Chinese Journal of Cancer

Original Article

Differential expression of genes involved in the epigenetic regulation of cell identity in normal human mammary cell commitment and differentiation

Danila Coradini[1], Patrizia Boracchi[1], Saro Oriana[2], Elia Biganzoli[1] and Federico Ambrogi[1]

Abstract

The establishment and maintenance of mammary epithelial cell identity depends on the activity of a group of proteins, collectively called maintenance proteins, that act as epigenetic regulators of gene transcription through DNA methylation, histone modification, and chromatin remodeling. Increasing evidence indicates that dysregulation of these crucial proteins may disrupt epithelial cell integrity and trigger breast tumor initiation. Therefore, we explored *in silico* the expression pattern of a panel of 369 genes known to be involved in the establishment and maintenance of epithelial cell identity and mammary gland remodeling in cell subpopulations isolated from normal human mammary tissue and selectively enriched in their content of bipotent progenitors, committed luminal progenitors, and differentiated myoepithelial or differentiated luminal cells. The results indicated that, compared to bipotent cells, differentiated myoepithelial and luminal subpopulations were both characterized by the differential expression of 4 genes involved in cell identity maintenance: *CBX6* and *PCGF2*, encoding proteins belonging to the Polycomb group, and *SMARCD3* and *SMARCE1*, encoding proteins belonging to the Trithorax group. In addition to these common genes, the myoepithelial phenotype was associated with the differential expression of *HDAC1*, which encodes histone deacetylase 1, whereas the luminal phenotype was associated with the differential expression of *SMARCA4* and *HAT1*, which encode a Trithorax protein and histone acetylase 1, respectively. The luminal compartment was further characterized by the overexpression of *ALDH1A3* and *GATA3*, and the down-regulation of *NOTCH4* and *CCNB1*, with the latter suggesting a block in cell cycle progression at the G_2 phase. In contrast, myoepithelial differentiation was associated with the overexpression of *MYC* and the down-regulation of *CCNE1*, with the latter suggesting a block in cell cycle progression at the G_1 phase.

Key words Mammary epithelial cell identity, epigenetic regulation, maintenance proteins, normal bipotent cells, committed luminal progenitors, differentiated myoepithelial/luminal cells

The mammary epithelium has been traditionally described as composed of polarized luminal and basally positioned myoepithelial cells organized to form lobulo-alveolar structures. According to the most recognized model, mature epithelial cells are generated from self-renewing undifferentiated mammary stem cells through a hierarchical process generating cell progeny that progressively differentiates into luminal and myoepithelial cells[1]. While mouse mammary stem cells are limited to the peripheral cap cells of terminal end buds, the basic structure of the human mammary gland is the terminal duct lobular unit, which includes the so-called breast epithelial stem cell zone. Essentially absent in lobules, stem cell zones are focally localized in ducts and contain scattered single stem cells or small clusters of cells composed of the progeny from a single progenitor[2].

Different from many other organs, the developmental program of the mammary gland is completed only postnatally when extensive

Authors' Affiliations: [1]Department of Clinical Sciences and Community Health, Medical Statistics, Biometry and Bioinformatics, University of Milan, Via Vanzetti 5, Milan 20133, Italy; [2]Senology Center, Ambrosiana Clinic, Cesano Boscone, Milan 20090, Italy.

Corresponding Author: Danila Coradini, Department of Clinical and Community Health Sciences, Medical Statistics, Biometry and Bioinformatics, University of Milan, Via Vanzetti 5, Milan 20133, Italy. Email: danila.coradini@gmail.com.

doi: 10.5732/cjc.014.10066

mammary gland remodeling occurs in response to ovarian hormones. Terminal differentiation is associated with the acquisition of a heritable tissue-specific cell identity that depends on the differential expression of a predetermined set of genes. The expression of these genes is under the epigenetic control of some proteins, collectively termed maintenance proteins[3], which act essentially through DNA methylation, histone modification, and chromatin remodeling[4]. The best characterized maintenance proteins belong to the Trithorax and Polycomb groups[5]. Organized as multifactor complexes, Trithorax and Polycomb proteins control the transcription of many genes involved in stem cell renewal, pluripotency, and embryonic development[6-8]. Consequently, dysregulation in the expression and activity of Trithorax and Polycomb group components may have dramatic effects on cell identity and trigger pathologic alterations, including neoplastic transformation. Hence, elucidating the interrelations among the genes involved as regulators in the establishment and maintenance of mammary epithelial cell identity is of paramount importance for identifying the crucial elements that, when altered, may cause neoplastic transformation and lead to breast cancer. To this end, by making use of a publicly accessible microarray dataset[9], we explored the expression pattern of a panel of genes known to be involved in the control of cell identity and mammary gland remodeling in cell subpopulations isolated from normal human mammary tissue and selectively enriched in their content of bipotent progenitors, committed luminal progenitors, and differentiated myoepithelial or differentiated luminal cells.

Materials and Methods

Cell samples

As reported by Raouf *et al*.[9], cell subpopulations were isolated from three different reduction mammoplasty samples. Gene expression was determined using the U133-X3P Human GeneChip oligo-based microarray (Affymetrix, Santa Clara, CA, USA), and the corresponding microarray dataset was publicly available at the ArrayExpress website (http://ebi.ac.uk/arrayexpress/) under accession number E-GEOD-11395.

Gene set selection

After an extensive literature review, we established a panel of 369 genes including genes associated with the establishment and maintenance of cell identity, and mammary gland remodeling as well as genes involved in cell fate decision, cell growth control, cell polarity and adhesion, and steroid and TGF-β signaling. Of these 369 genes, 8 had no corresponding probe sets on U133-X3P GeneChip. Therefore, the gene set was actually composed of 361 elements for 842 Affymetrix probe sets, as verified by GeneAnnot system v2.2 (http://bioinfo2.weizmann.ac.il/geneannot/), which additionally provided information about the quality of connection between each probe set and the corresponding gene in terms of sensitivity and specificity[10]. Specifically, the sensitivity, defined as the fraction of probes in a probe set that match a respective gene, is the number of matching probes in a given probe set divided by the total number of probes in this probe set. The specificity indicates to what extent the probes of a probe set bind to genes, and it sums up the number of matching probes while giving lower weight to probes that match additional genes, divided by the total number of probes that match any gene.

