Proteogenomic markers of chemotherapy resistance and response in triple-negative breast cancer

Meenakshi Anurag  
*Baylor College of Medicine*

Shunqiang Li  
*Washington University School of Medicine in St. Louis*

Ian S Hagemann  
*Washington University School of Medicine in St. Louis*

Jeremy Hoog  
*Washington University School of Medicine in St. Louis*

Foluso O Ademuyiwa  
*Washington University School of Medicine in St. Louis*

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/oa_4](https://digitalcommons.wustl.edu/oa_4)

Part of the [Medicine and Health Sciences Commons](https://digitalcommons.wustl.edu/oa_4)

Please let us know how this document benefits you.

**Recommended Citation**

Anurag, Meenakshi; Li, Shunqiang; Hagemann, Ian S; Hoog, Jeremy; Ademuyiwa, Foluso O; and et al., "Proteogenomic markers of chemotherapy resistance and response in triple-negative breast cancer." 

[https://digitalcommons.wustl.edu/oa_4/933](https://digitalcommons.wustl.edu/oa_4/933)

This Open Access Publication is brought to you for free and open access by the Open Access Publications at Digital Commons@Becker. It has been accepted for inclusion in 2020-Current year OA Pubs by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
ABSTRACT Microscaled proteogenomics was deployed to probe the molecular basis for differential response to neoadjuvant carboplatin and docetaxel combination chemotherapy for triple-negative breast cancer (TNBC). Proteomic analyses of pretreatment patient biopsies uniquely revealed metabolic pathways, including oxidative phosphorylation, adipogenesis, and fatty acid metabolism, that were associated with resistance. Both proteomics and transcriptomics revealed that sensitivity was marked by elevation of DNA repair, E2F targets, G2–M checkpoint, interferon-gamma signaling, and immune-checkpoint components. Proteogenomic analyses of somatic copy-number aberrations identified a resistance-associated 19q13.31–33 deletion where LIG1, POLD1, and XRCC1 are located. In orthogonal datasets, LIG1 (DNA ligase I) gene deletion and/or low mRNA expression levels were associated with lack of pathologic complete response, higher chromosomal instability index (CIN), and poor prognosis in TNBC, as well as carboplatin-selective resistance in TNBC preclinical models. Hemizygous loss of LIG1 was also associated with higher CIN and poor prognosis in other cancer types, demonstrating broader clinical implications.

SIGNIFICANCE: Proteogenomic analysis of triple-negative breast tumors revealed a complex landscape of chemotherapy response associations, including a 19q13.31–33 somatic deletion encoding genes serving lagging-strand DNA synthesis (LIG1, POLD1, and XRCC1), that correlate with lack of pathologic response, carboplatin-selective resistance, and, in pan-cancer studies, poor prognosis and CIN.
INTRODUCTION

Ten percent to 15% of breast cancers are designated triple-negative breast cancers (TNBC) because of low expression of HER2, the estrogen receptor (ER), and the progesterone receptor. TNBC exhibits high mortality and frequent chemotherapy resistance (1). A minority of TNBC cases are linked to hereditary homologous recombination defects (HRD), most commonly in the BRCA1 gene, and are treatable with PARP inhibitors (2). However, the majority of TNBC cases do not have an obvious hereditary explanation, and therefore the underlying DNA repair defects are more obscure (3). Cytotoxic chemotherapy is standard of care but is only partially effective; hence, lack of pathologic complete response (pCR) after neoadjuvant chemotherapy is frequent and associated with poor survival (4). Post non-pCR, salvage therapy with adjuvant capecitabine has modest efficacy (5). The programmed cell death receptor (PD1)-targeting antibody pembrolizumab is also approved for neoadjuvant TNBC treatment based on the results of the KEYNOTE-522 trial (6). In combination with neoadjuvant chemotherapy, pembrolizumab significantly prolongs event-free survival versus neoadjuvant chemotherapy alone (7). In contrast to metastatic TNBC, outcome improvements are not predicted by PD-L1 IHC in primary disease (8). Carboplatin also has efficacy in TNBC. The BrighTNess trial enrolled patients with stage II or III operable TNBC and randomized patient treatment to one of three arms prior to doxorubicin and cyclophosphamide: paclitaxel/carboplatin/veliparib (arm A), paclitaxel/carboplatin (arm B), or paclitaxel alone (arm C). Carboplatin-containing arms A and B showed significantly improved pCR compared with paclitaxel alone (53% and 58%, respectively, vs. 31%; ref. 9). The efficacy of carboplatin addition is supported by two other randomized neoadjuvant trials: CALGB 40603 (Alliance; ref. 10) and GeparSixto (11). Thus, in the absence of predictive markers for individual components of each regimen, the neoadjuvant treatment for TNBC involves up to seven different drugs.

Herein we describe the first study to deploy microscaled proteogenomics (12) to discover neoadjuvant chemotherapy...
response biomarkers in TNBC. Snap-frozen, optimal cutting temperature compound (OCT)-embedded core needle biopsies were accrued from patients enrolled in two clinical trials that investigated a simplified carboplatin and docetaxel regimen designed to be less toxic by omitting doxorubicin and cyclophosphamide (NCT02547987 and NCT02124902; ref. 13). This dataset included germline-matched tumor whole-exome DNA sequencing (WES), RNA sequencing (RNA-seq), and tandem mass tag (TMT)-based proteomics and phosphoproteomics. Analyses focused on the identification of biomarker associations with pCR, with the goal of identifying patients who would be better served with investigational drugs at diagnosis rather than suffer an ineffective standard of care. Multiple independent datasets were used to validate findings in the discovery analysis, including mRNA profiles of other TNBC clinical trials, IHC, preclinical therapeutic studies in patient-derived TNBC xenografts (PDX), and pan-cancer analysis using data from The Cancer Genome Atlas (TCGA).

RESULTS
Overview of the Proteogenomic Analysis Approach
OCT-embedded, snap-frozen core needle biopsies were accrued from consented patients with clinical stage 2 or 3 TNBC (70% Caucasian, 27% African American, and 3% other racial categories). Patients were subsequently treated with six cycles of neoadjuvant carboplatin and docetaxel combination chemotherapy (NCT02547987 and NCT02124902). Pretreatment samples from 59 patients had >25% tumor content (TC) and were ultimately analyzed. For 16 patients, an additional sample was obtained 48 to 72 hours after initiating chemotherapy. A Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) diagram demonstrates sample flow into different analytical pipelines (Fig. 1A). Using previously described BioTEXT sample processing and microscaled proteogenomics methods (12), frozen core biopsies were processed on a cryotome to produce 50-μm sections for analyte extraction interspersed with 5-μm sections to document TC. Alternating 50-μm sections were distributed into three different analyte preparation approaches to ensure even representation of analytes from different layers in the biopsy. Multiplex analyte enrichment was achieved for paired normal/tumor DNA exome sequencing (100x), RNA-seq, and quantitative, multiplexed (TMT) mass spectrometry (MS)–based proteomics and phosphoproteomics (Fig. 1B; Supplementary Tables S1–S3).

