

Assessing the efficacy of androgen receptor and Sox10 as independent markers of the triple-negative breast cancer subtype by transcriptome profiling

Khalid N. Al-Zahrani^{1,2}, David P. Cook^{1,2}, Barbara C. Vanderhyden^{1,2} and Luc A. Sabourin^{1,2}

¹Ottawa Hospital Research Institute, Centre for Cancer Therapeutics, Ottawa, Ontario, Canada

²Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Ontario, Canada

Correspondence to: Luc A. Sabourin, **email:** lsabourin@ohri.ca

Keywords: androgen receptor; Sox10; bioinformatics; biomarkers; breast cancer

Received: April 10, 2018

Accepted: August 13, 2018

Published: September 07, 2018

Copyright: Al-Zahrani et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

The Androgen Receptor (AR) has recently garnered a lot of attention as a potential biomarker and therapeutic target in hormone-dependent cancers, including breast cancer. However, several inconsistencies exist within the literature as to which subtypes of breast cancer express AR or whether it can be used to define its own unique subtype. Here, we analyze 1246 invasive breast cancer samples from the Cancer Genome Atlas and show that human breast cancers that have been subtyped based on their HER2, ESR1, or PGR expression contain four clusters of genes that are differentially expressed across all subtypes. We demonstrate that Sox10 is highly expressed in approximately one-third of all HER2/ESR1/PGR-low tumors and is a candidate biomarker of the triple-negative subtype. Although AR expression is acquired in many breast cancer cases, its expression could not define a unique subtype. Despite several reports stating that AR expression is acquired in HER2/ESR1/PGR triple-negative cancers, here we show that a low percentage of these cancers express AR (~20%). In contrast, AR is highly expressed in HER2-positive or ESR1/PGR-positive cancers (> 95%). Although AR expression cannot be used as an independent subtype biomarker, our analysis shows that routine evaluation of AR expression in tumors which express HER2, ESR1 and/or PGR may identify a unique subset of tumors which would benefit from anti-androgen based therapies.

INTRODUCTION

Breast cancer is the most common cancer in women, accounting for approximately 25% of all reported carcinomas worldwide [1]. However, the molecular basis for the various breast cancer subtypes is not fully understood and the vast heterogeneity that exists within each tumor makes treatment an even more difficult task.

Breast cancer subtypes are most commonly classified based on the expression of the Estrogen Receptor (ESR1), Progesterone Receptor (PGR), and/or the Human Epidermal Growth Factor Receptor-2 (HER2) [2–5]. ESR1 and PGR-positive breast cancers make up

the Luminal A (Ki-67-negative) and Luminal B (Ki-67-positive) subtypes [2–5]. The Luminal A and B subtypes are the most common, accounting for approximately 65% of all breast cancer cases [4]. The Luminal subtypes tend to have the best prognostic outcome of all subtypes and these tumors generally respond well to hormone therapy [6]. The HER2-positive subtype is characterized by overexpression and amplification of HER2 with a prevalence of approximately 25% and a poor prognosis due to its association with highly metastatic breast cancers [4, 7–9]. The monoclonal antibody, Trastuzumab (Herceptin), in combination with chemotherapy is currently the best form of treatment for HER2-positive breast cancers [10].

Lastly, the triple-negative breast cancers (TNBC) are defined by those that do not express HER2, ESR1, or PGR [5]. TNBCs account for approximately 10% of all cases and have the worst prognosis and survival rates, as major therapeutic targets have not yet been identified in this subtype [4]. As biomarkers and/or drivers have not been fully identified for this subtype, chemotherapy remains the standard of care treatment for systemic TNBC [11–13].

The Androgen Receptor (AR) is a steroid hormone nuclear receptor that is regulated through cytoplasmic binding of testosterone or dihydrotestosterone and subsequent translocation to the nucleus and activation of gene transcription [14]. In addition to its well characterized role in sexual development [15], AR has gained increasing attention as an important mediator of hormone-dependent cancers and a novel therapeutic target in breast cancer [16–18]. Increased expression, nuclear localization and/or phosphorylation of AR have now been observed in a few various breast cancer subtypes [17, 19, 20]. Several studies and systematic meta-analyses have now demonstrated that higher levels of AR are associated with a better prognostic outcome and a reduction in metastatic burden [21–23]. Our lab has previously identified that AR expression is acquired in a murine model of HER2-positive breast cancer following Periostin deletion [20]. We reported that these tumors were of a molecular apocrine histology and have since gained an interest in understanding how AR could be used as a possible biomarker of a specific subtype of breast cancer.

There is current controversy in the literature as to whether AR expression can be used as a biomarker for HER2-positive, ESR1-positive or TNBC, with much of the literature reporting the highest expression in TNBC tumors [21, 24–28]. Most reports indicating that AR is overexpressed in breast cancers are concluded based on immunohistochemical staining. A recent report, which included a review of 23 other TNBC studies, reported that

a high number of TNBC tumors acquire AR expression [26]. However, nine of these studies used a very low threshold of 1% positive cells to define acquired AR expression, with eleven others using 10% positive cells as the cut-off. Although the methods for detecting ESR1 or HER2 in breast cancer biopsies by immunohistochemistry are well established [29], AR immunohistochemistry performed on tumor biopsies is not routine or standardized leading to a wide variation in the reported number of AR-positive breast cancer cases [27, 28]. Given this large degree of variability, more consistent methods for determining AR status in the clinic are required.

To assess the subtype distribution of AR expression across human breast cancers, we interrogated The Cancer Genome Atlas containing RNA-Seq data from 1246 invasive human breast cancer samples. Using this dataset, we have shown that AR is highly expressed in over 95% of HER2-, ESR1- or PGR-positive tumors while TNBCs tend to express AR less frequently (~20%).

RESULTS

Luminal A/B, HER2-positive and triple-negative subtypes can be defined by four distinct gene expression signatures

We first aimed to stratify the RNA-Seq data available from The Cancer Genome Atlas according to the three major breast cancer subtypes: HER2-positive, ESR1/PGR-positive, and triple negative. Gene-level quantifications from RNA-Seq data of 1246 invasive breast cancer samples from The Cancer Genome Atlas [30] were clustered based on their expression of HER2, ESR1, and PGR. This resulted in three distinct putative subtypes: a HER2-positive, a Luminal A/B subtype which is ESR1/PGR-positive, and a TNBC subtype lacking expression of all three receptors (Figure 1).

