NIH-PA Author Manuscript

# Patterns of cell signaling pathway activation that characterize mammary development

**Eran R. Andrechek**, **Seiichi Mori**, **Rachel E. Rempel**, **Jeffrey T. Chang**, and **Joseph R. Nevins**[*]
Duke Institute for Genome Sciences and Policy, Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710, USA

## Abstract

Previous work has detailed the histological and biochemical changes associated with mammary development and remodeling. We have now made use of gene expression profiling, and in particular of the previously described signatures of cell signaling pathway activation, to explore the events associated with mammary gland development. We find that there is elevated E2F-specific pathway activity prior to lactation and relatively low levels of other important signaling pathways, such as RAS, MYC and SRC. Upon lactation and continuing into the involution phase, these patterns reverse with a dramatic increase in RAS, SRC and MYC pathway activity and a decline in E2F activity. At the end of involution, these patterns return to that of the adult non-lactating mammary gland. The importance of the changes in E2F pathway activity, particularly during the proliferative phase of mammary development, was confirmed through the analysis of mice deficient for various E2F proteins. Taken together, these results reveal a complex pattern of pathway activity in relation to the various phases of mammary gland development.

### Keywords

E2F transcription factor; Mammary gland; Microarray; Mouse

## INTRODUCTION

The vast majority of mammary gland development occurs after birth, when the epithelial network undergoes a rapid expansion to fill the mammary fat pad. Subsequently, when the adult animal cycles through pregnancy, lactation and involution, there are rounds of proliferation, differentiation and apoptosis corresponding to discrete stages of development. These changes are a result of the activation of numerous genetic pathways, some of which have been well characterized through the study of individual genes and their signaling networks.

A number of groups have analyzed the transcriptional changes associated with the various developmental phases of the mammary gland (Blanchard et al., 2007; Clarkson et al., 2004; Kouros-Mehr et al., 2006; Kouros-Mehr and Werb, 2006; Master et al., 2002; McBryan et al., 2007; Rudolph et al., 2007; Stein et al., 2004). Three reports are notable for their comprehensive coverage of the spectrum of development from virgin through pregnancy, lactation and involution (Clarkson et al., 2004; Master et al., 2002; Stein et al., 2004). In

[*]Author for correspondence (nevin001@mc.duke.edu).

each of these studies, self-organizing maps were used to identify groups of genes that were elevated in the various stages of mammary development. Through annotation of gene ontology, these reports noted a significant immune response with the onset of involution. Further examination of individual genes led to additional work in adaptive thermogenesis (Master et al., 2002), on annexin A8 in breast cancer (Stein et al., 2005) and on the composition of connexin channels in the mammary epithelium (Locke et al., 2004).

In addition to these datasets that encompass pregnancy, lactation and involution, several recent reports focused on pubertal ductal outgrowth (Kouros-Mehr et al., 2006; Kouros-Mehr and Werb, 2006; McBryan et al., 2007). Using a strategy to compare terminal end buds (TEBs) with ducts, it was noted that GATA3 was highly expressed (Kouros-Mehr and Werb, 2006) and was essential for mammary development (Kouros-Mehr et al., 2006). In addition, a recent report has examined transcription throughout puberty (McBryan et al., 2007). Taken together, there are now microarray datasets available that describe normal mammary development through all major stages.

Although these previous studies identified interesting genes, an alternate strategy is to look for higher-order structure in the expression data, particularly evidence of discrete biological function associated with cell signaling pathway activity. For example, Gene Set Enrichment Analysis (GSEA) is a method that uses previously established gene sets defining pathways or functions to interpret microarray expression data (Subramanian et al., 2005). Using gene sets defined through various methods including experimental perturbations and literature-based studies, GSEA provides a pathway discovery tool. A variation on GSEA, termed ASSESS, allows a measure of enrichment of gene sets across multiple samples (Edelman et al., 2006).

An alternative approach that we have previously described makes use of supervised methods of analysis and data from the experimentally controlled activation of specific oncogenic pathways to create signatures for each oncogene (Bild et al., 2006). These signatures can then be used to measure the probability of pathway activation in a tissue or cell line sample. Here, we have applied various pathway analysis techniques to the previously described mammary gland development dataset (Stein et al., 2004) and the pubertal dataset (McBryan et al., 2007). These analyses illustrate both the differential usage of various pathways in the discrete phases of mammary gland development and the broad applicability of these methods to a variety of developmental contexts.

## MATERIALS AND METHODS

### Animal use

*E2f1* (Field et al., 1996), *E2f2* (Murga et al., 2001), *E2f3* (Humbert et al., 2000) and *E2f4* (Rempel et al., 2000) knockout mice were maintained in accordance with institutional and federal guidelines. Lactating mammary glands were examined on the fifth day of lactation. To examine involution of the mammary gland, pups were removed from the lactating female on the fifth day of lactation. Transplant experiments were conducted as previously described (Deome et al., 1959). Briefly, mammary glands were removed from the donor mice and small, 1–2 mm$^3$ portions containing epithelium were isolated. Nu/nu mice were used as recipients and their endogenous epithelium was removed by excising the entire fat pad from the nipple to the inguinal lymph node. A small hole in the remaining fat pad was made with fine forceps and the donor epithelium was inserted. The mice were examined 25 days post-surgery for outgrowth effects.

## Histology

For wholemount analysis, the inguinal mammary gland was excised and stained with Harris Modified Hematoxylin. To assess outgrowth of the mammary epithelium, the distance from the nipple to the leading edge of the epithelium was measured, as was the distance from the nipple to the midpoint to the thoracic lymph node. The ratio of the distance of outgrowth and distance between the lymph node and nipple was calculated for the control. This ratio of outgrowth for the control was set to 100% and the various mutants were compared with this standard.

Samples for histological analysis were fixed in 10% formalin and were processed using standard procedures. Sections from involuting mammary glands were also used to assess cell death in a TUNEL analysis using the In Situ Cell Death Detection Kit POD (Roche Applied Science). Rabbit polyclonal anti-E2F3 (C18, Santa Cruz, 1:1000 dilution), rabbit polyclonal anti-E2F4 (A20, Santa Cruz, 1:300), mouse monoclonal anti-PCNA (PC10, Santa Cruz Biotechnology, 1:300) and mouse monoclonal anti-α smooth muscle actin (clone 1A4, Sigma Immunochemicals, 1:1000) were used in immunohistochemical experiments. For the mouse monoclonal antibodies, the mouse-on-mouse staining procedures were followed (Vector Laboratories). Secondary antibodies and blocking reagents were from the Vectastain Elite Kits (Vector Laboratories). Signal detection used DAB (Vector Laboratories) and a Hematoxylin counterstain.

## Quantitative RT-PCR

Mammary glands were excised, snap frozen in liquid nitrogen and stored at −80°C. Total RNA was isolated by guanidinium thiocyanate extraction and CsCl gradient sedimentation (Chirgwin et al., 1979). Quantitative RT-PCR was performed using a SYBR Green One-Step RT-PCR Kit (Qiagen) with the following primers (5′ to 3′): *E2f1* forward, CGATTCTGACGTGCTGCTCT and reverse, CAGCGAGGTACTGATGGTCA; *E2f2* forward, GCGCATCTATGACATCACCA and reverse, CGGGTGGGGTCTTCAAATAG; *E2f3a* forward, CCAGCAGCCTCTACACCAC and reverse, GGTACTGATGGCCACTCTCG; *E2f3b* forward, CTTTCGGAAATGCCCCTTACA and reverse, GGTACTGATGGCCACTCTCG; *E2f4* forward, CACTGAGGACGTCCAGAACA and reverse, GATGGGCACCTCTAGACTGG; *Gapdh* forward, TCATGACCACAGTGGATGCC and reverse, GGAGTTGCTGTTGAAGTCGC. Relative levels of product were calculated using the ΔΔCt method.

## E2F3-regulated genes

Fold expression differences from involution data for day 1 and day 2 between wild-type and *E2f3* heterozygous samples were calculated using Genespring (www.chem.agileut.com/Scripts/PDS.asp?lPage=27811). All target genes with a greater than 2-fold difference were used in GATHER to identify genes with an E2F consensus binding site in their promoter. To examine genes regulated by E2F3, mouse mammary HC11 cells were cultured under growth conditions (RPMI 1640 medium with 0.3 g/l L-glutamine, 10% fetal bovine serum, 20 mM HEPES, 10 μg/ml insulin and 10 ng/ml EGF) and apoptosis was induced through serum reduction to 0.1%, insulin and growth factor withdrawl. Chromatin immunoprecipitation was performed on growing and apoptotic cells as described (Zhu et al., 2004). Quantitative PCR using the ΔΔCT method was conducted for targets using the following primers (5′ to 3′): ribonucleotide reductase M2 forward, CGGAGAGCATGGCGAACGAGG and reverse, GCTCCTTAAAGGTCTTTGTGC; albumin forward, GGACACAAGACTTCTGAAAGTCCTC and reverse, TTCCTACCCCATTACAAAATCATA; *Hexb* forward, TCGGTCATCTGACTTGGTGA and reverse, ATGACTGCTCCGCGAGTCT; *Map3k4* forward, AGTCGAGTCACTCCCTCACC and reverse, CTCTCATCCGTGCACCAAG; *Mt1*

forward, ACTATGCGTGGGCTGGAG and reverse, GGTGACGCTTAGAGGACAGC; *Grim19* (*Ndufa13*) forward, GCGATCTGCGAGAAGAGTTC and reverse, TCGAGTTCATCGAAGTGTGC; *Rhob* forward, CAATCAAGCTAAGGCGAACC and reverse, AGAGTCCGGCCACTTTCTTA; *Trp53inp1* forward, CACGTGCGTTAGTGACAACC and reverse, GGGTCCTTTTGTTGTTGTGC.