Statistical analysis

Because some genes are recognized by more than a single probe set, each of which is characterized by an individual specificity and sensitivity that differently contribute to the gene expression value, prior to the analysis, a mean gene expression value was calculated after weighting each probe set for its own sensitivity and specificity. Namely, each gene expression value was multiplied by the semi sum of the sensitivity and specificity of the corresponding probe set. The differential gene expression between bipotent progenitors and differentiated basal, committed luminal, or differentiated luminal cells and between committed and differentiated luminal subpopulations was evaluated using an analysis of variance with pairing within each mammoplasty sample. To correct for multiple testing, the false discovery rate (FDR) with a cut-off of 10% was used[11]. All analyses were performed using the open source software R 2.11.1 package HDMD (http://www.R-project.org).

Results

Genes differentially expressed with an FDR < 0.1 in the total model

Of the 361 genes selected for the study, only 2 (*ALDH5A1* and *CCNE1*) were proven to be differentially expressed with an FDR < 0.1 in the total model, i.e., considering bipotent progenitors, committed luminal, differentiated luminal, and differentiated basal cell subpopulations together. In particular, the expression of *ALDH5A1*, which encodes aldehyde dehydrogenase 5 family, member A1, was significantly increased in committed luminal cells, whereas the expression of *CCNE1*, which encodes Cyclin E1, was dramatically decreased in differentiated basal cells.

Genes differentially expressed with a $P < 0.05$ (single model) in the basal and/or luminal compartment

In addition to *ALDH5A1* and *CCNE1*, 41 genes were proved to be differentially expressed in the basal and/or luminal compartment compared with bipotent progenitors (**Table 1**). Of these, 11 genes were differentially expressed (of note, in a similar manner) both in the basal and luminal compartments, 10 were differentially expressed exclusively in the basal cell subpopulation, and 20 were differentially expressed exclusively in the luminal compartment.

Genes differentially expressed in both compartments

The set of genes differentially expressed in both compartments

Table 1. Genes differentially expressed ($P < 0.05$) in the basal and/or luminal compartment

Gene symbol	Gene name	Basal compartment		Luminal compartment	
		Fold change (95% CI)	*P*	Fold change (95% CI)	*P*
Common between the basal and luminal compartments					
CBX6	Chromobox homolog 6	–4.77 (–7.89; –1.64)	0.010	–5.28 (–8.40; –2.16)	0.006
ID3	Inhibitor of DNA-binding 3	1.86 (0.55; 3.17)	0.013	–1.63 (–2.94; –0.32)	0.023
ID4	Inhibitor of DNA-binding 4	1.22 (0.44; 1.20)	0.009	–0.94 (–1.72; –0.16)	0.026
JAM3	Junctional adhesion molecule 3	–1.78 (–3.14; –0.42)	0.019	–2.19 (–3.56; –0.83)	0.008
MKI67	Antigen identified by monoclonal antibody Ki-67	–3.49 (–5.57; –1.41)	0.006	–2.11 (–4.19; –0.03)	0.048
NOTCH3	Notch homolog 3	0.86 (0.03; 1.70)	0.044	1.17 (0.34; 2.00)	0.014
PCGF2	Polycomb group ring finger 2	–1.81 (–3.08; –0.55)	0.013	–1.34 (–2.61; –0.08)	0.041
PFN2	Profilin 2	3.09 (0.36; 5.82)	0.032	2.86 (0.13; 5.59)	0.043
SMARCAD3	SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily d, member 3	–4.58 (–6.99; –2.16)	0.004	–3.86 (–6.28; –1.44)	0.008
SMARCE1	SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily e, member 1	2.22 (1.05; 3.39)	0.004	1.19 (0.02; 2.36)	0.047
TGFBR1	Transforming growth factor beta receptor 1	1.95 (0.90; 3.00)	0.004	1.63 (0.58; 2.68)	0.009
Exclusive for the basal compartment					
CADM1	Cell adhesion molecule 1	1.19 (0.14; 2.25)	0.032	–0.61 (–1.67; 0.44)	0.205
CCNE1	Cyclin E1	–4.38 (–5.64; –3.11)	0.0002	0.06 (–1.20; 1.33)	0.907
CLDN1	Claudin 1	1.49 (0.14; 2.85)	0.036	0.21 (–1.14; 1.56)	0.717
CRB3	Crumbs homolog 3	–2.84 (–4.24; –1.45)	0.002	–0.54 (–1.94; 0.86)	0.380
DPY30	Dpy-30 homolog (*C. elegans*)	1.30 (0.07; 2.53)	0.041	0.38 (–0.85; 1.61)	0.477
HDAC1	Histone deacetylase 1	–2.08 (–3.02; –1.15)	0.002	0.52 (–0.42; 1.45)	0.223
MFGE8	Milk fat globule-EGF factor 8	1.47 (0.07; 2.88)	0.043	0.99 (–0.41; 2.40)	0.134
MLL4	Myeloid/lymphoid or mixed-lineage leukemia 4	–1.96 (–2.84; –1.09)	0.002	0.32 (–0.55; 1.19)	0.405
MME	Membrane metallo-endopeptidase	0.82 (0.04; 1.59)	0.042	–0.01 (–0.79; 0.77)	0.976
MYC	v-Myc myelocytomatosis viral oncogene homolog (avian)	3.59 (0.17; 7.01)	0.043	0.11 (–3.31; 3.53)	0.938
PRKCZ	Protein kinase C, zeta	–2.58 (–5.11; –0.06)	0.047	0.42 (–2.10; 2.95)	0.696
Exclusive for the luminal compartment					
ALDH1A3	Aldehyde dehydrogenase 1 family, member A3	1.02 (–3.91; 5.94)	0.631	5.60 (0.68; 10.52)	0.032
ALDH5A1	Aldehyde dehydrogenase 5 family, member A1	–0.22 (–0.83; 0.39)	0.411	2.21 (1.60; 2.82)	0.0001
CCNB1	Cyclin B1	–0.21 (–1.27; 0.86)	0.652	–2.30 (–3.36; –1.23)	0.002
CLDN3	Claudin 3	0.08 (–3.86; 4.02)	0.962	–3.94 (–7.88; –0.01)	0.050
EMP1	Epithelial membrane protein 1	1.98 (–0.38; 4.34)	0.085	2.84 (0.49; 5.20)	0.026
EPCAM	Epithelial cell adhesion molecule	2.33 (–0.25; 4.90)	0.069	2.64 (0.06; 5.21)	0.046
GATA3	GATA-binding protein 3	0.31 (–0.63; 1.25)	0.445	1.30 (0.36; 2.24)	0.015
HAT1	Histone acetyltransferase 1	1.07 (–1.42; 3.56)	0.334	2.54 (0.05; 5.03)	0.047
HES6	Hairy and enhancer of split 6	0.46 (–0.33; 1.24)	0.204	0.78 (0.03; 1.57)	0.049
ITGB4	Integrin, beta 4	–0.96 (–2.62; 0.69)	0.205	–1.76 (–3.42; –0.10)	0.041
JAK3	Janus kinase 3	–1.34 (–2.94; 0.26)	0.087	–1.95 (–3.56; –0.35)	0.025
MUC1	Mucin 1, cell surface associated	0.63 (–2.19; 3.45)	0.604	2.91 (0.09; 5.73)	0.045
NOTCH4	Notch homolog 4	–0.74 (–4.79; 3.31)	0.669	–4.33 (–8.38; –0.28)	0.040
SMARCA4	SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily a, member 4	–0.17 (–0.97; 0.63)	0.623	0.94 (0.14; 1.75)	0.028
SOX10	SRY (sex-determining region Y)-box 10	0.21 (–0.98; 1.40)	0.681	1.40 (0.21; 2.59)	0.028
STAT5B	Signal transducer and activator of transcription 5B	–0.72 (–2.60; 1.16)	0.385	1.91 (0.03; 3.79)	0.047
TGFB3	Transforming growth factor, beta 3	0.02 (–0.53; 0.56)	0.943	1.40 (0.85; 1.95)	0.001

