(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 September 2023 (14.09.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/172913 A2

(51) International Patent Classification:
C12Q 1/6886 (2018.01) *A61K 41/00* (2020.01)

(21) International Application Number:
PCT/US2023/063863

(22) International Filing Date:
07 March 2023 (07.03.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/317,402 07 March 2022 (07.03.2022) US

(71) Applicants: **BAYLOR COLLEGE OF MEDICINE** [US/US]; One Baylor Plaza, Houston, Texas 77030 (US). **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US). **THE GENERAL HOSPITAL CORPORATION** [US/US]; 55 Fruit Street, Boston, Massachusetts 02114 (US).

(72) Inventors: **ANURAG, Meenakshi**; c/o Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030 (US). **ELLIS, Matthew J.**; c/o Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030 (US). **ZHANG, Bing**; c/o Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030 (US). **JAEHNIG, Eric**; c/o Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030 (US). **TRAN-HUYNH, Anh M.**; c/o Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030 (US). **SATPATHY, Shankha**; c/o THE BROAD INSTITUTE, INC., 415 Main Street, Cambridge, Massachusetts 02142 (US). **CARR, Steven**; c/o THE BROAD INSTITUTE, INC., 415 Main Street, Cambridge, Massachusetts 02142 (US). **KRUG, Karsten**; c/o THE BROAD INSTITUTE, INC., 415 Main Street, Cambridge, Massachusetts 02142 (US). **GILLETTE, Michael A.**; c/o THE GENERAL HOSPITAL CORPORATION, 55 Fruit Street, Boston, Massachusetts 02114 (US).

(74) Agent: **SISTRUNK, Melissa**; Norton Rose Fulbright US LLP, 1301 McKinney Suite 5100, Houston, Texas 77010 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH,

(54) Title: PROTEOGENOMIC MARKERS OF CHEMOTHERAPY RESISTANCE AND RESPONSE IN TRIPLE NEGATIVE BREAST CANCER

FIG. 1A

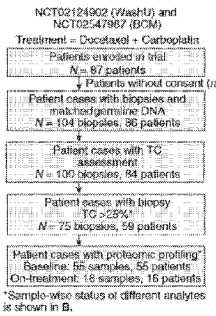
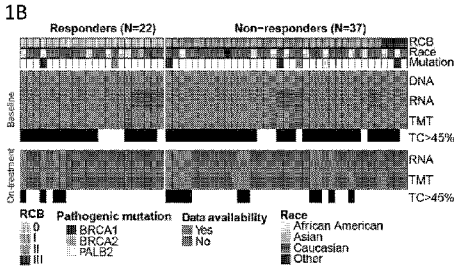


FIG. 1B



FIGs. 1A-1B

(57) Abstract: Microscaled proteogenomics was deployed to probe the molecular basis for differential response to neoadjuvant carboplatin & taxotere combination chemotherapy for triple-negative breast cancer (TNBC). Proteomic analyses of pre-treatment biopsies uniquely revealed that metabolic pathways including oxidative phosphorylation, fatty acid metabolism and glycolysis were resistance-associated. Both proteomics and transcriptomics revealed that sensitivity was marked by elevation of DNA repair, E2F targets, G2M checkpoint, interferon-gamma response, 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 involved in lagging strand synthesis) gene deletion and/or low mRNA expression were associated with lack of pathological complete response and poor prognosis in TNBC, as well as selective carboplatin-resistance in TNBC patient-derived xenograft models. Low expression or *LIG1* loss was also associated with higher chromosomal instability index (CIN) and poor prognosis in other cancer types, demonstrating that deletion of lagging-strand synthesis components has broad clinical significance.

WO 2023/172913 A2

[Continued on next page]

TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

PROTEOGENOMIC MARKERS OF CHEMOTHERAPY RESISTANCE AND RESPONSE IN TRIPLE NEGATIVE BREAST CANCER

[0001] This application claims priority to U.S. Provisional Patent Application Serial No. 63/317,402, filed March 7, 2022, which is incorporated by reference herein in its entirety.

BACKGROUND

[0002] This invention was made with government support under U01CA214125 awarded by the National Institutes of Health. The government has certain rights in the invention.

I. Technical Field

[0003] This disclosure relates at least to the fields of medicine, oncology, and proteogenomics.

II. Background

[0004] 10% to 15% of breast cancer is designated triple negative (TNBC) because of the lack of human epidermal growth factor receptor 2 (HER2), estrogen receptor (ER) and progesterone receptor (PR) expression. This disease subset exhibits high mortality and treatment resistance [1]. In a minority of cases, TNBC is linked to hereditary homologous recombination defects (HRD), most commonly in the BRCA1 gene. However, the majority of TNBC is sporadic, *i.e.* without an obvious hereditary association, and the driving DNA repair defects are more obscure [2]. Cytotoxic chemotherapy, the standard of care for many decades, is only partially effective, hence lack of pathological complete response (pCR) after neoadjuvant chemotherapy is frequent and is associated with poor survival [3]. Salvage therapy with adjuvant capecitabine has modest efficacy in this setting [4]. The programmed cell death receptor (PD1)-targeting antibody pembrolizumab is also approved for neoadjuvant/adjuvant TNBC treatment [5], but outcome improvements are unfortunately not accurately predicted by PD-L1-based immunohistochemistry (IHC) [6]. Thus, biomarker research in sporadic TNBC remains a critical endeavor. Embodiments herein describe studies deploying microscaled proteogenomics [7] to discover new 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 the

combination of carboplatin and docetaxel (NCT02547987 and NCT02124902) [8]. Tumor analyses included germline-matched tumor whole exome DNA sequencing (WES), RNA-seq, and tandem mass tag (TMT)-based proteomics and phosphoproteomics. Study endpoints were pCR and residual cancer burden (RCB). Multiple orthogonal data sets were used to validate the findings, including mRNA profiling in other TNBC clinical trials, immunohistochemistry (IHC) analyses, preclinical studies in patient-derived xenografts (PDX) and pan-cancer analyses using The Cancer Genome Atlas (TCGA) data.

SUMMARY

[0005] Embodiments herein concern methods of detecting genetic alterations and/or deletions. In some embodiments, methods concern methods of detecting deletion and/or amplification of one or more genes on cytoband 19q13.31-33 or 8q21.3 respectively in an individual. The genes may comprise DNA ligase 1 (LIG1), POLD1 and XRCC1 on cytoband 19q13.31-33 and RIPK2, RMDN1, CPNE3, DECR1, and OTUB6B on cytoband 8q21.3, in specific embodiments. In some embodiments, there are methods of treating an individual with a therapy other than a standard of care platinum-based drug when the individual has gene deletion and/or low mRNA expression and/or low protein level of LIG1. In specific cases, when a sample from the individual identifies gene deletion and/or low mRNA expression and/or low protein level of LIG1, the individual is administered a therapeutically effective amount of a therapy that is not a platinum-based drug. Particular embodiments provide a method of determining a treatment therapy for an individual with cancer, comprising the step of administering a therapy other than a platinum-based drug to an individual having gene deletion and/or low mRNA expression and/or low protein level of DNA ligase 1 (LIG1).

[0006] In some embodiments, the method comprises the step of combining into multiple vessels a plurality of sections from different regions of the biological sample, said biological sample optionally being a biopsy. The combining step may provide a mixture of cells and/or tissues from different regions of the biopsy for the plurality of sections. In some embodiments, the method comprises the step of isolating DNA from the plurality of sections in a first vessel, isolating RNA from the plurality of sections in a second vessel, and isolating protein from the plurality of sections in a third vessel. In some embodiments, the method comprises the step of analyzing the isolated DNA, analyzing the isolated RNA, and/or analyzing the isolated protein to determine the status of

at least one marker present in cells in the biopsy, wherein the marker comprises LIG1. The protein may be denatured, in certain embodiments.

[0007] In some embodiments, the method comprises the step of sectioning the biopsy. In some embodiments the sectioning and combining step are performed concurrently or separately. In some embodiments, a plurality of sections from a fourth vessel are used for sample quality analysis. In some embodiments, the plurality of sections for at least one of the 1st, 2nd, and/or 3rd vessels comprises nonadjacent sections from the different regions. In some embodiments, analyzing the isolated protein further comprises determining the phosphorylation status of LIG1. The phosphorylation status of any cancer marker may be determined by mass spectrometry, in specific embodiments.

[0008] In some embodiments, analyzing the isolated DNA, analyzing the isolated RNA, and analyzing the isolated protein determines the status of LIG1, POLD1 and/or XRCC1 in the biopsy. In some embodiments, when a status determination that LIG1, POLD1 and/or XRCC1 is deleted in any one, two, or all three of the vessels (in some embodiments, each containing DNA, RNA, or protein), the method further comprises the step of administering to the individual one or more therapies, wherein the therapies do not comprise a platinum-based drug. However, in some embodiments DNA and protein are analyzed from the same vessel. In some embodiments, deletion occurs in the gene (DNA) but may also manifest as lower expression in RNA and/or protein. The platinum-based drug may comprise carboplatin. In some embodiments, the individual has, or is suspected of having, a cancer.

[0009] Certain embodiments concern methods of determining the susceptibility of an individual having, or suspected of having, a cancer to a cancer treatment. In some embodiments, a sample from an individual is subjected to any method encompassed herein. In some embodiments, an individual is administered or not administered carboplatin. In some embodiments, an individual has or is suspected of having breast cancer, such as triple negative breast cancer. The individual may or may not have a family history of triple negative breast cancer.

[0010] Certain embodiments concern a method of treating an individual having, or suspected of having, a cancer using a biological sample from the individual comprising the steps of: combining into multiple vessels a plurality of sections from different regions of the biological sample, said biological sample optionally being a biopsy; isolating DNA from the plurality of sections in a first vessel, isolating RNA from the plurality of sections in a second vessel, and

isolating protein from the plurality of sections in a third vessel; analyzing the isolated DNA, analyzing the isolated RNA, and/or analyzing the isolated protein to determine the status of at least one marker present in cells in the biopsy, wherein the marker comprises LIG1; determining the LIG1 status in any one, two, or all three vessels having sections from different regions of the biological sample; and administering or not administering a particular therapeutic. In some embodiments, the sections are staggered so that the different vessels are as similar as possible even though they are collected from different sections; one embodiment is to minimize regional differences between the different analytes (DNA, protein, RNA) collected in each vessel by alternating the sections added to each.

[0011] In some embodiments, based on the LIG1 status an individual is not administered a platinum-based drug, inhibitor of EGFR (including tyrosine kinase inhibitors such as erlotinib and/or gefitinib and including monoclonal antibodies such as cetuximab and/or necitumumab) or PI3K inhibitor (such as copanlisib, duvelisib and/or idelalisib) when LIG1 is reduced or deleted in one or more of DNA, RNA, or protein. In some embodiments, an individual is administered one or more CDK2 inhibitors (such as Flavopiridol and/or TG02); one or more Chk2 inhibitors (4,4′-diacetyldiphenylurea-bis(guanylhydrazone) (NSC 109555); PV1019 (NSC 744039) [7-nitro-1H-indole-2-carboxylic acid {4-[1-(guanidinohydrazone)-ethyl]-phenyl}-amide]; and/or PHI-101); and/or one or more PARP inhibitors (such as olaparib, rucaparib, and/or niraparib) when LIG1 is reduced or deleted in one or more of DNA, RNA, or protein. In some embodiments, an individual is administered a platinum-based drug when the LIG1 status is determined to not be deleted in the sample. In some embodiments, an individual is not administered a platinum-based drug when the LIG1 status is determined to be deleted in the sample.

[0012] In some embodiments, there are methods of comparing or identifying tumors resistant or sensitive to one or more particular therapeutics based on the status of LIG1. When LIG1 RNA, protein, and/or DNA are depleted or reduced in level from a sample, the individual should not be administered a platinum-based drug, EGFR inhibitor, or PI3K inhibitor, and/or the individual should be administered a CDK2 inhibitor, Chk2 inhibitor, and/or PARP inhibitor.

[0013] Certain embodiments concern a method of determining an individual's susceptibility to a platinum-based drug comprising the steps of analyzing genomic DNA from the sample at a locus comprising LIG1 and/or measuring expression of LIG1 mRNA and/or protein in the sample.

[0014] Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating specific embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present disclosure. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0016] **FIGS. 1A-1F** show an overview of TNBC patient samples. (**FIG. 1A**) REMARK diagram showing pre- and on-treatment sample accrual schema from triple negative breast cancer patients enrolled in two clinical trials (NCT02547987 (BCM), NCT02124902 (WashU)) and treated with carboplatin and docetaxel in the neoadjuvant setting. * samples with <45% tumor content were later excluded from analysis based on evidence from data QC. (**FIG. 1B**) Overview of available omics datasets from 59 patients (22 tumors with pCR and 37 tumors without pCR). Pathogenic BRCA1/2 and PALB2 mutation status, residual cancer burden (RCB) and patient race are indicated via color-coded annotation tracks. (**FIG. 1C**) Venn-diagram showing overlap of gene IDs detected across multiple analytes and omics data profiled. (**FIG. 1D**) Hallmark metabolism pathways (PMID:26771021) are acutely induced by chemotherapy exclusively at the protein level. Scatter plot shows signed -log10 fdr from GSEA (PMID: 16199517) using the signed (by direction of change) -log10 p-values from paired Wilcoxon signed rank tests comparing RNA (x-axis) and protein levels (y-axis) for on-treatment (cycle 1 day 3) samples to matching baseline samples (n = 14). (**FIG. 1E**) MSigDB Hallmark metabolism pathways (PMID:26771021) are elevated in baseline non-pCR tumors at the protein level while immune and cell cycle pathways are elevated in baseline pCR tumors at both RNA and protein levels. Scatter plot shows the signed -log10 fdr values from Gene Set Enrichment Analysis (GSEA; PMID: 16199517) using ranked lists of signed (by direction of change) -log10 p-values from Wilcoxon rank sum tests comparing RNA (x-axis) and protein (y- axis) levels in non-pCR tumors to pCR tumors. (**FIG. 1F**) Cell cycle kinase targets

and PTM-SigDB phosphosites associated with genotoxic stress are enriched in pCR tumors relative to non-pCR tumors at baseline. Volcano plot shows results from Post-Translational Modification-Set Enrichment Analysis (PTM-SEA) using the signed -log10 p- values from Wilcoxon rank sum tests comparing phosphosite levels in non-pCR tumors to pCR tumors (PMID: 30563849).

[0017] **FIGS. 2A-2F** show proteogenomic features associated with pCR in TNBC tumors. (**FIG. 2A**) Proteogenomic features associated with the immune microenvironment are elevated in pCR tumors relative to non-pCR tumors. Heatmap shows protein-based Hallmark single sample GSEA (ssGSEA) scores (PMID:26771021; PMID: 19847166), protein-based immune modulator scores (PMID: 31433971; PMID: 33212010), RNA-based immune profiles (PMID: 24113773; PMID: 2582280; PMID: 29141660), and proteogenomic features for immune checkpoint genes that are targets of FDA-approved inhibitors. Within each group (pCR and non-pCR), samples are ordered by increasing immune stimulatory score. Asterisks indicate $p<0.05$ by Wilcoxon rank sum test comparing non-pCR to pCR tumors. (**FIG. 2B**) The immune stimulatory score is significantly higher in pCR tumors than non-pCR tumors ($p=0.01$, Wilcoxon rank sum test). Boxplots show interquartile range (IQR) with median marked in center. Whiskers indicate 1.5x IQR. (**FIG. 2C**) The immune stimulatory score is negatively correlated to chromosomal instability index (CIN; Spearman Rho = -0.612, p = 6.2e-6). Scatter plot shows immune stimulatory score on the x-axis and CIN on the y-axis. (**FIGs. 2D-2E**) Scatter plots showing correlation between PDL1 IHC levels with PDL1 protein (**FIG. 2D**) and phosphoprotein levels (**FIG. 2E**). (**FIG. 2F**) While cell cycle features are elevated in pCR tumors, a subset of non-pCR tumors have elevated levels of CDK4 target sites and Rb levels. Heatmap shows RNA- and protein-based multi-gene proliferation scores (MGPS; PMID: 28045625; PMID: 12058064), single sample PTM-SEA scores for CDK4 and CDK2 target sites (PMID: 30563849), ssGSEA scores for E2F targets (PMID:26771021; PMID: 19847166), and other proteogenomic features for genes regulating the G1/S transition of the cell cycle (see pathway on right). Box outlines samples from non-pCR tumors with high CDK4 activity and Rb phosphorylation. Asterisks indicate $p<0.05$ by Wilcoxon rank sum test comparing non-pCR to pCR tumors.

