Differentially Expressed Transcripts and Dysregulated Signaling Pathways and Networks in African American Breast Cancer

Paul A. Stewart[1], Jennifer Luks[1], Mark D. Roycik[1], Qing-Xiang Amy Sang[1,2*], Jinfeng Zhang[3*]

1 Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida, United States of America, **2** Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida, United States of America, **3** Department of Statistics, Florida State University, Tallahassee, Florida, United States of America

Abstract

African Americans (AAs) have higher mortality rate from breast cancer than that of Caucasian Americans (CAs) even when socioeconomic factors are accounted for. To better understand the driving biological factors of this health disparity, we performed a comprehensive differential gene expression analysis, including subtype- and stage-specific analysis, using the breast cancer data in the Cancer Genome Atlas (TCGA). In total, 674 unique genes and other transcripts were found differentially expressed between these two populations. The numbers of differentially expressed genes between AA and CA patients increased in each stage of tumor progression: there were 26 in stage I, 161 in stage II, and 223 in stage III. Resistin, a gene that is linked to obesity, insulin resistance, and breast cancer, was expressed more than four times higher in AA tumors. An uncharacterized, long, non-coding RNA, LOC90784, was down-regulated in AA tumors, and its expression was inversely related to cancer stage and was the lowest in triple negative AA breast tumors. Network analysis showed increased expression of a majority of components in p53 and BRCA1 subnetworks in AA breast tumor samples, and members of the aurora B and polo-like kinase signaling pathways were also highly expressed. Higher gene expression diversity was observed in more advanced stage breast tumors suggesting increased genomic instability during tumor progression. Amplified resistin expression may indicate insulin-resistant type II diabetes and obesity are associated with AA breast cancer. Expression of LOC90784 may have a protective effect on breast cancer patients, and its loss, particularly in triple negative breast cancer, could be having detrimental effects. This work helps elucidate molecular mechanisms of breast cancer health disparity and identifies putative biomarkers and therapeutic targets such as resistin, and the aurora B and polo-like kinase signaling pathways for treating AA breast cancer patients.

Introduction

In 2013, more than 200,000 women in the United States will be diagnosed with breast cancer (BRCa) [1]. Despite advances in treatment and earlier detection, nearly 40,000 women with BRCa will die from this disease [1]. In clinical practice, BRCa is divided into subgroups based on the expression of the estrogen receptor (ER), progesterone receptor (PR), and the status of gene amplification of human epidermal growth factor 2 receptor (HER2). The four predominant subtypes reported extensively in literature are categorized based on 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-). The luminal A and luminal B subtypes express ER or PR and have several treatment options due to their susceptibility to hormone-based adjuvant therapies that target these receptors, but the HER2 subtype, as its name implies, only has HER2 amplification which severely limits adjuvant treatment options [2]. Triple negative breast cancer (TNBC) lacks expression of ER and PR, and HER2 amplification and treatment is primarily limited to surgery and chemotherapy [3]. A subtype of 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 epidermal growth factor receptor) [4]. This phenotype is more likely to undergo metastasis and is associated with poorer prognosis [5].

African American (AA) women have a higher mortality than Caucasian American (CA) women with BRCa [1]. Although socioeconomic status has been shown to be an independent predictor of mortality, the increase in mortality of AA women with BRCa cannot be explained through socioeconomic status alone [6-8]. AA women still suffered from higher mortality in studies and in a randomized clinical trial where AA and CA women had equal access to care [9-11]. These observations suggest a biological basis for the elevated mortality of AA women, and this hypothesis is supported by a study that shows AA women are more likely to develop tumors with worse pathological characteristics (e.g., larger tumor size, less differentiation of cancerous cells) [12]. Other studies have shown that AA breast cancer patients are more likely to have detrimental tumor subtypes such as ER- and PR-, triple negative, and basal-like breast tumors [13-18].

There is no proven cause of health disparity, but studies indicate that genetic differences between AA and CA tumors do exist. Martin et al. first performed genome-wide mRNA expression analysis of tumor epithelium from 18 AA and 17 CA patients [19]. They found over 400 differentially expressed genes and identified a two-gene signature, putative phosphoserine phosphatase-like protein (PSPHL) and beta-crystallin B2 (CRYBB2), which could distinguish between populations. Field et al. conducted a study with 26 pairs of matched AA and CA BRCa patients and showed that molecular profiles differ between AA and CA breast tumors despite being matched by pathological characteristics. They detected more than two dozen differentially expressed genes that are involved in cellular growth, differentiation, invasion, metastasis, and immune response [20]. Additionally, a study by Grunda et al. using 12 pairs of matched AA and CA BRCa patients assessed a set of 84 genes involved in breast carcinoma prognosis, response to therapy, estrogen signaling, and tumor aggressiveness of age- and stage-matched AA and CA tumor samples [21]. They identified 20 differentially expressed genes from this set, leading to their conclusion that gene expression differences may play a role in increased metastatic potential, resistance to therapy, and worse clinical outcome in AA women.

These previous reports are consistent with one exception: ribosomal protein L13 (RPL13) was lower in AA women in the Martin et al. study but higher in the Field et al. study [19,20]. This could be explained by the small sample size or possibly by variation in age, stage, or subtype between the study populations. Moreover, these studies did not analyze differentially expressed genes by stage or subtype, possibly due to lack of patient samples. Since AA women are more likely to be afflicted by worse subtypes it is apparent that comprehensive studies that include stage and subtype analyses as well as larger sample size are necessary to elucidate the genetic factors responsible for BRCa health disparity.

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 AA and CA patient primary tumor samples. We chose NGS data because of its accuracy and the ability to use it to determine differential expression between non-coding RNAs in addition to coding RNAs. This study represents, to our knowledge, the first time that NGS data was used in combination with a large patient cohort to investigate expression differences of genes and other transcripts between AA and CA BRCa patients, and to our knowledge, this work represents the largest and most comprehensive study on health disparity in AA BRCa to date. To our knowledge, we are the first to identify differentially expressed genes by stage and subtype between AA and CA tumors in addition to differentially expressed subnetworks and pathways.

There are several options to test for differential expression. The Poisson distribution and the negative binomial distribution are both discrete probability distributions that can model count data from a gene sequencer over a finite period of time, and many previous methods for differential expression analysis of NGS data have been based on the Poisson distribution [22,23]. However, the Poisson distribution often under-estimates the actual variations of expression data because it assumes the mean is equal to the variance [24,25]. The negative binomial distribution can tolerate instances of large biological variance between replicates and is often used to cope with the over-dispersion issue by allowing mean and variance to be different. With this in mind, we chose to use DESeq, an R Bioconductor package, for our analyses because it uses the negative binomial distribution to derive a test for differential expression [26,27].

Results

Table **1** shows the breakdown of patient demographics obtained from TCGA. AA women had nearly double the mortality of their age- and stage-matched CA counterparts with 18.87% mortality compared to 10.28% CA mortality, and this increase in mortality is despite patient samples being age- and stage-matched. As mentioned in the Introduction, this disparity is likely caused by both socioeconomic and biological factors. Information regarding socioeconomic status or access to care is not available at TCGA. Our observed mortalities are lower than previously reported population data for both races: 18.2% for CA women and 26.8% for AA women [1]. This is likely caused by comparably fewer patients with late stage breast cancer donating their tumor samples to TCGA.

Differential expression analysis at the individual gene level was firstly conducted to identify genes and other transcripts that differ between AA and CA BRCa patients using DESeq; secondly, pathway analysis and gene set enrichment analysis with the Pathway Interaction Database (PID) and Gene Set Enrichment Analysis (GSEA) was performed; thirdly, Gene Expression Network Analysis (GXNA) was used to identify differentially expressed subnetworks; finally, patients were stratified to study differences in stage- or subtype-specific gene expression again with DESeq (Figure **1**).

In total, 674 unique genes and other transcripts were shown to be differentially expressed between the various comparisons

Table 1. Breast cancer patient clinical data from TCGA.

| Characteristics	Caucasian American (n=574)		African American (n=53)		Fisher's Exact Test
	Number	Pct.	Number	Pct.	P-value
Age					0.603
< 50	157	27.35%	18	33.96%	
50-64	252	43.90%	21	39.62%	
65+	165	28.75%	14	26.42%	
Tumor Stage					0.783
1	117	20.38%	11	20.75%	
2	337	58.71%	29	54.72%	
3	120	20.91%	13	24.53%	
Tumor Type					0.003
Luminal A	243	42.33%	16	30.19%	
Luminal B	62	10.80%	0	0.00%	
Triple Negative	67	11.67%	10	18.87%	
HER2 Type	17	2.96%	3	5.66%	
Other/NA	185	32.23%	24	45.28%	
Vital Status					0.066
Living	515	89.72%	43	81.13%	
Deceased	59	10.28%	10	18.87%	

doi: 10.1371/journal.pone.0082460.t001

performed (Table **S1**). The number of unique genes was computed by combining gene names from the overall, stages I-III, luminal A, HER2, and triple negative comparisons into a single list. We then removed duplicate gene names from the list, and this new list without duplicates was then counted to give the total number of unique genes. To identify trends in gene expression, primary breast tumor expression data between AA and CA patients irrespective of subtype or stage (i.e., the "overall" comparison) was first compared. Three hundred forty-two genes and other transcripts, 1.7% of the 20,531 genes and other transcripts assessed, were identified by DESeq as differentially expressed between AA and CA primary breast tumors ($P < 0.001$, Table **S2**). One hundred ten of these genes and other transcripts showed increased expression in AA tumors while 232 were decreased. Thirty-seven of the 342 genes and other transcripts exhibited significant differential expression (i.e., log_2 fold-change greater than 1.0 or less than -1.0), and selected genes and other transcripts with high fold-change and relevance to breast cancer can be seen in Table **2**.

