Age-associated gene expression in normal breast tissue mirrors qualitative age-at-incidence patterns for breast cancer

Jason R. Pirone[2*], Monica D'Arcy[1*], Delisha A. Stewart[3], William C. Hines[4], Melissa Johnson[7], Michael N. Gould[5], Paul Yaswen[4], D. Joseph Jerry[6,7], Sallie Smith Schneider[7], Melissa A. Troester[1,2,3].

[1]Department of Epidemiology, [2]Lineberger Comprehensive Cancer Center, [3]Department of Pathology and Lab Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States.
[4]Department of Cancer and DNA Damage Responses, Lawrence Berkeley National Laboratory, Berkeley, CA, United States.
[5]McArdle Laboratory, Department of Oncology, University of Wisconsin, Madison, WI, United States.
[6]Department of Veterinary and Animal Science, University of Massachusetts Amherst, Amherst, MA, United States.
[7]Pioneer Valley Life Sciences Institute, Springfield, MA, United States.

*These authors contributed equally.

Correspondence to:
Melissa A. Troester, Ph.D.
Department of Epidemiology, CB 7435
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
(919) 966-7408
troester@unc.edu

Running Title: Age-associated gene expression in normal breast and breast cancer

Keywords: aging, breast, cancer, gene expression, reduction mammoplasty

Abstract word count: 246

Text Word Count: 4237

Total number of figures and tables: 4

Conflict of interest statement: None declared

Category: Research Article – Biomarkers, -Omics & Systems Biology

ABSTRACT

Age is the strongest breast cancer risk factor, with overall breast cancer risk increasing steadily beginning at approximately 30 years of age. However, while risk of breast cancer is lower among younger women, young women's breast cancer tends to be more aggressive. Though several genomic and epidemiologic studies have shown higher prevalence of aggressive, estrogen-receptor negative breast cancer in younger women, the age-related gene expression that may predispose to these tumors is poorly understood. Characterizing age-related patterns of gene expression in normal breast tissues may provide insights on etiology of distinct breast cancer subtypes because it is from these tissues that the tumors arise. To identify age-related changes in normal breast tissue, 96 tissue specimens from reduction mammoplasty patients aged 14 to 70 were assayed by gene expression microarray. Significant associations between gene expression levels and age were identified for 802 probes (481 increased, 321 decreased with increasing age). Enriched functions included 'aging of cells', 'shape change', and 'chemotaxis', and enriched pathways included Wnt/beta-catenin signaling, Ephrin Receptor Signaling, and JAK/Stat Signaling. Applying the age-associated genes to publicly available tumor datasets, the age-associated pathways defined two groups of tumors with distinct survival. The hazard rates of young-like tumors mirrored that of high grade tumors in the Surveillance, Epidemiology and End Results Program, providing a biological link between normal aging and age-related tumor aggressiveness. These data show that studies of normal tissue gene expression can yield important insights about the pathways and biological pressures that are relevant during tumor etiology and progression.

INTRODUCTION

Age is the strongest demographic risk factor for human cancer overall (1, 2), with breast cancer rates steadily increasing with age. However, while tumors are less common in young women, younger women are more likely to have aggressive tumors. Young women's breast cancer is more often estrogen receptor negative, while estrogen receptor positive cancers are more prevalent in postmenopausal, older women (3, 4). Two previous gene expression studies have compared molecular features of breast cancer in younger and older women, focusing on the tumor gene expression (5, 6). These tumor gene expression studies have shown that tumors in younger women are associated with unique gene expression features, but many of these differences ultimately reflect the molecular subtypes of breast cancer. Because different types of breast cancers are more common in each age group, analyses across all tumors appeared to have been confounded by tumor subtype; after adjusting for grade and subtype, age-associated gene expression changes were no longer evident (7).

Persistent gaps in our understanding of age-associated changes in tumor aggressiveness remain. Research is needed to distinguish between characteristics of the malignancy and characteristics of the host (5, 8). As stated by Benz et al., "whether aging produces background effects from which the malignancy must be differentiated or contributes to the carcinogenic process is of fundamental importance" (9). Also, how the host-tumor interface changes with aging is poorly understood (7). To address these gaps, studies on normal breast tissue are needed. The aging of normal breast tissue has not been well studied, beyond changes such as postmenopausal involution (10). Aging-associated changes occur throughout lifespan, and not just at the onset of menopause, with potentially important consequences for the microenvironment of a nascent tumor. Indeed, there is widespread recognition that normal cellular environment, or tissue microenvironment, plays a role in tumor development and progression (11-13), but how aging may play a role in modifying this microenvironment is not known.

A conceptual model of cancer as an evolutionary process may be helpful in framing questions about how aging affects tumor etiology and progression. Evolutionary theories of cancer use language and principles from ecology, arguing that there are barriers to carcinogenesis that must be overcome by populations of tumor cells (14). Barriers to carcinogenesis derive from host biology and from normal tissue conditions that are present in the environment of the cancer cell prior to, or early in, disease (15). Because tumors evolve with selective pressure from their surrounding stroma, studies of normal breast tissue, which is typically more than 90% stroma by volume, may provide insights regarding the selective pressures faced by tumors during development and progression. In the current study, we hypothesized that age is associated with changes that may raise or lower barriers for tumor cell survival and evolution. Older vs. younger tissues represent distinct evolutionary environments, resulting in distinct cancer subtypes at different ages. According to this hypothesis, the signaling patterns present in the normal tissue of young women would be reflected in the signatures of aggressive tumors that are more common in young women.

To evaluate our hypothesis, we characterized gene expression in normal tissue using whole genome microarrays on 96 normal tissue specimens from reduction mammoplasty patients. We identified patterns and signaling changes that are associated with age in *normal tissue* of premenopausal women. We used whole tissue to allow for sufficiently large sample sizes to characterize the degree of inter-individual variation in aging, and we then tested whether our signature predicted age in isolated epithelium-enriched glands of normal tissue from women of different ages. We then evaluated the age-associated signature using publicly available *tumor tissue* gene expression data, asking whether age-associated gene expression from normal tissue can define distinct tumor groups with statistically significant differences in relapse-free and overall survival. The results of this investigation support the hypothesis that gene expression in young women's breast tissue creates an environment conducive to more aggressive tumor phenotypes.

METHODS

Patient characteristics. This study included women age 14 – 70 who were free of pathological diagnoses of the breast and who underwent reduction mammoplasty surgery at Baystate Medical Center in Springfield, Massachusetts between 2007 and 2009. Patient age and menopausal status are presented in Table 1. Institutional Review Boards (IRBs) at Baystate Medical Center and University of Massachusetts Amherst approved the study. Women consented to provide excess tissues not needed for diagnostic purposes and age and other demographic variables were measured by a telephone interview administered following surgery. Tissues from all patients were snap frozen and stored at -80° C prior to RNA isolation. To test age-associated signatures derived from microarrays on whole tissue, an independent data set of histologically normal breast tissues, obtained from surgically discarded reduction mammoplasty specimens, was provided by the UCSF Cancer Center and the Cooperative Human Tissue Network. These patients provided informed consent under an Institutional Review Board approved protocol. These samples were processed to isolate glands as described below.

