TITLE-Characterization of a naturally occurring breast cancer subset enriched in EMT and stem cell characteristics

Running title-Metaplastic breast cancer characterization

Bryan T. Hennessy[1-3]*^, Ana-Maria Gonzalez-Angulo[2-4*], Katherine Stemke-Hale[2,3], Michael Z. Gilcrease[5], Savitri Krishnamurthy[5], Ju-Seog Lee[2], Jane Fridlyand[6], Aysegul Sahin[5], Roshan Agarwal[2], Corwin Joy[7], Wenbin Liu[7], David Stivers[7], Keith Baggerly[7], Mark Carey[2,3], Ana Lluch[8], Carlos Monteagudo[9], Xiaping He[10], Victor Weigman[10], Cheng Fan[10], Juan Palazzo[11], Gabriel N. Hortobagyi[4], Laura K. Nolden[2], Nicholas J. Wang[6], Vicente Valero[4], Joe W. Gray[6], Charles M. Perou[10"], Gordon B. Mills[2,3"]

*co-first;"co-senior

^Corresponding author: Dept. of Gynecologic Medical Oncology, M.D.Anderson Cancer Center (MDACC), 1515 Holcombe Blvd., Houston, TX77030; Ph.-7135631792/Fax-7137451541; Email-bhennessy@mdanderson.org

1. Gynecologic Medical Oncology, MDACC
2. Systems Biology, MDACC
3. Kleberg Center for Molecular Markers, MDACC
4. Breast Medical Oncology, MDACC
5. Pathology, MDACC
6. Lawrence Berkeley National Laboratory, Berkeley, CA
7. Bioinformatics and Computational Biology , MDACC
8. Clinic Hospital, Valencia, Spain
9. University of Valencia, Spain
10. Lineberger Comprehensive Cancer Center, Chapel Hill, NC

11. Department of Pathology, Thomas Jefferson University, Philadelphia, PA

Abstract

Metaplastic breast cancers (MBCs) are aggressive, chemoresistant tumors characterized by lineage plasticity. To advance understanding of their pathogenesis and relatedness to other breast cancer subtypes, twenty-eight MBCs were compared with common breast cancers using comparative genomic hybridization, transcriptional profiling, reverse phase protein arrays and by sequencing for common breast cancer mutations. MBCs demonstrated unique DNA copy number aberrations compared with common breast cancers. *PIK3CA* mutations were detected in 9/19 MBCs (47.4%) versus 80/232 hormone receptor-positive cancers (34.5%; p=0.32), 17/75 *HER2*-positive samples (22.7%; p=0.04), 20/240 basal-like cancers (8.3%; p<0.0001) and 0/14 claudin-low tumors (p=0.004). Of 7 PI3K/AKT pathway phosphorylation sites, six were more highly phosphorylated in MBCs than in other breast tumor subtypes. The majority of MBCs displayed mRNA profiles different from those of most common including basal-like cancers. By transcriptional profiling, MBCs and the recently identified claudin-low breast cancer subset constitute related receptor-negative subgroups characterized by low expression of GATA3-regulated genes and of genes responsible for cell-cell adhesion with enrichment for markers linked to stem cell function and epithelial-mesenchymal transition (EMT). In contrast to other breast cancers, claudin-low tumors and most MBCs showed a significant similarity to a "tumorigenic" signature defined using CD44+/CD24- breast tumor-initiating stem cell-like cells. MBCs and claudin-low tumors are thus enriched in EMT and stem cell-like features, and may arise from an earlier, more chemoresistant breast epithelial precursor than basal-like or luminal cancers. *PIK3CA*

mutations, EMT and stem cell-like characteristics likely contribute to the poor outcomes of MBC and suggest novel therapeutic targets.

INTRODUCTION

Metaplastic breast carcinomas (MBCs) are aggressive estrogen receptor (ERα)-negative, progesterone receptor (PR)-negative, HER2-negative (triple-negative) tumors characterized by mesenchymal/sarcomatoid and/or squamous metaplasia of malignant breast epithelium (1-7). Because of limited understanding of their pathogenesis, MBCs are treated in the same fashion as basal-like or triple receptor-negative ductal cancers. However, while neoadjuvant chemotherapy is associated with high pathologic complete response rates in basal-like carcinomas, MBCs are usually chemoresistant (2).

Transcriptional profiling has defined breast cancer subtypes (8,9). The origin of luminal A and B tumors appears to be the mammary duct luminal epithelium with concomitant hormone receptor expression. Elevated HER2 expression defines a subgroup with a poor prognosis; however, the responsiveness of this subgroup to trastuzumab improves outcomes (10). In contrast, basal-like cancers likely represent multiple different subtypes arising from distinct precursor cells from those of other cancers. Some basal-like breast cancers likely arise from mammary myoepithelial cells. To date, basal-like cancers have not presented specific therapy targets.

As MBCs are triple-negative, they are distinct from luminal and *HER2*-amplified cancers. As they express some markers associated with basal-like cancers (e.g. epidermal growth factor receptor (EGFR), cytokeratins 5/6), MBCs are proposed to represent a form of basal-like breast cancer. However, distinct clinical features such as chemoresistance suggest that MBCs may represent a unique subtype (2,3).

We applied an integrated genomic-proteomic approach to determine mechanisms underlying metaplastic carcinogenesis and MBC chemoresistance along with the relatedness of MBCs to known breast cancer subtypes. Most MBCs showed a unique molecular profile and form a distinct subtype most closely related to a novel subset of receptor-negative breast cancers (claudin-low) characterized by loss of genes involved in cell-cell adhesion. An enrichment for stem cell-like and epithelial-mesenchymal transition (EMT) markers in MBCs (and claudin-low tumors) along with frequent genomic aberrations that activate the phosphatidylinositol-3-kinase (PI3K)/AKT pathway suggest reasons for MBC chemoresistance and that MBCs and claudin-low tumors may arise from more immature precursor cells than other breast cancers.

MATERIALS AND METHODS

Human tumors

Twenty-eight frozen grade 3 MBCs with sarcomatoid (19) or squamous (9) metaplasia were obtained from the Breast Tumor Bank at M.D.Anderson Cancer Center (MDACC) and from a collaborator in Valencia (A.L.). The diagnosis was reconfirmed by pathologists at MDACC (M.G./S.K.) (2,3). Frozen tissue was used for DNA extraction (28 tumors) and, where adequate frozen tumor tissue remained, for RNA and protein extraction (16 MDACC tumors) (11).

Three tumor cohorts were used for comparison to MBCs (Supplemental Figure 1). The first cohort, used for comparison of mutation frequency (547 tumors) and functional

proteomic profiles (693), was composed of 693 frozen primary breast tumors obtained under IRB-approved protocols from MDACC. These tumors were subdivided into clinically-defined subtypes as described previously (Table 1) (12).

A second cohort of 145 primary breast tumors was used for comparison to MBC gene copy number profiles herein (13,14). A third cohort (Lineberger Comprehensive Cancer Center (LCCC)) of 184 breast tumors and 9 normal breast tissues was used for comparison with MBC transcriptional profiles (8,9,15). There were no statistically significant differences in the proportion of patients with tumors of different stages between the cohorts.

Comparative genomic hybridization (CGH)

CGH profiles from the 28 MBCs were generated at Lawrence Berkeley National Laboratory (LBNL) using single nucleotide polymorphism (SNP)-based GeneChip® Human Mapping 50K Sty arrays (Affymetrix, Santa Clara, CA) and compared to BAC-CGH profiles of primary breast tumors previously generated and processed (J.F.) at LBNL using HumArray1.14/HumArray2.0 (13,14,16-18). MBC 50K data are available[i].

