Gene expression signature in organized and growth arrested mammary acini predicts good outcome in breast cancer

Marcia V. Fournier,[1,3] Katherine J. Martin,[2,3] Paraic A. Kenny,[1] Kris Xhaja,[2] Irene Bosch,[2] Paul Yaswen,[1] and Mina J. Bissell[1*]

1. Life Science Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720.
2. Center for Infectious Disease and Vaccine Research, University of Massachusetts Medical School, Worcester, Massachusetts 01655.
3. These authors contributed equally to this manuscript

Financial support: MVF, PY and MJB were supported by grants DE-AC03 SF0098, a Distinguished Fellowship award from the US Department of Energy, Office of Biological and Environmental Research, the National Cancer Institute 2 R01 CA064786-09, and an Innovator Award from the US Department of Defense Breast Cancer Research Program BC012005 to MJB. KJM, KX and IB were supported by NIAID grant U19-AI057319 to IB. PAK was supported by postdoctoral training fellowship DOD BCRP DAMD17-00-1-0224.

*Corresponding author: Mina J. Bissell, Department of Cancer Biology, Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720. E-mail: mjbissell@lbl.gov

Running title: A mammary gland differentiation signature predictive of good breast cancer outcome

Key words: microarray, 3D culture, extracellular matrix, gene expression profile, breast cancer.

ABSTRACT:

To understand how non-malignant human mammary epithelial cells (HMEC) transit from a disorganized proliferating to an organized growth arrested state, and to relate this process to the changes that occur in breast cancer, we studied gene expression changes in non-malignant HMEC grown in three-dimensional cultures, and in a previously published panel of microarray data for 295 breast cancer samples. We hypothesized that the gene expression pattern of organized and growth arrested mammary acini would share similarities with breast tumors with good prognoses. Using Affymetrix HG-U133A microarrays, we analyzed the expression of 22,283 gene transcripts in two HMEC cell lines, 184 (finite life span) and HMT3522 S1 (immortal non-malignant), on successive days post-seeding in a laminin-rich extracellular matrix assay. Both HMECs underwent growth arrest in G0/G1 and differentiated into polarized acini between days 5 and 7. We identified gene expression changes with the same temporal pattern in both lines. We show that genes that are significantly lower in the organized, growth arrested HMEC than in their proliferating counterparts can be used to classify breast cancer patients into poor and good prognosis groups with high accuracy. This study represents a novel unsupervised approach to identifying breast cancer markers that may be of use clinically.

INTRODUCTION:

Loss of growth control and disruption of tissue architecture are among the earliest hallmarks of cancer. We hypothesized that the gene expression changes that occur during the organization and growth arrest of cultured mammary acini would share similarities with breast tumors that had a good prognosis. To test this hypothesis, we examined temporal changes in gene expression in two non-malignant human mammary epithelial cells (HMEC) grown in a three-dimensional laminin-rich extracellular matrix (3D lrECM) assay (1). In this model, single non-malignant breast epithelial cells form polarized growth arrested multicellular structures that resemble acini over a period of several days. This transition from the unpolarized actively dividing state to a polarized non-dividing state is the reverse of what happens in the early stages of tumorigenesis.

We used two non-malignant HMEC, a non-immortalized HMEC strain,184 (2, 3), and spontaneously immortalized cell line, HMT3522-S1 (4), and monitored the changes in gene and protein expression as the cells formed acinar structures in 3D lrECM cultures. During the process of self-organization and withdrawal from the cell cycle, we noted a progressive hypophosphorylation of the retinoblastoma gene product (Rb) and an induction of the cyclin-dependent kinase (cdk) inhibitor $p27^{kip1}$. Analysis of the changes in gene expression in both cell lines allowed the identification of sets of commonly regulated genes.

While established breast cancer prognostic markers such as tumor size, grade, lymph node and hormone receptor status are useful in predicting survival in large populations (5-7), there is a pressing need to develop better prognostic signatures to predict recurrence and overall survival. A particular benefit would be the identification of

patients with good prognoses whose tumors are highly unlikely to recur and who nevertheless are being treated with cytotoxic chemotherapies (8). The advent of gene expression technologies has greatly aided the identification of molecular signatures with value for tumor classification and prognosis prediction (9-14) . Van de Vijver et al. have developed a 70-gene signature that effectively stratifies patients into good and poor prognosis groups (15, 16). Paik et al. have proposed a 21-gene signature with which to calculate a "recurrence score" that predicts the likelihood of recurrence in estrogen receptor (ER)-positive lymph node-negative patients (17). In each of these studies, the predictive signatures have been derived by using a training set of patients of known outcome, followed by testing these signatures in a validation set of patients. In contrast, our approach has been to identify directly the genes whose expression changes as cultured mammary epithelial cells transition from a disorganized to an organized state, and to then test as proof of principle the possible utility of these genes as prognostic markers in a validation set of patients.

MATERIALS AND METHODS:

Cell Culture.

Finite-lifespan 184 HMEC were obtained from reduction mammoplasty tissue and grown in a serum-free MCDB 170 medium (MEGM; Clonetics Division of BioWhittaker, Walkersville, MD, USA), as described previously (2, 18). HMT-3522 S1 mammary epithelial cells were cultured in H14 medium (DMEM/F12 containing 250 ng/ml insulin, 10 µg/ml transferrin, 2.6 ng/ml sodium selenite, 10^{-10} M estradiol, 1.4 x 10^{-6} M hydrocortisone, 10 ng/ml EGF and 5 µg /ml prolactin). The cells were cultured in a

3D laminin-rich extracellular matrix (Matrigel, BD Biosciences), as described (19). The colonies were isolated from the Matrigel in ice-cold PBS/5 mM EDTA after 3, 5, 7, and 10 days. For western blot analysis, the colonies were lysed in 150 mM NaCl, 1 % NP-40, 50 mM Tris, pH 8.0. RNA was isolated using the RNeasy kit (Qiagen), according to the manufacturer's instructions.

Indirect immunofluorescence and image acquisition.

Acinar structures were fixed on glass slides in methanol-acetone (1:1) at -20°C for 10 minutes and air dried. A primary block was performed in an IF Buffer (130 mM NaCl, 7 mM Na_2HPO_4, 3.5 mM NaH_2PO_4, 7.7 mM NaN_3, 0.1% bovine serum albumin, 0.2% Triton X-100, 0.05% Tween-20) + 10% goat serum for 1 hour at room temperature. A secondary block was performed in IF Buffer + 10% goat serum + 20µg/ml goat anti-mouse F(ab')2 fragment for 30 minutes. The primary antibody, rat anti-alpha-6 integrin (Chemicon International), was diluted 1:50 in the latter blocking buffer and incubated overnight (15-18 hours) at 4°C. Secondary antibodies conjugated with fluorescent dye FITC or Texas Red were diluted 1:100 in blocking buffer and incubated 1 hour at room temperature. Immunofluorescent images were acquired using an inverted microscope equipped with a digital camera and SPOT software. Confocal analysis was performed using a Zeiss 410 confocal microscopy system. The images presented are representative of two or more independent experiments. All the images were converted to TIFF format and arranged using Adobe Photoshop 5.0.

Flow cytometry.

To analyze the DNA content, acini were suspended in 300 µl 0.25% trypsin and incubated at 37°C for exactly 10 minutes. The dispersed cells were washed three times in

PBS and fixed in 40% ethanol at 4°C overnight. They were then incubated with 500 µg /ml RNaseA in PBS at 37°C for 30 minutes and stained with 69µM propidium iodide solution in PBS at room temperature for 30 minutes. The DNA content was determined by flow cytometry using a FACScan (Becton Dickinson, Hialeah, FL), and the data were analyzed with Cell Quest software (Becton Dickinson).

