

## Research Article

# Identification of Putative Ortholog Gene Blocks Involved in Gestant and Lactating Mammary Gland Development: A Rodent Cross-Species Microarray Transcriptomics Approach

**Maricela Rodríguez-Cruz,<sup>1</sup> Ramón M. Coral-Vázquez,<sup>2,3</sup>  
Gabriel Hernández-Stengele,<sup>1,2,3</sup> Raúl Sánchez,<sup>1</sup> Emmanuel Salazar,<sup>4</sup>  
Fausto Sanchez-Muñoz,<sup>5</sup> Sergio Encarnación-Guevara,<sup>4</sup> and Jorge Ramírez-Salcedo<sup>6</sup>**

<sup>1</sup> Unit of Medical Research in Nutrition, Hospital de Pediatría, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Avenida Cuauhtémoc 330, Col. Doctores, Delegación Cuauhtémoc, 06725 Mexico City, Mexico

<sup>2</sup> Sección de Estudios de Posgrado e Investigación, Escuela Superior de Medicina, Instituto Politécnico Nacional, Plan de San Luis y Díaz Mirón s/n, Col. Casco de Santo Tomas, Delegación Miguel Hidalgo, 11340 Mexico City, Mexico

<sup>3</sup> Subdirección de Enseñanza e Investigación, Centro Médico Nacional 20 de Noviembre, Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado, San Lorenzo 502 (2 piso), Col. Del Valle, Delegación Benito Juárez, 03100 Mexico City, Mexico

<sup>4</sup> Genomic Sciences Center, Universidad Nacional Autónoma de México, Avenida Universidad s/n, Col. Chamilpa, 62210 Cuernavaca, Morelos, Mexico

<sup>5</sup> Department of Immunology, Instituto Nacional de Cardiología Ignacio Chávez, Juan Badiano No. 1 Col. Sección XVI, 140080 Tlalpan, Mexico City, DF, Mexico

<sup>6</sup> Microarray Unit, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Delegación Coyoacán, 04510 Mexico City, Mexico

Correspondence should be addressed to Maricela Rodríguez-Cruz; [maricela.rodriguez.cruz@gmail.com](mailto:maricela.rodriguez.cruz@gmail.com) and Ramón M. Coral-Vázquez; [rmcoralv@prodigy.net.mx](mailto:rmcoralv@prodigy.net.mx)

Received 21 May 2013; Revised 3 September 2013; Accepted 4 September 2013

Academic Editor: Elena Pasyukova

Copyright © 2013 Maricela Rodríguez-Cruz et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The mammary gland (MG) undergoes functional and metabolic changes during the transition from pregnancy to lactation, possibly by regulation of conserved genes. The objective was to elucidate orthologous genes, chromosome clusters and putative conserved transcriptional modules during MG development. We analyzed expression of 22,000 transcripts using murine microarrays and RNA samples of MG from virgin, pregnant, and lactating rats by cross-species hybridization. We identified 521 transcripts differentially expressed; upregulated in early (78%) and midpregnancy (89%) and early lactation (64%), but downregulated in mid-lactation (61%). Putative orthologous genes were identified. We mapped the altered genes to orthologous chromosomal locations in human and mouse. Eighteen sets of conserved genes associated with key cellular functions were revealed and conserved transcription factor binding site search entailed possible coregulation among all eight block sets of genes. This study demonstrates that the use of heterologous array hybridization for screening of orthologous gene expression from rat revealed sets of conserved genes arranged in chromosomal order implicated in signaling pathways and functional ontology. Results demonstrate the utilization power of comparative genomics and prove the feasibility of using rodent microarrays to identification of putative coexpressed orthologous genes involved in the control of human mammary gland development.

## 1. Introduction

Mammals are the only animals that secrete a complex fluid from an elaborated skin gland to provide both innate protection and nourishment for their newborn. There are more

than 4,000 species of mammals with striking similarities in the structure and function of their mammary glands as well as in their unique milk components such as the caseins,  $\alpha$ -lactalbumin, lactoferrin, lactose, and milk fat.

Nevertheless, variations are exhibited in the arrangement and numbers of mammary gland, milk composition, and suckling strategies. Mammary gland development begins at puberty and is maintained throughout pregnancy until lactation. During these last stages, development compromises numerous overlapping programs such as branching morphogenesis, inductive stromal-epithelial interactions, programmed cell death, extracellular matrix remodeling, and hormone action [1].

Current knowledge of the molecular regulation of mammary development and lactation has largely been derived from the dissection of signaling networks in cell culture systems and phenotypic characterization of genetically altered mice as well as genomewide approaches such as microarrays. Nonetheless, to date, regulation of mammary gland development during pregnancy and lactation is incompletely understood. Lactation is regarded as one of the most remarkable products of evolution whose processes include the development of mammary tissue as well as the synthesis and secretion of milk [2]. Consequently, despite the fact that the development of mammary tissue and the synthesis and secretion of milk are considered as complex dynamic physiological processes, both must preserve overall common characteristics among mammals.

Considering the underlying assumption that important biological functions are often conserved across species, genes expressed across multiple species are likely to have conserved functions [3]. Given the completion of the DNA sequence of the human, mouse, and rat genomes [4], genes identified in microarray studies can be readily compared across species with respect to orthologous genes [5]. Therefore, a cross-species hybridization (CSH) experiment could provide significant information concerning probable conserved gene networks among mammals.

In a CSH experiment, there is hybridization of RNA from one or more (target) species to a microarray that contains DNA (cDNA or oligomers) from another (reference) species and represents a valuable tool for the identification of orthologous genes. Thus, a CSH microarray analysis offers the possibility of furthering our understanding of cross-species commonalities and differences that could lead to more effective use of animal models to understand the regulation of mammary gland development at the molecular level [6]. Dissection of unique patterns of expression of orthologous clusters of genes among species throughout distinct physiological time points along pregnancy and lactation could prove useful in the integrative analysis of the information available for discerning the molecular events underlying the regulation of mammary gland development and function that lead to milk synthesis.

In this study, bioinformatics techniques were applied to transcriptomic data. These data resulted from heterologous microarrays of target RNA samples derived from rat mammary gland during distinct stages of pregnancy and lactation in order to extrapolate and enhance the understanding on transcriptional module networks or coregulated functional gene groups conserved in rodents and in the development of the mammary gland in humans.

## 2. Materials and Methods

**2.1. Experimental Animals and Tissue Collection.** Fifteen female Sprague Dawley rats were obtained from the Animal Care Facility of Centro Médico Nacional Siglo XXI of the Mexican Institute of Social Security (IMSS) in Mexico City. Animals were housed at  $22 \pm 2^\circ\text{C}$  with a 12 h light/dark cycle with free access to water, and a purified diet was administered *ad libitum* during pregnancy and lactation. Dietary composition was previously reported by our group [7]. At 14 weeks of age, rats were randomly assigned to five groups representing distinct time points in mammary gland development: virgin (V), day 5 (P5) and, day 14 (P14) of pregnancy and day 1 (L1), and day 12 (L12) of lactation. Three rats were included in each group. Rats were mated and the same diet was administered during pregnancy and lactation. The day on which sperm was identified in vaginal smears was designated as day 1 of pregnancy and the day of parturition was designated as day 1 of lactation. Pregnant rats were housed individually. Litters were adjusted to eight pups per dam. No gender differentiation was made. Pups remained with their mother to stimulate milk synthesis. Rats were euthanized, and whole mammary tissue was removed from V, P5, P14, L1, and L12 rats. Tissue was immediately frozen in liquid nitrogen and stored at  $-70^\circ\text{C}$  for subsequent total RNA isolation or histological analysis.

### 2.2. Microarray Analysis

**2.2.1. Total RNA Isolation.** Total RNA was isolated from tissue (0.1-0.2 g) using TRIzol (Invitrogen, Carlsbad, CA, USA) following the method of Chomczynski and Sacchi [8]. Total RNA from mammary tissue was isolated from three different animals of each physiological period (V, P5, P14, L1, and L12), pooled, and kept in aliquots for later determination of purity and integrity. Total pooled RNA was used for microarray analysis and quantitative real-time PCR.

