S5).

Genes DE between pregnancy and lactation. Significant changes in gene expression levels in the mammary gland during pregnancy and lactation are likely to reflect the molecular pathways that play an important role in milk synthesis and secretion and, hence, lactation performance. We first examined microarray expression profiles of the known lactation genes (as reported by the Gene Ontology ID GO:0007595) (7), milk proteins, and genes associated with apical membrane calcium channels (70). Using the linear model approach (61), we found that the majority of these genes were significantly upregulated during lactation (FDR < 0.05) (Supplementary Table S1).

To identify the genes that are DE between pregnancy and lactation, we grouped the mammary microarray data based on the developmental stage (pregnancy or lactation) regardless of

1 The online version of this article contains supplemental material.
strain and conducted GSEA. The GSEA program ranks the genes and gene sets by calculating the ES and normalized enrichment score (NES) (67). The higher NES of the gene set indicates that expression levels of a higher proportion of the genes in the gene set are positively associated with the phenotype, whereas a low NES indicates more genes in the gene sets are negatively associated. This analysis revealed 2,211 gene sets (57% of the total selected gene sets as input) from MSigDB that were differentially expressed between P12 and L9 (FDR < 0.05), 85 gene sets (2.18%) were upregulated (Supplementary Data S6), and 2,126 gene sets (54%) were downregulated (Supplementary Data S7) at L9 relative to P12. Similar to the results of other studies (19, 36), only a small fraction of genes in the mammary transcriptome were upregulated during lactation compared with pregnancy. In this analysis, the top-ranked upregulated gene sets were categorized in relation to ion channel activities, G protein receptor activities, and neurotransmitter (Supplementary Data S6), whereas the top-ranked downregulated gene sets were associated with nucleic acid processing and cell cycle/cyclin activities (Supplementary Data S7). Similar observations have been made in a study of bovine lactation compared with pregnancy in which significant enrichment of cell proliferation-related functions was observed in the list of downregulated genes in early lactation compared with end of pregnancy, and a significant enrichment of signaling-related functions was observed in the list of genes upregulated (19). We also used the linear model approach to identify genes that are DE between pregnancy and lactation. In this case we escalated the degree of differential expression while maintaining statistical scores as before from

Fig. 1. Quantitative image analysis of mammary epithelial area at pregnancy day 12 (P12). A1 and A2: CBA. B1 and B2: QSi5. C: boxplot of mammary epithelial area in the CBA (n = 5) and the QSi5 (n = 6) mice. Ep, epithelial cells; F, adipose cells. The mean value of the mammary epithelial area was 9.97 ± 0.97% in the CBA and 8.15 ± 0.87% in the QSi5 mice at P12 (P = 0.198). Bars indicate 95% confidence intervals. Horizontal line within the box indicates the median value.
DAVID for functional annotation enrichment. Using a minimum fourfold increase during lactation, in an approach similar to that of Bionaz et al. (6), we found functional annotation clusters to be associated with protein posttranslational modification (disulfide bond and glycosylation) (ES  = 6.4), cation/metal ion transport (ES  = 2.03), fatty acid biosynthesis (ES  = 1.96), netrin domain (ES  = 1.93), whey acid protein glycoprotein biosynthetic and metabolic process/Golgi apparatus (ES  = 1.35). These findings were similar to the previous studies (2, 6, 57, 70) and represent a group of annotated genes that are consistent with changes that occur in the mammary gland during the lactation cycle.

**Gene expression profiles and pup weight gain.** Lactation performance is a quantitative trait in the variation in the genetic component attributable to many genes. We used the pup weight gain data from individual dams of CBA, QS15, and AIL litters as the phenotypic data for GSEA to identify genes that correlated with lactation performance. The analysis shows that only 10 gene sets (0.3% of total selected gene sets from MSigDB) were positively associated with pup weight gain (Supplementary Data S8), whereas 953 gene sets (31% of total selected gene sets from MSigDB) were negatively associated with pup weight gain (FDR < 0.05) (Supplementary Data S9). The top positively ranked genes were involved with ion channel activity, while the top negatively ranked genes were associated with RNA processing and cell proliferation.