The 342 differentially expressed genes and other transcripts from the overall comparison were searched against PID (Table **3**). PID places genes into known pathways and ranks them by the probability that their associated genes are enriched in the input gene list, and a low P-value indicates a greater chance of enrichment. The top PID pathway was the tumor protein p73 (p73) transcription factor network ($P = 4.75E-05$). p73 is a

tumor suppressor whose overexpression is seen in some aggressive breast cancer tumors and breast cancer cell lines [28]. The next most significant PID result was the aurora B signaling pathway ($P = 1.07E-04$). Aurora kinase B (AURKB) is a mitotic checkpoint kinase that mediates chromosome segregation and is overexpressed in a variety of cancers including lung, colorectal, and prostate [29-31]. The polo-like kinase (PLK1) signaling events pathway was also identified by PID ($P = 2.35E-04$). PLK1 is a kinase significantly overexpressed in TNBC that may be a potential therapeutic target for TNBC [32]. We saw differential expression in the BRCA1 associated RING domain 1 (BARD1) signaling events pathway ($P = 4.70E-04$). BARD1 interacts with tumor protein p53 (p53 or TP53); breast cancer 1, early onset (BRCA1); and breast cancer 2, early onset (BRCA2), all of which are important tumor suppressors in breast cancer [33-35]. The genes within the pathways identified by PID had almost unanimous increase in gene expression in AA tumors. The one exception was BARD1 signaling events pathway which had one gene, RAD50 homolog (S. cerevisiae) (RAD50), decrease in AA tumors. Additionally, we uploaded the genes from the Martin et al., Field et al., and Grunda et al. studies on health disparity to PID and compared the results with our own [19-21]. Out of all the previous works, our results only had the aurora B signaling pathway in common with the Martin et al. data [19].

GXNA was then used to identify differentially expressed subnetworks between AA and CA tumors. 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 two subnetworks, one involving p53 and another involving BRCA1, both with a majority of components upregulated in AA breast tumors ($P < 0.001$, Figure **2**). STRING was used to visualize GXNA data and to show potential gene product interactions beyond what is shown in the GXNA interactions. Gene Set Enrichment Analysis did not identify any differentially expressed gene sets (P-value > 0.05 or with unacceptable FDR; results not shown).

Stage-matched and then subtype-matched AA and CA tumors were next compared, but a lack of stage IV and luminal B AA tumors allowed us to only evaluate stages I through III and the luminal A, HER2, and triple-negative subtypes. Selected results from these comparisons are summarized in Table **4**. Interestingly, very few genes and other transcripts overlapped between the stage comparisons (Figure **3**). Twenty-six genes and other transcripts were differentially expressed in the stage I comparison (19 increased and 7 decreased in AA tumors, Table **S3A**), including BUB1 mitotic checkpoint serine/threonine kinase (BUB1; 1.05 log_2 fold-change; $P = 9.80E-04$) which is a potent prognostic factor for human breast cancer [36]. One hundred sixty-one genes were differentially expressed in the stage II comparison (134 increased and 27 decreased, Table **S3B**) including crystallin, beta B2 (CRYBB2; 2.00 log_2 fold-change; $P = 3.27E-10$) which was differentially expressed in AA breast, colorectal, and prostate cancers [19,20,37,38]. Two hundred twenty-three genes were differentially expressed in the stage III comparison (156 increased and 67 decreased, Table **S3C**), including estrogen receptor 1 (ESR1; -2.25 log_2 fold-change; $P = 8.80E-05$). Forty-
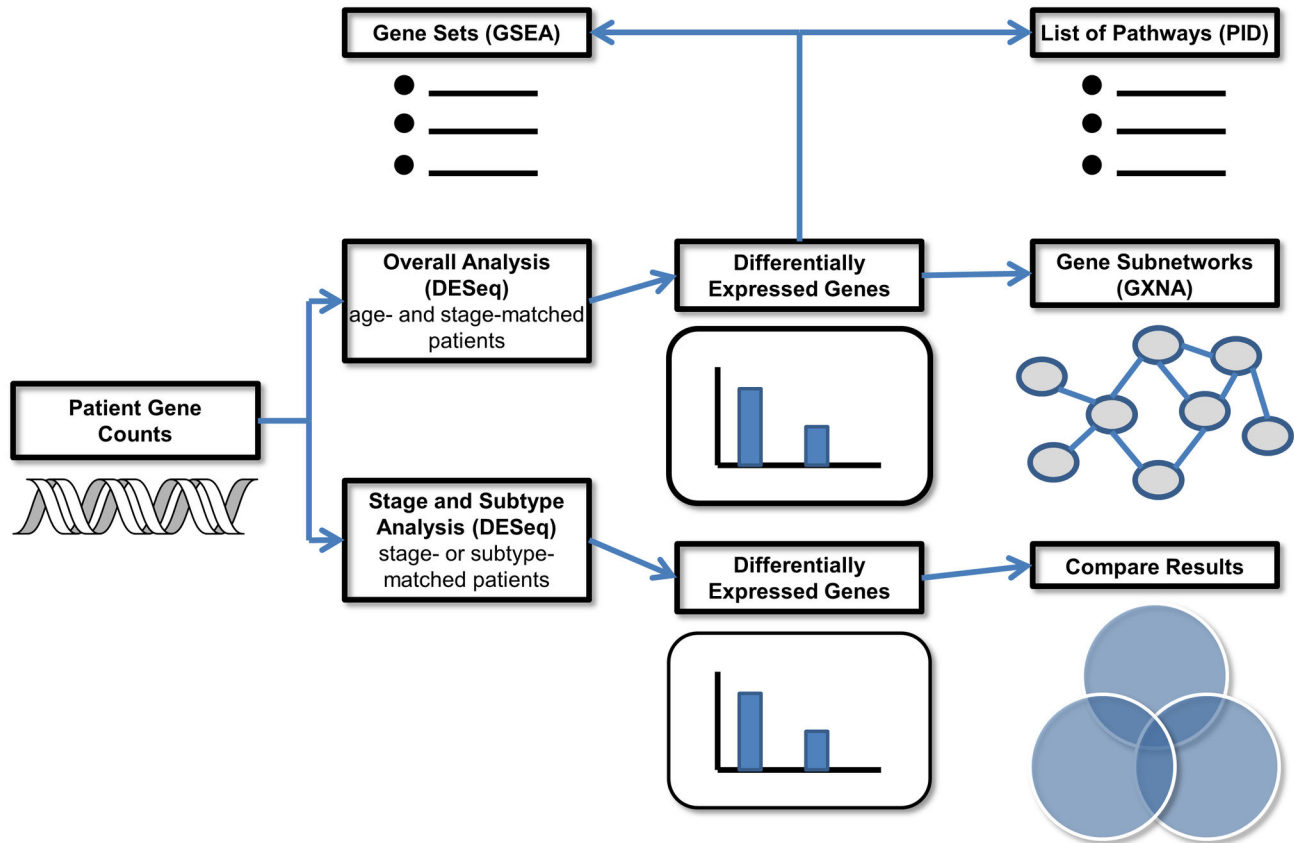
Figure 1. A flowchart summarizing the data analysis methodology.

six genes and other transcripts were differentially expressed in the luminal A comparison (39 increased and 7 decreased in AA tumors, Table **S3D**) including polycystin (PKD) family receptor for egg jelly (PKDREJ; -1.18 $\log_2$ fold-change; P = 9.56E-04) which is point mutated in breast cancer [39]. Twenty-five genes and other transcripts were differentially expressed in the HER2 comparison (21 increased and 4 decreased, Table **S3E**) including TCRgamma alternative reading frame protein (TARP; 8.74 $\log_2$ fold-change; P = 2.31E-04) which is a breast and prostate tumor-associated antigen [40]. Fifteen genes and other transcripts were differentially expressed in the triple negative comparison (10 genes increased and 5 decreased, Table **S3F**) including peptidoglycan recognition protein 1 (PGLYRP1; 2.99 $\log_2$ fold-change; P = 7.64E-04) whose expression may reflect immune cell response to developing breast tumors [41].

LOC90784, a transcript, was the only gene or transcript differentially expressed in more than one subtype (luminal A and triple negative). Additionally, it was differentially expressed in the overall, stage I, stage II, and stage III comparisons. This made it the only transcript or gene we found to be differentially expressed across six different comparisons. Interestingly, the expression of LOC90784 remained consistent across CA tumor stage and grade but was consistently lower in AA samples, and LOC90784 expression was inversely related to stage among

AA patient samples and was the lowest in triple negative AA tumors (Figure **4**).