(*To be continued*)

Table 1. Genes differentially expressed ($P < 0.05$) in the basal and/or luminal compartment (*continued*)

Gene symbol	Gene name	Basal compartment		Luminal compartment	
		Fold change (95% CI)	*P*	Fold change (95% CI)	*P*
THY1	Thy-1 cell surface antigen	−3.03 (−8.43; 2.38)	0.220	−5.75 (−11.2; −0.34)	0.041
TIAM1	T-cell lymphoma invasion and metastasis	0.35 (−1.01; 1.70)	0.553	1.54 (0.19; 2.89)	0.031
TJP3	Tight junction protein 3	0.60 (−2.63; 3.83)	0.664	3.44 (0.21; 6.67)	0.040
VIM	Vimentin	−0.98 (−3.19; 1.23)	0.320	−3.10 (−5.31; −0.89)	0.014

CI, confidence interval.

included some elements involved in the activation of cell differentiation (*TGFBR1, ID3, ID4,* and *NOTCH3*) and the epigenetic control of gene transcription (*CBX6, PCGF2, SMARCD3,* and *SMARCE1*) or coding for proteins associated with cell morphology (*PFN2* and *JAM3*) and cell proliferation (*MKI67*) (**Figure 1**).

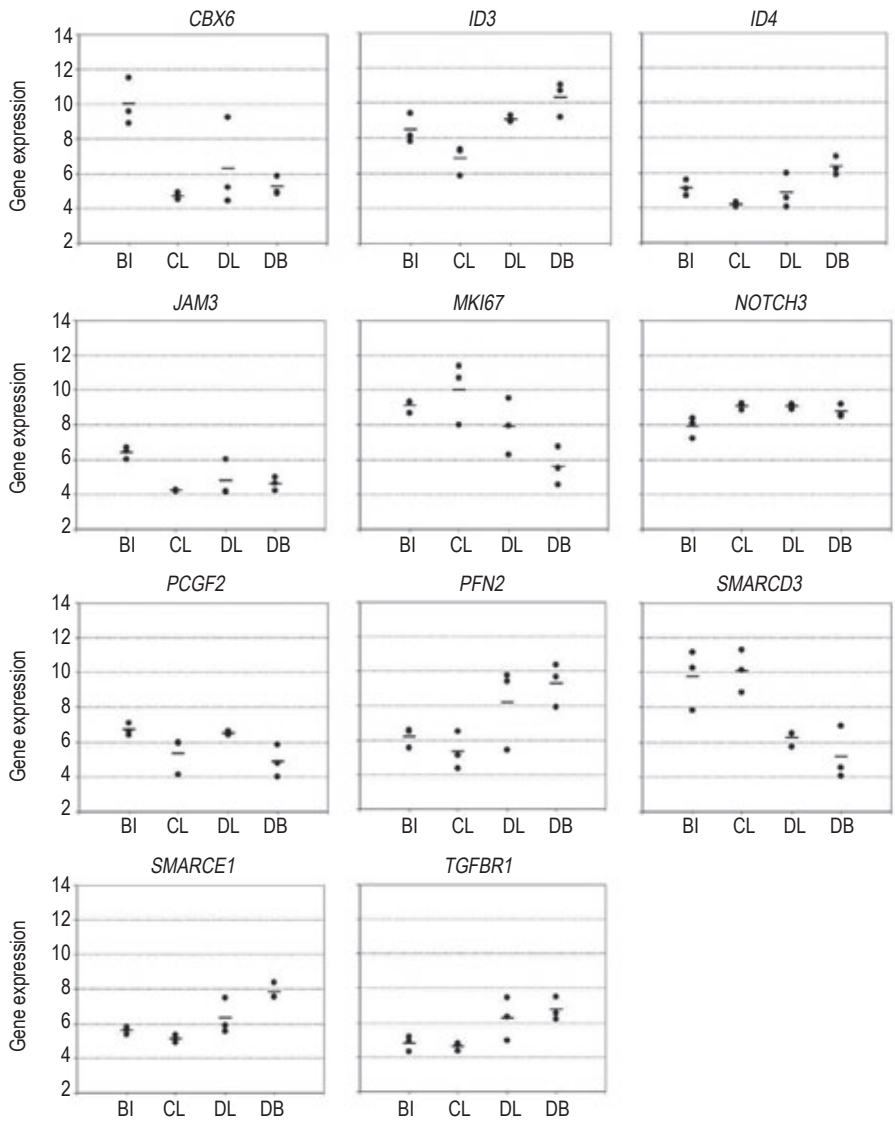

Figure 1. **Scatter plots of the 11 genes found differentially expressed in both the basal and luminal compartments compared with bipotent cells. BI, bipotent progenitors; CL, committed luminal cells; DL, differentiated luminal cells; DB, differentiated basal cells.**

Genes differentially expressed in the basal compartment

