Published OnlineFirst November 15, 2016; DOI: 10.1158/0008-5472.CAN-16-1959

Molecular and Cellular Pathobiology

Cancer Research

Integrative Comparison of mRNA Expression Patterns in Breast Cancers from Caucasian and Asian Americans with Implications for Precision Medicine

Yanxia Shi[1], Albert Steppi[2], Ye Cao[1], Jianan Wang[3], Max M. He[4,5], Liren Li[6], and Jinfeng Zhang[2]

Abstract

Asian Americans (AS) have significantly lower incidence and mortality rates of breast cancer than Caucasian Americans (CA). Although this racial disparity has been documented, the underlying pathogenetic factors explaining it are obscure. We addressed this issue by an integrative genomics approach to compare mRNA expression between AS and CA cases of breast cancer. RNA-seq data from the Cancer Genome Atlas showed that mRNA expression revealed significant differences at gene and pathway levels. Increased susceptibility and severity in CA patients were likely the result of synergistic environmental and genetic risk factors, with arachidonic acid metabolism and PPAR signaling pathways implicated in linking environmental and genetic factors. An analysis that also added eQTL data from the Genotype-Tissue Expression Project and SNP data from the 1,000 Genomes Project identified several SNPs associated with differentially expressed genes. Overall, the associations we identified may enable a more focused study of genotypic differences that may help explain the disparity in breast cancer incidence and mortality rates in CA and AS populations and inform precision medicine. *Cancer Res; 77(2); 423–33. ©2016 AACR.*

Introduction

In 2016, an estimated 246,660 new breast cancer cases will be diagnosed in the United States (1). Despite advances in treatment and earlier detection, 40,880 will die from this disease (1). Breast cancer is a very heterogeneous disease. Understanding its heterogeneity is the key for developing personalized treatment strategies and an inevitable step toward the complete cure of the disease. In clinical practice, breast cancer is divided into subgroups based on the expression of the estrogen receptor (ER), progesterone receptor (PR), and the status of gene amplification of HER2. The four predominant subtypes reported extensively in literature are categorized on the basis of these receptors: luminal A (ER^+ and/or PR^+, $HER2^-$), luminal B (ER^+ and/or PR^+, $HER2^+$), HER2 (ER^-, PR^-, $HER2^+$), and triple negative (ER^-, PR^-, $HER2^-$). A subtype of triple-negative breast cancer (TNBC) is the basal or basal-like phenotype that is negative for all three receptors but expresses specific basal markers (cytokeratin 5/6, cytokeratin 14, cytokeratin 17, and EGFR; ref. 2). This phenotype is more likely to undergo metastasis and is associated with poorer prognosis (3). Recent large-scale genomic studies have suggested that the whole spectrum of breast cancer subtypes can be much more complex than the four commonly known subtypes (4, 5).

The incidence and death rates of breast cancer differ among different race groups. For example, age-standardized breast cancer incidence and mortality rates for Asian American (AS) are 30% and 50% lower, respectively, than those in non-Hispanic whites (1). Various aspects of this health disparity have been studied in the past with qualitatively similar observations (6–8). AS are the fastest-growing racial/ethnic group in the United States, representing 6.3% of the population (20.0 million/318.7 million) in 2014 (9). Breast cancer is the most commonly diagnosed cancer and the second leading cause of cancer-related death among AS women, with a total of 11,090 new invasive cases and 1,180 deaths expected to occur in 2016 (1). There is a substantial variation in breast cancer occurrence within the Asian population, with lower rates among groups that have immigrated more recently (1, 8). These differences are thought to be related to the extent of adoption of western behaviors that increase breast cancer risk, such as a later age at childbirth, fewer births, and higher body weight (10). Despite the differences in socioeconomic status and lifestyles, genetic factors likely play roles in the health disparity between the two racial groups (1, 11). The identities of these

[1]Department of Medical Oncology, Sun Yat-sen University Cancer Center, Collaborative Innovation Center of Cancer Medicine, Sun Yat-sen University, State Key Laboratory of Oncology in South China, Guangdong, P.R. China. [2]Department of Statistics, Florida State University, Tallahassee, Florida. [3]No. 2 High School of East China Normal University, Shanghai, P.R. China. [4]Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, Wisconsin. [5]Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, Wisconsin. [6]Department of Colorectal Surgery, Sun Yat-sen University Cancer Center, Collaborative Innovation Center of Cancer Medicine, Sun Yat-sen University, State Key Laboratory of Oncology in South China, Guangdong, P.R. China.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Y. Shi, A. Steppi, and Y. Cao contributed equally to this article.

Corresponding Authors: Jinfeng Zhang, Department of Statistics, Florida State university, Tallahassee, FL 32306. Phone: 850-644-3218; Fax: 850-644-5271; E-mail: jinfeng@stat.fsu.edu; Liren Li, lilr@sysucc.org.cn; and Yanxia Shi, shiyx@sysucc.org.cn

doi: 10.1158/0008-5472.CAN-16-1959

Published OnlineFirst November 15, 2016; DOI: 10.1158/0008-5472.CAN-16-1959

genetic factors, however, are poorly understood. Consequently, the molecular basis, if there is any, for the change in incidence rate after immigration has been unknown. Such molecular level understanding may shed light on better understanding of tumorigenesis of breast cancer and help develop personalized treatments.

A recent study has compared the gene and miRNA expression between a Chinese breast cancer population and an Italian cohort (12). The study did not detect significant differences between the two patient groups. The study then concluded that "Transcriptional similarity across transethnic cohorts may simplify translational medicine approaches and clinical management of breast cancer patients worldwide." In a recent study, we compared the expression of genes and transcripts between Caucasian American (CA) and African American breast cancer patients and identified hundreds of differentially expressed genes and several interesting differentially regulated pathways between the two patient populations (13). The study identified a novel noncoding RNA, which could be useful for early detection of breast cancer (13). We believe a similar study comparing CA and AS breast cancer patients may also offer some clues on the biological factors causing the health disparity and better understanding of the biology of breast cancer.

We used next-generation sequencing (NGS) data from the Cancer Genome Atlas (TCGA) to determine differentially expressed genes and other transcripts between a large number of age- and stage-matched AS and CA patient primary breast tumor samples. Using NGS data, we can determine differential expression between noncoding RNAs in addition to coding RNAs. This study represented, to our knowledge, the first time that NGS data were used in combination with a large patient cohort to investigate expression differences of genes and other transcripts between AS and CA breast cancer patients, and this work represented the largest and most comprehensive genomic study on health disparity between CA and AS breast cancer to date, where differentially expressed genes by stage and subtype between AS and CA tumors were identified for the first time, in addition to differentially expressed subnetworks and pathways.

Generally speaking, three types of genetic factors can be produced in a genome-wide comparison: (i) genetic factors that are truly associated with cancer health disparity, which we call category one factors; (ii) genetic factors with significant differences between two different racial populations in general, unrelated to the cancer health disparity (category two factors); and (iii) genetic factors that appear simply by chance due to multiple comparisons (category three factors). Gene-level comparison will generally produce differentially expressed genes/transcripts in all these three categories. Category three factors can be removed by correcting for multiple comparisons using more stringent criteria, with the cost of also removing some of the category one factors (an issue of balancing type I and type II errors). To discern category two factors, we performed various analyses beyond individual gene level, such as gene set enrichment analysis, pathway enrichment analysis and network-based analysis. The differentially regulated gene sets/pathways/subnetworks were then examined in the context of literature to identify the category one factors.