Sample-level mRNA to protein correlations deteriorated in seven samples with an average TC below 45% (Supplementary Fig. S1A). Based on this cutoff, a total of nine samples with proteomics data (including one sample that lacked RNA and one sample that lacked both RNA and protein) were therefore excluded from further bioinformatic analyses. TMT11 multiplexes were linked using a pooled sample common reference to serve as a denominator for calculating peptide and phosphosite ratios (12). The common reference samples showed very strong correlations across multiplexes, indicating consistent data quality (Supplementary Fig. S1B). For each qualified sample, DNA, RNA, and protein level information was available for an average of 10,500 genes (Fig. 1C) and phosphoproteomic analysis quantified ~27,000 phosphorylation sites in ~5,000 distinct phosphoproteins (Fig. 1C).

Comparable with previous Clinical Proteomic Tumor Analysis Consortium (CPTAC) proteogenomic analyses, median per gene mRNA to protein correlation was 0.37 (ref. 14; Supplementary Fig. S1C). Genes with significant positive RNA-protein correlations were enriched for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways involved in cellular respiration as well as amino acid and lipid metabolism. Genes with lower correlations were enriched in pathways containing large protein complexes serving the spliceosome, replication, transcription, and pyrimidine metabolism (Supplementary Fig. S1C). Consistent with previous observations, protein data significantly outperformed RNA data for coexpression-based gene function predictions (Supplementary Fig. S1D; refs. 12, 15–17).

A pairwise analysis was also conducted using 14 cases with baseline high TC (out of 16 pairs) matched to a second high TC specimen collected 48 to 72 hours after treatment (only 13 pairs had RNA data; Fig. 1B). Whereas immune-related pathways were downregulated upon treatment at both the RNA and protein level, cell-cycle and metabolic pathways (except glycolysis) were significantly upregulated specifically at the protein level (Fig. 1D; Supplementary Table S4). Induction of DNA replication and repair pathways linked to the cell cycle was observed, likely in response to genotoxic stress triggered by chemotherapy exposure (18). This observation was also present in the phosphorylation site data (Supplementary Fig. S1E). Sets of phosphosites induced by treatment overlapped with those established to be induced by nocodazole and ionizing radiation treatment, which is logical in the setting of docetaxel and carboplatin exposure. Increases in phosphorylation were also detected for targets of the cell cycle and DNA damage kinases CDK1, CDK2, and ATM (Supplementary Fig. S1E).

Exploration of Proteogenomic Pathway Signatures and Response to Chemotherapy
Primary study endpoints were pCR and residual cancer burden (RCB) in the surgical specimen where 0 indicates pCR and I to III indicate increasing levels of residual disease (19). PAM50 intrinsic subtype (20), TNBC type (21), and racial categories lacked association with pCR, as did other cohort-specific clinical metadata (Supplementary Fig. S1F). Expected associations for pCR with germline mutations in the homologous recombination genes BRCA1/2 and PALB2 (22) or with homologous recombination deficiency–associated Catalogue of Somatic Mutations in Cancer (COSMIC) signature 3 were also not observed (23, 24). These negative findings emphasize the limitations of our study in terms of sample size. However, an elevated COSMIC signature 6 score, indicating a mismatch repair defect (23, 24), was associated with high RCB (II or III; P = 0.03; Supplementary Fig. S1G). Gene set enrichment analysis (GSEA) of proteogenomic features (Supplementary Table S5) that differed by pCR status indicated upregulation of the Molecular Signatures Database (MSigDB) Hallmark metabolic pathways, including oxidative phosphorylation, fatty acid metabolism, and adipogenesis in samples without pCR. These associations were observed in the proteomic data but not at the mRNA level (Fig. 1E; Supplementary Table S6). In contrast, immune signaling (interferon alpha and gamma response) and cell cycle (G1–M...
checkpoint and E2F and MYC target) pathways were elevated in pCR cases in both the proteomic and transcriptomic datasets (Fig. 1E). Enrichment analysis of differential phosphorylation sites [posttranslational modification-set enrichment analysis (PTM-SEA); ref. 25] logically demonstrated elevated phosphoproteome-driven signatures in samples from pCR cases for treatment with inhibitors that generate DNA damage (etoposide, hydroxyurea, and ionizing radiation; Fig. 1F). Elevated MARK2 target sites were enriched in non-pCR tumors (Fig. 1F), corroborating prior evidence for higher MARK2 levels in cisplatin resistance in other cancer types (26, 27). Consistent with significantly elevated cell-cycle pathways observed in pCR samples in the RNA and protein data, CDK1, 2, and 7 and CDC7 target phosphosites were also significantly higher in pCR samples (Fig. 1F). Further sample-wise investigation of cell-cycle proteogenomic features revealed that multigene proliferation scores (MGPS), single-sample GSEA (ssGSEA) scores, and PTM-SEA scores for cell cycle–related signatures were significantly higher in pCR samples compared to non-pCR samples (Fig. 1D).
pathways and cyclin-dependent kinases were higher in pCR but were variable in non-pCR (Supplementary Fig. S2A). Of note, a subset of non-pCR samples had elevated CDK4 activity and Rb phosphorylation (highlighted in the box in Supplementary Fig. S2A), and Rb phosphorylation was marginally higher in non-pCR tumors (Supplementary Fig. S2B). To study the therapeutic significance of these findings, TNBC cell lines from the DepMap resource were explored (www.depmap.org). In this database, higher Rb protein was associated with reduced carboplatin response but enhanced CDK4/6 inhibitor response (Supplementary Fig. S2C).

**Immunome Pathways and Response to Chemotherapy**

Because interferon alpha and gamma response signatures were elevated in samples from pCR cases, signals from the immune microenvironment were further explored (Fig. 2A). Protein-derived immune stimulatory scores, previously found to be well correlated with immune infiltration (14), as well as
PD-L1 RNA, protein, and phosphorylation levels, were significantly higher in pCR-associated samples (Fig. 2A and B). Nonsynonymous mutation load was associated with neither pCR (Wilcoxon rank sum P = 0.57, median for pCR = 77, median for non-pCR = 78) nor immune scores (Spearman rho = -0.17, P = 0.25), suggesting increased mutation burden was not a strong determinant of immune infiltration in this TNBC dataset. Rather, immune scores were significantly anticorrelated with chromosomal instability index (CIN; Spearman Rho = -0.61, P = 6.2e-6; Fig. 2C). Both PD-L1 protein and phosphoprotein levels significantly correlated with PD-L1 IHC (Fig. 2D and E). Similar correlations were also observed between PD-L1 RNA and IHC (Supplementary Fig. S2D). Representative IHC images for high and low PD-L1 staining are shown in Supplementary Fig. S2E and S2F, respectively.