Immunoblotting.

Whole-cell lysates (20–100 µg protein) were separated by SDS-PAGE and transferred electrophoretically to PVDF membranes (Millipore, Billerica, MA). After blocking in 10% nonfat dry milk for 90 minutes at room temperature, they were incubated with the following antibodies at 1:500 to1:1000 dilution in 2% nonfat dry milk for 2 hours at room temperature or overnight at 4°C: rabbit anti-phosphorylated Rb at S807/811 and S795 (Cell Signaling Technology, Beverly, MA), rabbit anti-phosphorylated Rb at S612 and T821 (Biosource, Camarillo, CA), and mouse anti-p27 ^{kip1} (BD Transduction Laboratories). The blots were then incubated with HRP-conjugated sheep anti-mouse IgG or anti-rabbit IgG at 1:2000 in 2% nonfat dry milk for 90 minutes at room temperature. The blots were developed using Super Signal West Pico Chemilumiscent Substrate (Pierce, Rockford, IL).

Survival analysis.

A database consisting of the microarray profiles of 295 human breast tumors with the associated clinical data (15) was obtained from Rosetta Inpharmatics (http://www.rii.com/publications/2002/nejm.html). For survival analysis of the 19 individual marker genes, the patients were stratified into quartiles for expression of each marker, and the survival curves were computed using the method of Kaplan and Meier.

Statistical significance was determined using the log-rank test. Statistical analyses were performed using Graphpad Prism. For survival analysis of the set of 249 marker genes, the patients were stratified into two groups using GeneSpring software by hierarchical cluster analysis with a distance metric of the expression pattern of all 249 genes. Kaplan-Meier survival curves, log-rank statistics, and the estimated hazard ratio for these two groups were computed using the Excel add-in EcStat.

Microarray hybridization and analysis.

Cell samples were harvested in duplicate at three time points, 3, 5, and 7 days, after seeding in lrECM. Purified total cellular RNA was biotin-labeled and hybridized to human oligonucleotide microarrays (Affymetrix HG-U133A), as previously described (20). Experiments with Affymetrix-present P-call rates of >30% were included in the analysis. Signal values from each of the 22,283 probe sets were calculated by means of robust multi-array analysis (RMA) (21) using Bioconductor (http://www.bioconductor.org) in the R computing environment. The signal values were inverse log2 transformed and then imported into GeneSpring software (SiliconGenetics); and each array was normalized to its median signal intensity. The genes were normalized to the mean of the 3-day time point for each cell type independently. For Method 1 of selecting significantly differential temporally co-regulated genes (Figure 2): Significantly up-regulated genes in each cell specimen were identified by first selecting the genes induced at least 1.5 fold in at least one of the six conditions and then performing an ANOVA analysis as a function of time. Variances were calculated using the cross-gene error model (GeneSpring), p-value cutoff 0.05, multiple testing correction: Benjamini and Hochberg False Discovery Rate. About 5% of the identified genes in each set would be

expected to pass this restriction by chance. Significantly down-regulated genes were identified in the same manner after normalizing to the 7-day time point. Genes that were up- or down-regulated early in each cell line were selected from the significantly up- or down-regulated gene lists. The early genes were defined as those with a mean expression at 5 days of at least 50% of their mean expression at 7 days. For Method 2 of selecting significantly differential, temporally co-regulated genes (Figure 6A): We first selected all genes that were at least 1.5 fold differential (up or down) in at least one of the four samples from the two later time points (days 5 and 7) for either cell specimen. We then performed an ANOVA as a function of time. Variances were not assumed equal (Welch ANOVA), p-value cutoff 0.05, multiple testing correction: Benjamini and Hochberg False Discovery Rate. From this list of genes that were significantly differential in either cell line, we then identified genes that were up- or down-regulated early (mean expression at 5 days of at least 50% of their mean expression at 7 days) or late in each cell line. We then identified those genes that were coordinately regulated in both cell lines.

RESULTS:

Temporal analysis of gene expression in two non-malignant human breast epithelial cells grown in 3D lrECM cultures.

To identify consistent changes in gene and regulatory protein expression levels, we characterized two independently derived non-malignant HMECs: one finite life-span strain (184) (2) and one spontaneously immortalized line (HMT3522 S1) (4). Both cells formed acinus-like structures with similar morphology and basal polarity when cultured

from single cells in lrECM (Figure 1A and 1B). We performed temporal studies to determine when, and in what order, changes in critical cell cycle regulatory molecules occurred in non-malignant HMECs in 3D lrECM cultures. Flow cytometric analyses indicated that, after undergoing a limited number of cell divisions, the majority of non-malignant S1 cells accumulated in the G0/G1 phase of the cell cycle by day 7 (Figure 1C). The product of the retinoblastoma susceptibility gene (Rb), a central player in the G1/S transition, is inactivated by phosphorylation, allowing cell cycle progression. Rb inactivation occurs through the sequential actions of cyclin D/cdks 4 and 6, and of cyclin E-cdk2 complexes (22). We analyzed cell cycle regulators known to affect G1 checkpoints in a time-dependent manner in S1 cells in 3D cultures. Phosphorylation of several sites on Rb was found to gradually decrease between days 5 and 10 in these non-malignant cells, consistent with the growth-suppressive role of the hypophosphorylated form (Figure 1D). Cyclins D1, E, and A, as well as their binding partners—cdks 4, 6, and 2—also decreased during this period (data not shown). In contrast, the cdk inhibitor $p27^{kip1}$ increased between days 5 and 10 (Figure 1D). The pronounced down-modulation of Rb phosphorylation and the elevation of $p27^{kip1}$ protein levels were changes observed in both the S1 and 184 cells (data not shown). Although other studies have used different HMEC (MCF10A) to determine how cells can escape normal growth control (23-26), the mechanisms by which mammary cells actually initiate and maintain growth arrest during the process of acini formation in the context of 3D-lrECM remain to be determined.

Global gene expression analysis of the time course of HMECs in 3D.

To probe systematically the molecular changes that accompany acinus formation, we analyzed the expression profiles of 22,283 transcripts using Affymetrix HG-U133A

microarrays. Microarray experiments were performed with biological duplicates using RNA samples harvested from S1 and 184 cells, after 3, 5, and 7 days' culture in lrECM. Since growth arrest and polarization occurred with similar kinetics in both cell types (Figure 3A), we reasoned that the gene expression changes important for these processes would follow a common temporal pattern in both cell lines, and that changes that were cell type-specific could be disregarded.

We first identified genes that showed at least 1.5-fold changes during the time course in the individual cell specimens (ANOVA, $p < 0.05$) (within this window, 363 genes were up-regulated and 117 genes down-regulated in 184 cells; 234 genes were up-regulated and 351 genes down-regulated in S1 cells). We then divided these lists into genes whose expression changed 'early' by our definition (between days 3 and 5) or 'late' (between days 5 and 7) in S1 and 184. Finally, we identified the genes from each temporal group that were common to both cell types (Figure 2). A total of 60 genes with common temporal patterns were identified, including 21 genes that were up-regulated early, 11 genes that were up-regulated late, 6 genes that were down-regulated early, and 22 genes that were down-regulated late (Figure 3B and 3C and supplementary material).

Correlation of the differentially expressed genes with survival of breast cancer patients.