RNA isolation and microarrays. Frozen specimens were cut over dry ice (approximately 100 mg whole tissue specimens) and RNA was isolated using standard manufacturer protocols for RNeasy midi kits or these same kits with a Qiazol extraction step. Higher yields were obtained with fatty, reduction mammoplasty tissues when using the Qiazol extraction. To test reproducibility of different extraction methods, a subset of samples were used to isolate RNA under both protocols. Cross-method intraclass correlation coefficients were similar to those obtained from replicate samplings by a single method. Agilent whole genome 4X44K catalog microarrays were used for all samples, with replicate samples assayed on the same platform or on custom 244k arrays. Only 4X44K probes were utilized in these analyses. All arrays were performed according to manufacturer protocols for linear amplification and two-color hybridization.

Reduction Mammoplasty Microarray Data Analysis. Spots that had an intensity of greater than 10 units in at least 80% of samples were selected for subsequent analyses. Data were lowess normalized and missing data were imputed using k-nearest neighbors with k=10. Duplicate arrays (N=15) were averaged, with those arrays with low correlation (< 0.7) removed from analysis (1 duplicate patient + 2 single of triplicate patients). A total of 114 microarrays representing 99 patients were included in the current analyses. All statistical analyses were performed in R, using Bioconductor packages. For each of the two analyses (age-associated gene expression and menopause-associated gene expression) using the full dataset (n=76 and n=99, respectively), the following prefiltering steps were applied: First, probes with no corresponding Entrez gene ID were eliminated. Second, probes with low variability, those with variability less than median variability across all probes, were eliminated. Linear regression was performed using LIMMA (16) to identify the maximum number of significant probes associated with chronological age (in decades) or menopausal status (pre- and perimenopausal vs. postmenopausal). Unadjusted p-values from LIMMA were used in conjunction with the qvalue package from Bioconductor to estimate q-values. A q-value < 0.10 (corresponding to a false discovery rate of 10%) was selected as evidence of statistical significance. Hierarchical clustering was used to visualize the data for age-associated genes, with samples ordered according to chronological age, and genes clustering according to Pearson correlation. Gene ontology analyses were performed using Ingenuity Pathway Analysis (IPA, Redwood City, CA).

We additionally tested a second dataset of isolated glands from reduction mammoplasty patients (N=30) against our signature. Tissue samples were minced and enzymatically dissociated using 0.1% w/v collagenase I in Dulbecco's Modified Eagle Medium at 37 °C for 12 to 18 h. Small tissue fragments (organoids), remaining after digestion, were collected by centrifugation at 100 g for two minutes. These organoids were stored frozen prior to RNA extraction. RNA was extracted using Qiagen RNeasy and

Affymetrix GeneChip Human Gene 1.0 ST microarrays were performed at University of Wisconsin, Madison. Microarray data was processed using Robust Multiarray Average in Bioconductor.

Confirmation of Age-Associated Expression Changes by Quantitative Reverse Transcription PCR. To confirm expression changes for CDKN2A and TP53 as identified in microarray data, 1 ug RNA per sample for 26 samples with remaining RNA was treated with genomic DNA Wipeout Buffer and reverse transcribed using the QuantiTect® Reverse Transcription kit (Qiagen) according to the manufacturer's protocol. Resultant cDNAs (10 ng) were plated, in triplicate, into a 96-well plate. Mastermix from the miScript SYBR Green PCR Kit (Qiagen) and miScript-derived primers for p16 (CDK2NA), p53 or GAPDH, were added to each well and real-time, quantitative PCR was performed on an Applied Biosystems 7900HT thermocycler. The q-RT-PCR program used was 15 minutes at 95°C, 40 cycles of 15 seconds each at 94°C, 30 seconds at 55°C, and 30 seconds at 72°C, with a denaturation cycle at the end. One statistical outlier was detected in regression diagnostics, and was removed from each of the p16 and p53 datasets to yield final datasets with 25 patients.

Analysis of Public Microarray Datasets and Test Sets of Isolated Glands. Our objective was to assess whether age-associated gene expression segregated tumors into clinically meaningful groups, based on the previous evidence that increasing age is associated with qualitative shifts in tumor subtype (5, 17). We hypothesized that gene expression patterns from normal aging would manifest themselves across different groups of tumors, such that more aggressive tumors would be more similar to younger, normal tissue. To test this hypothesis we projected the age-associated gene set onto two publicly available microarray datasets. These datasets were the NKI295 (18), Naderi et al. (19), and the UNC337 (20). To classify tumors as young-like or older-like, we applied methods described in Creighton et al. (21) to obtain a correlation coefficient describing the relation between each sample and the 802-probe age associated signature, we collapsed 802 probes by statistical mean to a list of unique entrez ids (N=719 genes). The vector summarizing the age signature on these genes was constructed by computing the difference in median gene expression in reduction mammoplasty

patients older than 39 versus those younger than 30 (median expression in younger minus median expression in older). Genes with differences less than zero (lower expression in young) were set to -1 and genes with differences larger than zero (higher expression in young) were set to 1. This vector of expected values was compared with the sample gene expression data and Pearson correlation coefficients were calculated. If Creighton correlations for a given tumor sample was positive, the patients were classified as young-like; if the correlation was negative, the patient was classified as older-like. Since all of the patients in the reduction mammoplasty data set and the NKI dataset were 55 or younger, all three tumor datasets were restricted to this age range for the combined analysis. Data for age restricted datasets were median-centered by gene and filtered to include only those genes with top 50% variability prior to performing correlation analysis. The 719 genes mapped to 380 variable genes in the NKI dataset, 2) 317 variable genes in the Naderi dataset, and 268 variable genes in the UNC337 dataset. Following classification, all samples were aggregated to a single dataset including 459 tumors.

Because we hypothesized that individuals with young-like tumors would mirror patterns of more aggressive tumors, we compared young-like tumors to aggressive tumors in SEER data. Data for women 55 or younger were selected from the SEER registry, years 1973-2009, [ASCII file: yr1973_2009.seer9], Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973-2009), National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2012, based on the November 2011 submission. To evaluate whether patterns of age at incidence and hazard rate over time in young-like tumors paralleled aggressive tumors in SEER, hazard curves were generated using the muhaz library in R. Young-like tumor behavior was compared with 'aggressive' (grade 3) tumors and 'less aggressive' SEER tumors (grade 1 or 2). Previous manuscripts (3) have demonstrated that for the SEER data as a whole, several clinical characteristics (ER status, PR status, race, grade, tumor size) duplicate the same general patterns for age at incidence and hazard rate, so grade was selected as representative.

