Genomic Alterations Associated with Recurrence and TNBC Subtype in High-risk Early Breast Cancers

Timothy R. Wilson¹*, Akshata R. Udyavar²*, Ching-Wei Chang³, Jill M. Spoerke¹, Junko Aimi¹, Heidi M. Savage¹, Anneleen Daemen², Joyce A. O'Shaughnessy⁴,⁵,⁶, Richard Bourgon² and Mark R. Lackner¹

Departments of ¹Oncology Biomarker Development, ²Bioinformatics and Computational Biology, ³Biostatistics, Genentech Inc., 1 DNA Way, South San Francisco, CA, USA. ⁴US Oncology, ⁵Baylor University Medical Center, ⁶Texas Oncology, Dallas, TX, USA. *These authors contributed equally to this work. Correspondence and request for materials should be addressed to T.R.W. (email: wilson.timothy@gene.com) or M.R.L. (email: lackner.mark@gene.com).

Running Title: Genomic Profiling of High-risk Early Breast Cancers

Keywords: early breast cancer, next generation sequencing, risk of recurrence, triple negative breast cancer, tumor mutational burden

Conflicts of interest: TRW, ARY, CWC, JMS, JA, HMS, AD, RB and MRL are employed by Genentech and have stocks in Roche. JOS has received speaking fees from Blue Earth Diagnostics and Novartis, consulting fees from AstraZeneca, Celgene, Eli Lilly, LaRoche, Lilly USA, Merck, Novartis and Pfizer, Honoraria from Seattle Genetics and
sponsored Travel from AstraZeneca, Blue Earth Diagnostics, Eli Lilly, Merck, Novartis and Seattle Genetics.
Abstract

The identification of early breast cancer (eBC) patients who may benefit from adjuvant chemotherapy has evolved to include assessment of clinicopathological features such as tumor size and nodal status, as well as several gene expression profiles for ER-positive, HER2-negative cancers. However, these tools do not reliably identify patients at the greatest risk of recurrence. The mutation and copy number landscape of triple-negative breast cancer (TNBC) subtypes defined by gene expression is also largely unknown, and elucidation of this landscape may shed light on novel therapeutic opportunities. The USO01062 Phase 3 clinical trial of standard chemotherapy (with or without capecitabine) enrolled a cohort of putatively high-risk patients based on clinical features, yet only observed a five-year disease-free survival event rate of 11.6%. In order to uncover genomic aberrations associated with recurrence a targeted next-generation sequencing (NGS) panel was used to compare tumor specimens from patients who had a recurrence event with a matched set who did not. The somatic mutation and copy number alteration landscapes of high-risk eBC patients were characterized and alterations associated with relapse were identified. Tumor mutational burden was evaluated but was not prognostic in this study, nor did it correlate with PDL1 or CD8 gene expression. However, TNBC subtypes had substantial genomic heterogeneity with a distinct pattern of genomic alterations and putative underlying driver mutations.
**Implications:** The present study uncovers a compendium of genomic alterations with utility to more precisely identify high-risk patients for adjuvant trials of novel therapeutic agents.

**Introduction**

Breast cancer is a highly heterogeneous disease which for decades has been subtyped and treated based on the immunohistochemical staining (IHC) of three receptors: estrogen receptor (ER), progesterone receptor (PR) and epidermal growth factor receptor 2 (ERBB2, HER2). With the seminal paper by Perou C. *et al.* (1), and follow on work from other groups (2-4), the breast cancer community began to appreciate the molecular heterogeneity that exists within breast cancer at the transcriptional level. Gene expression based classifiers such as PAM50 (5), MammaPrint (3) and others have been shown to provide additional prognostic information beyond traditional IHC based classification. More recently, Curtis and colleagues subtyped breast cancer into ten distinct integrative subtypes based on whole genome analysis of copy number alterations and gene expression, again showing distinct prognostic implications across subtypes (6).

Within the triple negative (TNBC) subtype of breast cancer, (i.e. those that stain negative by IHC for ER, PR and HER2) four to six distinct biological subtypes have been defined at the transcriptional level (7-9). For example, in the study by Lehmann *et al.*, the authors grouped TNBC into six subtypes: Basal-like 1 (BL1), Basal-like 2 (BL2), Immunomodulatory (IM), Mesenchymal-like (ML), Mesenchymal...
Stem-like (MSL) and Luminal AR (androgen receptor, LAR) (7). Interestingly, TNBC cell lines classified according to Lehmann et al. displayed different sensitivities to chemotherapeutics and/or targeted therapies, suggesting that different genomic alterations may drive each subtype. Additional work from other groups, including our own, has shown that the TNBC subtypes have implications for pathological complete response rates following neoadjuvant therapy (10,11) and for disease free survival (DFS) following adjuvant chemotherapy (12). To date, the mutational and copy number profiles associated with the TNBC subtype have not been fully characterized.