### Involution array analysis

Mammary RNA from day 5 of lactation and days 1–4 of involution was isolated (Chirgwin et al., 1979) from three wild-type and three *E2f3* heterozygous mice at each timepoint. RNA was hybridized to Operon 3.0 mouse arrays (www.operon.com/index.php?). In order to examine a phenotypic signature, two datasets containing mouse mammary involution data were downloaded (Clarkson et al., 2004; Stein et al., 2004).

### Dataset analysis

Unsupervised clustering analysis was performed with Cluster 3.0 (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/). Supervised clustering was performed with GeneCluster 2.0 (http://www.broad.mit.edu/cancer/software/genecluster2/gc2.html). GSEA was performed using GSEA 1.0 (http://www.broad.mit.edu/gsea/) (Subramanian et al., 2005). ASSESS (Edelman et al., 2006) was also employed.

### TEB/duct E2F promoter analysis

The top 200 genes, as judged by fold change amongst genes expressed more highly in the TEB than in the duct, were selected (Kouros-Mehr et al., 2006; Kouros-Mehr and Werb, 2006). These genes were used in GATHER [http://meddb01.duhs.duke.edu/gather/ (Chang and Nevins, 2006)] to predict transcription factor binding sites.

### Pathway signatures

Genomic pathway signatures were generated as previously described (Bild et al., 2006). Gene sets used in this analysis were described previously [GSE3151, GSE3158 (Bild et al., 2006)] or were generated from existing datasets. Additional datasets for pathway signatures were: STAT3 (Dauer et al., 2005), TNF [GSE2638 and GSE2639 (Viemann et al., 2006)], RHOA [GSE5913 (Berenjeno et al., 2007)], TGFβ [GSE1724 (Renzoni et al., 2004)], endotoxin [GSE3284 (Calvano et al., 2005)] and immune (Galon et al., 2006). The signatures were applied to several datasets including the developmental dataset [http://breast-cancer-research.com/content/supplementary/bcr753-s1.txt (Stein et al., 2004)] and the pubertal dataset [GSE6453 (McBryan et al., 2007)]. Generation of the E2F4 signature was derived from a comparison of wild-type and *E2f4*-null mouse embryonic fibroblasts (MEFs). Three wild-type and five *E2f4*-null embryos were used to generate multiple MEF plates for each cell line. RNA from each replicate was prepared independently using the Qiagen RNeasy protocol and was used on the Affymetrix Mouse Genome 430A 2.0 platform. The resulting data were used to generate a signature (Bild et al., 2006). However, the knockout samples were used in place of the GFP-infected training samples and the wild-type samples were used in place of the overexpression training samples to create a signature of E2F4 pathway use.

### Microarray data accession numbers

The accession number for the involution series microarray data is GSE11066. The accession number for the E2F4 MEF microarray data is GSE11039.

## RESULTS

### Gene expression patterns associated with mammary development

At birth, the mammary gland consists of a rudimentary epithelial tree and as puberty progresses the TEBs, which are found at the distal end of the epithelial branches, drive the growth into the fat pad to create a ductal network (Fig. 1). When the TEBs reach the terminus of the fat pad they regress, and the epithelial network becomes static with only small, estrus-related structural changes. Upon the induction of pregnancy, proliferation and additional branching occur throughout the mammary gland. After bearing pups, the mammary gland becomes fully engulfed by alveoli, which secrete milk into the ductal network. Upon removal of the pups, the mammary gland goes though involution, which encompasses both apoptosis and remodeling. When involution is complete, the resulting mammary gland is phenotypically similar to the virgin mammary gland. The schematic (Fig. 1A) illustrates this progression and is matched with wholemounts (Fig. 1B) and histology (Fig. 1C) for each stage. Previous studies have pointed to key activities regulating these events but are nevertheless limited in scope by the focus on particular genes. We now describe a more global view of the process, with a focus on cellular signaling pathway activity.

Unsupervised clustering of gene expression data from various stages of mammary gland development reveals patterns largely reflective of developmental stages, including separation of virgin and early pregnancy from later stages of pregnancy, as well as later stages of involution (Fig. 2A). Underscoring the rapid changes that are occurring on each successive day of involution, samples were separated on each day of involution in a stepwise progression away from the lactating samples. Interestingly, although mammary glands that have been involuting for 20 days appear to be almost indistinguishable from the non-pregnant adult gland in histological sections, the clustering analysis revealed that although this group of samples shared many characteristics of the virgin samples, there was maintenance of the cluster A genes at higher levels of expression, similar to those at active involution. In a similar analysis, samples from the developing mammary gland from early to late puberty were subjected to unsupervised clustering (Fig. 2B). This analysis clustered the individual timepoints together and in a linear progression that correlated with the time the sample was collected.

Examination of the individual clusters identified in the developmental time course (Fig. 2A) using various annotation sources, including GATHER, reveals distinct representation of functional groups in terms of gene ontology (Table 1). In particular, genes activated during the proliferative phases of pregnancy are enriched in cell cycle and proliferation-associated genes (cluster C). By contrast, genes activated during involution are enriched for various apoptotic-related activities (cluster A). Genes elevated during and associated with virgin development and early lactation are enriched for activities such as organ development and morphogenesis (cluster D). In addition, genes elevated during lactation are associated with activities such as localization and transport, presumably being involved with orientation of the secretory epithelium (cluster B).

To extend the analysis of the mammary cycle, supervised clustering was employed in which the developmental stages were used to identify the gene sets that best define each stage (Fig. 3). This reiterated the observation made in the unsupervised clustering that lactation and involution were easily separated and that there was a rapid, stepwise progression in each day of lactation and involution. In addition, the supervised clustering revealed distinctions between each day of pregnancy and illustrated the overlap between histologically similar samples, such as day 1 and day 2 of pregnancy. Interestingly, in contrast to the unsupervised clustering results for the involution day 20 samples, looking horizontally at the clustering

results showed that the supervised clustering for involution day 20 had the strongest overlap with the virgin samples. In addition, whereas the unsupervised clustering grouped virgin and pregnancy before day 12.5 together, the supervised clustering was able to identify unique sets of genes that identified each successive day of pregnancy. Taken together, the supervised and unsupervised clustering give a complementary view of the various changes associated with the developmental progression of the mammary gland, but do not provide information about which genetic pathways are employed.

### Enrichment of gene sets during mammary development

Given the limitations in interpretation of biological context for single genes, we have employed GSEA to explore the possible roles of annotated pathways and functional groups of genes. Given the striking difference between lactation and the successive days of involution in the clustering analysis (Fig. 2A and Fig. 3), we initially focused on this comparison. The GSEA analysis resulted in the identification of a number of gene sets that were enriched in either the lactation or involution samples. As an example (Fig. 4A), the gene set for androgen regulation is enriched in the lactation sample and is associated with the hormonal cues in lactation. Likewise, we noted a number of gene sets enriched in the involution samples with regulated cell death and immune functions involved in remodeling the mammary gland after weaning. This analysis was repeated for the pubertal sample using the division between early and late puberty identified through unsupervised clustering. For complete lists of GSEA-normalized enrichment scores for lactation/involution and early/late puberty, see Table S1 in the supplementary material.

One limitation of GSEA is the inability to account for sample variability. To address this issue, we employed ASSESS, an application based on GSEA methods, to analyze the lactating and involuting subset of the developmental dataset. Several interesting gene sets for lactation and involution are shown in Fig. 4B. Although gene sets elevated in GSEA were present, the inclusion of variability both strengthened and expanded the data. For instance, GSEA identified Brentani immune function as a gene set enriched in the involuting samples, as did ASSESS. However, with the inclusion of variability, it is clearly shown that the highly elevated normalized enrichment scores for this gene set only began on the second day of involution, consistent with well-established histological knowledge. Further, one would expect the matrix metalloproteinases to be identified in the remodeling phase of involution. Whereas GSEA did not identify this gene set, ASSESS illustrated that the matrix metalloproteinases were involved in involution in the latter stages, particularly during day 3 and day 4 when mammary gland remodeling occurs (Fig. 4B). For the complete results, see Fig. S1 and Table S2 in the supplementary material. In addition, ASSESS differentiated between early and late puberty (Fig. 4C; for complete results see Fig. S2 and Table S3 in the supplementary material). This result was an interesting contrast to the lactation and involution analysis in that there was less variability between sample timepoints, and clustering of the results cleanly broke the pubertal dataset into two distinct clusters (see Fig. S1 versus Fig. S2 in the supplementary material). These results illustrate how annotated pathways can define the biology that underpins developmental changes.

### Patterns of cell signaling pathway activity

As an alternative approach, we have described the use of expression signatures developed to reflect various aspects of cell function, including cell signaling pathway activation. The value in this approach is the ability to bring a higher level of understanding to the transitions of mammary development, going beyond simple gene lists to representations of defined biology. We have made use of these signatures to predict the probability of activation of these pathways in the samples representing distinct stages of mammary development.