These results share some agreement with previous studies (Table **5**). Fifteen genes out of 424 were found to be in common with those reported by Martin et al. [19]. An additional two genes were in common with those identified by Martin et al. but the results are contradictory (deoxyuridine triphosphatase or DUT expression was higher in AA samples in our results but lower in the Martin et al. results; Zic family member 1 or ZIC1 was lower in AA samples in our results but higher in the Martin et al. results) [19]. Three out of 27 genes were found to be in common with those found by Field et al. (CRYBB2, mitochondrial ribosomal protein L48 or MRPL48, and zinc finger protein 395 or ZNF395) [20], and one gene out of twenty was shared with those found by Grunda et al. (ESR1) [21]. It is not straightforward to assess the statistical significance of the agreements between these studies and our own so we cannot rule out that the agreements of our studies are purely by chance.

Discussion

We identified 342 differentially expressed genes and other transcripts (P < 0.001) between AA and CA primary breast cancer tumors in the overall comparison. Thirty-seven of these

Table 2. DESeq results showing significant changes of gene expression in age- and stage-matched AA tumors.

Gene	Log$_2$ Fold-Change	P-value	P-adj	Relevance	Citation
RETN	2.25	3.05E-06	2.80E-03	Expressed in inflammatory cells; insulin resistance in mice; increased in BRCa patients; high expression in breast cancer tissue associated with poor survival and malignant characteristics	[42-44]
CRYBB2	1.81	1.24E-13	1.25E-09	Differentially expressed in several studies investigating cancer health disparity	[19,20,37,38]
TREML4	1.79	8.87E-08	1.83E-04	Elicits T-cell immunity and tumor protection when coupled with the breast cancer antigen HER2	[72]
CXCL10	1.48	1.67E-04	2.87E-02	Chemokine; involved in breast tumor invasiveness and progression	[73]
PAX6	1.33	1.89E-05	9.57E-03	Facilitates regulatory roles in breast cancer cell line proliferation and tumor progression	[74]
BMP6	1.25	2.63E-04	3.59E-02	Promotes E-cadherin expression through repressing delta-EF1 in breast cancer cells; abnormally expressed and regulated by estrogen receptor alpha in breast cancer cells	[75,76]
EMR1	1.19	2.51E-04	3.48E-02	Hormone receptor; overexpressed in a cluster of genes associated with poor prognosis in breast cancer	[77]
APOBEC3B	1.13	1.10E-04	2.43E-02	Enzymatic source of mutation in breast cancer	[78]
SPATA18	-1.28	7.66E-06	5.53E-03	p53-inducible protein; controls mitochondrial quality	[79]
ADAMTS15	-1.29	1.82E-04	2.95E-02	Metalloprotease known to inhibit breast cancer cell migration	[49]
ABCC6	-1.37	1.10E-04	2.43E-02	Multidrug resistance-associated protein	[80]

Log$_2$ fold-change is for mean AA expression relative to mean CA expression. Log$_2$-fold change is calculated as log$_2$(AA/CA). Bold entries have been observed as differentially expressed previously in literature. Gene names: ABCC6 - ATP binding cassette, sub family C (CFTR/MRP), member 6; ADAMTS15 - ADAM metallopeptidase with thrombospondin type 1 motif, 15; APOBEC3B - apolipoprotein B mRNA editing enzyme, catalytic polypeptide like 3B; BMP6 - bone morphogenetic protein 6; CRYBB2 - crystallin, beta B2; CXCL10 - chemokine (C X C motif) ligand 10; EMR1 - egf like module containing, mucin like, hormone receptor like 1; PAX6 - paired box 6;RETN - resistin; SPATA18 - spermatogenesis associated 18; TREML4 - triggering receptor expressed on myeloid cells like 4.
doi: 10.1371/journal.pone.0082460.t002
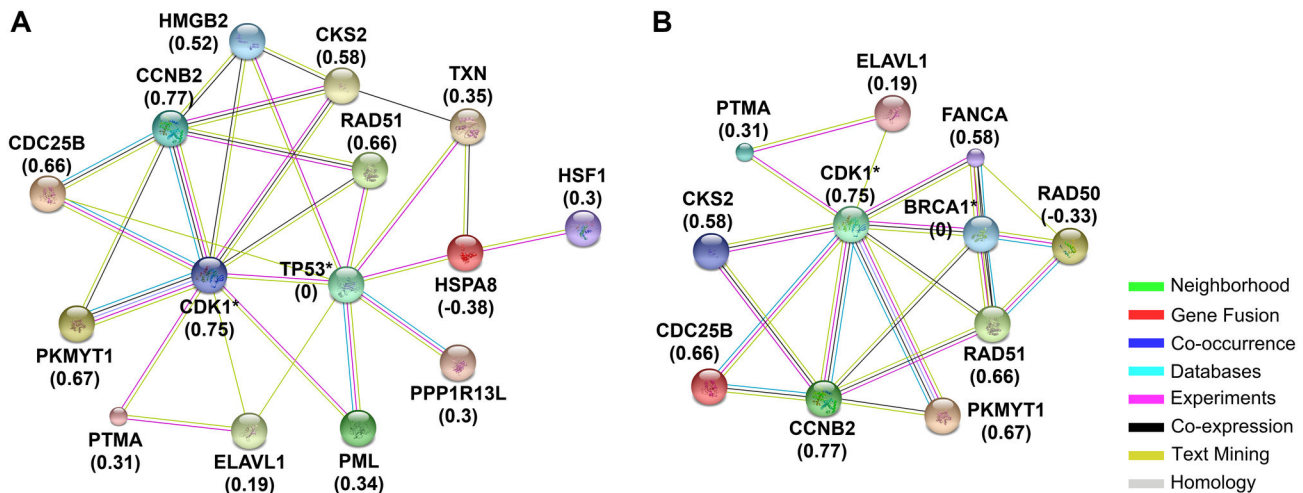
Figure 2. Differentially expressed subnetworks identified by Gene Expression Network Analysis. Subnetworks containing p53 (A) and BRCA1 (B) were differentially expressed in AA tumors. Subnetworks were identified using GXNA and visualized using STRING. Starred results were not differentially expressed but were included in the subnetwork by GXNA. Values in parentheses are the mean fold changes of log$_2$-transformed AA expression relative to CA expression, calculated as log$_2$(CA)/log$_2$(AA). Gene names: HMGB2: BRCA1 - breast cancer 1, early onset; CCNB2 - cyclin B2; CDC25B - cell division cycle 25B; CDK1 - cyclin-dependent kinase 1; CKS2 - CDC28 protein kinase regulatory subunit 2; ELAVL1 - ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R); FANCA - Fanconi anemia, complementation group A; HMGB2 - high mobility group box 2; HSF1 - heat shock transcription factor 1; HSPA8 - heat shock 70kDa protein 8; PKMYT1 - protein kinase, membrane. associated tyrosine/threonine 1; PML - promyelocytic leukemia; PPP1R13L - protein phosphatase 1, regulatory subunit 13 like; PTMA - prothymosin, alpha; RAD50 - RAD50 homolog (S. cerevisiae); RAD51 - RAD51 recombinase; TP53 - tumor protein p53; TXN – thioredoxin.
doi: 10.1371/journal.pone.0082460.g002

Table 4. DESeq results showing significant changes of gene expression in subtype- or stage-matched AA tumors.

Gene	Stage/Subtype	Log$_2$ Fold-Change	P-value	P-adj	Relevance	Citation
TARP	HER2	8.74	2.31E-04	5.01E-01	Breast and prostate tumor-associated antigen	[40]
PKDREJ	Luminal A	-1.18	9.56E-04	4.20E-01	Point mutated in breast cancers	[39]
PGLYRP1	Triple negative	2.99	7.64E-04	1.00E+00	Increased in expression may reflect immune cell response to developing breast tumor	[41]
BUB1	Stage I	1.05	9.80E-04	7.44E-01	Nuclear localization is a potent prognostic factor for human breast cancer	[36]
GALNT6	Stage I	-1.45	3.00E-04	4.03E-01	Deregulation of GALNT6 gene could be an early event during human breast carcinogenesis	[81]
CRYBB2	Stage II	2.00	3.27E-10	6.60E-06	Differentially expressed in several studies investigating cancer health disparity	[19,20,37,38]
FLJ45983	Stage III	-2.40	2.21E-04	5.66E-02	Putative uncharacterized protein; Hypermethylated in primary and metastatic tumors	[82]
ESR1	Stage III	-2.25	8.80E-05	4.91E-02	Estrogen receptor alpha	
FOXA1	Stage III	-1.15	3.48E-05	3.97E-02	Expression correlates with luminal subtype A breast cancer; significant predictor of cancer-specific survival in patients with ER-positive tumors	[83]
LRIG1	Stage III	-1.10	3.39E-04	6.11E-02	Estrogen-regulated growth suppressor; correlates with longer relapse-free survival in ERα-positive breast cancer; tumor suppressor	[84,85]
PLCG2	Stage III	1.00	1.08E-04	4.91E-02	Hypermethylated in metastatic breast cancer cell line	[86]