**Genes associated with lactation performance.** Our purpose was to identify genes associated with lactation performance and to group them according to patterns of expression as a means of subdividing them on the potential for functional similarity. The leading-edge genes from L9 that were correlated with lactation performance were extracted, and the expression patterns of these genes between pregnancy and lactation were used to group them. This resulted in four subsets: upregulated at L9 and positively correlated with lactation performance (G1); downregulated at L9 and positively correlated with lactation performance (G2); upregulated at L9 and negatively correlated with lactation performance (G3) and downregulated at L9 and negatively correlated with lactation performance (G4). There were 121 genes in G1, 35 genes in G2, 23 genes in G3, and 1,256 genes in G4 (Supplementary Data S10).

To further classify these genes, we analyzed each group for GO terms using statistical algorithms in DAVID. A stringent threshold value was applied for all biological terms: FDR < 0.001. Three functional annotation databases were used for the analysis, including GO_molecular_function (GO MF), GO_biological_process (GO BP) (4), and KEGG_pathway (KEGG) categories (http://www.kegg.jp/) (32). The three most significant GO MF annotations in G1 were “ion channel activity,” “substrate specific channel activity,” and “calcium activity,” while “G protein-coupled receptor protein signaling pathway,” “ion transport,” and “cell surface receptor-linked signal transduction” were the most significant GO BP terms. Two KEGG pathway categories passed the statistical threshold: “neuroactive ligand-receptor interaction” and “calcium signaling pathway.” Among the G2 gene group, the three most significant GO MF annotations were “ATP binding,” “adenyl ribonucleotide binding,” and “adenyl nucleotide binding.” The top-ranked GO BP terms were “cell cycle,” “M phase,” and “cell cycle process,” and the corresponding KEGG pathway categories were “DNA replication” and “cell cycle.” In the G3 gene group, no statistically significant GO MF, GO BP, or KEGG pathway terms were identified, whereas in the G4 group, the three highest ranked GO MF terms were “RNA binding,” “nucleotide binding,” and “RNA polymerase II carboxy-terminal domain kinase activity,” while the most significant GO BP terms were “RNA processing,” “mRNA metabolic process,” and “mRNA metabolic process;” and significant KEGG categories were “spliceosome,” “proteasome,” and “cell cycle.” A summary of these ontology categories is presented in Table 2.

**Grb10 gene expression levels are negatively correlated with lactation performance.** Previously, neogenin (Neo1), Igf2r, and Grb10 have been selected as strong candidate genes for lactation performance (52). In the current study, analysis of the microarray mammary gene expression data of the CBA and QS15 mice at P12 revealed that levels of Neo1 were higher (1.7-fold, P < 0.001) in the QS15 compared with the CBA females, whereas levels of Igf2r were lower (0.6-fold, P < 0.001) in QS15, and levels of Grb10 similar between these two strains. However, at L9, expression levels of Neo1 remained higher (2.7-fold, P < 0.001) in QS15 compared with CBA, and levels of Igf2r and Grb10 were lower in QS15 (0.6- and 0.3-fold respectively; P < 0.001), which is consistent with the previous observations (52). Further examination of microarray gene expression profiles in L9 mammary glands from CBA, QS15, and the AIL mice at L9 revealed that the expression of both Grb10 and Igf2r were negatively correlated with lactation performance (R = -0.56, P = 0.025 and R = -0.57, P =