To relate the process of acinar development in 3D lrECM cell cultures to the changes that occur in breast cancer, we examined the expression levels of the differentially regulated genes identified using our model using previously published microarray data for a panel of 295 breast cancer samples from the fresh-frozen-tissue bank of the Netherlands Cancer Institute, including 151 lymph node-negative and 144

lymph node-positive patient samples (15). Fifty five of the 60 genes selected in our 3D culture analysis were included on these microarrays. We looked at 5- and 10-year survival data and applied Student's t-test to determine how many of the genes modulated in 3D cultures showed survival-associated expression changes. T-tests were performed to determine whether the difference in the expression level of a given gene in two groups (e.g., patients who survived five years versus patients who did not) was large enough that it was not likely to be due to chance. The numbers and percentages of genes exhibiting significantly different expression in the tumors of patients with differential survival ($p < 0.05$) were tabulated for a) all the genes represented on the microarrays, b) genes selected on the basis of differential expression during the 3D lrECM timecourse, or c) randomly generated gene lists (Table 1). The percentage of genes with survival-associated expression changes was highest for those genes down-regulated late (between days 5 and 7) in the time course. The percentage for this gene list exceeded those of the unfiltered list of all 25,773 genes represented on the arrays, 5 random gene lists, and all other 3D lrECM gene lists. The list of genes that were down-modulated late in the lrECM timecourse showed a marked enrichment in genes whose expression level correlated to 5 (68%) and 10 years survival (53%). The levels of the majority of the late down-regulated genes were higher in patients who died within 5 or 10 years (ACTB, VRK1, ODC1, CKS2, FLJ10036, FLJ10540, FOXM1, RRM2, TRIP13, CDKN3, STK6, FLJ10517, TUBG1, ACTN1, TNFRSF6B and EPH2), while the levels of three genes (DUSP4, HBP17 and EIF4A1) were lower in these patients.

We identified 22 genes in all that were down-regulated in both HMECs between days 5 and 7 of lrECM culture. Of these, 19 genes were represented in the published data

set for the 295 patient tumor samples. We stratified the 295 tumors into quartiles based on the relative expression level of each of the genes in the selected set, and further analyzed the relationship of the expression level of each individual gene to survival (Figure 4). The resulting Kaplan-Meier curves showed that gene expression levels correlated significantly with outcome for 14 of the 19 selected markers. For 13 of the 14 markers, gene expression was lower in tumors from patients with better outcomes, while in one case (DUSP4) gene expression was lower in tumors from patients with poorer outcomes.

Collective gene signatures have the potential to discriminate among clinical endpoints more accurately than markers used individually. Hence, we tested the ability of our set of 19 genes to classify breast cancer patients into prognostic groups. We used hierarchical cluster analysis to separate the patients into groups and then determined the overall 10-year survival rates for these groups (Figure 5). The cluster analysis separated the patients into five groups, three of which had tumors that expressed comparatively lower levels of most of the 19 genes, and two of which expressed higher levels. The 10-year survival rates for these 5 groups were 95, 84, 67, 61, and 54% respectively.

To test whether other sets of genes down-regulated late in the lrECM timecourse identified by using other selection strategies would also include useful breast cancer markers, we applied a second selection strategy (Figure 6A) and tested the ability of the resulting gene set to predict breast cancer prognosis. This second method was less restrictive than the first, and resulted in the identification of 287 genes that were significantly down-regulated late in the 3D time course of both HMEC specimens (for complete gene list and gene expression information, see supplemental information).

Seventeen of the 22 genes selected using Method 1 were also included in the 287 genes selected using Method 2. We tested the ability of these genes to predict breast cancer prognosis by using hierarchical cluster analysis in the same set of previously published microarray data from 295 breast cancers (15). A large majority, 249 of the 287 genes were included on these microarrays. Of the 17 overlapping (methods 1 and 2) genes on the Affymetrix chips, 15 were present on the Rossetta chips: ACTB, ACTN1, CDK3, CKS2, DUSP4, EPHA2, HBP17, FOXM1, ODC1, RRM2, STK6, TNFRSF6B, TRIP13, TUBG1, VRK1.

Hierarchical cluster analysis using the 249 gene signature classified the samples into two groups of approximately equal numbers of tumors (Figure 6B). Overall 10 year survival rates were 90% (138 of 154) for the good prognosis group and 59% (83 of 141) for the poor prognosis group. To assess the significance of these predictions and take into account patients that could not be followed the entire length of the study, we performed a Kaplan-Meier analysis. The results show that the 249 gene profile was highly informative in identifying patients with poor outcome (log rank $p = 2.7 \times 10^{-10}$) (Figure 6C). The estimated hazard ratio for poor outcome (failure to survive) in the group with the poor prognosis signature as compared with the good prognosis signature was 4.7 [95% confidence interval 2.8 – 7.9].

DISCUSSION:

Three-dimensional (3D) laminin-rich (lrECM) cultures permit non-malignant cells to exhibit self-organizing properties. Such cultures provide models that allow the

study of processes that are aberrant in breast cancer (1, 19, 27, 28). To understand how breast epithelial cells transition from a proliferating, unorganized state to a resting, organized state, and to relate this process to the opposing changes that occur in clinical breast cancer, we performed genome-wide gene expression profiling for two non-malignant HMECs in 3D cultures, and utilized published data from a panel of 295 breast cancer samples. We show that genes that are down-regulated as HMEC transition from a proliferative to an acinar phenotype can be used collectively as signatures to predict clinical breast cancer prognosis.

Our approach represents a new way to identify genome-wide cancer prognostic markers. We based all the marker selection steps on HMEC cultured in 3D lrECM. The 3D model system provided a means to focus on epithelial cells themselves, and a defined, highly relevant biological process - the formation of breast acini. Whereas the stroma is absent, the 3D lrECM assays appears to substitute for myoepithelial cells and other signals that are needed to form an organized acini (29). Differentially expressed genes identified using this model system are likely to be functionally linked to the transformation-relevant process. Further, we have applied an unsupervised method (hierarchical cluster analysis) to classify the patient samples using selected markers. Hence, neither our method of marker selection nor our sample classification method relies on any clinical information.

Gene-expression profiling of tumors using DNA microarrays is a promising method for predicting prognosis and treatment response in cancer patients (17, 30-32). Two studies have recently employed genome-wide microarray analysis to identify gene signatures that predict prognosis in breast cancer. The profiles studied by both groups of

researchers were reported to be more powerful predictors of the outcome of disease than standard systems based on clinical and histological criteria. The study by van't Veer et al. (16) used the supervised classification of a primary data set of 78 tumor samples to identify a 70-gene signature that divided the samples into classes of poor and good prognosis. Validation of the classifier by a second overlapping set of 295 tumors showed that it accurately predicted 10-year survival in breast cancer patients (15). Overall, the 10-year survival rates were 54.6% ±4.4% and 94.5% ±2.6% for the poor and good prognosis groups, respectively. Their estimated hazard ratio for poor outcome (distant metastases) in the group with a poor-prognosis signature, as compared with the group with the good-prognosis signature, was 5.1 (95 % confidence interval, 2.9 to 9.0; $P<0.001$). In a similar study, Wang and collaborators (33) also used supervised classification in a training set of 115 breast tumor samples. They identified a 76-gene signature, which was verified in an independent set of 171 breast tumors. Their estimated hazard ratio for poor outcome (distant metastases within 5 years) in the group with the poor prognosis signature as compared with the group with the good prognosis signature was 5.7 (95% confidence interval, 2.6 to 12.4). Our 249-gene signature predicted 10-year survival rates of 59% and 90% for poor and good prognosis groups, respectively. The estimated hazard ratio for poor outcome (failure to survive) in the group with the poor prognosis signature as compared with the good prognosis signature was 4.7 (95% confidence interval 2.8 – 7.9).