Four microarray datasets generated using the custom-designed *Mus musculus* oligonucleotide array containing 65-mer probe sets representing 22,000 transcripts (Microarray Unit, Cellular Physiology Institute, UNAM, Mexico City) were analyzed. Each dataset represented distinct time points in mammary gland development such as P5, P14, L1, and L12. Histologically, the mammary proliferative stage is represented by P5, the secretory differentiation stage by P14, early lactation by L1, and full lactation by L12. Design of the microarray experiments is presented in Table S1 in supplementary materials available online at <http://dx.doi.org/10.1155/2013/624681>.

**2.2.2. Probe Preparation and Hybridization to Arrays.** Ten  $\mu\text{g}$  of total pooled RNA was reverse transcribed into cDNA incorporating dUTP-Cy3 or dUTP-Cy5 and using the CyScribe First-Strand cDNA labeling kit (Amersham Biosciences, Piscataway, NJ, USA). Using hybridization solution UniHyb (TeleChem International Inc., Sunnyvale, CA, USA), equal quantities of labeled cDNA were hybridized to the M22 K.01 microarray for 14 h at  $42^\circ\text{C}$ . Four hybridization assays were carried out as follows: (a) the fluorophore used

was dUTP-Cy3 for control nonpregnant virgin rats (V) and dUTP-Cy5 for P5, (b) dUTP-Cy3 for V and dUTP-Cy5 for P14, (c) dUTP-Cy3 for V and dUTP-Cy5 for L1, and (d) dUTP-Cy3 for V and dUTP-Cy5 for L12. Each hybridization assay was carried out in triplicate.

Data acquisition and analysis of array images were performed in ScanArray 4000 with its accompanying software ScanArray 4000 from Packard BioChips.

### 2.3. Data Analysis

**2.3.1. Global Analysis: Overview of Gene Expression.** Microarray data analysis was performed with free software genArise, which was developed in the Computing Unit of the Cellular Physiology Institute of the UNAM (<http://www.ifc.unam.mx/genarise/>). The goal of genArise is to identify which genes show good evidence of being differentially expressed. The software identifies differentially expressed genes by calculating an intensity-dependent *z*-score. Elements with a *z*-score >1.5 standard deviations are considered to be significantly and differentially expressed genes.

The complete set of raw Excel data files have been deposited at Gene Expression Omnibus (GEO) and are available on the GEO website (ID GEO GSE22545).

**2.3.2. Clustering Analysis for Gene Expression.** Gene lists were generated by a set of multiple comparisons among the distinct developmental stages and intersection in Venn diagrams. Two-way hierarchical clustering with average linkage and a range of 5 to 15 K-means classifications were used to group our time series data using open source software Cluster v3.0 [9]. Java TreeView was used to display the clustering results as dendrogram or heat map representations. We adopted the procedure as described in [10] to code the mean expression of a cluster at each stage as flat, decreased, and increased and converted it to numerical representation.

**2.3.3. Determination of Orthologous Genes.** Putative orthologous genes in rat, mouse, and human were identified from a genome comparative search with Roundup (<http://rodeo.med.harvard.edu/tools/roundup>). Roundup is an ortholog and phylogenetic profile retrieval tool backed by a massive repository of orthologous and associated evolutionary distances that were built using the reciprocal smallest distance algorithm [11]. The search was done with a stringent blast *E*-value threshold of  $1.0 \times 10^{-5}$  and a divergence threshold of 0.2.

**2.3.4. Gene Ontology Analysis.** The DAVID 2.0 Functional Annotation Tool (<http://david.abcc.ncifcrf.gov/summary.jsp>) was used to sort and arrange the similar, redundant, and heterogeneous annotation contents from a set of genes into defined functional groups. In the case of insufficient gene ontology information, published data on orthologous genes was used to classify the gene into a functional category.

**2.3.5. Pathway Analysis.** Pathway mapping was accomplished using the Kyoto Encyclopedia of Genes and Genomes

(KEGG) database of biological systems that integrates genomic, chemical, and systemic functional information (<http://www.genome.jp/kegg/kegg2.html>).

**2.3.6. Interaction Network Analysis.** Gene lists were converted to Human SwissProt IDs using tables from the Ensembl database, release 42 [12]. For each list of Human SwissProt IDs, interactions between those gene products were obtained from Online Predicted Human Interaction Database (OPHID) and postprocessed using custom scripts to determine all linkages in the network and to generate a network file. This network file was then explored using NAVIGaTOR v2.0.15, a program for large network analysis (<http://ophid.utoronto.ca/navigator/index.html>).

**2.3.7. Transcription-Factor Binding-Site Prediction.** Transcription-factor-binding site (TFBS) prediction was accomplished using CORE\_TF (Conserved and Over-Represented Transcription Factor binding sites), a web-based tool that identifies overrepresented TFBS in promoters from coexpressed genes aided by the evaluation of cross-species conservation.

**2.3.8. Quantitative Real-Time PCR.** We utilized qRT-PCR for validation of microarray results. We measured the relative transcript levels of 14 target genes, and five genes commonly used as references such as Glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*), beta-actin (*Actb*), and ribosomal large protein P0 (*Rlp0*) were used as high abundance internal controls as well as splicing factor arginine/serine-rich 1 (*Sfrs1*) and hypoxanthine guanine phosphoribosyl transferase 1 (*Hprt1*) as medium- and low-abundance internal controls, respectively (Table S2).

Quantitative real-time PCR was performed in a 20  $\mu$ L reaction with 5.0  $\mu$ L from 1/4 reverse transcription dilution using the LightCycler Probes Master mix (Roche Diagnostics, Mannheim, Germany) containing 0.2  $\mu$ M of mRNA-specific primers and 0.1  $\mu$ M corresponding UPL probe into LightCycler microplate wells under reduced light conditions. Each sample was run in triplicate.

**2.4. Calculations and Statistics.** The results are expressed as mean  $\pm$  SEM of at least three individual experimental observations. Data were tested for normality of distribution by the Kolmogorov-Smirnov test. Statistically significant differences among experimental groups (between the mean values of each group) were determined by an unpaired Students *t*-test (*z*-test), ANOVA, or a modified Fisher's exact test. Nonnormally distributed data were analyzed by the Mann-Whitney *U* test.

## 3. Results

**3.1. Histological Characteristics of Pregnant and Lactating Mammary Gland.** The rat mammary gland undergoes a series of dramatic phenotypic changes during pregnancy and lactation. In order to determine the integrity of the dissected inguinal mammary glands, a gross histological evaluation

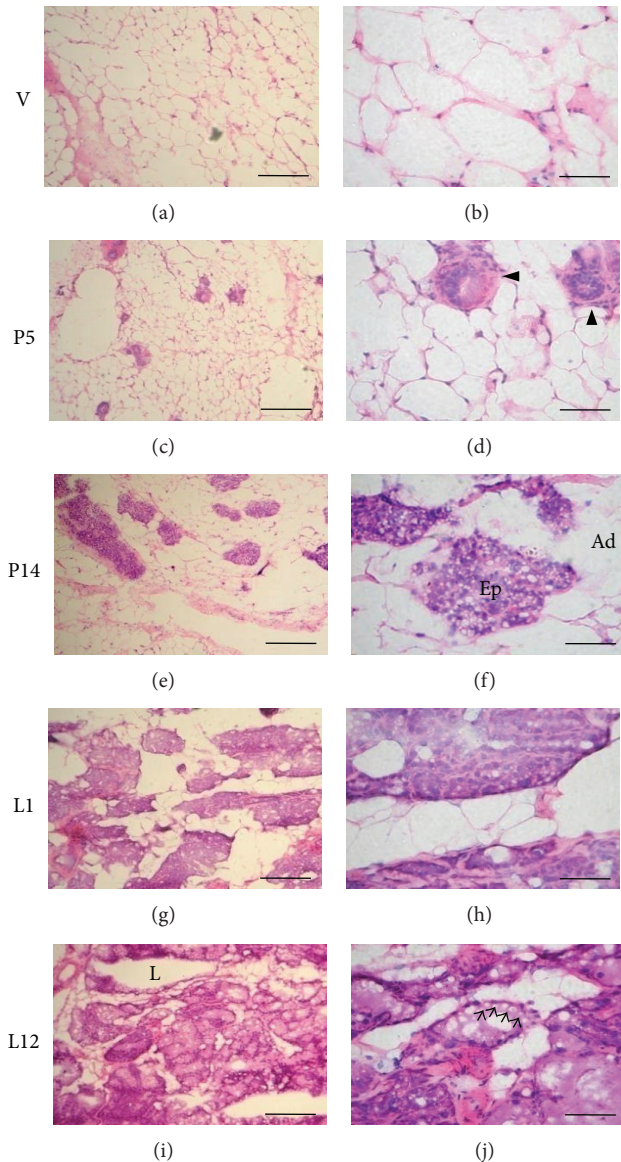


FIGURE 1: Histological features of the mammary gland from rats during pregnancy and lactation. Mammary glands were isolated from Sprague Dawley rats in (a, b) a nonpregnant virgin (V) stage; (c, d) day 5 (P5) and (e, f) day 14 (P14) of pregnancy; and (g, h) day 1 (L1) and (i, j) day 12 (L12) of lactation, cryosectioned, and stained with hematoxylin and eosin. Scale bars in (a), (c), (e), (g), and (i) = 50  $\mu\text{m}$ , whereas those in (b), (d), (f), (h), and (j) = 10  $\mu\text{m}$ . Adipose compartment (Ad), lobuloalveolar units ( $\blacktriangle$ ), epithelial compartment (Ep), luminal structures (L), and milk globules ( $\blackarrow$ ).