Surgical resection of the tumor followed by adjuvant therapy to eradicate micrometastatic lesions is potentially curative in patients with early breast cancer. Early screening, incorporation of hereditary risks and treatment improvements have dramatically improved survival (13,14); however, a subset of patients will still recur with metastatic disease. Defining the subset at very high risk for recurrence remains challenging, and slows the development of investigational agents that are attempting to show improvement in 5-year DFS, since the large majority of patients will not develop disease recurrence within five years. In order to enrich for patients most in need of novel therapies, adjuvant clinical trial attempt to enroll high-risk patients largely based on clinical features such as tumor size nodal status. The US001062 adjuvant phase III trial of standard chemotherapy, with or without capecitabine, enrolled 2,611 patients based on high-risk clinical features (T1–3, N1–2, M0; or T > 2 cm, N0, M0; or T > 1 cm, N0, M0 and both ER- and PR-negative).
Despite this enrichment strategy, the 5-year DFS rate was 88.4% (15), highlighting the need for additional means to define truly high-risk patients.

In the current study, we molecularly profiled tumors from the US001062 trial using the FoundationOne next generation sequencing (NGS) platform. We focused on those patients who experienced a DFS event and compared them to a demographically matched set of patients from this trial who did not experience a DFS event. We identify high-risk molecular traits that may be used to more accurately select patients at a high-risk of recurrence at five years. We also report the mutational, copy number alteration and rearrangement landscape of TNBC subtypes.

**Methods**

**US001062 (NO17629) study**

Patients were enrolled onto the parent study US001062, a randomized, open-label, multicenter, phase III trial comparing regimens of doxorubicin plus cyclophosphamide followed by either docetaxel or docetaxel plus capecitabine as adjuvant therapy for female patients with high-risk breast cancer (clinicaltrials.gov/show/NCT00089479) (15). Tissue samples were collected and analyzed following approval by the US Oncology Institutional Review Board and appropriate confirmation of written informed consent.

**Targeted next generation sequencing profiling**
Samples were submitted to a CLIA-certified, New York State-accredited, and CAP-accredited laboratory (Foundation Medicine, Cambridge, MA) for next-generation sequencing (NGS)-based genomic profiling using the Foundation Medicine FoundationOne comprehensive genomic panel (16). Tumor mutational burden (TMB) was determined by Chalmers ZR et al. (17). See Supplemental Methods for more detail.

**Prevalence of alterations within IHC subtypes**

Prevalence for each type of alteration was computed by calculating the sum of alterations in a given gene for all patients divided by the total number of patients in each IHC subtype. Relative prevalence in Fig 6 and Supplementary Fig 8 for TNBC subtypes was calculated by subtracting the prevalence of a given gene within each TNBC subtype by the prevalence of that gene in all TNBC patients.

**Mutual exclusivity, co-occurrence and enrichment of alterations within IHC subtypes**

For each type of alteration (short variant, amplification, loss, rearrangement), we considered a patient as “altered or 1” if the patient had an alteration in a particular gene, and “wild-type or 0” otherwise. Two or more alterations in the same gene in the same patient was considered as 1 for simplicity. Similarly, for pathway analysis, if any gene in a given pathway was altered, the patient was considered as “altered” or 1 for that pathway, and “wild-type or 0” otherwise. For pathway analysis, we
used the MSigDB 50 hallmark genesets (18) and filtered all genesets to only include the 401 genes that were part of the FoundationOne panel.

For each pair of genes within each IHC subtype, Fisher’s exact test was used to compute mutual exclusivity (negative log2 odds ratio) and co-occurrence (positive log2 odds ratio). The enrichment of alterations (given by log2 odds ratio) for individual genes or pathways across IHC subtypes was calculated by applying a Fisher’s exact test to a 2-by-2 contingency table. In all cases, raw p-values were corrected for multiple testing using the Benjamini-Hochberg method. Only log2 odds ratios with q-values < 0.2 are shown or annotated in Figure 1 and 2.

**Identification of alterations associated with high-risk of recurrence**

For the purpose of this analysis, we considered a patient as “altered or 1” if the patient had any type of alteration in a particular gene (mutation/copy number/rearrangement), and “wild-type or 0” otherwise. We first assessed the prognostic significance of all clinical covariates listed in Supplemental Table 2, in the 291 event-matched patients (in DFS) using the Cox-proportional hazards model, and identified lymph node status to be highly prognostic. Next for each IHC subtype, we computed hazard ratio for DFS using the Cox-proportional hazards model with lymph node status included as a covariate. We report only the prognostic alterations in the forest plot in Fig 3 that are altered in at least 2 patients (i.e. 2 or more patients) and have a raw p-value < 0.05. The raw p-values were corrected for multiple testing using the Benjamin-Hochberg method after filtering for genes that
had alterations occurring in at least 2 patients. The types of alterations in each gene are described in the oncoprint to prevalence plots in Supplementary Fig 3.