Our 249 gene signature overlapped by 11 genes with the 70 gene signature of van't Veer et al. (16) (including NUSAP1, UCHL5, RAMP, DC13, PRC1, COL4A2, PITRM1, CENPA, MELK, KNTC2, and MCM6) and by 7 genes with the 76 gene

signature of Wang et al (33) (including MTB, POLQ, SUPT16H, FEN1, DUSP4, PLK1, and SMC4L1). The van't Veer and Wang signatures overlapped by a single gene (cyclin E2). Our 19 gene signature had no genes in common with either of the previously published signatures.

Our 19-gene signature included several genes encoding proteins with roles in the cell cycle and in cell division. . Cell-cycle genes were previously identified as important markers of prognosis in ER-positive younger patients (34). In this earlier study, 50 cell-cycle-related genes divided 83 ER-positive younger patients into two groups of good versus poor prognosis. The overall 10-year survival rates were 46% and 96% for the poor and good prognosis groups, respectively. Similarly, we found that a core group of predominantly cell cycle and mitotic organizing center genes (CDKN3, RRM3, FLJ10540, FOXM1, STK6, TRIP13, EIF4A1, FLJ10036, VRK1, TUBG1, CKS2, FLJ10517) made a strong contribution to stratifying tumors into good versus poor prognostic groups. One gene from our 19-gene signature, STK6, which encodes Aurora-A (35, 36) was also included in the 50-gene signature of Dai *et al.* (34), while 15 genes from our 249-gene signature were included in the Dai *et al.* signature (including BM039, DKFZp762E1312, LOC51203, LOC51659, ID-GAP, KNSL6, PRC1, STK6, CDC45L, SNRPA1, H2AFZ, CENPA, CDC6, BIRC5, and BLM). In addition to cell-cycle genes, our prognostic genes also encoded products with other functions, including genes involved in cytoskeletal regulation (ACTB, ACTN1) (37), cell survival (TNFRSF6B) (38, 39), polyamine biosynthesis (ODC1) (40), and cell-cell interactions (EPHA2) (41). The genes in these additional functional groups were important in subdividing the patients into subgroups with differing survival rates.

In conclusion, we report that the gene expression changes that commonly occur in non-malignant HMEC grown in 3D lrECM cultures provide gene expression signatures that effectively stratify patients into prognostic groups according to overall survival rates. Our 249 gene signature achieved a hazard ratio of 4.7, which is comparable to hazard ratios achieved by large scale supervised breast cancer microarray studies. Our results underscore the relevance of 3D lrECM cultures for studies of malignant transformation, and suggest potentially valuable new biomarkers for further clinical evaluation.

ACKNOWLEDGEMENTS:

We would like to thank Martha Stampfer for kindly providing the 184 HMEC and Saira Mian and her for critical reading of the manuscript. We are grateful to other members of our laboratories for fruitful discussions. This work was supported by grants (DE-AC03 SF0098), and a Distinguished Fellowship award from the U.S. Department of Energy, Office of Biological and Environmental Research, the National Cancer Institute (2 R01 CA064786-09), and an Innovator Award from the U.S. Department of Defense Breast Cancer Research Program (BC012005) to M.J.B.; by a NIAID grant (U19-AI057319) to I.B.; and by a postdoctoral training fellowship (DOD BCRP DAMD17-00-1-0224) to P.A.K.

BIBLIOGRAPHY

1. Petersen OW, Ronnov-Jessen L, Howlett AR, and Bissell MJ. Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells. Proc Natl Acad Sci U S A 1992; 89: 9064-8.
2. Hammond SL, Ham RG, and Stampfer MR. Serum-free growth of human mammary epithelial cells: rapid clonal growth in defined medium and extended serial passage with pituitary extract. Proc Natl Acad Sci U S A 1984; 81: 5435-9.
3. Stampfer MR and Yaswen P. Culture systems for study of human mammary epithelial cell proliferation, differentiation and transformation. Cancer Surv 1993; 18: 7-34.
4. Briand P, Petersen OW, and Van Deurs B. A new diploid nontumorigenic human breast epithelial cell line isolated and propagated in chemically defined medium. In Vitro Cell Dev Biol 1987; 23: 181-8.
5. Hayes DF, Isaacs C, and Stearns V. Prognostic factors in breast cancer: current and new predictors of metastasis. J Mammary Gland Biol Neoplasia 2001; 6: 375-92.
6. Clark GM. Prognostic and Predictive Factors for Breast Cancer. Breast Cancer 1995; 2: 79-89.
7. Fitzgibbons PL, Page DL, Weaver D, et al. Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med 2000; 124: 966-78.
8. Goldhirsch A, Glick JH, Gelber RD, Coates AS, and Senn HJ. Meeting highlights: International Consensus Panel on the Treatment of Primary Breast Cancer. Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. J Clin Oncol 2001; 19: 3817-27.
9. Liang P and Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 1992; 257: 967-71.
10. Velculescu VE, Zhang L, Vogelstein B, and Kinzler KW. Serial analysis of gene expression. Science 1995; 270: 484-7.
11. Martin KJ, Graner E, Li Y, et al. High-sensitivity array analysis of gene expression for the early detection of disseminated breast tumor cells in peripheral blood. Proc Natl Acad Sci U S A 2001; 98: 2646-51.
12. Nacht M, Ferguson AT, Zhang W, et al. Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. Cancer Res 1999; 59: 5464-70.
13. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000; 406: 747-52.
14. Ramaswamy S and Golub TR. DNA microarrays in clinical oncology. J Clin Oncol 2002; 20: 1932-41.
15. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002; 347: 1999-2009.

16. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415: 530-6.
17. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004; 351: 2817-26.
18. Stampfer MR and Bartley JC. Induction of transformation and continuous cell lines from normal human mammary epithelial cells after exposure to benzo[a]pyrene. Proc Natl Acad Sci U S A 1985; 82: 2394-8.
19. Weaver VM, Petersen OW, Wang F, et al. Reversion of the malignant phenotype of human breast cells in three-dimensional culture and in vivo by integrin blocking antibodies. J Cell Biol 1997; 137: 231-45.
20. Warke RV, Xhaja K, Martin KJ, et al. Dengue virus induces novel changes in gene expression of human umbilical vein endothelial cells. J Virol 2003; 77: 11822-32.
21. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003; 4: 249-64.
22. Sherr CJ and Roberts JM. CDK inhibitors: positive and negative regulators of G1-phase progression. Genes Dev 1999; 13: 1501-12.
23. Muthuswamy SK, Li D, Lelievre S, Bissell MJ, and Brugge JS. ErbB2, but not ErbB1, reinitiates proliferation and induces luminal repopulation in epithelial acini. Nat Cell Biol 2001; 3: 785-92.
24. Debnath J and Brugge JS. Modelling glandular epithelial cancers in three-dimensional cultures. Nat Rev Cancer 2005; 5: 675-88.
25. Debnath J, Mills KR, Collins NL, Reginato MJ, Muthuswamy SK, and Brugge JS. The role of apoptosis in creating and maintaining luminal space within normal and oncogene-expressing mammary acini. Cell 2002; 111: 29-40.
26. Debnath J, Walker SJ, and Brugge JS. Akt activation disrupts mammary acinar architecture and enhances proliferation in an mTOR-dependent manner. J Cell Biol 2003; 163: 315-26.
27. Barcellos-Hoff MH, Aggeler J, Ram TG, and Bissell MJ. Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane. Development 1989; 105: 223-35.
28. Schmeichel KL and Bissell MJ. Modeling tissue-specific signaling and organ function in three dimensions. J Cell Sci 2003; 116: 2377-88.
29. Gudjonsson T, Villadsen R, Nielsen HL, Ronnov-Jessen L, Bissell MJ, and Petersen OW. Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. Genes Dev 2002; 16: 693-706.
30. Weigelt B, Hu Z, He X, et al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer Res 2005; 65: 9155-8.
31. Braun S, Vogl FD, Naume B, et al. A pooled analysis of bone marrow micrometastasis in breast cancer. N Engl J Med 2005; 353: 793-802.
32. Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 2005; 102: 3738-43.

33. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005; 365: 671-9.
34. Dai H, van't Veer L, Lamb J, et al. A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. Cancer Res 2005; 65: 4059-66.
35. Kimura M, Matsuda Y, Eki T, et al. Assignment of STK6 to human chromosome 20q13.2-->q13.3 and a pseudogene STK6P to 1q41-->q42. Cytogenet Cell Genet 1997; 79: 201-3.
36. Kimura M, Kotani S, Hattori T, et al. Cell cycle-dependent expression and spindle pole localization of a novel human protein kinase, Aik, related to Aurora of Drosophila and yeast Ipl1. J Biol Chem 1997; 272: 13766-71.
37. Gumbiner BM. Regulation of cadherin-mediated adhesion in morphogenesis. Nat Rev Mol Cell Biol 2005; 6: 622-34.
38. Pitti RM, Marsters SA, Lawrence DA, et al. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. Nature 1998; 396: 699-703.
39. Yu KY, Kwon B, Ni J, Zhai Y, Ebner R, and Kwon BS. A newly identified member of tumor necrosis factor receptor superfamily (TR6) suppresses LIGHT-mediated apoptosis. J Biol Chem 1999; 274: 13733-6.
40. Kay JE and Cooke A. Ornithine decarboxylase and ribosomal RNA synthesis during the stimulation of lymphocytes by phytohaemagglutinin. FEBS Lett 1971; 16: 9-12.
41. Lindberg RA and Hunter T. cDNA cloning and characterization of eck, an epithelial cell receptor protein-tyrosine kinase in the eph/elk family of protein kinases. Mol Cell Biol 1990; 10: 6316-24.

Tables

Table 1. Number of potential prognostic markers for breast cancer among the genes that were differential in the 3D time course (Method 1).

	Total no. of genes	Number of significantly* differential genes when Survival (5 yrs)	Survival (10 yrs)
All genes	25773	2418 (9.4%)	67 (0.26%)
Early up	20	7 (35%)	5 (25%)
Early down	5	1 (20%)	1 (20%)
Late up	11	3 (27%)	1 (9%)
Late down	19	**13 (68%)**	**10 (53%)**
Random list 1	20	0	1 (5%)
Random list 2	20	5 (25%)	3 (15%)
Random list 3	20	4 (20%)	0
Random list 4	20	0	0
Random list 5	20	0	0

*Student's t-test, $p<0.05$.
#Total number reflects genes included on the Rosetta microarrays; hence values in some cases are less than the total number included on the Affymetrix microarrays.
Bold: > 50%.

FIGURE LEGENDS

Figure 1. HMEC cultured in lrECM form polarized structures and arrest growth in G0/G1. A. Phase-contrast image of typical acinar structures formed by 184 cells at day 7 in lrECM. Structures reach dimensions of 20-40 µm diameter. Bars equal 25 µm. B. Indirect immunofluorescent image showing basal polarity of alpha-6 integrin (green) in 184 cells at day 7 in lrECM. Cell nuclei were stained with DAPI (red). C. Flow cytometric analyses of propidium iodide stained S1 cells indicated that the majority of the cells accumulated in the G0/G1 phase of the cell cycle by day 7 in lrECM. D. Western blot analyses of cell-cycle regulatory molecules in S1 cells at days 5, 7, 10, and 15 after suspension in lrECM indicated that total as well as specific phosphorylated forms (serines 807/811, serine795, serine 612, threonine 821) of Rb decreased , while p27 increased, over the time course. Ponceau staining indicated that equivalent amounts of total protein were loaded in each lane.

Figure 2. A scheme is provided outlining Method 1 used to select sets of temporally co-regulated genes.

Figure 3. Sixty genes were determined by Method 1 to show significant changes in expression in both HMEC specimens during the time course of culture in lrECM. A. Temporal pattern of growth arrest in S1 and HMEC184 cells. The percentage of cells in S-phase decreased significantly between days 3 and 7. B. Expression levels of the 60 genes that were coordinately expressed in both cell types. The genes were grouped into categories based on whether they showed coordinate up-regulation early, up-regulation

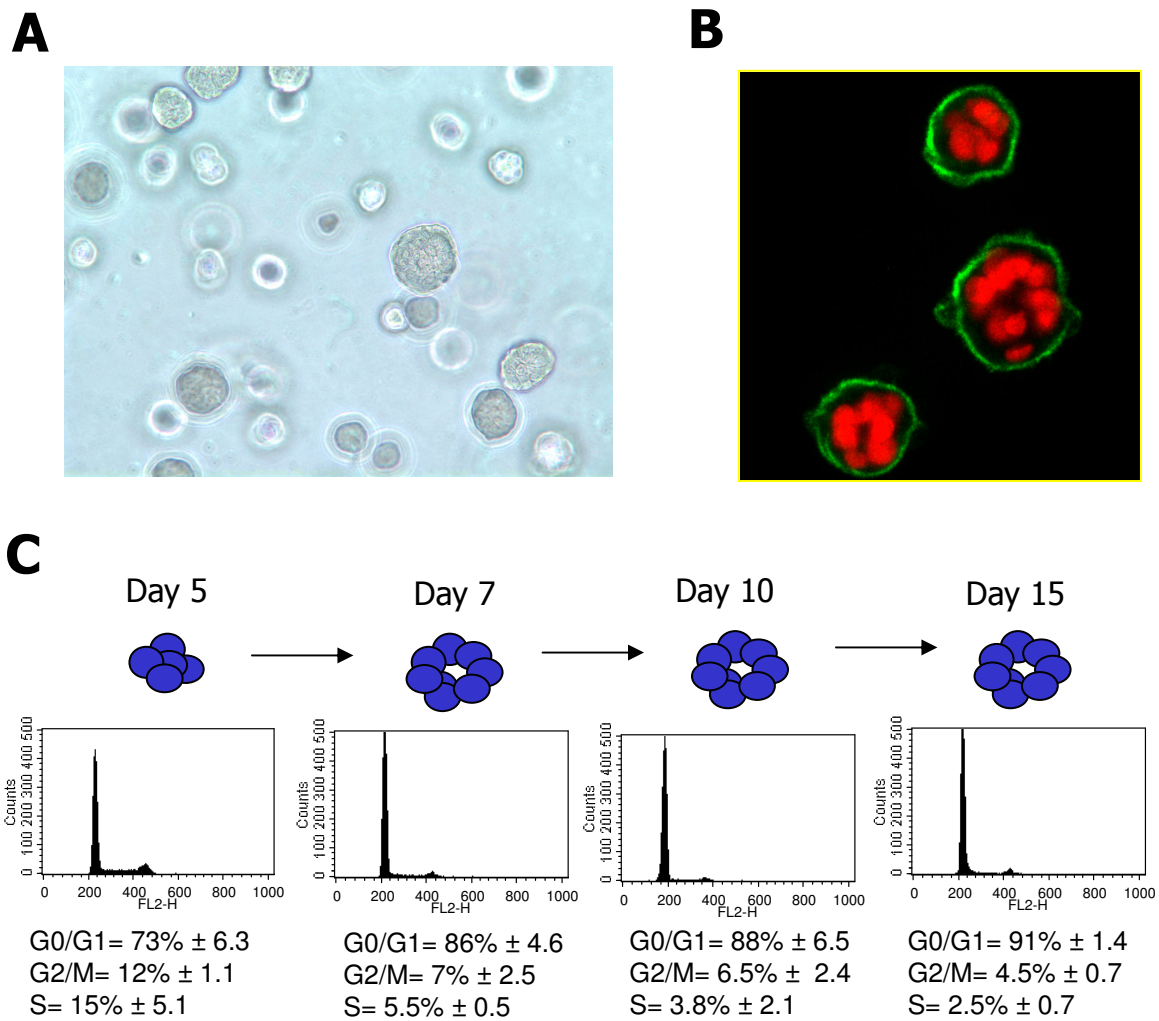
late, down-regulation early, or down-regulation late during the time course (see text for details). Note that the scale is logarithmic. C. Changes in the expression levels of the 60 individual genes are plotted and organized by hierarchical cluster analysis. Red = up-regulated, blue = down-regulated genes. Complete gene names and GenBank IDs are available (see supplementary material).