Among the category one factors, some are "drivers" that are upstream regulators causing health disparity, and some are "passengers" that are regulated by those drivers. Pathway and network-based analyses also help pinpoint the driver genetic factors and understand their mechanisms of actions.

We found from this study that there are widespread differences in the expression of genes and other transcripts between AS and CA primary breast tumor tissues. The health disparity is likely caused by an aggregated effect of all the associated biological factors (together with environmental factors), with some favoring one race and others favoring the other. The simple picture, where only one or a few genetic factors can explain the biological causes of the health disparity, may not exist. Both environmental and genetic factors work cooperatively to make CA population more breast cancer prone and also make the cancer of CA patients more severe. Specifically, arachidonic acid metabolism and PPAR signaling pathways may play central roles of linking environmental and genetic factors. Incorporation of eQTL and SNP data may help find SNPs associated with the differentially expressed genes and shed light on the genotype differences that drive the observed health disparity. Comparisons of mRNA expression of cancer between two races (or among multiple races) may offer a unique angle and some unique advantages for studying cancer biology and cancer heterogeneity compared to existing methods.

Materials and Methods

RNA-seq data from TCGA

Data used in this study were downloaded from Insilicom's BioKDE platform (insilicom.com), where processed, deidentified genomic and clinical data were downloaded from TCGA and integrated into a relational database for convenient queries. The deidentified patient data included up to 20,483 RNA sequence-derived gene expression values and clinical characteristics (e.g., age, cancer stage, receptor status). Fifty-seven AS tumor samples were matched by age and stage with 728 CA tumor samples, and a summary of the clinical characteristics of these patients is presented in Table 1. Subtypes were assigned to patient samples following ref. 13. The raw gene expression counts were normalized with reads per million (RPM), unless otherwise noted.

Differential gene expression

There are many methods for differential gene expression analysis based on different distribution assumptions (14–17). We chose to use DESeq2, an R Bioconductor package, for our analyses (18, 19). Patient gene counts input into DESeq2 were rounded to the nearest whole number but not normalized. Jobs were run on the Insilicom BioKDE server (insilicom.com). All the comparisons are AS versus CA, where increased expression (upregulation) means the expression in AS breast cancer is higher than the expression in CA breast cancer, and decreased expression (downregulation) means the expression in AS breast cancer is lower than the expression in CA breast cancer.

Sample matching

All the comparisons done in this study were matched by age and stage using R package, Matching (20). The matching was done differently than in our earlier study (13). The stage was matched exactly and age was matched in a way that any patients are considered matched if the difference of their ages is smaller or equal to 10. This is more flexible than the three categories used in the earlier study. The Matching package also performs optimal matching once the ratio of the two types of subjects is determined. To find the optimal ratio, we maximize the quantity, $nm/(n+m)$, where n is the number of patients of one race and m is the number

Table 1. Breast cancer patient clinical data from TCGA

	CA ($n = 728$)		AS ($n = 57$)		Fisher exact test
Characteristics	**Number**	**Percentage**	**Number**	**Percentage**	***P***
Age					0.013
<50	197	27.1%	23	40.4%	
50–64	311	42.7%	26	45.6%	
65+	220	30.2%	8	14.0%	
Tumor stage					0.17
0	1	0.137%	0	0	
1	134	18.4%	4	7.02%	
2	409	56.2%	41	71.93%	
3	165	22.7%	12	21.05%	
4	11	1.51%	0	0	
Unknown	8	1.10%	0	0	
Tumor type					0.0084
Luminal A	298	40.9%	21	36.84%	
Luminal B	77	10.6%	5	8.77%	
Triple negative	67	9.20%	8	14.04%	
HER2 type	19	2.61%	7	12.28%	
Unknown	267	36.7%	16	28.07%	
Vital status					0.022
Living	648	89.7%	56	98.2%	
Deceased	80	10.3%	1	1.75%	

NOTE: The total number of CA breast cancer patients is 728, of which 399 are matched to 57 AS patients by age and stage.

of patients of the other race. In this setup, we assume the power calculation of testing differential gene expression can be approximated with a two sample test of two normal samples with different sample sizes. We found that the new matching method produced notable differences, but its exact effect was not evaluated systematically.

Pathway and network analysis

Differentially expressed genes identified by DESeq2 were input into the server at humanmine.org (21). A separate analysis was also performed using Insilicom's Integrative Genomic Analysis pipeline. All the *P* values were corrected for multiple comparisons.

Gene Expression Network Analysis (GXNA) was used to identify subnetworks where differentially expressed genes are enriched compared with other parts of the whole gene/protein interaction network (22). Gene expression data input into GXNA was first $\log_2$-transformed so that differences in means corresponded to fold-change, and default GXNA parameters were used. STRING, a database of protein–protein interactions was used to visualize GXNA results (23).

Pathway visualization

The visualizations of differentially regulated pathways were made possible using the Pathview package in R (24). It allows users to easily spot up or downregulated genes in the pathway to understand how it is differentially regulated. The workflow of our differential gene expression analysis is given in Supplementary Fig. S1.

Association of SNPs to gene expression in breast cancer

Combining data obtained from Genotype-Tissue Expression (GTEx), Expression Quantitative Trait Loci (eQTL) studies, and the 1,000 Genomes project with the gene expression data from TCGA, we infer potential associations of certain SNPs to gene expressions in breast cancer. First, given a gene of interest, SNPs that are associated with expression of this gene are obtained from GTEx data portal for eQTL (http://www.gtexportal.org/home/eqtl), regardless of tissue types; second, the SNPs are used to compute the allele frequencies for races of interest using data from 1,000 Genomes project (http://www.1000genomes.org/); third, the effect sizes from eQTL and the allele frequencies are used to compute relative expressions of the gene (marginal expression without conditioning on other factors); finally, these relative expressions computed from allele frequencies and eQTL effect sizes are sorted to compare with the actual expressions observed in TCGA data.

Correlation analysis

For continuous numerical clinical features, Spearman rank correlation coefficients and two-tailed *P* values were estimated using "cor.test" function in R. For categories of continuous type of clinical data, Wilcoxon rank sum test was applied to compare their mean difference using "wilcox.test (continuous.clinical ~ as.factor(group), exact = FALSE)" function in R. This test is equivalent to the Mann–Whitney test.

Results

Table 1 shows the breakdown of the demographics of all the patients obtained from TCGA. CA women have a mortality rate of 10.3%, whereas the AS counterpart has a mortality rate of only 1.75% (Fisher exact test *P* value is 0.022). When matching age and stage, all the AS patients were kept and 399 CA patients were selected. The mortality rate of CA reduced to 8.77%. The Fisher exact test ($P = 0.044$) is still significant. As mentioned before, this disparity is likely caused by both socioeconomic and biological factors. Information regarding socioeconomic status or access to care is not available at TCGA to assess the effects of these factors to the health disparity.