Other statistical analyses

The p-values for multi-group comparisons in the boxplots in Fig 4 were computed using the Kruskal-Wallis test. The log-rank test was used to detect survival differences in the Kaplan-Meier curves for DFS. The oncoprint prevalence plots were generated using the ComplexHeatmap package in R. The lollipop plots in Supplementary Fig 5 were generated with the MutationMapper tool on www.cbioportal.org. All other plots were generated using the base or ggplot2 package in R. This study has been reported according to the Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) criteria (19).

Results

Targeted next generation profiling of high-risk early breast cancer patients.

In order to define the genomic landscape of a clinically-defined, high-risk early breast cancer population, we profiled the primary breast cancers of patients who experienced a DFS event following adjuvant chemotherapy within the US001062 study (15). Of the 2,611 enrolled patients, 1,181 patients had tumor tissue available for genomic profiling (12). Of the 1,181 patients, 145 patients experienced a DFS event and were selected for NGS profiling (Supplemental Fig 1). We also profiled a demographically matched control cohort of 146 samples from patients in this study who did not experience a DFS event, and a further 108 tumor samples from patients
who did not have a DFS event to increase the statistical precision of mutation prevalence estimates within IHC subtypes. As shown in Supplemental Table 1, the demographic and clinical characteristics were well balanced between the DFS event and control groups. Supplemental Fig 2 shows the overall genomic landscape of the clinically defined high-risk eBC population. The most prevalent alterations were consistent with previously published findings in TCGA (20).

To help interpret and organize this complex landscape, we next explored the genomic landscape within the IHC defined subtypes, in all the molecularly characterized tumors, regardless of patients’ clinical outcomes (Fig 1 and Supplemental Table 2). Within HER2+, HR+/- disease (hereafter referred to as HER2+, Fig 1A, n=57), the most frequently somatically mutated genes were TP53 (68%), PIK3CA (39%), and ARID1B (19%) and the most frequently copy number altered genes were ERBB2 (72%), CDK12 (58%), MYC (23%) and genes within the 11q13.3 (21%) and 17q (19%) loci. Within HR+, HER2- disease (hereafter referred to as HR+, Fig 1B, n=178), the most frequently somatically mutated genes were PIK3CA (45%), TP53 (30%) and MLL3 (24%), and the most frequently copy number altered genes were co-amplified genes within the 11q13.3 (21%), 8p11-12 (20%) and 8q (13%) loci. Within TNBC (Fig 1C, n=162), the most frequently somatically mutated genes were TP53 (93%), NOTCH1 (19%) and BRCA1 (18%), and the most frequently copy number altered genes were GATA3 (19%) and co-amplified genes within the 8q (19%) and 12q (11%) loci.
We next assessed the co-occurrence or mutual exclusivity of somatically altered genes within each IHC-defined subtype in all the molecularly characterized cancers. Within HER2+ disease (Fig 1A) significantly co-occurring events included ERBB2 with CDK12; 17q22-24 genes with 11q13.3 genes as previously described (21) or with SPOP; 8p11-12 genes and BRCA2 with CCNE1 and TP53BP1. Within HR+ disease (Fig 1B), co-occurrence events included 8p11-12 genes with MYST3 or with 11q13.3 genes; MLL2 with MLL3 and TP53 mutations with 8q genes. Within TNBC disease (Fig 1C), co-occurrence events included: 1q23 genes; 8q genes; 1q23 genes with 8q genes and 12p genes. These amplification events are in general agreement with previous reports (22-24). Few genes were found to exhibit statistically significant patterns of mutually exclusive mutation, with the only examples being TP53 with PRKAR1A, RAD51C, BRIP1, RNF43 or SPOP in HER2+ disease; and TP53 with PIK3CA or CDH1 in HR+ disease.

**Enrichment of genomic alterations and pathways by immunohistochemical subtype.**

We next assessed whether any mutational or copy number alterations were enriched within the different IHC subtypes (Fig 2). Directly comparing HER2+ to HR+ disease (Fig 2A) showed enrichment in HER2+ disease of alterations (short variants, copy number alterations and rearrangements) in RNF43, GNA13 and 17q genes. No enrichment of alterations was observed in HR+ compared to HER2+ disease. Comparing HER2+ disease to TNBC (Fig 2B) showed enrichment in HER2+ disease of PIK3CA, AURKA, RNF43, TOP2A and amplified genes on 17q and 11p13.3,
whereas enrichment of alterations in \( TP53, BRCA1, KDM5A, RB1 \) and \( NOTCH1 \) were observed in TNBC disease. Lastly, comparing HR+ disease to TNBC (Fig 2C) showed many statistically significant differences, including an enrichment of alterations in 11p13, 8p, \( PIK3CA, AURKA, CDH1 \) and \( MAP3K1 \) in HR+ disease, whereas alterations in \( TP53, BRCA1, RB1, PIK3C2G, PDCD1LG2, MYC, NOTCH1, PTEN, FGF6, RAD52, LYN, CCND3, CDKN1A \) and \( KRAS \) were preferentially seen in TNBC.