Log$_2$ fold-change is for mean AA expression relative to mean CA expression. Log$_2$-fold change is calculated as log$_2$(AA/CA). Bold entries have been previously observed as differentially expressed in literature. Gene names: CRYBB2 - crystallin, beta B2; CYP4F8 - cytochrome P450, family 4, subfamily F, polypeptide 8; ESR1 - estrogen receptor 1; FLJ45983 - Not Available; FOXA1 - forkhead box A1; GALNT6 - UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6); LRIG1 - leucine-rich repeats and immunoglobulin-like domains 1; PGLYRP1 - peptidoglycan recognition protein 1; PKDREJ - polycystin (PKD) family receptor for egg jelly; PLCG2 - phospholipase C, gamma 2 (phosphatidylinositol-specific), TARP - TCRgamma alternative reading frame protein.

doi: 10.1371/journal.pone.0082460.t004

Table 3. Differentially expressed genes in AA tumors are part of pathways with relevance to breast cancer.

| Pathway Name	Genes and Transcripts	P-value	P-adj
p73 transcription factor network	ADA, AEN, BUB1, CDK1, HSF1, PML, RAD51	4.75E-05	1.62E-02
Aurora B signaling	AURKB, BUB1, INCENP, KIF20A, KIF23	1.07E-04	1.83E-02
PLK1 signaling events	BUB1, CDC25B, CDK1, INCENP, KIF20A	2.35E-04	2.68E-02
Aurora C signaling	AURKB, INCENP	4.68E-04	3.21E-02
BARD1 signaling events	EWSR1, FANCA, RAD50, RAD51	4.70E-04	3.21E-02

The underlined entry was the only gene not increased in AA samples. Results from the Pathway Interaction Database (http://pid.nci.nih.gov/). Gene names: ADA - adenosine deaminase; AEN - apoptosis enhancing nuclease; AURKB - aurora kinase B; BUB1 - BUB1 mitotic checkpoint serine/threonine kinase; CDC25B - cell division cycle 25B; CDK1 - cyclin-dependent kinase 1; EWSR1 - EWS RNA-binding protein 1; FANCA - Fanconi anemia, complementation group A; HSF1 - heat shock transcription factor 1; INCENP - inner centromere protein antigens 135/155kDa; KIF20A - kinesin family member 20A; KIF23 - kinesin family member 23; PML - promyelocytic leukemia; RAD50 - RAD50 homolog (S. cerevisiae); RAD51 - RAD51 recombinase.

doi: 10.1371/journal.pone.0082460.t003

exhibited significant change (i.e., log$_2$ fold-change > 1.0 or < -1.0), and selected breast cancer-associated genes are shown (Table 2). Several genes have increased expression in AA tumors, including a few with direct ties to breast cancer such as resistin (RETN). RETN is an adipocytokine whose expression

level is significantly increased in AA tumors (2.25 log$_2$ fold-change; P = 3.05E-06). It is expressed in inflammatory cells in humans and plays a role in insulin resistance in mice [42,43]. Breast cancer patients show increased resistin concentrations over control [44], and a recent study indicates that high resistin expression in breast cancer tissue is associated with decreased survival and more malignant characteristics [45]. Its presence at such high expression levels suggests that resistin may play an important role in AA BRCa because of the strong link between obesity and cancer mortality [46-48]. Likewise, a number of genes have reduced expression in AA women, including ADAM metallopeptidase with thrombospondin type 1 motif, 15 (ADAMTS15; -1.29 log$_2$ fold-change; P = 1.82E-04). ADAMTS15 is a metalloprotease known to inhibit breast cancer cell migration [49], and it is possible that the reduced expression of this gene could have consequences in AA BRCa progression.

We noticed an increase in expression of CRYBB2 in AA samples (Table 2). This gene encodes a protein that is a major component of the vertebrate eye lens and whose mutation may cause cataracts [50]. Expression of this gene was increased in AA tumor samples [19], and its expression was increased in both nonmalignant and tumor samples [20]. Martin et al. used this gene to differentiate between AA and CA patients while Field et al. hypothesized that overexpression of this gene could possibly enhance AA tumorigenesis. Expression of CRYBB2 was also increased in a study comparing AA and CA colorectal cancer and in another study comparing AA and CA prostate cancer [37,38]. These last two results are intriguing since health disparity exists in both prostate cancer and colorectal cancer in the AA population. The fact that this gene is overexpressed in nonmalignant tissues could simply be a result
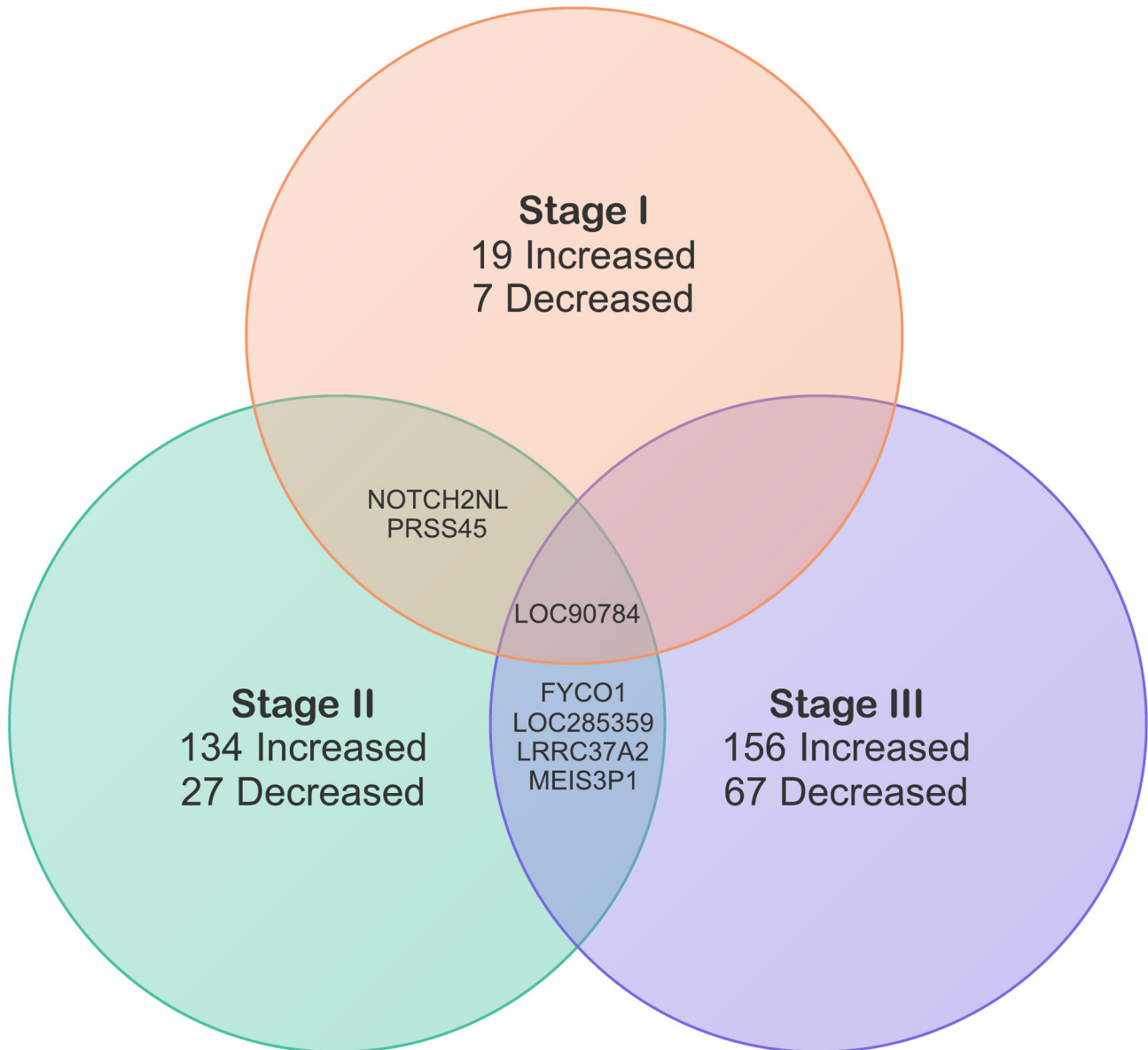
Figure 3. Venn diagram depicting overlap of differentially expressed genes and other transcripts between stage-matched African- and Caucasian-American tumors. Increases or decreases indicate AA gene expression and are relative to CA expression. Gene names: FYCO1 - FYVE and coiled-coil domain containing 1; LOC90784 - Not Available; LOC285359 - Not Available; LRRC37A2 - leucine rich repeat containing 37, member A2; MEIS3P1 - Meis homeobox 3 pseudogene 1; NOTCH2NL - notch 2 N-terminal like; PRSS45 - protease, serine, 45.

doi: 10.1371/journal.pone.0082460.g003

of racial differences, but we only observed differential expression of CRYBB2 in our overall and stage II comparisons. We would expect to see CRYBB2 in more comparisons if this gene was always differentially expressed between races. We suggest further studies into the role, if any, which CRYBB2 plays in cancer and health disparity.