The outline of the analyses we performed in this study is shown in Supplementary Fig. S1. First, differential expression analysis at the individual gene level was conducted to identify genes and other transcripts that differ between CA and AS breast cancer patients using DESeq2; second, pathway analysis and gene set enrichment analysis were performed using the server at humanmine.org and using R Pathview package; third, Gene Expression Network Analysis (GXNA) was used to identify differentially

Published OnlineFirst November 15, 2016; DOI: 10.1158/0008-5472.CAN-16-1959

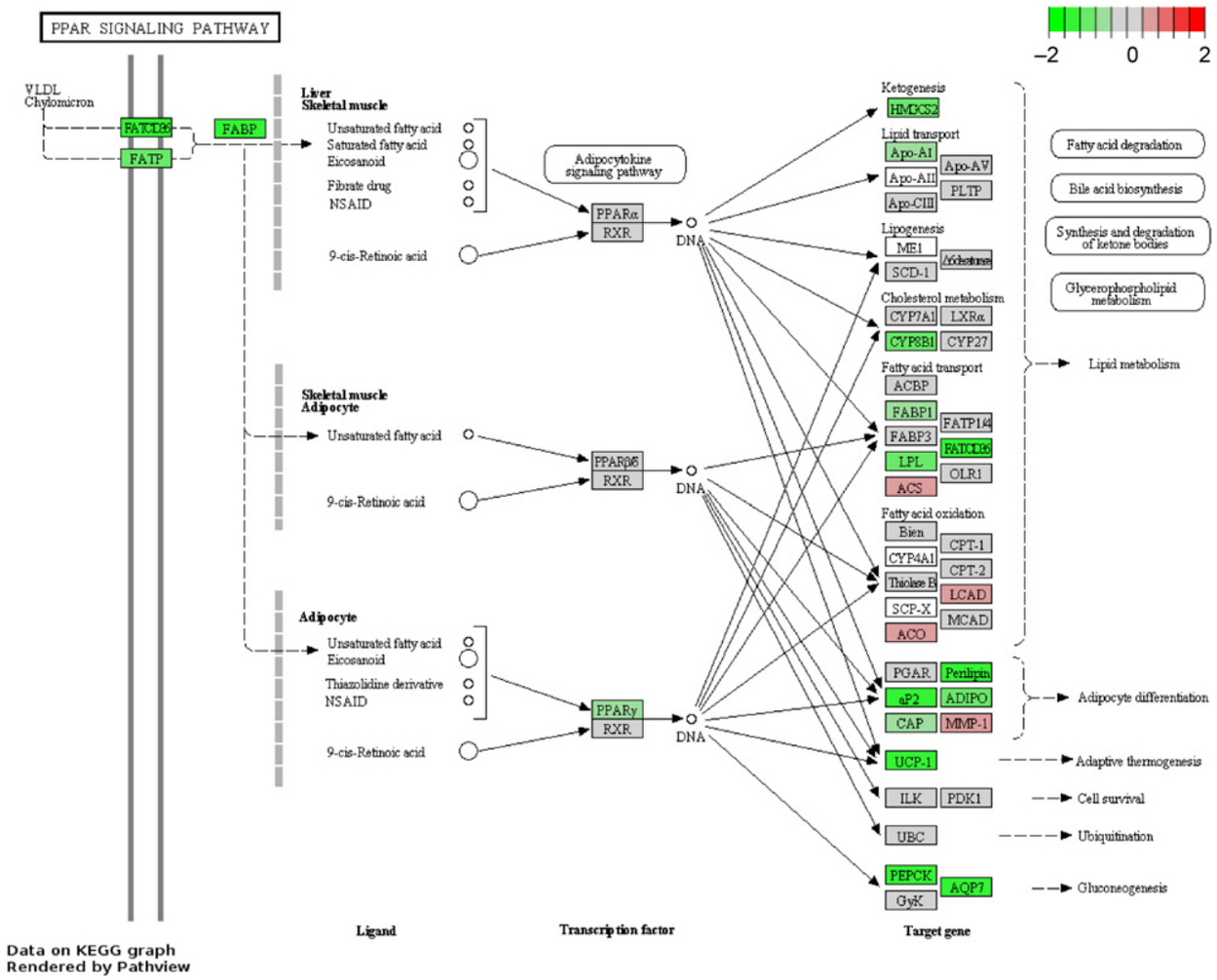

Figure 1.

Differential gene expression in PPAR signaling pathway. Green, log2-fold changes of downregulated genes in AS breast cancer patients; red, log2-fold changes of upregulated genes in AS breast cancer patients.

expressed subnetworks; finally, patients were stratified to study differences in stage- or subtype-specific gene expression again with DESeq2 (Fig. 1).

In total, 927 unique genes and other transcripts were shown to be differentially expressed (adjusted P value ≤ 0.05 and fold change ≥ 2) between the various comparisons performed (Supplementary Tables S1–S8). To identify overall trends in gene expression, primary breast tumor expression data between AS and CA patients irrespective of subtype or stage (i.e., the "overall" comparison) was first compared. Two hundred and eighty-four genes and other transcripts, 1.4% of the 20,483 genes and other transcripts assessed, were identified by DESeq2 as differentially expressed between AS and CA primary breast tumors (adjusted P value ≤ 0.05 and fold change ≥ 2; Supplementary Table S1). Forty-seven of these genes and other transcripts showed increased expression in AS tumors, whereas 236 were decreased. Selected genes and other transcripts with high fold-change and relevance to cancer are shown in Table 2.

The 284 significantly differentially expressed genes and other transcripts from the overall comparison were queried at humanmine.org. Humanmine.org has several widgets analyzing the given gene sets in terms of their enrichment in gene ontology (GO) term, pathways, protein domains, interactions, publications, and interaction networks. In this study, we looked at pathway, GO terms, and differentially regulated network modules. In pathway/GO enrichment analysis, genes were placed into known pathways/GO terms and ranked by the probability that the associated genes are enriched in the input gene list, and a low P value indicates a higher enrichment and greater chance of differential regulation. The significantly differentially regulated pathways (Table 3) are discussed in more details below.

PPAR signaling pathway

The top significantly differentially regulated KEGG pathway is PPAR signaling pathway (adjusted P value = 3.88e−4) with 10 genes significantly differentially expressed. PPARs are a group of nuclear receptor proteins that function as transcription

Table 2. Differentially expressed genes and transcripts between CA and AS breast cancer from overall compassion, with age and stage matched

Gene	Log_2 fold-change	P_{adj}	Relevance
PCSK1	−1.96	1.01e−10	Associated with obesity a known risk factor for breast cancer.
CIDEC	−1.84	1.83e−9	Expression correlates with BMI.
GLYAT	−1.8	2.89e−8	Associated with lean body mass.
SHC4	−1.77	1.74e−9	SHC4 signaling adaptor facilitates ligand-independent phosphorylation of the EGF receptor.
OR3A2	−1.75	4.5e−9	Upregulated in tobacco smokers. A breast cancer risk factor.
LEP	−1.72	6.75e−8	Associated with breast cancer risk, particularly in obese women. Plays a major role in the regulation of body weight.
PIWIL2	−1.71	2.71e−12	The combination of piR-932 and PIWIL2 may be a positive regulator in the process of breast cancer stem cells, and they both could be the potential targets for blocking the metastasis of breast cancer.
MGAM	−1.69	4.13e−9	Overexpressed in oral squamous cell carcinoma.
SLC19A3	−1.63	4.13e−9	Upregulation occurs to maintain nutrient uptake in hypoxic regions of tumors.
TDRD12	−1.63	5.07e−7	Hypomethylated in adenoid cystic carcinoma.
LY6D	−1.63	9.51e−7	Increased expression associated with poor outcomes in various cancers, including breast cancer.
SLC4A4	−1.59	5.16e−9	Regulates PH in hypoxic and acidic tumor environments, aiding in tumor growth and metastasis. A potential drug target.
PSG4	−1.59	2.4e−6	Pregnancy-specific glycoprotein. Expressed in pregnant women. Hormonal changes in pregnancy could be associated with breast cancer.
C8A	1.62	2.35e−7	Potential circulating biomarker for hepatocellular carcinoma.
RETN	2.05	1.62e−11	Differentially expressed in several studies investigating cancer health disparity.
UTS2	2.23	2.76e−14	There is strong evidence that suggest urotensin-II as the significant contributor of angiogenesis as well as cell proliferation and tumor biology.