As many individual genes can cause activation of a common pathway, we next grouped genes by pathway and assessed which pathways were differentially altered within the IHC subtypes. For pathway analysis, we used the Broad Molecular Signatures Database (MSigDB) 50 hallmark gene set collection (18). Comparing HER2+ to HR+ disease (Fig 2D, Supplementary Fig 3) showed enrichment in HER2+ disease of the p53 pathway, Wnt-beta catenin signaling, epithelial mesenchymal transition, peroxisome and pathways associated with cell survival including E2F targets, apoptosis and UV response. No pathway enrichment was noted in HR+ disease in comparison to HER2+. Comparing HER2+ BC to TNBC (Fig 2E, Supplementary Fig 3) showed an enrichment of pathways in TNBC, namely the allograft rejection and apical surface pathways. Other pathways enriched in TNBC compared to HER2+ BC were the DNA repair (e.g. homologous recombination), p53 pathway, PI3K/AKT/mTOR pathway, as well as fatty acid metabolism and oxidative phosphorylation. Comparing HR+ to TNBC tumors (Fig 2F, Supplementary Fig 3), we identified an enrichment of pathways in HR+ disease such as the complement and estrogen response late pathways. Conversely, E2F targets, myogenesis, interferon
gamma response, apical surface, PI3K/AKT/mTOR and hedgehog pathways were upregulated in TNBC tumors.

**Association of genomic alterations with disease free survival.**

The ability to identify patients who are most likely to experience a DFS event through genomic analysis could pave the way for designing adjuvant studies in specific high-risk populations, and could also be the basis for identification of therapeutically relevant targets in these patients. We directly compared the genomic landscape of patients who experienced a DFS event to a matched set of control patients (total 291 patients). As IHC status correlates with outcome in eBC, we controlled for its impact by specifically looking within each IHC subtype for enrichment of genomic alterations (Fig 3). We also controlled for lymph node status as this was found to associate with worse DFS outcomes in our data set.

Within HER2+ disease (Fig 3A, n=41), we identified 12 genomic alterations (mutations, copy number alterations and rearrangements) that produced a raw p-value < 0.05 for association with DFS, though only two genes (AR and MCL-1) yielded an adjusted p-value under our 0.20 FDR cutoff. These 12 genes were each altered in 5-7% of HER2+ tumors (Supplemental Fig 4). We conducted a co-occurrence analysis on the twelve genes with a raw p-value < 0.05 for association within disease; only two gene pairs were found to exhibit significant co-occurrence of alteration in the same tumor, HSD3B1 and LRP6, and INSR and JUN (Supplemental Fig 4). Within HR+ disease (Fig 3B, n=128), we identified 28 genomic alterations
that produced a raw p-value < 0.05 for association with DFS, twelve of which were associated with poor survival post-FDR correction. Most prevalent among the prognostic alterations were mutations in \textit{ATM} (12\%), \textit{EPHB6} (8\%), \textit{ALK} (8\%) and \textit{ERCC4} (8\%, Supplemental Fig 4). Among the genes shown in Fig 3B, two pairs of genes were found to mutually co-occur in the same tumor, these being \textit{H3F3A} and \textit{PARP1}, and \textit{EPHB6} and \textit{PARK2} (Supplemental Fig 4). Lastly, within TNBC (Fig 3C, n=120), we identified 21 genomic alterations that produced a raw p-value < 0.05 for association with DFS, seven of which were associated with poor prognosis in TNBC post-FDR correction. Most prevalent among the poor prognostic alterations were alterations in \textit{MAP3K1} (12\%), \textit{MST1R}, \textit{IKBKE} and \textit{EMSY} (4\%, Supplemental Fig 4). Conversely, genes associated with good prognosis at a raw p-value < 0.05, though not significant after multiple testing correction, were \textit{CREBBP} (HR 0.29) and \textit{BRCA1} (HR 0.24). Of the genes identified in Fig 3C, only one pair was found to significantly co-occur in the same tumor, namely \textit{LYN} and \textit{PRKDC} (Supplemental Fig 4).

Supplemental Fig 5 depicts the spatial distribution of single nucleotide variants associated in the genes that were prognostic post FDR correction and present in 5 or more samples.

\textbf{Association of tumor mutational burden, prognosis and immune markers.}

High tumor mutational burden (TMB) has been shown to correlate with clinical benefit from PD-1/PD-L1 checkpoint inhibition (25). Additionally, TMB derived by the FoundationOne targeted gene panel has been shown to act as a suitable surrogate for TMB derived via whole exome sequencing (26), and to correlate with
outcomes for atezolizumab in urothelial cancer (27). Therefore, we utilized data generated by the FoundationOne assay in our current study to determine whether TMB was associated with disease subtype, immune gene expression and DFS. Within IHC subtypes, median TMB was 1.5 fold higher in TNBC compared to HER2+ and HR+ (4.05 vs 2.7 and 2.7 respectively, Fig 4A). We next assessed TMB across the PAM50 subtypes (Fig 4B). Similar to TNBC, the basal-like subtype had the highest median TMB (4.5), along with the luminal B (4.5), followed by luminal A (2.7) and HER2-enriched (2.25) subtypes. TMB did not correlate significantly with Lehmann et al. TNBC subtypes (Fig 4C).