The set of genes associated with the basal subpopulation in an exclusive manner included elements encoding proteins involved in cell fate decision (*MFGE8* and *MYC*), the epigenetic control of gene transcription (*DPY30, HDAC1,* and *MLL4*), cell cycle progression (*CCNE1*), cell polarity (*CRB3* and *PRKCZ*), adhesion (*CADM1* and *CLDN1*), and morphology (*MME*) (**Figure 2**).

Genes differentially expressed in the luminal compartment

The set of genes specifically associated with the luminal subpopulation included elements coding for proteins involved in the TGF-β signaling pathway (*TGFB3*), cell fate decision (*ALDH1A3, ALDH5A1, HES6, NOTCH4,* and *SOX10*), the epigenetic control of gene transcription (*HAT1* and *SMARCA4*), cell cycle progression (*CCNB1*), cell adhesion (*CLDN3, ITGB4, TJP3,* and *TIAM1*) and morphology (*EMP1, EPCAM, GATA3, MUC1,* and *VIM*), and cytokines (*JAK3, STAT5B,* and *THY1*) (**Figure 3**). Consistent with the notion that mature basal cells have a pattern of expression similar to bipotent precursors, the luminal subpopulation showed a greater number of differentially expressed genes compared to the basal subpopulation.

Volcano plots

To allow a quick visual identification of the differentially expressed genes, we built a series of volcano plots combining, for each gene,

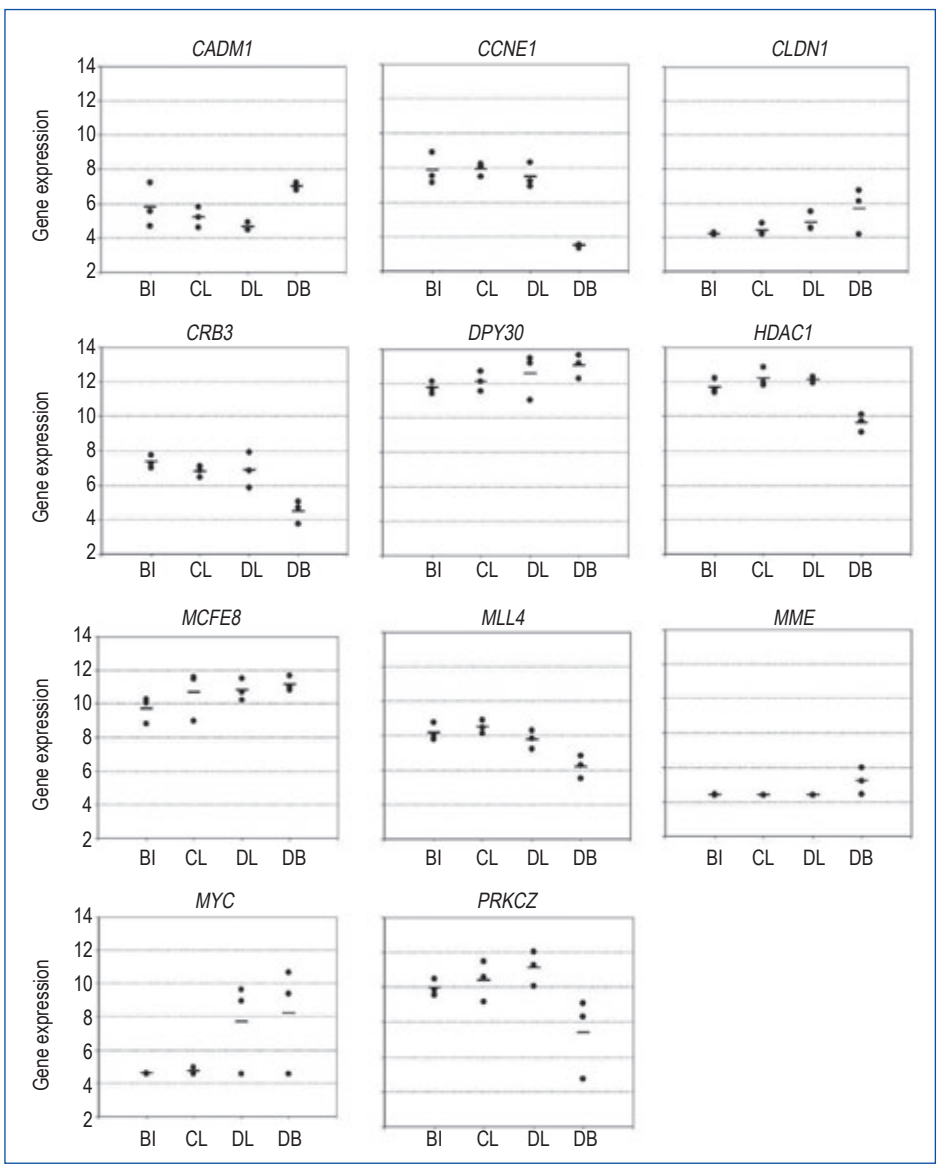

Figure 2. **Scatter plots of the genes found to be differentially expressed between differentiated basal cells and bipotent progenitors. Abbreviations as in the legends of Figure 1.**