factors that play essential roles in the regulation of cellular differentiation, development, and metabolism (carbohydrate, lipid, protein), and tumorigenesis of higher organisms (25, 26). Three types of PPARs have been identified: alpha, gamma, and delta (beta), where PPARγ is expressed in virtually all tissues. Endogenous ligands for the PPARs include free fatty acids and eicosanoids. The 10 genes in the pathway identified by the analysis are all significantly downregulated (fold change ≥ 2 and adjusted P value ≤ 0.05) in AS breast cancer tissues, indicating PPAR pathway is downregulated in AS breast cancer patients compared with CA patients. The PPAR pathway and the differential expression of associated genes drawn by Pathview can be visualized in Fig. 1.

The other significant KEGG pathway was neuroactive ligand–receptor interaction (adjusted P-value = 6.99e−3) with 15 genes in the list. This pathway does not appear to be closely related to breast cancer. It may be a pathway that is differentially regulated between AS and CA population in general (category two factors).

GXNA was then used to identify differentially expressed subnetworks between AS and CA tumors (13, 22). GXNA takes patient expression data as input and then outputs a ranked list of subnetworks that contain interactions involving the differentially expressed genes. GXNA identified a large connected subnetwork with 40 genes. Most of the genes in this subnetwork are downregulated in AS breast cancer patients. It contains some well-known cancer-related genes such as *TP53* and *ERBB2* (*HER2*). The network drawing tool at STRING database (23) was used to visualize the GXNA network and to show potential gene product interactions beyond those included in the GXNA interaction file (Fig. 2).

The 40 genes in the GXNA network were then used as input for humanmine.org server to identify enriched pathways (Table 3). Four pathways were significant (using 0.01 as cutoff for adjusted P values): arachidonic acid metabolism pathway ($P = 1.074$-4, 6 genes), long-term potentiation ($P = 0.0062$, 5 genes), pathways in cancer ($P = 0.0066$, 9 genes), and VEGF signaling pathway ($P = 0.0092$, 5 genes). Among them, long-term potentiation was not found to be related to breast cancer and may be a category two pathway.

Arachidonic acid metabolism pathway

The arachidonic acid pathway constitutes one of the main mechanisms for the production of pain and inflammation, as well as controlling homeostatic function. Pathview diagram with differentially expressed genes colored for arachidonic acid metabolism pathway is shown in Fig. 3.

Pathways in cancer

Several cancer-related pathways are differentially regulated between AS and CA breast cancer patients ($P = 6.59$e−03), which can be seen from the Pathview diagram for pathways in cancer (Supplementary Materials and Supplementary Fig. S2). Overall, through visual inspection, we can see that substantially more genes are downregulated in AS patients than upregulated, indicating overall the cancers in AS patients are probably not as severe as those in CA patients. This is consistent with the fact that the death rate for breast cancer is much lower for AS than CA.

Table 3. Differentially regulated pathways between AS and CA breast cancer

Pathway name	Representative genes and transcripts	P_{adj}
PPAR signaling pathway	*UCP1, SLC27A6, PCK1, CD36, AQP7, PLIN1, FABP4, CYP8B1, ADIPOQ, PLIN4*	3.88e−4
Neuroactive ligand–receptor interaction	*GLP2R, CCKBR, GLP1R, CRHR1, CHRM3, LHCGR, AGTR1, DRD2, GRIA1, CHRNB2, NPY5R, GRIK1, LEP, GPR156, NPY4R*	7.0e−03
Arachidonic acid metabolism	*PTGS2, GGT1, PTGDS, PLA2G4A, PTGIS, GPX3*	1.07e−04
Long-term potentiation	*PRKACB, ITPR1, PRKCA, GRIA1, GRIN2A*	6.20e−03
Pathways in cancer	*PTGS2, NFKB2, RAC1, IL6, TP53, ERBB2, PRKCA, VEGFD, CTNNB1*	6.59e−03
VEGF signaling pathway	*PTGS2, PLA2G4A, RAC1, PRKCA, SRC*	9.17e−03

Published OnlineFirst November 15, 2016; DOI: 10.1158/0008-5472.CAN-16-1959

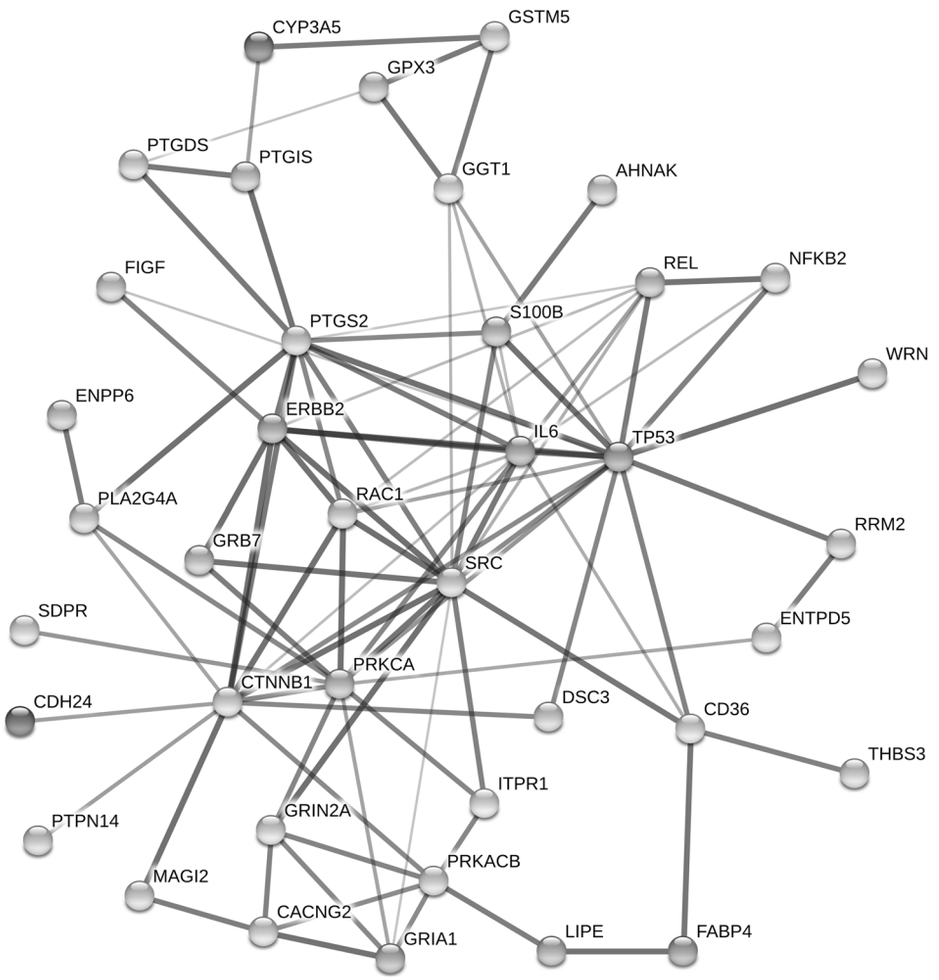

Figure 2.
The differentially expressed subnetwork obtained by GXNA.