TMB did not correlate with CD8 or PD-L1 gene expression (Figs 4D and E), and was not associated with DFS in the entire NGS population, or within the IHC defined subtypes (Supplemental Fig 6). However, CD8 gene expression, a marker for cytotoxic effector T-cells, strongly correlated with PD-L1 gene expression levels (Fig 4F). Although median CD8 gene expression was similar across the IHC subtypes (Supplemental Fig 7A), it was significantly prognostic in the entire BC population, where high expression was associated with good prognosis (Supplemental Fig 7), and this association was most pronounced in TNBCs (p=0.041), as previously observed (12). High CD8 gene expression weakly trended with better DFS in the HR+ and HER2+ subgroups, but this effect was not statistically significant. Within the Lehmann et al. TNBC subgroups, CD8 gene expression has highest in the IM and MSL groups and lowest in the BL2 group (Supplemental Fig 7B). Together, these data suggest that high CD8 gene expression, rather than TMB may represent an
immune activated tumor environment in breast cancer, particularly in various TNBC subtypes.

**Genomic landscape of TNBC molecular subtypes.**

TNBC is a heterogeneous disease and can be subtyped by gene expression into as many as six distinct molecular subtypes (7). We previously reported the prevalence and prognostic implications of the Lehmann et al. gene expression subtypes in this adjuvant population (12). However, the genomic landscape and underlying mutational drivers within each subtype are largely unknown. Therefore, we assessed the somatic mutation, copy number and rearrangement landscape within the Lehmann et al. defined subtypes (Fig 5 and Supplemental Fig 8A-F). All TNBC subtypes had a high prevalence of TP53 mutations ranging from 73-100%. In the BL1 subgroup, mutations in NOTCH1, ARID1B, BRCA1 and BRCA2 were common, as were copy number amplifications in GATA3, KDM5A and RAD52 (Fig 5A). Within the BL2 subgroup, which has a poor outcome, genomic amplifications were common in 8q locus genes (PREX2, RUNX1T1, NBN, LYN) and 12q locus genes (LRP6 and PIK3C2G, Fig 5B). The IM subtype, which has the best outcome, harbored mutations in CREBBP, PIK3C2B, and APC, and amplifications in GATA3 and AKT3 (Fig 5C). As the IM subgroup associated with a favorable outcome and contained a high prevalence of CREBBP and BRCA1 mutations as well as high CD8 gene expression, we next tested whether mutations in CREBBP and BRCA1 were independent correlates of outcome. When controlling for CD8 gene expression, only BRCA1 retained its association with prognosis (raw p-value = 0.046). The LAR subgroup
was mostly non-basal by PAM50 analysis, and enriched for the HER2 PAM50 subtype. The LAR subgroup was largely driven by PI3K signaling, as tumors had mutations in *PIK3CA, PIK3R1, mTOR, PTEN, IRS2* and *TSC2* along with copy number amplifications in *ERBB2* and *FGFR1* (Fig 5D). The M subgroup, which had the worst outcome, contained mutations in *IRS2* and *NOTCH1* and copy number amplifications in *CCNE1, CCND1* and *IL7R* (Fig 5E). Lastly, the MSL subgroup, which had a favorable outcome, displayed mutations in *ROS1, ATR* and *MET* and copy number amplifications in *SPTA1, DDR2* and *MYC* (Fig 5F).

We next assessed how the prevalence patterns of the most abundant alterations identified in Fig 1C are represented across the six TNBC subtypes. Out of the 42 most frequently altered genes in the entire TNBC cohort (Fig 1C), only seven displayed a significant (FDR < 0.2) differential mutation rate across the six TNBC subtypes (Fig 6). The MSL and BL2 subtypes more frequently harbor six of these seven TNBC genes. The remaining subtypes frequently show alterations in only two of the seven identified genes. Lastly, we assessed the prevalence of the TNBC prognostic genes associated with DFS from Fig 3C (Supplementary Fig 8G). Out of the 21 prognostic genes in TNBC (Fig 3C), only two genes *CREBBP* and *PRKDC* showed differential mutation rates across the six TNBC subtypes (raw p-value < 0.05), however, neither was significant post-FDR correction.

**Discussion**
A number of elegant studies describing the genomic and transcriptomic profiles of breast cancer and showing the spectrum of mutations and copy number alterations within IHC- and PAM50-defined subtypes have been published over the past several years (20,28,29). However, little is known about how these alterations affect clinical outcome. Using a targeted NGS panel, we profiled the genomic landscape of a clinically defined high-risk patient population that was enrolled onto the US001062 phase III adjuvant capecitabine trial. Specifically, we chose patients who had a recurrence event and we matched them demographically to a control set of samples from patients who did not have a recurrence event in order to uncover genomic traits that may be used to select high-risk patients for future adjuvant trials of novel agents and to potentially uncover therapeutic targets.