Creighton correlation-based classification methods were also used for analyzing age-associated gene expression in an independent test set of isolated glands. The Entrez IDs (144 genes) from the fold-change age signature were mapped to the University of Wisconsin data, with 138 probes identified in the isolated gland Affymetrix dataset. Creighton correlations were computed for each sample relative to the 144-gene age signature. The association between the Creighton correlation (coded as 'positive' if >= 0 or 'negative' if < 0) and the true age of tissue (< 30, 30-39, >39) was estimated using the nonzero correlation statistic, a 1 degree of freedom, Cochran-Mantel-Haenszel statistic, obtained from PROC FREQ in SAS 9.2.

Comparison of Age-Associated Signature with Previously Published Signatures. Previously, a signature of aging was published based on tumor gene expression data (5). In addition, a meta-analysis of common signatures of aging across many tissues was published (22). These two signatures, representing important previous work on aging in breast and across tissue types, were evaluated in our RM samples to test whether they also predicted age in these normal human breast tissues and to assess correlation with our aging signature. These published signatures were first mapped to our filtered dataset of reduction mammoplasty patients, resulting in 85 genes for the 145-gene Yau et al. (5) signature and 52 genes for the 73-gene de Magalhaes signature (22). A Creighton correlation was computed where genes with high expression in young were coded as 1 and high expression in older were coded as -1. If correlation was positive, these patients were young-like for a given signature. These signature based classes were then evaluated for association with patient age and for association with our RM-based young-like and older-like signature. Chi-square tests, or Fisher exact where cell counts were <5, were used to evaluate the statistical significance of these associations.

RESULTS

Age-associated gene expression in reduction mammoplasty patients. A substantial proportion of genes examined by microarray had expression changes associated with age. A False Discovery Rate (FDR) of 5

or 10% is commonly used in supervised analysis of microarray data, and at that level 2 or 4%, respectively, of genes are significantly associated with age. A striking 15% of genes were associated with age at FDR of 20%, demonstrating the broad change induced by aging. Figure 1A shows a one-dimensional (genes clustered only) heat map of gene expression for 802 age-associated genes (with FDR<10%) across 76 samples (representing 62 patients), ordered according to chronological age, with colored bars representing the age group by decade. The figure illustrates that there appears to be a qualitative shift in gene expression in the late thirties, with substantial inter-individual variation across individuals in any given age group.

To test this gene set in an independent dataset, we used gene expression data from isolated glands (enriched for mammary epithelia as described in Methods) of reduction mammoplasty patients. Because whole tissue was used in training, evaluation of this second epithelium-enriched dataset helped to evaluate whether age-related signatures were reflective of changes within epithelium or changes in tissue composition overall. Results in this independent test set shows that even in microdissected epithelium, the age-related changes observed in whole tissue are largely preserved. Figure 1B shows that the age-associated signature also is significantly associated with age in this dataset. There is a strong correlation between age and the expression of the aging signature, with young women (shown in green) tending to show positive correlation with the younger women's signature (9 of 14 samples positively correlated with the young signature) and the older women (shown in magenta) showing negative correlation (12 of 16 negatively correlated with young signature). Across all of the independent test samples, there was a strong trend toward decreasing correlation with the young age signature as age increases (odds ratio for young expression is 5.4 [95% CI: 1.1-26.0] comparing younger to older women). The association with age in this independent test set is particularly striking given the distinct specimen processing methods, microarray platform, and collection and processing at separate institutions. These data suggest that the signature we detected does not simply mirror changing tissue composition with age, but rather, it reflects age-related change within cellular compartments.

To evaluate the pathways that were differentially expressed with increasing age, we used Ingenuity Pathway Analysis as described in Methods. Among the 802 differentially expressed probes, several processes and pathways were associated with age (Table 2). Aging of cells, cell flattening, and shape change were significant processes, while JAK2-associatied hormone-like cytokine signaling, Wnt/β-catenin signaling, and Ephrin Receptor Signaling were significant pathways. The complete list of significant genes depicted in Figure 1, are shown in Supplemental Table 1 along with the average fold change comparing the median centered data for the youngest patients (< 30, n=20) to that of the oldest patients (>49, n=6).

Two genes were of particular interest: CDKN2A (p16) and TP53. We confirmed the direction of change for both of these genes by performing Q-RT-PCR on a subset of samples with remaining RNA (Supplemental Figure 1). The p16INK4a tumor suppressor has an established role in aging and its expression is known to increase with increasing age (23, 24). The links between p53 and aging are complex. While p53 plays a critical role in determining cellular senescence and in vitro lifespan (25), levels of activity decline with aging in rodents (26). The basis for loss of p53 function with aging could be attributed to the progressive impairment of mitochondrial OxPhos which was shown to abrogate p53 (27). And while aging causes decreased p53 activity, hyperactive p53 is associated with accelerated aging phenotypes (28). Our microarrays showed p53 levels decreasing with increasing age, which was also qualitatively confirmed by QPCR.

Menopause-associated gene expression in reduction mammoplasty patients. In contrast to the large number of genes found to be associated with age, there were relatively few that were associated with menopausal status. In a comparison between 76 pre/peri-menopausal women and 23 post-menopausal women, only 273 genes were statistically significant given an unadjusted p-value < .05. After adjustment for multiple testing (as was done for age-related genes), no genes were significantly associated with menopausal status (q value > 0.10 for all genes). Despite the weak association between menopausal status and gene expression, we performed an IPA analysis with the 273 genes that were significant given an unadjusted p-value less than 0.05.

We identified no Functional Annotation or Canonical Pathway categories that were differentially expressed with Benjamini-Hochberg $p<0.05$, further suggesting that these genes were not biologically significant. These results show that for the age range we studied (20-70), menopausal status did not strongly dichotomize the biological characteristics of breast tissue. However, our dataset contained a relatively small number of postmenopausal women (n=23) and a larger study with more balanced representation of pre- and post-menopausal women may be better powered to detect differences by menopausal status.

Age-associated gene expression in the breast cancer patients. According to evolutionary theories of cancer (12), tumors take advantage of the transcriptional programs and pathways that are active in normal tissue, leveraging existing programs to advance growth and survival. Thus we expected that pathways that showed age-associated expression in normal tissue would also be dysregulated in tumors. By applying the age-associated gene set from Figure 1 to three publicly available breast tumor microarray datasets, we identified two groups of patients. Similar to SEER data for 'aggressive' high grade vs. 'less aggressive' low grade tumors (Figure 2A), patients with young-like gene expression showed an left shift in the incidence distribution, documenting an earlier age at incidence pattern (Figure 2B). Aggressive tumors and those that had young-like expression also had peak hazard ratios early following diagnosis, followed by a declining hazard rate after peaking; this pattern was not observed for 'less aggressive' SEER tumors or older-like tumors, both of which had increasing hazard rates over time (Figure 2C and 2D). In sum, the patterns of age at incidence and hazard rate over time for young-like breast tumors are very similar to patterns presented for aggressive breast cancers based on SEER data (3).