VEGF signaling pathway

The VEGF signaling pathway is also downregulated significantly in AS breast cancer patients (adjusted P value = 9.17e−03). The VEGF signaling pathway regulates vascular development in the embryo (vasculogenesis) and new blood vessel formation (angiogenesis). Upregulation of angiogenesis is a key step in sustained tumor growth and may also be critical for tumor metastasis (27, 28).

Stage-matched and then subtype-matched AS and CA tumors were next compared (Supplementary Tables S3–S8). Because of sample size limit, we pooled stage I and II patients as early stage, and stage III and IV as late stage. Selected results from these comparisons are summarized in Supplementary Table S2 in Supplementary Materials. Interestingly, a small number of genes and other transcripts (totally 23) overlapped between early- and late-stage comparisons (Supplementary Fig. S3). Overlapping among overall comparison and subtype-specific comparison is also poor, similar to what was found in an earlier study comparing CA with African American breast cancer patients (13), indicating the differences within subtypes are likely specific to each subtype and in depth study may reveal more interesting biology of cancer health disparity between the two racial groups. The only gene that is differentially expressed in all the comparisons (overall, early stage, late stage, and subtype-specific comparisons) is *DDX11L2*, whose biological significance is unclear.

In summary, 267 genes and other transcripts were differentially expressed in the early-stage comparison (228 decreased and 39 increased in AS breast cancer patients), 137 genes and other transcripts were differentially expressed in the late-stage comparison (94 decreased and 43 increased), 507 genes and other transcripts were differentially expressed in luminal A comparison (439 decreased and 68 increased), 49 genes and other transcripts were differentially expressed in HER2 comparison (34 decreased and 15 increased), and 104 genes and other transcripts were differentially expressed in triple-negative comparison (10 decreased and 94 increased). An interesting observation is that for all the comparisons the numbers of downregulated genes (decreased) are significantly higher than those upregulated in AS breast cancer patients, except for triple-negative subtype, for which the number of upregulated genes is significantly higher than those downregulated.

SNPs associated with gene expression in breast cancer

A fundamental question in disease studies is to understand how genotypes affect phenotypes of the disease of interest. We combined gene expression data with eQTL data from GTEx project and SNP allele frequency information from 1,000 Genomes project (see Materials and Methods for details). We use gene *CYP1A1* to illustrate our approach. Because the expression of *CYP1A1* is

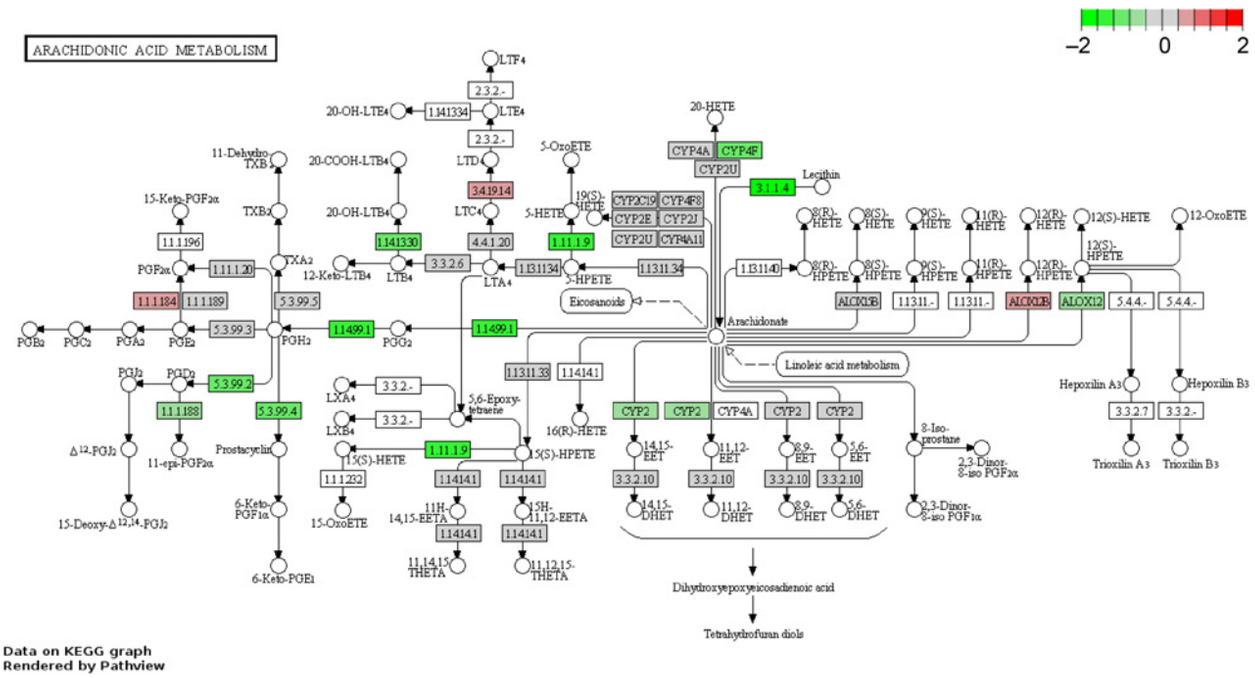

Figure 3.

Differential gene expression in arachidonic acid metabolism pathway. Green, downregulated genes in AS breast cancer patients; red, upregulated genes in AS breast cancer patients.

linked to ER (29) and arachidonic acid (AA) metabolism (30), the variants of *CYP1A1* identified in this study may be associated with tumorigenesis of breast cancer through metabolism of ER and AA. This analysis not only will help us to identify disease-associated genomic variants, but also point to plausible mechanisms for the associations.

We found one variant (ID: rs55646017) that is associated with expression of *CYP1A1* with an effect size of 0.36 and a *P* value of 1.5e−5. The reference allele is T and alternative allele is C. The frequencies of alternative allele in South Asian population, East Asian population, and European population are 0.29, 0.23, and 0.49, respectively. This agrees qualitatively with the observed expression difference between AS and CA breast cancer patients (downregulation in AS with a log2-fold change of −0.92). This variant may be a promising candidate in follow-up association studies for breast cancer. This approach can be applied to other differentially expressed genes to provide both additional evidences and mechanism explanations for potential associations.

Survival analysis for the differentially expressed genes

Survival analysis results for the 284 differentially expressed genes from the overall comparison were extracted from Broad GDAC of Broad Institute (http://gdac.broadinstitute.org/). We also performed survival analysis using the same approach (log-rank test in univariate Cox regression analysis with proportional hazards model) and obtained very similar results (Supplementary Materials). Totally, 28 genes have *P* values smaller than 0.05 (Supplementary Materials). Among them, *PTGS2* has a log-rank test *P* value of 0.026. *PTGS2* encodes the *COX-2* enzyme, which catalyzes the conversion of arachidonic acid to prostaglandins and other eicosanoids. Expression of *PTGS2* is also significantly correlated with pathologic stage ($P = 4.40e{-}05$) and pathologic T stage ($P = 0.00011$). *PTGS2* expression, which is undetectable in most normal tissues, is induced in response to hypoxia, inflammatory cytokines, tumor promoters, growth factors, and other stressors (31). *PTGS2* is involved in carcinogenesis, immune response suppression, inhibition of apoptosis, angiogenesis, and tumor cell invasion and metastasis. Recent studies have indicated that *PTGS2* genetic variation is associated with breast cancer susceptibility (32, 33). Furthermore, overexpression of *PTGS2* in patients with breast cancer is associated with a worse prognosis (34).