### Table 1. Significant gene groups from the gene functional classification analysis (DAVID); higher levels in QS15

<table>
<thead>
<tr>
<th>Affymetrix_3prime_IVT_ID</th>
<th>Gene Name</th>
<th>Enrichment Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene Group 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1439394_s_at, 1416664_at</td>
<td>centrosomal protein 55; Official Symbol: Cep55</td>
<td>1.55</td>
</tr>
<tr>
<td>1421963_a_at</td>
<td>shagoshin-like 1 (S. pombe); Official Symbol: Sgo1</td>
<td>1.40</td>
</tr>
<tr>
<td>1439510_at</td>
<td>predicted gene 8416; predicted gene 5593; cyclin B1; similar to cyclin B1; predicted gene 4870; Official Symbol: Ccnb1-rs1, Ccnb1</td>
<td>1.96</td>
</tr>
<tr>
<td>1448205_at, 1419943_s_at</td>
<td>intraflagellar transport 20 homolog (Chlamydomonas); Official Symbol: Ift20</td>
<td>1.93</td>
</tr>
<tr>
<td>1452242_at</td>
<td>cell division cycle 20 homolog (S. cerevisiae); Official Symbol: Cdc20</td>
<td>1.55</td>
</tr>
<tr>
<td>1452242_at</td>
<td>centrosomal protein 55; Official Symbol: Cep55</td>
<td>1.55</td>
</tr>
<tr>
<td>1439403_s_at</td>
<td>cell division cycle 25 homolog B (S. pombe); Official Symbol: Cdc25b</td>
<td>1.40</td>
</tr>
<tr>
<td>1451412_a_at</td>
<td>netrin domain (ES  = 1.93), whey acid protein glycoprotein biosynthetic and metabolic process/Golgi apparatus (ES  = 1.35)</td>
<td>1.35</td>
</tr>
<tr>
<td>1450900_at</td>
<td>cation/metal ion transport (ES  = 2.03), fatty acid biosynthesis (ES  = 1.96)</td>
<td>1.96</td>
</tr>
<tr>
<td>1451412_a_at</td>
<td>rho1 GTPase activating protein 1, suppressor of raf1; Reduced Dictyostelium discoideum; Official Symbol: Rho1</td>
<td>1.93</td>
</tr>
</tbody>
</table>
| 1434023_at               | “neuroactive ligand-receptor interaction” and “calcium signaling pathway.” Among the G2 gene group, the three most significant GO MF annotations were “ATP binding,” “adenyl ribonucleotide binding,” and “adenyl nucleotide binding.” The top-ranked GO BP terms were “cell cycle,” “M phase,” and “cell cycle process,” and the corresponding KEGG pathway categories were “DNA replication” and “cell cycle.” In the G3 gene group, no statistically significant GO MF, GO BP, or KEGG pathway terms were identified, whereas in the G4 group, the three highest ranked GO MF terms were “RNA binding,” “nucleotide binding,” and “RNA polymerase II carboxy-terminal domain kinase activity,” while the most significant GO BP terms were “RNA processing,” “mRNA metabolic process,” and “mRNA metabolic process;” and significant KEGG categories were “spliceosome,” “proteasome,” and “cell cycle.” A summary of these ontology categories is presented in Table 2.

Grb10 gene expression levels are negatively correlated with lactation performance. Previously, neogenin (Neo1), Igf2r, and Grb10 have been selected as strong candidate genes for lactation performance (52). In the current study, analysis of the microarray mammary gene expression data of the CBA and QS15 mice at P12 revealed that levels of Neo1 were higher (1.7-fold, P < 0.001) in the QS15 compared with the CBA females, whereas levels of Igf2r were lower (0.6-fold, P < 0.001) in QS15, and levels of Grb10 similar between these two strains. However, at L9, expression levels of Neo1 remained higher (2.7-fold, P < 0.001) in QS15 compared with CBA, and levels of Igf2r and Grb10 were lower in QS15 (0.6- and 0.3-fold respectively; P < 0.001), which is consistent with the previous observations (52). Further examination of microarray gene expression profiles in L9 mammary glands from CBA, QS15, and the AIL mice at L9 revealed that the expression of both Grb10 and Igf2r were negatively correlated with lactation performance (R = -0.56, P = 0.025 and R = -0.57, P =
strains of mice, there was a negative correlation with the analyzed for association with lactation performance in the six P

clear statistical difference (a greater variation in expression of Neo1 (Fig. 2 A) with no

and Grb10 (Fig. 2 B) between the strains (both

KEGG pathway

GO_biological process

GO:0007166 ~ cell surface receptor-linked signal transduction

66 55.93 5.61E-35 5.58 8.56E-32

KEGG pathway

mumu04020: Calcium signaling pathway 25 21.19 2.36E-88 17.37 1.92E-85

G2 (Downregulated at L9 and positively correlated with lactation performance)

GO_molecular function

GO:0005524 ~ ATP binding

16 48.48 2.00E-07 4.42 2.22E-04

GO:0032559 ~ adenyly ribonucleotide binding

16 48.48 2.33E-07 4.37 2.59E-04

GO:0030554 ~ adenylnucleotide binding

16 48.48 4.14E-07 4.16 4.89E-04

GO_biological process

GO:0007049 ~ cell cycle

22 66.67 2.75E-20 13.03 3.70E-12

GO:0000279 ~ M phase

15 45.45 3.77E-15 19.11 5.07E-12

GO:0022402 ~ cell cycle process

16 48.48 1.37E-14 14.66 1.85E-11

KEGG pathway

mumu03030: DNA replication

5 15.15 2.82E-06 43.19 0.002078

mumu04110: Cell cycle

6 18.18 2.23E-05 14.75 0.016413

G4 (Downregulated at L9 and positively correlated with lactation performance)

GO_molecular function

GO:0003723 ~ RNA binding

113 9.80 3.82E-17 2.27 5.90E-14

GO:0000166 ~ nucleotide binding

259 22.46 1.44E-13 1.51 2.22E-10

False discovery rate (FDR) <0.05. GO, gene ontology; L, lactation day; G, group.