**Metabolic Pathway Analysis and Response to Chemotherapy**

As noted above, (Fig. 1D and E) metabolic pathway enrichment appeared specific to proteomic data (with FDR correction). Both GSEA (Fig. 1E) and ssGSEA showed that differential metabolic pathways, including oxidative phosphorylation, adipogenesis, and fatty acid metabolism, as
well as glycolysis, were significantly higher in pretreatment tumors without subsequent pCR (Fig. 3A). Further analyses of the individual protein levels identified many chemotherapy resistance–associated metabolic proteins, such as those directly involved in the tricarboxylic acid cycle (ACO2, FH, MDH2, SUCLG1, SUCLG2, PDP1, and DLAT), the electron transport chain (SDHC and UQCR10), fatty acid metabolism (CRAT, ACADS, ACAT1, DECR1, ECHS1, and HADHB), and amino acid catabolism (ALDH6A1, HMGCL, DBT, and BCKDHB; Fig. 3B). Although pCR-associated metabolic pathway scores were more robust at the proteomic data level than transcriptomic data, this did not equate to lack of mRNA and protein correlation for all metabolism gene products associated with non-pCR. A subset (29 of 43) from the relevant Hallmark metabolic pathways showed sufficient protein–mRNA correlation to allow independent validation of metabolic gene expression associations with pCR at the mRNA level (Fig. 3B) in the BrighTNess trial dataset (9). In this study, patients in arms A and B received combination treatment with carboplatin and paclitaxel plus/minus veliparib (addition of which did not affect outcomes), as well as subsequent treatment with doxorubicin and cyclophosphamide (9). Baseline RNA expression data for the subset of metabolism-associated resistance genes with high mRNA–protein correlation were used for association with pCR status on data from both arms A and B combined. For this analysis, the two carboplatin- and paclitaxel-containing arms were fusionable in other datasets, the association of LIG1, XRCC1, and ERCC2 mRNA with pCR and RCB was evaluated (Fig. 4D). While single-copy loss (GISTIC $\leq$ 1) status in a subset of patients showed enrichment in the DNA damage repair (DDR) pathway GSEA of the genes on cytoband 19q13.31–33 (amplified) and 8q21.3 (deleted) cytobands was elevated and suppressed, respectively, in non-pCR versus pCR tumors (Fig. 4B). Four genes located at 8q21.3, RMDN1, CPNE3, DECR1, and OTUD6B, showed higher mRNA and protein expression in non-pCR tumors (Supplementary Fig. S4B). In addition, RIPK2, which may mediate metastasis in advanced breast cancer (28), also located on 8q21.3, was significantly higher in non-pCR tumors but only at the protein level. Similarly, four genes located on 19q13.31–33, LIG1, PPPSC, BCL3, and NOSIP, showed lower mRNA and protein expression in non-pCR tumors (Fig. 4C). Both mRNA and protein level expression from these coordinate downregulated genes were confirmed to be suppressed in association with single-copy $LIG1$ loss (GISTIC $= -1$) status in a subset of non-pCR–associated samples (Supplementary Fig. S4C). Hallmark pathway GSEA of the genes on cytoband 19q13.31–33 showed enrichment in the DNA damage repair (DDR) pathway, with $LIG1$, XRCC1, POLD1, and ERCC2 comprising the leading-edge genes (Fig. 4D). LIG1 showed the strongest association with treatment response at the protein level, followed by POLD1 (Fig. 4D).

To determine whether these observations were reproducible in other datasets, the association of $LIG1$, XRCC1, POLD1, and ERCC2 mRNA with pCR and RCB was evaluated at the mRNA level in the BrighTNess trial. For this analysis, the two carboplatin- and paclitaxel-containing arms were

Figure 2. (Continued) D and E, Scatter plots showing the Pearson correlation between PD-L1 IHC levels with PD-L1 protein (D) and phosphoprotein levels (E). pCR cases are shown in green and non-pCR in orange.
Proteogenomic Markers of Chemotherapy Resistance in TNBC

**Figure 3.** Proteogenomic features associated with the lack of pCR in TNBC tumors. A, Heat map showing ssGSEA normalized enrichment score for metabolic Hallmark pathways that are significantly higher in non-pCR cases, arranged by RCB 0 (pCR) and RCB I/II/III (non-pCR). Shown are the four pathways that showed significant enrichment at either the RNA or protein level in Fig. 1E. Single-sample pathway enrichment scores were assessed at the level of mRNA (yellow), protein (blue), and phosphoprotein (red). The Wilcoxon rank sum test was used to compare scores for non-pCR vs. pCR scores; *, \( P < 0.05 \). NA, not available. B, Membership of differentially regulated proteins to pathways highlighted in A. Proteins (rows) belonging to a given pathway (columns) are shown in light green. The differential expression (DE) at protein and mRNA levels for each gene along with mRNA–protein correlation scores are shown as signed \(-\log_{10} P\) value (signedp). C, A multiomics metabolic gene signature derived for genes that are correlated at mRNA and protein levels was further investigated in patients treated with carboplatin and paclitaxel in the BrighTNess clinical trial (treatment arms A and B) for which RNA-seq data were available. The mean mRNA expression score for this signature was significantly higher in higher RCB tumors.
Figure 4. Discovery of DNA repair and replication components enriched in non-pCR TNBC tumors. A, Cytobands enriched in genes differentially expressed between non-pCR and pCR for both mRNA and protein. To identify upregulated or downregulated features overrepresented in certain cytobands within the chromosome, GSEA was used to identify regions from chromosomal location databases enriched with differential genes [GSEA input was ranked expression list (signed $-\log_{10} P$ value) from Wilcoxon rank sum tests]. Overrepresented cytobands that were either enriched or depleted using differentially expressed mRNA and protein are indicated in B, and the overlapping sets were used for further analysis. B, Plot showing significantly enriched or depleted cytobands obtained by running differential mRNA and protein ranked lists through GSEA. NES, normalized enrichment score. Genes downregulated in non-pCR samples corresponding to cytoband 19q13.31–33 are indicated in C. C, Venn diagram showing differential (non-pCR vs. pCR) mRNA and proteins located on cytoband 19q13.3. (continued on following page)
Proteogenomic Markers of Chemotherapy Resistance in TNBC

3.5

D

Proteogenomic Markers of Chemotherapy Resistance in TNBC
the four DDR genes located within 19q13.31–33, or partial response (CR + SD) as opposed to samples associated with a complete or partial response (CR + PR). Combined to parallel the docetaxel and carboplatin treatment in this study (9). LIG1 and POLD1 were confirmed to be significantly downregulated in baseline tumor samples from patients who experienced residual disease (Fig. 4E). Similar differences were not observed in the paclitaxel-only arm, although the sample size was smaller (treatment arm C, Wilcoxon, P = 0.11). Finally, a trial where a modest number of patients were treated with single-agent cisplatin neoadjuvant therapy was interrogated (31). Consistent with the other datasets, LIG1 mRNA levels were significantly lower in samples associated with stable or progressive disease (SD + PD) as opposed to samples associated with a complete or partial response (CR + PR). Of the four DDR genes located within 19q13.31–33, LIG1 was the most consistently associated with chemotherapy resistance and poor metastasis-free survival across datasets (Fig. 4E and F; Supplementary Fig. S4E and S4F).