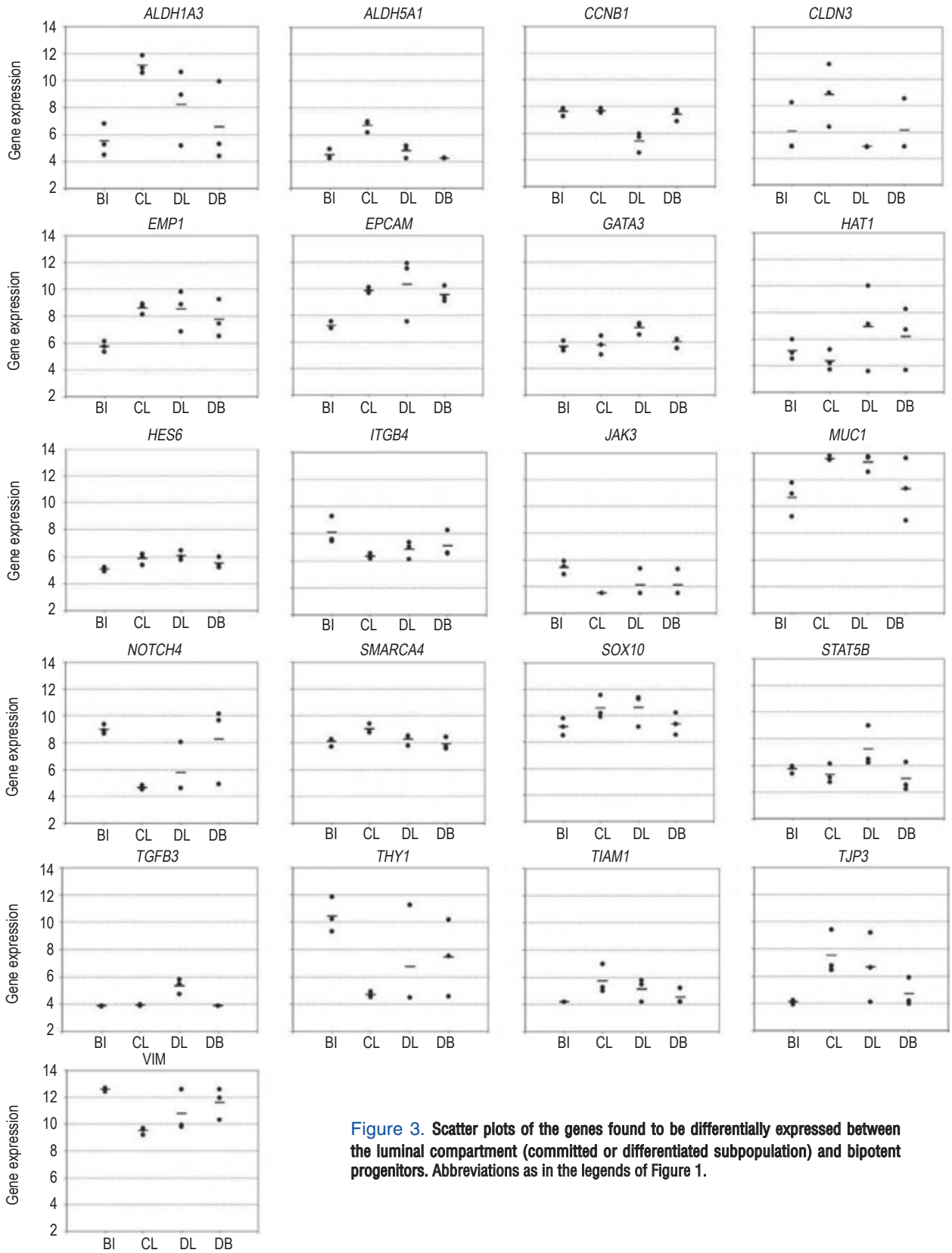

Figure 3. **Scatter plots of the genes found to be differentially expressed between the luminal compartment (committed or differentiated subpopulation) and bipotent progenitors. Abbreviations as in the legends of Figure 1.**

the statistical significance with the magnitude of its change compared to a reference class (expressed as fold change). **Figure 4** shows three of these volcano plots: the upper and middle panels illustrate the genes differentially expressed between differentiated basal or committed luminal cells and bipotent progenitors, whereas the lower panel illustrates the genes differentially expressed between committed and differentiated luminal subpopulations. To avoid misleading conclusions, we considered as biologically significant only genes having a fold change higher than 2. According to such a criterion, with respect to bipotent progenitors, mature basal cells were represented by 10 genes, 5 of which (*CBX6, MKI67, PFN2, SMARCAD3,* and *SMARCE1*) were in common with the luminal subpopulation, suggesting their modulation is associated with epithelial differentiation rather than with basal- or luminal-restricted commitment. With regard to the exclusive genes (*CCNE1, CRB3, HDAC1, MYC,* and *PRKCZ*), all but MYC were proven to be underexpressed, with a special mention for the dramatic down-regulation of CCNE1 (**Figure 2**). As shown in **Figure 4**, 11 genes characterized committed luminal cells (middle panel); an additional 7 genes were specifically associated with differentiated luminal cells (lower panel). Interestingly, the volcano plots indicated that, excluding the genes specifically associated with the luminal phenotype (namely, the overexpressed *EMP1, EPCAM, MUC1, and TJP3* and the underexpressed *VIM* and *THY1*), the committed cells were characterized by the overexpression of *ALDH1A3* and *ALDH5A1* and the underexpression of *NOTCH4*, and the differentiated cells were characterized by the overexpression of *HAT1* and the underexpression of *CCNB1* and *CLDN3*.

Discussion

We performed this *in silico* study to specifically explore the expression pattern of genes involved in the control of cell identity and mammary gland remodeling in normal bipotent progenitors, committed luminal cells, and differentiated basal and luminal cells isolated from reduction mammoplasty samples. Although the study suffers from the limitation common to the majority of studies involving only gene expression profiling, i.e., a lack of validation of the observed mRNA modulations at the protein level, our results provide interesting suggestions about the relationships among the genes that define epithelial cell identity and those involved in lineage differentiation. In particular, the findings indicate a portion of genes as common to the basal and luminal compartments, with each cell subpopulation being represented by an exclusive set of genes.