We saw elevated expression of all members of the aurora B signaling pathway identified by PID (P = 1.07E-04). Aurora B kinase (AURKB), a member of the pathway sharing its name, had nearly doubled expression in AA compared to CA (0.89 $\log_2$ fold-change; P = 3.49E-05; Table **S2**). AURKB is a mitotic checkpoint kinase that mediates chromosome segregation and is overexpressed in a variety of cancers including lung, colorectal, and prostate [29-31], and it interacts with the aforementioned BARD1 pathway [35]. There has been no evidence that AURKB is overexpressed in breast cancer, but aurora kinase family inhibitors have been suggested as treatments for cancer. Breast cancer cell lines treated with the
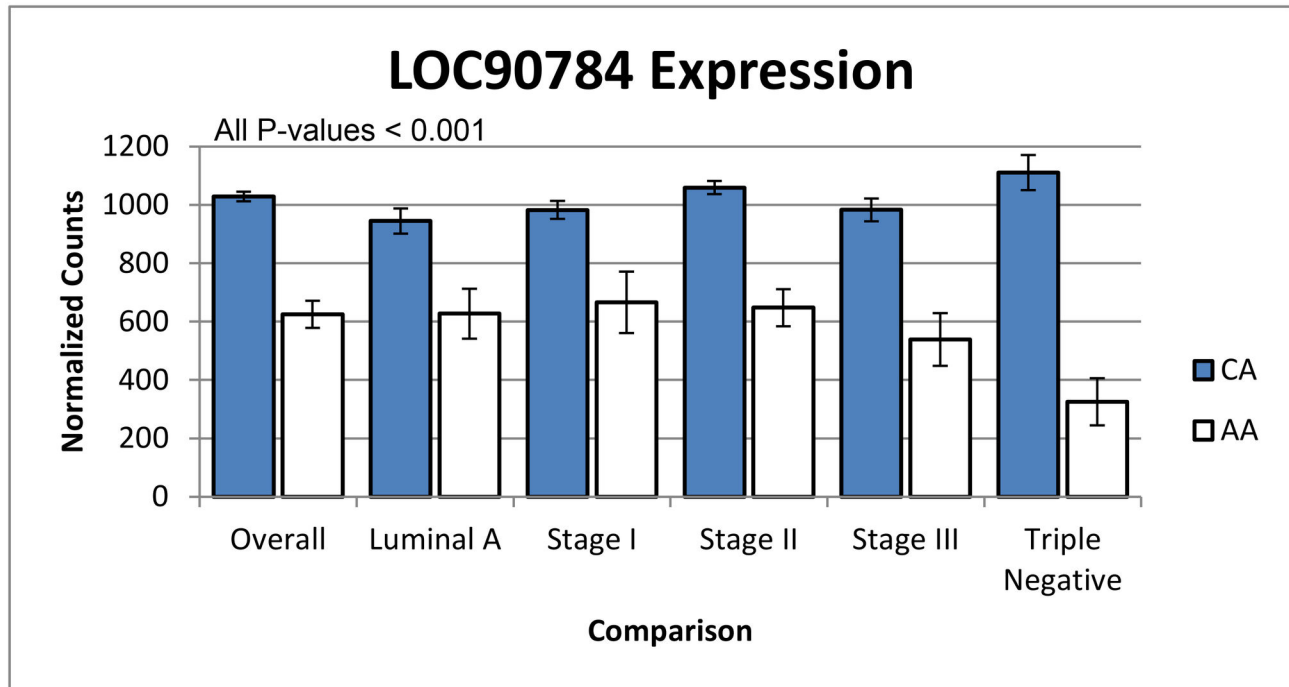
Figure 4. LOC90784, a long non-coding RNA, was differentially expressed across comparisons of African- and Caucasian-American tumors. The expression of this transcript was consistent in CA tumors, but its expression was inversely related to stage and was lowest in AA with TNBC. P-values for the comparisons were overall: 1.46E-14, luminal A: 1.98E-04, stage I: 6.61E-04, stage II: 3.63E-08, stage III: 2.57E-05, and triple negative: 1.86E-09 (see Supplementary Tables). Error bars are standard error.

doi: 10.1371/journal.pone.0082460.g004

highly selective AURKB inhibitor Barasertib show antineoplastic effects [51-53]. Based on the increased expression of AURKB and other aurora B signaling pathway members in AA tumors, we speculate that this drug may be used as an effective treatment for African American breast cancer patients.

We used GXNA to identify differentially expressed subnetworks in AA BRCa. Subnetworks differ from pathway analysis results because they consist of genes that share interactions with one-another across multiple pathways, and this allows for the identification of differentially expressed subnetworks that are not limited to a single pathway. We identified two differentially expressed subnetworks related to BRCa with GXNA, one centered on p53 (P = 0.001) while the other involved BRCA1 (P = 0.001, Figure **2**). Both of these pathways had a majority of genes with increased expression in AA samples. p53 is well-studied tumor suppressor that can activate DNA repair as well as initiate apoptosis [54], and p53 status is also an independent predictor of survival in AA women with BRCa [55]. Although p53 was not differentially expressed in our results, many of its signaling neighbors were increased. BRCA1 is an important DNA repair enzyme whose loss is associated with dramatically increased BRCa incidence [56], and BRCA1 and BRCA2 mutations are seen in high-risk AA women with BRCa [57]. Like p53, BRCA1 was not differentially expressed in our results but many of its signaling neighbors were increased. The GXNA results are strengthened by our pathway analysis result that showed increased activity in the

BARD1 pathway whose members interact with both p53 and BRCA1 [33-35]. These results strongly suggest increased activity in the p53 and BRCA1 subnetworks in AA women, but these results are somewhat contradictory: higher expression or activity in two pathways related to tumor suppression and DNA repair should logically have a beneficial effect and should not contribute to health disparity. However, we suggest that there may be dysregulation up or downstream of these pathways as a result of a mutation, and the increase in activity of these pathways could be explained by the patient's cells attempting to compensate for dysregulation. p53-expressing tumors are more frequent in AA women with TNBC when compared to their CA counterparts with TNBC [58], so the increase in expression of the p53 subnetwork could be partially explained by this fact.

Our stage III comparison showed that AA tumors had significantly lower ESR1 (-2.25 $\log_2$ fold-change; P = 8.80E-05; Table **4**). This comparison was our only comparison where ESR1 was differentially expressed between AA and CA samples, and Grunda et al. saw a similar decrease in expression of ESR1 in their study [21]. We believe this finding suggests that AA women may have reduced ER expression in later stages or that AA women may be more susceptible to mutations causing the ER- phenotype.

LOC90784, a transcript, was the only gene or transcript differentially expressed in the overall, stage I, stage II, stage III, luminal A, and triple negative comparisons. The expression of

Table 5. Comparison of increases and decreases in AA gene expression between presented results and previous findings.

Gene	Results (674 Genes)	Martin (17 of 424 Genes)	Field (3 of 27 Genes)	Grunda (1 of 20 Genes)
AK2	increase	increase		
APP	increase	increase		
BUB1	increase	increase		
CKS2	increase	increase		
CRYBB2	increase	increase	increase	
CXCL10	increase	increase		
DNAJC15	increase	increase		
DUT	increase	decrease		
ESPL1	increase	increase		
ESR1	decrease			decrease
HLA-DQB1	increase	increase		
ISG20	increase	increase		
KIF20A	increase	increase		
MRPL48	increase		increase	
SNED1	decrease	decrease		
TAP2	increase	increase		
UBE2C	increase	increase		
ZIC1	decrease	increase		
ZNF395	decrease	decrease	decrease	

Italicized genes were inconsistent with literature while blank entries were not observed. Comparisons are an increase in decrease in expression of AA samples relative to CA samples. The number of genes in parentheses refers to the number of genes in common with the presented results compared to the total number of differentially expressed genes from each study. Gene names: AK2 - adenylate kinase 2; APP - amyloid beta (A4) precursor protein; BUB1 - BUB1 mitotic checkpoint serine/threonine kinase; CKS2 - CDC28 protein kinase regulatory subunit 2; CRYBB2 - crystallin, beta B2; CXCL10 - chemokine (C-X-C motif) ligand 10; DNAJC15 - DnaJ (Hsp40) homolog, subfamily C, member 15; DUT - deoxyuridine triphosphatase; ESPL1 - extra spindle pole bodies homolog 1 (S. cerevisiae); ESR1 - estrogen receptor 1; HLA-DQB1 - major histocompatibility complex, class II, DQ beta 1; ISG20 - interferon stimulated exonuclease gene 20kDa; KIF20A - kinesin family member 20A; MRPL48 - mitochondrial ribosomal protein L48; SNED1 - sushi, nidogen and EGF-like domains 1; TAP2 - transporter 2, ATP-binding cassette, sub-family B (MDR/TAP); UBE2C - ubiquitin-conjugating enzyme E2C; ZIC1 - Zic family member 1; ZNF395 - zinc finger protein 395.

doi: 10.1371/journal.pone.0082460.t005

LOC90784 was very similar across the CA population regardless of stage or subtype, but its expression was inversely related to stage among AA samples and was lowest in AA women with TNBC (Figure **4**). Little is known about this transcript other than it is found on chromosome 2 in humans (2p11.2), is 3,653 bases long, shares 100% sequence with polymerase (RNA) I polypeptide A, 194kDa (POLR1A), and produces long, non-coding RNA (lncRNA) of undetermined function (NCBI Gene ID: 90784, updated on 16-Aug-2013). Non-coding RNA do not encode for protein, yet are alternatively spliced and processed into smaller products [59]. A search of miRBase (accessed March 13th, 2013) revealed that LOC90784 is not processed into any known microRNA [60]. However, lncRNAs are expressed in a disease-specific manner that makes them potential therapeutic targets since they are implicated in BRCa and promoting metastasis [61-63]. Another possibility is that LOC90784 is a miscategorized, alternatively spliced variant of POLR1A, but we have no evidence of this and further experiments are needed.