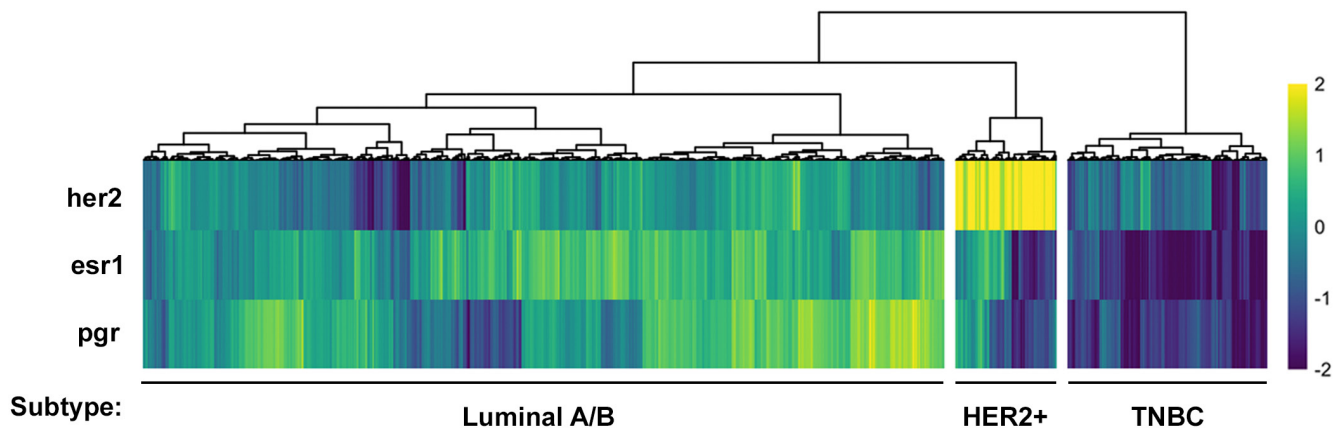


Figure 1: Stratification of human breast cancer samples from *The Cancer Genome Atlas* database into three putative breast cancer subtypes. Receptor heatmap of HER2, ESR1 and PGR expression (z-score of log₂ counts) across all samples. Each column represents the expression values for an individual patient sample. Hierarchical clustering was used to group patients by the expression patterns of each receptor. This clustering revealed a group of putative TNBC, HER2+ and Luminal A/B patients.

Next, we sought to define an expanded gene expression signature associated with these subtypes to identify novel biomarkers. Additionally, we were interested in assessing whether AR could be used to define its own unique subtype. To perform this, we identified differentially expressed genes (ANOVA, FDR < 0.01) with a minimum log₂ fold change of 4 (16-fold) across subtypes. This resulted in four clusters of genes with distinct expression patterns between subtypes, which we have termed “Marker Clusters”, comprising 112 genes (Figure 2; Table 1).

Luminal A/B and HER2-positive breast cancers have a defined gene signature

The Marker Cluster 3 corresponded to genes whose expression is highest in the Luminal A/B subtype and contained the canonical markers ESR1 and PGR (Figure 3, Table 1). Other genes within this cluster include CCDC170, WNK4 and AGR3 (Table 1) which have been reported to be implicated in these cancers [31–33].

Marker Cluster 4 genes are enriched in HER2-positive breast cancers and includes ERBB2 (HER2), GRB7 and CEACAM6 (Figure 3, Table 1). This subtype was well-defined, with all samples expressing higher levels of HER2 than the mean of all samples. As with the Luminal cluster of genes, most genes making up this cluster have been heavily implicated in HER2-positive breast cancer [34, 35].

Two marker clusters define a TNBC gene signature

Two remaining Marker Clusters corresponded to gene expression patterns that define the TNBC subtype. Firstly, we identified Marker Cluster 1, termed the “mosaic” cluster, as the grouping of genes that were significantly increased in both the HER2-positive and Luminal A/B subtypes when compared to the TNBC subtype (Figure 3, Table 1). Therefore, this cluster represents genes whose RNA expression values are lowest in the TNBC subtype. Interestingly, this cluster contains AR and one of its target genes, PIP [36] which are both expressed at much lower levels in TNBC tumors when compared to the other two subtypes (Figure 3). Further investigation into AR expression reveals that only two major isoforms are expressed in human breast cancers, with both isoforms showing a significantly lower level of expression in the TNBC subtype compared to both other subtypes (Supplementary Figure 1). Additionally, HER2-positive and Luminal A/B tumors express that same isoforms of AR suggesting a similar function for AR in these subtypes.

Given that the “mosaic” Marker Cluster contains genes that are generally downregulated in TNBC tumors compared to HER2-positive and Luminal A/B tumors, we investigated Marker Cluster 2, which contained genes that are significantly enriched in the TNBC subtype. The most striking differences in our differential expression

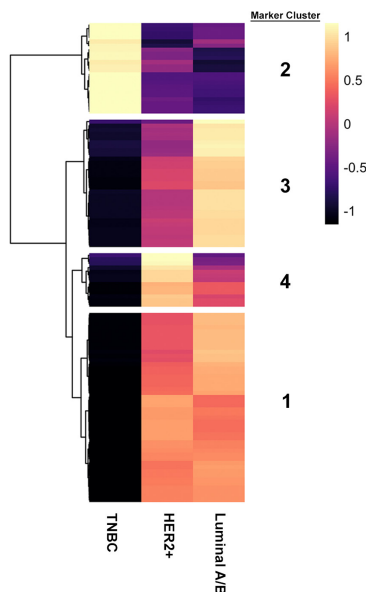


Figure 2: Four distinct marker clusters can be stratified from three putative breast cancer subtypes. Marker heatmap showing the average expression (z-score) of 112 marker genes across the three molecular subgroups identified in panel A. Each row represents an individual gene. Marker genes were identified by using an ANOVA to identify genes that are differentially expressed across molecular subtypes. Differentially expressed genes were defined as those with an FDR-adjusted p-value < 0.01 and a log₂ fold-change > 4 between the lowest and highest expressing subtype. Hierarchical clustering was used to define four groups of marker genes with similar expression patterns across subtypes.

Table 1: Top 20 identified genes for each Marker Cluster identified across HER2+, Luminal A/B and TNBC subtypes

Marker Cluster 1 (Mosaic)	Marker Cluster 2 (High in TNBC)	Marker Cluster 3 (Luminal A/B)	Marker Cluster 4 (HER2+)
AR	SOX10	PGR	ERBB2
PIP	ROPN1	ESR1	CEACAM5
CLC7A2	GABRP	CCDC170	CEACAM6
ABCC8	VGLL1	GRPR	ABCC11
F7	MSLN	WNK4	NXPH1
BCAS1	CA9	CHAD	ABCC12
CA12	GABBR2	AFF3	GRB7
BPIFB2	HORMAD1	GFRA1	PNMT
DHRS2	ZIC1	CPB1	MUCL1
TMC5	ART3	CLSTN2	LRRC26
ARG2	FABP7	PGLYRP2	UGT2B11
GATA3	A2ML1	KCNJ3	PP14571
SCGB2A2	FDCSP	NEK10	DSCAM-AS1
LRRC31	PRAME	GRIK3	
MLPH	KRT16	DNALI1	
FOXA1	SBSN	SERPINA6	
HMGCS2	PPP1R14C	SYT9	
TTC6	CT83	NAT1	
CYP4B1	NKX1-2	CST9	
SYTL5		AGR3	

The four Marker Clusters (and the putative group that they define) are provided with the top genes identified for each cluster with an FDR-adjusted p-value <0.01 and a log₂ fold-change > 4 between the lowest and highest expressing subtypes. For clusters with more than 20 genes, only the top 20 are shown here. These genes correspond to those identified by hierarchical clustering in Figure 2.

analysis were the genes that show increased expression in the absence of HER2, ESR1 and PGR. This can be observed in the first column of the marker heatmap in Figure 2 as well as the top rows of marker heatmap averaged across clusters in Figure 3A. SOX10 is among the most differentially expressed genes within the TNBC subtype (Figure 3C) and has gained increasing attention as a novel biomarker for these tumors over the past decade [37, 38]. Within the TNBC cluster, tumors tend to express higher levels of SOX10 than any other subtype (Figure 3C), providing further evidence that SOX10 is in fact a biomarker of TNBC. This is strikingly similar to the 73.3% of Luminal A/B tumors with PGR expression – a bona fide Luminal A/B marker – greater than its mean across all samples.