We also evaluated whether young like vs. older-like were correlated with particular tumor characteristics (Table 3). In fact, young-like tumors were more likely to have clinically aggressive characteristics, with statistically significant associations in the largest of the three datasets (NKI dataset): ER negative (p=0.02), high grade (p=0.005), larger (p=0.04). Substantial, though non-significant, associations in the same direction (more

aggressive tumors given young-like gene expression class) were observed for Naderi et al. and UNC 337 datasets. Considering the combined dataset, significant associations held with numerous clinically aggressive phenotypes (large size, high grade, and young age). The strongest association was for high tumor grade, which was moderately associated with being young-like in all three datasets, including UNC (p=0.17) and Naderi (p=0.09) data sets, and in the combined analysis (p=0.01). Young-like tumors were also more prevalent among young women in all datasets except for the Naderi et al., where the young-like signature did not correlate with patient age. However, this dataset (n=52) had an older patient population (mean age of 47 compared with UNC and NKI which both had mean age of 44). These results document that the normal biology of younger women is reflected in more aggressive tumors that are more common in this age group.

Evaluation of correlations with previous age-related signatures. A previous study had evaluated common signatures of aging, across 27 different studies including mouse, rats and humans (22). While mammary gland/breast was not specifically studied in these datasets, common signatures were identified for tissues such as heart, lung, brain, muscle, kidney and liver. We evaluated whether this aging signature was correlated with normal aging human breast signatures in our RM study. As shown in Table 4, young-like samples based on the de Magalhaes signature, were both younger chronologically and more likely to be young-like according to our signature. While previous studies of aging human breast were not available for comparison, we evaluated one tumor-based signature that carefully evaluated age-related gene expression among breast tumors (5). This tumor-based age signature was selected because it restricted to only ER positive tumors, potentially avoiding some of the problems of confounding by tumor subtype as described in Anders et al. (7). The Yau et al. signature also was significantly associated with our young-like signature, but was not associated with patient age when projected onto normal tissues. The weaker correlation with age for this signature may reflect the fact that the tumor biology evolves and diverges from the normal age-related biology as the tumor progresses.

CONCLUSIONS

Molecular and cellular studies of breast tissue clearly show that there are important compositional and morphometric changes in aging breast tissue (29). During the premenopausal period, a decline in ovarian function causes regressive changes in both epithelial structures and in stroma, with declining epithelial tissue and lobular volume from the third to sixth decades of life, independent of previous reproductive history (30). The observation that aging-associated change is a process spanning decades rather than a simple dichotomy is also reflected by our observation that self-report of menopausal symptoms was not strongly associated with changes in breast tissue. Pre- versus post-menopausal status is categorized on the basis of menopausal symptoms in most studies and not on the basis of tissue level change. Future research should consider misclassification error induced by using menopausal status as a surrogate for tissue level change or should consider alternative measures of ovarian function, rather than using reported menopausal symptoms as the basis for dichotomizing complex tissue-level biological processes.

Age-associated changes in gene expression in histologically normal breast tissue have not been reported previously. This is a striking gap in the literature given that aging-associated gene expression profiles have been reported previously for human fibroblasts and lymphocytes (31, 32) as well as brain (33), kidney (34), and skeletal muscle (35, 36). A recent meta-analysis has compared aging-related changes across species and tissues, but without inclusion of mammary gland (22). In our data, age-associated gene expression was functionally linked with previous 'aging' gene expression categories (as shown in our ontological analyses), but also included individual genes of particular interest. CDKN2A (p16) is recognized as a biomarker and effector of mammalian aging (24), and its upregulation is accompanied by changes in telomere length. As expected from the previous association studies in other tissues and mechanistic studies conducted in vitro (Bazarov et al., 2010), we found increasing expression of p16 with age and decreasing expression of TERT with age. Transcripts of the gene coding tumor suppressor p53 (TP53) also changed with age, with p53 expression

declining in older patients. Links between p53 and aging have focused largely on studies in mouse models of breast cancer (26), but it has been demonstrated in human studies that p53 mutations are more common in younger women and the basal-like tumors that occur more frequently in the young (37). These observations raise the hypothesis that increased mutation frequency in young women may reflect both greater activity of p53 in young women and a resulting pressure to inactivate p53 in young women. Further, as reviewed in Reinhardt and Schumacher, p53 and its target genes function as important regulators of cancer prevention and aging (38).

Other interesting developmentally-regulated pathways were also altered with age in adult breast tissue. The JAK2-associated hormone-like cytokine signaling, Wnt/β-catenin signaling, and Ephrin Receptor signaling were differentially expressed in older versus younger women. The hormone-dependent JAK2 signaling alterations (including higher expression of STAT5A), may reflect changes in ovarian function/estrogen signaling over time; this pathway regulating mammary gland development is responsive to estrogen and progesterone in mouse models (39). The latter two pathways (Wnt/β-catenin and Ephrin receptor signaling) are known to be involved in maintaining stem cell dynamics in cancer (40, 41), but their specific roles in histologically normal human breast tissue are relatively unexplored. Given that mammary progenitor cells are a rare cell population, these signals are unlikely to be derived specifically from stem cell populations, but may reflect the role of both pathways (and cross talk between them) in tissue architecture or cellular differentiation [reviewed in (42) and (43), for Ephrins and Wnt pathway, respectively]. While the details of mechanism are not enlightened by our results, we hypothesize that the changes we observe reflect alterations in survival and proliferation potential of the normal cell types that are susceptible to carcinogenesis. Alteration of these signals and normal tissue homeostasis with age may dictate pathways to malignancy and determine the aggressiveness of resulting tumors. A recent interesting commentary has emphasized the importance of altered homeostasis in age-dependent cancer rates, countering the previous notion that oncogenic mutation rates alone limit carcinogenesis at young ages (44).

Our study also identified trends that are relevant for the epidemiology of breast cancer and aging. Previous analyses of epidemiologic data have used breast cancer incidence data to draw inferences about the rate of aging of breast tissue (45, 46). These papers have suggested that the rate of aging is most rapid in the early years after menarche and before the first pregnancy, decreases with each subsequent birth, and decreases further with menopause. Use of very large datasets and anchoring of changes to particular reproductive events allowed for resolution of these trends in spite of substantial inter-individual variability. While aging-related changes in undiseased tissue are a more direct route to studying aging in tissue, these studies are currently underpowered to try to dissect the composite and interactive effects of multiple demographic and exposure variables as correlates of age-related phenotypes. However, by examining the gene expression in Figure 1, it is clear that there is substantial heterogeneity in the population. Some young patients' tissues show gene expression patterns more similar to older women. Premature expression of signatures reflective of older biology might predict earlier increased risk of breast cancer. In the future, it may be possible to evaluate age-related signatures as biomarkers of early breast cancer risk, particularly given that a larger number of epidemiologic studies are now collecting histologically normal tissue from both diseased and unaffected individuals. Recent studies with peripheral blood T-lymphocytes have demonstrated that molecular markers of aging do show associations with health behaviors such as physical activity and smoking status (24). Perhaps even more importantly, though, the biomarkers of aging can give us greater understanding of mechanisms of etiology.