Correlation analysis of differentially expressed genes with clinical variables

The correlations between clinical variables and the differentially expressed genes were also extracted from Broad GDAC of Broad Institute (http://gdac.broadinstitute.org/). We examined the following clinical variables: pathologic stage, pathologic T stage, pathologic N stage, pathologic M stage, and the number of lymph nodes. Overall, 129 genes are significantly associated with at least one clinical variable (including survival time). Several genes are highly correlated with the stage of breast cancer: *MMRN1* ($P = 1.83e{-}07$), *PLIN1* ($P = 2.60e{-}05$), *PTGS2* ($P = 4.40e{-}05$), and *TMEM132C* ($P = 1.67e{-}05$). Two genes are correlated with metastasis (M-stage): *KLK5* ($P = 0.0074$) and *PTGS2* ($P = 0.00082$). The detailed result is included in Supplementary Materials.

Discussion

The heterogeneity of breast cancer is manifested in its racial disparities, as in many other aspects of the disease. Racial

Published OnlineFirst November 15, 2016; DOI: 10.1158/0008-5472.CAN-16-1959

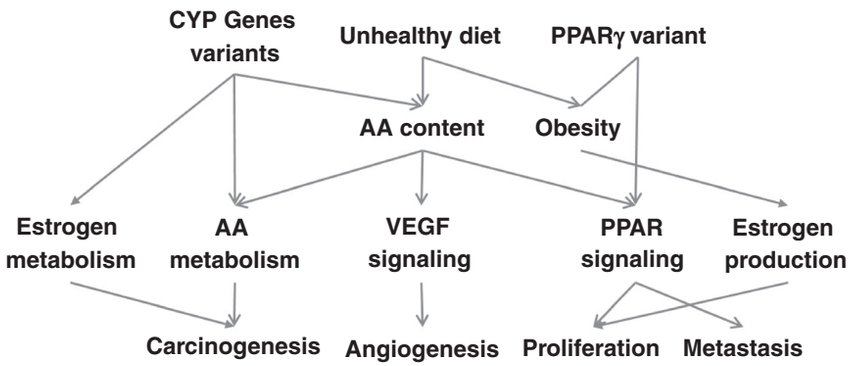

Figure 4.

The environmental and genetic factors that may explain the health disparity between CA and AS breast cancer and a potential link between obesity and breast cancer risk. (i) Unhealthy diets cause higher AA (arachidonic acid) content in the body, which cause the activation or upregulation of several pathways, including AA metabolism, VEGF signaling, and PPAR signaling, which are related to cancer development and migration. (ii) Known genetic variants in CYPs may cause the differential expression of these genes and different metabolic patterns of AA, which, in turn, affect AA content and AA metabolism pathway. (iii) A PPARγ variant exists in Chinese population that is associated with severe obesity. The variant may also exist in people in western countries. The variant may cause the differential regulation of PPAR signaling pathway in the two race groups. (iv) Both estrogen production and metabolism are important for breast cancer risk and genetic differences between AS and CA population may affect the estrogen metabolism, which, in turn, may have affected their breast cancer incidence rate.

disparities may provide a unique angle to investigate the heterogeneity of breast cancer. In this study, we compared mRNA expression between CA and AS breast cancer primary tumor samples using RNAseq data obtained from TCGA. We matched the patients on age and stage and performed overall, early stage, late stage, and subtype-specific comparisons. We have identified a large number of differentially expressed genes and other transcripts for each comparison. We focused on the analysis of overall comparison in this study. The results from other comparisons also deserve more detailed analysis and interpretation, especially the comparison of triple-negative patients, which showed unique expression patterns compared with other comparisons. In the comparison of triple-negative subtype, AS triple-negative breast cancer patients have significantly larger number of upregulated genes than the CA counterpart whereas in other comparisons AS patients have significantly lower number of upregulated genes than downregulated genes. In the discussion below, downregulation or upregulation refers to the regulation changes in AS breast cancer. Pathway and network analysis using significantly differentially expressed genes revealed several differentially regulated pathways between the two racial groups, namely, PPAR signaling pathway, AA metabolism pathway, pathways in cancer, and VEGF pathway.

The biological factors associated with the breast cancer health disparity between AS and CA

The difference in incidence and mortality rate of breast cancer between AS and CA populations is likely the aggregated effect of multiple factors, including genetic, lifestyle, and environmental factors, with some favoring one race and others favoring the other. The lower incidence and mortality rate of breast cancer in AS population is likely because the factors favoring AS outweigh those favoring CA. With a set of carefully chosen and optimized analysis methods, our study revealed a complex picture with interconnected genetic and environmental factors leading to a plausible explanation for the health disparity between AS and CA populations.

We first list the observations/findings made in previous studies (1, 8, 29, 35–41), followed by the evidences we have collected in this study before drawing the hypotheses.

Observations/findings made in previous studies:

1. AS have significantly lower incidence and mortality rate of breast cancer than CA (1).
2. The incidence rate of AS increases with the time they live in United States (1, 8).
3. Unhealthy diet with more meats leads to more consumption of arachidonic acid (35). United States have more meat consumption than Asian populations (36).
4. A variant (SNP) on PPARγ is associated with severe obesity in Chinese population (37).
5. Inhibitors of the 20-HETE–producing enzymes of the Cytochrome (CYP450)4A and CYP4F families can block the proliferation of breast cancer cell lines (38). Many variants of CYP family of genes are known, which are linked to different activities toward their substrates (39).
6. *PLIN1* (perilipin) positivity was found to be an independent poor prognostic factor for metastatic breast cancer (40).
7. Metabolic activation of carcinogenic estrogens plays a critical role in estrogen-induced carcinogenesis (41). Estrogen is metabolized in breast tissue by the product of *CYP1A1* gene (29).
8. *PTGS2* (*COX-2*) is involved in carcinogenesis, immune response suppression, inhibition of apoptosis, angiogenesis and tumor cell invasion, and metastasis (31). Recent studies have indicated that *PTGS2* genetic variation is associated with breast cancer susceptibility (32, 33). Furthermore, overexpression of *PTGS2* in patients with breast cancer is associated with a worse prognosis (34).

Evidences obtained in this study (down- or upregulation refers to the expression changes in AS breast cancer):

1. Arachidonic acid metabolism pathway is differentially regulated between AS and CA breast cancer with some key genes downregulated.
2. PPAR signaling pathway is significantly downregulated. A key downregulated gene is *PPARγ*.