One emerging trend was that 26 genes and other transcripts were identified as differentially expressed in stage I, then 161 in stage II, and then 223 in stage III. This suggests that stage I tumors are more genetically similar to one another and that similarity is lost in higher stages. This increase in differentially expressed genes may be due to the increased rate of changes over time in a dysfunctional, cancerous cell due to genomic instability [64], but this finding may also suggest that AA women with BRCa have a higher rate of mutation than their CA counterparts. It would have been interesting to observe if this trend was maintained in stage IV tumors, but no stage IV AA data was available for analysis at the time of writing.

This study further provides molecular insights on breast cancer health disparity and identifies several genes, other transcripts, and pathway targets for additional investigation. We observed increases in differentially expressed genes in later stages of cancer, suggesting higher genomic instability in African American patients. Several genes were differentially expressed including resistin, an adipocytokine involved in insulin resistance, obesity, and breast cancer, which had over four times higher expression in AA patients. A long, non-coding RNA, LOC90784, which may have protective activity in breast cancer, was significantly reduced in AA patients and was the lowest in TNBC. There was also higher expression in a number of signaling pathways related to cancer, such as aurora B, BRCA1, and p53, which suggests up- or down-stream dysregulation within these signal transduction pathways.

Methods

RNA-Seq Data from TCGA

Processed, de-identified, patient BRCa data (TCGA "Level 3" data designation) was downloaded from TCGA. The de-identified patient data included up to 20531 RNA sequence-derived gene expression values and clinical characteristics (e.g., age, cancer stage, receptor status). Custom Python scripts (The Python Software Foundation, http://www.python.org) were used to sort and aggregate patient data for analysis. Fifty-three AA tumor samples were matched by

age and stage with 574 CA tumor samples, and a summary of the clinical characteristics of these patients is presented in Table **1**. The mean ages were 58 years for CA patients and 55 years for AA patients. Subtypes were assigned to patient samples according to receptor status from TCGA clinical data: luminal A (ER+ and/or PR+, HER2-), luminal B (ER+ and/or PR +, HER2+), HER2 (ER-, PR-, HER2+), and triple negative (ER-, PR-, HER2-). Normalized patient gene expression counts were used unless otherwise noted.

Differential Gene Expression

DESeq version 1.10.1 was used to test for differentially expressed genes from patient data using the R statistical programming environment (version 2.15.2) [26,27]. DESeq uses the negative binomial distribution to derive a test for differential expression. The DESeq function "estimateDispersions()" was utilized with the argument sharingMode="gene-est-only" due to the relatively large number of patient samples, except for the HER2 comparison. Patient gene counts input into DESeq were rounded to the nearest whole number but not normalized. Jobs were run on the Florida State University High Performance Computing Cluster.

Pathway and Network Analysis

Differentially expressed genes identified by DESeq were input into PID (accessed February 11th, 2013) batch query tool and searched against the NCI-Nature curated data set [65]. PID identified pathways and constituent genes and assigned P-values to each pathway. The parameters used to generate P-values were the size of the query list, the number of molecules in a given pathway, and the number of molecules in the entire database. The P-value calculation for the batch query tool is based on a hypergeometric cumulative distribution function and was not corrected for multiple-hypothesis testing [66]. PID contains 137 pathways at the time of writing.

Gene Expression Network Analysis (GXNA) was used to identify subnetworks where differentially expressed genes are enriched compared to other parts of the whole gene/protein interaction network [67]. 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 [68]. STRING takes a single gene product or list of gene products and returns a diagram of known and potential interactions.

Gene Set Enrichment Analysis (GSEA) was used to identify differentially expressed gene sets including pathways [69,70]. Gene sets consist of groups of genes that may share common biological function, chromosomal location, or regulation. GSEA differs from network analysis in that it does not utilize topological or gene interaction information.

False Discovery Rates (FDR)

A cutoff of P < 0.001 was used for reported results to reduce the number of false positives. If possible, the Benjamini-Hochberg procedure was used to control for FDR and is reported in the results [71]. Benjamini-Hochberg is used in DESeq output by default. GXNA adjusts for FDR using the more stringent family-wise error rate method. PID did not give a means to adjust for FDR, so Benjamini-Hochberg was again used to adjust the P-values. P-values that have been adjusted for FDR are referred to as adjusted P-values or P_{adj}.

Supporting Information

Table S1. **Log$_2$ fold-change of differentially expressed genes and other transcripts across all comparisons.**
(XLSX)

Table S2. **Differentially expressed genes and other transcripts identified from the overall comparison using DESeq.**
(XLSX)

Table S3. A. Differentially expressed genes and other transcripts identified in stage I tumors with DESeq. B. Differentially expressed genes and other transcripts identified in stage II tumors with DESeq. C. Differentially expressed genes and other transcripts identified in stage III tumors with DESeq. D. Differentially expressed genes and other transcripts identified in luminal A tumors with DESeq. E. Differentially expressed genes and other transcripts identified in HER2 tumors with DESeq. F. Differentially expressed genes and other transcripts identified in triple negative tumors with DESeq.
(XLSX)

Acknowledgements

Author Contributions

Conceived and designed the experiments: PAS JL MDR QXS JZ. Performed the experiments: PAS JL MDR QXS JZ. Analyzed the data: PAS JL MDR QXS JZ. Contributed reagents/materials/analysis tools: PAS JL MDR QXS JZ. Wrote the manuscript: PAS JL MDR QXS JZ.

References

1. (2013) American Cancer Society. Cancer Facts & Figures 2013. Atlanta: American Cancer Society.

2. Davoli A, Hocevar BA, Brown TL (2010) Progression and treatment of HER2-positive breast cancer. Cancer Chemother Pharmacol 65: 611-623. doi:10.1007/s00280-009-1208-1. PubMed: 20087739.

3. Carey LA (2011) Directed therapy of subtypes of triple-negative breast cancer. Oncologist 16 Suppl 1: 71-78. doi:10.1634/theoncologist. 2010-0227. PubMed: 2127844321138955.

4. Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR et al. (2009) Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. Clin Cancer Res 15: 2302-2310. doi: 10.1158/1078-0432.CCR-08-2132. PubMed: 19318481.

5. Rakha EA, Reis-Filho JS, Ellis IO (2008) Basal-like breast cancer: a critical review. J Clin Oncol 26: 2568-2581. doi:10.1200/JCO. 2007.13.1748. PubMed: 18487574.

6. Newman LA, Griffith KA, Jatoi I, Simon MS, Crowe JP et al. (2006) Meta-analysis of survival in African American and white American patients with breast cancer: ethnicity compared with socioeconomic status. J Clin Oncol 24: 1342-1349. doi:10.1200/JCO.2005.03.3472. PubMed: 16549828.

7. Newman LA, Mason J, Cote D, Vin Y, Carolin K et al. (2002) African-American ethnicity, socioeconomic status, and breast cancer survival: a meta-analysis of 14 studies involving over 10,000 African-American and 40,000 White American patients with carcinoma of the breast. Cancer 94: 2844-2854. doi:10.1002/cncr.10575. PubMed: 12115371.

8. Fiscella K, Franks P (1997) Poverty or income inequality as predictor of mortality: longitudinal cohort study. BMJ 314: 1724-1727. doi:10.1136/ bmj.314.7096.1724. PubMed: 9185498.

9. Jatoi I, Becher H, Leake CR (2003) Widening disparity in survival between white and African-American patients with breast carcinoma treated in the U. S. Department of Defense healthcare system. Cancer 98: 894-899. doi:10.1002/cncr.11604. PubMed: 12942554.

10. Wojcik BE, Spinks MK, Optenberg SA (1998) Breast carcinoma survival analysis for African American and white women in an equal-access health care system. Cancer 82: 1310-1318. doi:10.1002/ (SICI)1097-0142(19980401)82:7. PubMed: 9529023.

11. Albain KS, Unger JM, Crowley JJ, Coltman CA, Hershman DL (2009) Racial Disparities in Cancer Survival Among Randomized Clinical Trials Patients of the Southwest Oncology. Group - Journal of the National Cancer Institute 101: 984-992. doi:10.1093/jnci/djp175.

12. Elledge RM, Clark GM, Chamness GC, Osborne CK (1994) Tumor biologic factors and breast cancer prognosis among white, Hispanic, and black women in the United States. J Natl Cancer Inst 86: 705-712. doi:10.1093/jnci/86.9.705. PubMed: 7908990.