of the characteristic cytomorphological features were determined through hematoxylin-eosin staining (Figure 1). Four time points (pregnancy days 5 and 14; lactation days 1 and 12) were selected to represent distinct periods in mammary gland development. Histologically, the mammary proliferative stage is represented by P5, the secretory differentiation stage by P14, early lactation by L1, and full lactation by L12.

As reported elsewhere [13], initial changes observed during pregnancy include an increase in ductal branching

and the formation of alveolar buds (Figures 1(c) and 1(d)). The latter half of pregnancy is characterized by the expansion of alveolar buds to form clusters of lobuloalveolar units followed by the differentiation of these structures into presecretory structures. By day 14 of pregnancy, there is a readily apparent increase in the size of the epithelial compartment (Ep) (Figures 1(e) and 1(f)), and expansion of the epithelium (whereas the adipose compartment decreases) continues until the epithelial compartment predominates by onset of lactation (Figures 1(g) and 1(h)). By day 12 of lactation in the rat, the mammary gland is producing copious amounts of milk [7]. As expected, examination of the histology of the mammary gland at this stage reveals prominent luminal structures (L) and ducts and few adipocytes visible at this time (Figures 1(i) and 1(j)).

**3.2. Global Analysis: Overview of Gene Expression.** In this study, we analyzed expression profiles of 22,000 transcripts using murine microarrays and RNA samples of MG from virgin, pregnant, and lactating rats by cross-species hybridization. We first identified the total number of genes differentially expressed throughout distinct time points in mammary gland development such as P5, P14, L1, and L12.

A total of 807 oligonucleotide probe sets representing 521 annotated genes showed differential expression in at least one of four physiological time points evaluated, taking into consideration a mean  $z$ -score cutoff value of 1.50 standard deviations using GenArise.

During early pregnancy (day 5), 158 transcripts were differentially expressed. Most of these transcripts (123, 77.8%) were upregulated, suggesting a feasible tendency in the direction of gain of function versus the virgin stage (V). Likewise, in mid-pregnancy (day 14), as opposed to the virgin stage, the number of transcripts with an altered expression maintained a similar value (133 transcripts; 89.26% upregulated). During early and mid-lactation (days 1 and 12), 342 and 461 transcripts were differentially expressed, corresponding to a percentage of 64.0 and 38.4 overexpressed, respectively (Figure 2(a)).

To further illustrate the differences and commonalities among the four physiological time points, changes in gene expression were also interpreted with a Venn diagram. As shown in Figure 2(b), the descriptive table of the Venn diagram denotes the number of genes showing upregulation ( $\uparrow$ ) or downregulation ( $\downarrow$ ) uniquely at pregnancy (day 5 or 14) or lactation (day 1 or 12) and differential expression at a combination of stages. Venn diagram analysis indicated that 47.2% (381/807) of all the differentially expressed transcripts presented an average significant  $z$ -score fold change ( $z > \pm 1.5$ ) exclusively during either or both time points of lactation. Interestingly, among the 381 altered gene transcripts during lactation, 64.8% (247/381) were found downregulated, implying as previously stated by Lemay et al. (2007) [2] that mammary epithelial cells become biofactories not by gain of function but by a broad suppression of function to effectively push all cell resources towards a very few important tasks. All the gene sets that shared spatial and temporal distributions (overlapping changes in expression) are listed in additional data files (Table S3).

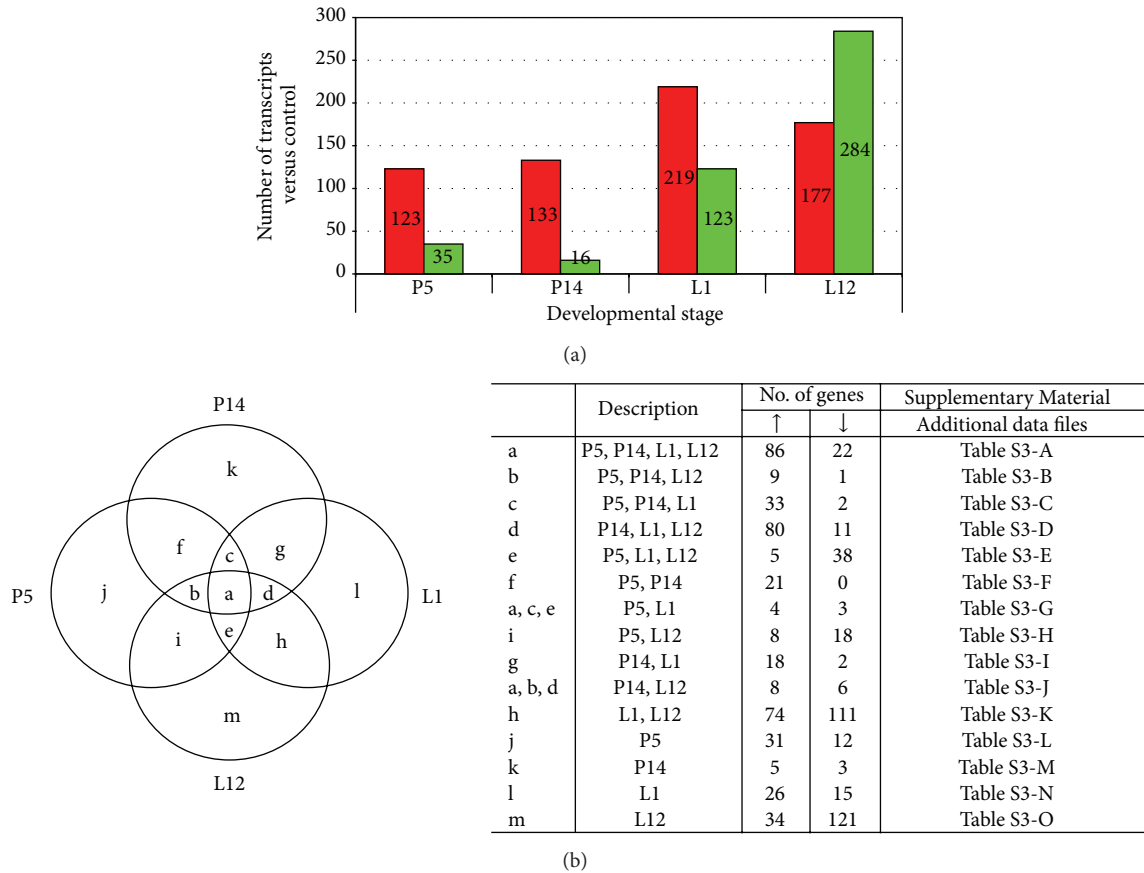


FIGURE 2: Overview of gene expression during mammary development through early pregnancy to mid-lactation. (a) Number of altered transcripts of each developmental stage versus a virgin stage (fold change ( $z$ -score)  $>1.5$ ). Low expression is represented by green and high expression is represented by red. (b) Venn diagram and representative table illustrating the number of significant altered gene transcripts and overlaps among different reproductive stages.

**3.3. Clustering Analysis for Gene Expression.** To determine global alterations in gene expression across developmental stages of the mammary gland from early pregnancy to mid-lactation, we performed a complete-linkage hierarchical clustering with an Euclidean distance similarity metric on the expression profiles of the differentially expressed genes (annotated and EST) across all four time points. The expression profiles of the 807 genetic elements resulted in six predominant clusters on a dendrogram (designated clusters 1–6 in Figure 3). All the gene cluster sets are enumerated in Table S4.