3. Cancer-related pathways activated by PPAR signaling pathway are also downregulated, such as VEGF signaling and WNT signaling pathways (Supplementary Fig. S2).
4. *CYP4F*, the cytochrome P450 gene that metabolizes arachidonic acid into 20-HETE, a carcinogen, is downregulated. The other two genes in this pathway, *CYP4A* and *CYP2U*, are marginally downregulated.
5. *PLIN1*, a downstream gene of *PPARγ*, is downregulated. *PLIN1* is also strongly correlated with the stage of breast cancer ($P = 2.60e{-}05$).
6. *LEP* (leptin) and *IL6* are both downregulated. These adipokines are both secreted by adipocytes and can act in either autocrine or paracrine manners to increase production of aromatase, which is directly related to increased synthesis of estrogen (42).
7. *CYP19A1* (aromatase), the enzyme for synthesizing estrogen, is downregulated.
8. *CYP1A1*, the enzyme for metabolizing estrogen in breast tissues (29), is downregulated.
9. *PTGS2* is downregulated in AS breast cancer patients. Survival analysis showed that *PTGS2* is marginally significant for survival. Correlation analysis showed that *PTGS2* is significantly correlated with pathologic stage and T-stage of breast cancer.

The combined discoveries in both previous and current studies lead us to the hypothesis that breast cancer health disparity between CA and AS populations is driven by the interaction between environmental (lifestyle) and a few specific genetic factors (Fig. 4). The fact that the incidence rate for AS increases with the time lived in the United States suggests a role the environmental factor plays in this health disparity. Asians tend to have a healthier diet with less meat consumption compared with Americans. Asians are less likely to suffer from obesity than CAs. The average BMI is also lower in the Asian population. The downregulation of many genes in the AA metabolism pathway in AS breast cancer (Fig. 3) may be due to the less AA content in their tumor tissues (and possibly normal tissues as well). The genetic variants in CYP family genes metabolizing AA and its metabolites may also play a role in the differential regulation of the AA metabolism pathway (39, 43, 44). Recent studies have suggested that AA and its metabolite pathway play a critical role in the development and progression of breast cancer (45–47). It was also suggested that AA-activated mTOR signaling also plays critical roles in the angiogenesis and tumorigenesis of breast cancer (48). Taken together, AA metabolism may play a key role in the breast cancer health disparity between AS and CA populations. More experimental studies would be required to elucidate how genetic and environmental factors cooperatively affect AA metabolism and its downstream pathways.

A PPARγ variant was found to be associated with severe obesity in the Chinese population (37). The variant may also exist in the CA population and be associated with obesity in that population. PPARγ and its downstream gene, *PLIN1* (perilipin), are all downregulated in AS (Fig. 1). Genetic factors, such as PPARγ variant, may also play a role in the health disparity.

A well-known link between obesity and breast cancer risk is estrogen (49). However, the molecular level mechanism for the association has been poorly understood. We found that estrogen signaling pathway was not significantly differentially regulated between AS and CA breast tumors, indicating the mechanism for the estrogen linkage may not be a simple one. It has been shown in a previous study that metabolic activation of carcinogenic estrogens plays a critical role in estrogen-induced carcinogenesis (41). Estrogen is metabolized in breast tissue by *CYP1A1* (29). We found that *CYP1A1* is downregulated in AS breast cancer patients, which may have also contributed to the breast cancer health disparity. This observation suggests that among obese women, those with a *CYP1A1* variant that metabolizes estrogen more effectively may have higher risk of breast cancer than those with a *CYP1A1* variant that metabolizes estrogen poorly. This hypothesis can be tested in future studies. The lack of BMI and obesity information in TCGA makes a more in depth study on the link between obesity and breast cancer risk infeasible. Future research using datasets with such information will likely deepen our understanding in this important public health issue.

The Asian population represents 60% of the human population worldwide. A comparison between AS and CA breast cancer could also help researchers to understand the genetic differences between Caucasian and Asian populations and how such differences may affect their breast cancer risk and responses to treatment. Several of our discoveries have implications for drug discovery and development of diagnostic and prognostic tools. We discuss two of them below.

Arachidonic acid metabolism or PPAR signaling pathway as potential drug targets

Both the AA metabolism and PPAR signaling pathways are significantly differentially regulated between AS and CA breast cancer. The most notable changes are some downregulated genes in both pathways in AS breast cancer. Clinical trials for drugs targeting those genes/proteins that are significantly differentially expressed between breast cancer cells of the two racial groups would likely see different outcomes depending upon the racial makeup of the subjects. These clinical trials should consider race as a confounding factor. It may also be beneficial to have a personalized medicine component that takes the expressions of the drug targets or those associated variants into account.

PLIN1 as a prognostic factor

In a previous study, *PLIN1* positivity was found to be an independent poor prognostic factor for metastatic breast cancer (40). To use *PLIN1* as a biomarker in clinical practice, it needs to be validated in a real clinical setting. Because of the large expression difference of *PLIN1* between the breast cancers of CA and AS populations, the chances for it to pass clinical validation may be quite different in western and Asian countries. Again, clinical trials involving *PLIN1* may benefit from having a personalized medicine component that takes the expressions of the biomarker or those associated variants into account.

In summary, our study revealed wide-spread transcriptome differences between CA and AS breast cancer and discovered a few promising candidates for future studies aiming at developing more personalized strategies for breast cancer.

Disclosure of Potential Conflicts of Interest

J. Zhang is the CEO of Insilicom LLC and has ownership interest (including patents) in Insilicom LLC. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: Y. Shi, Y. Cao, L. Li, J. Zhang
Development of methodology: Y. Shi, Y. Cao, J. Zhang

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): Y. Shi
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Y. Shi, A. Steppi, J. Wang, M.M. He, J. Zhang
Writing, review, and/or revision of the manuscript: Y. Shi, A. Steppi, Y. Cao, M.M. He, L. Li, J. Zhang
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Y. Shi, J. Zhang
Study supervision: L. Li, J. Zhang

Acknowledgments

We thank Insilicom (www.insilicom.com) for providing computing resources, curated data, and analysis pipelines for this research.

Grant Support

Dr. Y. Shi received support from Science and Technology Foundation of Guangdong province (2013B021800062 and 2012B06170008) and the International Collaboration Foundation of Science and Technology Planning Project of Guangdong Province of China (2016A050502015). Dr. M.M. He received support from the NIH (UL1TR000427) and the Marshfield Clinic Research Foundation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 22, 2016; revised October 20, 2016; accepted November 3, 2016; published OnlineFirst November 15, 2016.