13. Amend K, Hicks D, Ambrosone CB (2006) Breast cancer in African-American women: differences in tumor biology from European-American women. Cancer Res 66: 8327-8330. doi: 10.1158/0008-5472.CAN-06-1927. PubMed: 16951137.

14. Gapstur SM, Dupuis J, Gann P, Collila S, Winchester DP (1996) Hormone receptor status of breast tumors in black, Hispanic, and non-Hispanic white women. An analysis of 13,239 cases. Cancer 77: 1465-1471. doi:10.1002/(SICI)1097-0142(19960415)77:8. PubMed: 8608530.

15. Stead LA, Lash TL, Sobieraj JE, Chi DD, Westrup JL et al. (2009) Triple-negative breast cancers are increased in black women regardless of age or body mass index. Breast Cancer Res 11: R18. doi: 10.1186/bcr2279. PubMed: 19320967.

16. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V (2007) Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer 109: 1721-1728. doi:10.1002/cncr. 22618. PubMed: 17387718.

17. Harris LN, Broadwater G, Lin NU, Miron A, Schnitt SJ et al. (2006) Molecular subtypes of breast cancer in relation to paclitaxel response and outcomes in women with metastatic disease: results from CALGB 9342. Breast Cancer Res 8: R66. doi:10.1186/bcr1622. PubMed: 17129383.

18. Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y et al. (2007) Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. Cancer 110: 876-884. doi: 10.1002/cncr.22836. PubMed: 17620276.

19. Martin DN, Boersma BJ, Yi M, Reimers M, Howe TM et al. (2009) Differences in the tumor microenvironment between African-American and European-American breast cancer patients. PLOS ONE 4: e4531. doi:10.1371/journal.pone.0004531. PubMed: 19225562.

20. Field LA, Love B, Deyarmin B, Hooke JA, Shriver CD et al. (2012) Identification of differentially expressed genes in breast tumors from African American compared with Caucasian women. Cancer 118: 1334-1344. doi:10.1002/cncr.26405. PubMed: 21800289.

21. Grunda JM, Steg AD, He Q, Steciuk MR, Byan-Parker S et al. (2012) Differential expression of breast cancer-associated genes between stage- and age-matched tumor specimens from African- and Caucasian-American Women diagnosed with breast cancer. BMC Res Notes 5: 248. doi:10.1186/1756-0500-5-248. PubMed: 22616718.

22. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18: 1509-1517. doi:10.1101/gr. 079558.108. PubMed: 18550803.

23. Wang L, Feng Z, Wang X, Wang X (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26: 136-138. doi:10.1093/bioinformatics/btp612. PubMed: 19855105.

24. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D et al. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320: 1344-1349. doi:10.1126/science.1158441. PubMed: 18451266.

25. Robinson MD, Smyth GK (2007) Moderated statistical tests for assessing differences in tag abundance. Bioinformatics 23: 2881-2887. doi:10.1093/bioinformatics/btm453. PubMed: 17881408.

26. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106. doi:10.1186/ gb-2010-11-10-r106. PubMed: 20979621.

27. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80. doi:10.1186/ gb-2004-5-10-r80. PubMed: 15461798.

28. Zaika AI, Kovalev S, Marchenko ND, Moll UM (1999) Overexpression of the wild type p73 gene in breast cancer tissues and cell lines. Cancer Res 59: 3257-3263. PubMed: 10397274.

29. Takeshita M, Koga T, Takayama K, Ijichi K, Yano T et al. (2013) Aurora-B overexpression is correlated with aneuploidy and poor prognosis in non-small cell lung cancer. Lung Cancer 80: 85-90. doi: 10.1016/j.lungcan.2012.12.018. PubMed: 23313006.

30. Tatsuka M, Katayama H, Ota T, Tanaka T, Odashima S et al. (1998) Multinuclearity and increased ploidy caused by overexpression of the aurora- and Ipl1-like midbody-associated protein mitotic kinase in human cancer cells. Cancer Res 58: 4811-4816. PubMed: 9809983.

31. Ditchfield C, Johnson VL, Tighe A, Ellston R, Haworth C et al. (2003) Aurora B couples chromosome alignment with anaphase by targeting BubR1, Mad2, and Cenp-E to kinetochores. J Cell Biol 161: 267-280. doi:10.1083/jcb.200208091. PubMed: 12719470.

32. Maire V, Némati F, Richardson M, Vincent-Salomon A, Tesson B et al. (2013) Polo-like kinase 1: a potential therapeutic option in combination with conventional chemotherapy for the management of patients with triple-negative breast cancer. Cancer Res 73: 813-823. doi: 10.1158/0008-5472.CAN-12-2633. PubMed: 23144294.

33. Wu LC, Wang ZW, Tsan JT, Spillman MA, Phung A et al. (1996) Identification of a RING protein that can interact in vivo with the BRCA1 gene product. Nat Genet 14: 430-440. doi:10.1038/ng1296-430. PubMed: 8944023.

34. Dong Y, Hakimi MA, Chen X, Kumaraswamy E, Cooch NS et al. (2003) Regulation of BRCC, a holoenzyme complex containing BRCA1 and BRCA2, by a signalosome-like subunit and its role in DNA repair. Mol Cell 12: 1087-1099. doi:10.1016/S1097-2765(03)00424-6. PubMed: 14636569.

35. Ryser S, Dizin E, Jefford CE, Delaval B, Gagos S et al. (2009) Distinct roles of BARD1 isoforms in mitosis: full-length BARD1 mediates Aurora B degradation, cancer-associated BARD1beta scaffolds Aurora B and BRCA2. Cancer Res 69: 1125-1134. doi: 10.1158/0008-5472.SABCS-1125. PubMed: 19176389.

36. Takagi K, Miki Y, Shibahara Y, Nakamura Y, Ebata A et al. (2013) BUB1 immunolocalization in breast carcinoma: its nuclear localization as a potent prognostic factor of the patients. Horm Cancer 4: 92-102. doi:10.1007/s12672-012-0130-x. PubMed: 23288590.

37. Jovov B, Araujo-Perez F, Sigel CS, Stratford JK, McCoy AN et al. (2012) Differential gene expression between African American and European American colorectal cancer patients. PLOS ONE 7: e30168. doi:10.1371/journal.pone.0030168. PubMed: 22276153.

38. Wallace TA, Prueitt RL, Yi M, Howe TM, Gillespie JW et al. (2008) Tumor immunobiological differences in prostate cancer between

African-American and European-American men. Cancer Res 68: 927-936. doi:10.1158/0008-5472.CAN-07-2608. PubMed: 18245496.

39. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J et al. (2006) The consensus coding sequences of human breast and colorectal cancers. Science 314: 268-274. doi:10.1126/science.1133427. PubMed: 16959974.

40. Wolfgang CD, Essand M, Vincent JJ, Lee B, Pastan I (2000) TARP: a nuclear protein expressed in prostate and breast cancer cells derived from an alternate reading frame of the T cell receptor gamma chain locus. Proc Natl Acad Sci U S A 97: 9437-9442. doi:10.1073/pnas. 160270597. PubMed: 10931945.

41. Pitteri SJ, Kelly-Spratt KS, Gurley KE, Kennedy J, Buson TB et al. (2011) Tumor microenvironment-derived proteins dominate the plasma proteome response during breast cancer induction and progression. Cancer Res 71: 5090-5100. doi:10.1158/1538-7445.AM2011-5090. PubMed: 21653680.

42. Patel L, Buckels AC, Kinghorn IJ, Murdock PR, Holbrook JD et al. (2003) Resistin is expressed in human macrophages and directly regulated by PPAR gamma activators. Biochem Biophys Res Commun 300: 472-476. doi:10.1016/S0006-291X(02)02841-3. PubMed: 12504108.

43. Steppan CM, Bailey ST, Bhat S, Brown EJ, Banerjee RR et al. (2001) The hormone resistin links obesity to diabetes. Nature 409: 307-312. doi:10.1038/35053000. PubMed: 11201732.

44. Sun CA, Wu MH, Chu CH, Chou YC, Hsu GC et al. (2010) Adipocytokine resistin and breast cancer risk. Breast Cancer Res Treat 123: 869-876. doi:10.1007/s10549-010-0792-4. PubMed: 20177966.

45. Lee YC, Chen YJ, Wu CC, Lo S, Hou MF et al. (2012) Resistin expression in breast cancer tissue as a marker of prognosis and hormone therapy stratification. Gynecol Oncol 125: 742-750. doi:10.1016/j.ygyno.2012.02.032. PubMed: 22370603.

46. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ (2003) Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med 348: 1625-1638. doi:10.1056/NEJMoa021423. PubMed: 12711737.

47. Calle EE, Kaaks R (2004) Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer 4: 579-591. doi:10.1038/nrc1408. PubMed: 15286738.

48. Zhang C, Rexrode KM, van Dam RM, Li TY, Hu FB (2008) Abdominal obesity and the risk of all-cause, cardiovascular, and cancer mortality: sixteen years of follow-up in US women. Circulation 117: 1658-1667. doi:10.1161/CIRCULATIONAHA.107.739714. PubMed: 18362231.