For comparison with LBNL tumors, the 28 MBC SNP chips were mapped to BAC resolution. This approach has been validated by comparing data derived using both platforms to analyze breast cancer cell lines (not shown). LBNL tumors were remapped to the May04 freeze from UCSC and regions around each BAC clone were defined as within a half distance to each neighboring clone or to the beginning or end of the

[i] **ftp://beamish.lbl.gov/njwang/**

chromosome if telomeric. A median expression value was then obtained for SNPs in each BAC region. Missing values were assigned if <5 SNPs mapped to a particular region. Each array was recentered to have a median of 0. The resulting values were segmented using Circular Binary Segmentation (CBS) followed by a merge-level procedure to combine segmented levels across the genome. Each missing value was assigned the value of its corresponding segment. Gain/loss events and fraction of genome altered were calculated. After this resolution reduction (median=18 SNPs/BAC;mean=30), the mean variability estimate was 0.25. Similar analyses beginning with the CBS steps were performed on the original dChip processed data. We used a Fisher's test to measure the difference in copy number at probes on each side of genes encoding PI3K/AKT pathway components. These p-values were used to fit a beta-uniform mixture (BUM) model to determine significance at a given false discovery rate (FDR).

To directly compare the 50K SNP and older BAC platforms, DNA extracted from five MBCs was also run using the BAC platform. This confirmed a high concordance for the matched data derived from the two platforms (not shown).

Detection of mutations

DNA was extracted from 547 MDACC breast tumors along with 14 LCCC claudin-low breast tumors and 19 MBCs with sufficient remaining DNA for mutation detection (9,11,12). Following whole genome amplification, *p53/PTEN* genes were resequenced (19). *CTNNB1* exon 3 (the most common site of mutations) was amplified from genomic DNA using a forward primer located at the 5' portion and a reverse primer at the 3' end of

the exon. A tumor sample with a known *CTNNB1* mutation was amplified and sequenced in parallel with tumor samples as a positive control. A SNP-based approach (Sequenom (San Diego, CA) MassArray) was used to detect mutations in *PIK3CA*, *KRAS* and E17K mutations in the *AKT1/2/3* genes (12,20). This approach is unsuitable for detection of mutations that are not 'hot spot' mutations but is particularly suitable to mutation detection in breast cancer where stromal 'contamination' is prevalent (21).

Reverse phase protein lysate array (RPPA)

RPPA was applied with the antibodies in Supplemental Table 1 to compare PI3K/AKT and mitogen activated protein kinase (MAPK) pathway activation in protein lysates derived from 16 MBCs versus 693 common breast cancers (Supplemental Table 2) (22-25). The expression of each antibody in a sample was corrected for protein loading using the average expression levels of all probed proteins.

Transcriptional profiling

Total RNA was isolated by phenol-chloroform extraction (Trizol, GIBCO/BRL), and mRNA was purified by either magnetic separation using Dynabeads (Dynal) or the Invitrogen FastTrack 2.0 Kit. Twelve of 16 MBC RNA samples with RNA integrity numbers (RINs)>6 were assayed on Agilent oligomicroarrays at LCCC and compared with a published Agilent microarray data set also previously assayed and processed (C.M.P.) at LCCC (8,9,15). The microarray and clinical data are available at UNC Microarray Database and in the Gene Expression Omnibus (GSE10885). Expression

Analysis Systematic Explorer (EASE) was applied to perform functional analysis of gene lists.

Mapping gene expression onto regions of MBC copy number change

Using a Significance Analysis of Microarray (SAM)-defined list of MBC-defining genes, we determined the chromosomal location of each gene to link with the CGH data. Probes with an undefined chromosomal position were discarded from further analyses. CBS was applied to the pre-processed MBC copy number data to determine breakpoints for aberrations (26). The CBS calls made were as follows: class.segment<- segment (class.cna, alpha=0.05, p.method="perm", nperm=1000, trim= 0.05, undo.splits="sdundo", undo.SD=2, verbose=2). Using CBS output, a plot of segment intensities versus segment markers was used to determine an intensity boundary threshold of 0.12. Segments with intensity values beyond this threshold were flagged as gained or lost based on the sign of intensity and parsed from the original CBS output. By applying a customized R script to this output, segments from each sample were collated and regions were assigned that had varying levels of overlap between the MBC patients. Cutoffs were made for regions with aberrations in ≥1/3 of MBCs tested. SAM genes with chromosomal locations that were contained within these gains and losses were determined and plotted.

Comparison of the MBC and claudin-low transcriptional profiles with a $CD44^{+}/CD24^{-/low}$ breast cancer cell profile

We compared breast tumor transcriptional signatures with a "tumorigenic" signature (Creighton et al., submitted) that was derived by comparing gene expression profiles of flow-sorted $CD44^{+}/CD24^{-/low}$ cancer cells with profiles of all other sorted cells (CD44-/CD24+ and CD44-/CD24- combined). For each tumor, an "R-value" was derived in relation to the "tumorigenic" signature, which was defined as the Pearson's correlation between the "tumorigenic" gene signature pattern (using "1" and "-1", for up and down, respectively) and tumor expression values. Tumors with high R-values would tend to have both high expression of many of the genes high in "tumorigenic" cells and low expression of many of the genes low in "tumorigenic" cells (and vice versa for tumors with low R-values).

STATISTICAL ANALYSIS

R[ii] and NCSS/PASS software (Kaysville, Utah) were used. Reported p-values are two sided. ANOVA and T-tests for gene expression data were performed using SAS. For clustering, we used CLUSTER and TREEVIEW (University of Glasgow, Scotland) softwares.

RESULTS

MBCs possess patterns of DNA copy number gains and losses that are distinct from those in common breast cancers

MBCs demonstrated a high level of genomic instability based on fraction of the genome altered and number of transitions. However, MBCs demonstrated a unique set of aberrations compared with common breast cancers (Figure 1) (13,14). Specifically, gains

[ii] http://cran.r-project.org

of distal chromosome 1p/5p and loss of 3q were common in MBCs but rare in other breast cancers (13,14). Conversely, alterations that occur in most breast cancers, such as gain of chromosome 1q and loss of 16q, were uncommon in MBC (13,14). In particular, as compared with basal-like tumors, MBCs exhibited more frequent amplification of 1p/11q/12q/14q/19p/19q/22q and increased frequency of loss at 1q/2p/3q/8q. There was retention of 5q/9q/15q/16p/17p/17q/19p/19q/20q/22q as compared with basal-like tumors. Overall, MBCs did not display similar alterations to basal-like cancers and showed substantial differences from common breast cancers, compatible with MBCs representing a distinct subgroup.

MBCs possess distinct patterns of somatic mutations from basal-like breast cancers

PIK3CA mutations were detected in 9/19 MBCs (47.4%) as compared to 80/232 hormone receptor-positive cancers (34.5%;p=0.32), 17/75 *HER2*-amplified samples (22.7%;p=0.04), 20/240 triple-negative cancers (8.3%;p<0.0001) and 0/14 claudin-low tumors (p=0.004) (Table 1) (12). The *PIK3CA* mutation frequencies in the common subtypes are compatible with those frequencies reported in the literature, with the exception of that in claudin-low tumors which has not been reported (27). One *PTEN* mutation was detected in a MBC (5%) that did not have a *PIK3CA* mutation. *p53* mutations were detected in 6/19 MBCs (32%). No mutation in exon 3 of *CTNNB1* was detected in 19 MBCs.