At the population level, our genomic analysis has commonalities with previous findings. Within HER2+ breast cancer, 72% of the samples showed ERBB2 copy number amplification by the FoundationOne assay, suggesting significant but nonetheless imperfect concordance between NGS-based amplification assessment and traditional scoring by IHC. Within TNBC, mutations in the tumor suppressor TP53 were the most common genetic alteration, and they occurred throughout the coding region of the gene. Loss-of-function mutations in TP53 result in cells being more reliant on the Chk1 pathway to repair DNA breaks (30). Pre-clinical studies suggest that Chk1 inhibitors can potentiate the cytotoxic effects of chemotherapies, such as gemcitabine (31), and such combinations are currently being assessed clinically (32). Also in TNBC, we identified DDR2, RAD52 and KDM5A amplification
together with \textit{PTEN} loss, as well as \textit{BRCA1} and \textit{RB1} mutations, which suggest increased reliance on other DNA repair mechanisms and an impaired ability to repair DNA following certain therapeutics. These results are particularly intriguing in light of recent reports suggesting that patients with higher genomic instability and TMB levels have an increased likelihood of responding to checkpoint inhibition (33).

Cancer pathways can be regulated by many genes to achieve the same biological effect. Using the MSigDB hallmark pathway genesets, we found that distinct pathways were altered within the three IHC subtypes. Myogenesis, oxidative phosphorylation and PI3K-AKT-MTOR pathways were altered with significantly higher frequency in TNBCs, compared to HER2+ and HR+ tumors. Activation of PI3K-AKT-MTOR signaling was also shown in TCGA (20), which reported higher PI3K pathway activity in the basal subtype of breast cancer, a large majority of which are TNBCs. PI3K pathway targeted therapies such as the AKT inhibitor ipatasertib have been shown to prolong DFS in TNBC in combination with chemotherapy (34).

By comparing the somatic mutation, copy number alteration and rearrangement landscapes between patients who experienced a recurrence event and those who did not, we identified numerous alterations that were associated with clinical outcome. Although in some cases the mutations were of low prevalence, our findings show several key alterations that may be targetable and associated with...
relapse in this analysis. For example, in HR+ disease, we found that alterations in cell cycle/DNA response and repair genes, such as CDK8, CDK4, CDKN2B, ATM and ERCC4, were associated with increased rates of recurrence. These results are timely given the beneficial effects of CDK4/6 inhibitors in HR+ metastatic breast cancer (35), with the results from adjuvant studies eagerly anticipated.

In HER2+ disease, we found that AR and MCL-1 were significantly associated with poor prognosis. Agents that target AR are routinely used for the treatment of men with prostate cancer (36), and small molecule inhibitors targeting MCL-1 are under clinical evaluation (ClinicalTrials.gov NCT02675452), both of which perhaps could be utilized in these genomically altered HER2+ breast cancers.

Within TNBC, pro-tumorigenic inflammatory cytokine signaling genes such as IKBKE and MAP3K1 were significantly associated with poor prognosis (37,38). MAP3K1, which regulates JNK activation and cell migration (39), has recently been identified as a driver gene in metastatic breast cancer samples from the SAFIR01 clinical trial (40). Within TNBC, alterations in BRCA1 and CREBBP were associated with a better DFS, although these findings were not statistically significant after multiple testing correction. However, amplification of EMSY, a nuclear protein that binds and represses BRCA2 and increases genomic instability (41), was significantly associated with poor prognosis in TNBC. This finding raises the question of whether patients whose early stage TNBCs harbor an EMSY amplification might benefit from the addition of carboplatin to standard chemotherapy.
Cancer immunotherapy checkpoint inhibitors, such as those that target PD-L1 and PD-1, can unlock the patient’s own immune system to unleash an anti-cancer response, particularly in cancers that have a high mutational load such as lung cancer and melanoma (42,43). We found that TMB was highest in the TNBC subtype, and when classifying samples by PAM50 subtype, we further showed that the luminal B tumors have a similar TMB to the basal-like subtype. We found that gene expression of the activated T cell marker, \( CD8 \), did not correlate with TMB, suggesting that TMB may not be a predictive marker of benefit from immune activity in breast cancer, which may be in part due to the low TMB levels in breast cancer.