[0018] **FIGS. 3A-3C** show proteogenomic features associated with lack of pCR and that are altered upon treatment in TNBC tumors. (**FIG. 3A**) Heatmap showing single-sample GSEA NES for metabolic Hallmark pathways that are significantly higher in non-pCR cases, arranged by

RCB0 (pCR) and RCBI/II/III (non-pCR). Shown are enrichment scores for 4 pathways as assessed at the level of mRNA (yellow), protein (blue) and phospho-protein (red). Wilcoxon rank sum test was used to compare scores for non-pCR Vs pCR scores, * $p<0.05$. (**FIG. 3B**) 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 at protein and mRNA levels for each gene along with mRNA-protein correlation scores are shown as signed -log10 p-value (signedp). (**FIG. 3C**) A multi-omics metabolic gene signature derived for genes that are correlated at mRNA and protein level was further investigated in patients treated with carboplatin and paclitaxel in the BrighTNess clinical trial. The mean mRNA expression score for this signature was significantly higher in higher RCB tumors.

[0019] **FIGS. 4A-4E** show discovery of DNA repair and replication components enriched in non-pCR TNBC tumors. (**FIG. 4A**) Workflow depicts the strategy that was used to identify enriched or depleted chromosomal cytobands using mRNA and protein lists ranked by signed -log10 p-value derived from differential expression for non-pCR vs. pCR samples. (**FIG. 4B**) Plot showing significantly enriched or depleted cytobands obtained by running differential mRNA and protein ranked lists through GSEA. (**FIG. 4C**) Venn-diagram showing differential (non-pCR vs. pCR) mRNA and proteins located on cytoband 19q13.3. (**FIG. 4D**) Over-representation analysis (ORA) shows that differential 19q13.31-33 genes are enriched with Hallmark DNA repair pathway genes. Downregulation of these DNR repair genes at mRNA and protein level in non-pCR cases is shown in the bar chart on the right as signed -log10 p-values from Wilcoxon rank sum tests. (**FIG. 4E**) Boxplot showing RNA expression of DNA-repair genes located on 19q13.31-33 in the previously published BrighTNess clinical trial (PMID: 29501363), where patients were treated with carboplatin and paclitaxel. Wilcoxon rank sum test was used to compare residual disease (RD) cases to pCR cases. (**FIG. 4F**) Forest plot showing hazard ratios (HR) and p-values for metastasis free survival associated with LIG1, POLD1, XRCC1 and ERCC2. HR is based on categorizing samples using median expression cutoff for each gene in the Hatzis dataset (PMID: 21558518).

[0020] **FIG. 5** shows proteogenomic features associated with LIG1. Heatmap showing copy number, mRNA and protein levels of LIG1, which are significantly (Wilcoxon rank sum test) lower (blue) in non-pCR tumors. Corresponding boxplots show that tumors with low level copy loss of LIG1 (GISTIC = -1, likely single copy number loss) display significantly higher

chromosomal instability and multi-gene proliferation scores and significantly lower Signature 3 (COSMIC mutational signature associated with homologous recombination defect) than tumors that that are WT for or show gain of CNV (GISTIC >= 0). T-test and wilcoxon rank sum test were used to compare LIG1 loss cases to LIG1 intact (WT/Gain) cases.

[0021] **FIGS. 6A-6D** show LIG1 association with advanced TNBC disease in preclinical models. (**FIG. 6A**) Proteogenomic status of LIG1, POLD1 and XRCC1 in three PDX models derived from longitudinal biopsies from the same TNBC patient 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 one year after treatment initiation (WHIM75). Mutation and copy number derived from WES and RNA from RNAseq (WES and RNAseq as previously described in https://doi.org/10.1101/2021.08.20.457116), and protein from TMT proteomics generated by this current study. Bottom panels show representative western blots from 3 biological repeats for LIG1, POLD1 and XRCC1 protein levels. GAPDH is used as loading control. (**FIG. 6B**) 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 highest LIG1 protein levels, was most sensitive to carboplatin while 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 vs Day 0 for each treatment arm. (**FIG. 6C**) Boxplots showing LIG1 mRNA levels in TNBCs PDXs categorized into complete response (CR) and non-CR groups. After 4 of carboplatin weeks treatment, CR was defined as PDXs with non-palpable 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. (**FIG. 6D**) Association between LIG1 copy number loss and treatment response in patient-derived xenograft organoids (PDXOs) obtained from the BCaPE database (PMID 27641504). Carboplatin and docetaxel are highlighted in red.

[0022] **FIGS. 7A-7E** show pan-cancer analysis of LIG1 loss. (**FIG. 7A**) Kaplan-Meier curve showing significantly reduced (Log-rank P value) progression free survival (PFS) for tumors with single copy loss of LIG1 (HETLOSS, GISTIC <=-1 indicated in orange) in the TCGA pan-cancer cohort. (**FIG. 7B**) Boxplot 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). (**FIG. 7C**) Violin plot showing significantly lower (Wilcoxon rank sum test) COSMIC Signature 3 scores in LIG1 loss tumors (shown in teal). (**FIG. 7D**) Forest plot showing impact of LIG1 copy number loss on PFS by cancer type along with LIG1 WT/gain/loss frequency, HR and corresponding p-value. (**FIG. 7E**) Boxplot showing significantly higher (Wilcoxon rank sum test) FGA (representing chromosomal instability) in tumors that had LIG1 copy number loss versus tumors with either wild-type LIG1 or with LIG1 copy number gain. Shown are the only 5 cancers (HNSCC, UCEC, COAD, PRAD and KIRP) that displayed significant association between LIG1 loss and adverse PFS.

[0023] **FIGS. 8A-8G** show cohort characteristics. (**FIG. 8A**) Proportion plot showing distribution of samples obtained from baseline tumors grouped by PAM50 (Basal, HER2, LumA, LumB), TNBC subtyping (BL1-basal like 1, BL2-basal like 2, IM- immunomodulatory, LAR-luminal androgen receptor, M-mesenchymal, MSL-mesenchymal stem like, UNS-unspecified tumors) and Race (AA-african american, C-caucasian, Others). The left panel shows all samples, and the right panel shows samples segregated into two groups based on pCR. P-values were obtained using Fisher's exact test. (**FIG. 8B**) Correlation between average tumor content as assessed by immunohistochemistry (IHC) and mRNA- protein correlation for all baseline samples for which there were paired mRNA and protein data. Samples indicated by teal points were excluded from this study due to poor tumor content and poor mRNA-protein correlation (tumor content <45%). (**FIG. 8C**) Pairwise correlation analysis depicting correlations across (green) and within (red) TMT-plexes at the protein (left) and phosphopeptide level (right). A common reference sample from the prospective BRCA study (PMID: 33212010) was measured in each TMT plex as part of this study and showed consistently high correlations across all 8 TMT plexes. Pairwise within-plex correlations of tumor samples are shown in red. (**FIG. 8D**) Gene-wise mRNA-protein correlations ordered from most negative to most positive by Spearman rank correlation coefficient are shown in the top panel. Signed log10 p-values were used as input for GSEA using the KEGG pathway database. Top pathways enriched for genes that had positive correlation between mRNA and protein are shown in the middle panel while top pathways enriched for genes with lower correlation are shown in the bottom panel. (**FIG. 8E**) Gene function prediction performance quantified by AUROC shows co-expression networks based on proteomic profiles outperformed RNA-based networks for predicting KEGG pathway membership. Dotted lines indicate 10% increase or decrease in prediction performance. (**FIG. 8F**) Genotoxic stress

sites (sites induced by nocodazole, UV, and ionizing radiation) and ATM and CDK1/2 target sites are acutely induced by chemotherapy. Volcano plot shows results from Post-Translational Modification-Set Enrichment Analysis (PTM-SEA; PMID: 30563849) using the signed (by direction of change) -log10 p-values from paired Wilcoxon signed tests comparing phosphosite levels in on-treatment (cycle 1 day 3) tumors to matching baseline tumors (n = 13). (**FIG. 8G**) Heatmap showing distribution of COSMIC mutational signatures in the baseline (pre-treatment) samples. Signatures 6 and 10, which are associated with MMR and POLE mutations respectively, were higher in RCB II/III categories (ANOVA test, $p<0.05$).

[0024] **FIGS. 9A-9D** show PDL1 IHC, distribution of Rb phosphorylation, and association of Rb protein and drug sensitivity in cell lines. (**FIGs. 9A-9B**) Representative images for patient samples with high (A) and low (B) PDL1 IHC staining. (**FIG. 9C**) The distribution of Rb phosphorylation levels in non-pCR tumors (brown) is overlapping with but also shifted towards higher levels than in pCR tumors (green). Density plots for each group are shown. (**FIG. 9D**) Plot showing Pearson correlations between Rb protein by TMT profiling in TNBC cell lines and responses from all approved drugs in CTRP (Cancer Therapeutics Response Portal), GDSC (Genomics of Drug Sensitivity in Cancer), and PRISM (Profiling Relative Inhibition Simultaneously in Mixtures) databases from the DepMap (Dependency Map) resource. Note that, of all Rb protein-drug response correlations, platinum compounds such as cisplatin and carboplatin, as well as an alkylating agent that induces DNA damage, temozolomide, were among the most positively correlated pairs (higher Rb protein in TNBC cell lines with increasing resistance to those drugs) while response to palbociclib was negatively correlated (higher Rb protein in TNBC cell lines with decreasing resistance to palbociclib).

[0025] **FIG. 10** shows metabolic multi-omic signature and results from PTM-SEA. The multi-omics metabolic gene signature identified in this study was further investigated in patients treated with carboplatin and paclitaxel in the BrighTNess clinical trial. The mean mRNA expression score for this signature was significantly higher in residual disease. P-value was calculated using the Wilcoxon rank sum test.

[0026] **FIGS. 11A-11D** show proteogenomic prioritization of candidates driving chemo-resistance in TNBC. (**FIG. 11A**) Chromosome-wise differences in the SCNA landscape in non-pCR and pCR samples are shown as the top two tracks. Amplification and deletion events are indicated by red and blue respectively. The bottom panels show differential mRNAs and proteins

up- and down-regulated in non-pCR compared to pCR cases arranged based on their chromosomal location to align with the SCNA tracks. (**FIG. 11B**) Venn-diagram showing differential (non-pCR vs. pCR) mRNA and proteins located on cytoband 8q21.3. (**FIG. 11C**) Circos (proteogenomic circos) plots for chromosome 19 showing signed -log10 p-values for mRNA and protein based on Wilcoxon rank-sum tests comparing non-pCR to pCR cases and frequency of amplifications and deletions (CNV) in non-pCR and pCR. Genes annotated in blue on the outermost track represent those to be low at the CNV, mRNA, and protein level from (C). (**FIG. 11D**) Kaplan-Meier (KM) curve depicting metastasis free survival of patients with TNBC breast cancer from the Hatzis dataset. Tumors are categorized into high and low LIG1 based on mean RNA expression cutoff.

[0027] **FIGS. 12A-12E** show features associated with *LIG1* genomic status. (**12A-12B**) Plot showing results for Spearman rank correlation between LIG1 copy number and LIG1 mRNA expression (**FIG. 12A**) and LIG1 protein level (**FIG. 12B**). The samples are colored according to treatment response. (**FIG. 12C**) Boxplot showing immune stimulatory scores across LIG1 copy number and chemotherapy response categories. P-values are calculated using the Wilcoxon rank-sum test. (**FIG. 12D**) LIG1 loss tumors have higher CDK1/2 activity than non-loss tumors. Volcano plot shows results from Post-Translational Modification-Set Enrichment Analysis (PTM-SEA; PMID: 30563849) using the signed - log10 p-values from Wilcoxon rank sum tests comparing phosphosite levels in LIG1 loss (GISTIC = -1) tumors to tumors with GISTIC = 0-2 as input. (**FIG. 12E**) Heatmap showing ssPTMSEA scores for significantly differential kinases and pathways shown in D. LIG1 single copy loss and RCB class are shown as additional rows.

[0028] **FIGS. 13A-13D** show LIG1 association with advanced TNBC disease in preclinical models and independent cohorts. (**FIG. 13A**) Proportion plot showing distribution of LIG1 copy number alterations in TCGA-Breast (primary disease) and INSERM (metastatic disease) cohorts. P-value was obtained using Fisher's exact test. (**FIG. 13B**) Quantification by densitometry using the same western blots for LIG1, POLD1 and XRCC1, along with GAPDH as a control, in WHIM 68, 74 and 75 as shown in FIG. 6A. Normalized LIG1, POLD1, and XRCC1 protein abundance from each PDX was calculated by dividing levels of each protein by their corresponding GAPDH signals. These values were then normalized to WHIM68 (primary disease). (**FIG. 13C**) Tumor volumes in PDX models treated with 4 weekly cycles of vehicle or 20 mg/kg docetaxel. 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 vs Day 0 for each treatment arm. (**FIG. 13D**) Boxplots

showing LIG1 mRNA levels in TNBCs PDXs categorized into complete response (CR) and non-CR groups. After 4 weeks of docetaxel treatment, CR was defined as PDXs with non-palpable tumors and non-CR defined for PDXs with residual tumors with measurable dimensions. Wilcoxon rank sum test was used to compare the two groups.

[0029] **FIGS. 14A-14B** show *LIG1* loss across TCGA Pan-Cancer cohort. (**FIG. 14A**) LIG1 single copy loss (shallow deletion, GISTIC <= -1) frequency in Testicular Germ Cell Tumors (TGCT) subtypes. The TCGA-TGCT dataset was used for this analysis. (**FIG. 14B**) Chromosomal Instability (CIN) scores in LIG1 WT/gain and loss in 3 CPTAC cancer cohorts. HSCC: Head and Neck Squamous Cell Carcinoma (PMID: 33417831) and UCEC: Endometrial Carcinoma (PMID: 32059776), and COAD: Colon Cancer PMID: 31031003). Tumors with LIG1 loss had significantly (Wilcoxon test) higher levels of CIN scores than tumors without LIG1 loss. Boxplots show the interquartile range (IQR) for the Chromosomal Instability index (CIN) for each group with median marked in center. Whiskers indicate 1.5x IQR. P-values are from Wilcoxon rank sum tests comparing LIG1 loss tumors (GISTIC = -1) to tumors with GISTIC = 0-2.

[0030] **FIG. 15** demonstrates that *LIG1* loss tumors have higher CDK2 activity than non-loss tumors. Volcano plot shows results from Post-Translational Modification-Set Enrichment Analysis using the signed -log10 p-values from Wilcoxon rank sum tests comparing phosphosite levels in *LIG1* loss with *LIG1* WT or gain cases.

[0031] **FIGS. 16A-16B** (**FIG. 16A**) Reanalyzed data from three independent CRISPR (clustered regularly interspersed palindromic repeats) screens to identify genes and pathways that mediate cellular resistance to Olaparib, a clinically approved PARP inhibitor (Zimmerman et al, 2018). The Venn-diagram shows that *LIG1* is one of 13 genes that showed increased sensitivity to PARP inhibitors across all the three models when suppressed. (**FIG. 16B**) Olaparib treatment of the TNBC cell line MDA-MB-468 with (siLIG1) and without (siNC) *LIG1* suppression using small interfering RNA (N=3).

[0032] **FIGS. 17A-17C.** (**FIG. 17A**) boxplot showing *LIG1* estimates (by FISH) by manual counting of 2000 cells from the three PDX models from 2 slides. Evidence suggests higher *LIG1*-loss heterogeneity in primary (WHIM68,74) as compared to the metastatic site (WHIM75). (**FIG. 17B**) Dot plots showing a higher % of LIG1-haploid cells in non-CR than in CR patients (**FIG. 17B**) with BRCA1/2 mutated cases (**FIG. 17C**) without BRCA1/2 mutated cases. 200 cells were manually counted for each of the 50 TNBC PDXs.

DETAILED DESCRIPTION

I. Definitions

[0033] Throughout this application, the term "about" is used to indicate that a value includes the inherent variation of error for the measurement or quantitation method.

[0034] The use of the word "a" or "an" when used in conjunction with the term "comprising" may mean "one," but it is also consistent with the meaning of "one or more," "at least one," and "one or more than one."