Cluster 1 (C1) represents 37.35% of the total 807 genetic elements. This major trend is a decline in gene expression during mid-pregnancy that remains low during lactation. Cluster 2 (C2) describes 25.77% of the total data population and is characterized by a linear decrement in gene expression towards mid-lactation. However, positive  $z$ -score values are retained with respect to the reference stage (virgin). The remainder of the clusters (C3–C6) appears to explain between 3.09 and 17.84% of the data variation. In C3 and C4, gene expression rises exponentially from early pregnancy, reaching a plateau during mid-lactation. However, the slope of the curve is even steeper in C3 in comparison to C4. In cluster

5 (C5), 65 elements matched the profile outline (inverted sigmoid form) of major trend C1 although the reduction tendency was less marked. In cluster 6 (C6), expression was roughly unchanged during pregnancy and lactation. Even so, the relative abundance of transcripts remained in a higher proportion than the reference virgin stage as described for C2.

This transcriptional profile, involved in the mammary development program identified in rat, could be conserved in other mammals like mouse. Consequently, in order to delineate potential groups of coreregulated genes, final cluster membership was determined by a K-means analysis based on the preestimated number (six) of gene clusters.

K-means clustering revealed six distinct clusters (K1–K6) that distinguished up from down, early from middle, and transient from sustained changes in expression (Figure 4; Table S5). Each of the six clusters was designated with its unique trajectory expression profile signature across stages (pregnancy days 5 and 14, lactation days 1 and 12) as presented in Figure 4.

There were two major groups of 245 under- and 175 overexpressed tags during lactation only (Table S5, K1: 1,1,0,0 and K3: 1,1,2,2 according to the procedure of Rudolph et al., [10] to code the mean expression, see Section 2). Among

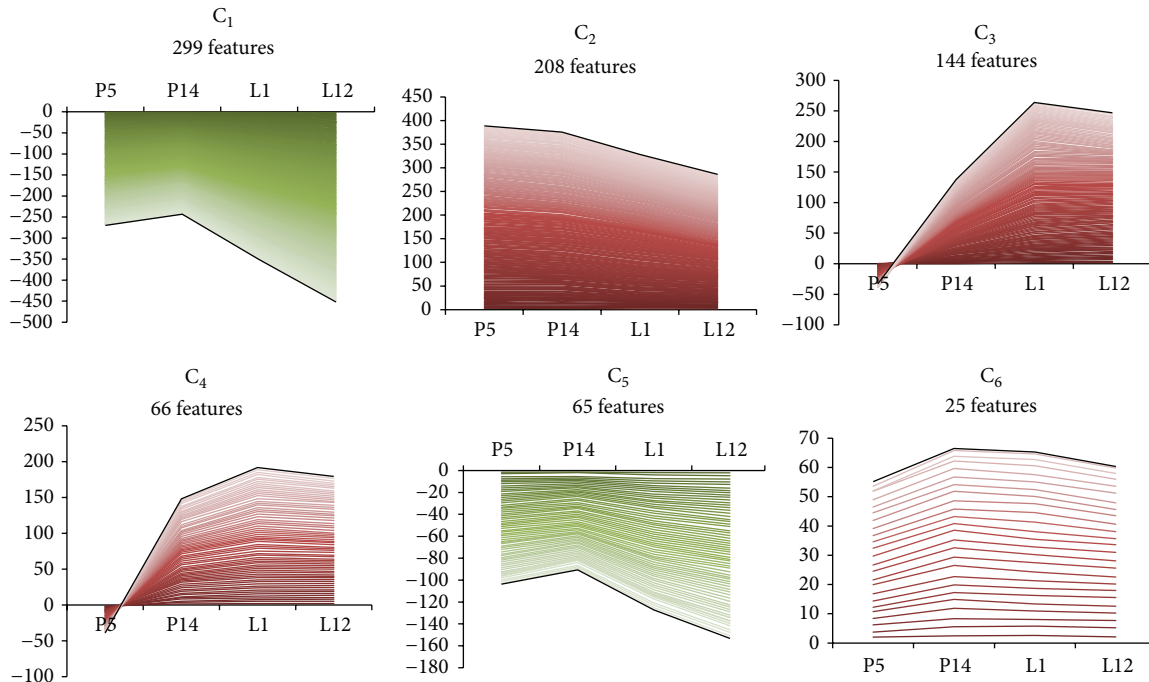


FIGURE 3: Unsupervised hierarchical clustering of the differentially expressed gene list. Gene expression profiles. Gene expression data from the 807 tags comprising the significant altered gene transcript list ( $z$ -score  $> \pm 1.50$ ) were best represented by six clusters consisting of distinct up (red) and down (green) patterns of expression. Developmental stage time points and fold change are indicated on the  $x$ - and  $y$ -axes, respectively. The number of features (genetic elements) in each cluster is indicated. A black pseudoline representing the general (average) pattern of expression has been superimposed on each cluster.

the typical upregulated genes of lactation stage are the milk protein (casein alpha (*Csn1s1*, *Csn1s2a*), casein beta (*Csn2*), and whey acidic protein (*Wap*)) and biogenesis genes that mainly concern glucose and lipid metabolism (*Akr1c6*, *Aldob*, *Ugt2b1*, *Plb1*, *ApoE*, and *Sult2b1*) and transcriptional regulation (*Stat5a*, *Pou2af1*) [2]. Among those genes found significantly downregulated only during lactation, several play an important role in the regulation of apoptosis, mediation of metastatic behavior (epithelial-mesenchymal transition), or ubiquitin-mediated protein catabolism (lysosome degradation) in the mammary gland including *Igfbp5*, *Mmp2*, and *Ube2r2* [14, 15]. One hundred forty-nine genes were upregulated exclusively during early pregnancy (K2: 2,1,1,1) such as *Esr1*, *Esr2*, *Tshr*, and *Oxt*. These participate in the transduction of hormonal status [2, 16] involved in the modulation of important physiological processes such as carbohydrate metabolism (*Creb3l4*, *Hkl1*, and *Coasy*) [17]; glutathione metabolism (*Ggt1*, *Mgst1*, and *Gstm6*) [18]; cell differentiation (*Foxa1*, *Mtap7*, *Gdf1*, *Twist2*, *Hey1*, *Dll4*, and *Pcaf*) [19–23], stromal-epithelial communication (cell-cell junctions) (*Cldn10*, *Mpp5*, and *Epb4*) [24], and cell adhesion (*Matn1*, *Krt71*, *Mpzl2*, and *Dscaml1*) [25].

The smallest group of 24 genes were significantly upregulated exclusively at the onset of lactation (K4: 1,1,2,1) such as *Lpo*, *Cd8a*, and *Irs1*, important for lactogenesis, particularly in milk production capabilities and related immunotropic constituents (antigen-specific CD8+ T cells) found in colostrum [26, 27]. One hundred nineteen genes were downregulated from early pregnancy (K5: 1,1,1,0). For example, *Acta1*, *Flnc*,

and *Pax7*, which are either restricted to muscular tissues or involved in myogenic development and cellular differentiation [28, 29], are included in this group. According to the trajectory profile signature, 95 additional genes were found upregulated at all stages evaluated (K6: 2,2,2,2). Interestingly, most of the overexpressed genes in this group include general transcription and translation (including spliceosome and protein folding) machinery factors (*Eif4a2*, *Eif2ak1*, *Etf1*, *Taf1*, *Ercc2*, *Sart1*, *Ppih*, and *Dbr1*) [30, 31] as well as structural (*Itga5*, *Actg1*, *Add2*, *Cldnd2*, *Rptn*, and *Triobp*) [32] and basal metabolic genes (*Pank4*, *Agpat5*, *Cyp24a1*, and *Phyh*) [33].

**3.4. Determination of Orthologous Genes.** Once the gene clusters were properly defined, identification of orthologous gene transcripts among the time course differentially expressed gene list subsets was critical for reliable comparison of gene function and subsequent determination of probably conserved transcriptional modules implicated in biological processes during the development of mammary tissue. According to genome comparative RoundUp orthologous database of a total of 448 transcripts upregulated and 371 downregulated, 213 (upregulated) and 183 (downregulated) genes were identified as orthologous to rat and/or human. The remainder of the genes was discarded or removed from subsequent analysis due to lack of similarity, insufficient information, or unknown identifiers. A complete list of orthologous genes from each dataset was compiled (Table S6).