Using unbiased microarray gene expression technologies, the TNBC subtype has emerged as the most heterogeneous of breast cancers with the claudin-low (44,45), Basal-A/B and HER2-enriched subtypes (46). Lehmann and colleagues identified six molecular TNBC subtypes, which they subsequently modified to a four-subtype classifier by removing the tumor-associated microenvironment, including the stromal TILs, by micro-dissection of the tumor specimens (7,8). We previously applied the original Lehmann et al. gene classifier to our adjuvant US001062 population and showed its relationship to clinical outcome (12). In the current analysis, we report the distinct mutation and copy number alterations within the six TNBC subtypes. The LAR subtype, which has previously been reported to contain the majority of \( PIK3CA \) mutations present within TNBC (12,47), appears almost
completely driven by PI3K signaling, suggesting that the LAR population may be potentially responsive to PI3K-mTOR inhibitors, perhaps in the presence of anti-androgen receptor therapies. Within the M subtype, IRS2 was a frequently altered gene. IRS2 is a substrate for Insulin Receptor Kinase 1, inhibitors of which have been tested in clinical trials, with unfavorable results, albeit in biomarker unselected populations (48). It is possible that the IRS2-altered M sub-population could benefit from inhibition of the IGFR pathway. Recently, Barecehe and colleagues utilized the METABRIC and TCGA datasets to analyze the mutational and copy number landscape in TNBC and made similar observations (49). Specifically, they noted that LAR tumors were associated with higher TMB and PI3K pathway activation, M tumors had activated EGFR and Notch signaling and BL1 tumors showed copy number losses for BRCA1/2 and RB1. Given that each Lehmann subtypes may comprise approximately 10-15% of TNBC, and that TNBC itself represents 15 to 18% of all breast cancers, the development of targeted therapies in these patient populations will likely require phase II testing in small subgroups of molecularly selected patients whose breast cancers have been screened for multiple clinical trial-qualifying alterations.

Although the total number of patients in the USO01062 trial is large, the statistical power in our study was nonetheless limited by the low number of progression events and our selection of only a subset of patients’ tumors for genetic analysis. Furthermore, we carried out many of our analyses within IHC-defined subtypes, further reducing our power to detect differences between patients who experienced
a progression event versus those who did not. In addition, our study was not powered to formally statistically test the differences in genomic alterations within the six TNBC subtypes nor to correlate these alterations with outcomes. In spite of these limitations, our study does have detailed clinical outcomes on the patients enrolled onto the USO01062 trial, and it identifies multiple, novel genomic associations that warrant further testing in an independent data set.

In conclusion, using a targeted NGS approach, we characterized the somatic mutation and copy number alteration landscape of high-risk early breast cancer patients in a prospective phase 3 trial where all patients received state of the art adjuvant chemotherapy. We describe the mutational landscape and enrichment patterns for certain alterations at the single-gene and pathway levels, and within IHC-defined subtypes. We discovered genomic traits associated with disease recurrence that may be used to select high-risk patients in future studies. We show that TMB did not correlate with clinical outcome overall, nor within any IHC subtype. Finally, we show that the Lehmann et al. TNBC subtypes have distinct mutational landscapes and we uncovered several previously unrecognized alterations that may be therapeutically relevant in this patient population.
References


23. Giltane JM, Hutchinson KE, Stricker TP, Formisano L, Young CD, Estrada MV, et al. Genomic profiling of ER+ breast cancers after short-term estrogen...


blind, placebo-controlled, phase 2 trial. The Lancet Oncology
35. Finn RS, Martin M, Rugo HS, Jones S, Im SA, Gelmon K, et al. Palbociclib and
Letrozole in Advanced Breast Cancer. The New England journal of medicine
survival with enzalutamide in prostate cancer after chemotherapy. The New
cytokine network impairs triple-negative breast cancer growth. The Journal
of clinical investigation 2014;124(12):5411-23 doi 10.1172/JCI75661.
38. Kyriakis JM, Avruch J. Mammalian MAPK signal transduction pathways
activated by stress and inflammation: a 10-year update. Physiological
39. Pham TT, Angus SP, Johnson GL. MAP3K1: Genomic Alterations in Cancer and
Function in Promoting Cell Survival or Apoptosis. Genes & cancer 2013;4(11-
Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis.
PloS medicine 2016;13(12):e1002201 doi 10.1371/journal.pmed.1002201.
links the BRCA2 pathway to sporadic breast and ovarian cancer. Cell
Nivolumab versus chemotherapy in patients with advanced melanoma who
progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised,
controlled, open-label, phase 3 trial. The Lancet Oncology 2015;16(4):375-
84 doi 10.1016/S1470-2045(15)70076-8.
Atezolizumab versus docetaxel in patients with previously treated non-
small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised
controlled trial. Lancet 2017;389(10066):255-65 doi 10.1016/S0140-
6736(16)32517-X.
Phenotypic and molecular characterization of the claudin-low intrinsic
10.1186/bcr2635.
epithelial-to-mesenchymal transition interactome gene-expression signature
is associated with claudin-low and metaplastic breast cancer subtypes.
Proceedings of the National Academy of Sciences of the United States of
breast cancer cell lines for the study of functionally distinct cancer subtypes.