[0035] The phrase "and/or" means "and" or "or". To illustrate, A, B, and/or C includes: A alone, B alone, C alone, a combination of A and B, a combination of A and C, a combination of B and C, or a combination of A, B, and C. In other words, "and/or" operates as an inclusive or.

[0036] The words "comprising" (and any form of comprising, such as "comprise" and "comprises"), "having" (and any form of having, such as "have" and "has"), "including" (and any form of including, such as "includes" and "include") or "containing" (and any form of containing, such as "contains" and "contain") are inclusive or open-ended and do not exclude additional, unrecited elements or method steps.

[0037] The term "cancer," as used herein, may be used to describe a solid tumor, metastatic cancer, or non-metastatic cancer. In certain embodiments, the cancer may originate in the bladder, blood, bone, bone marrow, brain, breast, colon, esophagus, duodenum, small intestine, large intestine, colon, rectum, anus, gum, head, kidney, liver, lung, nasopharynx, neck, ovary, pancreas, prostate, skin, stomach, testis, tongue, or uterus. In some embodiments, the cancer is recurrent cancer. In some embodiments, the cancer is Stage I cancer. In some embodiments, the cancer is Stage II cancer. In some embodiments, the cancer is Stage III cancer. In some embodiments, the cancer is Stage IV cancer.

[0038] The compositions and methods for their use can "comprise," "consist essentially of," or "consist of" any of the ingredients or steps disclosed throughout the specification. Compositions and methods "consisting essentially of" any of the ingredients or steps disclosed limits the scope of the claim to the specified materials or steps which do not materially affect the basic and novel characteristic of the claimed invention.

[0039] As used herein, a "protein" or "polypeptide" refers to a molecule comprising at least five amino acid residues.

[0040] As used herein, the term "wild-type" refers to the endogenous version of a molecule that occurs naturally in an organism.

[0041] It is contemplated that any embodiment discussed in this specification can be implemented with respect to any method or composition of the invention, and vice versa. Furthermore, compositions of the invention can be used to achieve methods of the invention.

II. Genetic Alterations

[0042] Certain embodiments herein concern detecting genetic alterations. The genetic alteration may be in one or more cells in an individual. The genetic alteration may directly or indirectly cause or contribute to oncogenicity of the cells. The genetic alteration may directly or indirectly cause or contribute to drug resistance in the cells. In some embodiments, the genetic alteration directly or indirectly causes or contributes to resistance to a platinum-based drug. The platinum-based drug may comprise carboplatin, cisplatin, oxaliplatin, nedaplatin, heptaplatin, and/or lobaplatin, as examples. The drug may be a platinum pro-drug and/ or metabolites such as satraplatin and/or JM518, JM559, JM383, JM118. In some embodiments, the genetic alteration comprises one or more alterations to genomic DNA in the cell, one or more alterations to expression of at least one gene or regulatory molecule, and/or one or more alterations to a biomolecule in the cell, including a protein and/or nucleic acid. In some embodiments, the genetic alteration comprises alteration of a genomic locus. The genomic locus may comprise the 8q21.3 and/or 19q13.31-33 cytobands. The genomic locus may comprise the *LIG1, PPP5C, BCL3,* and/or *NOSIP* genes.

[0043] Certain embodiments concern detecting one or more genomic deletions. The deletion may be a deletion in all or part of a genomic locus that comprises the 8q21.3 and/or 19q13.31-33 cytobands. The 19q13.31-33 genomic locus may comprise the *LIG1, POLD1, XRCC1, PPP5C, BCL3,* and/or *NOSIP* genes. The 8q21.3 genomic locus may comprise the *RIPK2, RMDN1, CPNE3, DECR1*, and/or *OTUB6B* genes. The deletion (19q13.3) or amplification (8q21.2) may cause or contribute to drug resistance in the cells. In some embodiments, the deletion directly or indirectly causes or contributes to resistance to a platinum-based drug. The platinum-based drug may comprise carboplatin, cisplatin, oxaliplatin, nedaplatin, heptaplatin, and/or lobaplatin.

[0044] In some embodiments, a genetic alteration and/or deletion to LIG1 in a cell and/or tumor results in drug resistance in said cell and/or tumor. In specific embodiments, Lig1 loss is a

marker of lack of response to certain therapies, such as platinum-based drugs. The drug resistance may comprise a complete or partial lack of response to the drug. The genetic alteration and/or deletion may cause the cell and/or tumor not to respond to the drug in a manner that a cell and/or tumor not having the genetic alteration and/or deletion may have. For example, a cytotoxic drug may have less of an effect on a cell and/or tumor having a LIG1 alteration and/or deletion compared to a cell and/or tumor not having the alteration and/or deletion. The drug may comprise a platinum-based drug.

[0045] Embodiments herein refer to LIG1, *LIG1*, DNA Ligase 1, which may comprise the gene, its expressed mRNA, and/or protein. One skilled in the art would understand that references to one may comprise any of the biomolecules. For example, a reference to alteration of LIG1 may refer to an alteration of the gene in genomic DNA and/or an alteration to the expressed mRNA and/or alteration to the expression of protein.

[0046] Detection of the genetic alteration and/or deletion may be achieved through any method known in the art, including any method disclosed herein. In some embodiments, detection occurs through one or a combination of the proteogenomic methods disclosed herein. The proteogemoic methods are discussed in PCT Application No. PCT/US2020/045962, which is incorporated by reference herein in its entirety.

[0047] Compositions, systems, kits, and methods described herein may be useful for detecting one or more genetic alterations. One skilled in the art would understand that the methods disclosed herein may be used for determining genetic alterations that cause or contribute to oncogenicity and/or drug resistance or the like.

[0048] Certain embodiments concern assays useful for diagnosing and/or treating cancer. The cancer may specifically be of the following histological type, though it is not limited to these: neoplasm, malignant; carcinoma; carcinoma, undifferentiated; giant and spindle cell carcinoma; small cell carcinoma; papillary carcinoma; squamous cell carcinoma; lymphoepithelial carcinoma; basal cell carcinoma; basal-like 1 triple-negative breast cancer; basal-like 2 triple-negative breast cancer; pilomatrix carcinoma; transitional cell carcinoma; papillary transitional cell carcinoma; adenocarcinoma; gastrinoma, malignant; cholangiocarcinoma; hepatocellular carcinoma; combined hepatocellular carcinoma and cholangiocarcinoma; trabecular adenocarcinoma; adenoid cystic carcinoma; adenocarcinoma in adenomatous polyp; adenocarcinoma, familial polyposis coli; solid carcinoma; carcinoid tumor, malignant; branchiolo-alveolar adenocarcinoma; papillary

adenocarcinoma; chromophobe carcinoma; acidophil carcinoma; oxyphilic adenocarcinoma; basophil carcinoma; clear cell adenocarcinoma; granular cell carcinoma; follicular adenocarcinoma; papillary and follicular adenocarcinoma; nonencapsulating sclerosing carcinoma; adrenal cortical carcinoma; endometroid carcinoma; skin appendage carcinoma; apocrine adenocarcinoma; sebaceous adenocarcinoma; ceruminous adenocarcinoma; mucoepidermoid carcinoma; cystadenocarcinoma; papillary cystadenocarcinoma; papillary serous cystadenocarcinoma; mucinous cystadenocarcinoma; mucinous adenocarcinoma; signet ring cell carcinoma; infiltrating duct carcinoma; medullary carcinoma; lobular carcinoma; inflammatory carcinoma; paget's disease, mammary; acinar cell carcinoma; adenosquamous carcinoma; adenocarcinoma w/squamous metaplasia; mesenchymal-like; mesenchymal stem-like; immunomodulatory; luminal androgen receptor; thymoma, malignant; ovarian stromal tumor, malignant; thecoma, malignant; granulosa cell tumor, malignant; androblastoma, malignant; sertoli cell carcinoma; leydig cell tumor, malignant; lipid cell tumor, malignant; paraganglioma, malignant; extra-mammary paraganglioma, malignant; pheochromocytoma; glomangiosarcoma; malignant melanoma; amelanotic melanoma; superficial spreading melanoma; malignant melanoma in giant pigmented nevus; epithelioid cell melanoma; blue nevus, malignant; sarcoma; fibrosarcoma; fibrous histiocytoma, malignant; myxosarcoma; liposarcoma; leiomyosarcoma; rhabdomyosarcoma; embryonal rhabdomyosarcoma; alveolar rhabdomyosarcoma; stromal sarcoma; mixed tumor, malignant; mullerian mixed tumor; nephroblastoma; hepatoblastoma; carcinosarcoma; mesenchymoma, malignant; brenner tumor, malignant; phyllodes tumor, malignant; synovial sarcoma; mesothelioma, malignant; dysgerminoma; embryonal carcinoma; teratoma, malignant; struma ovarii, malignant; choriocarcinoma; mesonephroma, malignant; hemangiosarcoma; hemangioendothelioma, malignant; kaposi's sarcoma; hemangiopericytoma, malignant; lymphangiosarcoma; osteosarcoma; juxtacortical osteosarcoma; chondrosarcoma; chondroblastoma, malignant; mesenchymal chondrosarcoma; giant cell tumor of bone; ewing's sarcoma; odontogenic tumor, malignant; ameloblastic odontosarcoma; ameloblastoma, malignant; ameloblastic fibrosarcoma; pinealoma, malignant; chordoma; glioma, malignant; ependymoma; astrocytoma; protoplasmic astrocytoma; fibrillary astrocytoma; astroblastoma; glioblastoma; oligodendroglioma; oligodendroblastoma; primitive neuroectodermal; cerebellar sarcoma; ganglioneuroblastoma; neuroblastoma; retinoblastoma; olfactory neurogenic tumor; meningioma, malignant; neurofibrosarcoma; neurilemmoma, malignant; granular cell tumor, malignant;

malignant lymphoma; hodgkin's disease; hodgkin's; paragranuloma; malignant lymphoma, small lymphocytic; malignant lymphoma, large cell, diffuse; malignant lymphoma, follicular; mycosis fungoides; other specified non-hodgkin's lymphomas; malignant histiocytosis; multiple myeloma; mast cell sarcoma; immunoproliferative small intestinal disease; leukemia; lymphoid leukemia; plasma cell leukemia; erythroleukemia; lymphosarcoma cell leukemia; myeloid leukemia; basophilic leukemia; eosinophilic leukemia; monocytic leukemia; mast cell leukemia; megakaryoblastic leukemia; myeloid sarcoma; and hairy cell leukemia.

III. Detecting a Genetic Signature

[0049] Particular embodiments concern the methods of detecting a genetic signature in an individual. The genetic signature may identify a treatment regimen for an individual and/or may identify whether an individual will have poor metastasis-free survival. In particular embodiments, one or more genes from cytoband 19q13.31-33 are analyzed to either determine an appropriate treatment and/or determine whether or not the individual will have poor metastasis-free survivial. In specific embodiments, when a sample from an individual identifies gene deletion and/or low mRNA expression and/or low protein level of LIG1; POLD1; and/or XRCC1, the individual is administered a therapeutically effective amount of a particular therapy, including of a therapy that is not a platinum-based drug. In some embodiments, when a sample from an individual identifies gene deletion and/or low mRNA expression and/or low protein level of LIG1; POLD1; and/or XRCC1, the individual will have poor metastasis-free survival. In either case, the individual may be administered a therapeutically effective amount of one or more therapies that are not a platinum-based drug.

[0050] A genetic signature may encompass the genetic DNA and/or its expression into mRNA and/or protein. In some embodiments, the method for detecting the genetic signature may include selective oligonucleotide probes, arrays, allele-specific hybridization, molecular beacons, restriction fragment length polymorphism analysis, enzymatic chain reaction, flap endonuclease analysis, primer extension, 5'-nuclease analysis, oligonucleotide ligation assay, single strand conformation polymorphism analysis, temperature gradient gel electrophoresis, denaturing high performance liquid chromatography, high-resolution melting, DNA mismatch binding protein analysis, surveyor nuclease assay, sequencing, RNASeq, BEAMing, or a combination thereof, for example. The method for detecting the genetic signature may include fluorescent in situ

hybridization, comparative genomic hybridization, arrays, polymerase chain reaction, sequencing, or a combination thereof, for example. The detection of the genetic signature may involve using a particular method to detect one feature of the genetic signature and additionally use the same method or a different method to detect a different feature of the genetic signature. Multiple different methods independently or in combination may be used to detect the same feature or a plurality of features.

[0051] The present disclosure concerns methods that facilitate analysis of biological samples for accurate treatment or prognosis for an individual. In particular cases, an individual that has cancer or is suspected of having cancer is subject to methods of the disclosure to provide a correct assessment to allow for selection of one or more suitable treatments. The methods greatly reduce the risk of inaccurate treatment regimens at least in part because they employ proteogenomics that encompass analysis of DNA, RNA, and protein as part of the evaluation for the individual. The particular methods of the disclosure, however, reduce the level of required tissue and include uniform distribution of sample parts (such as sections) each for the DNA analysis, RNA analysis, and protein analysis. The DNA analysis, RNA analysis, and protein analysis may or may not occur in parallel, although in particular cases the different analyses occur generally concomitantly. In specific cases, following distribution of the sample sections for the DNA analysis, RNA analysis, and protein analysis, mass spectrometry may be utilized to analyze the proteome and/or phosphoproteome from the biological sample. In specific cases, analysis of the proteome and/or phosphoproteome utilizes a scale of tissue on the order of micrograms.

[0052] Certain embodiments of the disclosure concern at least one biological sample of any kind, such as a biopsy of any kind, taken from an individual, including any individual encompassed herein. The individual may have cancer, may be suspected of having cancer, or may be at increased risk for having cancer compared to the general population (for example, a personal or family history, a smoker, the elderly, exposure to the sun or environmental conditions, a combination thereof, and so forth). The individual may be a research subject, including any mammal that is part of a research study. A biopsy may be obtained as part of a routine preventative practice or as part of a directed concern or suspected indication for the onset of cancer. The biopsy may be taken from the breast, for example from a mass in the breast. Alternatively nipple aspirate may provide the biological sample.

A. Biological Samples

[0053] The one or more biological samples may be taken from the individual at any time, such as before, after, and/or simultaneously with diagnosis of a cancer, or such as before, after, and/or simultaneously with the administration of one or more therapies. At least one of the biological samples taken from the individual may be taken from a tumor or other cancer cells present in the individual. A tumor may or may not be benign or suspected of being benign. In some embodiments, at least one of the biological samples taken from the individual may be taken from non-cancerous tissue or other biological material in the individual. The biological sample may be taken from the individual using any method known in the art, including a(n) needle biopsy (including core-needle biopsy), guided biopsy, aspiration biopsy, surgical biopsy, core biopsy, open biopsy, punch biopsy, sentinel lymph node biopsy, shave biopsy, endoscopic biopsy, or a combination thereof. In some embodiments, the biological sample is taken from the individual by a core-needle biopsy, including a core-needle biopsy using a 14 gauge, 15 gauge, 16 gauge, 17 gauge, 18 gauge, 19 gauge, 20 gauge, 21 gauge, or 22 gauge needle. The biological sample may be from tissue, bone, blood, serum, plasma, urine, stool, sputum, saliva, semen, vaginal fluids, mucus, fat, or a combination thereof. In specific embodiments, a biological sample comprises a biopsy that is taken from a mass in a breast of an individual.

[0054] The biological sample may be prepared, processed, stored, handled, and/or fixed using any method known in the art. The sample may or may not be stored prior to processing. In some embodiments, at least one biological sample is embedded in optimal cutting temperature (OCT) medium. In some embodiments, at least one biological sample is stored in cryogenic storage, such as at a temperature lower than -80°C, lower than -70°C, lower than -60°C, lower than -50°C, lower than -40°C, lower than -30°C, or lower than -20°C. In some embodiments, at least one biological sample is prepared and/or sectioned using a microtome, such as a crytostat. In some embodiments, the sectioning is done at a temperature lower than -30°C, lower than -20°C, or lower than -10°C. In some embodiments, the sectioning is done at a temperature between -30°C to -10°C, or between -23°C to -15°C. In some embodiments, the biological sample is sectioned at a thickness between 3 microns and 100 microns, or between 4 microns and 100 microns, or between 5 microns and 100 microns, or between 3 microns and 50 microns, or between 4 microns and 50 microns, or between 5 microns and 50 microns. However, the biological sample may be sectioned to any thickness that is suitable for practicing the methods of the disclosure. In some embodiments, one or more sections

from the biological samples are placed into one or more vessels suitable for comprising the sections.