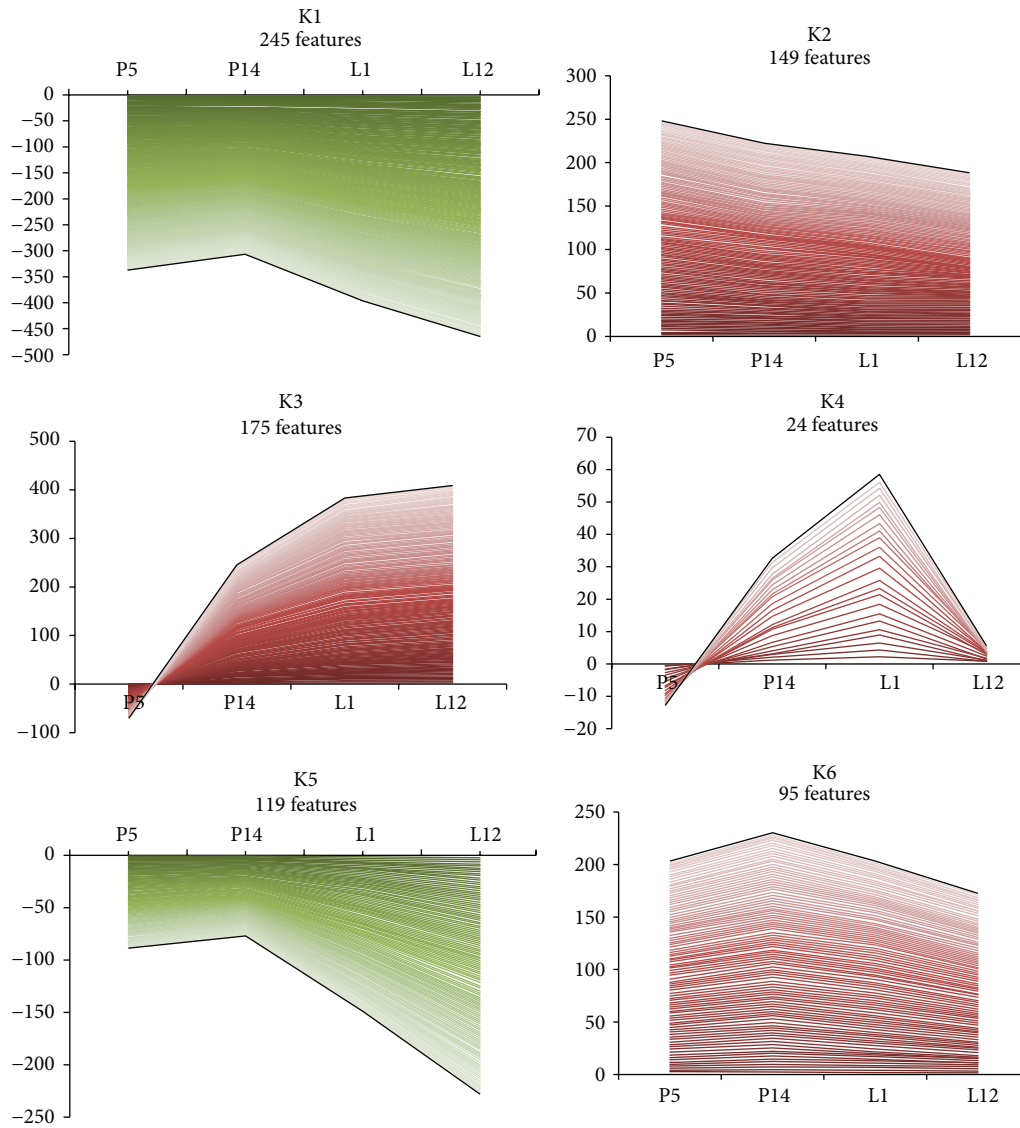


FIGURE 4: K-means clustering of the differentially expressed gene list with trajectory expression profile signatures. Gene expression data from the 807 tags comprising the significant altered gene transcript list ( $z$ -score  $> \pm 1.50$ ) were best represented by six K-means clusters consisting of distinct up (red) and down (green) patterns of expression. Developmental stage time points and fold change are indicated on the  $x$ - and  $y$ -axes, respectively. The numbers of tags in each cluster are indicated. A black pseudoline representing the general (average) pattern of expression has been superimposed on each cluster.

Among the upregulated orthologous genes to rat and/or human are those encoding to milk proteins, carbohydrate and lipid metabolism, transcriptional factors [17, 34], transduction of hormones [16], glutathione metabolism [18], and cell differentiation [19, 20]. Others are associated with stromal-epithelial communication [24], cell adhesion [25], lactogenesis [26], and general transcription and translation machinery factors [30] as well as structural [32] and basal metabolic genes [33–35].

Among those genes found significantly downregulated, several play an important role in the regulation of apoptosis, mediation of metastatic behavior (epithelial-mesenchymal transition), or ubiquitin-mediated protein catabolism (lysosome degradation) [14, 15]. Also, genes restricted to muscular

tissues or involved in myogenic development and cellular differentiation [28, 36] are downregulated.

**3.5. Confirmation Studies.** Taking into consideration their temporal expression profile signature and the fact that they represent different K-means cluster, 14 genes were selected for real-time PCR analysis (Table S2). Results show that the expression trends were consistent with the results from the microarray analysis. Correlation analysis showed good agreement between real-time RT-PCR and microarray analysis. Microarray results for all 14 genes tested were confirmed by real-time RT-PCR with regard to direction and significance of change (Figure 5).