Strikingly, therefore, 10/19 (53%) MBCs demonstrated PI3K/AKT pathway mutations. This was significantly different from the mutation rate in triple-negative/basal-like

cancers in particular. Further, other PI3K/AKT pathway genomic aberrations were more frequent in MBCs. Based on two SNP probes closest to each end of genes encoding PI3K/AKT components, *AKT1* (chromosome 14), *AKT2* (chromosome 19) and *RPS6KB2* (p70S6K, chromosome 11) demonstrated more frequent copy number gain (Fisher's exact test $p<0.05$ at a 1% FDR) in MBCs compared with other subtypes. This was also supported at the protein level using RPPA data (Supplemental Table 2). Thus, PI3K/AKT pathway aberrations likely play a major role in MBC pathophysiology, further suggesting that MBCs and basal-like tumors are distinct.

PI3K pathway activation in MBC

The high frequency of genomic aberrations in PI3K/AKT pathway genes implicates this pathway in MBC pathogenesis. This pathway has already been implicated in breast cancer resistance to multiple therapies (28-30). Using RPPA, phosphorylation of most core PI3K/AKT pathway proteins was elevated in MBCs compared with at least one other breast tumor subtype (Figure 2), with the exception of phosphorylated p70S6K. The most likely explanation for the latter discrepancy is phosphorylation of p70S6K by kinases other than core PI3K/AKT pathway kinases (31,32). Overall, however, key PI3K/AKT pathway components are generally more highly phosphorylated in MBCs than in most other breast tumors, which parallels the results of the genomic analyses and could contribute to the poor outcomes associated with MBC (Supplemental Figure 2).

Only GSK3 phosphorylation was higher in MBCs possessing mutant versus wild-type *PIK3CA*/*PTEN* genes (P=0.01). The failure to demonstrate an association between

PI3K/AKT pathway mutations and activation is potentially due to the small number of tumors analyzed. However, alterations in PI3K/AKT pathway activation by processes independent of mutations, by other interacting pathways or by signaling modulation through feedback loops may have prevented identification of statistically significant associations (33). Compatible with the latter contention, we have previously demonstrated that *PIK3CA* mutational status is not correlated with AKT phosphorylation in hormone receptor-positive breast cancers or cell lines (12).

We also quantified expression and phosphorylation of several PI3K/AKT pathway-activating and -related proteins (Supplemental Table 2). EGFR expression was lower in MBCs compared with other breast cancer subtypes. HER2 levels were also significantly decreased in MBCs relative to hormone receptor-positive, triple negative and in particular HER2-positive tumors. Cyclin E1, a PI3K/AKT pathway target, was present at higher levels in MBCs compared with hormone receptor-positive and HER2-positive tumors (P=0.0005, 0.02, respectively) (34,35). The Y box-binding protein 1 (YB1) has been implicated in chemoresistance and is located in the chromosome 1p amplicon in MBC (Figure 1) (36). Indeed *YB1* was more frequently amplified (at 1% FDR rate) in MBCs than in other tumors, and YB1 protein expression was also higher in MBCs.

MBC transcriptional profiles are distinct from those of basal-like breast cancers and related to those of claudin-low breast tumors

On unsupervised hierarchical clustering, the majority of MBCs displayed markedly different mRNA profiles from those of most common breast cancers including basal-like

cancers (not shown). To explore the relationships between MBCs and the intrinsic breast cancer subtypes, twelve MDACC MBCs were compared with 184 breast tumors and 9 normal breast samples by hierarchical clustering using a combination of four intrinsic gene lists (8,9,15,37). MBCs were somewhat heterogeneous in this analysis (Supplemental Figure 3). Two MBCs clustered with basal-like tumors, two with a novel subtype of receptor-negative tumors that is characterized by loss of a cluster of genes that encode proteins involved in cell-cell adhesion (claudin-low tumors), two clustered within the normal-like group, and six formed a novel subgroup with characteristics intermediate between those of basal-like and claudin-low tumors (Supplemental Figure 3).

To assess the significance of this clustering pattern, we applied "SigClust" to test the null hypothesis that any group of samples contained within a common dendrogram branch constitutes a single group (38). This analysis showed that the dendrogram branch containing 6 MBCs, along with some previously assayed tumors, represents a distinct group. Upon re-analysis of the histology of the latter tumors, four showed metaplastic and/or spindloid features compatible with the tumors representing MBCs (Supplemental Figure 4). Further, mouse mammary tumors that are similar to human claudin-low tumors demonstrate a spindloid morphology (9).

The gene set that defines claudin-low and MBC tumors (Supplemental Figure 3C) was determined by Gene Ontology (GO) analysis to be enriched for the terms “tight junction”, “intercellular junction”, “apicolateral plasma membrane” and “cell junction”. Of 29 claudin-low cluster genes, 13 are positively regulated by GATA3 and none by ERα

(hypergeometic mean p-value<0.01) (39). In addition to lacking genes involved in cell-cell adhesion and polarity, MBCs and claudin-low tumors lack luminal genes including *GATA3* (Supplemental Figure 3F) and *HER2*, and demonstrate inconsistent expression of genes associated with basal-like tumors (Supplemental Figure 3D). T-tests confirmed that the expression of the claudin-low cluster of genes (Supplemental Figure 3C) was lower in MBCs versus other breast cancers (Supplemental Table 3).

MBCs and claudin-low tumors express high levels of stem cell and EMT markers

SAM was utilized to identify a MBC vs. common breast tumor expression signature (40,41). This analyses resulted in 556 'up' and 373 'down' genes, with a FDR of <1 gene (Supplemental Table 4). Almost 33% of the 'SAM up' and 50% of the 'SAM down' genes mapped to regions of copy number aberration (CNA-see below). In a GO analysis using EASE, the top six enriched biological processes in the 'up' gene list were "cell communication", "cell adhesion", "signal transduction", "cellular process", "cell-cell adhesion" and "intracellular protein transport", while "protein transport", "intracellular transport", "transport", "male meiosis" and "ubiquitin-dependent protein catabolism" were enriched in the 'down' gene list (40). The claudin-low gene cluster showed a statistically significant overlap with the SAM-defined 'down' genes (15/29 genes, hypergeometic mean analysis p-value<0.001). Genes near the top of the MBC 'up' list that have been previously implicated in carcinogenesis included *ALK*, Crystallin γ and the master regulator of epithelial-mesenchymal transition (EMT), *TWIST1* (42-45).

Breast cancers are thought to contain a minority population of tumor initiating/stem cell-like cells with high CD44 but low or undetectable levels of CD24 (CD44+/CD24-); these cells have higher tumorigenic capacity than other purified populations of tumor-derived cells (46). Their phenotype and the low responsiveness of MBCs to chemotherapy suggest that MBCs might possess stem cell-like characteristics. Indeed, MBCs had markedly elevated CD44/CD24 and CD29/CD24 ratios compared to other breast cancers, with the exception of claudin-low tumors (Figure 3). This "electronic stem cell signature" is also differentially expressed between fluorescence-activated cell sorted (FACS) human breast tumor-initiating cells versus normal breast epithelial cells (Supplemental Figure 5) (46,47).