In each case, the probability of pathway activation is displayed as a color map, red indicating a high probability of pathway activation and blue a low probability of pathway activation (Fig. 5). We first examined pathways known to be differentially regulated during mammary gland development and function, including STAT3 and p63 (TRP63 – Mouse Genome Informatics). STAT3 showed a low level of pathway activation until the onset of involution, at which time the STAT3 pathway made the immediate transition to a very high probability of activation (Fig. 5A). This is fully consistent with the well-established role for STAT3 in mammary involution and serves as an important validation of the utility of pathway predictions in a developmental setting. Moreover, given that the expression level of *Stat3* is only changed slightly (see Fig. S3 in the supplementary material), it is the phosphorylation of STAT3 that is crucial, illustrating the strength of analyzing a pathway, rather than a specific gene. As an additional test, p63 was examined because its expression was previously identified as being associated with lactation (Beaton et al., 1997) (Fig. 5A). The pathway prediction illustrates the elevated probability of p63 pathway activation during lactation, consistent with the original descriptions of p63 in the mammary gland.

Given the STAT3 and p63 results, we then examined the activation state for a number of additional cell signaling pathways (Bild et al., 2006; Black et al., 2003; Huang et al., 2003). This analysis revealed a number of interesting trends that were associated with defined stages of mammary development (Fig. 5B). Specifically, the E2F1-3 pathway signatures were all elevated during pregnancy and were decreased during involution. This coincided with the retinoblastoma (*Rb; Rb1*)-knockout signature, consistent with the control of E2F as a major Rb function and with the annotations that indicated an enrichment for proliferation activities in these samples (Fig. 2A, cluster B and Table 1). These results also indicate that although E2F1 and E2F3 are active during proliferation associated with pregnancy, they are not active during the lactation process. A signature for E2F4 was also developed by comparing *E2f4*-null with wild-type mouse embryonic fibroblasts (MEFs). The resulting E2F4 signature is roughly the inverse of the E2F1-3 patterns, consistent with the fact that E2F4 has been shown to function as a transcriptional repressor, whereas E2F1-3 are transcriptional activators. As such, this result suggests that not only is there a period of E2F-specific activation followed by loss of the activators, but that this is accompanied by active repression of the target genes through the action of E2F4. In sharp contrast to the Rb/E2F1-3 profiles, signatures of RAS (HRAS1), SRC, β-catenin, and MYC were low in the pregnancy samples and then dramatically elevated upon lactation and continuing through early stages of involution (Fig. 5B).

To extend these results, training sets were derived from available datasets in other tissues and cell types to generate signatures for TNF, RHOA and TGFβ. We then assessed the mammary development dataset with these signatures (Fig. 5C). This revealed that there was an elevation of TNF in the late stages of pregnancy followed by an abrupt termination of TNF pathway activation with the onset of lactation. By contrast, RHOA (as well as RHOB and RHOC, data not shown) was elevated in lactation and in the initial phase of involution. TGFβ was elevated in early stages of pregnancy, but was then downregulated for the duration of lactation.

In a similar manner, datasets that defined phenotypic states were used to generate predictions in the mammary dataset (Fig. 5C). Here we have shown that a predictor built on breast cancer basal/luminal cell types identifies the lactating mammary gland as being highly luminal in nature. Further predictions with both an endotoxin signature and an immune signature suggest that there is a strong immune response in lactation as well as the expected elevated signature in involution.

The same sets of pathway predictors were used on the pubertal dataset to assess the importance of the various pathways (Fig. 5D). Although the probabilities of pathway activation were not as strong as for the developmental dataset, the changes associated with mammary gland development were reflected in the puberty pathway probabilities. For instance, the E2F1 and E2F3 pathway probabilities were elevated in the latter stages of puberty, as were MYC, TGFβ and β-catenin. Taken together, the signatures examined in puberty and the developmental time courses illustrate the utility of examining pathways in interpreting the underlying biology.

**A role for E2F activities in mammary gland development**

In light of the distinct pattern of E2F pathway activity seen during the proliferative phase of mammary development in both the developmental dataset (Fig. 5B) and the pubertal dataset (Fig. 5D), we have directly addressed the importance of these activities through the analysis of mice deficient for various members of the E2F family. Initially, we demonstrated E2F expression in the mammary gland (see Fig. S3 in the supplementary material). Using a wholemount staining technique, we examined gross morphology and extent of growth of the various mammary glands lacking E2F proteins at 4 weeks of age. By comparison with wild-type controls (*n*=14), mammary glands from *E2f1* knockouts (*n*=12), *E2f3* heterozygotes (*n*=12) and *E2f4* knockouts (*n*=4) exhibited a reduction in the extent of fat pad penetration (Fig. 6A). Whereas the *E2f1* knockout and *E2f3* heterozygous glands lagged behind the control (81% and 66% of wild-type growth, respectively) (Fig. 6B), the *E2f4* knockout was severely stunted in outgrowth (31% of wild-type growth) (Fig. 6B) and did not yet exhibit normal branching. The *E2f3*-knockout line was not significantly different from the wild-type control in proportional outgrowth (*n*=4) (Fig. 6B); however, the *E2f3*-null epithelium exhibited the same lower measured distance as the heterozygous epithelium. These data suggest that the location of the lymph node and nipple relative to each other might have changed slightly in the *E2f3*-null mice. Similar results, but with reduced outgrowth in the *E2f3* knockouts, were observed when epithelial growth was measured at 8 weeks of development (Fig. 6B). In addition to the outgrowth delays, branching defects were noted in the majority of the mutant mice at 8 weeks of development that persisted into adulthood (Fig. 6A). Mammary epithelial branch points were counted in a defined area adjacent to the lymph node at 8 weeks. This revealed a significant reduction in branching, as compared with controls (*n*=6), for *E2f1*-knockout (*n*=3), *E2f3* heterozygous (*n*=5), *E2f3*-knockout (*n*=4) and *E2f4*-knockout mice (*n*=6).

Based on a wholemount examination of mammary epithelium transplants, the proliferation and differentiation defects appeared to be a cell-autonomous phenotype for the various E2F-deficient mice (*n*=3 for each genotype) (Fig. 6C). In contrast to the controls, *E2f1*-knockout epithelium transplanted into a nu/nu recipient exhibited reduced outgrowth and branching. More strikingly, *E2f3*-knockout and *E2f4*-knockout transplants lagged far behind the controls. Quantitation revealed that there were significant differences in *E2f1*-knockout, *E2f3* heterozygous, *E2f3*-knockout and *E2f4*-knockout transplants (Fig. 6D). These results clearly indicate that the epithelial effects noted in the various mutant mice are cell-autonomous and are in agreement with the proliferative role for the E2Fs identified in the pathway predictions.

Further evidence for a role of E2F activities in the proliferative phase of mammary development was indicated by an analysis of gene expression patterns developed to profile TEB formation. Comparing RNA expression in dissected TEB structures with that in ductal structures revealed a pattern of expression unique to the TEBs. Analysis of this group of genes for various sources of annotation revealed a very significant enrichment for genes containing E2F binding sites in their promoters (Table 2 and see Fig. S4 in the supplementary material).

Pathway analysis also suggested a role for E2Fs in the involution stage, and previous work has pointed to a role for E2F1 in apoptosis induction in fibroblast cultures. However, the loss of *E2f1* resulted in no discernable changes in mammary gland involution by comparison with the control (Fig. 7A). Interestingly, *E2f3* heterozygous mice exhibited a delay in involution, which was most striking in the initial phase of involution (Fig. 7A). These results indicate that partial or complete loss of E2F3 delays apoptosis in the involuting mammary gland, a result also seen in TUNEL assays (Fig. 7B,C).

Finally, we also made use of a signature generated to reflect the involution process (see Fig. S5 in the supplementary material) and then used the signature to predict the probability of this phenotype in the samples from wild-type and $E2f3^{+/-}$ mice. As shown in Fig. 7D, there was a marked reduction in the involution signature in the $E2f3^{+/-}$ mice, corresponding to the phenotypic delay noted in the histology results.

In order to examine the involution microarray data for targets that could be potentially regulated directly by E2F3, the wild-type and $E2f3^{+/-}$ data from involution day 1 and day 2 were compared for fold differences. Genes with greater than a 2-fold difference were used to identify differentially expressed genes with E2F consensus sites in their promoters (see Table S4 in the supplementary material). The expression pattern of these genes was examined and genes with strong E2F binding sites and variable expression patterns were chosen for further analysis in a chromatin immunoprecipitation (ChIP) assay. HC11 mouse mammary epithelial cells that were either proliferating or undergoing apoptosis were examined by ChIP. Compared with the ribonucleotide reductase M2 positive control, there were indeed genes with differential expression, the promoters of which were bound by E2F3 (see Fig. S6 and Table S4 in the supplementary material).

## DISCUSSION

Numerous studies have provided detailed descriptions of the morphological events associated with mammary gland development, starting with extensive epithelial cell expansion following birth. These stages of mammary gland morphogenesis are accompanied by dramatic changes in gene expression that reflect the action of various genes and pathways that elicit theses events. Previous studies using microarray analysis of mammary gland development have predominantly focused on the regulation of individual genes or a specific stage of development. These changes agree with the global analysis of expression changes we noted through clustering experiments. Further, the annotation of some of the discrete clusters within this analysis revealed gene ontology groups consistent with the phenotypic state. Although interesting, these data did not provide insight into the mechanism of action of the various stages.