Figure 4. Fourteen of the nineteen genes down-regulated late during acinar morphogenesis showed significant correlations with patient survival. 295 patients were grouped into quartiles based on the relative expression of each selected gene in their corresponding tumors. Survival of each quartile was plotted according to the method of Kaplan and Meier. P-values represent the outcomes of the log-rank tests between the upper and lower quartiles.

Figure 5. The set of 19 genes down-regulated late during acinar morphogenesis can be used to accurately cluster 295 breast cancer samples into prognostic groups. The rows represent the relative expression levels of the genes listed on the right. The columns represent individual patients. The scale of relative expression levels is the same as that shown in Figure 3C. Genes and tumor samples were arranged by a hierarchical cluster analysis using a Pearson metric. Dendrograms at the top and left reflect the degree of relatedness of the samples and genes, respectively. Dendrogram branch colors at the top indicate different prognostic groups. Genes in bold were significantly associated with survival (Kaplan-Meier analysis, $p < 0.05$, Figure 4). The panel below the cluster diagram indicates clinical parameters for the 295 patient samples. The black regions represent the

the percentages of patients surviving 10 years in each of the five prognostic groups are shown below.

Figure 6. A second set of 249 genes down-regulated late in the time course of HMECs in 3D lrECM cultures classifies good versus poor prognosis in breast cancer patients. A. A scheme is provided outlining Method 2 used to select sets of temporally co-regulated genes. B. The set of 249 down-regulated genes identified by Method 2 clustered the 295 breast cancer samples into two prognostic groups. Genes are arranged by a hierarchical cluster analysis using a Pearson metric and samples are arranged using a distance metric. Clinical parameters are indicated below the diagram. C. Overall survival in the two groups was plotted by the method of Kaplan and Meier. P-value represents the outcome of the log-rank tests between the good and poor prognosis groups.



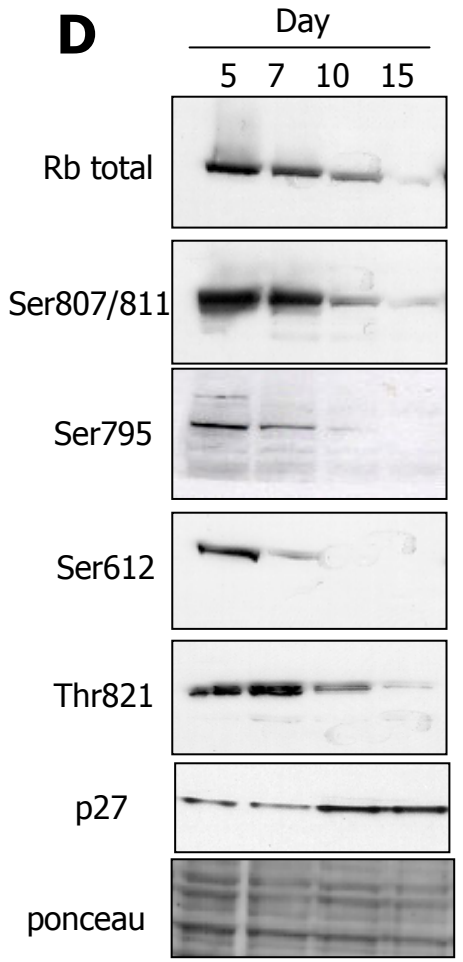
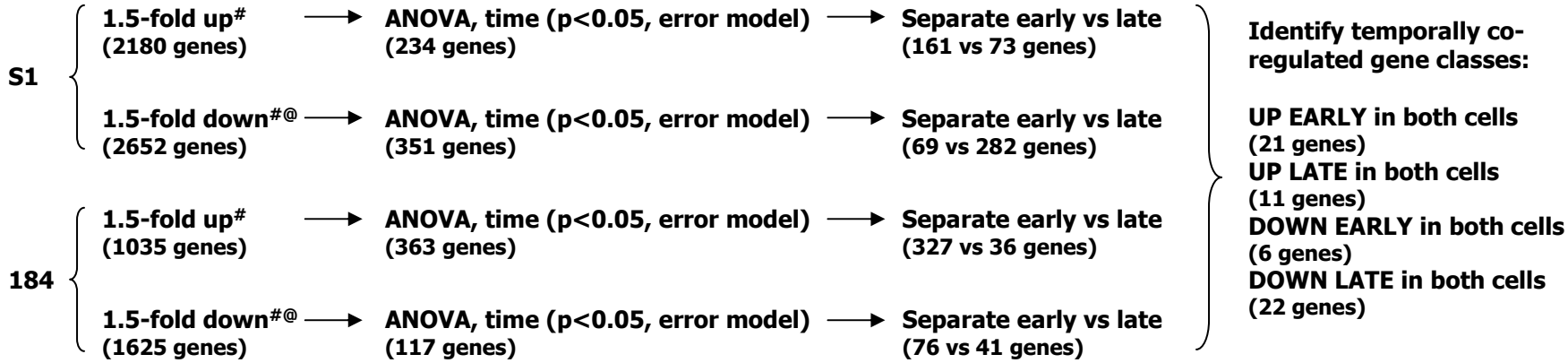


Figure 1. Fournier et al.

Identification of temporally co-regulated genes (Method 1):



#Genes at least 1.5-fold differential in at least 1 of 6 samples (2 experimental duplicates per time point).
@The GeneSpring Cross Gene Error Model works by fitting a smooth curve to replicate variability resulting in error estimates that are based on pooling variability information for genes at similar expression levels. Since this method preferentially identifies positively changing (up-regulated) genes, the mean of the 7 day samples was used as the reference in cases where we wished to identify down-regulated genes. In all other cases, the mean of the 3 day samples was used as the reference.

Figure 2. Fournier et al.