0.02, respectively). Moreover, gene coexpression pattern analysis showed that Grb10 was classified in the G4 coexpression group and Igf2r was in the G3.

To further investigate the relationship between lactation performance and gene expression levels of Neo1, Grb10, and Igf2r, qRT-PCR was utilized to measure expression levels in mammary tissue across six inbred mouse strains at L9, namely QSi5, FVB, AKR, C57, CBA, and DBA1/J. The results revealed significant differences in the levels of Igf2r (Fig. 2 A) and Grb10 (Fig. 2 C) between the strains (both P < 0.001) but a greater variation in expression of Neo1 (Fig. 2 A) with no clear statistical difference (P = 0.106). However, a 3.3-fold greater level of Neo1 in QSi5 mice was observed compared with CBA mice (P = 0.008) (Fig. 2A), which is similar to the results from the microarray analysis. In addition, expression levels of Grb10 were 2.8-fold higher in the CBA mice compared with the QSi5 (P = 0.006) (Fig. 2C), and no difference in the Igf2r expression levels was observed between the CBA and QSi5 mice (P = 0.914). When Igf2r and Grb10 were analyzed for association with lactation performance in the six strains of mice, there was a negative correlation with the expression levels of Grb10 (R = 0.84, P = 0.036) (Fig. 3). We should note that all of the results presented here for qRT-PCR are based on comparative levels relative to one housekeeping gene, Ppia.

**DISCUSSION**

In the resting mammary gland, adipocytes comprise the majority of cells in the gland (47). During pregnancy, the mammary gland undergoes rapid growth of lobuloalveolar structure and size. Wang et al. (74) reported that in FVB mice 60% of the mammary tissue was epithelium. In the current study, morphometric analysis of the percentage of the mammary epithelial area revealed that during midpregnancy, the mammary gland still consists of a large proportion of adipose tissue, and no significant difference in the mammary epithelium per gram of tissue between the CBA and QSi5 mice at P12 (~10%). Similar percentages of the epithelial area of the mammary gland in both strains at P12 indicate that similar proportions of the epithelial cells and the adipocytes were sampled for the mammary transcriptome analysis, which would minimize variations in the expression profiles due to heterogeneity of the cellular components in the mammary gland during pregnancy. At this stage, in which mammary cells proliferate exponentially (41, 66), although the percentage of mammary epithelial area in both mouse strains was similar, the ongoing molecular events at the cellular level were markedly different. A previous microarray study of gene expression of the mammary gland development over the lactation cycle (10) has identified that genes associated with cell cycle and mitosis were two of the significant functional clusters that are upregulated only during midpregnancy. In the current study, mRNA profiling of the mammary gland of both strains at P12 showed that genes associated with cell cycle and mitosis were significantly enriched within the genes expressed at higher levels in the QSi5 mice. The gene functional classification analysis identified a total of eight genes associated with cell cycle. Previous findings revealed that Cdc20 (38) and Sgo1l (33) are
involved in the ubiquitin degradation process and formation of the anaphase promoting complex during cell division. The phosphatase Cdc25b plays a key role in meiosis (39), and the protein encoded by Ccnb1 is involved in regulation of mitosis (42). Cep55 (17), Cep120 (78), and Smek2 (8) participate in centrosome organization during cell division. The Ift20 protein has been implicated in coupling extracellular and proliferative events in some ductal cells (31). The results suggest a greater level of cell growth and division in mammary glands of QSi5 mice at midpregnancy, which is consistent with larger glands and greater productive capacity in these mice compared with the CBA strain (52).

One of the advantages of using GSEA for analysis is its capacity to process data derived from independent studies and across different platforms (67). This approach allowed a combined analysis of microarray data from mammary glands of CBA, AIL, and QSi5 mice and identification of groups of genes that are associated with lactation performance. Initially we used the linear model approach to identify genes that are differentially expressed between pregnancy and lactation. When this approach was combined with analysis of those genes that are most highly variable between pregnancy and lactation, annotation analysis characterized the group into well-recognized pathways associated with metabolism, protein translation and posttranslational modification, lipid biosynthesis, ion transport, and whey acidic protein. These findings are consistent across a number of studies that have examined gene expression during lactation cycle (2, 6, 57, 70). These profiles appear to be consistent even though tissue heterogeneity can cause problems in the longitude transcriptome expression analysis of mammary tissue. A previous study by Wang et al. (74) indicates that expression data of the heterogeneous tissues including nonlactating mammary tissue may be adjusted according to the proportion of the cellular components prior to ontology analysis. Several computation methods have been developed to overcome this issue (69, 70), and there are still issues with the computational methodology and modeling (69, 73). In the current study, we utilized the GSEA to identify the leading-edge genes within the expression profiles that were primarily correlated with lactation performance. By applying this criteria we selected away from exploring those genes that may have been most affected by tissue heterogeneity and focused on any genes that may be considered to be correlated with lactation performance, regardless of their origin and independent of mechanistic considerations.