Sox10 can be used as an independent biomarker of the TNBC subtype

The current strategy for identifying the TNBC subtype is to assess patients for the absence of HER2, ESR1 and PGR, which is the approach that we used when first analyzing the TCGA dataset (Figure 1). However, identifying a gene or set of genes whose expression defines a subtype could improve clinical diagnosis, may provide a better insight into the disease and allow for the identification of novel therapeutics. One of the strongest candidates for a bona fide biomarker of the TNBC subtype that we have identified is Sox10 (Figure 3A, 3C; Table 1). To further validate whether Sox10 is a biomarker of the TNBC subtype, we performed unsupervised clustering of the TCGA dataset based on the expression of HER2,

ESR1, PGR and SOX10. We predicted that if Sox10 was a biomarker of the TNBC subtype, adding its expression analysis to the clustering we performed with HER2/ESR1/PGR alone (Figure 1) would not change the proportion of patient samples within each putative subtype. Confirming this hypothesis, the three putative subtypes remained effectively identical when SOX10 expression was added to the clustering analysis (Figure 4). Although SOX10 expression is highest in the TNBC subtype, it is interesting to note that most HER2-positive samples have very low expression of Sox10 while the Luminal A/B subtype has more variable SOX10 expression (Figure 4).

Sox10 correlates with a more basal/stem-like phenotype in the TNBC but not luminal A/B subtype

The TNBC subtype is often associated with increased stem cell activity which may result in these tumors being more resistant to conventional drug therapies [11, 39]. We have also identified that SOX10 is amongst the most differentially upregulated genes in the TNBC subtype (Table 1). Interestingly, SOX10 transcriptional activity has been shown to be sufficient to reprogram pluripotent cells into a multipotent state and more specifically has been shown to regulate the stem/progenitor activity of mammary epithelial cells [40, 41]. Therefore, we aimed to assess whether SOX10 expression was correlated with an increase in stemness. To perform this analysis, we utilized a recent study in which the

stemness indices for all TCGA samples were calculated [42] and plotted stemness index against SOX10 expression (Figure 5). Interestingly, high SOX10 expression correlated with high stemness only in the TNBC subtype, whereas Luminal A/B tumors which had higher levels of SOX10 tended towards a lower stemness index (Figure 5). These data suggest that SOX10 alone is not the only driver of the stem-like phenotype or that Luminal A/B and HER2 tumors possess a pro-differentiation program that is able to overcome the basal/stem-like state driven by SOX10.

Androgen receptor expression is low in TNBC but acquired in a subset of HER2, ESR1 or PGR-positive tumors

AR expression was observed to be significantly increased in the mosaic Marker Cluster, which was characterized by having higher expression in HER2-positive and Luminal A/B subtypes when compared to TNBCs (Figure 3B). Therefore, we set out to identify the extent of AR expression within all three breast cancer subtypes. To address this, we compared AR expression with the strongest identified marker of each subtype: HER2, PGR or SOX10 (Figure 6A-6F). Consistent with our clustering, HER2 (Figure 6A, 6D), PGR (Figure 6B, 6E) and SOX10 (Figure 6C, 6F) were most expressed in the HER2-positive, Luminal A/B and TNBC subtypes, respectively. Interestingly, AR expression tends to be higher in the HER2- and PGR-positive samples but lower in the SOX10-positive cases as seen by a leftward shift of

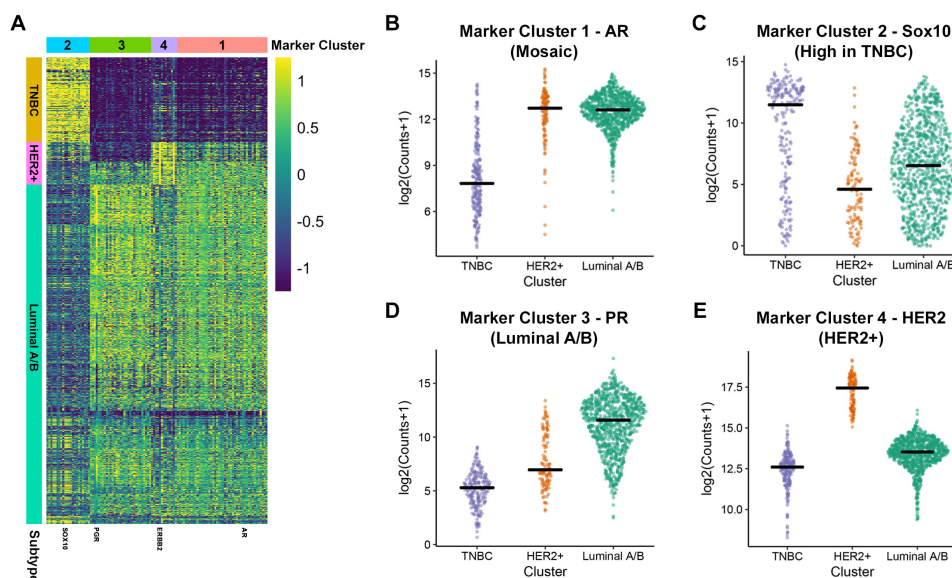


Figure 3: Identification of gene markers for HER2+, Luminal A/B and TNBC. (A) Marker heatmap showing the expression (z-score of log₂ counts) of 112 marker genes. Each row represents the expression of a single gene across each patient sample (columns). The rows of the heatmap are ordered identically to the receptor heatmap in Figure 1. (B-E) Expression of a representative identified marker across molecular subtypes is shown for each of the four identified marker groups. Each plot provides the distribution of log₂ expression values of AR (B), SOX10 (C), PGR (D) and HER2 (E) across subtypes. The black line represents the median expression value for each subtype.

SOX10-high/AR-low samples (Figure 6C). Together, these data suggest that AR expression is lower in the majority of TNBC samples.

When comparing HER2 and AR expression 93.9% of HER2-positive patients express high levels of both receptors (Figure 6D, 6G). Next, we compared PGR and