If a group of tumors has distinct mechanisms of etiology, then the progression of these tumors may depend upon the degree to which mechanistically relevant pathways are expressed in the normal tissue from which these tumors arise. In other words, if gene expression microenvironments in younger tissue apply selective pressures or create optimal conditions for specific subtypes, then (1) these subtypes would be more common in younger women, and (2) the tumors would be expected to differentially express the pathways

common to young breast tissue. Regarding the first point, several papers have documented that different subtypes of tumors are more prevalent in younger women (e.g. ER-negative and basal-like cancer), which is echoed in our data showing that tumors with young-like signatures were more likely to have aggressive clinical features. Regarding the second point, our work illustrates that the young-like tumors have distinct age at incidence patterns and hazard rates over time, similar to the incidence and hazard rate patterns produced by stratifying on aggressive clinical characteristics. Thus our work provides a strong biological link between aging processes and the etiology of aggressive breast cancer subtypes.

While links between age-associated gene expression and epidemiologic age at incidence patterns are informative, there are several caveats to our analysis. First, we used public microarray datasets to evaluate our age-associated signature classes in comparison with high grade/aggressive tumors in SEER data. While the data convincingly recapitulate the SEER patterns for age-at-incidence and hazard rate over time, it must be noted that these public datasets are not population-based samples and therefore may have substantial distributional differences from SEER in both age and tumor characteristics. While we restricted the age-range of tumors in our analysis to improve the comparability across datasets, the lack of a population-based tumor gene expression data for evaluating age-dependent signatures limits our comparability with SEER data. Second, all of the microarray datasets used were modestly sized. Therefore, it was impossible to stratify on relevant demographic variables such as race, and we were unable to detect weaker changes in gene expression with age. However, these analyses are an important first step toward characterizing some of the strongest changes induced in aging breast tissue. Third and finally, our reduction mammoplasty tissues were not microdissected prior to analysis and therefore it is likely that we identified only changes that are common to both stromal and epithelial cell types. Given that our signature derived from stroma-rich whole tissue predicted chronological age in isolated epithelium, we expect these changes are shared between tissues. Indeed aging changes may be highly stereotyped across tissue and cell types and highly conserved across organisms, given that our signature

correlated with a signature derived from multiple species and multiple tissues, both stroma and epithelium rich (22).

Continued and future research may consider whether other breast cancer risk factors perturb particular breast-cancer related pathways in normal tissue. For example, if particular pathways are altered in normal breast tissue according to body mass index (47) or parity (48), epidemiologic studies could assay these pathways in tumors and stratify cases according to whether they express these pathways. This would help to delineate the distinct causal paths that contribute to heterogeneous phenotypes of breast cancer. Case-only studies have been used to identify etiologic heterogeneity with respect to particular pathways (49), and more recently, concordance of phenotypes between first and second primary cancers has been used to establish etiologic distinctiveness (50). Our work demonstrates that evaluating pathways in both normal tissue and in tumors can help advance our understanding of etiologic distinctiveness.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of Maureen Lahti at Baystate Medical Center for her contributions to patient data collection.

GRANT SUPPORT

This project was supported by grants from the National Cancer Institute and National Institutes of Environmental Health Sciences (U01-ES019472, R01-CA138255, R01-ES017400, U01-ES019548, U01-ES019466), a Breast SPORE (P50CA058223), a grant from the Avon Foundation, and the University Cancer Research Fund at the University of North Carolina.

REFERENCES

1. Edwards BK, Howe HL, Ries LA, Thun MJ, Rosenberg HM, Yancik R, et al. Annual report to the nation on the status of cancer, 1973-1999, featuring implications of age and aging on U.S. cancer burden. Cancer. 2002;94:2766-92.
2. Benz CC, Yau C. Ageing, oxidative stress and cancer: paradigms in parallax. Nat Rev Cancer. 2008;8:875-9.
3. Anderson WF, Jatoi I, Devesa SS. Distinct breast cancer incidence and prognostic patterns in the NCI's SEER program: suggesting a possible link between etiology and outcome. Breast Cancer Res Treat. 2005;90:127-37.
4. Anderson WF, Pfeiffer RM, Dores GM, Sherman ME. Comparison of age distribution patterns for different histopathologic types of breast carcinoma. Cancer Epidemiol Biomarkers Prev. 2006;15:1899-905.
5. Yau C, Fedele V, Roydasgupta R, Fridlyand J, Hubbard A, Gray JW, et al. Aging impacts transcriptomes but not genomes of hormone-dependent breast cancers. Breast Cancer Res. 2007;9:R59.
6. Anders CK, Hsu DS, Broadwater G, Acharya CR, Foekens JA, Zhang Y, et al. Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. J Clin Oncol. 2008;26:3324-30.
7. Anders CK, Fan C, Parker JS, Carey LA, Blackwell KL, Klauber-DeMore N, et al. Breast carcinomas arising at a young age: unique biology or a surrogate for aggressive intrinsic subtypes? J Clin Oncol. 2011;29:e18-20.
8. Cohen HJ. The cancer aging interface: a research agenda. J Clin Oncol. 2007;25:1945-8.
9. Benz CC. Impact of aging on the biology of breast cancer. Crit Rev Oncol Hematol. 2008;66:65-74.
10. Milanese TR, Hartmann LC, Sellers TA, Frost MH, Vierkant RA, Maloney SD, et al. Age-related lobular involution and risk of breast cancer. J Natl Cancer Inst. 2006;98:1600-7.
11. Egeblad M, Nakasone ES, Werb Z. Tumors as organs: complex tissues that interface with the entire organism. Dev Cell. 2010;18:884-901.
12. Merlo LM, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. Nat Rev Cancer. 2006;6:924-35.
13. Allinen M, Beroukhim R, Cai L, Brennan C, Lahti-Domenici J, Huang H, et al. Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell. 2004;6:17-32.
14. Gatenby RA, Gillies RJ, Brown JS. Evolutionary dynamics of cancer prevention. Nat Rev Cancer. 2010;10:526-7.
15. Bissell MJ, Hines WC. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. Nat Med. 2011;17:320-9.
16. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology. 2004;3:Article3.
17. Anders C, Carey LA. Understanding and treating triple-negative breast cancer. Oncology (Williston Park). 2008;22:1233-9; discussion 9-40, 43.
18. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002;347:1999-2009.
19. Naderi A, Teschendorff AE, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG, et al. A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene. 2007;26:1507-16.
20. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12:R68.