B. Production of DNA, RNA, and/or Protein

[0055] Certain embodiments of the present disclosure concern methods of generating DNA, RNA, and/or protein from at least one biological sample, such as a biopsy. In some embodiments, the DNA, RNA, and/or protein are isolated in individual vessels, including vessels comprising one or more sections of the biological sample(s). The DNA, RNA, and/or protein isolated into individual vessels may have originated from different regions of the biological sample, such that the isolated DNA, RNA, and/or protein in an individual vessel comprises DNA, RNA, and/or protein originating from different regions of the biological sample. In some embodiments, at least one biological sample contained in a vessel is used for the preparation of the section(s) for microscopic analysis.

[0056] In certain embodiments, some sections of the biological sample are combined with other sections of the biological sample. The combining of some sections with other sections may comprise combining sections of one region of the biological sample with at least one other region of the biological sample. In certain embodiments, non-adjacent sections from the biological sample are combined, including non-adjacent sections from different regions of the biological sample.

[0057] In some embodiments, at least three sections are generated from a biopsy, followed by adding any one of the three sections to a first vessel, adding any one of the two remaining sections to a second vessel, and adding the remaining section to a third vessel. The process may be repeated indefinitely. The processes may be repeated until a sufficient number of sections, from the biological samples, for practice of the disclosure are generated and placed into vessels. A sufficient number may be the number required to produce sufficient RNA, DNA, and/or protein for analysis. In some embodiments, four sections are generated followed by, in any order: adding any one of the four sections to a first vessel, adding any one of the four sections not in the first vessel to a second vessel, adding any one of the four sections not in the first or second vessel to a third vessel, and preparing any one of the four sections not in the first, second, or third vessel for microscopic analysis. The process may be repeated indefinitely. The processes may be repeated until a sufficient number of sections, from the biological samples, for practice of the disclosure are

generated and placed into vessels and/or prepared for microscopic analysis. A sufficient number may be the number required to produce sufficient RNA, DNA, and/or protein for analysis.

[0058] In some embodiments, a sufficient number of sections for practice of the disclosure will produce approximately between 10 μg to 45 μg of isolated protein. In some embodiments, a sufficient number of sections for practice of the disclosure will produce approximately 10 μg, 15 μg, 20 μg, 25 μg, 30 μg, 35 μg, 40 μg, 45 μg, or more than 45 μg of isolated protein. In some embodiments, a sufficient number of sections for practice of the disclosure will produce approximately between 0.1 μg to 1 μg of isolated DNA. In some embodiments, a sufficient number of sections for practice of the disclosure will produce approximately 0.1 μg, 0.2 μg, 0.3 μg, 0.4 μg, 0.5 μg, 0.6 μg, 0.7 μg, 0.8 μg, 0.9 μg, 1.0 μg, or more than 1.0 μg of isolated DNA. In some embodiments, a sufficient number of sections for practice of the disclosure will produce approximately between 0.1 μg to 1 μg of isolated RNA. In some embodiments, a sufficient number of sections for practice of the disclosure will produce approximately 0.1 μg, 0.2 μg, 0.3 μg, 0.4 μg, 0.5 μg, 0.6 μg, 0.7 μg, 0.8 μg, 0.9 μg, 1.0 μg, or more than 1.0 μg of isolated RNA.

[0059] Certain embodiments of the disclosure concern the isolation of DNA, RNA, and/or protein from one or more sections generated from at least one biological sample. The DNA, RNA, and/or protein may be isolated using any method known in the art. In some embodiments, the DNA may be isolated from one or more sections of at least one biological sample by digesting the section(s) with a proteinase and an RNase, then purifying the DNA, such as by ethanol precipitation and/or a column purification system. In some embodiments, the RNA may be isolated from one or more sections of at least one biological sample by incubating the section(s) with an RNA extraction reagent, such as TRIzol reagent. The TRIzol reagent incubated sections may be sonicated and the organic layer may be extracted using an organic solvent, such as chloroform. The resulting RNA may be dissolved in a suitable solution (including water) and further purified, such as by ethanol precipitation and/or a column purification system. In some embodiments, protein, including native and/or denatured protein, may be isolated from one or more sections of at least one biological sample such as by, optionally precipitating the sections with ethanol, followed by incubation with a suitable lysis buffer. In some embodiments, the DNA, RNA, and/or protein isolated are subjected to quality control analysis.

C. Biological Sample Analysis of DNA, RNA, and/or Protein

[0060] Certain embodiments of the present disclosure concern methods for analyzing biological samples taken from an individual, including any individual encompassed herein. In some embodiments, DNA, RNA, and/or protein isolated from one or more sections of at least one biological sample is analyzed. Analyzing DNA, RNA, and/or protein may comprise, for example, any PCR technique (such as allele-specific PCR, qPCR, RT-qPCR, multiplex PCR, and/or digital PCR), the use of restriction enzymes (such as for restriction fragment length polymorphism analysis or the like), any sequencing technique (such as Sanger sequencing, next generation sequencing, high throughput sequencing, deep sequencing, nanopore sequencing, exome sequencing, and/or single cell sequencing), Northern blotting, Western blotting, Southern blotting, flow cytometry, mass spectrometry, NMR spectroscopy, electrophoresis, or a combination thereof. Additionally, DNA may also be isolated from blood samples from the same individual to predict if a genomic alteration is somatic or germline.

[0061] In some embodiments, DNA may be analyzed by sequencing. The DNA may be prepared for sequencing by any method known in the art, such as library preparation, hybrid capture, sample quality control, product-utilized ligation-based library preparation, or a combination thereof. The DNA may be prepared for any sequencing technique, including whole exome sequencing. In some embodiments, a unique genetic readout for each sample may be generated by genotyping one or more highly polymorphic SNPs. In some embodiments, sequencing, such as 76 base pair, paired-end sequencing, may be performed to cover approximately 70%, 75%, 80%, 85%, 90%, 95%, 99%, or greater percentage of targets at more than 20x, 25x, 30x, 35x, 40x, 45x, 50x, or greater than 50x coverage. In certain embodiments, mutations, SNPS, INDELS, copy number alterations (somatic and/or germline), or other genetic differences may be identified from the sequencing, such as whole exome sequencing, using at least one bioinformatics tool, including Strelka2, Mutect2, CARNAC, Pindel , GISTIC,, any R package (including CopywriteR) and/or Annovar.

[0062] In some embodiments, RNA may be analyzed by sequencing. The RNA may be prepared for sequencing by any method known in the art, such as poly-A selection, cDNA synthesis, stranded or nonstranded library preparation, or a combination thereof. The RNA may be prepared for any type of RNA sequencing technique, including stranded specific RNA sequencing. In some embodiments, sequencing may be performed to generate approximately 10M, 15M, 20M, 25M, 30M, 35M, 40M or more reads, including paired reads. The sequencing may be

performed at a read length of approximately 50 bp, 55 bp, 60 bp, 65 bp, 70 bp, 75 bp, 80 bp, 85 bp, 90 bp, 95 bp, 100 bp, 105 bp, 110 bp, or longer. In some embodiments, raw sequencing data may be converted to estimated read counts (RSEM), fragments per kilobase of transcript per million mapped reads (FPKM), and/or reads per kilobase of transcript per million mapped reads (RPKM). In some embodiments, one or more bioinformatics tools may be used to infer stroma content, immune infiltration, and/or tumor immune cell profiles, such as by using upper quartile normalized FPKM data.

[0063] In particular embodiments, protein from samples is analyzed, including denatured protein. The protein may be analyzed by mass spectrometry. The protein may be prepared for mass spectrometry using any method known in the art. In specific embodiments, and regardless of the methods of analyzing protein, the protein is digested to produce peptides that are then analyzed. Protein, including any isolated protein encompassed herein, may be treated with DTT followed by iodoacetamide. The protein may be incubated with at least one peptidase, including an endopeptidase, proteinase, protease, or any enzyme that cleaves proteins. In some embodiments, protein is incubated with the endopeptidase, LysC and/or trypsin. The protein may be incubated with one or more protein-cleaving enzymes at any ratio, including a ratio of µg of enzyme to µg protein at approximately 1:1000, 1:100, 1:90, 1:80, 1:70, 1:60, 1:50, 1:40, 1:30, 1:20, 1:10, 1:1, or any range between. In some embodiments, the cleaved proteins may be purified, such as by column purification. In certain embodiments, purified peptides may be snap-frozen and/or dried, such as dried under vacuum. In some embodiments, the purified peptides may be fractionated, such as by reverse phase chromatography or basic reverse phase chromatography. Fractions may be combined for practice of the methods of the disclosure. In some embodiments, one or more fractions, including the combined fractions, are subject to enrichment based on one or more post-translational modifications, such as phosphopeptide enrichment, including phospho-enrichment by affinity chromatography and/or binding, ion exchange chromatography, chemical derivatization, immunoprecipitation, co-precipitation, or a combination thereof. The entirety or a portion of one or more fractions, including the combined fractions and/or phospho-enriched fractions, may be subject to mass spectrometry. In some embodiments, the raw mass spectrometry data may be processed and normalized using at least one relevant bioinformatics tool. In specific cases, the protein is analyzed with mass spectrometry instead of antibody-based analysis.

[0064] In one embodiment, one can carry out further enrichments from the unbound fraction from the phospho-enrichment step to obtain e.g., acetylated peptide or other modified peptides.

D. Sample Preparation

[0065] In certain aspects, methods involve obtaining a sample from a subject. The methods of obtaining provided herein may include methods of biopsy such as fine needle aspiration, core needle biopsy, vacuum assisted biopsy, incisional biopsy, excisional biopsy, punch biopsy, shave biopsy or skin biopsy. In certain embodiments the sample is obtained from a biopsy from esophageal tissue by any of the biopsy methods previously mentioned. In other embodiments the sample may be obtained from any of the tissues provided herein that include but are not limited to non-cancerous or cancerous tissue and non-cancerous or cancerous tissue from the serum, gall bladder, mucosal, skin, heart, lung, breast, pancreas, blood, liver, muscle, kidney, smooth muscle, bladder, colon, intestine, brain, prostate, esophagus, or thyroid tissue. Alternatively, the sample may be obtained from any other source including but not limited to blood, sweat, hair follicle, buccal tissue, tears, menses, feces, or saliva. In certain aspects of the current methods, any medical professional such as a doctor, nurse or medical technician may obtain a biological sample for testing. Yet further, the biological sample can be obtained without the assistance of a medical professional.

[0066] A sample may include but is not limited to, tissue, cells, or biological material from cells or derived from cells of a subject. The biological sample may be a heterogeneous or homogeneous population of cells or tissues. The biological sample may be obtained using any method known to the art that can provide a sample suitable for the analytical methods described herein. The sample may be obtained by non-invasive methods including but not limited to: scraping of the skin or cervix, swabbing of the cheek, saliva collection, urine collection, feces collection, collection of menses, tears, or semen.

[0067] The sample may be obtained by methods known in the art. In certain embodiments the samples are obtained by biopsy. In other embodiments the sample is obtained by swabbing, endoscopy, scraping, phlebotomy, or any other methods known in the art. In some cases, the sample may be obtained, stored, or transported using components of a kit of the present methods. In some cases, multiple samples, such as multiple esophageal samples may be obtained for diagnosis by the methods described herein. In other cases, multiple samples, such as one or more

samples from one tissue type (for example esophagus) and one or more samples from another specimen (for example serum) may be obtained for diagnosis by the methods. In some cases, multiple samples such as one or more samples from one tissue type (e.g. esophagus) and one or more samples from another specimen (e.g. serum) may be obtained at the same or different times. Samples may be obtained at different times are stored and/or analyzed by different methods. For example, a sample may be obtained and analyzed by routine staining methods or any other cytological analysis methods.

[0068] In some embodiments the biological sample may be obtained by a physician, nurse, or other medical professional such as a medical technician, endocrinologist, cytologist, phlebotomist, radiologist, or a pulmonologist. The medical professional may indicate the appropriate test or assay to perform on the sample. In certain aspects a molecular profiling business may consult on which assays or tests are most appropriately indicated. In further aspects of the current methods, the patient or subject may obtain a biological sample for testing without the assistance of a medical professional, such as obtaining a whole blood sample, a urine sample, a fecal sample, a buccal sample, or a saliva sample.

[0069] In other cases, the sample is obtained by an invasive procedure including but not limited to: biopsy, needle aspiration, endoscopy, or phlebotomy. The method of needle aspiration may further include fine needle aspiration, core needle biopsy, vacuum assisted biopsy, or large core biopsy. In some embodiments, multiple samples may be obtained by the methods herein to ensure a sufficient amount of biological material.

[0070] General methods for obtaining biological samples are also known in the art. Publications such as Ramzy, Ibrahim Clinical Cytopathology and Aspiration Biopsy 2001, which is herein incorporated by reference in its entirety, describes general methods for biopsy and cytological methods. In one embodiment, the sample is a fine needle aspirate of a esophageal or a suspected esophageal tumor or neoplasm. In some cases, the fine needle aspirate sampling procedure may be guided by the use of an ultrasound, X-ray, or other imaging device.

[0071] In some embodiments of the present methods, the molecular profiling business may obtain the biological sample from a subject directly, from a medical professional, from a third party, or from a kit provided by a molecular profiling business or a third party. In some cases, the biological sample may be obtained by the molecular profiling business after the subject, a medical professional, or a third party acquires and sends the biological sample to the molecular profiling

business. In some cases, the molecular profiling business may provide suitable containers, and excipients for storage and transport of the biological sample to the molecular profiling business.

[0072] In some embodiments of the methods described herein, a medical professional need not be involved in the initial diagnosis or sample acquisition. An individual may alternatively obtain a sample through the use of an over the counter (OTC) kit. An OTC kit may contain a means for obtaining said sample as described herein, a means for storing said sample for inspection, and instructions for proper use of the kit. In some cases, molecular profiling services are included in the price for purchase of the kit. In other cases, the molecular profiling services are billed separately. A sample suitable for use by the molecular profiling business may be any material containing tissues, cells, nucleic acids, genes, gene fragments, expression products, gene expression products, or gene expression product fragments of an individual to be tested. Methods for determining sample suitability and/or adequacy are provided.

[0073] In some embodiments, the subject may be referred to a specialist such as an oncologist, surgeon, or endocrinologist. The specialist may likewise obtain a biological sample for testing or refer the individual to a testing center or laboratory for submission of the biological sample. In some cases the medical professional may refer the subject to a testing center or laboratory for submission of the biological sample. In other cases, the subject may provide the sample. In some cases, a molecular profiling business may obtain the sample.

E. Single Nucleotide Polymorphism (SNP) Detection

[0074] Particular embodiments of the disclosure concern methods of detecting a SNP in an individual. One may employ any of the known general methods for detecting SNPs for detecting the particular SNP in this disclosure, for example. Such methods include, but are not limited to, selective oligonucleotide probes, arrays, allele-specific hybridization, molecular beacons, restriction fragment length polymorphism analysis, enzymatic chain reaction, flap endonuclease analysis, primer extension, 5'-nuclease analysis, oligonucleotide ligation assay, single strand conformation polymorphism analysis, temperature gradient gel electrophoresis, denaturing high performance liquid chromatography, high-resolution melting, DNA mismatch binding protein analysis, surveyor nuclease assay, sequencing, or a combination thereof.

[0075] In some embodiments of the disclosure, the method used to detect the SNP comprises sequencing nucleic acid material from the individual and/or using selective oligonucleotide probes.

Sequencing the nucleic acid material from the individual may involve obtaining the nucleic acid material from the individual in the form of genomic DNA, complementary DNA that is reverse transcribed from RNA, or RNA, for example. Any standard sequencing technique may be employed, including Sanger sequencing, chain extension sequencing, Maxam-Gilbert sequencing, shotgun sequencing, bridge PCR sequencing, high-throughput methods for sequencing, next generation sequencing, RNA sequencing, or a combination thereof. After sequencing the nucleic acid from the individual, one may utilize any data processing software or technique to determine which particular nucleotide is present in the individual at the particular SNP.