EMT is characterized by the up-regulation of vimentin and of E-cadherin-repressor molecules (snail/slug/twist) with downregulation of E-cadherin and other cell adhesion molecules (45,48,49). These events occur in MBCs and in claudin-low tumors (Figure 3/Supplemental Table 3). In MBCs, *TWIST1* and snail homolog 2 (*SNAI2*/*SLUG*) were expressed at high levels, while *SNAI3* was overexpressed in claudin-low tumors. Thus, claudin-low tumors and MBCs may be enriched for stem cell-like and EMT markers, features that may contribute to poor patient outcomes (Supplemental Figure 2) (2,3).

Genes that are altered at the genomic and transcriptional levels in MBC

Three hundred and six genes (Supplemental Table 5) in the SAM-derived MBC transcriptome localized to areas of chromosomal gain and loss in at least 33% of MBCs (Supplemental Figure 6a,b). Thus, these genes demonstrate coordinate changes at the

DNA and RNA levels in MBC. Functional analysis of these genes (Supplemental Tables 6 and 7) demonstrated that, among genes that are amplified and overexpressed in MBCs, components of three major branches (JNK, MAPK, p38) that compose the MAPK signaling pathway are significantly overrepresented in MBCs. Compatible with this finding, phosphorylation of three of four assessed protein components of these pathways (JNK (p=0.06), MEK (p=0.003) and p38 (p=0.0008) but not ERK1/2 (p=NS)) was higher in MBCs versus all other breast cancers.

Comparison of the MBC and claudin-low transcriptional profiles with a $CD44^{+}/CD24^{-/low}$ breast cancer cell profile

Given their enrichment for stem cell markers, we compared the transcriptional signatures of MBCs and of other breast cancers with a "tumorigenic" signature (Creighton et al., submitted) that was derived by comparing gene expression profiles of flow-sorted $CD44^{+}/CD24^{-/low}$ breast tumor cells with profiles of all other sorted cells (CD44-/CD24+ and CD44-/CD24-). In contrast to other breast cancers, except for tumors of the "claudin-low" subtype, most of the MBCs showed a clear association with the "tumorigenic" signature (Figure 4). Further, of 373 and 217 down-regulated genes in the MBC and "tumorigenic" signatures, respectively, there were 29 shared genes (Supplemental Table 8, $p=1x10^{-13}$ for the overlap); 5 of these 29 genes were components of the claudin-low gene cluster. In addition, as has been demonstrated with the claudin-low signature (Creighton et al, submitted), the MBC signature is enriched in post-docetaxel and post-letrozole treatment specimens (Supplemental Figure 7). These data collectively suggest that, at diagnosis, MBCs and claudin-low tumors possess transcriptional features that are

enriched in highly purified breast tumor-initiating and chemoresistant breast cancer cell fractions[50], the latter also compatible with an enrichment for stem cell-like activity.

DISCUSSION

MBCs are aggressive, chemoresistant tumors associated with poor outcomes (2,3). Although uncommon, MBCs account for several hundred new breast cancer cases every year in the U.S thus representing a therapeutic dilemma for oncologists. With only retrospective case reviews as a basis for making recommendations, it has not been possible to define therapy guidelines. Thus, we sought to determine the relationship of MBCs to common breast cancers, and in particular to basal-like breast cancers given the common assumption that MBCs are basal-like cancers. We also sought to determine whether the underlying pathophysiology of MBCs would result in the identification of new drug targets.

Supplemental Figure 8 summarizes the features of MBC defined in this study that have potential clinical and therapeutic utility. Due to low expression of hormone receptors and HER2, as well as expression of some basal epithelial markers, MBCs have been proposed to represent a form of basal-like breast cancer (5). However, based on the integrated analyses herein, most MBCs likely represent an independent subtype that is distinct from basal-like cancers. Their transcriptional profiles are most closely related to claudin-low cancers, a novel subgroup of receptor-negative breast cancers that are clearly different from basal-like cancers (Supplemental Figure 3). CGH profiles, their enrichment for

stem-cell like markers and their PI3K/AKT pathway activation status also differentiate MBCs from basal-like cancers. MBCs, like claudin-low cancers, express high levels of EMT markers and demonstrate elevated CD44/CD24 and CD29/CD24 ratios, which have been proposed to represent breast cancer stem cell-like markers (46). Indeed, a recent study detected a direct and causative link between EMT and the gain of epithelial stem cell properties (51). These features likely contribute to the lineage plasticity of MBCs on light microscopy and to their limited chemoresponsiveness (2,3). Claudin-low features, including EMT and stem cell-like properties, also potentially contribute to the aggressive phenotype of MBCs. This is supported by the significant overlap between the MBC signature and the "tumorigenic" signature (Creighton et al., submitted), with overlapping genes including five of the claudin-low genes (Supplemental Figure 3C). The MBC, the "tumorigenic" and the claudin-low signatures are all enriched in residual post-treatment chemoresistant breast tumors. Thus, MBCs and claudin-low breast tumors may arise from a more primitive and chemoresistant "stem" cell than luminal or basal-like tumors.

The pattern of chromosomal gains and losses in MBCs is distinct from that in other breast cancers including basal-like cancers. This unique pattern suggests that the processes underlying metaplastic carcinogenesis are distinct from those associated with other breast cancer subtypes. MBCs demonstrate a high frequency of mutation, amplification and activation of PI3K/AKT pathway components. This is markedly different from basal-like breast cancers, where we and others have shown that PI3K/AKT pathway genomic mutations are uncommon[iii] (11,27). We also did not detect *PIK3CA* mutations in 14 Claudin-low breast cancers pointing to differences between claudin-low breast cancers

[iii] **www.sanger.ac.uk**

and MBCs. The frequency of *PIK3CA* and *PTEN* mutations combined with amplification of *AKT* and *p70S6K* suggest that PI3K/AKT pathway activation is critical to metaplastic carcinogenesis. Activation of this pathway, along with enrichment for tumor-initiating/stem cells, may underlie the chemoresistance and poor outcomes associated with MBC (28-30). MBCs also demonstrate a high frequency of amplification, overexpression and activation of MAPK pathway components. As particularly important genes and targets in cancer are likely to be aberrant at the level of the genome, transcriptome, and proteome, the PI3K/AKT and MAPK pathways are therefore potentially attractive therapy targets in MBC. As inhibitors of the PI3K/AKT and MAPK pathways are now in clinical trials, it will be of interest to determine whether these inhibitors will sensitize MBCs to cytotoxic drugs (52). As there are no MBC cell lines or animal models available, it will be necessary to develop a consortium approach to test this hypothesis in patients.

In contrast to a recently published study, mutations in exon 3 of *CTNNB1* were not identified in 19 MBCs in our study (53). This may in part relate to the fact that our study was restricted to high grade MBCs while the previous study included a significant proportion of lower grade MBCs, tumors that behave in a less aggressive fashion than high grade MBCs.

This study has several potential limitations. Although a subset (6/12 i.e. 50%) of MBCs constitute a significantly related group of tumors as defined by SigClust, it is clear that MBCs form a somewhat heterogeneous group of receptor-negative breast cancers in

terms of their molecular characteristics. Just as breast tumors that are defined as ductal are clearly heterogeneous based on receptor status and transcriptional profiling, it is not surprising that MBCs represent a molecularly heterogeneous group. MBCs are microscopically heterogeneous with this study being limited to tumors with squamous and sarcomatoid metaplasia (1-7). The rarity of MBC precluded analysis of other histologic variants. In addition, it is possible that MBC represents multiple different diseases. There were no clear correlation between the pattern of gene copy number change and the histologic appearance of the tumors analyzed.