Although genome-scale measures of gene expression provide a global picture of the expression changes associated with mammary development, and clearly have tremendous power to reveal subtle phenotypes not otherwise detected, it is also a significant challenge to extract and understand the underlying biology highlighted by these profiles. The work we describe here has made use of methods of expression analysis that relate profiles (signatures) to defined function, and then evaluate the extent to which this signature is represented in a given biological sample. Importantly, these signatures can be assessed quantitatively in a given sample to determine the extent to which that signature is present and thus the extent to which the biological phenotype represented by the signature is present.

The power of the expression signature is twofold. First, it is portable in the sense that it can be assayed in different contexts. That is, an expression signature developed in a cell culture context can be measured in a tumor or developmental stage, providing the capacity to link

otherwise heterologous systems. A cell culture phenotype, such as pathway activation, is difficult to represent in a diverse sample, such as a tumor. By contrast, the expression profile that represents pathway activation in cell culture can then be used to interrogate the expression data from a tumor. Second, it is quantitative and can be assessed in the context of other signatures to identify patterns of signatures. This latter characteristic is perhaps of greatest significance because it provides the opportunity to not only go beyond single genes, but to also go beyond single pathways to begin to describe more complex interactions that ultimately define networks.

The analyses we report here suggest that although cell signaling pathways such as Rb/E2F, RAS, MYC, SRC and others are all generally associated with cell proliferation, there is a sharp distinction in the activity of these pathways during the course of mammary gland development. In particular, E2F activity was seen as prominent during early stages prior to lactation, whereas other pathways, including MYC, β-catenin, RAS and SRC, were generally silent during this phase and then abruptly activated upon lactation. These latter pathways remained active through involution and then returned to baseline levels as the mammary gland returned to the pre-lactation state.

The elevation of MYC expression during lactation is consistent with the precocious lactation induced by overexpression of MYC in the mammary gland during pregnancy (Blakely et al., 2005). Conversely, *Src* transgenics have deficiencies in lactation (Webster et al., 1995) and *Src*-knockouts have deficiencies in mammary outgrowth (Kim et al., 2005), corresponding to an elevated pathway signature in the puberty dataset, but to date experiments have not directly assessed the role of SRC in lactation. Other work has demonstrated that TNF expression is increased during pregnancy and is decreased during lactation (Varela and Ip, 1996), correlating with the pathway probability for TNF. The TGFβ signature was elevated in early pregnancy but was eliminated during lactation, consistent with the lactational failure of WAP-TGFβ, designed to express TGFβ during pregnancy and lactation (Jhappan et al., 1993). Moreover, the early pregnancy signature associated with TGFβ is in accordance with the theory that TGFβ acts in pregnancy to prevent premature milk synthesis and the associated apoptosis, but is downregulated in lactation to allow milk synthesis and secretion (Robinson et al., 1993). In contrast to these pregnancy-associated signatures, the pathway probability for Rho showed a high level of activation during lactation, consistent with the role of the Rho family in mediating signals from integrins and hormones in milk synthesis (Akhtar and Streuli, 2006). Interestingly, when both an endotoxin infection and an immune signature were applied there was activation of these states in the involuting samples, consistent with prior studies (Clarkson et al., 2004; Master et al., 2002; Stein et al., 2004). Further, these immune signatures were also present in the lactating sample, an observation not made in the previous microarray studies. The presence of an immune signature in the lactating mammary gland is consistent with the presence of host defense proteins secreted into milk (Smolenski et al., 2007) and the use of pathways such as NF-κB and Stat in both lactation and inflammation (Vorbach et al., 2006).

The observation that E2F1 and E2F3 loss-of-function affect proliferation is consistent with the majority of in vitro work that details the importance of transcriptional activator E2F proteins in mediating proliferation and was predicted by the pathway signatures. Functional overlap and compensation by other family members is likely to mute some of the effects of mutating individual E2Fs. However, given the distinct role identified for E2F3 in mediating DNA replication and mitotic activity in vitro (Kong et al., 2007), it is not unexpected that the E2F3 phenotype is the most striking amongst those of the transcriptional activators. The defect in outgrowth of the *E2f4*-null mammary epithelium was striking and displayed an even more severe phenotype than that of *E2f3*-null mice. Although previous work has suggested a role for E2F4 as the predominant E2F transcriptional repressor in the regulation

of cell fate and differentiation, recent studies have shown that loss of E2F4 function can also result in a reduction in proliferation as well as altered cell fate decisions (Kinross et al., 2006). Further evidence for a role of E2F activities in the mammary gland proliferative process came from the observation of enrichment of E2F elements in the promoters of genes induced in the TEB. Future work might reveal the precise mechanism by which select E2Fs can regulate these branching and outgrowth processes and why certain E2Fs are essential for this process.

In addition, although previous work has established E2F1 as a trigger for p53 (TRP53)-mediated apoptosis in fibroblast cultures (DeGregori et al., 1995; Hallstrom and Nevins, 2003; Kowalik et al., 1998; Kowalik et al., 1995), the pathway analysis suggested that there might be an E2F3-specific role during involution, based on the changes in involution at days 3 and 4 in the E2F3 signature. Although it is possible that E2F1-mediated apoptosis, like that mediated by p53, might also be affected by background, our results suggest that E2F1 is not required for involution (Jerry et al., 1998; Li et al., 1996; Matthews and Clarke, 2005). By contrast, and consistent with the pathway analysis, we demonstrated that perturbations in E2F3 levels resulted in a delay in involution-mediated apoptosis. Although previous studies have illustrated a role for E2F3 in mediating apoptosis, they have done so in an *Rb*-null background (Tsai et al., 1998; Ziebold et al., 2001) and proposed that apoptosis is induced once E2F3 expression exceeds a certain threshold. By contrast, our results show that E2F3 expression is lowered in involution as compared with lactation and then mediates apoptosis. Given the switch of E2F3 to E2F4 in DNA binding at the lactation/involution switch (Gadd et al., 2001), our results suggest that the balance between E2F-mediated transcriptional activation and repression is critical for the induction of mammary apoptosis. To test whether genes involved in involution are bound by E2F3, we performed a ChIP analysis and identified a number of promoters bound by E2F3. Interestingly, several of these targets (*Trp53inp1*, *Map3k4* and *Hexb*) have previously been observed to be involved in mediating apoptosis (Huang et al., 1997; Takekawa and Saito, 1998; Tomasini et al., 2005). Unfortunately, because the *E2f4*-null mice fail to generate significant numbers of secretory alveoli (data not shown), we were unable to accurately examine involution to test the pathway analysis prediction. However, the results observed in the E2F-knockout mice served to validate the pathway activation predictions that indicated that the E2Fs have a role in mammary gland development.

Taken together, we have shown that using a computational model in conjunction with genome-scale measurements of gene expression enables the prediction of a role for various genetic pathways in a developmental context. Importantly, the utility of these predictive models was tested in vivo, revealing important insights into the differential regulation of mammary development by the E2F family of transcription factors. Given the nature of this method, it will immediately lend itself to examination of pathways in other contexts, whether in the development of a tissue or in the aberrant development of a tumor.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

# References

Akhtar N, Streuli CH. Rac1 links integrin-mediated adhesion to the control of lactational differentiation in mammary epithelia. J Cell Biol. 2006; 173:781–793. [PubMed: 16754961]

Beaton A, Wilkins RJ, Wheeler TT. Lactation-associated and prolactin-responsive changes in protein synthesis in mouse mammary cells. Tissue Cell. 1997; 29:509–516. [PubMed: 9364800]

Berenjeno IM, Nunez F, Bustelo XR. Transcriptomal profiling of the cellular transformation induced by Rho subfamily GTPases. Oncogene. 2007; 26:4295–4305. [PubMed: 17213802]

Bild A, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature. 2006; 439:353–357. [PubMed: 16273092]

Black EP, Huang E, Dressman H, Ishida S, West M, Nevins JR. Distinct gene expression phenotypes of cells lacking Rb and Rb family members. Cancer Res. 2003; 63:3716–3723. [PubMed: 12839964]

Blakely CM, Sintasath L, DCruz CM, Hahn KT, Dugan KD, Belka GK, Chodosh LA. Developmental stage determines the effects of MYC in the mammary epithelium. Development. 2005; 132:1147–1160. [PubMed: 15689376]

Blanchard A, Shiu R, Booth S, Sorensen G, DeCorby N, Nistor A, Wong P, Leygue E, Myal Y. Gene expression profiling of early involuting mammary gland reveals novel genes potentially relevant to human breast cancer. Front Biosci. 2007; 12:2221–2232. [PubMed: 17127459]

Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, et al. A network-based analysis of systemic inflammation in humans. Nature. 2005; 437:1032–1037. [PubMed: 16136080]

Chang JT, Nevins JR. GATHER: a systems approach to interpreting genomic signatures. Bioinformatics. 2006; 22:2926–2933. [PubMed: 17000751]

Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry. 1979; 18:5294–5299. [PubMed: 518835]

Clarkson RW, Wayland MT, Lee J, Freeman T, Watson CJ. Gene expression profiling of mammary gland development reveals putative roles for death receptors and immune mediators in post-lactational regression. Breast Cancer Res. 2004; 6:R92–R109. [PubMed: 14979921]

Dauer DJ, Ferraro B, Song L, Yu B, Mora L, Buettner R, Enkemann SA, Jove R, Haura EB. Stat3 regulates genes common to both wound healing and cancer. Oncogene. 2005; 24:3397–3408. [PubMed: 15735721]