[0076] In some embodiments, the nucleotide at the particular SNP is detected by selective oligonucleotide probes. The probes may be used on nucleic acid material from the individual, including genomic DNA, complementary DNA that is reverse transcribed from RNA, or RNA, for example. Selective oligonucleotide probes preferentially bind to a complementary strand based on the particular nucleotide present at the SNP. For example, one selective oligonucleotide probe binds to a complementary strand that has an A nucleotide at the SNP on the coding strand but not a G nucleotide at the SNP on the coding strand, while a different selective oligonucleotide probe binds to a complementary strand that has a G nucleotide at the SNP on the coding strand but not an A nucleotide at the SNP on the coding strand. Similar methods could be used to design a probe that selectively binds to the coding strand that has a C or a T nucleotide, but not both, at the SNP. Thus, any method to determine binding of one selective oligonucleotide probe over another selective oligonucleotide probe could be used to determine the nucleotide present at the SNP.

[0077] One method for detecting SNPs using oligonucleotide probes comprises the steps of analyzing the quality and measuring quantity of the nucleic acid material by a spectrophotometer and/or a gel electrophoresis assay; processing the nucleic acid material into a reaction mixture with at least one selective oligonucleotide probe, PCR primers, and a mixture with components needed to perform a quantitative PCR (qPCR), which could comprise a polymerase, deoxynucleotides, and a suitable buffer for the reaction; and cycling the processed reaction mixture while monitoring the reaction. In one embodiment of the method, the polymerase used for the qPCR will encounter the selective oligonucleotide probe binding to the strand being amplified and, using endonuclease activity, degrade the selective oligonucleotide probe. The detection of the degraded probe determines if the probe was binding to the amplified strand.

[0078] Another method for determining binding of the selective oligonucleotide probe to a particular nucleotide comprises using the selective oligonucleotide probe as a PCR primer, wherein the selective oligonucleotide probe binds preferentially to a particular nucleotide at the SNP position. In some embodiments, the probe is generally designed so the 3' end of the probe pairs with the SNP. Thus, if the probe has the correct complementary base to pair with the particular nucleotide at the SNP, the probe will be extended during the amplification step of the PCR. For example, if there is a T nucleotide at the 3' position of the probe and there is an A nucleotide at the SNP position, the probe will bind to the SNP and be extended during the amplification step of the PCR. However, if the same probe is used (with a T at the 3' end) and there is a G nucleotide at the SNP position, the probe will not fully bind and will not be extended during the amplification step of the PCR.

[0079] In some embodiments, the SNP position is not at the terminal end of the PCR primer, but rather located within the PCR primer. The PCR primer should be of sufficient length and homology in that the PCR primer can selectively bind to one variant, for example the SNP having an A nucleotide, but not bind to another variant, for example the SNP having a G nucleotide. The PCR primer may also be designed to selectively bind particularly to the SNP having a G nucleotide but not bind to a variant with an A, C, or T nucleotide. Similarly, PCR primers could be designed to bind to the SNP having a C or a T nucleotide, but not both, which then does not bind to a variant with a G, A, or T nucleotide or G, A, or C nucleotide respectively. In particular embodiments, the PCR primer is at least or no more than 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,3 5, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, or more nucleotides in length with 100% homology to the template sequence, with the potential exception of non-homology the SNP location. After several rounds of amplifications, if the PCR primers generate the expected band size, the SNP can be determined to have the A nucleotide and not the G nucleotide.

[0080] In some embodiments, a germline or somatic SNP in the LIG1 gene of an individual predicts susceptibility or resistance to a chemotherapy, including a platinum-based chemotherapy.

F. Copy Number Variation Detection

[0081] Particular embodiments of the disclosure concern methods of detecting a copy number variation (CNV) of a particular allele. One can utilize any known method for detecting CNVs to detect the CNVs. Such methods include fluorescent in situ hybridization, comparative genomic hybridization, arrays, polymerase chain reaction, sequencing, or a combination thereof, for example. In some embodiments, the CNV is detected using an array, wherein the array is capable of detecting CNVs on the entire X chromosome and/or all targets of miR-362. Array platforms such as those from Agilent, Illumina, or Affymetrix may be used, or custom arrays could be designed. One example of how an array may be used includes methods that comprise one or more of the steps of isolating nucleic acid material in a suitable manner from an individual suspected of having the CNV and, at least in some cases from an individual or reference genome that does not have the CNV; processing the nucleic acid material by fragmentation, labelling the nucleic acid with, for example, fluorescent labels, and purifying the fragmented and labeled nucleic acid material; hybridizing the nucleic acid material to the array for a sufficient time, such as for at least 24 hours; washing the array after hybridization; scanning the array using an array scanner; and analyzing the array using suitable software. The software may be used to compare the nucleic acid material from the individual suspected of having the CNV to the nucleic acid material of an individual who is known not to have the CNV or a reference genome.

[0082] In some embodiments, detection of a CNV is achieved by polymerase chain reaction (PCR). PCR primers can be employed to amplify nucleic acid at or near the CNV wherein an individual with a CNV will result in measurable higher levels of PCR product when compared to a PCR product from a reference genome. The detection of PCR product amounts could be measured by quantitative PCR (qPCR) or could be measured by gel electrophoresis, as examples. Quantification using gel electrophoresis comprises subjecting the resulting PCR product, along with nucleic acid standards of known size, to an electrical current on an agarose gel and measuring the size and intensity of the resulting band. The size of the resulting band can be compared to the known standards to determine the size of the resulting band. In some embodiments, the amplification of the copy number (CN) will result in a band that has a larger size than a band that is amplified, using the same primers as were used to detect the CNV, from a reference genome or an individual that does not have the CNV being detected. The resulting band from the CN amplification may be nearly double, double, or more than double the resulting band from the reference genome or the resulting band from an individual that does not have the CNV being

detected. In some embodiments, the deletion of the CN will result in a band that has a smaller size than a band that is deleted, using the same primers as were used to detect the CNV, from a reference genome or an individual that does not have the CNV being detected. The resulting band from the CN deletion may be nearly half, half, or less than half the resulting band from the reference genome or the resulting band from an individual that does not have the CNV being detected. In some embodiments, the CNV can be detected using nucleic acid sequencing. Sequencing techniques that could be used include, but are not limited to, whole genome sequencing, whole exome sequencing, and/or targeted sequencing.

[0083] In some embodiments, a germline or a somatic CNV in the LIG1 gene of an individual predicts susceptibility or resistance to a chemotherapy, including a platinum-based chemotherapy.

G. DNA Sequencing

[0084] In some embodiments, DNA may be analyzed by sequencing. The DNA may be prepared for sequencing by any method known in the art, such as library preparation, hybrid capture, sample quality control, product-utilized ligation-based library preparation, or a combination thereof. The DNA may be prepared for any sequencing technique. In some embodiments, a unique genetic readout for each sample may be generated by genotyping one or more highly polymorphic SNPs. In some embodiments, sequencing, such as 76 base pair, paired-end sequencing, may be performed to cover approximately 70%, 75%, 80%, 85%, 90%, 95%, 99%, or greater percentage of targets at more than 20x, 25x, 30x, 35x, 40x, 45x, 50x, or greater than 50x coverage. In certain embodiments, mutations, SNPS, INDELS, copy number alterations (somatic and/or germline), or other genetic differences may be identified from the sequencing using at least one bioinformatics tool, including Strelka2, Mutect2, CARNAC, Pindel , GISTIC, any R package (including CopywriteR) and/or Annovar.

H. RNA Sequencing

[0085] In some embodiments, RNA may be analyzed by sequencing. The RNA may be prepared for sequencing by any method known in the art, such as poly-A selection, cDNA synthesis, stranded or nonstranded library preparation, or a combination thereof. The RNA may be prepared for any type of RNA sequencing technique, including stranded specific RNA sequencing. In some embodiments, sequencing may be performed to generate approximately 10M,

15M, 20M, 25M, 30M, 35M, 40M or more reads, including paired reads. The sequencing may be performed at a read length of approximately 50 bp, 55 bp, 60 bp, 65 bp, 70 bp, 75 bp, 80 bp, 85 bp, 90 bp, 95 bp, 100 bp, 105 bp, 110 bp, or longer. In some embodiments, raw sequencing data may be converted to estimated read counts (RSEM), fragments per kilobase of transcript per million mapped reads (FPKM), and/or reads per kilobase of transcript per million mapped reads (RPKM). In some embodiments, one or more bioinformatics tools may be used to infer stroma content, immune infiltration, and/or tumor immune cell profiles, such as by using upper quartile normalized RSEM data.

I. Proteomics

[0086] In some embodiments, protein may be analyzed by mass spectrometry. The protein may be prepared for mass spectrometry using any method known in the art. Protein, including any isolated protein encompassed herein, may be treated with DTT or TCEP followed by iodoacetamide. The protein may be incubated with at least one peptidase, including an endopeptidase, proteinase, protease, or any enzyme that cleaves proteins. In some embodiments, protein is incubated with the endopeptidase, LysC and/or trypsin. The protein may be incubated with one or more protein cleaving enzymes at any ratio, including a ratio of µg of enzyme to µg protein at approximately 1:1000, 1:100, 1:90, 1:80, 1:70, 1:60, 1:50, 1:40, 1:30, 1:20, 1:10, 1:1, or any range between. In some embodiments, the cleaved proteins may be purified, such as by column purification. In certain embodiments, purified peptides may be snap-frozen and/or dried, such as dried under vacuum. In some embodiments, the purified peptides may be fractionated, such as by reverse phase chromatography or basic reverse phase chromatography. Fractions may be combined for practice of the methods of the disclosure. In some embodiments, one or more fractions, including the combined fractions, are subject to phosphopeptide enrichment, including phospho-enrichment by affinity chromatography and/or binding, ion exchange chromatography, chemical derivatization, immunoprecipitation, co-precipitation, or a combination thereof. The entirety or a portion of one or more fractions, including the combined fractions and/or phospho-enriched fractions, may be subject to mass spectrometry. In some embodiments, the raw mass spectrometry data may be processed and normalized using at least one relevant bioinformatics tool.

J. Detection Kits and Systems

[0087] One can recognize that based on the methods described herein, detection reagents, kits, and/or systems can be utilized to detect the SNP and/or the CNV related to the genetic signature for diagnosing an individual (the detection either individually or in combination). The reagents can be combined into at least one of the established formats for kits and/or systems as known in the art. As used herein, the terms "kits" and "systems" refer to embodiments such as combinations of at least one SNP detection reagent, for example at least one selective oligonucleotide probe, and at least one CNV detection reagent, for example at least one PCR primer. The kits could also contain other reagents, chemicals, buffers, enzymes, packages, containers, electronic hardware components, etc. The kits/systems could also contain packaged sets of PCR primers, oligonucleotides, arrays, beads, or other detection reagents. Any number of probes could be implemented for a detection array. In some embodiments, the detection reagents and/or the kits/systems are paired with chemiluminescent or fluorescent detection reagents. Particular embodiments of kits/systems include the use of electronic hardware components, such as DNA chips or arrays, or microfluidic systems, for example. In specific embodiments, the kit also comprises one or more therapeutic or prophylactic interventions in the event the individual is determined to be in need of.

[0088] In specific embodiments, the kit may comprise one or both of a composition for detecting a polymorphism and a composition for detecting a CNV. In certain embodiments, the polymorphism detected is polymorphism rs4567 as represented by position 101 of SEQ ID NO:1. In certain embodiments, the CNV detected is represented by SEQ ID NO:2. The composition in the kit for detecting the polymorphism may be selected from the group consisting of oligonucleotide, one or more primers suitable for amplifying the polymorphism, one or more sequencing reagents, and a combination thereof. The composition in the kit for detecting the CNV may be selected from the group consisting of one or more primers suitable for amplifying the polymorphism, one or more sequencing reagents, one or more arrays, and a combination thereof.

IV. Administration of Therapeutic Compositions

[0089] The therapy provided herein may comprise administration of one or a combination of therapeutic agents, such as a first cancer therapy and a second cancer therapy. The therapies may be administered in any suitable manner known in the art. In some embodiments, the first and second cancer treatment may be administered sequentially (at different times) or concurrently (at

the same time). In some embodiments, the first and second cancer treatments are administered in a separate composition. In some embodiments, the first and second cancer treatments are in the same composition.

[0090] In some embodiments, the first therapy and the second therapy are administered substantially simultaneously. In some embodiments, the first therapy and the second therapy are administered sequentially. In some embodiments, the first therapy, the second therapy, and a third therapy are administered sequentially. In some embodiments, the first therapy is administered before administering the second therapy. In some embodiments, the first therapy is administered after administering the second therapy.

[0091] Embodiments of the disclosure relate to compositions and methods comprising therapeutic compositions. The different therapies may be administered in one composition or in more than one composition, such as 2 compositions, 3 compositions, or 4 compositions. Various combinations of the agents may be employed.

[0092] In some embodiments, the cancer therapy is administered intravenously, intramuscularly, subcutaneously, topically, orally, transdermally, intraperitoneally, intraorbitally, by implantation, by inhalation, intrathecally, intraventricularly, or intranasally. In some embodiments, the antibiotic is administered intravenously, intramuscularly, subcutaneously, topically, orally, transdermally, intraperitoneally, intraorbitally, by implantation, by inhalation, intrathecally, intraventricularly, or intranasally. The appropriate dosage may be determined based on the type of disease to be treated, severity and course of the disease, the clinical condition of the individual, the individual's clinical history and response to the treatment, and the discretion of the attending physician.

[0093] The treatments may include various "unit doses." Unit dose is defined as containing a predetermined-quantity of the therapeutic composition. The quantity to be administered, and the particular route and formulation, is within the skill of determination of those in the clinical arts. A unit dose need not be administered as a single injection but may comprise continuous infusion over a set period of time. In some embodiments, a unit dose comprises a single administrable dose.

[0094] The quantity to be administered, both according to number of treatments and unit dose, depends on the treatment effect desired. An effective dose is understood to refer to an amount necessary to achieve a particular effect. In the practice in certain embodiments, it is contemplated that doses in the range from 10 mg/kg to 200 mg/kg can affect the protective capability of these

agents. Thus, it is contemplated that doses include doses of about 0.1, 0.5, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, and 200, 300, 400, 500, 1000 μg/kg, mg/kg, μg/day, or mg/day or any range derivable therein. Furthermore, such doses can be administered at multiple times during a day, and/or on multiple days, weeks, or months.

[0095] In certain embodiments, the effective dose of the pharmaceutical composition is one which can provide a blood level of about 1 μM to 150 μM. In another embodiment, the effective dose provides a blood level of about 4 μM to 100 μM.; or about 1 μM to 100 μM; or about 1 μM to 50 μM; or about 1 μM to 40 μM; or about 1 μM to 30 μM; or about 1 μM to 20 μM; or about 1 μM to 10 μM; or about 10 μM to 150 μM; or about 10 μM to 100 μM; or about 10 μM to 50 μM; or about 25 μM to 150 μM; or about 25 μM to 100 μM; or about 25 μM to 50 μM; or about 50 μM to 150 μM; or about 50 μM to 100 μM (or any range derivable therein). In other embodiments, the dose can provide the following blood level of the agent that results from a therapeutic agent being administered to a subject: about, at least about, or at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 μM or any range derivable therein. In certain embodiments, the therapeutic agent that is administered to a subject is metabolized in the body to a metabolized therapeutic agent, in which case the blood levels may refer to the amount of that agent. Alternatively, to the extent the therapeutic agent is not metabolized by a subject, the blood levels discussed herein may refer to the unmetabolized therapeutic agent.

[0096] Precise amounts of the therapeutic composition also depend on the judgment of the practitioner and are peculiar to each individual. Factors affecting dose include physical and clinical state of the patient, the route of administration, the intended goal of treatment (alleviation of symptoms versus cure) and the potency, stability and toxicity of the particular therapeutic substance or other therapies a subject may be undergoing.

[0097] It will be understood by those skilled in the art and made aware that dosage units of μg/kg or mg/kg of body weight can be converted and expressed in comparable concentration units of μg/ml or mM (blood levels), such as 4 μM to 100 μM. It is also understood that uptake is species and organ/tissue dependent. The applicable conversion factors and physiological

assumptions to be made concerning uptake and concentration measurement are well-known and would permit those of skill in the art to convert one concentration measurement to another and make reasonable comparisons and conclusions regarding the doses, efficacies and results described herein.