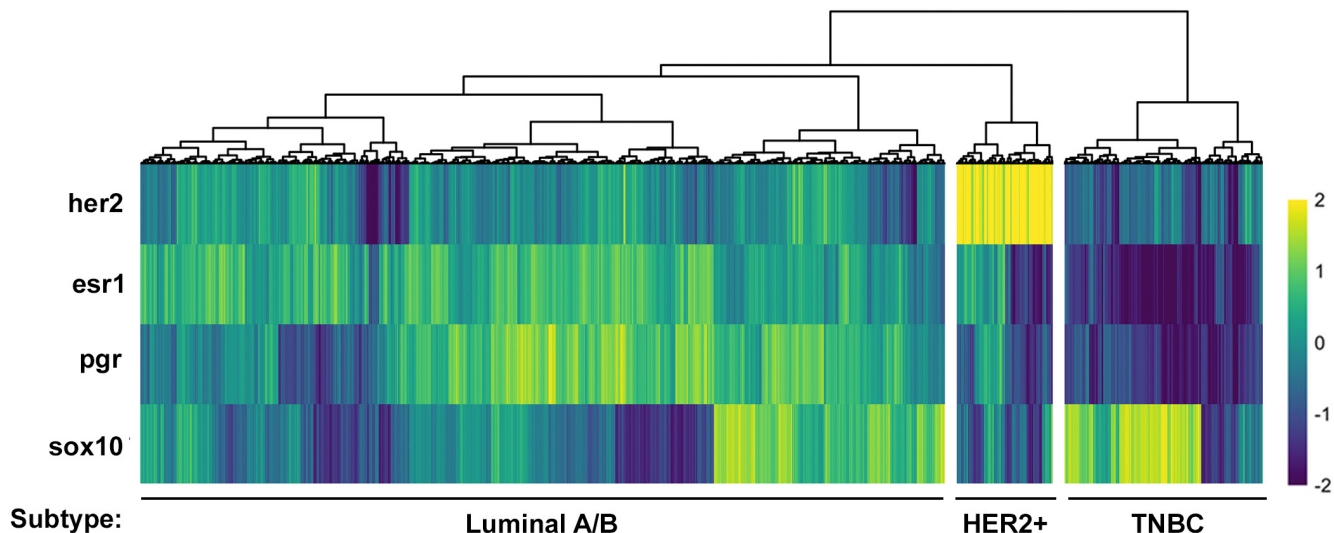


Figure 4: SOX10 expression can be used to independently define a TNBC subtype. Heatmap of HER2, ESR1, PGR and SOX10 expression (z-score of log₂ counts) across all samples. Each column represents the expression values for an individual patient sample. Hierarchical clustering was used to group patients by the expression patterns of each gene. This clustering revealed a group of putative TNBC, HER2+ and Luminal A/B patients which was almost identical to that in Figure 1 suggesting that SOX10 expression defines a TNBC subtype.

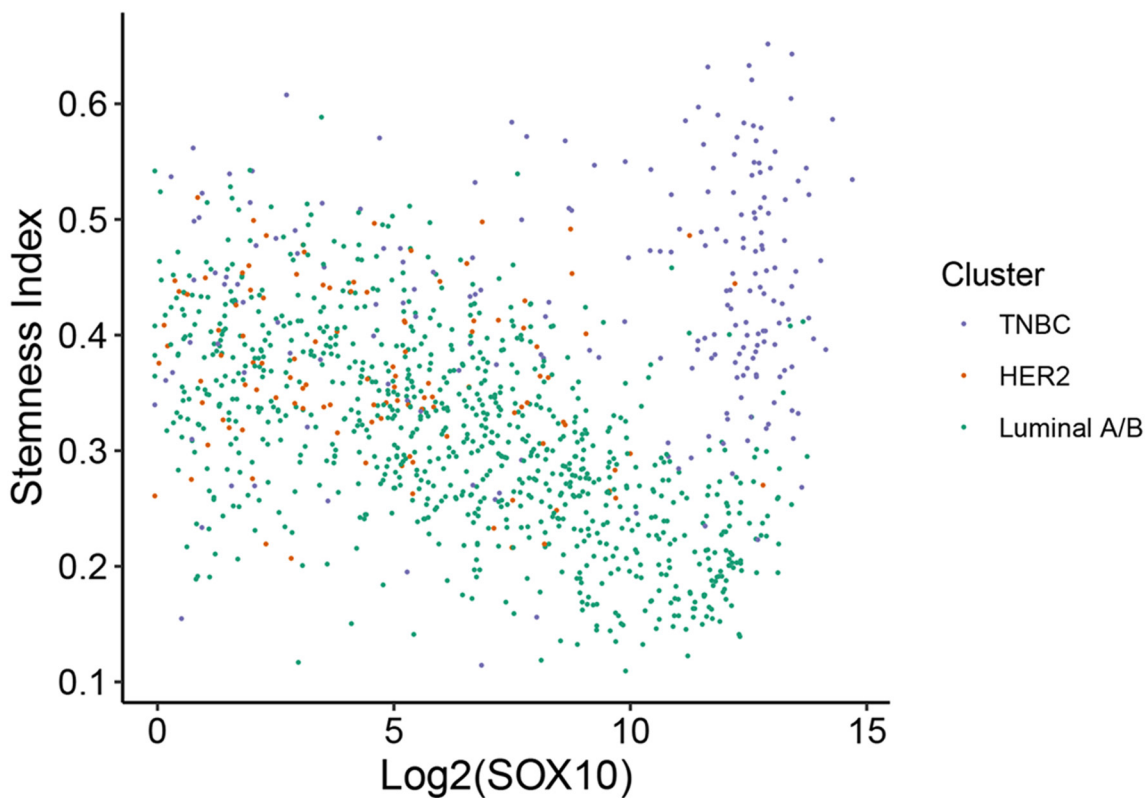


Figure 5: SOX10 expression correlates with a more basal/stem-like phenotype in TNBC. Stemness index for all 1246 patients from the TCGA dataset was previously calculated [42] and plotted against SOX10 expression (log₂). Individual samples were color coded according to their identified molecular subtype (Figure 1).

AR expression within the Luminal A/B subtype. Here, we observed that under 2% of the entire subtype has low expression of AR, with over 98% expressing high levels of AR independent of PGR expression levels (Figure 6E). If we consider only Luminal tumors with high PGR expression, AR is highly expressed in 99.2% of those samples (Figure 6H). Lastly, we compared AR expression with the novel TNBC marker, SOX10. Within this subtype we observe that 71.3% of TNBC tumors display low levels of AR, suggesting that AR independently cannot be considered a biomarker for the TNBC subtype (Figure 6C, 6F, 6I).

To provide further evidence for the lack of AR expression in the TNBC subtype, we utilized our hierarchical clustering of patient samples based on HER2, ESR1, PGR and SOX10 and assessed the expression of AR across each patient (Figure 7). The patient samples are ordered identically to Figure 4, with AR expression not affecting the clustering. This confirmed that AR expression is in fact lowest in the SOX10-positive TNBC patient samples, inversely correlated with SOX10 expression (Figure 7). In fact, the few sporadic patients in the TNBC subtype with low SOX10 expression have a higher than average expression of AR (Figure 7).

To corroborate the findings drawn from our analysis, we performed immunohistochemical staining of a human breast cancer TMA for both SOX10 and AR. In accordance with our previous analysis, Sox10 histochemistry revealed an increased staining intensity and more prominent nuclear localization in the TNBC tumors than within the HER2 and Luminal A/B cores (Figure 8A, 8C). Additionally, quantification of the AR staining showed no significant differences between subtypes, although qualitatively a weaker cytoplasmic signal was noticed within TNBC tumor cores when compared to the other subtypes (Figure 8A, 8B). Therefore, at the level of both the transcriptome and proteome, SOX10 acts a strong marker of the TNBC subtype.