Molecular Features of TNBCs Harboring LIG1 Deletion

The associations between LIG1 deletion and/or reduced expression with tumor pathophysiologic features were further investigated in the discovery set. Low LIG1 copy-number levels (GISTIC = −1) were observed in 8 of the 31 (∼26%) tumors without pCR (Fig. 5). LIG1 copy-number log ratios were strongly and positively correlated with the level of both LIG1 mRNA (Pearson, R = 0.67, P = 2.8e−06) and LIG1 protein (R = 0.55, P = 8.2e−05; Supplementary Fig. SSA and SSB). At the genomic level, COSMIC HRD signature 3 was lower in tumors with LIG1 loss (t test, P = 0.01; Fig. 5). In contrast, tumors harboring LIG1 loss exhibited significantly higher CIN (t test, P = 0.0003; Fig. 5). Although no significant differences were observed in immune stimulatory scores when LIG1 loss tumors were compared other...
tumors, tumors with LIG1 loss had lower immune stimulatory scores when compared with tumors that were associated with pCR (P = 0.01; Supplementary Fig. S5C). At the level of phosphosite expression-based PTM-SEA (25) analysis, the IL33 pathway was significantly downregulated in LIG1-loss tumors (Supplementary Fig. S5D and S5E). Tumors with LIG1 loss also had significantly higher protein-based proliferation scores (p-MGPS; Wilcoxon P = 0.004; Fig. 5) as well as upregulation of CDK1/2 activity (Supplementary Fig. S5D) in PTM-SEA analysis of differential phosphosites (25), supporting increased cell-cycle activity (FDR P < 0.05). Collectively, these results suggest that the loss of LIG1 is associated with a constellation of poor prognosis features including higher proliferation rates, a less active immune microenvironment, and higher copy-number instability. Furthermore, when the phosphoproteomic data were examined, signatures of EGFR (gefitinib) and PI3K (wortmannin) perturbations were significantly enriched in LIG1-loss tumors but in a negative direction (Supplementary Fig. SS and SSE). Because LIG1-loss tumors have suppressed EGFR and PI3K signaling, they may be less responsive to EGFR, PI3K, or AKT inhibition.

**LIG1 and Chemotherapy Response in Model Systems**

When chemotherapy-resistance biomarkers are identified, the question arises as to whether the biomarker relationship is drug selective. Model systems can be useful in this regard because patients almost always receive multiple drugs. Another concern is whether a biomarker is associated with intrinsic resistance, acquired resistance, or both. Patient data suggested a higher frequency of LIG1 copy loss in metastatic disease (Supplementary Fig. S6A). Patterns of 19q13.31–33 loss during malignant progression were therefore explored using three orthotopic PDX models generated from a single patient on the discovery trial NCT02544987. WHIM68 grew from the pretreated breast primary tumor, WHIM74 from a surgical sample accrued after 5 months of neoadjuvant carboplatin and docetaxel, and WHIM75 from a liver metastasis that appeared 1 year after treatment initiation. Proteogenomic analysis...
revealed a progressive loss of LIG1 at the copy-number, mRNA, and protein levels as the tumor progressed to a chemotherapy-resistant state (Fig. 6A). Progressive loss of LIG1 protein was confirmed by Western blotting (Fig. 6A; Supplementary Fig. S6B) along with similar reductions of POLD1 and XRCC1 protein expression. Consistent with the progressive loss of chemotherapy sensitivity observed clinically, WHIM68, which expressed the highest LIG1 protein levels, was most sensitive to carboplatin, whereas WHIM74 and 75, which displayed progressive LIG1 loss at the copy-number, mRNA, and protein levels, were insensitive to carboplatin treatment. *P* values derived from a general linear model within each PDX were computed using estimated mean log2 fold changes in tumor volume at day 28 versus day 0 for each treatment arm. (continued on next page)

Figure 6. LIG1 association with advanced TNBC disease in preclinical models. A, Proteogenomic status of LIG1, POLD1, and XRCC1 in three PDX models derived from longitudinal biopsies from the same patient with TNBC prior to any treatment (WHIM68), at the time of surgery after completing 5 months of neoadjuvant carboplatin and docetaxel (WHIM74), and from a liver metastasis 1 year after treatment initiation (WHIM75). Mutation and copy-number data were derived from WES and RNA from RNA-seq, and protein data were obtained from TMT proteomics generated by this current study. Bottom, representative Western blots from three biological repeats for LIG1, POLD1, and XRCC1 protein levels. GAPDH was used as a loading control. WT, wild-type. *, stop codon. B, Tumor volume was measured in three PDX models. Black and red lines indicate changes in tumor volume in PDXs treated with vehicle and carboplatin, respectively. WHIM68, with the highest LIG1 protein levels, was most sensitive to carboplatin, whereas WHIM74 and 75, which displayed progressive LIG1 loss at the copy-number, mRNA, and protein levels, were insensitive to carboplatin treatment. *P* values derived from a general linear model within each PDX were computed using estimated mean log2 fold changes in tumor volume at day 28 versus day 0 for each treatment arm. (continued on next page)
Figure 6. (Continued) C, Box plots showing LIG1 mRNA levels in TNBC PDXs categorized into CR and non-CR groups. After 4 weeks of carboplatin treatment, CR was defined as PDXs with nonpalpable tumors, and non-CR was defined as PDXs with residual tumors with measurable dimensions. The Wilcoxon rank sum test was used to compare the two groups. RSEM, expected counts from RNA-seq by expectation-maximization. D, Association between LIG1 copy-number loss and treatment response in PDX organoids (PDXO) obtained from the BCaPE database. Carboplatin and docetaxel are highlighted in red.
Proteogenomic Markers of Chemotherapy Resistance in TNBC

[ref. 33; Breast Cancer PDTX Encyclopaedia (BCaPE) database]. This dataset demonstrated that LIG1 copy-number loss was uniquely correlated with carboplatin resistance among over 100 drugs tested (Fig. 6D).

**LIG1 Copy-Number Loss Is Associated with Poor Progression-Free Survival and CIN across Multiple Cancer Types**

Gene copy-number analysis of tumors characterized by TCGA demonstrated that LIG1 single-copy loss is present in other cancer types. In the TCGA pan-cancer dataset, LIG1 heterozygous loss was associated with poor progression-free survival (PFS; Fig. 7A; \( P < 0.0001 \)), significantly higher fraction genome altered (Fig. 7B), and lower signature 3 scores (suggesting proficient homologous recombination; Fig. 7C). Cancer types driving these relationships include endometrial carcinoma (hazard ratio = 2.23, \( P = 0.02 \)), head and neck squamous cell carcinoma (hazard ratio = 1.46, \( P = 0.03 \)), prostate adenocarcinoma (hazard ratio = 2.07, \( P = 0.02 \)), colon adenocarcinoma (hazard ratio = 1.75, \( P = 0.03 \)), and most convincingly renal papillary cell carcinoma (hazard ratio = 4, \( P = 0.0001 \); Fig. 7D). Despite a marginal association between PFS and LIG1 loss in testicular germ cell tumors, the seminoma subtype, which demonstrates exquisite sensitivity to carboplatin (34), displayed no cases of LIG1 loss (Supplementary Fig. S7A). Higher genomic instability was observed with LIG1 loss in several other cancers (Fig. 7E, fraction genome altered in TCGA cohorts; Supplementary Fig. S7B, CIN in CPTAC cohorts).