[0098] In certain instances, it will be desirable to have multiple administrations of the composition, e.g., 2, 3, 4, 5, 6 or more administrations. The administrations can be at 1, 2, 3, 4, 5, 6, 7, 8, to 5, 6, 7, 8, 9, 10, 11, or 12 week intervals, including all ranges there between.

[0099] The phrases "pharmaceutically acceptable" or "pharmacologically acceptable" refer to molecular entities and compositions that do not produce an adverse, allergic, or other untoward reaction when administered to an animal or human. As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, anti-bacterial and anti-fungal agents, isotonic and absorption delaying agents, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredients, its use in immunogenic and therapeutic compositions is contemplated. Supplementary active ingredients, such as other anti-infective agents and vaccines, can also be incorporated into the compositions.

[0100] The active compounds can be formulated for parenteral administration, e.g., formulated for injection via the intravenous, intramuscular, subcutaneous, or intraperitoneal routes. Typically, such compositions can be prepared as either liquid solutions or suspensions; solid forms suitable for use to prepare solutions or suspensions upon the addition of a liquid prior to injection can also be prepared; and, the preparations can also be emulsified.

[0101] The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions; formulations including, for example, aqueous propylene glycol; and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. In all cases the form must be sterile and must be fluid to the extent that it may be easily injected. It also should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi.

[0102] A pharmaceutical composition can include a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of

the required particle size in the case of dispersion, and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various anti-bacterial and anti-fungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

[0103] Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various other ingredients enumerated above, as required, followed by filtered sterilization or an equivalent procedure. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques, which yield a powder of the active ingredient, plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0104] Administration of the compositions will typically be via any common route. This includes, but is not limited to oral, or intravenous administration. Alternatively, administration may be by orthotopic, intradermal, subcutaneous, intramuscular, intraperitoneal, or intranasal administration. Such compositions would normally be administered as pharmaceutically acceptable compositions that include physiologically acceptable carriers, buffers or other excipients.

[0105] Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically or prophylactically effective. The formulations are easily administered in a variety of dosage forms, such as the type of injectable solutions described above.

A. Cancer Therapy

[0106] In some embodiments, the method further comprises administering a cancer therapy to the patient. The cancer therapy may be chosen based on the assay measurements, including any or all of the proteogenomic methods, alone or in combination with a clinical risk score calculated for the patient. In some embodiments, the cancer therapy comprises a local cancer therapy. In

some embodiments, the cancer therapy excludes a systemic cancer therapy. In some embodiments, the cancer therapy excludes a local therapy. In some embodiments, the cancer therapy comprises a local cancer therapy without the administration of a system cancer therapy. In some embodiments, the cancer therapy does not comprise a platinum-based drug, such as does not comprise carboplatin, cisplatin, oxaliplatin, nedaplatin, heptaplatin, and/or lobaplatin. In some embodiments, the cancer therapy comprises an immunotherapy, which may be an immune checkpoint therapy. Any of these cancer therapies may also be excluded. Combinations of these therapies may also be administered.

B. Surgery

[0107] Approximately 60% of persons with cancer will undergo surgery of some type, which includes preventative, diagnostic or staging, curative, and palliative surgery. Curative surgery includes resection in which all or part of cancerous tissue is physically removed, excised, and/or destroyed and may be used in conjunction with other therapies, such as the treatment of the present embodiments, chemotherapy, radiotherapy, hormonal therapy, gene therapy, immunotherapy, and/or alternative therapies. Tumor resection refers to physical removal of at least part of a tumor. In addition to tumor resection, treatment by surgery includes laser surgery, cryosurgery, electrosurgery, and microscopically-controlled surgery (Mohs' surgery).

[0108] Upon excision of part or all of cancerous cells, tissue, or tumor, a cavity may be formed in the body. Treatment may be accomplished by perfusion, direct injection, or local application of the area with an additional anti-cancer therapy. Such treatment may be repeated, for example, every 1, 2, 3, 4, 5, 6, or 7 days, or every 1, 2, 3, 4, and 5 weeks or every 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 months. These treatments may be of varying dosages as well.

V. Biomolecules

A. Proteins

[0109] In some embodiments, wild-type versions of a protein or polypeptide or a modified protein are assayed. A "modified protein" or "modified polypeptide" or a "variant" refers to a protein or polypeptide whose chemical structure, particularly its amino acid sequence, is altered with respect to the wild-type protein or polypeptide. In some embodiments, a modified/variant

protein or polypeptide has at least one modified activity or function (recognizing that proteins or polypeptides may have multiple activities or functions). It is specifically contemplated that a modified/variant protein or polypeptide may be altered with respect to one activity or function yet retain a wild-type activity or function in other respects, such as immunogenicity.

[0110] Where a protein is specifically mentioned herein, it is in general a reference to a native (wild-type) or recombinant protein. The protein may be isolated directly from the organism of which it is native, produced by recombinant DNA/exogenous expression methods, or produced by solid-phase peptide synthesis (SPPS) or other in vitro methods. In particular embodiments, there are isolated nucleic acid segments and recombinant vectors incorporating nucleic acid sequences that encode a polypeptide (e.g., an antibody or fragment thereof). The term "recombinant" may be used in conjunction with a polypeptide or the name of a specific polypeptide, and this generally refers to a polypeptide produced from a nucleic acid molecule that has been manipulated in vitro or that is a replication product of such a molecule.

[0111] In certain embodiments the size of a protein or polypeptide (wild-type or modified) may comprise, but is not limited to, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2250, 2500 amino acid residues or greater, and any range derivable therein, or derivative of a corresponding amino sequence described or referenced herein. It is contemplated that polypeptides may be mutated by truncation, rendering them shorter than their corresponding wild-type form, also, they might be altered by fusing or conjugating a heterologous protein or polypeptide sequence with a particular function (e.g., for targeting or localization, for enhanced immunogenicity, for purification purposes, etc.). As used herein, the term "domain" refers to any distinct functional or structural unit of a protein or polypeptide, and generally refers to a sequence of amino acids with a structure or function recognizable by one skilled in the art.

[0112] The polypeptides, proteins, or polynucleotides encoding such polypeptides or proteins of the disclosure may include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,

22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 (or any derivable range therein) or more variant amino acids or nucleic acid substitutions or be at least 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% (or any derivable range therein) similar, identical, or homologous with at least, or at most 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 300, 400, 500, 550, 1000 or more contiguous amino acids or nucleic acids, or any range derivable therein, of a wild-type protein or gene.

[0113] The nucleotide as well as the protein, polypeptide, and peptide sequences for various genes have been previously disclosed, and may be found in the recognized computerized databases. Two commonly used databases are the National Center for Biotechnology Information's Genbank and GenPept databases (on the World Wide Web at ncbi.nlm.nih.gov/) and The Universal Protein Resource (UniProt; on the World Wide Web at uniprot.org). The coding regions for these genes may be assayed and/or amplified and/or expressed using the techniques disclosed herein or as would be known to those of ordinary skill in the art.

1. Variant Polypeptides

[0114] Amino acid sequence assayed by the methods of the disclosure can be substitutional, insertional, or deletion variants. A variation in a polypeptide of the disclosure may affect 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31,

32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more non-contiguous or contiguous amino acids of the protein or polypeptide, as compared to wild-type. A variant can comprise an amino acid sequence that is at least 50%, 60%, 70%, 80%, or 90%, including all values and ranges there between, identical to any sequence provided or referenced herein. A variant can include 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more substitute amino acids.

[0115] Deletion variants typically lack one or more residues of the native or wild type protein. Individual residues can be deleted or a number of contiguous amino acids can be deleted. A stop codon may be introduced (by substitution or insertion) into an encoding nucleic acid sequence to generate a truncated protein.

[0116] Insertional mutants typically involve the addition of amino acid residues at a non-terminal point in the polypeptide. This may include the insertion of one or more amino acid residues. Terminal additions may also be generated and can include fusion proteins which are multimers or concatemers of one or more peptides or polypeptides described or referenced herein.

[0117] Substitutional variants typically contain the exchange of one amino acid for another at one or more sites within the protein or polypeptide, and may be designed to modulate one or more properties of the polypeptide, with or without the loss of other functions or properties. Substitutions may be conservative, that is, one amino acid is replaced with one of similar chemical properties. "Conservative amino acid substitutions" may involve exchange of a member of one amino acid class with another member of the same class. Conservative substitutions are well known in the art and include, for example, the changes of: alanine to serine; arginine to lysine; asparagine to glutamine or histidine; aspartate to glutamate; cysteine to serine; glutamine to asparagine; glutamate to aspartate; glycine to proline; histidine to asparagine or glutamine; isoleucine to leucine or valine; leucine to valine or isoleucine; lysine to arginine; methionine to leucine or isoleucine; phenylalanine to tyrosine, leucine or methionine; serine to threonine; threonine to serine; tryptophan to tyrosine; tyrosine to tryptophan or phenylalanine; and valine to isoleucine or leucine. Conservative amino acid substitutions may encompass non-naturally occurring amino acid residues, which are typically incorporated by chemical peptide synthesis rather than by synthesis in biological systems. These include peptidomimetics or other reversed or inverted forms of amino acid moieties.

[0118] Alternatively, substitutions may be "non-conservative", such that a function or activity of the polypeptide is affected. Non-conservative changes typically involve substituting an amino acid residue with one that is chemically dissimilar, such as a polar or charged amino acid for a nonpolar or uncharged amino acid, and vice versa. Non-conservative substitutions may involve the exchange of a member of one of the amino acid classes for a member from another class.

EXAMPLES

[0119] The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

I. Example 1: LIG1 deletion predicts chemotherapy resistance, chromosomal instability, and poor prognosis in triple negative breast cancer

[0120] Proteogenomic analysis of neoadjuvant chemotherapy resistance in triple negative breast cancer reveals a complex landscape of response associates, including a 19q13.31-33 somatic deletion affecting lagging-strand DNA synthesis genes that associates with lack of pathological complete response, carboplatin-selective chemotherapy resistance, poor prognosis and chromosomal instability in multiple cancer types.

[0121] Sporadic Triple Negative Breast Cancer (TNBC) represents 10-15% of breast cancers (BC) worldwide. Cytotoxic chemotherapy remains the standard of care and adjuvant PD1 therapy is not biomarker guided. Pathological complete response (pCR) often not achieved and portends poor survival (PMID: 28135148). Predictive markers for individual chemotherapy agents has proved elusive and the use of carboplatin remains particularly controversial.

[0122] Microscaled proteogenomics (MPG) (PMID: 31988290) is a discovery technique that integrates DNA and RNA sequencing with mass spectrometry based proteomics. MPG was applied to snap-frozen TNBC clinical trial core needle biopsies obtained before treatment with carboplatin and docetaxel (PMID: 34173924, BCM NCT02547987; WashU: NCT02124902). Proteogenomic

profiling included whole exome sequencing (WES), RNA-seq, and tandem mass tag-based proteomics and phosphoproteomics. Clinical endpoints were pathological complete response (pCR) and residual cancer burden (RCB). Standard non-parametric statistical tests were employed to identify proteogenomic analytes associated with these endpoints.

[0123] Poorly understood recurrent copy number aberrations (CNA) are a pathogenomic feature of TNBC. Integrated proteogenomic analysis was applied to identify CNA with concordant changes at the mRNA and protein levels associated with pCR status. Genes located within a recurrent interstitial deletion at chromosomal location 19q13.31-33 were the most significantly down-regulated at mRNA and protein level in chemotherapy resistant cases. 19q13.31-33 encodes multiple DNA damage response (DDR) genes; however, only LIG1, a DNA ligase required for lagging strand synthesis and DNA repair, showed concordant changes at both the mRNA and protein level. In multiple independent TNBC data sets, LIG1 deletion was associated lack of pCR and poor metastasis-free survival. Additionally, in the BrighTNess trial (PMID: 29501363) lower LIG1 mRNA levels were associated with increased chemotherapy resistance (pCR and Residual cancer burden 0-III; p=0.0008 and 0.003 respectively). In PDX-derived short-term cultures (PMID: 27641504) and PDXs treated with docetaxel or carboplatin, a specific association of carboplatin resistance with LIG1 deletion was observed. LIG1 depleted tumors did not harbor elevated scores for homologous recombination defect signature (COSMIC mutational signature 3), suggesting LIG1 loss is an orthogonal pathway for TNBC pathogenesis. The high chromosomal instability index in LIG1 deletion tumors in the TNBC study was robustly reproduced in multiple datasets (including TCGA-Breast $p<0.0001$; Metastatic breast cancer project $p<0.0001$). LIG1 copy number deletion was also associated with poor progression free survival, and high chromosomal instability in multiple other cancers (including TCGA-UCEC (HR=2.23), TCGA-HNSC (HR=1.46), TCGA-PRAD (HR=2.07), TCGA- COAD (HR=1.75) and TCGA-KIRP (HR=4.00)).

[0124] Deletion of LIG1 is associated with chromosomal instability in TNBC and specifically occurs in tumors without evidence for defects in homologous recombination. The role of LIG1 deletion as a negative predictive marker for carboplatin, and a discussion of how LIG1 deletion causes chromosomal instability and tumorigenesis (PMID: 12124343) is encompassed herein.

[0125] Microscaled proteogenomics was deployed to probe the molecular basis for differential response to neoadjuvant carboplatin & taxotere combination chemotherapy for triple negative

breast cancer (TNBC). Proteomic analyses of pretreatment biopsies uniquely revealed that metabolic pathways including oxidative phosphorylation, fatty acid metabolism and glycolysis were resistance associated. Proteogenomic analyses of somatic copy number aberrations identified a resistance-associated 8q21.3 gain and 19q13.31-33 deletion where LIG1, POLD1 and XRCC1 are located. In orthogonal datasets, LIG1 (DNA ligase I involved in lagging strand synthesis) gene deletion and/or low mRNA expression were associated with lack of pathological complete response and poor prognosis in TNBC, as well as selective carboplatin-resistance in TNBC patient-derived xenograft models. Low expression or LIG1 loss was also associated with higher chromosomal instability index (CIN) and poor prognosis in other cancer types, demonstrating that deletion of lagging strand synthesis components has broad clinical significance.

II. Example 2: Proteogenomic overview of TNBC.

[0126] OCT-embedded snap-frozen core needle biopsies were accrued from consented patients with TNBC (70% Caucasian, 27% African American, and 3% other ethnicities). Samples from 59 patients taken before treatment had >25% tumor content. For 16 patients, an additional sample was obtained 48 to 72 hours after initiating chemotherapy. All patients received six cycles of neoadjuvant carboplatin and docetaxel combination chemotherapy (NCT02547987, NCT02124902). A REMARK diagram provides information on sample flow into different analytical pipelines (**FIG.1A**). Using the previously described BioTEXT and MiProt methods [7], samples were processed on a cryotome to produce both 50um sections for analyte extraction and 5um sections (after each set of six 50 um sections) to document tumor content. 50um sections were alternated into three different analyte preparation approaches to ensure even representation of analytes from different layers in the biopsy. Multianalyte extraction allowed for paired normal/tumor DNA exome sequencing (100X), RNA sequencing and quantitative, multiplexed (TMT) mass spectrometry (MS)-based proteomics/phosphoproteomics as previously described [7] (**FIG.1B**).

[0127] As expected, RNA-based PAM50 [9] and TNBCtype [10] intrinsic subtyping of baseline samples revealed "Basal-like" to be the most prevalent subtype (75%) although all Lehman TNBC subtypes were represented (**FIG.8A**). Sample-level mRNA and protein correlations deteriorated in samples with an average tumor content (TC) below 45% (**FIG.8B**), and these samples were excluded from further bioinformatics analyses. As previously reported [7], TMT11

plexes were linked using a common reference which served as a common denominator for calculating protein and phosphosite ratios. As a measure of excellent data quality across plexes, common references across multiple plexes showed very strong correlations (**FIG.8C**). For each qualified sample, DNA, RNA and protein level information were available for an average of 10,500 genes (**FIG.1C**). Overall, phosphoproteomic analysis quantified ~27,000 phosphorylation sites in ~5,000 distinct phosphoproteins (**FIG.1C**). Median mRNA and protein correlation per gene in qualified samples was 0.37, similar to previous CPTAC proteogenomic studies [11] (**FIG.8D**). Genes with significant negative RNA-protein correlations were enriched in KEGG pathways involved in cellular respiration and amino-acid and lipid metabolism. Genes with lower correlations were enriched in pathways containing large protein complexes such as spliceosome, replication, transcription, and pyrimidine metabolism pathways (**FIG.8D**). Protein data significantly outperformed RNA data in co-expression-based gene function prediction (**FIG.8E**), consistent with previous observations [7, 12-14].