DISCUSSION

To assess the subtype distribution of AR amongst the three major breast cancer subtypes, we stratified The Cancer Genome Atlas into putative HER2-positive, Luminal A/B and Triple Negative subtype using the available RNA-seq data [30]. Interestingly, molecular subtyping of breast cancers in the clinic is routinely based

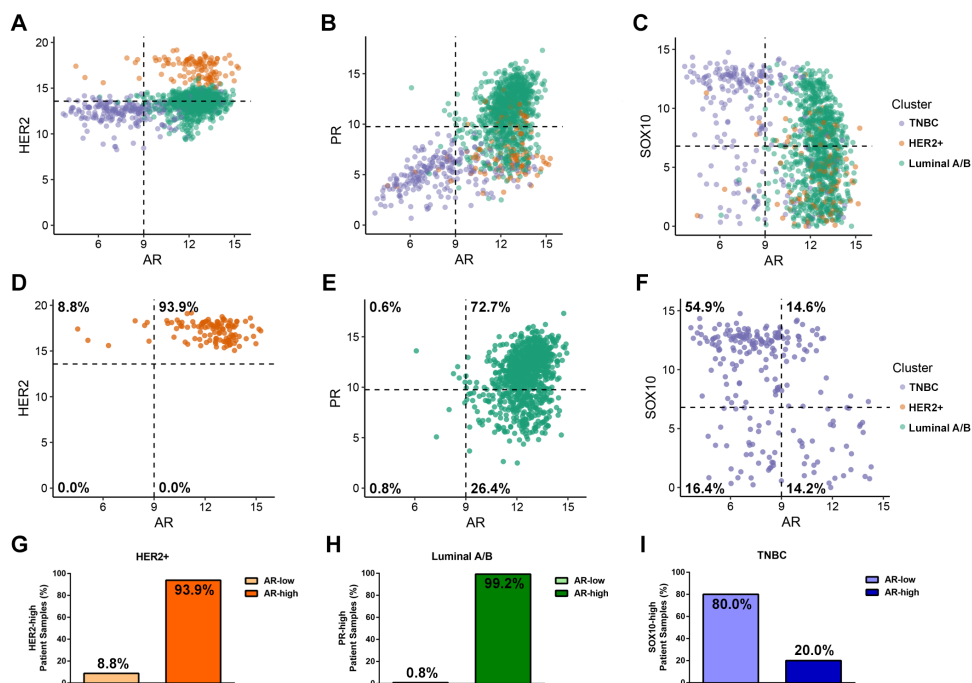


Figure 6: Androgen Receptor is expression is highest in the HER2-positive and Luminal A/B subtypes. (A-C) Scatter plots showing AR expression (log2) and the expression of markers for each molecular subtype. HER2 was used as a marker of the HER2+ subtype (A), PR for the Luminal A/B subtype (B) and Sox10 for the TNBC subtype (C). Each point represents an individual patient sample and is colour coded according to which subtype that sample stratified with from Figure 1. The dotted lines along the axes is used to visually quadrant the data points. For the subtype markers, this line corresponds to the mean expression value of the indicated gene across all 1246 patient samples, and for AR, the line was manually set at a value that splits the bimodal distribution of AR expression observed across cell types. (D-F) Identical plots to (A-C), but reduced to only samples from the HER2-positive (D), Luminal A/B (E) or TNBC (F) subtypes. The percentage of patient samples that fall within each quadrant of expression is provided. (G-I) The percentage of AR-low or AR-high expressing tumor samples within the HER2-positive (G), Luminal A/B (H) or TNBC (I) subtypes which express high levels of their corresponding markers (HER2, PGR and SOX10, respectively).

on immunohistochemical analysis of HER2-, ESR1- and PGR-expression from tumor biopsies and is rarely assessed at a transcriptional level [29]. Consistent with the immunohistological molecular subtyping, our hierarchical clustering of patient samples based on HER2, ESR1

and PGR expression resulted in a similar distribution of subtypes, with most samples falling within the ESR1/PGR-positive luminal subtype and TNBC making up the minority of samples [2–9].

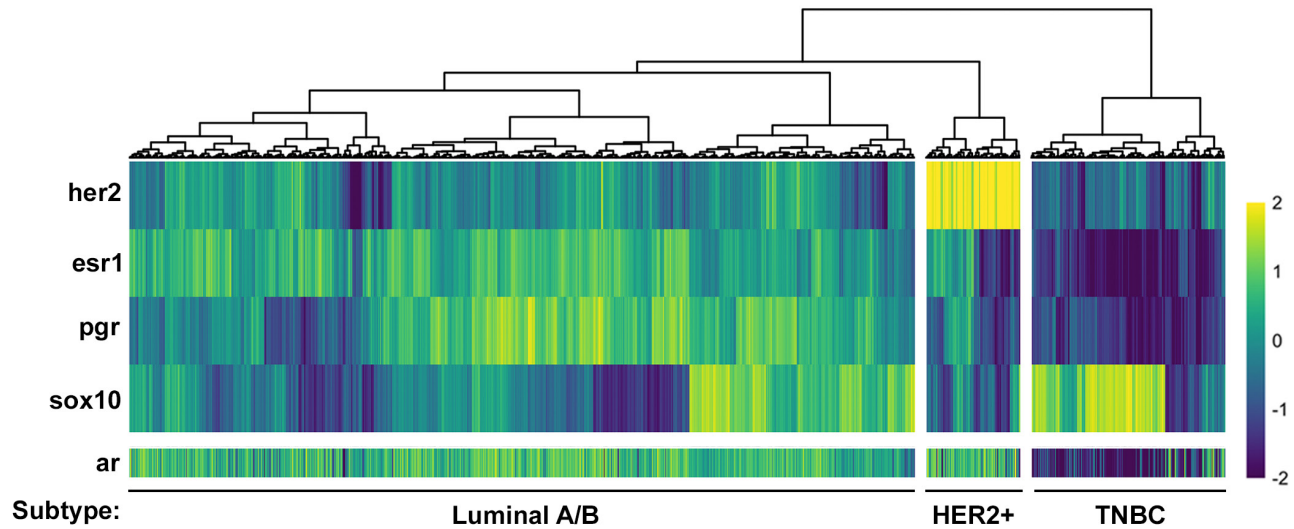


Figure 7: Androgen Receptor expression is inversely correlated with the TNBC marker, Sox10. AR expression data (z-score of log₂ counts) was plotted for each patient sample from an unsupervised clustering of the TCGA dataset using HER2, ESR1, PGR and SOX10 expression and are lowest in the TNBC subgroup.