References

1. American Cancer Society. Cancer Facts & Figures 2016. Atlanta, GA: American Cancer Society; 2016.
2. Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, et al. Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. Clin Cancer Res 2009;15:2302–10.
3. Rakha EA, Reis-Filho JS, Ellis IO. Basal-like breast cancer: a critical review. J Clin Oncol 2008;26:2568–81.
4. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70.
5. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346–52.
6. Jin H, Pinheiro PS, Xu J, Amei A. Cancer incidence among Asian American populations in the United States, 2009–2011. Int J Cancer 2016;138: 2136–45.
7. Sineshaw HM, Gaudet M, Ward EM, Flanders WD, Desantis C, Lin CC, et al. Association of race/ethnicity, socioeconomic status, and breast cancer subtypes in the National Cancer Data Base (2010–2011). Breast Cancer Res Treat 2014;145:753–63.
8. Gomez SL, Quach T, Horn-Ross PL, Pham JT, Cockburn M, Chang ET, et al. Hidden breast cancer disparities in Asian women: disaggregating incidence rates by ethnicity and migrant status. Am J Public Health 2010;100 Suppl 1:S125–31.
9. Colby S, Ortman J. Projections of the size and composition of the U.S. population: 2014 to 2060. Washington, DC: Census Bureau; 2014. p. 25–2243.
10. Keegan TH, Gomez SL, Clarke CA, Chan JK, Glaser SL. Recent trends in breast cancer incidence among 6 Asian groups in the Greater Bay Area of Northern California. Int J Cancer 2007;120:1324–9.
11. Iqbal J, Ginsburg O, Rochon PA, Sun P, Narod SA. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. JAMA 2015;313:165–73.
12. Huang X, Dugo M, Callari M, Sandri M, De Cecco L, Valeri B, et al. Molecular portrait of breast cancer in China reveals comprehensive transcriptomic likeness to Caucasian breast cancer and low prevalence of luminal A subtype. Cancer Med 2015;4: 1016–30.
13. Stewart PA, Luks J, Roycik MD, Sang QX, Zhang J. Differentially expressed transcripts and dysregulated pathways in African American Breast Cancer. PLoS One 2013;8:e82460.
14. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 2008;18:1509–17.
15. Wang L, Feng Z, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 2010;26:136–8.
16. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 2008;320:1344–9.
17. Robinson MD, Smyth GK. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics 2007;23:2881–7.
18. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106.
19. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.
20. Sekhon JS. Multivariate and propensity score matching software with automated balance optimization: the matching package for R. J Stat Softw 2011;42:1–52.
21. Smith RN, Aleksic J, Butano D, Carr A, Contrino S, Hu F, et al. InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. Bioinformatics 2012;28:3163–5.
22. Nacu S, Critchley-Thorne R, Lee P, Holmes S. Gene expression network analysis and applications to immunology. Bioinformatics 2007;23:850–8.
23. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013;41(Database issue): D808–15.
24. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics 2013;29: 1830–1.
25. Tyagi S, Gupta P, Saini AS, Kaushal C, Sharma S. The peroxisome proliferator-activated receptor: a family of nuclear receptors role in various diseases. J Adv Pharm Technol Res 2011;2:236–40.
26. Feige JN, Gelman L, Michalik L, Desvergne B, Wahli W. From molecular action to physiological outputs: peroxisome proliferator-activated receptors are nuclear receptors at the crossroads of key cellular functions. Prog Lipid Res 2006;45:120–59.
27. Hicklin DJ, Ellis LM. Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis. J Clin Oncol 2005;23:1011–27.
28. Ferrara N, Hillan KJ, Gerber HP, Novotny W. Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. Nat Rev Drug Discov 2004;3:391–400.
29. Tsuchiya Y, Nakajima M, Yokoi T. Cytochrome P450-mediated metabolism of estrogens and its regulation in human. Cancer Lett 2005; 227:115–24.
30. Schwarz D, Kisselev P, Ericksen SS, Szklarz GD, Chernogolov A, Honeck H, et al. Arachidonic and eicosapentaenoic acid metabolism by human CYP1A1: highly stereoselective formation of 17(R),18(S)-epoxyeicosatetraenoic acid. Biochem Pharmacol 2004;67:1445–57.
31. Schmedtje JF Jr, Ji YS, Liu WL, DuBois RN, Runge MS. Hypoxia induces cyclooxygenase-2 via the NF-kappaB p65 transcription factor in human vascular endothelial cells. J Biol Chem 1997;272:601–8.
32. Dossus L, Kaaks R, Canzian F, Albanes D, Berndt SI, Boeing H, et al. PTGS2 and IL6 genetic variation and risk of breast and prostate cancer: results from the Breast and Prostate Cancer Cohort Consortium (BPC3). Carcinogenesis 2010;31:455–61.
33. Langsenlehner U, Yazdani-Biuki B, Eder T, Renner W, Wascher TC, Paulweber B, et al. The cyclooxygenase-2 (PTGS2) 8473T>C polymorphism is associated with breast cancer risk. Clin Cancer Res 2006;12:1392–4.
34. Festa-Vasconcellos JS, Piranda DN, Amaral LM, Indio-do-Brasil V, Koifman S, Vianna-Jorge R. Polymorphisms in cyclooxygenase-2 gene and breast cancer prognosis: association between PTGS2 haplotypes and histopathological features. Breast Cancer Res Treat 2012;132:251–8.
35. Li D, Ng A, Mann NJ, Sinclair AJ. Contribution of meat fat to dietary arachidonic acid. Lipids 1998;33:437–40.

36. The Food and Agricultural Organization of the United States. Meat consumption per capita by country. Rome, Italy: Food and Agriculture Organization of the United Nations; 2009.
37. Chen HH, Lee WJ, Fann CS, Bouchard C, Pan WH. Severe obesity is associated with novel single nucleotide polymorphisms of the ESR1 and PPARγ locus in Han Chinese. Am J Clin Nutr 2009;90:255–62.
38. Alexanian A, Sorokin A. Targeting 20-HETE producing enzymes in cancer - rationale, pharmacology, and clinical potential. Onco Targets Ther 2013;6:243–55.
39. Rodriguez-Antona C, Ingelman-Sundberg M. Cytochrome P450 pharmacogenetics and cancer. Oncogene 2006;25:1679–91.
40. Jung YY, Kim HM, Koo JS. Expression of lipid metabolism-related proteins in metastatic breast cancer. PLoS One 2015;10:e0137204.
41. Bhat HK, Calaf G, Hei TK, Loya T, Vadgama JV. Critical role of oxidative stress in estrogen-induced carcinogenesis. Proc Natl Acad Sci U S A 2003;100:3913–8.
42. Purohit A, Newman SP, Reed MJ. The role of cytokines in regulating estrogen synthesis: implications for the etiology of breast cancer. Breast Cancer Res 2002;4:65–9.
43. Zhou SF, Liu JP, Chowbay B. Polymorphism of human cytochrome P450 enzymes and its clinical impact. Drug Metab Rev 2009;41:89–295.
44. Preissner SC, Hoffmann MF, Preissner R, Dunkel M, Gewiess A, Preissner S. Polymorphic cytochrome P450 enzymes (CYPs) and their role in personalized therapy. PLoS One 2013;8:e82562.
45. Navarro-Tito N, Robledo T, Salazar EP. Arachidonic acid promotes FAK activation and migration in MDA-MB-231 breast cancer cells. Exp Cell Res 2008;314:3340–55.
46. Kennett SB, Roberts JD, Olden K. Requirement of protein kinase C micro activation and calpain-mediated proteolysis for arachidonic acid-stimulated adhesion of MDA-MB-435 human mammary carcinoma cells to collagen type IV. J Biol Chem 2004;279:3300–7.
47. Singh-Ranger G, Salhab M, Mokbel K. The role of cyclooxygenase-2 in breast cancer: review. Breast Cancer Res Treat 2008;109:189–98.
48. Wen ZH, Su YC, Lai PL, Zhang Y, Xu YF, Zhao A, et al. Critical role of arachidonic acid-activated mTOR signaling in breast carcinogenesis and angiogenesis. Oncogene 2013;32:160–70.
49. Cleary MP, Grossmann ME. Minireview: obesity and breast cancer: the estrogen connection. Endocrinology 2009;150:2537–42.

Published OnlineFirst November 15, 2016; DOI: 10.1158/0008-5472.CAN-16-1959

Cancer Research
The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association for Cancer Research

Integrative Comparison of mRNA Expression Patterns in Breast Cancers from Caucasian and Asian Americans with Implications for Precision Medicine

Yanxia Shi, Albert Steppi, Ye Cao, et al.

Cancer Res 2017;77:423-433. Published OnlineFirst November 15, 2016.