**DISCUSSION**

The absence of a baseline pCR predictor is a persistent unmet need for the precision treatment of TNBC. Patients without pCR suffer prolonged exposures to toxic and ineffective treatment and therefore do not receive alternative treatment soon enough. Additionally, PD-L1 IHC assays have failed to predict the benefit of immune-checkpoint blockade in TNBC (35). Thus, alternative biomarkers for antitumor immunity are required. This study suggests that integrated proteogenomic characterization provides more extensive information on the immune microenvironment that could be used to complement PD-L1 IHC. Although a TMT-based proteomic assay for PD-L1 would not be practical, targeted proteomic assays optimized for quantitative measurement using heavy isotope–labeled peptides for multiple immune response components are an efficient and cost-efficient approach that could complement IHC (36). Secondly, we observed a novel association for baseline oxidative phosphorylation and fatty acid metabolism gene products with chemoresistance in TNBC. These findings are supported by functional studies in TNBC model systems demonstrating a role for oxidative phosphorylation and fatty acid metabolism as drivers of TNBC chemoresistance (37, 38). In fact, fatty acid synthase inhibition using the proton pump inhibitor omeprazole in combination with neoadjuvant chemotherapy in patients with TNBC is currently being evaluated in a phase II trial (NCT02595372). pCR prediction models could therefore be strengthened by the inclusion of protein level analysis of these pathways. The cellular origin of these resistance-associated metabolic signals is unresolved. An additional possibility is immunosuppressive tumor-associated macrophages with high lipid content (39). A third class of potential pCR predictors are G2-M checkpoint components, E2F regulators, and MYC target pathways. For example, TNBC tumors with high/intact Rb protein and phosphorylation levels have lower pCR rates and lower levels of proliferation and E2F target gene expression than tumors with the loss of Rb protein (14). CDK4/6 or CDK2 inhibitors could therefore be an alternative treatment for RB intact TNBC. Finally, the proteomic analysis clearly assists in the prioritization of genomic chromosomal alterations associated with pCR status, exemplified herein by the identification of LIG1 as a TNBC chemotherapy-resistance and multicancer-type poor prognosis marker. The finding from preclinical models that LIG1 loss is a selective biomarker for carboplatin resistance is provocative. The use of carboplatin adds toxicity to an already toxic anthracycline-based regimen and could potentially be avoided in LIG1-depleted tumors.

Regarding LIG1 loss as a potential pathogenic event in TNBC, there are already mechanistic studies of LIG1 loss that support this hypothesis. LIG1 encodes an ATP-dependent DNA ligase that seals DNA nicks during replication, recombination, and a variety of DNA damage responses (40). Of the three DNA ligases in the human genome (LIG1, 3, and 4), LIG1 is the main enzyme responsible for ligating Okazaki fragments during lagging-strand synthesis at the replication fork during S-phase (41–43). LIG1 also ligates single-stranded or double-stranded DNA breaks in various DDR pathways, including long-patch base-excision repair, nucleotide-excision repair, and alternative nonhomologous end-joining repair (44, 45). A phenotype for LIG1 deficiency in humans was first identified in an immunodeficient patient with homozygous germline hypomorphic LIG1 alleles causing impaired Okazaki fragment ligation (46). Insufficient LIG1 activity results in the accumulation of replication intermediates that cause single-stranded and double-stranded breaks (DSB; refs. 47, 48), ultimately leading to reduced genome integrity. In transgenic mice, hypomorphic LIG1 alleles were associated with high susceptibility to cancer formation (49). However, the relevance of these observations can be challenged in the setting of TNBC, because single-copy LIG1 loss observed in our studies may not produce sufficient functional deficiency to generate a phenotype. However, codeletion of LIG1, POLQ1, and XRCC1 on 19q13.31–33 may produce a hemizygous compound deficiency phenotype because all three genes serve lagging-strand synthesis. XRCC1 is particularly noteworthy because LIG3/XRCC1 provides a backup pathway for LIG1-mediated DNA ligation during DNA repair and lagging-strand DNA synthesis (50).

The presence of LIG1 loss was found to be orthogonal to the HRD mutational signature 3. Consequently, LIG1-deficient cells may be required to be proficient in DSB repair, that is, HRD and LIG1 loss are orthogonal routes to TNBC pathogenesis, and this potentially could explain the correlation with carboplatin insensitivity. The PDX study (Fig. 6) hints at this, as the model derived from the pretreatment sample (WHIM68) had a BRCA2 frameshift mutation and no LIG1 loss, and the subsequent carboplatin-resistant lines (WHIM74 and 75) had lost the BRCA2 mutation and gained a LIG1 hemizygous deletion. It remains unclear why LIG1 loss is so strongly associated with chromosomal instability across cancer types, and mechanistic studies connecting these events
Shown are the only five cancers (HNSC, UCEC, COAD, PRAD, and KIRP) that displayed a significant association between loss and adverse prognosis.

**Figure 7.** Pan-cancer analysis of LIG1 loss. **A,** Kaplan–Meier curve showing significantly reduced (log-rank P value) PFS for tumors with single-copy loss of LIG1 (HETLOSS, GISTIC ≤ −1, indicated in orange) in the TCGA pan-cancer cohort. **B,** Box plot showing higher fraction genome altered (FGA) in tumors with LIG1 copy-number-loss tumors (shown in teal) relative to tumors that were LIG1 wild-type or displayed LIG1 gain (shown in orange). **C,** Violin plot showing significantly lower (Wilcoxon rank sum test) COSMIC signature 3 scores in LIG1-loss tumors (shown in teal) relative to tumors that were LIG1 wild-type or displayed LIG1 CNA status.