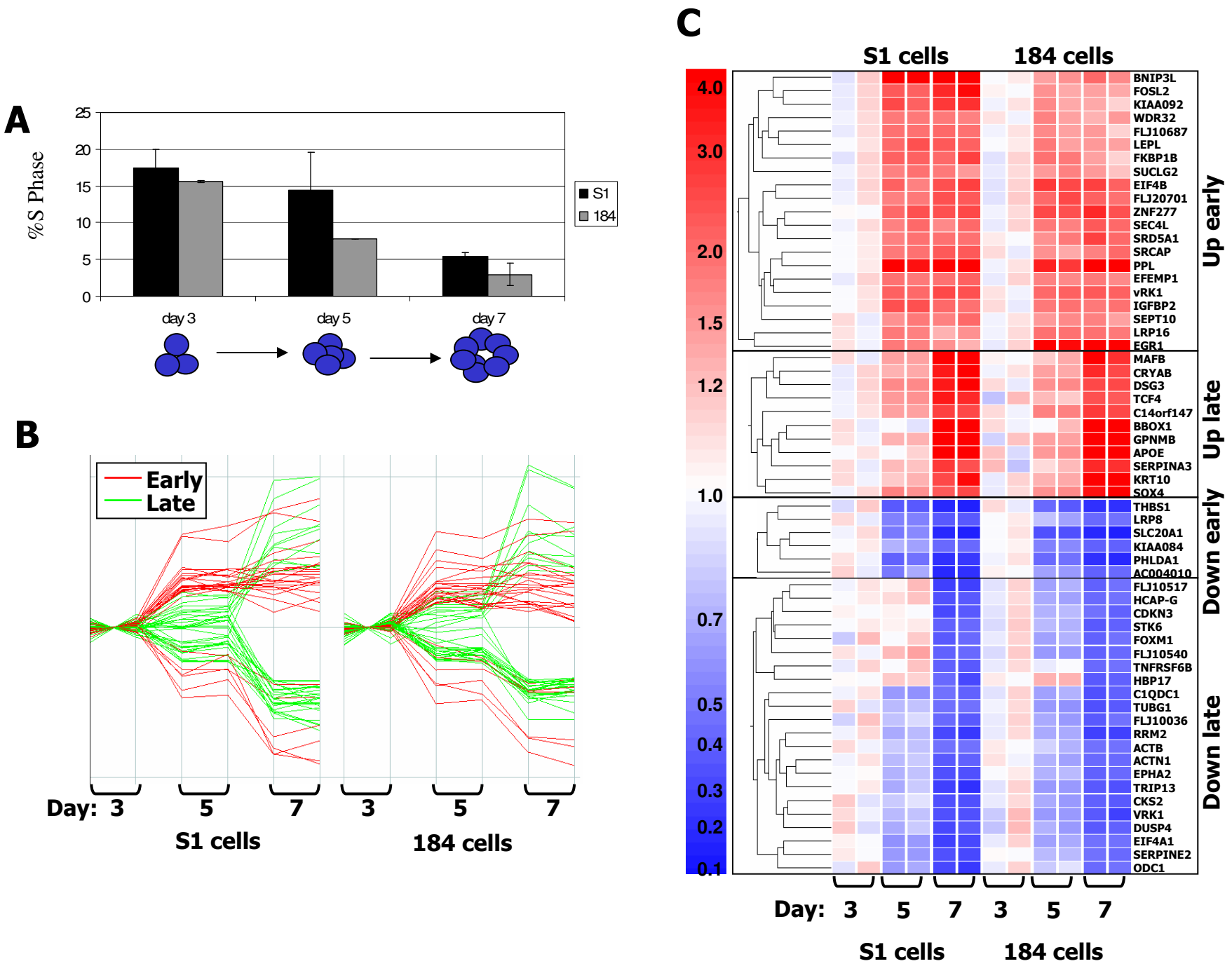


Figure 3. Fournier et al.

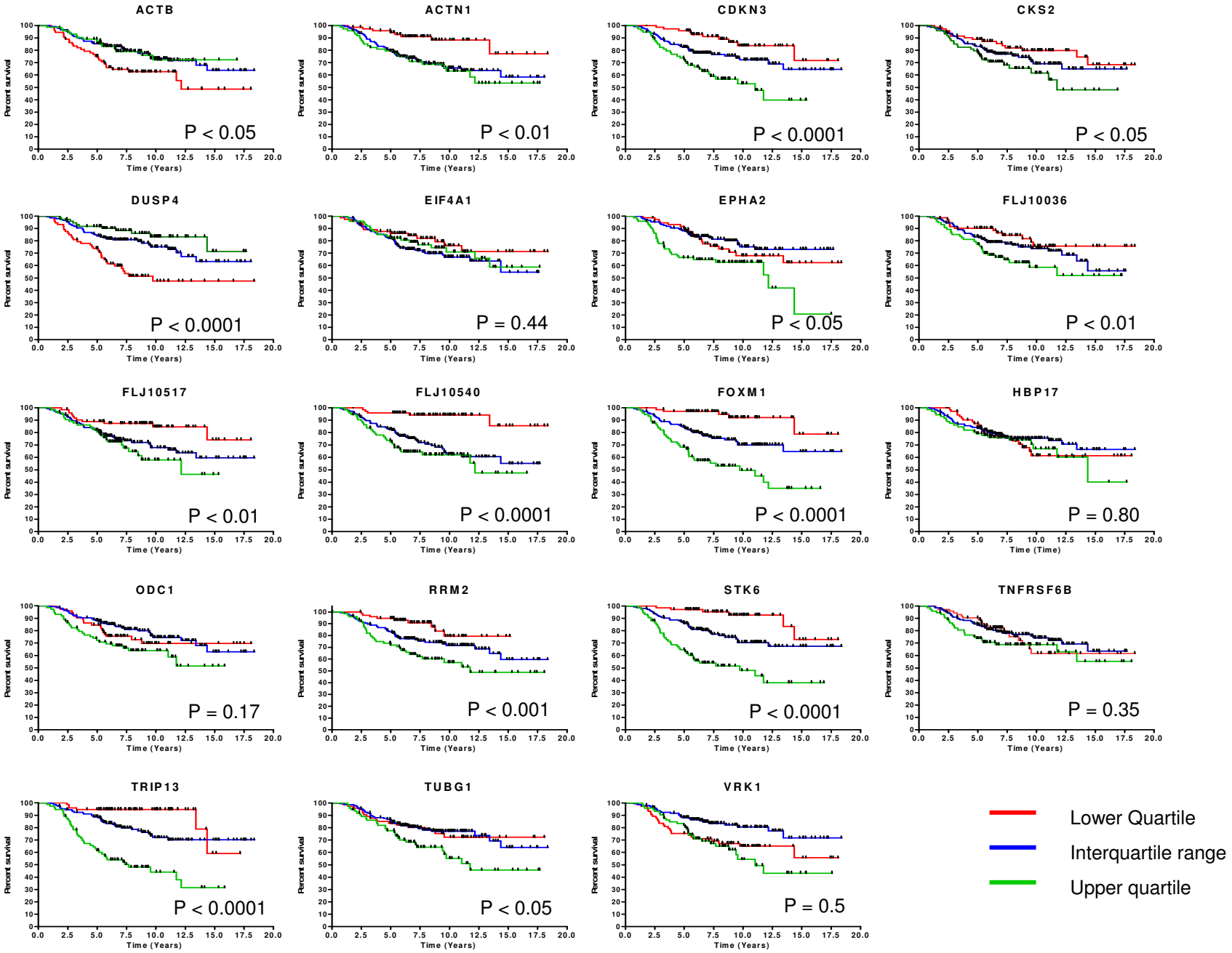


Figure 4. Fournier et al.