The major conclusions of this manuscript are: 1. MBCs are molecularly distinct from other breast cancers, 2. despite their relative histologic uniformity, MBCs are molecularly heterogeneous, and 3. Claudin-low breast cancers are likely the most closely related ductal breast cancer subset to MBCs. The molecular mechanisms underlying metaplastic carcinogenesis are likely different from those associated with other breast cancer subtypes including basal-like cancers. By gene expression analysis, MBCs and claudin-low tumors share common features that suggest related cellular origins, potentially from a more primitive cell than that implicated as a precursor to luminal or basal-like tumors. It is likely that MBCs, and potentially claudin-low tumors, define a novel chemoresistant triple-negative breast cancer subgroup that exhibits a signature similar to that of breast tumor-initiating cells and of residual common breast tumor cells isolated after patient treatment. The frequency of PI3K/AKT pathway aberrations argues that this pathway should be explored as a therapeutic target in MBC. A challenge to advancing therapy for

MBC patients is the infrequency of this disease. However, a centralized clinical trial effort is a feasible venture that will improve patient outcomes.

Acknowledgements

Supported by Kleberg Center for Molecular Markers at MDACC, Cancer Center Support Grant (CA16672) at MDACC, The University of Texas M. D. Anderson Cancer Center Physician Scientist Program and The McNair Scholars Program supported by The Robert and Janice McNair Foundation to BTH, NCI K23-CA121994, R21-CA120248 and NCI Breast SPORE program (P50-CA116199) to AMG, NCI CA116199 and CA099031 to GBM, Komen Foundation grant FAS0703849 to GBM, AMG and BTH, by NCI Breast SPORE program (P50-CA58223-09A1), RO1-CA-101227-01, the V Foundation for Cancer Research, and the Breast Cancer Research Foundation to CMP and by the Director, Office of Science, Office of Biological & Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, by the National Institutes of Health, National Cancer Institute grants P50 CA 58207, and the U54 CA 112970 grants to JWG.

We thank Len Pennacchio and Jan-Fang Cheng of Lawrence Berkelely National Laboratory for Sanger sequencing. We also thank Jenny Chang and Chad Creighton for sharing data and providing analysis concerning their tumorigenic signature in Creighton et al Submitted.

REFERENCES

1. Tavassoli FA. (1999). Pathology of the Breast, 2nd edn (Hong Kong: Appleton and Lange), 1999.
2. Hennessy BT, Giordano S, Broglio K, et al. Biphasic metaplastic sarcomatoid carcinoma of the breast. Ann Oncol 2006;17:605-13.
3. Hennessy BT, Krishnamurthy S, Giordano S, et al. Squamous cell carcinoma of the breast. J Clin Oncol 2005;23:7827-35.
4. Reis-Filho JS, Milanezi F, Steele D, et al. Metaplastic breast carcinomas are basal-like tumours. Histopathology 2006;49:10-21.
5. Wargotz ES, Deos PH, Norris HJ. Metaplastic carcinomas of the breast. II. Spindle cell carcinoma. Hum Pathol 1989;20:732–40.
6. Foschini MP, Dina RE, Eusebi V. Sarcomatoid neoplasms of the breast: proposed definitions for biphasic and monophasic sarcomatoid mammary carcinomas. Semin Diagn Pathol 1993;10:128–36.
7. Gutman H, Pollock RE, Janjan NA, Johnston DA. Biologic distinctions and therapeutic implications of sarcomatoid metaplasia of epithelial carcinoma of the breast. J Am Coll Surg 1995;180:193–9.
8. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001;98:10869–74.

9. Herschkowitz JI, Simin K, Weigman VJ, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol 2007;8:R76.
10. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 2005;353:1673-84.
11. Pinkel D, Segraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 1998;20:207-11.
12. Stemke-Hale K, Gonzalez-Angulo AM, Lluch A, et al. An integrative genomic and proteomic analysis of PIK3CA, PTEN and AKT mutations in breast cancer. Cancer Res 2008;68:6084-91.
13. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 2006;10:529-41.
14. Fridlyand J, Snijders AM, Ylstra B, et al. Breast tumor copy number aberration phenotypes and genomic instability. BMC Cancer 2006;6:96.
15. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 2006;7:96.
16. Snijders AM, Schmidt BL, Fridlyand J, et al. Rare amplicons implicate frequent deregulation of cell fate specification pathways in oral squamous cell carcinoma. Oncogene 2005;24:4232-42.
17. Snijders AM, Nowak N, Segraves R, et al. Assembly of microarrays for genome-wide measurement of DNA copy number. Nat Genet 2001;29:263-4.

18. Janne PA, Li C, Zhao X, et al. High-resolution single-nucleotide polymorphism array and clustering analysis of loss of heterozygosity in human lung cancer cell lines. Oncogene 2004;23:2716–26.
19. Tartaglia M, Pennacchio LA, Zhao C, et al. Gain-of-function SOS1 mutations cause a distinctive form of Noonan syndrome. Nat Genet 2007;39:75-9.
20. Thomas RK, Baker AC, Debiasi RM, et al. High-throughput oncogene mutation profiling in human cancer. Nat Genet 2007;39:347-51.
21. West RB, Nuyten DS, Subramanian S, et al. Determination of stromal signatures in breast carcinoma. PLoS Biol 2005;3:e187.
22. Tibes R, Qiu Y, Lu Y, Hennessy B, Mills GB, Kornblau S. Reverse phase protein array (RPPA): Validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. Mol Cancer Ther 2006;5:2512-21.
23. Liang J, Shao SH, Xu ZX, et al. The energy sensing LKB1-AMPK pathway regulates p27kip1 phosphorylation mediating the decision to enter autophagy or apoptosis Nat Cell Biol 2007;9:218-24.
24. Hu J, He X, Baggerly KA, Coombes KR, Hennessy BT, Mills GB. Non-parametric quantification of protein lysate arrays. Bioinformatics 2007;23:1986-94.
25. Hennessy BT, Lu Y, Poradosu E, et al. Quantified pathway inhibition as a pharmacodynamic marker facilitating optimal targeted therapy dosing: Proof of principle with the AKT inhibitor perifosine. Clin Cancer Res 2007;13:7421-31.

26. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 2007;23:657-63.
27. Saal LH, Holm K, Maurer M, et al. PIK3CA mutations correlate with hormone receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma. Cancer Res 2005;65:2554-9.
28. Stassi G, Garofalo M, Zerilli M, et al. PED mediates AKT-dependent chemoresistance in human breast cancer cells. Cancer Res 2005;65:6668-75.
29. Mondesire WH, Jian W, Zhang H, Ensor J, Hung MC, Mills GB, Targeting mammalian target of rapamycin synergistically enhances chemotherapy-induced cytotoxicity in breast cancer cells. Clin Cancer Res 2004;10:7031-42.
30. Knuefermann C, Lu Y, Liu B, et al. HER2/PI-3K/Akt activation leads to a multidrug resistance in human breast adenocarcinoma cells. Oncogene 2002;22:3205-12.
31. Le Pabic H, L'Helgoualc'h A, Coutant A, et al. Involvement of the serine/threonine p70S6 kinase in TGF-beta1-induced ADAM12 expression in cultured human hepatic stellate cells. J Hepatol 2005;43:1038-44.
32. Grove JR, Price DJ, Banerjee P, Balasubramanyam A, Ahmad MF, Avruch J. Regulation of an epitope-tagged recombinant Rsk-1 S6 kinase by phorbol ester and erk/MAP kinase. Biochemistry 1993;32:7727-38.
33. Wan X, Harkavy B, Shen N, Grohar P, Helman LJ. Rapamycin induces feedback activation of Akt signaling through an IGF-1R-dependent mechanism. Oncogene 2007;26:1932-40.