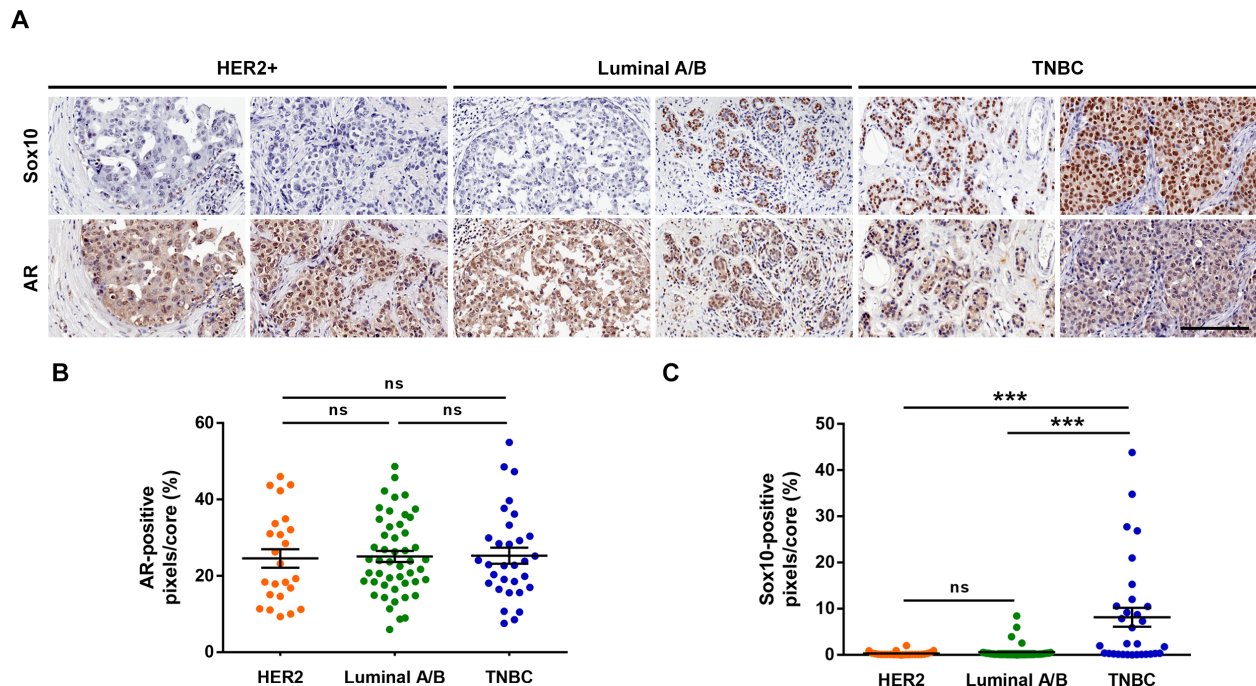


Figure 8: SOX10 histochemistry can be used as a predictor of TNBC. (A) Human breast cancer TMA BR20810 was purchased from US Biomax. Serial slides containing 104 breast cancer cases were stained for SOX10 and AR. Two representative cores for each subtype are shown. (B, C) Each core was quantified for the positive-pixel intensity for both AR (B) and SOX10 (C) staining. No differences were observed between subtypes for AR staining. However, a significant increase in SOX10 staining intensity was observed in the TNBC tumor cores. Scale bar = 200 μm.

To identify potential biomarkers within each molecular subtype we chose to perform our analysis using a log₂ fold change of 4 (16-fold) cut-off. This was chosen as any biomarkers defining these subtypes should have a robust difference in expression. This cut-off resulted in 112 significantly changed genes between each subtype, with no more than 25 genes comprising any one cluster. We believe that these clusters form a manageable list of genes that can be considered as potential biomarkers for each group. Although arbitrarily defined, the identification of HER2 and ESR1/PGR as markers of their corresponding subtype validates our approach. More stringent cut-offs, such as log₂ fold change of 5 (32-fold) or greater may be applied to this data set to identify more rigorous biomarkers of each subtype.

Interestingly, our analysis identified distinct gene signatures for both the Luminal A/B and HER2-subtypes. A number of these genes have been validated as either important for the progression of their respective subtype or reported to be a biomarker for that disease [31–35]. One caveat to our approach is that we cannot determine whether the expression of any given gene is downregulated in two of the three subtypes that we defined (e.g. promoter hypermethylation) or, conversely, whether that gene is induced in one subtype (e.g. chromosome duplication). One possible method to address these issues is to include matched normal tissue as a baseline measure of gene expression to elucidate whether expression is gained or lost during tumorigenesis. These gene signatures may be very important in deriving more-thorough subtype classifications beyond the hormone receptors as well as novel therapeutic strategies in the treatment of these cancer subtypes.

One of the most induced genes within the TNBC subtype was SOX10 showing that it may be used as a potential biomarker of the TNBC subtype. This also raises the interesting possibility that SOX10 targeted therapies may be beneficial in treating TNBC patients. Although interfering with transcription factor activity can be challenging therapeutically, it has proven to be successful in a number of clinical trials (reviewed in [43]). Furthermore, identifying specific target genes, pathways and processes regulated by SOX10 in the context of breast cancer may provide a novel therapeutic approach in the treatment of TNBCs.

Although SOX10 is highly expressed in the TNBC cluster, we have also observed that a subset of Luminal A/B tumors also express SOX10 (Figure 3C, 4, 6C). Initially this observation would argue against SOX10 being an independent marker of the TNBC subtype. However, Luminal A/B tumors which have high levels of SOX10 also express high levels of AR. Assessing tumors for both AR and SOX10 may provide a novel method for distinguishing TNBC tumors from the small proportion of Luminal tumors that express SOX10. Therefore, screening Luminal A/B patients for SOX10 may provide a new avenue for treatment for endocrine therapy resistant tumors. Further

gene expression analyses could reveal whether these SOX10-positive luminal cases dichotomize the Luminal A and Luminal B patients. Of note, the Luminal A/B patients fall within three major subgroups which can be defined by high ESR1-expression with low PGR-expression, low ESR1-expression with high PGR-expression or a median level of expression of both receptors (Figure 1, 3). Interestingly, high SOX10-expression within the Luminal A/B subtype falls within this last subgroup with an average expression of both ESR1 and PGR.

A dichotomy exists in the current literature as to whether AR can be used as a biomarker or a “fourth receptor” along with HER2, ESR1 and PGR to define the TNBC subtype, with many sources claiming AR to be a TNBC marker [21, 24, 25]. Although many of these studies report that a small percentage of TNBC cases are of a luminal androgen receptor positive (LAR) molecular classification, we believe that caution should be taken in the use of “biomarker” in the context of the whole TNBC subtype as we have shown that SOX10 and AR are often reciprocally expressed. The high degree of correlation between AR and HER2, ESR1 or PGR lends further support to the so-called Quadruple Negative Breast Cancer (QNBC) hypothesis, which may be a stratification of the TNBC subtype with high expression of SOX10 [44]. Although these QNBC tumors would probably not respond to anti-AR based therapeutics, the 20-30% of TNBC cancers that do in fact express AR might and most likely belong to the LAR molecular subgroup [25].

As AR is often co-expressed with HER2 specifically in the HER2-positive subtype, we believe that these patients may also be responsive to anti-androgen based therapies. Currently, Herceptin is the best form of treatment for HER2-positive breast cancers, but often patients develop resistance due to constitutive activation of downstream HER2 signalling [10]. Targeting of AR may have a large impact as a novel therapeutic strategy for treating Herceptin resistant HER2-positive breast cancers or could even be considered as a strong candidate for combination therapy. As with the HER2-positive subtype, the majority of Luminal A/B tumors express high levels of AR. Although these cancers have the best prognostic outcome and respond relatively well to hormone therapy, they may also benefit from anti-androgen based therapies.

AR expression is elevated in samples that have high expression of HER2, ESR1, or PGR. Therefore, it is unlikely that these receptors are all independently responsible for maintaining AR expression. Similarly, this would suggest that it is unlikely that SOX10 is acquired in TNBC tumors due to the loss of AR. We have also corroborated this by immunohistochemical analysis of primary patient tumor cores where we observed no differences in AR expression between HER2, Luminal A/B, or TNBC samples despite the TNBC subtype showing a significant increase in SOX10 expression (Figure 8). Additionally, SOX10 expression did not

correlate with increased stemness within the Luminal A/B or HER2-positive subtypes despite having a strong correlation in TNBC tumors (Figure 5), suggesting that these tumors may possess pro-oncogenic signals that bypass the basal/stem-like phenotype of TNBC tumors and keep them in a differentiated state. Understanding the mechanisms of SOX10 induction in breast cancers and the signaling that regulated SOX10 activity is a subject of ongoing research and will shed light on potentially new therapeutic targets for TNBC cancers.

Taken together, our data suggests that AR is expressed in many breast carcinomas, however unlike SOX10, it cannot be used as an independent biomarker. Here, our RNA-seq based approach for profiling the three major breast cancer subtypes matches what is currently reported in the literature with immunohistochemical based profiling. Additionally, by using RNA-seq data we performed a genome-wide exploration for markers associated with each subtype. However, our analysis from this same dataset suggests that AR expression is variable across all subtypes. Although a discordance between the levels of RNA transcripts and protein expression does exist, further analyses are required to determine whether current immunohistochemical approaches can accurately predict AR-positivity in breast cancer. Together, we have shown that AR is not a good biomarker for the existing subtypes of breast cancer, although its expression could be used to define further subtype stratification.