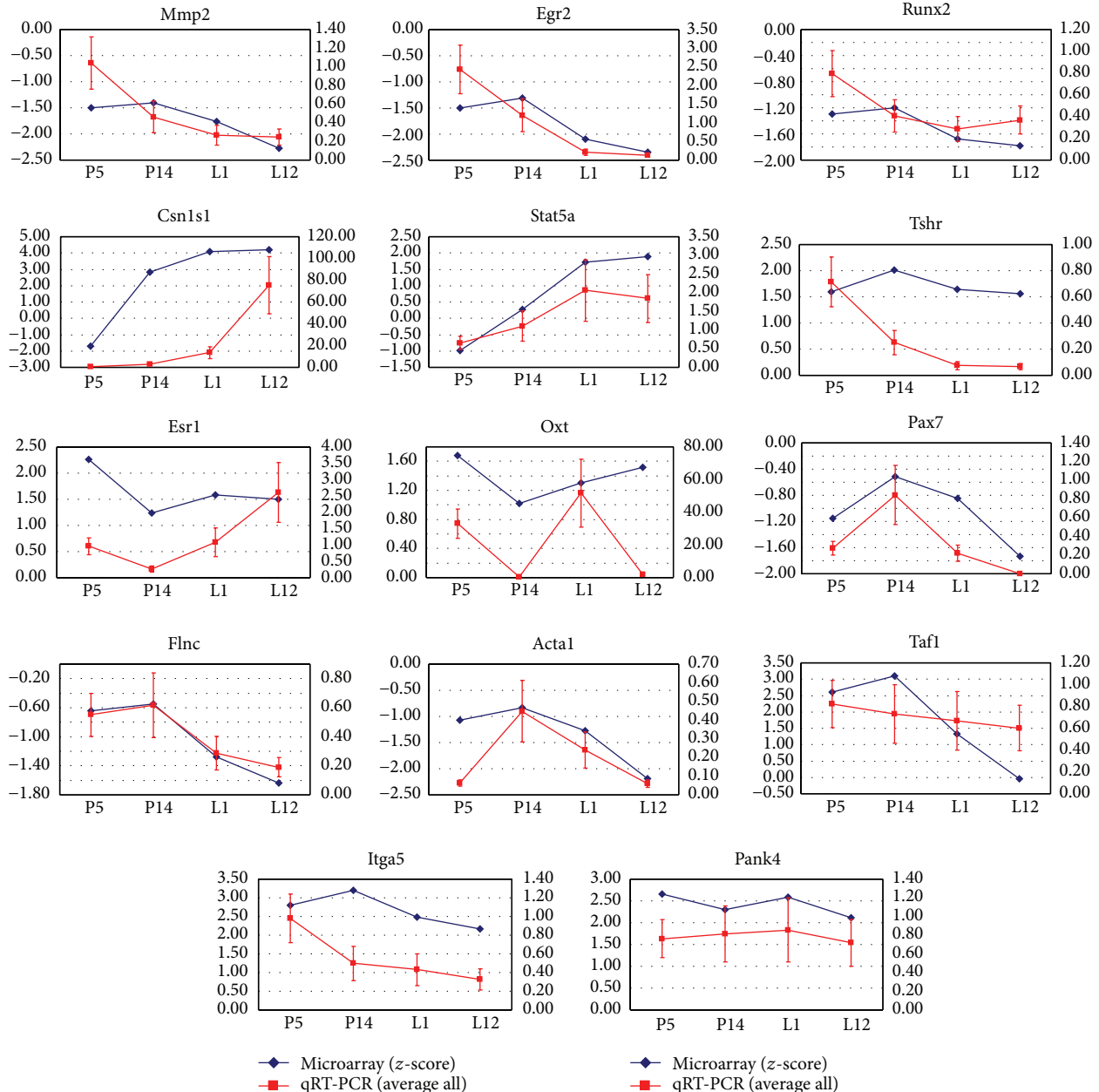


FIGURE 5: Validation of differentially expressed genes identified by microarray using qRT-PCR. Quantitative real-time-PCR was performed as in *Experimental Procedures*, and the calculated mean log ratio for each gene in their corresponding developmental stage was compared with the mean z-score from the microarray analysis during that same period. A positive value indicates a greater mRNA abundance in a developmental stage than in the control group, whereas a negative value indicates a lower mRNA abundance in a developmental stage than in the control group. Lines represent data obtained by qRT-PCR (red, *right axis*) and microarray analysis (blue, *left axis*), respectively, whereas the *x-axis* represents the physiological time point. The average fold change relative to time-matched virgin controls for three animals per group is shown. Genes are indicated by their official gene symbols. Data are presented as the mean of three independent experiments, each performed in triplicate. Error bars represent the SD for the average fold change. The correlation information in each analysis is indicated by Spearman rho value ( $r_s$ ). All 18 pairs of genes had an absolute correlation value of  $r_s = 0.49$  with a  $P = 7.59E - 06$ . ND: not detected.

#### 4. Discussion

Structural and functional homologies of specific genes are important. Conservation of functional blocks of genes is likely to be more important in a cross-species comparison. We found distinct blocks of significantly differentially

expressed genes within different cytogenetic regions of the rat with homologous chromosomal segments in human and mouse. However, human, mouse, and rat have different chromosomal arrangements. Genes in these blocks appear in contiguous cytogenetic regions, irrespective of species and chromosomal location. This finding is not surprising



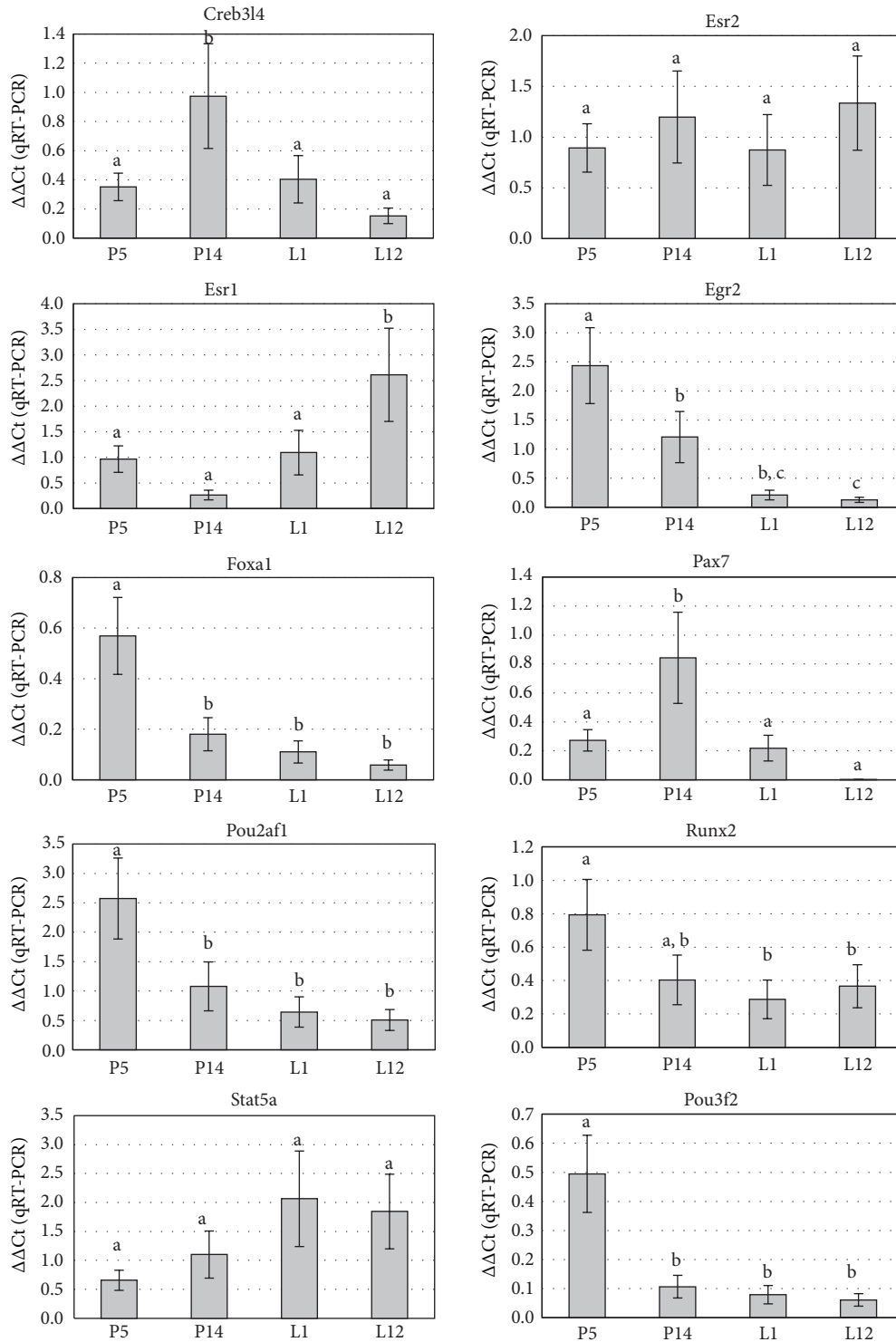


FIGURE 6: Gene expression profiles of the putative transcriptional factors identified by microarray using qRT-PCR. Quantitative real time-PCR was performed as in *Experimental Procedures*, and the calculated mean log ratio for each gene in their corresponding developmental stage was determined. A positive value indicates a greater mRNA abundance in a developmental stage than in the control group, whereas a negative value indicates a lower mRNA abundance in a developmental stage than in the control group. Bars represent data obtained by qRT-PCR (*y*-axis), whereas the *x*-axis represents the physiological time point. The average fold change relative to time-matched virgin controls for three animals per group is shown. Genes are indicated by their official gene symbols. Data are presented as the mean of three independent experiments, each performed in triplicate. Bars represent the average ( $\pm$ SD) fold change. Different letters indicate statistically significant differences between developmental stages ( $P < 0.05$ ). Note that the scales of the *y*-axis vary among genes. ND: not detected.

considering the close evolutionary distance between the species where 278 orthologous segments are reported to be shared between human and rat, and 280 segments are reported to be shared between human and mouse [4]. It is proposed that these gene blocks may be significant for mammary gland development and maintenance and progression of lactation across human, rat, and mouse. For example, genes in the blocks may be coordinately expressed to share transcription programs as stated in previous studies [37]. The argument may be made against the feasibility of using rodent data to draw inferences to human mammary gland gene expression. However, our objective in this study was to utilize the best available sources of information such as rat gene expression data during mammary development and mapping data to develop hypotheses on putative functional gene blocks conserved across species, underlying similarities despite reported differences in the architecture and hormonal control of mammary glands between rats and other rodents and between rats and humans [38, 39].