DeGregori J, Kowalik T, Nevins JR. Cellular targets for activation by the E2F1 transcription factor include DNA synthesis and G1/S regulatory genes. Mol Cell Biol. 1995; 15:4215–4224. [PubMed: 7623816]

Deome KB, Faulkin LJ Jr, Bern HA, Blair PB. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. Cancer Res. 1959; 19:515–520. [PubMed: 13663040]

Edelman E, Porrello A, Guinney J, Balakumaran B, Bild A, Febbo PG, Mukherjee S. Analysis of sample set enrichment scores: assaying the enrichment of sets of genes for individual samples in genome-wide expression profiles. Bioinformatics. 2006; 22:e108–e116. [PubMed: 16873460]

Field SJ, Tsai FY, Kuo F, Zubiaga AM, Kaelin WG Jr, Livingston DM, Orkin SH, Greenberg ME. E2F-1 functions in mice to promote apoptosis and suppress proliferation. Cell. 1996; 85:549–561. [PubMed: 8653790]

Gadd M, Pisc C, Branda J, Ionescu-Tiba V, Nikolic Z, Yang C, Wang T, Shackleford GM, Cardiff RD, Schmidt EV. Regulation of cyclin D1 and p16(INK4A) is critical for growth arrest during mammary involution. Cancer Res. 2001; 61:8811–8819. [PubMed: 11751403]

Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, Tosolini M, Camus M, Berger A, Wind P, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science. 2006; 313:1960–1964. [PubMed: 17008531]

Hallstrom TC, Nevins JR. Specificity in the activation and control of transcription factor E2F-dependent apoptosis. Proc Natl Acad Sci USA. 2003; 100:10848–10853. [PubMed: 12954980]

Huang E, Ishida S, Pittman J, Dressman H, West M, Nevins JR. Gene expression phenotypic models that predict the activity of oncogenic pathways. Nature Genet. 2003; 34:226–230. [PubMed: 12754511]

Huang JQ, Trasler JM, Igdoura S, Michaud J, Hanal N, Gravel RA. Apoptotic cell death in mouse models of GM2 gangliosidosis and observations on human Tay-Sachs and Sandhoff diseases. Hum Mol Genet. 1997; 6:1879–1885. [PubMed: 9302266]

Humbert PO, Verona R, Trimarchi JM, Rogers C, Dandapani S, Lees JA. E2f3 is critical for normal cellular proliferation. Genes Dev. 2000; 14:690–703. [PubMed: 10733529]

Jerry DJ, Kuperwasser C, Downing SR, Pinkas J, He C, Dickinson E, Marconi S, Naber SP. Delayed involution of the mammary epithelium in BALB/c-p53null mice. Oncogene. 1998; 17:2305–2312. [PubMed: 9811461]

Jhappan C, Geiser AG, Kordon EC, Bagheri D, Henninghausen L, Roberts AB, Smith GH, Merlino G. Targeting expression of a transforming growth factor beta 1 transgene to the pregnant mammary gland inhibits alveolar development and lactation. EMBO J. 1993; 12:1835–1845. [PubMed: 8491177]

Kim HG, Laing M, Muller W. c-Src-null mice exhibit defects in normal mammary gland development and ERalpha signaling. Oncogene. 2005; 24:5629–5636. [PubMed: 16007215]

Kinross KM, Clark AJ, Iazzolino RM, Humbert PO. E2f4 regulates fetal erythropoiesis through the promotion of cellular proliferation. Blood. 2006; 108:886–895. [PubMed: 16861343]

Kong LJ, Chang JT, Bild AH, Nevins JR. Compensation and specificity of function within the E2F family. Oncogene. 2007; 26:321–327. [PubMed: 16909124]

Kouros-Mehr H, Werb Z. Candidate regulators of mammary branching morphogenesis identified by genome-wide transcript analysis. Dev Dyn. 2006; 235:3404–3412. [PubMed: 17039550]

Kouros-Mehr H, Slorach EM, Sternlicht MD, Werb Z. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. Cell. 2006; 127:1041–1055. [PubMed: 17129787]

Kowalik TF, DeGregori J, Schwarz JK, Nevins JR. E2F1 overexpression in quiescent fibroblasts leads to induction of cellular DNA synthesis and apoptosis. J Virol. 1995; 69:2491–2500. [PubMed: 7884898]

Kowalik TF, DeGregori J, Leone G, Nevins JR. E2F1-specific induction of apoptosis and p53 accumulation is modulated by mdm2. Cell Growth Differ. 1998; 9:113–118. [PubMed: 9486847]

Li M, Hu J, Heermeier K, Hennighausen L, Furth PA. Apoptosis and remodeling of mammary gland tissue during involution proceeds through p53-independent pathways. Cell Growth Differ. 1996; 7:13–20. [PubMed: 8788029]

Locke D, Stein T, Davies C, Morris J, Harris AL, Evans WH, Monaghan P, Gusterson B. Altered permeability and modulatory character of connexin channels during mammary gland development. Exp Cell Res. 2004; 298:643–660. [PubMed: 15265710]

Master SR, Hartman JL, D'Cruz CM, Moody SE, Keiper EA, Ha SI, Cox JD, Belka GK, Chodosh LA. Functional microarray analysis of mammary organogenesis reveals a developmental role in adaptive thermogenesis. Mol Endocrinol. 2002; 16:1185–1203. [PubMed: 12040007]

Matthews JR, Clarke AR. p53 mediates a default programme of mammary gland involution in the absence of STAT3. Oncogene. 2005; 24:3083–3090. [PubMed: 15735683]

McBryan J, Howlin J, Kenny PA, Shioda T, Martin F. ERalpha-CITED1 co-regulated genes expressed during pubertal mammary gland development: implications for breast cancer prognosis. Oncogene. 2007; 26:6406–6419. [PubMed: 17486082]

Murga M, Fernandez-Capetillo O, Field SJ, Moreno B, Borlado LR, Fujiwara Y, Balomenos D, Vicario A, Carrera AC, Orkin SH, et al. Mutation of E2F2 in mice causes enhanced T lymphocyte proliferation, leading to the development of autoimmunity. Immunity. 2001; 15:959–970. [PubMed: 11754817]

Rempel RE, Saenz-Robles MT, Storms R, Morham S, Ishida S, Engel A, Jakoi L, Melhem MF, Pipas JM, Smith C, et al. Loss of E2F4 activity leads to abnormal development of multiple cellular lineages. Mol Cell. 2000; 6:293–306. [PubMed: 10983977]

Renzoni EA, Abraham DJ, Howat S, Shi-Wen X, Sestini P, Bou-Gharios G, Wells AU, Veeraraghavan S, Nicholson AG, Denton CP, et al. Gene expression profiling reveals novel TGFbeta targets in adult lung fibroblasts. Respir Res. 2004; 5:24. [PubMed: 15571627]

Robinson SD, Roberts AB, Daniel CW. TGF-beta suppresses casein synthesis in mouse mammary explants and may play a role in controlling milk levels during pregnancy. J Cell Biol. 1993; 120:245–251. [PubMed: 8416990]

Rudolph MC, McManaman JL, Phang T, Russell T, Kominsky DJ, Serkova NJ, Stein T, Anderson SM, Neville MC. Metabolic regulation in the lactating mammary gland: a lipid synthesizing machine. Physiol Genomics. 2007; 28:323–336. [PubMed: 17105756]

Smolenski G, Haines S, Kwan FY, Bond J, Farr V, Davis SR, Stelwagen K, Wheeler TT. Characteristisation of host defence proteins in milk using a proteomic approach. J Proteome Res. 2007; 6:207–215. [PubMed: 17203965]

Stein T, Morris JS, Davies CR, Weber-Hall SJ, Duffy MA, Heath VJ, Bell AK, Ferrier RK, Sandilands GP, Gusterson BA. Involution of the mouse mammary gland is associated with an immune cascade and an acute-phase response, involving LBP, CD14, and STAT3. Breast Cancer Res. 2004; 6:89–92. [PubMed: 14979913]

Stein T, Price KN, Morris JS, Heath VJ, Ferrier RK, Bell AK, Pringle MA, Villadsen R, Petersen OW, Sauter G, et al. Annexin A8 is up-regulated during mouse mammary gland involution and predicts poor survival in breast cancer. Clin Cancer Res. 2005; 11:6872–6879. [PubMed: 16203777]

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gilletee MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005; 102:15545–15550. [PubMed: 16199517]

Takekawa M, Saito H. A family of stress-inducible GADD45-like proteins mediate activation of the stress-responsive MTK1/MEKK4 MAPKKK. Cell. 1998; 95:521–530. [PubMed: 9827804]

Tomasini R, Seux M, Nowak J, Bontemps C, Carrier A, Dagorn JC, Pebusque MJ, Iovanna JL, Dusetti NJ. TP53INP1 is a novel p73 target gene that induces cell cycle arrest and cell death by modulating p73 transcriptional activity. Oncogene. 2005; 24:8093–8204. [PubMed: 16044147]

Tsai KY, Hu Y, Macleod KF, Crowley D, Yamasaki L, Jacks T. Mutation of E2f-1 suppresses apoptosis and inappropriate S phase entry and extends survival of Rb-deficient mouse embryos. Mol Cell. 1998; 2:293–304. [PubMed: 9774968]

Varela LM, Ip MM. Tumor necrosis factor-alpha: a multifunctional regulator of mammary gland development. Endocrinology. 1996; 137:4915–4924. [PubMed: 8895364]