**Table D**

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>LIG1 WT/gain</th>
<th>LIG1 loss</th>
<th>LIG1 loss %</th>
<th>HR</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCS</td>
<td>39</td>
<td>17</td>
<td>30.36</td>
<td>0.58</td>
<td>0.1</td>
</tr>
<tr>
<td>LGG</td>
<td>255</td>
<td>256</td>
<td>50.1</td>
<td>0.64</td>
<td>0.002</td>
</tr>
<tr>
<td>MESO</td>
<td>72</td>
<td>15</td>
<td>17.24</td>
<td>0.64</td>
<td>0.5</td>
</tr>
<tr>
<td>GBM</td>
<td>481</td>
<td>94</td>
<td>46.35</td>
<td>0.79</td>
<td>0.08</td>
</tr>
<tr>
<td>PAAD</td>
<td>161</td>
<td>22</td>
<td>12.02</td>
<td>0.97</td>
<td>0.7</td>
</tr>
<tr>
<td>LUSC</td>
<td>350</td>
<td>137</td>
<td>28.13</td>
<td>0.89</td>
<td>0.5</td>
</tr>
<tr>
<td>READ</td>
<td>133</td>
<td>22</td>
<td>14.19</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>OV</td>
<td>278</td>
<td>294</td>
<td>51.4</td>
<td>0.98</td>
<td>0.8</td>
</tr>
<tr>
<td>LUAD</td>
<td>321</td>
<td>190</td>
<td>37.18</td>
<td>0.98</td>
<td>0.9</td>
</tr>
<tr>
<td>SKCM</td>
<td>288</td>
<td>79</td>
<td>21.53</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ESCA</td>
<td>131</td>
<td>51</td>
<td>28.02</td>
<td>1.03</td>
<td>0.9</td>
</tr>
<tr>
<td>LIHC</td>
<td>324</td>
<td>43</td>
<td>11.72</td>
<td>1.06</td>
<td>0.6</td>
</tr>
<tr>
<td>BRCA</td>
<td>867</td>
<td>203</td>
<td>18.97</td>
<td>1.09</td>
<td>0.7</td>
</tr>
<tr>
<td>SARC</td>
<td>171</td>
<td>82</td>
<td>32.41</td>
<td>1.11</td>
<td>0.5</td>
</tr>
<tr>
<td>CESC</td>
<td>257</td>
<td>36</td>
<td>12.29</td>
<td>1.13</td>
<td>0.7</td>
</tr>
<tr>
<td>STAD</td>
<td>364</td>
<td>74</td>
<td>16.89</td>
<td>1.26</td>
<td>0.3</td>
</tr>
<tr>
<td>BLCA</td>
<td>344</td>
<td>64</td>
<td>15.69</td>
<td>1.29</td>
<td>0.2</td>
</tr>
<tr>
<td>TGCT</td>
<td>91</td>
<td>58</td>
<td>38.93</td>
<td>1.44</td>
<td>0.3</td>
</tr>
<tr>
<td>HNSC</td>
<td>421</td>
<td>96</td>
<td>18.57</td>
<td>1.46</td>
<td>0.03</td>
</tr>
<tr>
<td>COAD</td>
<td>397</td>
<td>40</td>
<td>9.15</td>
<td>1.75</td>
<td>0.03</td>
</tr>
<tr>
<td>PRAD</td>
<td>452</td>
<td>37</td>
<td>7.57</td>
<td>2.07</td>
<td>0.02</td>
</tr>
<tr>
<td>UCEC</td>
<td>467</td>
<td>56</td>
<td>10.71</td>
<td>2.23</td>
<td>6e−04</td>
</tr>
<tr>
<td>KIRP</td>
<td>265</td>
<td>18</td>
<td>6.36</td>
<td>4</td>
<td>1e−04</td>
</tr>
</tbody>
</table>

**E**

Forest plot showing the impact of LIG1 copy-number loss on PFS by cancer type along with LIG1 wild-type (WT)/gain/loss frequency, HR, and corresponding P value. BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; COAD, colon adenocarcinoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PRAD, prostate adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; UCEC, uterine corpus endometrial carcinoma; UCS, uterine cervix carcinoma. **F** Box plot showing significantly higher (by Wilcoxon rank sum test) FGA (representing chromosomal instability) in tumors that had LIG1 copy-number loss versus tumors with either wild-type LIG1 or LIG1 copy-number gain. Shown are the only five cancers (HNSC, UCEC, COAD, PRAD, and KIRP) that displayed a significant association between LIG1 loss and adverse prognosis.
are an important next step. However, it can be hypothesized that cells that enter mitosis with unrepaired lagging strands are at risk for chromosomal breakage, illicit chromosomal fusion events, and aneuploidy.

In conclusion, our findings emphasize the potential of microscaled proteogenomic approaches for the investigation of cancer treatment resistance. Follow-up mechanistic studies are clearly warranted, not just for LIG1-related biology but also, for example, the role of lipid-related metabolic signatures in chemotherapy resistance. However, the lack of complete mechanistic insight does not diminish the clinical importance of novel chemotherapy drug-selective predictive biomarkers in a setting where a genomic approach or transcriptomic analyses have yet to produce actionable models.

METHODS

Clinical Sample Collection

Eligible patients for the two clinical trials (NCT02547987 and NCT02124902) included pre- or postmenopausal women at least 18 years old with clinical stages II/III ER-negative and HER2-negative (0 or 1+ by IHC or FISH negative) invasive breast cancer. The study was approved by the Institutional Review Board at both participating sites, WashU and BCM, and written informed consent from the patients was obtained. The studies were conducted in accordance with recognized ethical guidelines and followed the Declaration of Helsinki and Good Clinical Practice guidelines. All patients were uniformly treated (without randomization or blinding) with neo-adjuvant intravenous docetaxel 75 mg/m² and carboplatin every 21 days for 6 cycles with granulocyte colony-stimulating factor support (13). Research tumor biopsies for correlative studies were obtained at baseline prior to chemotherapy and on cycle 1, day 3. On-treatment biopsy on cycle 1, day 3 and biopsy at the time of relapse were optional. Details of the clinical cohort have recently been published (13). Treatment response information was provided by clinical teams associated with these trials and RCB was calculated using the RCB calculator (http://www3.mdanderson.org/app/medcalc/index.cfm?pagemain=rcbcalc3).

IHC

For IHC, cut tissue sections (5 mm) on charged glass slides were baked for 10 to 12 hours at 58°C in a dry slide incubator, deparaffinized in xylene, and rehydrated via an ethanol step gradient. The IHC slides were stained for CD3 and PD-L1. Pathology slide scoring was performed using established professional guidelines for TNBC when appropriate. All IHC results were evaluated against positive and negative tissue controls. See Supplementary Data and Methods for more details.

Genomic Analysis

WES. Tumor DNA was extracted from fresh-frozen biopsies and matched leukocyte germline DNA from blood samples. WES data were generated for 59 unique baseline DNA samples using the Illumina platform. For this, paired-end libraries were constructed as described previously (51) with the modifications described in the Supplementary Data and Methods “Whole exome sequencing (WES)” section.

RNA-seq Data. Transcriptome data were generated for 60 samples in this study. For this, strand-specific, poly-A+ RNA-seq libraries for sequencing on the Illumina platform were prepared as previously described (52). See the Supplementary Data and Methods “RNA-Seq data” section for additional details. Between 59.96 and 112.62M total reads were generated for these 60 samples. The average strand specificity and RNA rate were 97.04% and 1.79%, respectively. The transcripts for 22,868 to 27,856 genes were detected in these samples.

The paired-end reads were mapped to the human genome version GRCh38.d1.v1 (from Genomic Data Commons) using STAR-2.7.1a. Gene expression estimation was performed using RSEM-1.3.1, and expected counts from RNA-seq by expectation-maximization (RSEM) and fragments per kilobase per million mapped reads (FPKM) values were upper-quartile normalized. Unless otherwise noted, gene median-centered log-transformed RSEM values were used for the analyses presented here.

Somatic and Copy-Number Variant Calling. Somatic variants were called using paired tumor and blood normal from WES data. Tools used for somatic variant calling were Strelka2, MuTect2, CARNAC, and Pindel (v.0.2.56). Filtering steps are described in the Supplementary Data and Methods “Somatic and copy number variant calling” section. Similarly, germline mutations were called by comparing normal WES against the reference genome. Hg19.USC.adl_miR.140312.refgene was used to map the copy-number information to genes. COSMIC mutational signature scores for every sample were estimated using deconstructSigs (53).