21. Creighton CJ, Casa A, Lazard Z, Huang S, Tsimelzon A, Hilsenbeck SG, et al. Insulin-like growth factor-I activates gene transcription programs strongly associated with poor breast cancer prognosis. J Clin Oncol. 2008;26:4078-85.
22. de Magalhaes JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics. 2009;25:875-81.
23. Janzen V, Forkert R, Fleming HE, Saito Y, Waring MT, Dombkowski DM, et al. Stem-cell ageing modified by the cyclin-dependent kinase inhibitor p16INK4a. Nature. 2006;443:421-6.
24. Tsygankov D, Liu Y, Sanoff HK, Sharpless NE, Elston TC. A quantitative model for age-dependent expression of the p16INK4a tumor suppressor. Proc Natl Acad Sci U S A. 2009;106:16562-7.
25. Bond JA, Wyllie FS, Wynford-Thomas D. Escape from senescence in human diploid fibroblasts induced directly by mutant p53. Oncogene. 1994;9:1885-9.
26. Feng Z, Hu W, Teresky AK, Hernando E, Cordon-Cardo C, Levine AJ. Declining p53 function in the aging process: a possible mechanism for the increased tumor incidence in older populations. Proc Natl Acad Sci U S A. 2007;104:16633-8.
27. Compton S, Kim C, Griner NB, Potluri P, Scheffler IE, Sen S, et al. Mitochondrial dysfunction impairs tumor suppressor p53 expression/function. J Biol Chem. 2011;286:20297-312.
28. Donehower LA. Does p53 affect organismal aging? J Cell Physiol. 2002;192:23-33.
29. Walker RA, Martin CV. The aged breast. J Pathol. 2007;211:232-40.
30. Hutson SW, Cowen PN, Bird CC. Morphometric studies of age related changes in normal human breast and their significance for evolution of mammary cancer. J Clin Pathol. 1985;38:281-7.
31. Geigl JB, Langer S, Barwisch S, Pfleghaar K, Lederer G, Speicher MR. Analysis of gene expression patterns and chromosomal changes associated with aging. Cancer Res. 2004;64:8550-7.
32. Tan Q, Zhao J, Li S, Christiansen L, Kruse TA, Christensen K. Differential and correlation analyses of microarray gene expression data in the CEPH Utah families. Genomics. 2008.
33. Lu T, Pan Y, Kao SY, Li C, Kohane I, Chan J, et al. Gene regulation and DNA damage in the ageing human brain. Nature. 2004;429:883-91.
34. Rodwell GE, Sonu R, Zahn JM, Lund J, Wilhelmy J, Wang L, et al. A transcriptional profile of aging in the human kidney. PLoS Biol. 2004;2:e427.
35. Welle S, Brooks AI, Delehanty JM, Needler N, Bhatt K, Shah B, et al. Skeletal muscle gene expression profiles in 20-29 year old and 65-71 year old women. Exp Gerontol. 2004;39:369-77.
36. Welle S, Brooks AI, Delehanty JM, Needler N, Thornton CA. Gene expression profile of aging in human muscle. Physiol Genomics. 2003;14:149-59.
37. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. Jama. 2006;295:2492-502.
38. Reinhardt HC, Schumacher B. The p53 network: cellular and systemic DNA damage responses in aging and cancer. Trends Genet. 2012.
39. Santos SJ, Haslam SZ, Conrad SE. Estrogen and progesterone are critical regulators of Stat5a expression in the mouse mammary gland. Endocrinology. 2008;149:329-38.
40. Incassati A, Chandramouli A, Eelkema R, Cowin P. Key signaling nodes in mammary gland development and cancer: beta-catenin. Breast Cancer Res. 2010;12:213.
41. Genander M, Frisen J. Ephrins and Eph receptors in stem cells and cancer. Curr Opin Cell Biol. 2010;22:611-6.
42. Micalizzi DS, Farabaugh SM, Ford HL. Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. J Mammary Gland Biol Neoplasia. 2010;15:117-34.
43. Andres AC, Ziemiecki A. Eph and ephrin signaling in mammary gland morphogenesis and cancer. J Mammary Gland Biol Neoplasia. 2003;8:475-85.

44. Degregori J. Challenging the axiom: does the occurrence of oncogenic mutations truly limit cancer development with age? Oncogene. 2012.
45. Pike MC, Krailo MD, Henderson BE, Casagrande JT, Hoel DG. 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. Nature. 1983;303:767-70.
46. Rosner B, Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. J Natl Cancer Inst. 1996;88:359-64.
47. Sun X, Casbas-Hernandez P, Bigelow C, Makowski L, Joseph Jerry D, Smith Schneider S, et al. Normal breast tissue of obese women is enriched for macrophage markers and macrophage-associated gene expression. Breast Cancer Res Treat. 2012;131:1003-12.
48. Belitskaya-Levy I, Zeleniuch-Jacquotte A, Russo J, Russo IH, Bordas P, Ahman J, et al. Characterization of a genomic signature of pregnancy identified in the breast. Cancer Prev Res (Phila). 2011;4:1457-64.
49. Begg CB, Zhang ZF. Statistical analysis of molecular epidemiology studies employing case-series. Cancer Epidemiol Biomarkers Prev. 1994;3:173-5.
50. Begg CB. A strategy for distinguishing optimal cancer subtypes. Int J Cancer. 2011;129:931-7.

TABLE 1: Demographic characteristics of women*

	Cases N	%
Menopausal Status†		
Premenopausal‡	62	81.6
Perimenopausal§	14	18.4
Missing	0	
Race & Hispanic Ethnicity		
White, non-Hispanic	48	63.2
Black, non-Hispanic	6	7.9
Hispanic	19	25.0
Other	3	3.9
Missing	0	
Age#		
20-29	20	26.3
30-39	33	43.4
40-49	17	22.4
>=50	6	7.9
Missing	0	

* Only included individuals >= 20 who are not menopausal
† 23 women who were postmenopausal were analyzed in pre/peri vs. postmenopausal analyses, but excluded from age analyses and from Table 1.
‡ Premenopausal if reports regular periods, or very young age on exogenous hormones, or very young age with hysterectomy, both ovaries preserved
§ Perimenopausal if last menstrual period < 1 year before interview
# excluded are those < 20 years of age.