Viemann D, Goebeler M, Schmid S, Nordhues U, Klimmek K, Sorg C, Roth J. TNF induces distinct gene expression programs in microvascular and macrovascular human endothelial cells. J Leukoc Biol. 2006; 80:174–185. [PubMed: 16617158]

Vorbach C, Capecchi MR, Penninger JM. Evolution of the mammary gland from the innate immune system? BioEssays. 2006; 28:606–616. [PubMed: 16700061]

Webster MA, Cardiff RD, Muller WJ. Induction of mammary epithelial hyperplasias and mammary tumors in transgenic mice expressing a murine mammary tumor virus/activated c-src fusion gene. Proc Natl Acad Sci USA. 1995; 92:7849–7853. [PubMed: 7544006]

Zhu W, Giangrande P, Nevins JR. E2Fs link the control of G1/S and G2/M. EMBO J. 2004; 23:4615–4626. [PubMed: 15510213]

Ziebold U, Reza T, Caron A, Lees JA. E2F3 contributes both to the inappropriate proliferation and to the apoptosis arising in Rb mutant embryos. Genes Dev. 2001; 15:386–391. [PubMed: 11230146]
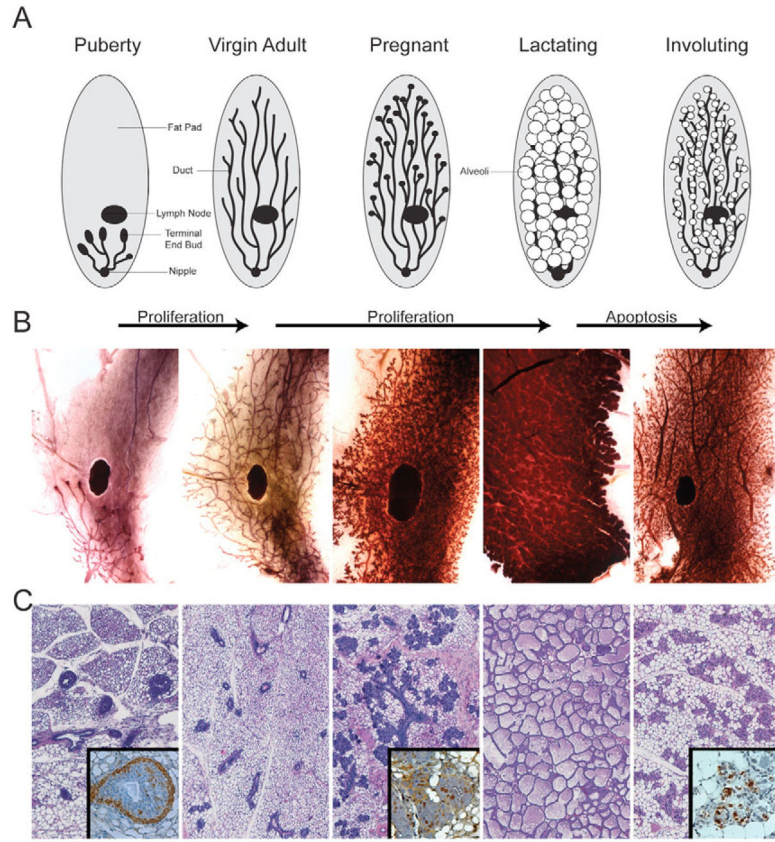
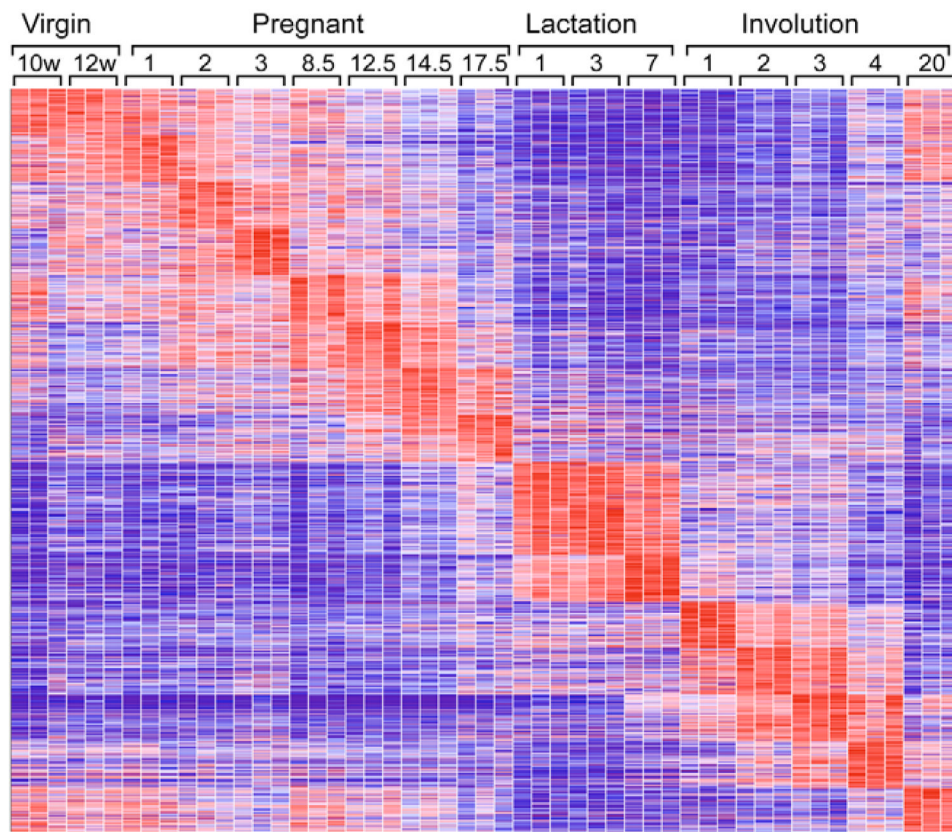**Fig. 1. Mammary gland development**
(**A**) Schematic illustrating the development of the mouse mammary gland. Each stage of the schematic is accompanied by (**B**) matched timepoints of mammary glands in wholemount preparations and (**C**) histological sections. Insets in the histological panels show (from left to right) examples of immunohistochemistry for α smooth muscle actin to highlight the cap cells (puberty), PNCA to illustrate rapid growth (pregnant) and TUNEL to illustrate regulated cell death (involution).
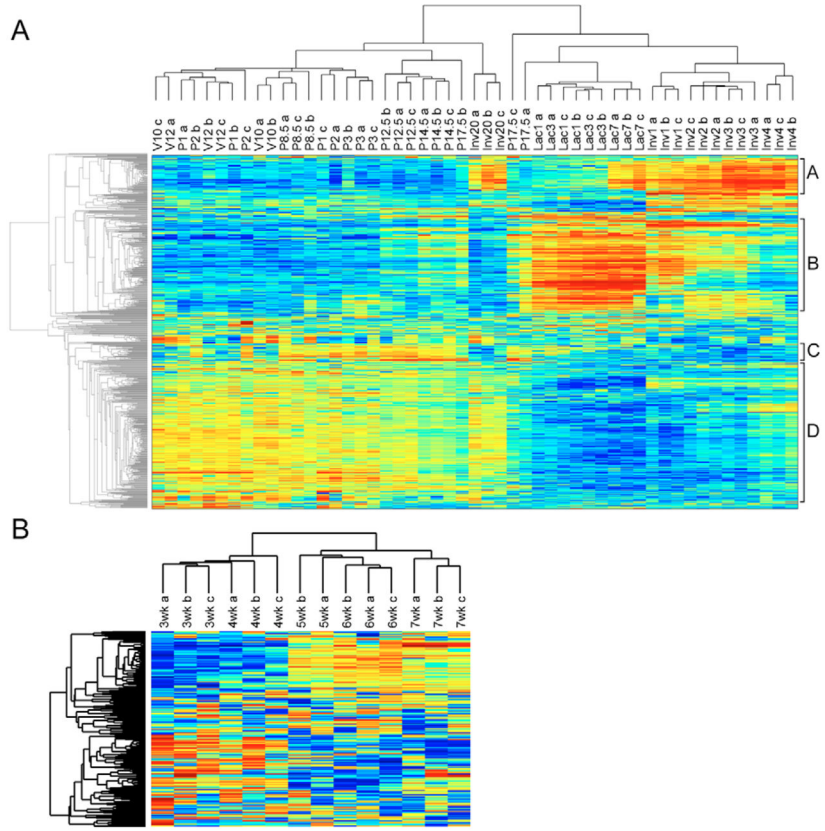
**Fig. 2. Unsupervised analysis of mouse mammary gland development**
(**A**) The developmental dataset including virgin (V) at 10 weeks and 12 weeks of age, pregnancy (P) at 1, 2, 3, 8.5, 12.5, 14.5 and 17.5 days post coitum, lactation (Lac) at 1, 3 and 7 days following partition, and involution (Inv) samples at days 1–4 and 20 following weaning were used in unsupervised clustering. The various clusters denoting involution, lactation, early and late pregnancy that were used in a search of gene ontology are indicated on the right and are denoted A–D. (**B**) The pubertal dataset including samples from 3 to 7 weeks of development, as also used in unsupervised clustering.