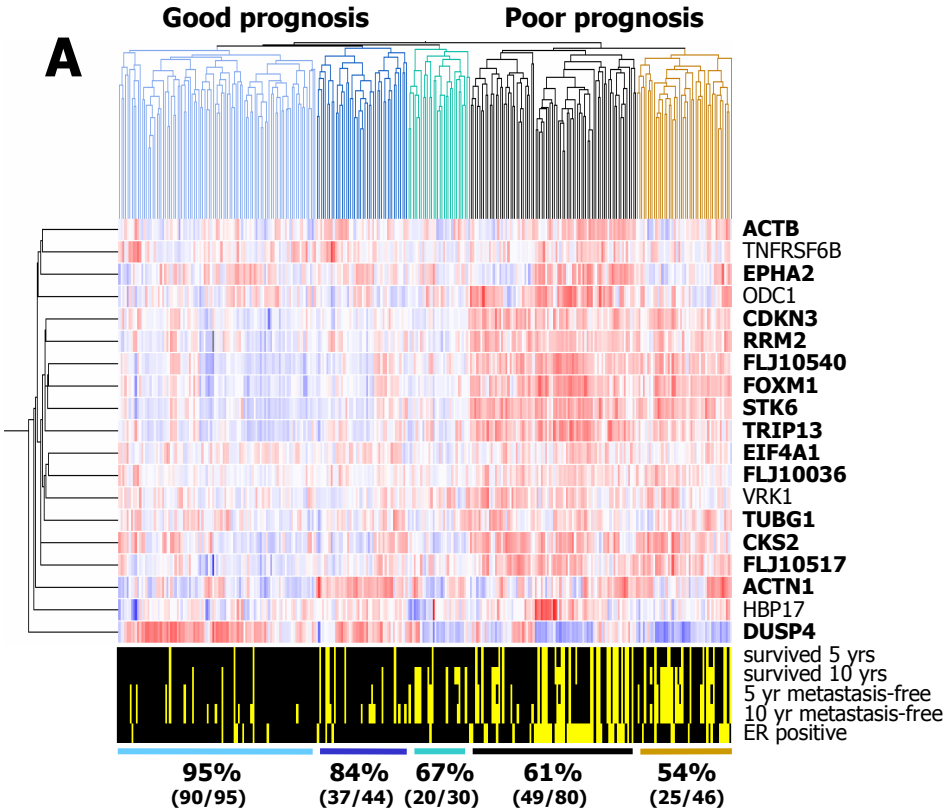
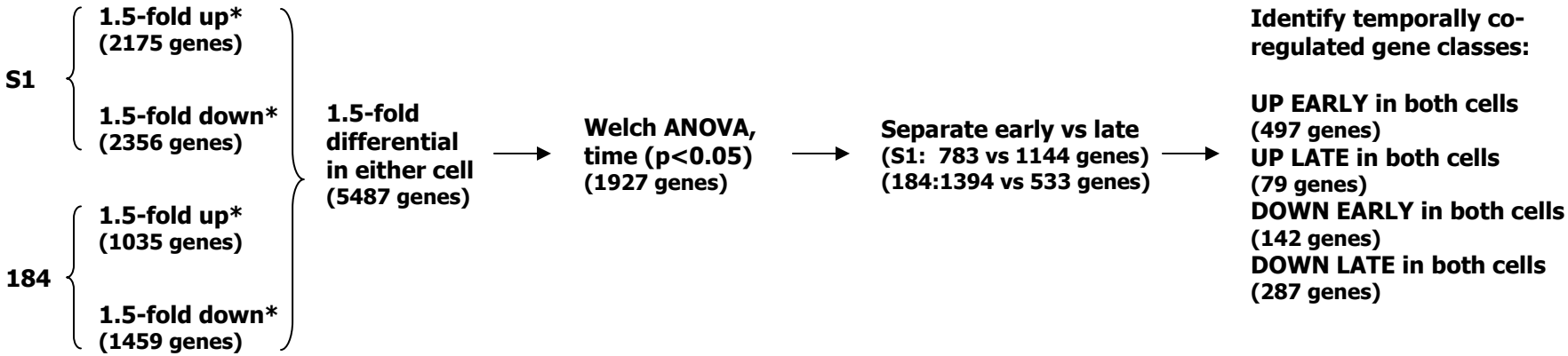


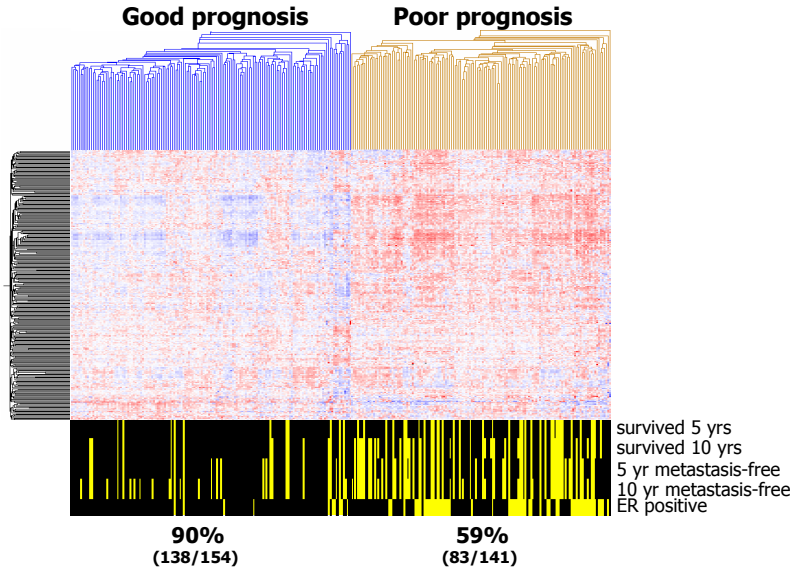
Figure 5. Fournier et al.

A **Identification of temporally co-regulated genes (Method 2):**



*Genes at least 1.5-fold differential in at least 1 of the later time points, 5 and 7 days.

B



C

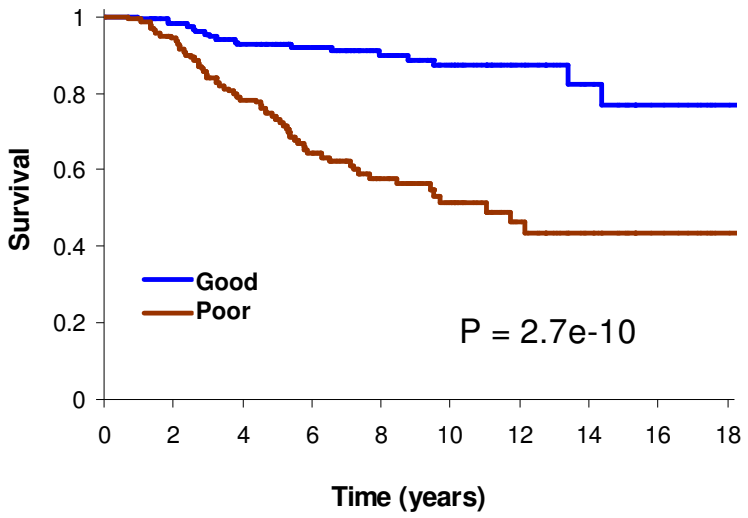


Figure 6. Fournier et al.