34. Lee SR, Park JH, Park EK, Chung CH, Kang SS, Bang OS. Akt-induced promotion of cell-cycle progression at G2/M phase involves upregulation of NF-Y binding activity in PC12 cells. J Cell Physiol 2005;205:270-7.
35. Reichert M, Saur D, Hamacher R, Schmid RM, Schneider G. Phosphoinositide-3-kinase signaling controls S-phase kinase-associated protein 2 transcription via E2F1 in pancreatic ductal adenocarcinoma cells. Cancer Res 2007;67:4149-56.
36. Huang J, Tan PH, Li KB, Matsumoto K, Tsujimoto M, Bay BH. Y-box binding protein, YB-1, as a marker of tumor aggressiveness and response to adjuvant chemotherapy in breast cancer. Int J Oncol 2005;26:607-13.
37. Perreard L, Fan C, Quackenbush JF, et al. Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. Breast Cancer Res 2006;8:R23.
38. Liu Y, Hayes DN, Nobel A, Marron JS. Statistical Significance of Clustering for High Dimension Low Sample Size Data. Journal of the American Statistical Association 2008; 103:1281-93.
39. Oh DS, Troester MA, Usary J, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. J Clin Oncol 2006;24:1656-64.
40. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116-21.
41. Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003;4:P3.
42. Chiarle R, Voena C, Ambrogio C, Piva R, Inghirami G. The anaplastic lymphoma kinase in the pathogenesis of cancer. Nat Rev Cancer 2008 ;8:11-23.

43. Sun H, Ma Z, Li Y, et al. Gamma-S crystallin gene (CRYGS) mutation causes dominant progressive cortical cataract in humans. J Med Genet 2005 ;42:706-10.
44. Moyano JV, Evans JR, Chen F, et al. AlphaB-crystallin is a novel oncoprotein that predicts poor clinical outcome in breast cancer. J Clin Invest 2006;116:261-70.
45. Yang J, Mani SA, Donaher JL, et al. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell 2004;117:927-39.
46. Liu R, Wang X, Chen GY, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. New Engl J Med 2007;356:217-26.
47. Takahashi K, Tanabe K, Ohnuki M, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 2007;131:861-72.
48. Shipitsin M, Campbell LL, Argani P, et al. Molecular definition of breast tumor heterogeneity. Cancer Cell 2007;11:259-73.
49. Lien HC, Hsiao YH, Lin YS, et al. Molecular signatures of metaplastic carcinoma of the breast by large-scale transcriptional profiling: identification of genes potentially related to epithelial-mesenchymal transition. Oncogene 2007;26:7859-71.
50. Li X, Lewis MT, Huang J, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J Natl Cancer Inst 2008;100:672-9.
51. Mani SA, Guo W, Liao MJ, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 2008;133:704-15.
52. Hennessy BT, Smith DL, Ram PT, Lu Y, Mills GB. Exploiting the PI3K/AKT pathway for cancer drug discovery. Nature Rev Drug Discov 2005;4:988-1004.

53. Hayes MJ, Thomas D, Emmons A, Giordano TJ, Kleer CG. Genetic changes of wnt pathway genes are common events in metaplastic carcinomas of the breast. Clin Cancer Res 2008;14:4038-44.

Table 1. Frequency of mutations in the *PIK3CA*, *AKT1, 2 and 3*, *PTEN* and *KRAS* genes in human breast cancers. The *PIK3CA*/*AKT1, 2, 3*/*PTEN* mutation frequency in hormone receptor-positive (HR+), HER2-positive and triple-negative (TN) cancers, used herein as a comparator for metaplastic and claudin-low tumors, has been published (12).

Tumor subtype	***PIK3CA* catalytic domain**	***PIK3CA* other**	***PIK3CA* total**	***PTEN***	***AKT1* E17K**	***AKT2/3* E17K**	***KRAS***
All human breast tumors excluding metaplastic and claudin-low cancers	**73/547 (13.3%)**	**44/547 (8.0%)**	**117/547 (21.4%)**	**2/88 (2.3%)**	**6/418 (1.4%)**	**0/418 (0%)**	**0/418 (0%)**
Human breast HR+	48/232 (20.7%)	32/232 (13.8%)	80/232 (34.5%)	2/58 (3.4%)	6/232 (2.6%)	0/232 (0%)	0/232 (0%)
Human breast HER2+	13/75 (17.3%)	4/75 (5.3%)	17/75 (22.7%)	0/10 (0%)	0/75 (0%)	0/75 (0%)	0/75 (0%)
Human breast TN	12/240 (5.0%)	8/240 (3.3%)	20/240 (8.3%)	0/20 (0%)	0/111 (0%)	0/111 (0%)	0/111 (0%)
Human breast claudin-low	0/14 (0%)	0/14 (0%)	0/14 (0%)	-	0/14 (0%)	0/14 (0%)	0/14 (0%)
Human breast metaplastic	4/19 (21.1%)	5/19 (26.3%)	9/19 (47.4%)	1/19 (5.3%)	0/19 (0%)	0/19 (0%)	1/19 (5.3%)

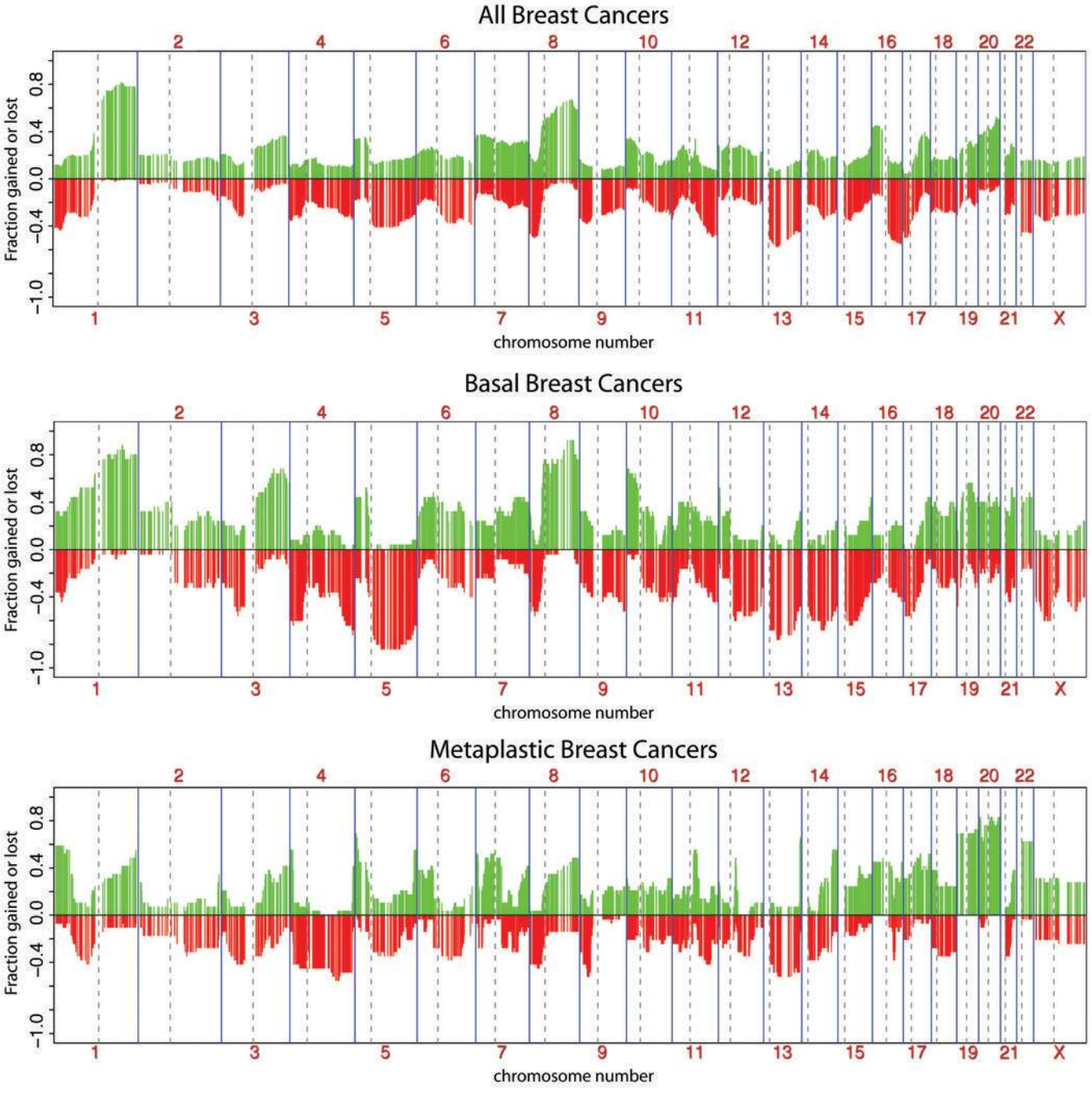
Figure legends:

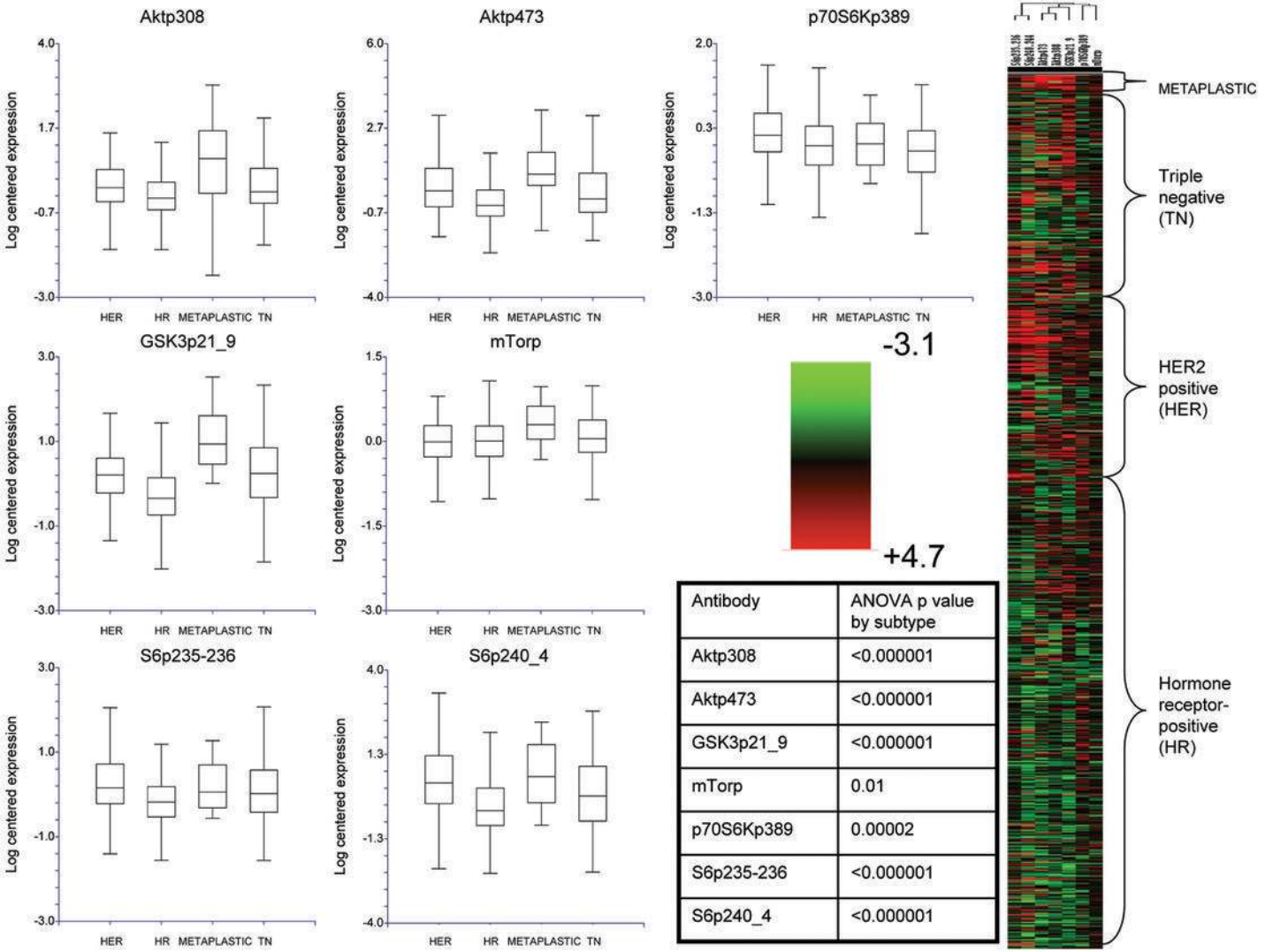
Figure 1. Gene copy number changes. Gains and losses in all common breast cancers (including luminal, *HER2*-amplified and basal-like cancers) and basal-like cancers alone versus metaplastic breast tumors were determined. The chromosomes are subdivided into arms and ordered from left to right, beginning with 1p, 1q and ending with X.

Figure 2. Functional proteomics of metaplastic breast cancer (MBC). The comparative expression of seven core phosphatidylinositol-3-kinase (PI3K)/AKT pathway phosphoproteins in 383 hormone receptor-positive (HR), 142 HER2-positive (HER), 168 triple-negative (TN) breast tumors and 16 MBCs was determined. ANOVA-analysis of variance; GSK3-glycogen synthase kinase 3; XpY-phosphorylation of protein X at amino acid(s) Y.