**Fig. 3. Supervised clustering of mouse mammary gland development**
The developmental dataset used in Fig. 2A was analyzed through supervised clustering.
Virgin timepoints are measured in weeks, the remainder in days.
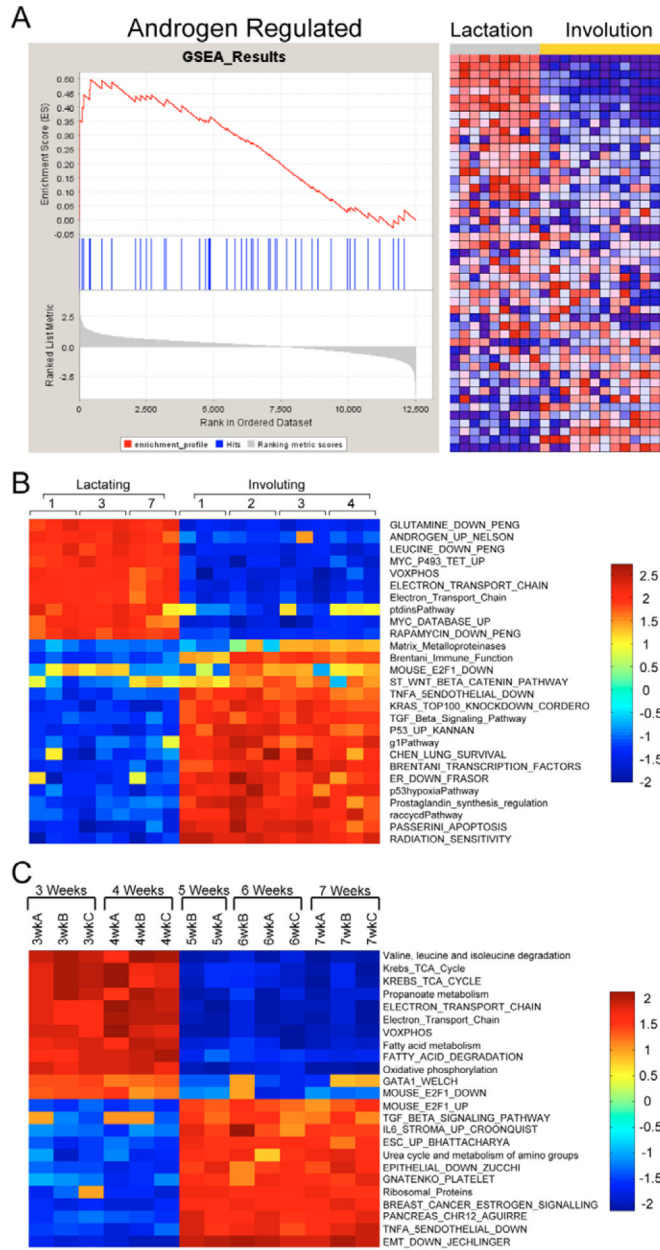
**Fig. 4. Gene set enrichment during mouse mammary development**

(**A**) Lactation (days 1, 3 and 7) and involution (days 1–4) samples were used in GSEA. An example of the enrichment analysis is shown with the random walk for the androgen-regulated gene set as well as the heatmap of the individual genes that compose the gene set. (**B**) ASSESS, a version of GSEA, was applied to the same data with the normalized enrichment scores being reported on a sample-by-sample basis for each gene set. Results from selected gene sets are shown. (**C**) In addition, the pubertal dataset was examined, with the two phenotypic classes based on the differential sample clustering from Fig. 2B, and the normalized enrichment scores are shown for a number of gene sets.
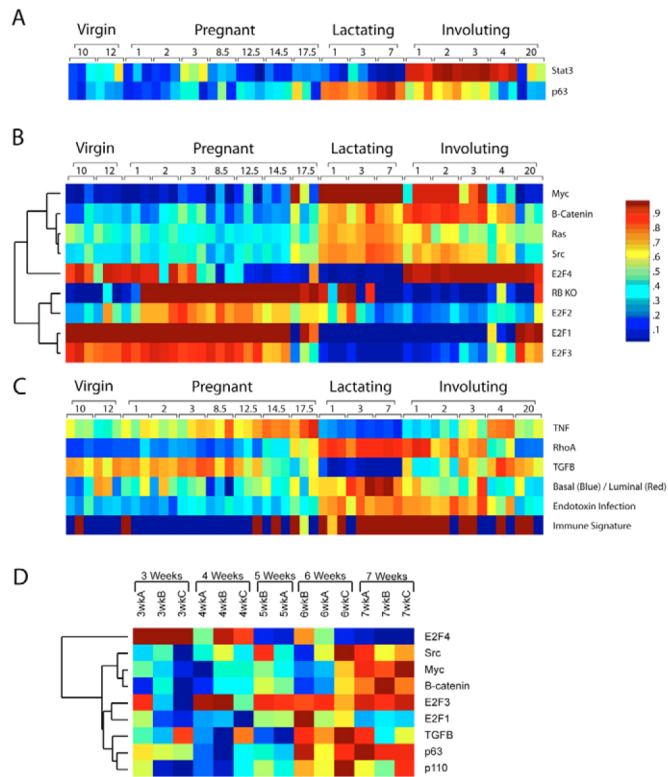
**Fig. 5. Pathway activation probabilities**

(**A**) The probability of pathway activation was examined in the developmental and pubertal datasets. STAT3 and p63, well-established mammary markers, were initially used to validate this method in the mammary development dataset. (**B**) In addition, a variety of genetic pathways were analyzed for their probability of activation in the developmental dataset, including MYC, β-catenin, RAS, SRC, a knockout signature for Rb, E2F1, E2F2, E2F3 and E2F4. (**C**) Other phenotypic and genetic pathways were also examined for the probability of their activation including TNF, RHOA, TGFβ, basal/luminal status, endotoxin infection and an immune signature. (**D**) In addition, the pubertal data were subjected to the same analysis and relevant results are shown.
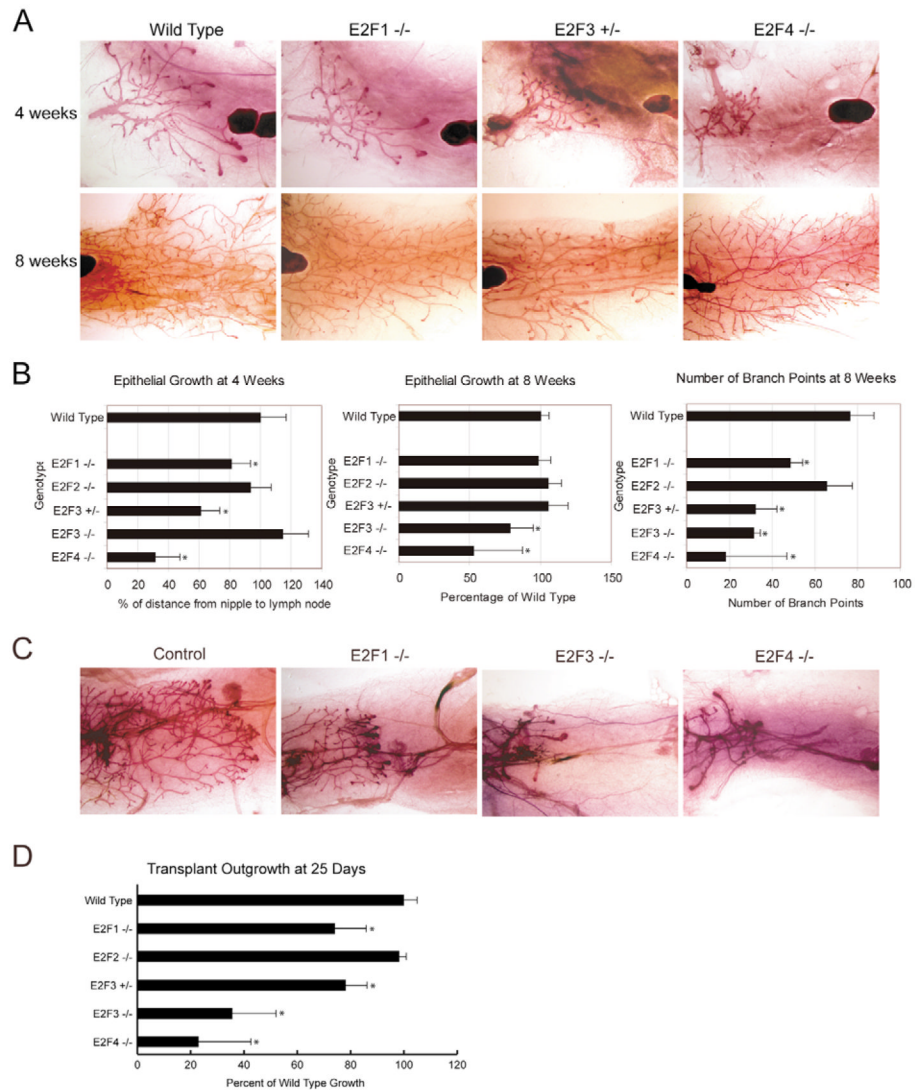
**Fig. 6. Mammary outgrowth is defective in E2F mutant mice**