In an effort to further characterize potential highly coregulated gene blocks, we combined transcription-factor binding-site prediction [40] along the promoters of each gene member with the detection of expression profiles of annotated altered transcripts categorized as nucleic acid binding protein. Several families of transcription factors were identified (Table S7). For the most part, zinc finger domain/motif proteins were the most widely represented. The presence of *cis*-elements found with CORE\_TF (<http://www.LGTC.nl/CORE.TF>) in the promoters of the genes *Slc44a4*, *Ppt2*, and *B3galt4* that comprises the conserved block D15 (Table S8) along with the cotranscription of mRNAs that encoded for *trans*-regulator elements suggests that they are most likely modulated by transcription factors *Runx2*, *Creb3l4*, *Pou3f2*, and *Pou2af1*. Correspondingly, the gene members of block U1 may possibly be coregulated by transcription factors *Stat5a*, *Foxa1*, *Creb3l4*, and *Pou2af1*. In the same manner, other gene blocks (U3, U8, U9, D1, D9, and D14, Table S8) were found most likely co-regulated by a minor number of transcription factors (*Foxa1*, *Creb3l4*, *Pou2af1*, or *Egr2*). Hence, identification of conserved overrepresented upstream motifs unravels putative regulatory elements for transcription (Figure 6) in at least half of the gene block members reported in this study. Consequently, these results strongly substantiate the maintenance of comparable transcriptional regulation programs among the predicted coexpressed modules.

Because cotranscription of genes in conserved blocks may allow concerted expression of gene products involved in the same response or pathway [41], integration of this type of analysis enables the discovery of putative evolutionary conserved regulatory networks among mammals. Thus, the co-regulated clusters we proposed may indeed be conserved transcriptional modules through evolution, at least between rodents and primates.

Heterologous hybridization experiments on any microarray are of limited use for genes that have undergone rapid evolutionary change in coding regions, large rearrangements, and duplication [42]. Long oligonucleotide-based microarray platform may be more suitable for cross-species gene

expression studies than a short oligonucleotide-based system [43]. This comparative approach is based on the assumption that similar gene sequences in closely related species allow a reasonably reliable detection of many orthologous genes. For instance, according to several independent and unrelated studies carried out on comparable 50 to 60-mer oligonucleotide microarrays, cross-hybridization was observed only with genes with 50%–75% overall sequence identity, respectively [44, 45]. Considering that orthologous genes between human and mouse and between human and rat both have a mean of ~85% sequence identity [46], validity of the results obtained in this study—despite the problems encountered by CSH—in comparison to SSH seems upholding. In fact, the nucleotide sequence alignment confirmed an >75.3% homology at least for the transcript members of the distinct gene blocks described, depending on the sequence evaluated among primates and rodents reinforcing the notion of attaining valid biological results. In addition, similar expression trends for distinct probe sets for one corresponding gene (data not shown) seem to largely substantiate the certainty and reproducibility of hybridization results obtained in this study.

Because of the challenges inherent to CSH data, their confirmation by other techniques is essential [43]. In addition to qRT-PCR, orthologous gene expression profiles with syntenic regions of rat, mouse, and human chromosomes reinforce another confirmation method that potentially substantiates the CSH results obtained in this study. Nonetheless, further validation of the results must be carried out by using CSH of human RNA to mouse oligonucleotide arrays.

This study provides access to a prevalidated platform for analyzing transcriptional changes in rat mammary gland. This paper will hopefully spur an increase of mammary gland CSH transcriptome analysis, thus adding to our knowledge base of this interesting evolutionary feature among mammals. However, although we acknowledge the multitude of aspects that can be elucidated by traditional SSH transcriptome analysis, we believe the biggest potential of the presented microarray lies in the multispecies-type studies described. We demonstrated that data analysis strategies such as the combination of orthologous gene expression profiles and chromosome mapping in conjunction with directed promoter transcription-factor binding-site prediction presented here can add strength to conclusions and help identify systems and responses that are conserved across the mammal *taxa*. The possibility of studying the evolutionary depth of transcriptional regulation adds a new dimension to comparative transcriptomic, particularly identification of differentially coregulated gene blocks mapped to highly conserved syntenic chromosomal regions, which is important in mammary gland development using CSH experiments among mammal species.

## Conflict of Interests

All authors declare that they have no financial or personal relationships with other persons or organizations that could inappropriately influence this work.

## Acknowledgments

This study was supported by the Coordinación de Investigación Médica en Salud, Instituto Mexicano del Seguro Social (IMSS), Mexico (Grant no. FIS/IMSS/PROT/531), and Consejo Nacional de Ciencia y Tecnología (CONACYT), Mexico (Grant no. 101012). The authors thank Dr. Carlos Pérez Plasencia and Dr. Omar K. Ruvalcaba Salazar for their careful reading of this paper and useful comments that improved the presentation; Simón Guzmán León, Jose Luis Santillán Torres, and Lorena Chávez González from the DNA Microarray Facility for technical assistance in the microarray determination; and Gerardo Coello, Lina Riego, Gustavo Corral, Patricia Gómez, and Emmanuel Plata for GenArise software assistance. The authors acknowledge Sharon Morey, Scientific Communications, for providing editorial assistance.