Of note, among the genes differentially expressed in both compartments (**Figure 1**), *TGFBR1*, *ID3*, *ID4*, and *NOTCH3* were identified, all of which are known to be involved in the activation of the differentiation program. *TGFBR1*, *ID3*, and *ID4*, in fact, encode TGF-β receptor 1 and two transcription factors involved in the canonical TGF-β/Smad signaling pathway, whereas *NOTCH3* encodes a single-pass type I membrane protein having a receptor function for several membrane-bound ligands (*Jagged 1*, *Jagged 2,* and *Delta 1*) during the cell fate determination of stem/progenitor cells in mammary differentiation[12]. Interestingly, although *ID3* and *ID4* expression levels were decreased in committed luminal cells, in keeping with the

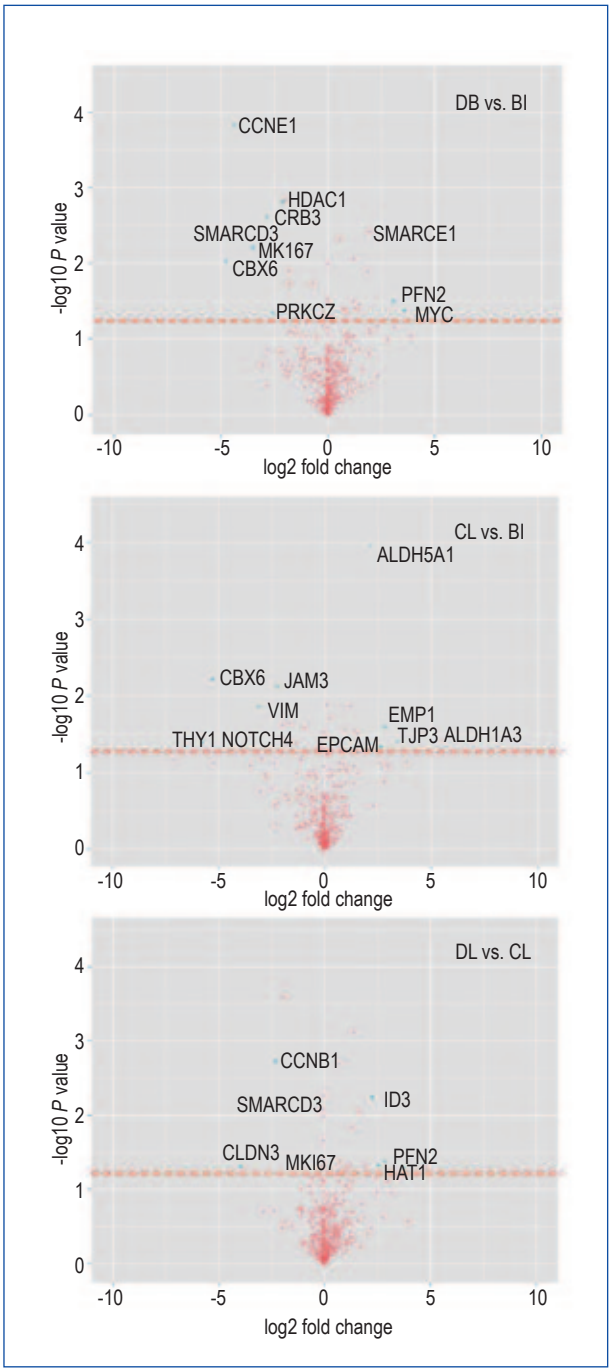

Figure 4. **Volcano blots of the genes differentially expressed between differentiated basal cells and bipotent progenitors (upper panel), between committed luminal cells and bipotent progenitors (medium panel), or between differentiated luminal and committed luminal cells (lower panel). Abbreviations as in the legends of Figure 1. Combining the statistical significance with the magnitude of the change (expressed as fold change), the volcano plot enables a quick visual identification of the genes differentially expressed. The dashed red line shows where the *P* = 0.05. Genes having a fold change higher than 2 are depicted in blue.**

established role of inhibitor of DNA-binding (ID) proteins in enforcing the undifferentiated state of embryonic and somatic stem cells[13,14], their expression levels increased in mature cells, which is in line with the tumor-suppressor activity exerted by some ID proteins, especially ID4[15].

Consistent with their common epithelial commitment, both compartments showed the overexpression of *PFN2*, which encodes Profilin 2, a structural protein involved in cytoskeleton organization, and the underexpression of *JAM3*, which codes for a junctional adhesion molecule specific to high endothelial cells. In addition, in agreement with the association between cell cycle arrest and differentiation, both compartments exhibited the underexpression of *MKI67*, which codes for the cell proliferation marker Ki-67.

Regarding genes coding for cell identity maintenance proteins, 4 genes were proven to be differentially expressed in both compartments: *CBX6* and *PCGF2*, belonging to the Polycomb group, and *SMARCD3* and *SMARCE1*, belonging to the Trithorax group. Genetic studies have established that, in humans, Polycomb proteins are organized into two distinct complexes: polycomb repressive complex (PRC)-1 and PRC-2, which act cooperatively as transcriptional repressors. In particular, PRC-2 first catalyzes the trimethylation of histone H3 at lysine 27, and then PRC-1 promotes the monoubiquitination of histone H2A at lysine 119, blocking transcriptional elongation[16]. Consistent with this repressive process, *CXB6* and *PCGF2* were underexpressed in differentiated cells, thereby allowing the transcription of the genes involved in cell differentiation. In contrast, Trithorax proteins can act as transcriptional activators or repressors depending on the specific context and on the multiprotein complex in which they are included[17]. Hence, it is not surprising that we observed an apparent contrasting expression pattern of *SMARCD3* (underexpressed) and *SMARCE1* (overexpressed), which code for 2 regulatory proteins, Baf60c and Baf57. Recent studies have demonstrated that Baf57 depletion caused the transcriptional misregulation of various cell cycle-related genes[18], whereas the stable re-expression of the gene induced cell cycle arrest and apoptosis in breast cancer cells[19]. These findings suggest a role for Baf57 as a tumor suppressor. The present findings support this hypothesis because we found that the overexpression of *SMARCE1* was associated with the underexpression of *MKI67*.