Fig. 2. Boxplots of qRT-PCR expression profiles of Neo1 (A), Igf2r (B), Grb10 (C) in the lactating mammary gland tissue of the QSi5, FVB, AKR, C57, CBA, and DBA/1J mice. ANOVA tests for between strain variation revealed significant differences in the expression of Igf2r ($P < 0.001$) and Grb10 ($P < 0.001$) but not in the expression of Neo1 ($P = 0.106$). Bars indicate 95% confidence intervals. Horizontal line within the box indicates the median value. Circle with a cross indicates the mean value.

Fig. 3. The negative correlation between the strain mean values of lactation performance and Grb10 expression levels. (Pearson correlation coefficient $R = -0.84$, $P = 0.036$.)
Genes expressed in the mammary gland during lactation were considered to be influential in milk production and secretion and compared with genes that were correlated with milk yield as measured by pup growth. Genes that satisfied both criteria were regarded as genes that potentially contributed to lactation performance. The G protein-coupled receptors (GPCRs) of the “neuroactive ligand-receptor interaction” pathway in the KEGG database was identified as the most significantly enriched functional group within the G1 genes. In the study by Clarkson et al. (10), expression of genes associated with G protein signaling was significantly upregulated during the midlactation period compared with pregnancy and involution. This pathway is fundamental to many developmental processes and physiological activities. The GPCR family is a large transmembrane protein receptor family that responds to a range of stimuli to activate signal transduction pathways and, ultimately, cellular and physiological responses. A number of molecules belonging to the GPCR family have also been demonstrated to control mammary development. An example taken from within the candidate gene list is that of the glycoprotein hormones, alpha polypeptide (CGA). CGA is the common alpha subunit of choric gonadotropin (CG), luteinizing hormone, follicle stimulating hormone, and thyroid stimulating hormone. It is well documented that these four glycoprotein hormones play key roles in reproduction, metabolism, and body growth as well as mammary gland development (reviewed in Refs. 46, 49, 68). Another example from within the candidate gene list are that of the galanin receptor (Galar)1 and 3, being receptors of the neuropeptide galanin (Gal). Gal and receptors are differentially expressed in mammary gland during lactation compared with pregnancy and involution (45). A deficiency in Gal in mice leads to abnormal mammary lobuloalveolar development and eventually diminished lactation (45).

Another significant pathway identified in the G1 gene group was the calcium signaling pathway. In the study by Lemay et al. (37), three major gene coexpression patterns in the mammary gland across the lactation cycle were identified. Calcium-mediated signaling and ion channel activity were among the most significant functional terms identified in a group of genes with the expression profiles belonging to one of the major coexpressed patterns in which genes were upregulated from midpregnancy, reaching the highest level at parturition and across the lactation period, and then downregulated at the onset of involution. Calcium is extensively transported to the lactating mammary gland as a constituent of the secreted milk. Calcium also acts as a second messenger in the GPCR signal transduction, which mediates downstream pathways. Although the transport of calcium in mammary epithelial cells has previously been studied (3, 70), and the mechanisms underlying calcium transport in the lactating mammary gland are elucidated, the level of ionized calcium and its regulation in cytosol remain unclear (59). One of the G1 genes in the calcium signaling pathway, ryanodine receptor 3 (Ryr3) is an intracellular calcium release channel. Ryr3 was initially discovered in brain tissue and was later also found to be expressed in a range of tissues (44). The expression profile of Ryr3 was notable in the current study, and whereas previous studies of Ryr3 have focused on its roles in neurotransmission and muscle contraction (reviewed in Ref. 81), there is no information about a potential role for Ryr3 in the mammary gland.