TABLE 1a: Average Body Mass Index (BMI) by age category

	N	BMI
Age		
20-29	15	30.4
30-39	16	29.4
40-49	13	30.1
>= 50	2	30.5
Missing	30	

Table 2. Gene Ontology Categories Enriched Among Age-Associated Genes

Ingenuity Category	p-value[a]	Molecules
Functions Annotation		
Aging of cells	9.4e-5	NOX4, PDCD4, SOD1, TBX2, TP53
Cell flattening of cell lines	9.4e-5	CDKN2A, FIS1, PLD1, SMARCA4, SMARCB1
Shape change	1.2e-4	ANGPT1, AP1S2, ARFIP2, ARHGAP15, CASP10, CD28, CDC42EP4, CDKN2A, DLC1, EFNA1, EPHA2, EPHB4, FADD, FIS1, KITLG, MARK2, MYH9, NEDD4L, PACSIN2, PLD1, PLXNB1, RAP1B, RAP2C, ROCK2, RPS6KB1, SEMA3A, SLIT2, SMARCA4, SMARCB1, TEK, TNFSF11
Developmental process of melanocytes	3.2e-4	CTNNB1, EFNA1, KITLG, MITF, TP53
Nectortizing enterocolitis	3.2e-4	PDE11A, PDE3B, PDE4A, PDE4D, TLR4
Assembly of cellular protrusion	5.8e-4	ARFIP2, CCDC88A, RALA, RHOQ, SLC9A3R1, SLIT2
Retraction of cellular protrusion	6.6e-4	ARHGAP32, EFNA1, IL6R, ITGA3, MYH9, RHOQ, ROCK2
Canonical Pathways		
Role of JAK2 in Hormone-Like Cytokine Signaling	1.1e-3	SHC1, STAT5A, GHR, TYK2, SOCS2, JAK2
Wnt/β-catenin Signaling	2.9e-3	TP53, SOX4, CDKN2A, SOX12, MARK2, SOX17, CSNK1E, FZD8, FZD4, SOX8, DVL2, CTNNB1, ACVR1C, WNT5B, SOX5
Ephrin Receptor Signaling	5.2e-3	RAP1B, ACTR2, EPHB4, ANGPT1, GNB2L1, LIMK2, JAK2, ITGA3, GNG7, GNG10, EFNA1, ROCK2, SHC1, SDCBP, GRIN2C, EPHA2

[a]Benjamini-Hochberg, multiple testing adjusted p-value.

Table 3. Characteristics of young (≤55 years old) breast cancer patients according to young-like and old-like signature class, public microarray datasets from NKI , Naderi et al., and UNC.

	NKI			Naderi et al.			UNC 337			Combined		
	Young-like (N)	Old-like (N)	p-value	Young-like (N)	Old-like (N)	p-value	Young-like (N)	Old-like (N)	p-value	Young-like (N)	Old-like (N)	p-value
ER status												
Positive	106	120		17	15		29	27		150	164	
Negative	44	25	*0.02*	8	12	*0.52*	31	23	*0.69*	87	56	*0.07*
Missing							0	2		0	2	
Size												
< 2 cm	49	65		15	17		12	15		78	95	
>= 2 cm	101	80	*0.04*	10	10	*0.95*	48	36	*0.35*	159	126	*0.02*
Missing							0	1		0	1	
Tumor Grade												
Well differentiated/1	31	44		9	5		2	5		42	54	
Intermediate/2	45	56		5	13		20	10		70	79	
Poorly differentiated/3	74	45	*5.0 e-3*	11	9	*0.09*	37	30	*0.16*	122	84	*0.01*
Missing							1	7		1	7	
Subtype												
Basal	31	15		4	3		22	14		57	32	
ERBB2	20	29		4	7		12	7		36	43	
Luminal A	42	44		10	13		16	16		68	73	
Luminal B	44	37		4	3		9	11		57	51	
Normal	13	20	*0.05*	3	1	*0.66*	1	4	*0.34*	17	25	*0.05*
Age												
< 30	1	3		0	0		3	2		4	5	
30-39	35	24		2	2		13	10		50	36	
40-49	102	81		13	15		31	25		146	121	
50-55	12	37	*1.2 e-4*	10	10	*1*	13	15	*0.87*	35	62	*6.5e-3*

Table 4: Association between published age signatures based classification (columns), chronological age, and reduction mammoplasty age signature class.

	Older-like (N)	Young-like (N)	p-value
de Magalhaes*			
RM Older-like	28	10	
RM Young-like	13	25	*1.3 e-3*
20-29	4	16	
30-39	15	18	
40-49	12	5	
50+	2	4	*0.018*
Yau†			
RM Older-like	22	16	
RM Young-like	11	27	*0.02*
20-29	11	9	
30-39	14	19	
40-49	11	6	
50+	5	1	*0.21*

*de Magalhaes et al. (22) meta-analysis of aging signature includes 52 intersecting genes with RM signature.
† Yau et al (5) includes 85 intersecting genes with RM signature.

FIGURE LEGENDS

Figure 1. Age-associated gene expression in whole tissue (n=76) and in isolated glands (n=30). (A) 802 probes were collapsed to unique genes by averaging and a hierarchical cluster analysis (genes-only) shows two distinct gene expression groups. Some heterogeneity is observed within groups, but an overall trend is evident. (B) The gene expression signature was then used to predict age in an independent test set of isolated glands. Glands from younger patients were more likely to show positive correlation with the young signature from whole glands, while older patients were more likely to have negative correlations.

Figure 2. Age-at-incidence distribution and hazard rate over time are similar for aggressive tumors in Surveillance Epidemiology and End Results and young-like tumors in public tumor gene expression data. (A) Grade was used to stratify 'aggressive' (poorly differentiated, grade 3) tumors and 'less aggressive' tumors, with a left shift in the age distribution for 'aggressive tumors. (B) The young-like tumors mirror the left shift seen with aggressive tumors, providing a biological link between age and tumor aggressiveness. (C) Aggressive tumors have a unique hazard function in SEER data, with an early peak in hazard rate (2-5 years depending on tumor characteristic modeled) followed by a decreasing hazard rate, while less aggressive tumors (grade 1 or 2 in this example), have linearly increasing hazard rate with years following diagnosis. (D) Similarly, young-like tumors have a peak hazard early (prior to 5 years following diagnosis). Tumors with older-like gene expression show the characteristic linear increase in hazard rate over time. These hazard rate curves show that young-like tumors represent more aggressive breast cancer.