With regard to genes differentially expressed in the basal compartment, our results indicate that normal myoepithelial differentiation is associated with the block of cell progression at the G_1 phase of the cell cycle, as suggested by the down-regulation of *CCNE1*. Cyclin E1, which is involved in the control of cell cycle progression at the G_1/S boundary[20], activates Cyclin-dependent kinase (Cdk)2 to overcome the G_1/S restriction point. Conversely, low expression of Cyclin E1 prevents cell cycle progression and mitosis. As high levels of Cyclin E have been associated with basal-like breast cancer and a poor outcome[21], the present findings support the attractive hypothesis that the constitutive reactivation of *CCNE1* expression might represent a crucial event in basal-like cancer development and progression. In apparent contrast with *CCNE1* silencing, we observed the overexpression of *MYC*, which encodes a transcription factor involved in cell proliferation and is known as potent activator of tumorigenesis in a variety of cancers, including breast cancer[22]. Specifically, Myc controls the G_1/S transition through the activation of the Cyclin E/Cdk2 complex, which in turn represses the expression of Cdk inhibitor $P21^{Cip1}$. This apparently paradoxical association between *MYC* overexpression and *CCNE1* underexpression is however consistent with the proposed context-dependent role of Myc in normal mammary epithelium. In fact, Myc can promote mammary epithelial terminal differentiation or cell proliferation according to the developmental stage[23]. The present findings support this stage-dependent role of *MYC* in promoting terminal differentiation: although silenced in bipotent progenitors, it was proven to be overexpressed in differentiated cells.

Regarding genes coding for cell identity maintenance proteins, in addition to the above descriptions, mature basal cells were found to underexpress *HDAC1*, which codes histone deacetylase 1, an enzyme able to induce histone modification via the deacetylation of lysine residues on histone tails. Because histone deacetylase 1 promotes gene silencing when associated with Baf60c and Baf57, the observation that *HDAC1* is underexpressed in differentiated cells, thus allowing gene transcription, is not surprising.

As opposed to basal cells, in which cell cycle arrest is associated with *CCNE1* down-regulation, our results indicate that luminal terminal differentiation is associated with the down-regulation of *CCNB1*, which encodes Cyclin B1, the Cyclin involved in the control of cell cycle progression at the G_2/M boundary. Furthermore, the observation that, in contrast to mature cells, committed luminal cells constitutively express high levels of *CCNB1* (similar to bipotent cells, **Figure 3**) suggests that attaining luminal differentiation specifically requires the block of cell cycle progression before entering mitosis. Interestingly, the underexpression of *CCNB1*, which was observed in mature luminal cells, accounts for the clinical evidence that Cyclin B1 overexpression in breast cancer is associated with aggressive tumors (high tumor grade, large tumor size, and high metastasis probability) and a poor prognosis[24].

Consistent with cell cycle arrest and terminal luminal differentiation [supported by the differential expression of several genes involved in cell adhesion (*CLDN3*, *ITGB4*, *TJP3*, and *TIAM1*) and morphology (*EMP1*, *EPCAM*, *GATA3*, *MUC1*, and *VIM*)], mature cells overexpressed *HAT1*, which encodes histone acetyltransferase 1, the enzyme able to acetylate lysine residues on histone tails to promote the relaxation of chromatin condensation and allow gene transcription.

Of note, compared to mature cells, committed luminal cells were characterized by the differential expression of 3 genes involved in cell fate decision: *ALDH1A3*, *ALDH5A1* (both overexpressed), and *NOTCH4* (underexpressed). *ALDH1A3* and *ALDH5A1* code for 2 isoenzymes belonging to the family of aldehyde dehydrogenases (ALDHs) but with different functions. ALDH1A3 is a retinaldehyde dehydrogenase that catalyzes the biosynthesis of retinoic acid (a strong differentiation inducer), whereas ALDH5A1 is a mitochondrial NAD(+)-dependent succinic semialdehyde dehydrogenase. ALDH family members have been extensively investigated and are involved in several processes, including development; *ALDH1A3*, in particular, has been shown to increase at the point of progenitor

cell commitment to the luminal lineage[25]. Consistent with this experimental evidence, our results indicate that although *ALDH1A3* expression is low in bipotent progenitors, the gene shows a sharp increase in committed luminal cells (**Figure 3**). Interestingly, the finding also supports the hypothesis that ALDH1A3 activates the luminal cell differentiation program as part of the retinoid signaling pathway in concert with Gata-3, the essential regulator of mammary gland morphogenesis and luminal cell differentiation [26], which we found to be overexpressed in differentiated luminal cells. ALDH1A3, in fact, catalyzes the biosynthesis of retinoic acid in mammary cells[27], and experimental studies have demonstrated that approximately two-thirds of human genes with binding sites for Gata-3 contain retinoic acid regulatory elements[28]. Regarding the down-regulation of *NOTCH4*, which is known to be critical for maintaining stem/progenitor cell activity[29], the present findings corroborate the notion of a diverse functional role of Notch receptors in the luminal lineage restriction process[30], as indicated by the complementary pattern of *NOTCH3* and *NOTCH4* expression.

It is noteworthy that the present findings also corroborate two significant aspects of luminal/basal commitment and differentiation: the co-presence of both luminal and basal cytokeratins in undifferentiated and primitive luminal/myoepithelial progenitors and the lack of estrogen receptors in differentiated luminal cells. In fact, in agreement with the current opinion suggesting that the phenotype-restricted expression of cytokeratins is a late step in the process of lineage-specific differentiation[31], we found that all cell subpopulations, bipotent progenitors included, constitutively expressed high levels of both basal- (*KRT5*, *KRT6B*, *KRT14*, *KRT16*, and *KRT17*) and luminal-associated (*KRT7*, *KRT8*, *KRT18*, and *KRT19*) cytokeratins. Furthermore, consistent with the notion that estrogen receptor expression is a terminal event in luminal-specific differentiation, the *ESR1* gene was found to be silenced in all subpopulations.