MATERIALS AND METHODS

Access of the cancer genome atlas database

Gene-level RNA-seq (v2) counts from the cohort of invasive breast carcinoma samples from The Cancer Genome Atlas (n=1246, [30]) were collected using the R package recount2 [45]. Counts were scaled to account for differences in library size across samples.

Stratification of molecular breast cancer subtypes

HER2, ESR1 and PGR expression (z-score of log2 counts) was calculated from each of the 1246 invasive breast carcinoma samples. Hierarchical clustering was used to group patients by the expression pattern of each receptor, revealing a group of putative TNBC, HER2-positive and Luminal A/B patients.

Identification of marker genes

Marker genes were identified using an ANOVA to identify genes that are differentially expressed across all molecular subtypes. Differentially expressed genes were filtered with a Benjamini-Hochberg-adjusted p-value <

0.01 (ANOVA) across all subtypes and a minimum log2 fold change of 4 (16-fold) between the subtype with the lowest expression and the one with the highest expression.

Immunohistochemistry

Human breast cancer tissue microarrays (BR20810, US Biomax) were deparaffinized and subject to antigen retrieval in 10 mM citrate buffer (pH 6.0) for 10 minutes in a pressure cooker. Endogenous peroxidase was quenched with 3% hydrogen peroxide for 15 minutes. Sections were blocked in 5% goat serum in PBS for an hour and incubated with either Sox10 (NBP2-44474, Novus) or AR (ab74272, abcam) primary antibody overnight at 4°C followed by incubation with the appropriate HRP-conjugated secondary at room temperature for 30 minutes. Antibodies were incubated in blocking solution. Sections were then incubated in DAB substrate (Sigma Aldrich) and counterstained with Haematoxylin. Sections were dehydrated in ethanol prior to clearing in xylene before mounting the slides. Positive pixels were enumerated using an Aperio Scanscope.

Data analysis scripts

All analysis scripts required to reproduce these findings are available at: https://github.com/dpcook/tcga_breast_cancer.

Abbreviations

AR: Androgen Receptor; HER2: Human Epidermal Growth Factor Receptor; ESR1: Estrogen Receptor 1; PGR: Progesterone Receptor; TNBC: Triple Negative Breast Cancer; FDR: False Discovery Rate; LAR: Luminal Androgen Receptor Positive; QNBC: Quadruple Negative Breast Cancer.

Author contributions

KNA conceptualized, analyzed and interpreted the data and wrote the manuscript. DPC produced the analysis scripts, generated and analyzed the data. BCV critically reviewed the manuscript. LAS conceptualized and interpreted the data and wrote the manuscript.

ACKNOWLEDGMENTS

We would like to acknowledge Dr. Doug Gray and Dr. Michele Ardolino for critical reading of this manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

KNA is funded by a Canadian Breast Cancer Foundation Doctoral Fellowship. This work was supported by grants from the Canadian Breast Cancer Foundation and the Cancer Research Society.

REFERENCES

1. Ghoncheh M, Pournamdar Z, Salehiniya H. Incidence and Mortality and Epidemiology of Breast Cancer in the World. *Asian Pac J Cancer Prev*. 2016; 17:43-6.
2. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, et al. Molecular portraits of human breast tumours. *Nature*. 2000; 406:747-52. <https://doi.org/10.1038/35021093>.
3. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001; 98:10869-74. <https://doi.org/10.1073/pnas.191367098>.
4. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS, Perou CM, Ellis MJ, Nielsen TO. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst*. 2009; 101:736-50. <https://doi.org/10.1093/jnci/djp082>.
5. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ, Panel Members. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol*. 2011; 22:1736-47. <https://doi.org/10.1093/annonc/mdr304>.
6. Brenton JD, Carey LA, Ahmed AA, Caldas C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J Clin Oncol*. 2005; 23:7350-60. <https://doi.org/10.1200/JCO.2005.03.3845>.
7. Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, Muller WJ. Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proc Natl Acad Sci U S A*. 1992; 89:10578-82.
8. Mansour EG, Ravdin PM, Dressler L. Prognostic factors in early breast carcinoma. *Cancer*. 1994; 74:381-400.
9. Dankort DL, Muller WJ. Signal transduction in mammary tumorigenesis: a transgenic perspective. *Oncogene*. 2000; 19:1038-44.
10. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, McShane LM, Paik S, Pegram MD, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med*. 2007; 131:18-43.
11. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, Ollila DW, Sartor CI, Graham ML, Perou CM. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res*. 2007; 13:2329-34. <https://doi.org/10.1158/1078-0432.CCR-06-1109>.
12. Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, Nielsen TO, Gelmon K. Metastatic behavior of breast cancer subtypes. *J Clin Oncol*. 2010; 28:3271-7. <https://doi.org/10.1200/JCO.2009.25.9820>.
13. Brouckaert O, Wildiers H, Floris G, Neven P. Update on triple-negative breast cancer: prognosis and management strategies. *Int J Womens Health*. 2012; 4:511-20. <https://doi.org/10.2147/IJWH.S18541>.
14. Mooradian AD, Morley JE, Korenman SG. Biological actions of androgens. *Endocr Rev*. 1987; 8:1-28. <https://doi.org/10.1210/edrv-8-1-1>.
15. Batch JA, Williams DM, Davies HR, Brown BD, Evans BA, Hughes IA, Patterson MN. Role of the androgen receptor in male sexual differentiation. *Horm Res*. 1992; 38:226-9.
16. Hickey TE, Robinson JL, Carroll JS, Tilley WD. Minireview: The androgen receptor in breast tissues: growth inhibitor, tumor suppressor, oncogene? *Mol Endocrinol*. 2012; 26:1252-67. <https://doi.org/10.1210/me.2012-1107>.
17. Takeshita T, Omoto Y, Yamamoto-Ibusuki M, Yamamoto Y, Iwase H. Clinical significance of androgen receptor and its phosphorylated form in breast cancer. *Endocr Relat Cancer*. 2013; 20:L15-21.
18. Chang C, Lee SO, Yeh S, Chang TM. Androgen receptor (AR) differential roles in hormone-related tumors including prostate, bladder, kidney, lung, breast and liver. *Oncogene*. 2014; 33:3225-34. <https://doi.org/10.1038/onc.2013.274>.
19. Ren Q, Zhang L, Ruoff R, Ha S, Wang J, Jain S, Reuter V, Gerald W, Giri DD, Melamed J, Garabedian MJ, Lee P, Logan SK. Expression of androgen receptor and its phosphorylated forms in breast cancer progression. *Cancer*. 2013; 119:2532-40. <https://doi.org/10.1002/cncr.28092>.
20. Sriram R, Lo V, Pryce B, Antonova L, Mears AJ, Daneshmand M, McKay B, Conway SJ, Muller WJ, Sabourin LA. Loss of periostin/OSF-2 in ErbB2/Neu-driven tumors results in androgen receptor-positive molecular apocrine-like tumors with reduced Notch1 activity. *Breast Cancer Res*. 2015; 17:7. <https://doi.org/10.1186/s13058-014-0513-8>.
21. Sutton LM, Cao D, Sarode V, Molberg KH, Torgbe K, Haley B, Peng Y. Decreased androgen receptor expression is associated with distant metastases in patients with androgen receptor-expressing triple-negative breast carcinoma. *Am J Clin Pathol*. 2012; 138:511-6. <https://doi.org/10.1309/AJCP8AVF8FDPTZLH>.
22. Qu Q, Mao Y, Fei XC, Shen KW. The impact of androgen receptor expression on breast cancer survival: a retrospective study and meta-analysis. *PLoS One*. 2013; 8:e82650. <https://doi.org/10.1371/journal.pone.0082650>.