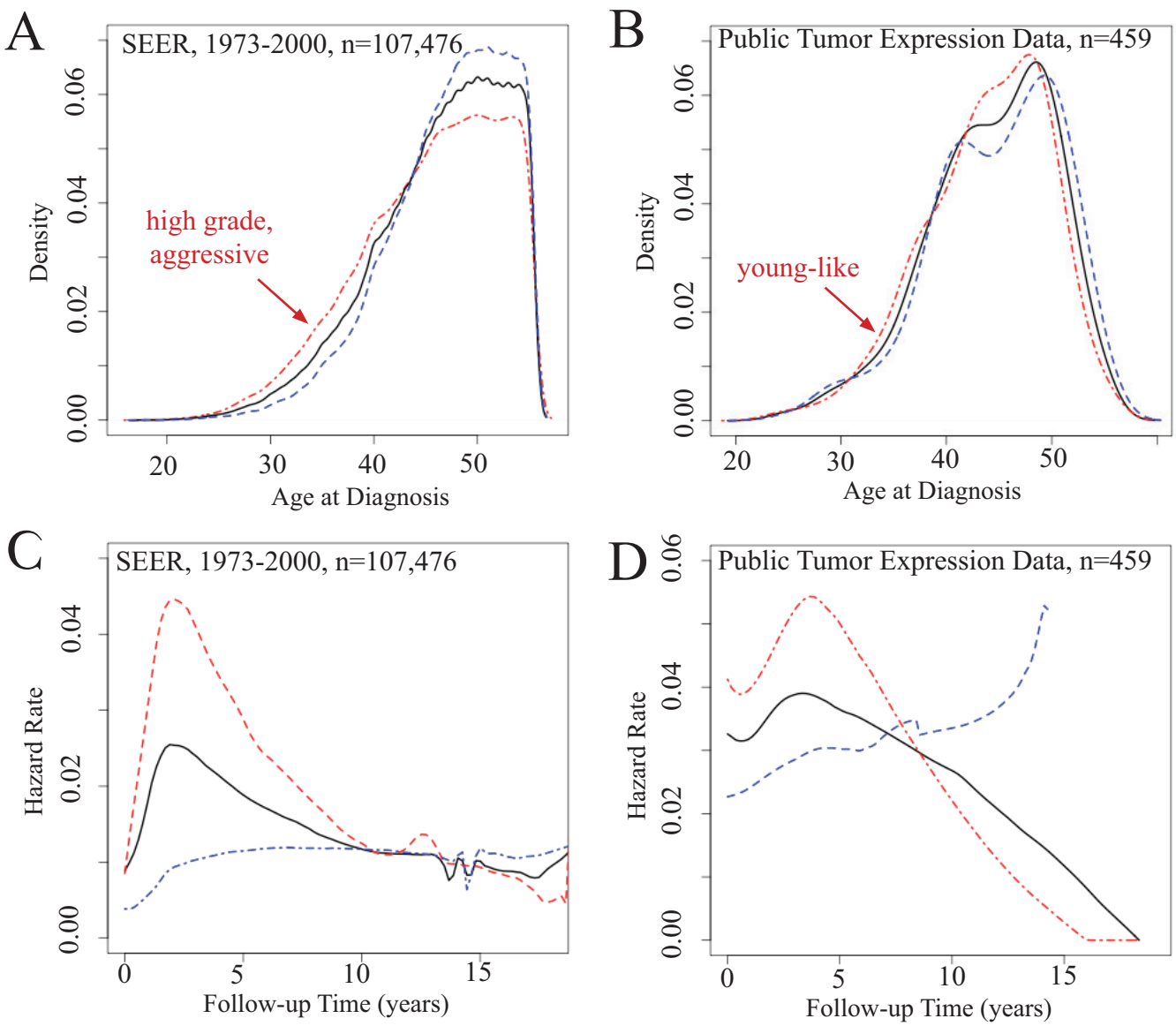
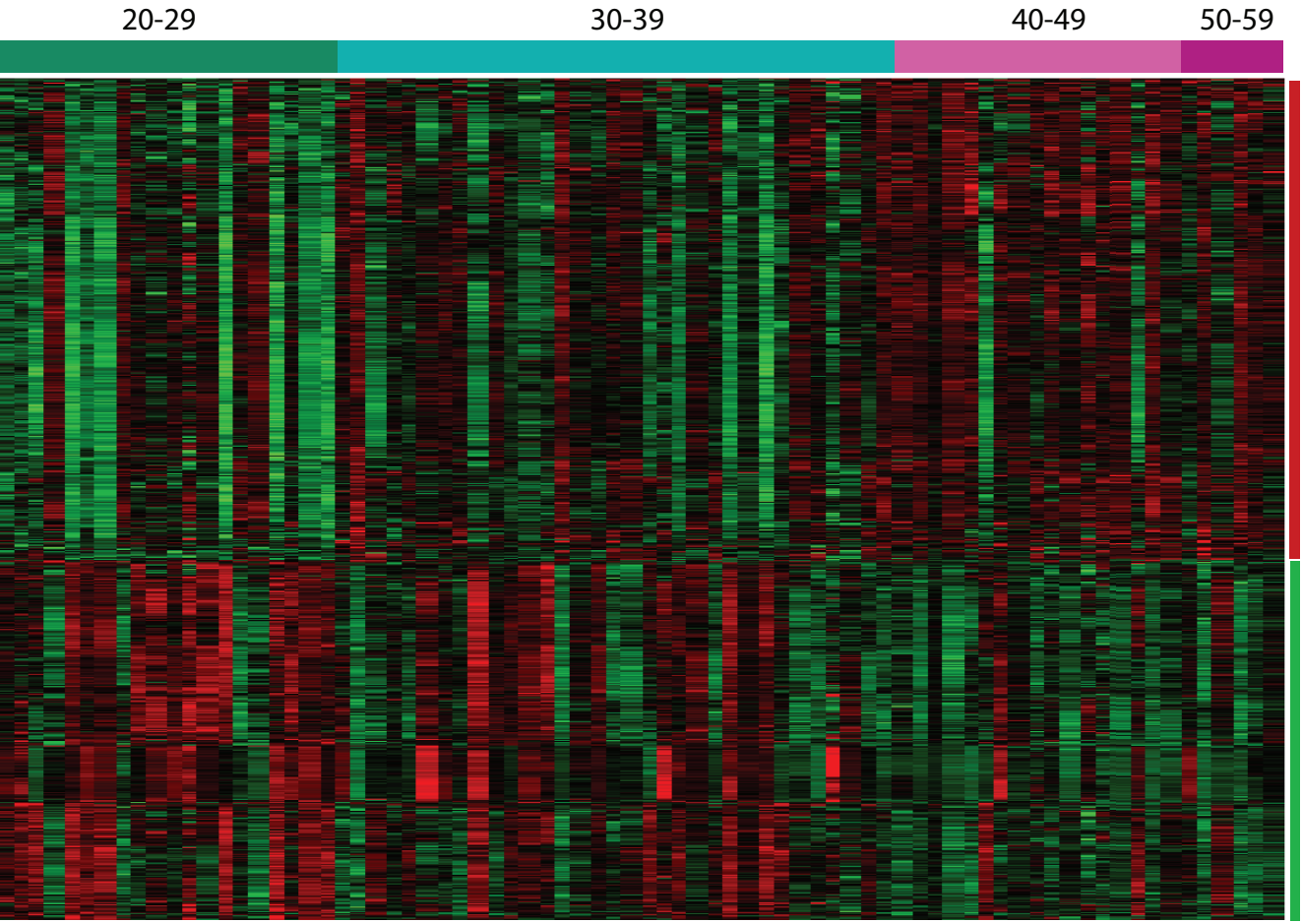
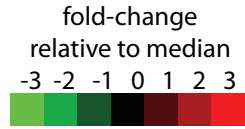

A
SEER, 1973-2000, n=107,476
high grade,
aggressive
Density
0.00
0.02
0.04
0.06
20
30
40
50
Age at Diagnosis
B
Public Tumor Expression Data, n=459
young-like
Density
0.00
0.02
0.04
0.06
20
30
40
50
Age at Diagnosis
C
SEER, 1973-2000, n=107,476
Hazard Rate
0.00
0.02
0.04
0
5
10
15
Follow-up Time (years)
D
Public Tumor Expression Data, n=459
Hazard Rate
0.00
0.02
0.04
0.06
0
5
10
15
Follow-up Time (years)

Figure 1

A. Gene expression in whole tissue





B. Prediction of young vs. old in independent test set of isolated glands

positive correlation with young signature

negative correlation with young signature