(**A**) Wholemounts from wild-type, *E2f1*-knockout, *E2f3* heterozygous and *E2f4*-knockout mice are shown at 4 and 8 weeks of age. (**B**) The extent of mammary epithelial outgrowth was quantitated at 4 weeks of development. Relative to the control, there was a significant difference in *E2f1*-knockout (*P*=0.0017), *E2f3* heterozygous (*P*<0.0001) and *E2f4*-knockout (*P*<0.0001) mice. This outgrowth analysis was repeated at 8 weeks of age, when the *E2f3*-knockout and *E2f4*-knockout exhibit a growth delay (*P*=0.059 and *P*=0.02, respectively) relative to the control. In addition, ductal branching was quantitated at 8 weeks and was significant for each of the E2F mutant strains, except the *E2f2*-knockout mice [*P*=0.05 (*E2f1*), 0.00008 (*E2f3* heterozygous), 0.0003 (*E2f3* knockout) and 0.001 (*E2f4* knockout)]. (**C**) Transplants of control and E2F mutant mammary epithelium in nu/nu recipients as shown in a wholemount analysis. (**D**) Quantitation of transplant outgrowth in E2F mutants relative to wild-type control. *E2f1*-knockout ductal extension was reduced (to 74% of control, *P*=0.046), whereas no defects were observed for *E2f2* knockouts. *E2f3*-knockout and *E2f3* heterozygous mice had 36% (*P*=0.014) and 78% growth (*P*=0.043), respectively, of wild-type growth. *E2f4* knockouts were most severely affected with only 23% of control growth (*P*=0.010). Asterisk denotes statistical significance.
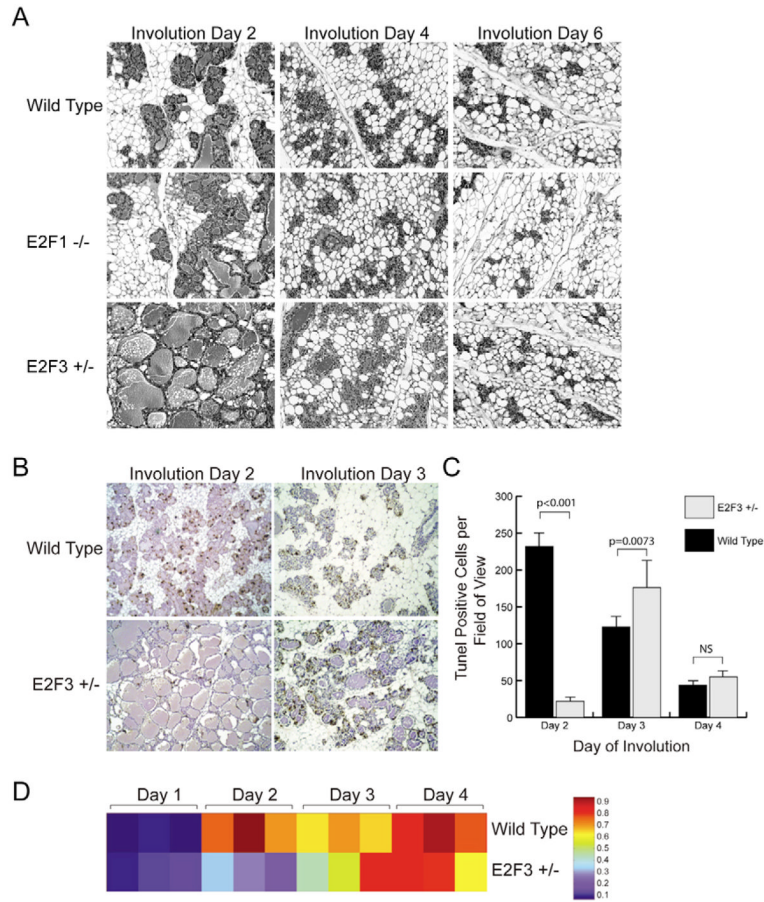
**Fig. 7. Apoptotic delays during involution in *E2f3* mutant mice**

(**A**) Involution in the wild-type control, *E2f1*-knockout and *E2f3* heterozygous samples. (**B**) To examine apoptosis in the wild type and *E2f3* mutant, a TUNEL analysis was conducted at day 2 and day 3 of involution. TUNEL-positive cells stain brown and the sections are lightly counterstained with Hematoxylin. (**C**) The results of the TUNEL staining were quantitated, with the day 2 and day 3 differences being significant ($P < 0.0001$ and $P = 0.0073$, respectively). (**D**) An involution signature from a wild-type mammary involution dataset was generated and the probability of fitting the signature was assessed for the wild type and *E2f3* mutant at successive days of involution. Red indicates a high probability of matching the involution day 4 signature; blue indicates a high probability of matching the involution day 1 signature.

**Table 1**

Gene ontology for clusters from unsupervised clustering

| Annotation | Bayes factor |
| --- | --- |
| **Cluster A** | |
| GO:0009607 [4]: response to biotic stimulus | 46.51 |
| GO:0006955 [4]: immune response | 41.26 |
| GO:0006952 [5]: defense response | 40.17 |
| GO:0006950 [4]: response to stress | 30.89 |
| GO:0043207 [5]: response to external biotic stimulus | 30.78 |
| GO:0050776 [5]: regulation of immune response | 29.23 |
| GO:0009613 [5]: response to pest, pathogen or parasite | 28.57 |
| GO:0006954 [5]: inflammatory response | 28.11 |
| GO:0051246 [5]: regulation of protein metabolism | 26.57 |
| **Cluster B** | |
| GO:0007166 [5]: cell surface receptor linked signal transduction | 10.18 |
| GO:0050875 [3]: cellular physiological process | 9.89 |
| GO:0007186 [6]: G-protein coupled receptor protein signaling pathway | 9.88 |
| GO:0006810 [4]: transport | 7.07 |
| GO:0051234 [4]: establishment of localization | 7.04 |
| GO:0051179 [3]: localization | 6.83 |
| GO:0009581 [5]: detection of external stimulus | 6.74 |
| GO:0007600 [5]: sensory perception | 6.64 |
| GO:0050877 [4]: neurophysiological process | 6.06 |
| GO:0050896 [3]: response to stimulus | 5.78 |
| **Cluster C** | |
| GO:0000279 [6]: M phase | 14.29 |
| GO:0007049 [5]: cell cycle | 11.92 |
| GO:0000280 [7]: nuclear division | 11.34 |
| GO:0007067 [8]: mitosis | 9.91 |
| GO:0000087 [7]: M phase of mitotic cell cycle | 9.86 |
| GO:0030261 [7]: chromosome condensation | 9.85 |
| GO:0008283 [4]: cell proliferation | 9.38 |
| GO:0000278 [6]: mitotic cell cycle | 7.9 |
| GO:0000088 [9]: mitotic prophase | 5.48 |
| GO:0007076 [8]: mitotic chromosome condensation | 5.48 |
| **Cluster D** | |
| GO:0007155 [4]: cell adhesion | 13.98 |
| GO:0007600 [5]: sensory perception | 13.97 |
| GO:0007275 [2]: development | 13.64 |

| Annotation | Bayes factor |
|---|---|
| GO:0042330 [5]: taxis | 12.21 |
| GO:0006935 [6]: chemotaxis | 12.21 |
| GO:0009581 [5]: detection of external stimulus | 11.96 |
| GO:0006954 [5]: inflammatory response | 10.1 |
| GO:0009653 [3]: morphogenesis | 8.89 |
| GO:0009887 [4]: organogenesis | 7.86 |
| GO:0048513 [3]: organ development | 7.72 |

Individual genes from each of the four clusters identified by unsupervised clustering (Fig. 2A) were used to search in gene ontology resources to identify major classes of genes. GATHER was used to define the ontology of each cluster using the Bayes factor as a measure of significance.

**Table 2**

Transcription factor binding sites enriched in TEB versus duct

| # | Annotation | Total genes | Bayes factor |
|---|------------|-------------|--------------|
| 1 | **E2F1_Q6: E2F1** | 123 | 15.34 |
| 2 | KROX_Q6 | 101 | 13.37 |
| 3 | SOX9_B1: SOX (SRY-related HMG box) | 61 | 9.67 |
| 4 | SP3_Q3 | 138 | 9.39 |
| 5 | MRF2_01: modulator recognition factor 2 | 71 | 9.23 |
| 6 | MAZ_Q6 | 142 | 8.75 |
| 7 | **E2F_Q3_01** | 86 | 8.48 |
| 8 | OCT1_06: octamer-binding factor 1 | 62 | 8.11 |
| 9 | MAZR_01: MAZ related factor | 98 | 8.07 |
| 10 | SOX5_01: SOX5 | 48 | 7.81 |
| 11 | **E2F1_Q6_01** | 86 | 7.6 |
| 12 | POU6F1_01 | 1 | 6.57 |
| 13 | NMYC_01: N-MYC | 59 | 6.27 |
| 14 | ELK1_02: ELK1 | 148 | 5.59 |
| 15 | USF_01: upstream stimulating factor | 35 | 5.49 |
| 16 | **E2F_Q4: E2F** | 81 | 5.33 |
| 17 | NRF2_01: nuclear respiratory factor 2 | 48 | 5.33 |
| 18 | EVI1_04: ectopic viral integration site 1 encoded factor | 101 | 4.7 |
| 19 | **E2F1_Q3_01** | 68 | 4.67 |
| 20 | MYC_Q2 | 51 | 4.61 |

From a dataset comparing microdissected TEBs to ducts, the expression of various genes was compared for fold change [log2(Cy5 TEB/Cy3 stroma) – log2(Cy5 duct/Cy3 stroma)]. For the top 200 genes elevated in the TEB sample, the genes were analyzed using GATHER to assess which transcription factor binding sites they contained based on the TRANSFAC database. The top 20 results are shown for these 200 genes, with the number of genes from this dataset and the significance indicated by the Bayes factor. E2F factors are in bold.