Conclusions

Although preliminary, the present findings indicated that the commitment and differentiation of normal human mammary cells are associated with the differential expression of a narrow panel of genes coding for some regulatory proteins belonging to the Polycomb and Trithorax groups or that are involved in the epigenetic control of gene transcription. In addition, the results provide evidence that basal differentiation is associated with the block of cell progression before entering mitosis, whereas luminal differentiation is associated with blockage at the G_1 phase, during which Gata-3, the essential regulator of mammary gland morphogenesis, promotes and governs the expression of a plethora of genes involved in the establishment and maintenance of the luminal phenotype.

Received: 2014-05-05; revised: 2014-07-12;
accepted: 2014-07-17.

References

[1] Villadsen R, Fridriksdottir AJ, Rønnov-Jessen L, et al. Evidence for a stem cell hierarchy in the adult human breast. J Cell Biol, 2007,177:87–101.

[2] Boecker W, Moll R, Poremba C, et al. Common adult stem cells in the human breast give rise to glandular and myoepithelial cell lineages: a new cell biological concept. Lab Invest, 2002,82:737–746.

[3] Brock HW, Fisher CL. Maintenance of gene expression patterns. Dev Dynamics, 2005,232:633–655.

[4] Vermaak D, Ahmad K, Henikoff S. Maintenance of chromatin states: an open-and-shut case. Curr Opin Cell Biol, 2003,15:266–274.

[5] Ringrose L, Paro R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu Rev Genet, 2004,38:413–443.

[6] Kennison JA. The Polycomb and Trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. Annu Rev Genet, 1995,29:289–303.

[7] Lee TI, Jenner RG, Boyer LA, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell, 2006,125:301–313.

[8] Jacobs JJ, van Lohuizen M. Polycomb repression: from cellular memory to cellular proliferation and cancer. Biochim Biophys Acta, 2002,1602:151–161.

[9] Raouf A, Zhao Y, To K, et al. Transcriptome analysis of the normal human mammary cell commitment and differentiation process. Cell Stem Cell, 2008,3:109–118.

[10] Ferrari F, Bortoluzzi S, Coppe A, et al. Novel definition files for human GeneChips based on GeneAnnot. BMC Bioinformatics, 2007,8:446.

[11] Benjamini Y, Hochberg Y. Controlling the false discovery rate. A practical and powerful approach to multiple testing. J Royal Stat Soc, 1995,57:284–300.

[12] Dontu G, Jackson KW, McNicholas E, et al. Role of Notch signaling in cell-fate determination of human mammary stem/progenitor cells. Breast Cancer Res, 2004,6:R605–R615.

[13] Nam H.S, Benezra R. High levels of Id1 expression define B1 type adult neural stem cells. Cell Stem Cell, 2009,5:515–526.

[14] Ying QL, Nichols J, Chambers I, et al. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell, 2003,115:281–292.

[15] Umetani N, Mori T, Koyanagi K, et al. Aberrant hypermethylation of ID4 gene promoter region increases risk of lymph node metastasis in T1 breast cancer. Oncogene, 2005,24:4721–4727.

[16] Spermann A, van Lohuizen M. Polycomb silencers control cell fate, development and cancer. Nat Rev Cancer, 2006,6:846–856.

[17] Johnson CN, Adkins NL, Georgel P. Chromatin remodeling complexes: ATP-dependent machines in action. Biochem Cell Biol, 2005,83:405–417.

[18] Hah N, Kolkman A, Ruhl DD, et al. A role for BAF57 in cell cycle–dependent transcriptional regulation by the SWI/SNF chromatin remodeling complex. Cancer Res, 2010,70:4402–4411.

[19] Wang L, Baiocchi RA, Pal S, et al. The BRG1- and hBRM-associated

factor BAF57 induces apoptosis by stimulating expression of the cylindromatosis tumor suppressor gene. Mol Cell Biol, 2005, 25:7953–7965.
[20] Massague J. G_1 cell-cycle control and cancer. Nature, 2004, 432:298–306.
[21] Voduc D, Nielsen TO, Cheang MC, et al. The combination of high cyclin E and Skp2 expression in breast cancer is associated with a poor prognosis and the basal phenotype. Hum Pathol, 2008,39:1431–1437.
[22] Xu J, Chen Y, Olopade OI. MYC and breast cancer. Genes Cancer, 2010,1:629–640.
[23] Blakely CM, Sintasath L, D'Cruz CM, et al. Developmental stage determines the effects of MYC in the mammary epithelium. Development, 2005,132:1147–1160.
[24] Suzuki T, Urano T, Miki Y, et al. Nuclear cyclin B1 in human breast carcinoma as a potent prognostic factor. Cancer Sci, 2007,98:644–651.
[25] Eirew P, Kannan N, Knapp DJHF, et al. Aldehyde dehydrogenase activity is a biomarker of primitive normal human mammary luminal cells. Stem Cells, 2012,30:344–348.
[26] Asselin-Labat ML, Sutherland KD, Barker H, et al. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. Nat Cell Biol, 2007,9:201–209.
[27] Rexer BN, Zheng WL, Ong DE. Retinoic acid biosynthesis by normal human breast epithelium is via aldehyde dehydrogenase 6, absent in MCF-7 cells. Cancer Res, 2001,61:7065–7070.
[28] Hua S, Kittler R, White KP. Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. Cell, 2009,137:1259–1271.
[29] Dontu G, Jackson KW, McNicholas E, et al. Role of Notch signaling in cell-fate determination of human mammary stem/progenitor cells. Breast Cancer Res, 2004,6:R605–R615.
[30] Bouras T, Pal B, Vaillant F, et al. Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. Cell Stem Cell, 2008,3:429–441.
[31] Bocker WR, Moll C, Poremba R, et al. Common adult stem cells in the human breast give rise to glandular and myoepithelial cell lineages: a new cell biological concept. Lab Invest, 2002,82:737–746.