Lactation performance in mice may be determined by both the size and milk production efficiency of the mammary gland (51, 55). During the lactation period the main cellular activities in the mammary gland shift from proliferation to biosynthesis, metabolism, and secretion. This was reflected in the present study, with genes involved in the cell cycle, the most significant pathway affected, and overall, genes associated with cell proliferation were downregulated between pregnancy and lactation. The finding in the current study is consistent with a switch in emphasis on the increase in mammary gland mass during midpregnancy toward milk production in the lactating gland and is supported by similar gene expression changes observed in other studies (10, 37). Interestingly, expression of genes involved in M-phase and DNA replication during the cell cycle was positively correlated with lactation performance. One possible explanation of this relates to mammary gland development. This may reflect a higher DNA replication and mitotic activity of mammary cells in QSi5 mice compared with CBA mice at P12, which would produce the greater volume of mammary gland tissue that is observed in QSi5 mice (55) and hence an increased lactation capacity.

By a combined in silico approach, a number of potential candidate genes for lactation performance have been identified. The protein encoded by the histone acetyltransferase 1 (HAT1) gene, classified in G4, is an example from these candidates that may be validated by comparison to previous studies. This molecule has been postulated as one of the 27 major hubs of the comprehensive lactation network in a previous study (37). The hub proteins represent the key nodes for connecting other proteins within the lactation regulatory network, and hence they tend to be highly conserved (22). HAT1 has an important function as an epigenetic regulator in chromatin assembly (71, 80), and although there is a lack of specific information about the role of HAT1 in mammary development, epigenetic modification is inhibited during lactation (6). HAT1 expression in the present study was negatively correlated with lactation performance, and this gene may be functionally significant in the regulation of milk production.

Microarray gene expression profiles from the mammary glands of lactating CBA and QSi5 mice have shown that Neo1, Igf2r, and Grb10 are DE between these two strains of mice, and they have been proposed as potential lactation performance associated genes in a previous study (52). Neo1 was originally discovered in neural cells and also plays a crucial role in mammary development (63). Grb10 and Igf2r are imprinted genes and act as negative regulators in the insulin and IGF signaling pathways (35, 40). In the present study, quantitative RT-PCR profiling further confirmed a negative correlation between levels of Grb10 gene expression and lactation performance, with a negative trend between the level of Igf2r expression and lactation performance amongst the six strains of mice. However, no significant difference in Igf2r expression as determined by qRT-PCR was observed between CBA and QSi5 mice at L9. In each case these results reflect a relative change compared with one housekeeping gene, Ppia, and this may therefore limit interpretation of the data. In previous studies, IGFBP5, one of the negative regulators of IGFR signaling has been proposed as a key regulator of mammary gland events during lactation and involution (10, 57). The IGF signaling pathway has been demonstrated to play a central role in embryonic and postnatal mammary development, particularly in modulating mammary cell proliferation and differentiation, and as an inhibitor of apoptosis (23, 26). IGF-1 transgenic animal models have demonstrated a
positive effect of the IGF signaling pathway on maternal lactation performance (24). In the normal, lactating mammary gland, ex- cessive IGF-I and IGF-2 levels leads to enhanced lactation capacity (50). Moreover, Grb10 also binds to insulin receptor (IR) to induce IR degradation, which subsequently inhibits the insulin signaling pathway (40, 53). Studies have shown that the insulin signaling pathway is critically important in mammary gland development and milk synthesis in different species (7, 25, 48, 60, 68). However, expression levels of those genes directly involved in the IGF and insulin pathways were not correlated with lactation performance in the current study, rather it was the inhibitors that were significant. This implies that inhibition of the IGF signaling pathway may have a negative impact on lactation performance.

In summary, by using mouse strain comparison and close examination of mammary gland transcriptome profiling, we provide a set of pathways and functional ontologies associated with lactation performance. These pathways, and the key regulatory genes that comprise the pathways, provide a basis for investigating molecular mechanisms underlying variation in lactation performance.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: J.W., P.R., and P.W. conception and design of research; J.W. and P.R. performed experiments; J.W., P.R., and P.W. analyzed data; J.W., P.R., and P.W. interpreted results of experiments; J.W. prepared figures; J.W. drafted manuscript; J.W., P.R., I.C.M., R.M.T., and P.W. edited and revised manuscript; P.W. approved final version of manuscript.

REFERENCES


70. Wang H, Owens JD, Shih JH, Li MC, Bonner RF, Mushinski JF. Histological staining methods preparatory to laser capture microdissection significantly affect the integrity of the cellular RNA. *BMC Genomics* 7: 9, 2006.