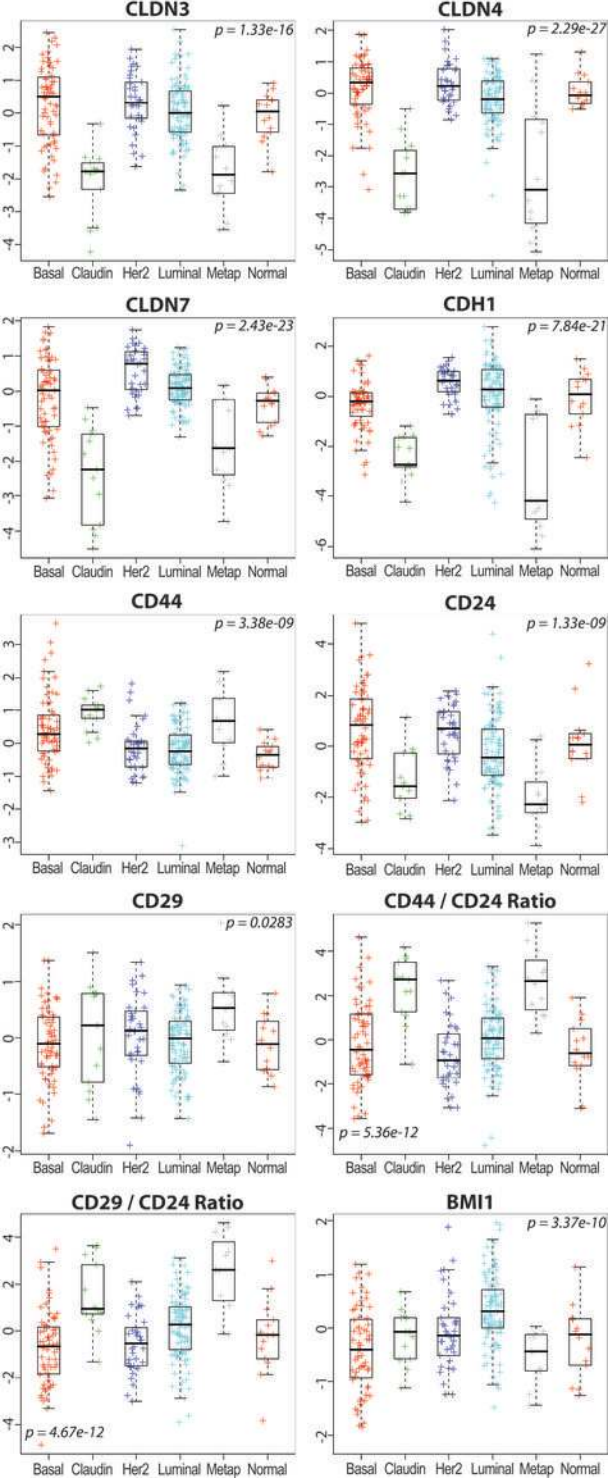
Figure 3. Expression of Claudin-Low and Stem Cell Markers in Breast Cancer Subgroups. Using data from transcriptional profiling, metaplastic (metap) and claudin-low tumors express low levels of claudins CLDN3, CLDN4 and CLDN7, of CDH1 (E-cadherin) and high CD44/CD24 and CD29/CD24 ratios. The P-values shown are from ANOVA (analysis of variance). The p-values for the metaplastic-basal comparison of stem cell markers were 0.41 (BMI1), 0.19 (CD44), 0.004 (CD29), 0.00009 (CD24), 0.00006 (CD44/CD24 ratio) and 0.00007 (CD29/CD24 ratio).

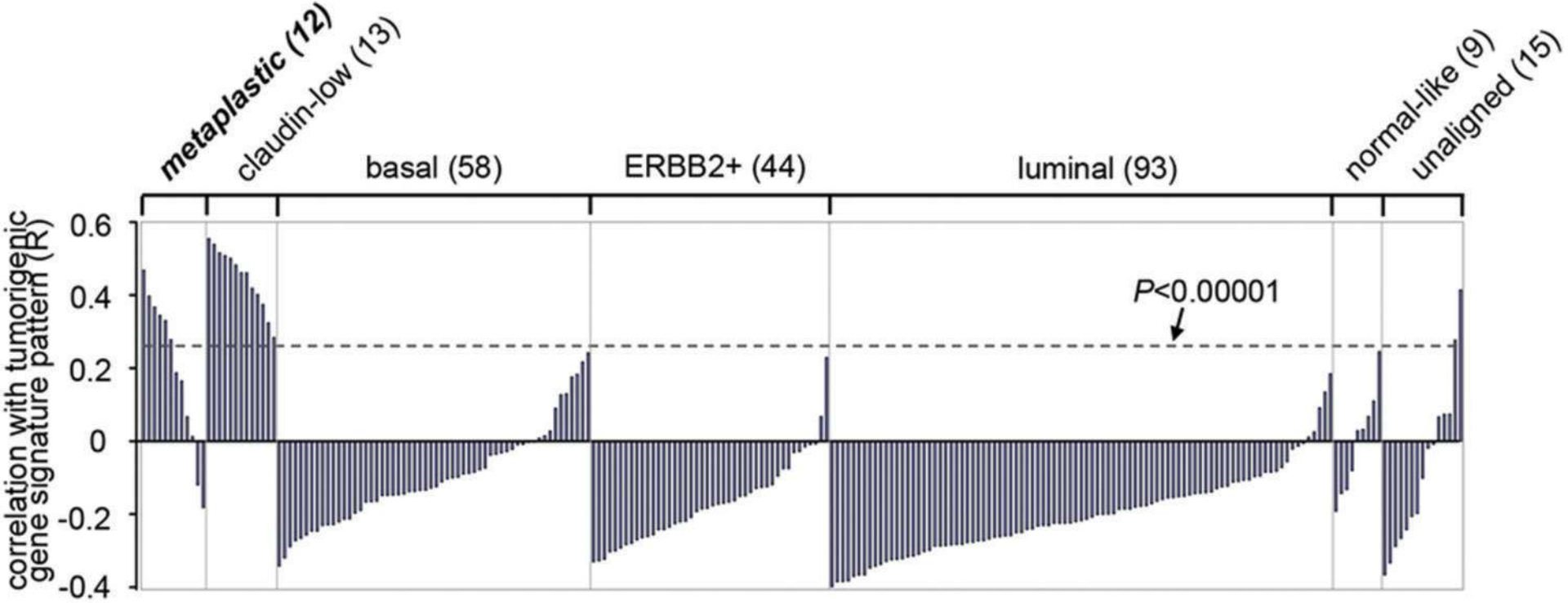
Figure 4. A $CD44^{+}/CD24^{-/low}$ tumorigenic gene signature is enriched in human breast tumors of the "claudin-low" and metaplastic (MBC) subtypes. Correlation between the "tumorigenic" signature pattern (Creighton et al., submitted) (using "1"and "-1", for up and down genes, respectively) and each MBC as well as each tumor in the gene expression profile dataset by Herschowitz et al (9). R-values above red dotted line are significant ($p<0.00001$).


All Breast Cancers
Basal Breast Cancers
Metaplastic Breast Cancers
Fraction gained or lost
0.8
0.4
0.0
−0.4
−1.0
2
4
6
8
10
12
14
16
18
20
22
1
3
5
7
9
11
13
15
17
19
21
X
chromosome number



Antibody	ANOVA p value by subtype
Aktp308	<0.000001
Aktp473	<0.000001
GSK3p21_9	<0.000001
mTorp	0.01
p70S6Kp389	0.00002
S6p235-236	<0.000001
S6p240_4	<0.000001


CLDN3
p = 1.33e-16
CLDN4
p = 2.29e-27
CLDN7
p = 2.43e-23
CDH1
p = 7.84e-21
CD44
p = 3.38e-09
CD24
p = 1.33e-09
CD29
p = 0.0283
CD44 / CD24 Ratio
p = 5.36e-12
CD29 / CD24 Ratio
p = 4.67e-12
BMI1
p = 3.37e-10
Basal Claudin Her2 Luminal Metap Normal


metaplastic (12)
claudin-low (13)
basal (58)
ERBB2+ (44)
luminal (93)
normal-like (9)
unaligned (15)
correlation with tumorigenic gene signature pattern (R)
0.6
0.4
0.2
0
-0.2
-0.4
P<0.00001