49. Wagstaff L, Kelwick R, Decock J, Arnold H, Pennington C et al. (2010) ADAMTS15 metalloproteinase inhibits breast cancer cell migration. Breast Cancer Research 12: 15. doi:10.1186/bcr2482.

50. Graw J (2009) Genetics of crystallins: cataract and beyond. Exp Eye Res 88: 173-189. doi:10.1016/j.exer.2008.10.011. PubMed: 19007775.

51. Gully CP, Zhang F, Chen J, Yeung JA, Velazquez-Torres G et al. (2010) Antineoplastic effects of an Aurora B kinase inhibitor in breast cancer. Mol Cancer 9: 42. doi:10.1186/1476-4598-9-42. PubMed: 20175926.

52. Gautschi O, Heighway J, Mack PC, Purnell PR, Lara PN et al. (2008) Aurora kinases as anticancer drug targets. Clin Cancer Res 14: 1639-1648. doi:10.1158/1078-0432.CCR-07-2179. PubMed: 18347165.

53. Keen N, Taylor S (2004) Aurora-kinase inhibitors as anticancer agents. Nat Rev Cancer 4: 927-936. doi:10.1038/nrc1502. PubMed: 15573114.

54. Vousden KH, Lane DP (2007) p53 in health and disease. Nat Rev Mol Cell Biol 8: 275-283. doi:10.1038/nrm2147. PubMed: 17380161.

55. Dookeran KA, Dignam JJ, Ferrer K, Sekosan M, McCaskill-Stevens W et al. (2010) p53 as a marker of prognosis in African-American women with breast cancer. Ann Surg Oncol 17: 1398-1405. doi:10.1245/s10434-009-0889-3. PubMed: 20049641.

56. Calderon-Margalit R, Paltiel O (2004) Prevention of breast cancer in women who carry BRCA1 or BRCA2 mutations: a critical review of the literature. Int J Cancer 112: 357-364. doi:10.1002/ijc.20429. PubMed: 15382059.

57. Pal T, Vadaparampil S, Betts J, Miree C, Li S et al. (2008) BRCA1/2 in high-risk African American women with breast cancer: providing genetic testing through various recruitment strategies. Genet Test 12: 401-407. doi:10.1089/gte.2007.0108. PubMed: 18752448.

58. Davion S, Sullivan M, Rohan S, Siziopikou KP (2012) p53 Expression in Triple Negative Breast Carcinomas: Evidence of Age-Related and Racial Differences. Journal Cancer Therapy 3: 649-654. doi:10.4236/jct.2012.325084.

59. Mattick JS, Makunin IV (2006) Non-coding RNA. Human molecular genetics 15 Spec No 1: R17-29

60. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39: D152-D157. doi:10.1093/nar/gkr817. PubMed: 21037258.

61. Prensner JR, Chinnaiyan AM (2011) The emergence of lncRNAs in cancer biology. Cancer Discov 1: 391-407. Available online at: doi:10.1158/2159-8290.CD-11-0209 PubMed: 22096659.

62. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM et al. (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464: 1071-1076. doi:10.1038/nature08975. PubMed: 20393566.

63. Gutschner T, Diederichs S (2012) The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol 9: 703-719. doi:10.4161/rna.20481. PubMed: 22664915.

64. Negrini S, Gorgoulis VG, Halazonetis TD (2010) Genomic instability--an evolving hallmark of cancer. Nat Rev Mol Cell Biol 11: 220-228. doi:10.1038/nrm2858. PubMed: 20177397.

65. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M et al. (2009) PID: the Pathway Interaction. Database - Nucleic Acids Research 37: D674-D679. doi:10.1093/nar/gkn653.

66. Wang X, Stewart PA, Cao Q, Sang QX, Chung LW et al. (2011) Characterization of the phosphoproteome in androgen-repressed human prostate cancer cells by Fourier transform ion cyclotron resonance mass spectrometry. J Proteome Res 10: 3920-3928. doi:10.1021/pr2000144. PubMed: 21786837.

67. Nacu S, Critchley-Thorne R, Lee P, Holmes S (2007) Gene expression network analysis and applications to immunology. Bioinformatics 23: 850-858. doi:10.1093/bioinformatics/btm019. PubMed: 17267429.

68. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M et al. (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41: D808-D815. doi:10.1093/nar/gks1094. PubMed: 23203871.

69. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102: 15545-15550. doi:10.1073/pnas.0506580102. PubMed: 16199517.

70. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP (2007) GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics 23: 3251-3253. doi:10.1093/bioinformatics/btm369. PubMed: 17644558.

71. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B:_Statistical Methodology 57: 289-300.

72. Hemmi H, Zaidi N, Wang B, Matos I, Fiorese C et al. (2012) Treml4, an Ig superfamily member, mediates presentation of several antigens to T cells in vivo, which provide immunity to HER2 protein. Journal of Immunology 188: 1147-1155. doi:10.4049/jimmunol.1102541.

73. Mulligan AM, Raitman I, Feeley L, Pinnaduwage D, Nguyen LT et al. (2013) Tumoral lymphocytic infiltration and expression of the chemokine CXCL10 in breast cancers from the Ontario Familial Breast Cancer Registry. Clinical Cancer Research : an Official Journal of the American Association for Cancer Research 19: 336-346. doi:10.1158/1078-0432.CCR-11-3314.

74. Zong X, Yang H, Yu Y, Zou D, Ling Z et al. (2011) Possible role of Pax-6 in promoting breast cancer cell proliferation and tumorigenesis. BMB Rep 44: 595-600. doi:10.5483/BMBRep.2011.44.9.595. PubMed: 21944253.

75. Yang S, Du J, Wang Z, Yuan W, Qiao Y et al. (2007) BMP-6 promotes E-cadherin expression through repressing deltaEF1 in breast cancer cells. BMC Cancer 7: 211. doi:10.1186/1471-2407-7-211. PubMed: 17997862.

76. Zhang M, Wang Q, Yuan W, Yang S, Wang X et al. (2007) Epigenetic regulation of bone morphogenetic protein-6 gene expression in breast cancer cells. J Steroid Biochem Mol Biol 105: 91-97. doi:10.1016/j.jsbmb.2007.01.002. PubMed: 17574840.

77. Bertucci F, Nasser V, Granjeaud S, Eisinger F, Adelaïde J et al. (2002) Gene expression profiles of poor-prognosis primary breast cancer correlate with survival. Hum Mol Genet 11: 863-872. doi:10.1093/hmg/11.8.863. PubMed: 11971868.

78. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM et al. (2013) APOBEC3B is an enzymatic source of mutation in breast cancer. Nature 494: 366-370. doi:10.1038/nature11881. PubMed: 23389445.

79. Kitamura N, Nakamura Y, Miyamoto Y, Miyamoto T, Kabu K et al. (2011) Mieap, a p53-inducible protein, controls mitochondrial quality by repairing or eliminating unhealthy mitochondria. PLOS ONE 6: e16060. doi:10.1371/journal.pone.0016060. PubMed: 21264228.

80. Kool M, van der Linden M, de Haas M, Baas F, Borst P (1999) Expression of human MRP6, a homologue of the multidrug resistance protein gene MRP1, in tissues and cancer cells. Cancer Res 59: 175-182. PubMed: 9892204.

81. Berois N, Mazal D, Ubillos L, Trajtenberg F, Nicolas A et al. (2006) UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-6 as a new immunohistochemical breast cancer marker. J Histochem Cytochem 54: 317-328. doi: 10.1369/jhc.5A6783.2005. PubMed: 16260590.

82. Piotrowski A, Benetkiewicz M, Menzel U, Díaz de Ståhl T, Mantripragada K et al. (2006) Microarray-based survey of CpG islands identifies concurrent hyper- and hypomethylation patterns in tissues derived from patients with breast cancer. Genes Chromosomes Cancer 45: 656-667. doi:10.1002/gcc.20331. PubMed: 16575877.

83. Badve S, Turbin D, Thorat MA, Morimiya A, Nielsen TO et al. (2007) FOXA1 expression in breast cancer--correlation with luminal subtype A and survival. Clin Cancer Res 13: 4415-4421. doi: 10.1158/1078-0432.CCR-07-0122. PubMed: 17671124.

84. Krig SR, Frietze S, Simion C, Miller JK, Fry WH et al. (2011) Lrig1 is an estrogen-regulated growth suppressor and correlates with longer relapse-free survival in ERalpha-positive breast cancer. Mol Cancer Res 9: 1406-1417. doi:10.1158/1541-7786.MCR-11-0227. PubMed: 21821674.

85. Wang Y, Poulin EJ, Coffey RJ (2013) LRIG1 is a triple threat: ERBB negative regulator, intestinal stem cell marker and tumour suppressor. Br J Cancer 108: 1765-1770. doi:10.1038/bjc.2013.138. PubMed: 23558895.

86. Rodenhiser DI, Andrews J, Kennette W, Sadikovic B, Mendlowitz A et al. (2008) Epigenetic mapping and functional analysis in a breast cancer metastasis model using whole-genome promoter tiling microarrays. Breast Cancer Res 10: R62. doi:10.1186/bcr2060. PubMed: 18638373.