For somatic CNA analysis, bam files were processed by the CopywriteR package (54) to derive log, tumor-to-normal copy-number ratios, and the circular binary segmentation algorithm (55) implemented in the CopywriteR package was used for the copy-number segmentation with the default parameters. CIN for each chromosome in each sample was inferred from the segmentation data using a weighted-sum approach in which the absolute values of the log, ratios of all segments within a chromosome were weighted by the segment length and summed up (16). The genome-wide chromosome instability index (CIN) was derived by adding up the instability scores for all 22 autosomes in each sample. MSISensor (56) was used to calculate somatic microsatellite instability (MSI) counts. GISTIC2 (57) was used to retrieve gene-level copy-number values and call significant CNAs in the cohort. A threshold of 20.3 was applied to log, copy-number ratio to identify gene-wise gain or loss of copy number, respectively. Each gene of every sample was assigned a thresholded copy-number level that reflected the magnitude of its deletion or amplification. These are integer values ranging from −2 to 2, where 0 means no amplification or deletion of magnitude greater than the threshold parameters described above. Amplifications are represented by positive numbers: 1 means amplification above the amplification threshold; 2 means amplification larger than the arm-level amplifications observed in the sample. Deletions are represented by negative numbers: −1 means deletion beyond the threshold; −2 means deletions greater than the minimum arm-level copy number observed in the sample. For the pan-cancer analysis, GISTIC value ±2 exceeds the high-level thresholds for amplifications/deep deletions, and that with ±1 exceeds the low-level thresholds but not the high-level thresholds. The low-level thresholds are just the “amphfresh” and “delTfresh” noise threshold input values to GISTIC (typically 0.1 or 0.3) and are the same for every threshold.

Proteomics Data Generation and Analysis

Proteomic Sample Preparation. Samples were prepared for proteomic analysis as described in a previous microscaled proteogenomic study (12) with minimal alterations. The details are described in the Supplementary Data and Methods “Proteomic sample preparation” section. For TMT labeling, a total of 30 μg peptides in 100 μL 50 mmol/L HEPES, pH 8.5, were labeled with 240 μg TMT reagent for an 8:1 TMT:peptide ratio and incubated at 25°C for 1 hour. Excess TMT reagent was quenched by incubating with 5 μL 5% hydroxylamine (Sigma) for 15 minutes. Samples within each plex were combined according to the ratios determined to achieve sample representation within ±15% error margin to all other samples. The combined
peptides were desalted on a 100-mg tC18 Sep-Pak (Waters), eluted with 50% acetonitrile/0.1% FA, and dried in a vacuum centrifuge.

**Experimental Design for Proteomics and Phosphoproteomics.** Samples were analyzed in a TMT11 format as described above. To measure relative protein and phosphosite expression, common references were constructed. The first core common reference consisted of peptide material from all clinical core samples, such that an even proportion was contributed for each of the 60 patients. The second common reference (“prospective BCCA CR”) was from a previous large cohort breast cancer proteomics study (14). Protein and phosphosite expression was reported as the TMT intensity ratio between each sample and the core common references within each plex. For analysis of clinical core samples, eight TMT 11-plexes each contained peptides from nine core needle biopsies in the first nine channels. If available, paired pre- and posttreatment tumor samples from a patient were grouped within the same 11-plex. As a quality control measure, we obtained protein and phosphopeptide ratios between the prospective BCCA CR and the core common reference, and the results are shown in Supplementary Fig. S1C.

**Basic Reverse-Phase Fractionation and Phosphoenrichment**

For basic reverse-phase fractionation, ~330 μg of peptides were dissolved in 500 μL of 5 mM ammonium formate and 5% acetonitrile using an offline Agilent 1260 LC with a 30-cm long, 2.1-mm inner diameter C18 column, running at 200 μL per minute in a total of 72 fractions, and further concatenated into 18 fractions for proteome analysis and 6 fractions for Fe⁺ immobilized metal affinity chromatography-based phosphoproteomics analysis. The details of this method have been described (12) and appear in the Supplementary Data and Methods “Basic reverse phase fractionation and phosphoenrichment” section.

**Proteomic Data Acquisition and Processing**

Proteome and phosphoproteome data acquisition was performed with a Proxeon nLC-1200 coupled to Thermo Lumos instrumentation with parameters described in the Supplementary Data and Methods “Proteomic data acquisition and processing” section.

Raw files were searched against the human (clinical cores) or humanRefSeq protein databases complemented with 553 small open reading frames (smORF) and common contaminants (Human: RefSeq.20171003_Human_refSeq_20171003_ref_553smORFs.fasta) using Spectrum Mill (Broad Institute) and parameters described in the Supplementary Data and Methods “Proteomic data acquisition and processing” section.

**Quantification, Normalization, and Filtering of Proteomics Data**

Before the calculation of protein and phosphopeptide ratios, reporter ion signals were corrected for isotopic impurities. Relative abundances of proteins and phosphosites were selected as the median of TMT reporter ion intensity ratios from all peptide spectral matches (PSM) matching to the protein or phosphosite. PSMs were excluded if they lacked a TMT label, had a precursor ion purity <50%, or had a negative delta forward-reverse score. To normalize across 11-plex experiments, TMT intensities were divided by the common reference for each protein and phosphosite. Log₂ TMT ratios were further normalized by median centering and median absolute deviation scaling. Proteins and phosphosites quantified in fewer than 30% of samples (i.e., missing in >70% of samples) were removed from the respective datasets.

**PDX Proteomics Data Generation and Analysis**

For the PDX experiment, cryopulverized PDX tumor tissues were lysed and digested as described above. Peptides (50 μg) were dissolved in 200 μL 50 mMol/L HEPES, pH 8.5, and labeled with 400 μg of the TMT reagent. TMT sample generation, basic reverse fractionation, and proteomic analysis were performed identically to that of clinical core biopsies. Raw files were searched against the human and mouse (PDX samples) UniProt protein databases complemented with 553 smORFs and common contaminants (human and mouse: UniProt. human.mouse.20171228.RlsnrNF.553smORFs.264contams.fasta) using the Spectrum Mill subgroup-specific option described in the Supplementary Data and Methods “PDX proteomics data generation and analysis” section.

**Data Quality Control and Differential Expression and Pathway Enrichment Analysis**

Samples with estimated TC below 45% were entirely removed from the dataset due to a lack of RNA-to-protein correlation in these samples (Supplementary Fig. S1B). The Wilcoxon rank sum test in R was used to identify genes (RNA), proteins, phosphosites, and phosphoproteins (mean of all sites on a given protein) that were differential between samples from pCR and non-pCR cases (Supplementary Table S5) and between samples with LGI loss (GISTIC = −1) and those without a loss (LGI wild-type/gain GISTIC ≥ 0). WebGestaltR (58) and PTM-SEA (25) were used to identify MSigDB Hallmark pathways (gene-level data) and posttranslational modification (PTM) signature sets (phosphosite-level data), respectively, that show enrichment in pCR or non-pCR tumors by applying the GSEA/PTM-SEA algorithms to signed (by direction of change) log₁₀ P values from the differential expression analysis (Supplementary Table S6). Additionally, the ssGSEA R package (59, 60) was applied to RNA, protein, and phosphoprotein data, and scores for Hallmark pathways were obtained for individual samples (Supplementary Table S6). Normalized enrichment scores (NES) were utilized for visualization purposes. The Wilcoxon signed rank test in R was used for paired differential analysis of on-treatment to baseline measurements for RNA, protein, phosphosite, and phosphoprotein data for 14 patients with matched on-treatment and baseline biopsies (only 13 had matched RNA data). GSEA using WebGestaltR and PTM-SEA was applied to signed log₁₀ transformed P values from this analysis. PTM-SEA was also applied to phosphosite log₂ TMT ratios for each baseline sample to obtain single-sample kinase activity scores (NES for kinase target PTM sets).