Figure Legends

**Figure 1. Genomic landscape within IHC BC subtypes.**

Figure displays the genomic landscape of (A) HER2+, (B) HR+ and (C) TNBC denoting the most frequently mutated genes in each IHC subtype. Left panels show log2 odds ratio for co-occurrence of mutation in gene pairs, with mutually exclusive events corresponding to a negative log2 odds ratio (violet color), and co-occurrence events corresponding to a positive log2 odds ratio (green color). Only gene pairs with adjusted p-value < 0.2 are colored in the plots. Middle panels are tile plots showing the detailed genomic landscape of the most frequently altered genes, within each IHC subtype, PAM50 status and HER2/ER/PR status measured by IHC are shown as relevant at the top of each plot. Commonly co-amplified genes on same loci are grouped and denoted by black brackets. Right panels are barplots denoting the overall prevalence of genes delineating the individual prevalence of the four types of alterations - short variant (green), amplification (red), loss (blue) and rearrangement (orange).

**Figure 2. Enrichment of genomic alterations and hallmark pathways by IHC subtype.**

Volcano plots show the enrichment of overall alterations (short variants, copy number alterations and rearrangements) for single genes (A-C) or MSigDB (18) Hallmark pathway gene sets (D-F). The x-axis represents the log2 odds ratio that a gene would be altered in one IHC subtype over the other as indicated by each title.
Title subtype A vs. subtype B indicates that highlighted genes/pathways on the left-hand side of the graph are higher expressed in subtype A, while genes/pathways highlighted on the right-hand side of the graph are higher expressed in subtype B. The y-axis represents the negative log10 Benjamini and Hochberg FDR adjusted p-value, applied to each panel independently. The horizontal grey line denotes adjusted p-value of 0.2 and two vertical grey lines denote log2 odds ratio of 1. Only the genes and pathways that have log2 odds ratio > 1 and adjusted p-value < 0.2 are annotated.

**Figure 3. Association of genomic alterations with disease free survival.**

Forest plots displaying hazard ratios (with unadjusted 95% confidence intervals) of genes whose alteration (mutation, copy number, and rearrangement) was associated with disease free survival in HER2+ (A), HR+ (B) and TNBC (C). For each gene, the table shows the number of recurring patients with alterations, the total number of patients with alterations, the hazard ratio point estimate, and raw and adjusted p-values. All genes with raw p-values < 0.05 are shown in the plots, and those with adjusted p-value < 0.2 — corresponding to the most robust signal — are denoted by stars.

**Figure 4. Correlation of tumor mutational burden, prognosis and immune markers.**

Tumor mutational burden (TMB, mutations per megabase) stratified by IHC (A), PAM50 (B) and Lehmann et al. TNBC (C) subtypes. The p-values denote significance
by the Kruskal-Wallis test. Association of TMB with \textit{CD8} gene expression (D), TMB with \textit{PDL1} gene expression (E) and correlation of \textit{CD8} with \textit{PDL1} gene expression (F), across all BC patients. Each dot is a patient. Pearson’s correlation coefficient, linear model fit (blue line) and associated p-values (for non-zero slope) are shown on each plot.

Figure 5. Genomic landscape of TNBC molecular subtypes.

Plots showing the genomic landscape of the frequently mutated genes within each Lehmann et al TNBC subtypes. (A) Basal-like 1 (BL1), (B) Basal-like 2 (BL2), (C) Immunomodulatory (IM), (D) Luminal AR (androgen receptor, LAR), (E) Mesenchymal-like (ML), and (F) Mesenchymal Stem-like (MSL). The frequency of the combined alterations and the individual prevalence of the four types of alterations - short variant (green), amplification (red), loss (blue) and rearrangement (orange) for each gene is shown as percentages and barplots of counts.

Figure 6. Distribution of prevalent TNBC genes in the Lehmann et al. TNBC subtypes.

Figure shows significant genes that are frequently mutated in TNBC from Fig 1C, that have a significant difference in prevalence within the TNBC subtypes by Fisher’s test (adjusted p-value < 0.2). Dot size denotes prevalence of mutation of a gene within a given subtype. Color shows the relative prevalence of a given gene across Lehmann et al. TNBC subtypes, i.e., the degree to which mutation rate in a
given TNBC subtype is less or greater than the average rate across all TNBC subjects.

The adjusted p-values for the genes are indicated on the right side.
Figure 1

A

HER+ N=57

B

HR+ N=178

C

TNBC N= 162

Prevalence

Log2 o dds ratio

Co-occurrence

Mutual exclusivity

Alte ration type

Amplification

Loss

Short variant

Rear ranging

HER2/ IHC

Positive

Negative

PR/ IHC

Positive

Negative
Figure 2
Figure 3

(A) HER2+ : DFS

(B) HR+ : DFS

(C) TNBC : DFS

Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.
Figure 4
Figure 5

Downloaded from mcr.aacrjournals.org on January 2, 2021. © 2018 American Association for Cancer Research.
Figure 6

Prevalent genes in TNBC

<table>
<thead>
<tr>
<th>Gene</th>
<th>BL1</th>
<th>BL2</th>
<th>IM</th>
<th>LAR</th>
<th>M</th>
<th>MSL</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>MYC</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>PREX2</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>RUNX1T1</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>NBN</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>ROS1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>CREBBP</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Relative Prevalence

Prevalence

Downloaded from mcr.aacrjournals.org on January 2, 2021. © 2018 American Association for Cancer Research.