[0128] Analytes were also examined in a comparison of 14 samples collected after 48-72 hours of treatment to baseline samples matched to the same patients (only 13 pairs had RNA data). Whereas genes downregulated in the on-treatment samples were enriched with immune-related pathways at both RNA and protein levels, cell cycle and metabolic pathways, with the exception of glycolysis, were significantly upregulated after treatment specifically at the protein level (**Fig. 1D**). Induction of DNA replication and repair pathways linked to the cell cycle likely arises in response to genotoxic stress triggered by chemotherapy [15], an observation also reflected in the phosphosite data (**FIG.8F**). Sets of phosphosites induced by perturbation with UV light, nocodazole, and ionizing radiation, as well as CDK1, CDK2, and ATM targets were induced by chemotherapy in this cohort (**FIG.8F**).

III. Example 3: Molecular associates of response to chemotherapy

[0129] Study endpoints were pCR and residual cancer burden (RCB). RCB ranged from 0-3 (referred to as 0, I, II, III), with 0 indicating pCR and III indicating extensive residual disease [16]. PAM50, TNBCtype, and race lacked association with pCR, as did other cohort-specific metadata (**FIG.8A**). At the genomic level, the homologous recombination defect (HRD) score (Signature 3 [17]) did not associate with chemotherapy response (**FIG.8G**). Lack of association with pCR was also observed for mutations in homologous recombination (HR) genes

implicated in TNBC progression, *BRCA1/2* or *PALB2,* and for the COSMIC mutation signature[18, 19] score indicative of defective HR (Signature 3). An elevated mismatch repair defect (MMRD) signature (Signature 6 [17]) score was associated with RCB II or III (p=0.03, **FIG.8G**). Genes with higher protein but not mRNA expression in non-pCR cases showed enrichment for MSigDB Hallmark metabolic pathways including oxidative phosphorylation, fatty-acid metabolism, and adipogenesis. In contrast, genes with both higher mRNA and protein expression in pCR cases showed enrichment for immune (interferon alpha and gamma response) and cell cycle (G2M checkpoint and E2F and MYC target) pathways (**FIG.1E**). Enrichment analysis of differential phosphorylation sites (PTM-SEA [20]) showed elevated phosphoproteome-driven signatures for treatment with inhibitors targeting microtubule depolymerization or with DNA damage-inducing agents such as etoposide, hydroxyurea and ionizing radiation in pCR tumors (**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 [21, 22]. Consistent with elevated cell cycle pathways observed for pCR samples in the RNA and protein data, CDK1, 2, and 7 and CDC7 target phosphosites were also significantly higher in pCR. A notable exception was CDK4, for which target sites were significantly higher in non-pCR.

IV. Example 4: Immune activity and E2F targets are higher at baseline in tumors that achieve pCR

[0130] Elevated expression of interferon response pathways in pCR is consistent with previous reports highlighting that immune infiltration is predictive of response to chemotherapy [23]. A heatmap detailing baseline proteogenomic features associated with the immune microenvironment is provided (**FIG.2A).** Protein-derived immune stimulatory scores, previously found to be well correlated with immune infiltration [11], as well as PD-L1 RNA, protein, and phosphorylation levels were significantly higher in pCR tumors (**Fig. 2B**). Both PD-L1 protein and phosphoprotein levels significantly correlated with PD-L1 IHC levels (**Fig. 2D-E**). Representative IHC images for high and low PD-L1 staining are shown in **FIG.9A** and **B**, respectively. Non-synonymous 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 is not a strong determinant of immune infiltration in this data set. Rather, the

immune score was significantly anti-correlated with chromosomal instability (CIN) (Spearman Rho=-0.61, p=6.2c-6; **FIG.2C**).

[0131] Chemotherapy selectively targets actively dividing cells, but despite enrichment of cell cycle- related pathways in tumors achieving pCR (**FIG. 1E**), the CDK4 target set was one of the most significantly downregulated phosphosite sets in pCR-associated tumors. This prompted a sample-wise investigation of cell-cycle proteogenomic features (**FIG. 2F**). In general, RNA- and protein-based multi-gene proliferation scores (MGPS), single-sample GSEA (ssGSEA) scores for cell cycle-related pathways, PTM-Signature Enrichment Analysis (PTM-SEA) scores for cyclin dependent kinases, and other cell cycle features were high in pCR but variable in non- pCR. However, a subset of non-pCR samples with elevated CDK4 activity also had high levels of Rb phosphorylation (highlighted by box in Fig. 2F). While not significantly differential by Wilcoxon rank sum test (p=0.12), a one-tailed Kolmogorov-Smirnov test indicated the Rb phosphorylation distribution was shifted towards higher levels in non-pCR than in pCR tumors, most likely because of this subset (p=0.047, **FIG. 9C**). This suggests that CDK4/6 inhibitors may be effective for this subset, a hypothesis suggested by others [23-25] and consistent with retrospective analysis of TNBC cell lines treated with CDK4/6 inhibitors and platinum-based chemotherapeutic agents (**FIG.9D**).

V. Example 5: Metabolic pathway proteins are higher at baseline in non-pCR tumors

[0132] Despite some heterogeneity amongst individual patient tumors, ssGSEA scores for differential metabolic pathways including oxidative phosphorylation, adipogenesis, fatty acid metabolism, and glycolysis were significantly higher in pre-treatment tumors that did not achieve pCR (RCB0) (**FIG.3A**). As with the on-treatment vs. baseline comparison (**FIG. 1D**), metabolic pathway enrichment was specific to proteomic data, which may account for the lack of these associations in prior mRNA-based analyses. Metabolism-related proteins significantly associated with resistance included many mitochondrial proteins such as those directly involved in the tricarboxylic acid (TCA) cycle (ACO2, FH, MDH2, SUCLG1, SUCLG2, PDP1, DLAT), the electron transport chain (SDHC, UQCR10), fatty acid metabolism (CRAT, ACADS, ACAT1, DECR1, ECHS1, HADHB), and amino acid catabolism (ALDH6A1, HMGCL, DBT, BCKDHB). While the majority of metabolic genes contributing to these pathway scores showed association with non-pCR only at the protein level, a subset of metabolic genes from the differential Hallmark

pathways (**FIG.3B**). RNA expression of this subset of 29 genes was tested against pCR status in an independent cohort of patients who received combination treatment with carboplatin and docetaxel in the BrighTNESS TNBC trial [26]. Geometric mean scores derived from mRNA data were significantly higher for non-pCR cases as compared to pCR cases (Wilcoxon rank sum test, p=0.003; **FIG.10A**) in this large data set (N=359 including patients from trial arms A and B). Additionally, significantly higher scores were observed in higher RCB categories (ANOVA test, p= 0.0024; **FIG.3C**).

VI. Example 6: Functional copy number aberrations are associated with treatment response

[0133] The somatic mutation landscape of TNBC is dominated by chromosomal instability (CIN), which produces recurrent copy number aberrations [27]. The overall copy number landscape in this dataset was typical for TNBCs, and alignment of the differential expression results for RNA and protein by position in the genome demonstrated chromosomal regions where the differential expression of multiple genes is determined by CNA (**FIG. 12A**). In an effort to determine whether response correlated with particular CNA locations, GSEA was utilized to statistically evaluate relationships between cytoband location and upregulated or downregulated gene expression (**FIG.4A**). Individual genes were subsequently examined to identify examples that were up- or down- regulated with respect to pCR status. This unbiased prioritization effort demonstrated that expression of gene products from the 8q21.3 and 19q13.31-33 cytobands were elevated and suppressed, respectively, in non-pCR relative to pCR tumors (**FIG.4B**), suggesting the possibility of functionally relevant copy number gain of 8q21.3 and copy number loss of 19q13.31-33 specifically in the non-pCR tumors. Four genes located at 8q21.3, *RMDN1*, *CPNE3*, *DECR1*, and *OTUB6B*, showed higher mRNA and protein expression in non-pCR tumors (**FIG.11B**). In addition, *RIPK2*, which mediates metastasis in patients with advanced breast cancer [27], 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, PPP5C, BCL3,* and *NOSIP*, showed lower mRNA and protein expression in non-pCR tumors (**FIG.4B**). The downregulated genes and proteins were also confirmed to be frequently lost (GISTIC = -1, potential single copy loss) at the CNV level in a subset of non-pCR tumors (**FIG.11C**). Hallmark pathway GSEA analysis of the genes on cytoband 19q13.31-33 showed enrichment in the DNA

damage repair (DDR) pathway at the mRNA level, with LIG1, XRCC1, POLD1 and ERCC1 comprising the leading-edge genes (**FIG.4C**). LIG1 showed the strongest association with treatment response at the protein level, followed by POLD1 (**FIG.4C**).

[0134] To determine whether these observations are reproducible in other data sets, the association of LIG1, XRCC1, POLD1 and ERCC1 with pCR and RCB was evaluated at the mRNA level in the BrighTNess trial (combining the two carboplatin and paclitaxel containing arms to mimic the docetaxel and carboplatin treatment in the discovery data sets) [26] (LIG1, POLD1, and XRCC1 were confirmed to be significantly down-regulated in baseline tumor samples from patients who experienced residual disease (**FIG.4D**). RNA expression levels for LIG1 and XRCC1 were also significantly associated with poor metastasis-free survival in the TNBC subset of another chemotherapy-treated TNBC patient cohort (Hatzis dataset [28]) (**FIG.4E**). In summary, out of four DDR genes on cytoband 19q13.3, LIG1 was the most consistently associated with chemotherapy resistance and poor metastasis-free survival in multiple independent datasets (**FIG.4D, 4E, 11D**).

VII. Example 7: Molecular features of TNBCs harboring LIG1 deletion

[0135] To further investigate the association of LIG1 deletion and/or reduced expression with tumor pathophysiologic features, additional analysis of the TNBC proteogenomic data was undertaken. Low *LIG1* copy number levels (GISTIC = -1) were observed in eight of thirty-one (~26%) tumors without pCR (**FIG.5**). LIG1 copy number log ratios are strongly and positively correlated with both LIG1 RNA (Pearson, R = 0.67, p = 2.8e–06) and protein (R = 0.55, p = 8.2e–05) levels (**FIG.12A- B**). At the genomic level, COSMIC Signature 3 (proposed etiology: defective in DNA repair by homologous recombination) was significantly lower in tumors with LIG1 loss (**FIG.5**). Tumors harboring LIG1 loss also exhibited higher chromosomal instability (CIN) scores. Consistent with the aforementioned anti-correlation between CIN and the immune stimulatory modulator score, tumors with LIG1 loss had lower immune stimulatory (IM) scores than LIG1 WT/copy number gain (GISTIC >= 0) non-pCR (p>0.05) and pCR (p=0.01) tumors (**FIG.12C**). The IL33 pathway was significantly upregulated in tumors without LIG1 loss, further indicating a less active immune environment associated with LIG1 loss (**FIG.12D-E**). Tumors with LIG1 loss also had significantly higher protein-based proliferation scores (pMGPS, **FIG.5**) as well as upregulation of CDK1/2 activity (**FIG.12D**) by Single-sample PTM-SEA [20],

indicating increased cell cycle activity. These results collectively suggest that loss of LIG1 is associated with a reduced immune response, higher proliferation rates and chemotherapy resistance. Signatures of EGFR (gefitinib) and PI3K (wortmannin) inhibition were significantly enriched in LIG1 loss tumors (**FIG.12D-E**), suggesting that LIG1 loss tumors have lower EGFR and PI3K signaling and therefore unlikely to be sensitive to treatment with this class of agent.

VIII. Example 8: LIG1 copy number loss is associated with poor progression-free survival and CIN across multiple cancer types

[0136] Gene copy number analysis of tumors characterized by TCGA demonstrates that LIG1 single copy loss is present in other cancer types. In the TCGA "pan-cancer" data set, LIG1 heterozygous loss was associated with poor progression-free survival (PFS) (**FIG.7A**, $p<0.0001$), significantly higher CIN (Fraction genome altered, **FIG.7B**), and lower signature 3 score (suggesting proficient homologous recombination, **FIG.7C**). Cancer types driving these relationships include endometrial carcinoma (HR=2.23, p=0.02), head & neck squamous cell carcinoma (HR=1.46, p=0.03), prostate adenocarcinoma (HR=2.07, p=0.02), colon adenocarcinoma (HR=1.75, p=0.03) and renal papillary cell carcinoma (HR=4, p= 0.0001) (**FIG.7D**). Despite marginal association of PFS with LIG1 loss in testicular germ cell tumors (TGCT), the seminoma subtype, which demonstrates exquisite sensitivity to carboplatin [29], displayed no cases of LIG1 loss (**FIG.14A**). Higher CIN (fraction genome altered) was also observed in several individual cancers as shown in **FIG.7E** (TCGA cohort) and S7B (CPTAC cohort), though not in the CPTAC COAD dataset. These analyses suggest that the predictive utility of LIG1 loss for response to platinum-based chemotherapy may not be exclusive to TNBC.

IX. Example 9: Discussion of Embodiments

[0137] While the molecular basis for the variable chemotherapy response in TNBC has been extensively studied at the DNA and RNA levels, investigations have yet to produce useful tools to guide the use of chemotherapy in TNBC. The proteogenomic analysis disclosed herein provides at least five new perspectives. First, integrated proteogenomic information provides extensive information on the immune microenvironment that could be used as alternatives to PDL1 IHC. Second, oxidative phosphorylation and fatty acid metabolism were strongly associated with chemotherapy resistance but uniquely at the proteomic level, indicating models to predict

chemotherapy response could be strengthened by the inclusion of proteomic data. Third, G2M checkpoint components, E2F regulation and MYC target pathways provide an additional rich resource of biomarkers for predictive model-building. Fourth, proteomics can help prioritize genomic observations in tumors with large numbers of chromosomal aberrations such as TNBC demonstrated by the identification of LIG1 as chemotherapy resistance marker. Finally, proteomic analyses of preclinical models suggest that LIG1 loss is a selective biomarker for carboplatin resistance. The use of carboplatin in TNBC remains controversial and could be potentially avoided in LIG1-depleted tumors. Further investigation of LIG1/XRCC1/POLD1 expression and 19q13.31-33 deletion status in randomized trials of carboplatin should therefore be pursued in an effort to reduce carboplatin use.

[0138] There are already ample prior mechanistic studies of LIG1 that suggest that reduced function destabilizes the genome and promotes cancer formation. LIG1 encodes an ATP-dependent DNA ligase that seals DNA nicks during replication, recombination, and a variety of DNA damage responses [30]. In the ligase family, it was first discovered as the main enzyme responsible for ligating Okazaki fragments together during lagging-strand synthesis at the replication fork during S-phase [31-33]. LIG1 also ligates single-stranded or double-stranded DNA breaks in various DNA damage repair pathways including long-patch base-excision repair, nucleotide-excision repair, and alternative non-homologous end-joining repair [34, 35]. 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 [36]. Insufficient LIG1 activity results in the accumulation of replication intermediates that cause single-stranded (SSB) and double-stranded breaks (DSB) [37, 38], ultimately leading to reduced genome integrity. In transgenic systems hypomorphic LIG1 alleles are associated with high susceptibility to cancer formation [39]. However, the relevance of LIG1 hypomorphic alleles can be questioned because single copy loss is most likely observed, with the remaining LIG1 allele intact, in sporadic TNBC and other cancers, raising the question of whether single copy loss produces sufficient functional deficiency to generate a phenotype. One obvious possibility is that co-deletion of POLD1 and XRCC1 produce a compound deficiency phenotype since all three genes serve lagging strand synthesis. XRCC1 is particularly noteworthy because LIG3/XRCC1 provides a backup pathway for LIG1 during DNA repair and lagging strand synthesis but is known to be less efficient than LIG1 [40]. An investigation of pathways that compensate for LIG1 loss

may also help address the question of carboplatin insensitivity. LIG1 loss was found to be orthogonal to HRD signatures. Consequently, LIG1 cells may still be paradoxically proficient at double strand break repair. The PDX study hints at this, as the model derived from the baseline study had a BRCA2 frameshift mutation and no LIG1 loss, but the BRCA2 mutation was undetectable in the both the PDX from treatment-resistant residual disease at surgery and the PDX from a liver metastasis, both of had gained a 19q13.3/LIG1 hemizygous deletion. It remains unclear why LIG1 loss is so strongly associated with chromosomal instability across many cancer types. We have not demonstrated cause and effect in this paper, but clearly cells that enter mitosis with unrepaired strand gaps are at risk for chromosomal breakage and illicit chromosomal fusion events, likely driving aneuploidy.