**Functional Prediction Based on Gene Coexpression**

Coexpression network construction using mRNA and protein expression data and network-based gene function prediction for KEGG pathways were performed as previously described (15) using OmicsEV (https://github.com/bzhanglab/OmicsEV).

**Multigene Proliferation and Immune Profiling Scores**

RNA-based MGPS were calculated as described previously (14, 61) by averaging the gene median-centered log₂, RSEM data for all genes previously characterized as cycle-regulated (62) in each sample. Protein-based MGPS were generated for each sample by averaging log₂ TMT ratios for all proteins that showed significant correlation with the RNA-based MGPS (Pearson correlation, P < 0.01 after Benjamin-Hochberg FDR correction). Immune profile and microenvironment scores were inferred from the FPKM version of the RNA-seq data using ESTIMATE (63), Cibersort (ref. 64; run in absolute mode), and xCell (65). Protein-based immune modulator scores were calculated as described previously (14) by averaging log₂ TMT ratios for expert-curated sets of immune modulators belonging to three categories: immune stimulatory, immune inhibitory, and HLA (66).

**Immunoblotting**

Fresh-frozen WHIM68, WHIM74, and WHIM75 tumors were cryopulverized (Covaris CP02) and then lysed in RIPA buffer. Lysates were botted for LIG1 (cat. #18051-1-AP, ProteinTech, 1:1,000), POLD1 (cat. #15656-1-AP, ProteinTech, 1:1,000), or XRCC1 (cat. #ab134056,
Proteogenomic Markers of Chemotherapy Resistance in TNBC

Data Availability

The genomics and transcriptomics data have been deposited in the database of Genotypes and Phenotypes (dbGaP) under the accession code phs002505.v1, and the proteomics data are accessible through the NCI Proteomics Data Commons (https://pdc.cancer.gov/pdc/) with the accession identifiers PDC000408 (TNBC biopsies proteome raw files), PDC000409 (TNBC biopsies phosphoproteome raw files), and PDC000410 (TNBC PDX proteome raw files). MS raw files can also be accessed via MassIVE (https://massive.ucsd.edu/) with the accession identifier MSV000089758.

Authors’ Disclosures

M. Anurag reports grants from the NHI (U01CA214125 and P50 CA186784-06) during the conduct of the study, as well as a patent for proteogenomic markers of chemotherapy resistance and response in TNBC pending. E.J. Jaehning reports grants from the NCI and the Cancer Prevention & Research Institute of Texas during the conduct of the study, as well as a patent for proteogenomic markers of chemotherapy resistance and response in TNBC pending. K. Krug reports a patent for Provisional Application No. 63/317,402 pending to the Broad Institute. J.T. Lei reports grants from the NIH during the conduct of the study. D.M. Muzny reports grants from the NCI and the Cancer Prevention & Research Institute of Texas during the conduct of the study, as well as personal fees from StemMed, Ltd. outside the submitted work. M.T. Lewis reports grants from the NCI and the Cancer Prevention & Research Institute of Texas during the conduct of the study, as well as personal fees from StemMed, Ltd., Tvardi Therapeutics Inc., and StemMed Holdings outside the submitted work. M. Nemati Shafaee reports personal fees from Eli Lilly, Moderna, AstraZeneca, and Sanofi and grants from Pfizer outside the submitted work. S. Satpathy reports personal fees from Pfizer and personal fees from MacroGenics, Seagen, AstraZeneca, and Novartis outside the submitted work. F.O. Ademuyiwa reports personal fees from Pfizer and Gilead, grants from NeolmmuneTech, Cardinal Health, Biotheranostics, QED, AstraZeneca, and Athenex outside the submitted work. S. Satpathy reports a patent for proteogenomic markers of chemotherapy resistance and response in TNBC pending. M. Ellis reports grants from NIH NCI U01214125, NIH NCI P50186784, RR140033, NIH NCI P30 125123, RP210227, and U24CA210954, other support from Lisa and Ralph Eads, and grants and other support from the Korell family during the conduct of the study; personal fees from AstraZeneca outside the submitted work; and a patent for US/094,898 issued and with royalties paid from Varayce, a patent for CA2,630,974 issued and with royalties paid from Varayce, a patent for EU06844413.2 issued and with royalties paid from Varayce, and a patent for U.S. Provisional 63/317,402 pending. No disclosures were reported by the other authors.

Authors’ Contributions


Acknowledgments

The authors acknowledge the following funding support: U24CA160034 (S.A. Carr), U24CA210986 (S.A. Carr and M.A. Gillette), U01CA214125 (M.J. Ellis /M. Anurag, S.A. Carr), U24CA210979 (D.R. Mani), R03OD032626 (D.R. Mani), U24CA210954 (B. Zhang), P50 CA186784-06 (M.J. Ellis), U54CA224083 (S. Li), U54 CA224076 (M.T. Lewis), U24 CA226110 (M.T. Lewis), P30 NCI-CA125123 (C.K. Osborne), RR160027 [B. Zhang, Cancer Prevention & Research Institute of Texas (CPRIT) Scholar in Cancer Research], RR RR140027 (M.J. Ellis, CPRIT Scholar in Cancer Research), RR200009 (G.V. Echeverria, CPRIT Scholar in Cancer Research), RP170691 (CPRIT Core Facilities Support Grant), GH0005083 Breast Cancer Research (S. Li). The NCI-SPORE Career Enhancement Award to M. Anurag (part of P50 CA160034), T32CA203690 support to J.T. Lei, 1S10OD028671-01 to G. Miles, and K12 CA167540 to F.O. Ademuyiwa.
are also acknowledged. M.J. Ellis is a Susan G. Komen Foundation Scholar. M.J. Ellis and B. Zhang are McNair Scholars supported by the McNair Medical Institute at The Robert and Janice McNair Foundation. This work was also supported by generous gifts from the Krell family, Lisa and Ralph Eads, and the Washington University School of Medicine, Department of Oncology (Dr. John Dipsos). The authors are most grateful to the NCI CPTAC along with patients and caregivers participating in this study. The authors thank the patient advocates (GRASP Huddle) for their insightful comments and suggestions.

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and so to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

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

Received February 18, 2022; revised June 8, 2022; accepted August 18, 2022; published first August 24, 2022.

REFERENCES


Proteogenomic Markers of Chemotherapy Resistance in TNBC


43. Paul-Konietzko K, Thomale J, Arakawa H, Iliakis G. DNA ligases I