23. Vera-Badillo FE, Templeton AJ, de Gouveia P, Diaz-Padilla I, Bedard PL, Al-Mubarak M, Seruga B, Tannock IF, Ocana A, Amir E. Androgen receptor expression and outcomes in early breast cancer: a systematic review and meta-analysis. *J Natl Cancer Inst.* 2014; 106:djt319. <https://doi.org/10.1093/jnci/djt319>.
24. Gucalp A, Traina TA. Triple-negative breast cancer: role of the androgen receptor. *Cancer J.* 2010; 16:62-5. <https://doi.org/10.1097/PPO.0b013e3181ce4ae1>.
25. Rampurwala M, Wisinski KB, O'Regan R. Role of the androgen receptor in triple-negative breast cancer. *Clin Adv Hematol Oncol.* 2016; 14:186-93.
26. Safarpour D, Pakneshan S, Tavassoli FA. Androgen receptor (AR) expression in 400 breast carcinomas: is routine AR assessment justified? *Am J Cancer Res.* 2014; 4:353-68.
27. Park S, Koo J, Park HS, Kim JH, Choi SY, Lee JH, Park BW, Lee KS. Expression of androgen receptors in primary breast cancer. *Ann Oncol.* 2010; 21:488-92. <https://doi.org/10.1093/annonc/mdp510>.
28. Christenson JL, Trepel JB, Ali HY, Lee S, Eisner JR, Baskin-Bey ES, Elias AD, Richer JK. Harnessing a Different Dependency: How to Identify and Target Androgen Receptor-Positive Versus Quadruple-Negative Breast Cancer. *Horm Cancer.* 2018; 9:82-94. <https://doi.org/10.1007/s12672-017-0314-5>.
29. Zaha DC. Significance of immunohistochemistry in breast cancer. *World J Clin Oncol.* 2014; 5:382-92. <https://doi.org/10.5306/wjco.v5.i3.382>.
30. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012; 490:61-70. <https://doi.org/10.1038/nature11412>.
31. Veeraraghavan J, Tan Y, Cao XX, Kim JA, Wang X, Chamness GC, Maiti SN, Cooper LJ, Edwards DP, Contreras A, Hilsenbeck SG, Chang EC, Schiff R, et al. Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers. *Nat Commun.* 2014; 5:4577. <https://doi.org/10.1038/ncomms5577>.
32. Cornen S, Guille A, Adelaide J, Addou-Klouche L, Finetti P, Saade MR, Manai M, Carbuccia N, Bekhouche I, Letessier A, Raynaud S, Charafe-Jauffret E, Jacquemier J, et al. Candidate luminal B breast cancer genes identified by genome, gene expression and DNA methylation profiling. *PLoS One.* 2014; 9:e81843. <https://doi.org/10.1371/journal.pone.0081843>.
33. Garczyk S, von Stillfried S, Antonopoulos W, Hartmann A, Schrauder MG, Fasching PA, Anzeneder T, Tannapfel A, Ergonen Y, Knuchel R, Rose M, Dahl E. AGR3 in breast cancer: prognostic impact and suitable serum-based biomarker for early cancer detection. *PLoS One.* 2015; 10:e0122106. <https://doi.org/10.1371/journal.pone.0122106>.
34. Bivin WW, Yergiyev O, Bunker ML, Silverman JF, Krishnamurti U. GRB7 Expression and Correlation With HER2 Amplification in Invasive Breast Carcinoma. *Appl Immunohistochem Mol Morphol.* 2017; 25:553-8. <https://doi.org/10.1097/PAI.0000000000000349>.
35. Tsang JY, Kwok YK, Chan KW, Ni YB, Chow WN, Lau KF, Shao MM, Chan SK, Tan PH, Tse GM. Expression and clinical significance of carcinoembryonic antigen-related cell adhesion molecule 6 in breast cancers. *Breast Cancer Res Treat.* 2013; 142:311-22. <https://doi.org/10.1007/s10549-013-2756-y>.
36. Baniwal SK, Little GH, Chinge NO, Frenkel B. Runx2 controls a feed-forward loop between androgen and prolactin-induced protein (PIP) in stimulating T47D cell proliferation. *J Cell Physiol.* 2012; 227:2276-82. <https://doi.org/10.1002/jcp.22966>.
37. Cimino-Mathews A, Subhawong AP, Elwood H, Warzecha HN, Sharma R, Park BH, Taube JM, Illei PB, Argani P. Neural crest transcription factor Sox10 is preferentially expressed in triple-negative and metaplastic breast carcinomas. *Hum Pathol.* 2013; 44:959-65. <https://doi.org/10.1016/j.humpath.2012.09.005>.
38. Ivanov SV, Panaccione A, Nonaka D, Prasad ML, Boyd KL, Brown B, Guo Y, Sewell A, Yarbrough WG. Diagnostic SOX10 gene signatures in salivary adenoid cystic and breast basal-like carcinomas. *Br J Cancer.* 2013; 109:444-51. <https://doi.org/10.1038/bjc.2013.326>.
39. Bertucci F, Finetti P, Cervera N, Esterni B, Hermitte F, Viens P, Birnbaum D. How basal are triple-negative breast cancers? *Int J Cancer.* 2008; 123:236-40. <https://doi.org/10.1002/ijc.23518>.
40. Kim YJ, Lim H, Li Z, Oh Y, Kovlyagina I, Choi IY, Dong X, Lee G. Generation of multipotent induced neural crest by direct reprogramming of human postnatal fibroblasts with a single transcription factor. *Cell Stem Cell.* 2014; 15:497-506. <https://doi.org/10.1016/j.stem.2014.07.013>.
41. Dravis C, Spike BT, Harrell JC, Johns C, Trejo CL, Southard-Smith EM, Perou CM, Wahl GM. Sox10 Regulates Stem/Progenitor and Mesenchymal Cell States in Mammary Epithelial Cells. *Cell Rep.* 2015; 12:2035-48. <https://doi.org/10.1016/j.celrep.2015.08.040>.
42. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelsken J, Omberg L, Gevaert O, Colaprico A, Czerwinska P, Mazurek S, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell.* 2018; 173:338-54 e15. <https://doi.org/10.1016/j.cell.2018.03.034>.
43. Bhagwat AS, Vakoc CR. Targeting Transcription Factors in Cancer. *Trends Cancer.* 2015; 1:53-65. <https://doi.org/10.1016/j.trecan.2015.07.001>.
44. Hon JD, Singh B, Sahin A, Du G, Wang J, Wang VY, Deng FM, Zhang DY, Monaco ME, Lee P. Breast cancer molecular subtypes: from TNBC to QNBC. *Am J Cancer Res.* 2016; 6:1864-72.
45. Collado-Torres L, Nellore A, Kammers K, Ellis SE, Taub MA, Hansen KD, Jaffe AE, Langmead B, Leek JT. Reproducible RNA-seq analysis using recount2. *Nat Biotechnol.* 2017; 35:319-21. <https://doi.org/10.1038/nbt.3838>.