## References

- [1] T. W. Hale and P. Hartmann, *Textbook on Human Lactation*, Hale Publishing, Amarillo, Tex, USA, 2007.
- [2] D. G. Lemay, M. C. Neville, M. C. Rudolph, K. S. Pollard, and J. B. German, "Gene regulatory networks in lactation: identification of global principles using bioinformatics," *BMC Systems Biology*, vol. 1, article 56, 2007.
- [3] X. J. Zhou and G. Gibson, "Cross-species comparison of genome-wide expression patterns," *Genome Biology*, vol. 5, no. 7, article 232, 2004.
- [4] R. A. Gibbs, G. M. Weinstock, M. L. Metzker et al., "Genome sequence of the Brown Norway rat yields insights into mammalian evolution," *Nature*, vol. 428, no. 6982, pp. 493–521, 2004.
- [5] D. L. Wheeler, T. Barrett, D. A. Benson et al., "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Research*, vol. 35, no. 1, pp. D5–D12, 2007.
- [6] H. Koltai and C. Weingarten-Baror, "Specificity of DNA microarray hybridization: characterization, effectors and approaches for data correction," *Nucleic Acids Research*, vol. 36, no. 7, pp. 2395–2405, 2008.
- [7] M. Rodríguez-Cruz, R. Sánchez, M. Bernabe-García, J. Maldonado, M. del Prado, and M. López-Alarcón, "Effect of dietary levels of corn oil on maternal arachidonic acid synthesis and fatty acid composition in lactating rats," *Nutrition*, vol. 25, no. 2, pp. 209–215, 2009.
- [8] P. Chomczynski and N. Sacchi, "Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction," *Analytical Biochemistry*, vol. 162, no. 1, pp. 156–159, 1987.
- [9] J. H. Kim, "Chapter 8: biological knowledge assembly and interpretation," *PLOS Computational Biology*, vol. 8, no. 12, Article ID e1002858, 2012.
- [10] M. C. Rudolph, J. L. McManaman, L. Hunter, T. Phang, and M. C. Neville, "Functional development of the mammary gland: Use of expression profiling and trajectory clustering to reveal changes in gene expression during pregnancy, lactation, and involution," *Journal of Mammary Gland Biology and Neoplasia*, vol. 8, no. 3, pp. 287–307, 2003.
- [11] T. F. DeLuca, I.-H. Wu, J. Pu et al., "Roundup: a multi-genome repository of orthologs and evolutionary distances," *Bioinformatics*, vol. 22, no. 16, pp. 2044–2046, 2006.
- [12] P. Flicek, B. L. Aken, K. Beal et al., "Ensembl 2008," *Nucleic Acids Research*, vol. 36, no. 1, pp. D707–D714, 2008.
- [13] S. M. Anderson, M. C. Rudolph, J. L. McManaman, and M. C. Neville, "Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis!," *Breast Cancer Research*, vol. 9, no. 1, article 204, 2007.
- [14] D. J. Flint, M. Boutinaud, C. B. A. Whitelaw, G. J. Allan, and A. F. Kolb, "Prolactin inhibits cell loss and decreases matrix metalloproteinase expression in the involuting mouse mammary gland but fails to prevent cell loss in the mammary glands of mice expressing IGFBP-5 as a mammary transgene," *Journal of Molecular Endocrinology*, vol. 36, no. 3, pp. 435–448, 2006.
- [15] Y. Ning, B. Hoang, A. G. P. Schuller et al., "Delayed mammary gland involution in mice with mutation of the insulin-like growth factor binding protein 5 gene," *Endocrinology*, vol. 148, no. 5, pp. 2138–2147, 2007.
- [16] A. Morani, M. Warner, and J.-Å. Gustafsson, "Biological functions and clinical implications of oestrogen receptors alpha and beta in epithelial tissues," *Journal of Internal Medicine*, vol. 264, no. 2, pp. 128–142, 2008.
- [17] S. B. Aicha, J. Lessard, M. Pelletier, A. Fournier, E. Calvo, and C. Labrie, "Transcriptional profiling of genes that are regulated by the endoplasmic reticulum-bound transcription factor AIBZIP/CREB3L4 in prostate cells," *Physiological Genomics*, vol. 31, no. 2, pp. 295–305, 2007.
- [18] J. Ålander, J. Lengqvist, P. J. Holm et al., "Microsomal glutathione transferase 1 exhibits one-third-of-the-sites-reactivity towards glutathione," *Archives of Biochemistry and Biophysics*, vol. 487, no. 1, pp. 42–48, 2009.
- [19] H. Nakshatri and S. Badve, "FOXO1 in breast cancer," *Expert Reviews in Molecular Medicine*, vol. 11, p. e8, 2009.
- [20] D. R. Magnan, D. V. Spacek, N. Ye, Y.-C. Lu, and T. R. King, "The male sterility and histoincompatibility (mshi) mutation in mice is a natural variant of microtubule-associated protein 7 (Mtap7)," *Molecular Genetics and Metabolism*, vol. 97, no. 2, pp. 155–162, 2009.
- [21] C. T. Rankin, T. Bunton, A. M. Lawler, and S.-J. Lee, "Regulation of left-right patterning in mice by growth/differentiation factor-1," *Nature Genetics*, vol. 24, no. 3, pp. 262–265, 2000.
- [22] J. C. Martinez, M. M. Müller, H. Turley et al., "Nuclear and membrane expression of the angiogenesis regulator delta-like ligand 4 (DLL4) in normal and malignant human tissues," *Histopathology*, vol. 54, no. 5, pp. 598–606, 2009.
- [23] C. Zhu, Y.-R. Qin, D. Xie et al., "Characterization of tumor suppressive function of P300/CBP-associated factor at frequently deleted region 3p24 in esophageal squamous cell carcinoma," *Oncogene*, vol. 28, no. 31, pp. 2821–2828, 2009.
- [24] A. C. Zemke, J. C. Snyder, B. L. Brockway et al., "Molecular staging of epithelial maturation using secretory cell-specific genes as markers," *American Journal of Respiratory Cell and Molecular Biology*, vol. 40, no. 3, pp. 340–348, 2009.
- [25] C. Nicolae, Y.-P. Ko, N. Miosge et al., "Abnormal collagen fibrils in cartilage of matrilin-1/matrilin-3-deficient mice," *Journal of Biological Chemistry*, vol. 282, no. 30, pp. 22163–22175, 2007.
- [26] D. L. Hadsell, W. Olea, N. Lawrence et al., "Decreased lactation capacity and altered milk composition in insulin receptor substrate null mice is associated with decreased maternal body mass and reduced insulin-dependent phosphorylation of mammary Akt," *Journal of Endocrinology*, vol. 194, no. 2, pp. 327–336, 2007.

- [27] C. J. Watson, "Immune cell regulators in mouse mammary development and involution," *Journal of Animal Science*, vol. 87, no. 13, pp. 35–42, 2009.
- [28] I. Dalkilic, J. Schienda, T. G. Thompson, and L. M. Kunkel, "Loss of filaminC (FLNc) results in severe defects in myogenesis and myotube structure," *Molecular and Cellular Biology*, vol. 26, no. 17, pp. 6522–6534, 2006.
- [29] Z. Luan, Y. Liu, T. J. Stuhlmiller, J. Marquez, and M. I. García-Castro, "SUMOylation of Pax7 is essential for neural crest and muscle development," *Cellular and Molecular Life Sciences*, vol. 70, no. 10, pp. 1793–1806, 2013.
- [30] R. C. Wek and D. R. Cavener, "Translational control and the unfolded protein response," *Antioxidants and Redox Signaling*, vol. 9, no. 12, pp. 2357–2371, 2007.
- [31] M. C. Thomas and C. M. Chiang, "The general transcription machinery and general cofactors," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 41, no. 3, pp. 105–178, 2006.
- [32] I. A. Belyantseva, B. J. Perrin, K. J. Sonnemann et al., "γ-Actin is required for cytoskeletal maintenance but not development," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 24, pp. 9703–9708, 2009.
- [33] N. Saito, Y. Suhara, D. Abe et al., "Synthesis of 2α-propoxy-1α,25-dihydroxyvitamin D3 and comparison of its metabolism by human CYP24A1 and rat CYP24A1," *Bioorganic and Medicinal Chemistry*, vol. 17, no. 13, pp. 4296–4301, 2009.
- [34] C. J. Watson and K. Neoh, "The Stat family of transcription factors have diverse roles in mammary gland development," *Seminars in Cell and Developmental Biology*, vol. 19, no. 4, pp. 401–406, 2008.
- [35] Y. Li, Y. Chang, L. Zhang et al., "High glucose upregulates pantothenate kinase 4 (PanK4) and thus affects M2-type pyruvate kinase (Pkm2)," *Molecular and Cellular Biochemistry*, vol. 277, no. 1-2, pp. 117–125, 2005.
- [36] D. Horst, S. Ustanina, C. Sergi et al., "Comparative expression analysis of Pax3 and Pax7 during mouse myogenesis," *International Journal of Developmental Biology*, vol. 50, no. 1, pp. 47–54, 2006.
- [37] H. Fang, W. Tong, R. Perkins et al., "Bioinformatics approaches for cross-species liver cancer analysis based on microarray gene expression profiling," *BMC Bioinformatics*, vol. 6, no. 2, article S6, 2005.
- [38] M. D. Aupperlee, A. A. Drolet, S. Durairaj, W. Wang, R. C. Schwartz, and S. Z. Haslam, "Strain-specific differences in the mechanisms of progesterone regulation of murine mammary gland development," *Endocrinology*, vol. 150, no. 3, pp. 1485–1494, 2009.
- [39] C. J. Watson and W. T. Khaled, "Mammary development in the embryo and adult: a journey of morphogenesis and commitment," *Development*, vol. 135, no. 6, pp. 995–1003, 2008.
- [40] M. S. Hestand, M. van Galen, M. P. Villerius, G.-J. B. van Ommen, J. T. den Dunnen, and P. A. C. Hoen, "CORE.TF: a user-friendly interface to identify evolutionary conserved transcription factor binding sites in sets of co-regulated genes," *BMC Bioinformatics*, vol. 9, article 495, 2008.
- [41] C. F. Singer, G. Hudelist, W. Lamm, R. Mueller, K. Czerwenka, and E. Kubista, "Expression of tyrosine kinases in human malignancies as potential targets for kinase-specific inhibitors," *Endocrine-Related Cancer*, vol. 11, no. 4, pp. 861–869, 2004.
- [42] S. C. P. Renn, N. Aubin-Horth, and H. A. Hofmann, "Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray," *BMC Genomics*, vol. 5, article 42, 2004.
- [43] C. Bar-Or, M. Bar-Eyal, T. Z. Gal, Y. Kapulnik, H. Czosnek, and H. Koltai, "Derivation of species-specific hybridization-like knowledge out of cross-species hybridization results," *BMC Genomics*, vol. 7, article 110, 2006.
- [44] J. Adjaye, R. Herwig, D. Herrmann et al., "Cross-species hybridisation of human and bovine orthologous genes on high density cDNA microarrays," *BMC Genomics*, vol. 5, article 83, 2004.
- [45] M. D. Kane, T. A. Jatko, C. R. Stumpf, J. Lu, J. D. Thomas, and S. J. Madore, "Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays," *Nucleic Acids Research*, vol. 28, no. 22, pp. 4552–4557, 2000.
- [46] K. A. Frazer, L. Elnitski, D. M. Church, I. Dubchak, and R. C. Hardison, "Cross-species sequence comparisons: a review of methods and available resources," *Genome Research*, vol. 13, no. 1, pp. 1–12, 2003.