[0139] In conclusion, the embodiments of the proteogenomic analysis of TNBC demonstrates the considerable value of this approach.

X. Example 10: CDK2 and Chk2 inhibitor to treat *LIG1* depleted tumors

[0140] Tumors with *LIG1* depletion also had significantly higher protein-based proliferation scores as well as upregulation of CDK1 and 2 activity (**FIG. 15**, [1]) by single-sample PTM-SEA [8], indicating increased cell cycle activity. These results collectively indicate that loss of *LIG1* is associated with a reduced immune response, higher proliferation rates, genomic instabilities, and chemotherapy resistance. Signatures of EGFR (gefitinib) and PI3K (wortmannin) inhibition were significantly enriched in *LIG1* depleted tumors, indicating that *LIG1* depletion tumors have lower EGFR and PI3K signaling and are therefore unlikely to be sensitive to treatment with these kinase inhibitors. In contrast, CDK1 and CDK2 are targets in *LIG1* depleted tumors, in some embodiments. Because targeting CDK1 in cancer patients may lead to toxicity in normal proliferating cells[9], one can focus on CDK2 inhibition. Similarly, preliminary data also indicates elevated levels of Chk2 targets in *LIG1* depleted TNBCs, and one can characterize Chk2 inhibitors in this setting.

XI. Example 11: Link between LIG1 loss and PARP inhibitor sensitivity

[0141] PARP inhibition causes synthetic lethality in breast cancers with germline *BRCA1* and *BRCA2* mutations [10, 11] and is used routinely in clinical practice for metastatic breast

cancer[12]. Meta-analysis of CRISPR screens aimed at identifying determinants of PARPi sensitivity (Zimmerman et al., 2018) showed that LIG1 was one of the thirteen high confidence determinants for which loss showed consistent association with PARPi sensitivity in three different cell line models (**FIG. 16A**).

[0142] Other studies have reported that inhibition of LIG1 activates PARP and promotes PAR synthesis at DNA replication sites[13],[14]. In cells treated with PARPi, un-ligated Okazaki fragments persist, potentially leading to replication conflicts in the subsequent S phase[13]. The effect was determined of PARP inhibition with Olaparib on the MB-MDA-468 TNBC cell line with and without LIG1 suppression by siRNA. LIG1 suppression further reduced tumor cell viability upon Olaparib treatment (**FIG. 16B**).

[0143] Also profiled was a highly relevant set of 3 PDXs derived from biopsies of the same patient at three different timepoints: pre-treatment (WHIM68 -- *LIG1* high), at surgery after chemotherapy (WHIM74 – *LIG1* heterogenous) and from a liver metastatic site (WHIM75 – LIG1 low) using RNA, DNA, and proteomic profiling.

[0144] While bulk RNA and proteomic approaches work optimally, the inventors also analyzed *LIG1* in these 3 PDX models by FISH (**FIG. 17A**). Recapitulating the primary disease, the tumors were highly heterogenous although this heterogeneity was lower in the metastatic tumor that showed the greatest loss of LIG1 in the proteogenomics data (**FIG. 17A**). Next, the *LIG1* FISH assay was extended to 50 TNBC PDXs. Here, a subset of non-CRs had more *LIG1*-haploid cells (**FIG. 17B**). As BRCA1/2 mutated TNBC have been known to be sensitive to carboplatin, FISH results were re-analyzed by excluding BRCA1/2 mutant cases as confounding factors. Interestingly, non-CR PDXs have a higher % of *LIG1*-haploid cells (FIG. 17C). This aligns well with patient data as *LIG1* loss cases exhibit efficient homologous recombination as reported previously.

XII. Example 12: Methods for Certain Embodiments

A. Clinical sample collection

[0145] Eligible patients for the two clinical trials (NCT02547987 (BCM), NCT02124902 (WashU) included pre or post-menopausal 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 IRB at both participating sites WashU and BCM and followed the Declaration of Helsinki and Good Clinical Practice guidelines. The protocol and informed consent documents were approved by WUSM and BCM. All patients were uniformly treated (without randomization or blinding) with neoadjuvant intravenous docetaxel 75 mg/m2 and carboplatin AUC 6 cycled every 21 days for 6 cycles, with granulocyte colony-stimulating factor support [8]. Research tumor biopsies for correlative studies were obtained at baseline prior to chemotherapy, and on cycle 1 day 3 (C1D3). On-treatment biopsy on C1D3 and biopsy at time of relapse were optional. Details of the clinical cohort have been recently published [8]. Treatment response information was provided by clinical teams associated with these trials and residual cancer burden was calculated using Residual Cancer Burden calculator at the website of M.D. Anderson Cancer Center).

B. Immunohistochemistry

[0146] For immunohistochemistry (IHC) cut tissue sections (5mm) on charged glass slides were baked for 10-12 hours at 58C in a dry slide incubator, deparaffinized in xylene and rehydrated via an ethanol step gradient. For CD3, heat-induced antigen retrieval steps were performed at pH 9.0. The primary antibody was incubated at room temperature for 1 hour [CD3 (polyclonal, Dako, 1:100)] followed by a standard chromogenic staining protocol with the Envision Polymer-HRP anti-mouse/3,30diaminobenzidine (DAB, Dako) process. Slides were counterstained in Harris hematoxylin. PDL1 staining was performed using the PD-L1 IHC 22C3 pharmDx kit per regulated protocol on the Dako Autolink-48 platform (SK006; Agilent). Pathology slide scoring was performed using established professional guidelines for TNBC, when appropriate. All immunohistochemistry results were evaluated against positive and negative tissue controls.

C. Genomic analysis

[0147] **Whole exome sequencing (WES):** Tumor DNA was extracted from fresh-frozen biopsies and matched leukocyte germline DNA from blood samples. WES data was generated for 59 unique baseline DNA samples using the Illumina platform. For this, paired-end libraries were constructed as described previously[41] with the following modifications. Samples were barcoded at ligation step using Illumina unique dual barcodes adapters (Cat# 20022370) and were amplified 6-8 cycles using the Library Amplification Ready-mix containing KAPA HiFi DNA Polymerase

(Kapa Biosystems, Inc). For capture enrichment, libraries were pooled in equimolar ratios in groups of 10 and were hybridized in solution to the HGSC VCRome 2.1 design [42]. To this design, exome coverage across >3,500 clinically relevant genes that are previously <20X (~2.72Mb) was supplemented. Enriched libraries were sequenced on the NovaSeq 6000 instrument using the S4 reagent kit (300 cycles) to generate 2x150bp paired-end reads. For these 110134 samples, on average, 11.01 Gb of unique sequence data was generated with 97.3% of the bases in the exome design coverage to 20x read depth or greater.

[0148] RNA-Seq data: Transciptome data was 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 (PMID: 25360585). Briefly, poly-A+ mRNA was extracted from 1 μg total RNA, followed by fragmentation and first strand cDNA synthesis. The resultant cDNA was end-repaired, A-tailed and ligated with Illumina Dual barcode adapters. Libraries were sequenced on NovaSeq 6000 instruments using the S4 reagent kit (300 cycles) to generate 2x150bp paired-end reads. Between 59.96 and 112.62M total reads were generated for these 60 samples. The average strand-specificity and rRNA rate was 97.04% and 1.79% respectively. The transcripts for 22868 to 27856 genes were detected in these samples.

[0149] The paired-end reads were mapped to the human genome version GRCh38.d1.vd1 (From GDC) using STAR-2.7.1a. Gene expression estimation was performed using RSEM-1.3.1, and RSEM and FPKM values were upper-quartile normalized. Unless otherwise noted, gene median-centered log2-transformed RSEM values were used for the analyses presented here.

[0150] **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 are Strelka2, Mutect2, CARNAC, and Pindel (v 0.2.5b9). Variants reported by these tools were filtered using GATK VariantFiltration (v 3.8.0) with parameters window 35, cluster 3, FS > 30.0, and QD < 2.0. We kept SNVs called by any 2 callers among Strelka2, Mutect2 and indels called by any 2 callers among Mutect2, CARNAC, and Pindel. To merge SNVs and indels, a 10X coverage cutoff was applied for both tumor and normal sequence depth. We also filtered somatic SNVs and indels by a minimal variant allele frequency (VAF) of 0.02. Then annovar (v 04.16.2018) was used to annotate remaining variants. Somatic mutations were called comparing tumor DNA against matched blood normal DNA using WES data. Similarly, germline mutations were called by comparing normal WES against the reference genome. Hg19.UCSC.add_miR.140312.refgene was

used to map the copy number information to genes. COSMIC mutational signature scores for every sample were estimated using deconstructSigs [43].

[0151] For somatic copy number alteration analysis, bam files were processed by the CopywriteR package [44] to derive log2 tumor-to-normal copy number ratios, and the circular binary segmentation (CBS) algorithm [45] implemented in the CopywriteR package was used for the copy number segmentation, with the default parameters. Chromosomal instability for each chromosome in each sample was inferred from the segmentation data using a weighted-sum approach in which the absolute values of the log2 ratios of all segments within a chromosome were weighted by the segment length and summed up [13]. The genome-wide chromosome instability index (CIN) was derived by adding up the instability scores for all 22 autosomes in each sample. Next, GISTIC2[46] was used to retrieve gene-level copy number values and call significant copy number alterations in the cohort. A threshold of +/-0.3 was applied to log2 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 reflects 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.

[0152] For the Pancancer analysis GISTIC value +/- 2 exceed the high-level thresholds for amplifications/deep deletions, and those with +/- 1 exceed the low-level thresholds but not the high-level thresholds. The low-level thresholds are just the 'ampthresh' and 'delthresh' noise threshold input values to GISTIC (typically 0.1 or 0.3) and are the same for every thresholds.

D. Proteomics data generation and analysis

[0153] **Proteomic sample preparation**: Samples were prepared for proteomic analysis as described in a previous microscaled proteogenomic study [7]. Protein lysates in 8 M urea were reduced with 1 mM DTT for 45 min, then alkylated with iodoacetamide (IAA) for 45 min protected from light. Before digestion, urea was diluted to a final concentration of 2 M with 50 mM Tris-HCl pH 8.5. Protein lysates were then treated with endopeptidase LysC (Promega) at a 1:40

enzyme mass to BCA-estimated protein mass ratio, followed by overnight treatment with trypsin (Promega) at a 1:30 ratio. Both digestions were performed at 25C. To desalt the digestion mixture, peptides were acidified to 1% formic acid (FA), and purified with a 50 mg tC18 Sep-Pak cartridge (Waters). Peptides were eluted with 50% acetonitrile/0.1% FA. Peptide concentration was measured by Nanodrop (Thermo) using 280 nm absorbance. To evaluate peptide sample quality and digestion efficiency, 0.5 ug peptides from each sample were run on a nLC1200 coupled to Q-Exactive + LC- MS instrumentation (Thermo). The remainder of the eluted peptides were then snap-frozen and dried with a vacuum centrifuge.

[0154] **TMT labeling:** A total of 30 ug peptides in 100 uL 50 mM HEPES, pH 8.5, were labeled with 240 ug TMT reagent for an 8:1 TMT:peptide ratio and incubated at 25C for 1 hour. Before quenching excess TMT, labeling was assessed by stage-tip desalting 1 uL per sample and running 0.5 ug peptides on a 30 min gradient. Mixing ratios were assessed by pooling and stage-tip desalting 2 uL from each sample, and running 0.5 ug peptides on a 110 min gradient. All possible TMT site labeling was required to be over 97%, and mixing ratios were balanced to be within +/- 15% when compared to the common references in each 11-plex (see Experimental design for proteomics and phospho-proteomics). Excess TMT reagent was quenched by incubating with 5 uL 5% hydroxylamine (Sigma) for 15 min. Samples within each plex were combined according to the ratios determined by the mixing controls to achieve equal sample representation within each plex. 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.

[0155] **Experimental design for proteomics and phospho-proteomics**: 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 per each of the 60 patients. The second common reference ("prospective BRCA CR") was from a previous large cohort breast cancer proteomics study [11]. Protein and phosphosite expression were reported as the TMT intensity ratio between each sample and the common references within each plex. For analysis of clinical core samples, eight TMT 11-plexes each contained peptides from 9 core needle biopsies in the first 9 channels. If available, paired pre- and post-treatment tumor samples from a patient were grouped within the same 11-plex. As a quality control measure,

protein and phosphopeptide ratios were obtained between prospective BRCA common reference and the core common reference, and the results are shown in FIG.8C.

E. Basic reverse phase fractionation and phosphoenrichment

[0156] For basic reverse phase fractionation, ~330 ug of peptides were dissolved in 500 uL of 5 mM ammonium formate and 5% acetonitrile. Peptides were fractionated using an offline Agilent 1260 LC with a 30 cm long, 2.1 mm inner diameter C18 column, running at 200 uL per minute. Peptides were separated across 72 fractions, which were then concatenated into 18 fractions. 2 ug peptides from each of the 18 fractions were set aside for whole proteome analysis, of which 0.67 ug was analyzed from each fraction. For phospho-enrichment, the 18 fractions were further concatenated to 6 fractions (~50 ug per fraction).

[0157] Phosphopeptides were enriched using Fe3+ immobilized metal affinity chromatography (IMAC) as previously described [7]. Ni-NTA (Qiagen) beads were washed three times with HPLC grade water followed by incubation with 100 mM EDTA (Sigma) to strip nickel from the beads. Beads were washed three more times with HPLC grade water and incubated in 10 mM FeCl3 (Sigma) for 30 min. Fe3+-loaded beads were resuspended in a 1:1:1 solution of methanol, acetonitrile, and 0.01% acetic acid in water. Dried peptides were resuspended to a final volume of 500 uL with 50% acetonitrile and 0.1% trifluoroacetic acid (TFA) followed by 100% acetonitrile and 0.1% TFA for a final concentration of 80% acetonitrile. Each reconstituted fraction was added to 20 uL of 50% bead slurry and incubated while rotating end-over-end at room temperature for 30 min. Beads were then spun down and supernatant removed. The beads were transferred to a conditioned C18 stage- tip in 200 uL of 80% acetonitrile and 0.1% TFA. Phosphopeptides were eluted from the beads with 500 mM potassium phosphate, pH 7, onto the C18, washed with 1% formic acid, and eluted into an autosampler vial with 50% acetonitrile and 0.1% FA.

F. Proteomic data acquisition and processing

[0158] Proteome and phosphoproteome data acquisition was performed with a Proxeon nLC-1200 coupled to Thermo Lumos instrumentation. For proteomic analysis, peptides were run on a 110 min gradient with 86 min of effective gradient (6 to 30% buffer B containing 90% ACN and 0.1%FA). For phosphoproteomics analysis, two injections were run per fraction: a first injection

over a 90 min gradient with 70 min of effective gradient (6 to 30% buffer B containing 90% ACN and 0.1% FA), and a second injection over a 140 min gradient with 120 min of effective gradient (6 to 35% buffer B containing 90% ACN and 0.1% FA). The acquisition parameters for the 110 min proteome and 90 min phosphoproteome methods are: MS1 resolution = 60,000, MS1 injection time = 50 s, MS2 resolution = 50,000, MS2 injection time = 105 s, AGC 6e4. For the 140 min phosphoproteome injections, the MS2 injection time was increased to 250 s, and the MS2 AGC decreased to 5e4. A cycle time of 2 s was used for all methods.

[0159] Raw files were searched against the human (clinical cores) or humanRefSeq protein databases complemented with 553 small-open reading frames (smORFs) and common contaminants (Human: RefSeq.20171003_Human_ucsc_hg38_cpdb_mito_259contamsnr_553smORFS.fasta), using Spectrum Mill (Broad Institute) as previously described in detail [7]. Lysine was required to be fully TMT-labeled while N-termini were allowed to be under-labeled for TMT quantification using the "Full, Lys only" option. Carbamidomethylation of cysteines was set as a fixed modification, and N-terminal protein acetylation, oxidation of methionine (Met-ox), de-amidation of asparagine, hydroxylation of proline, TMT over-labeling of serine, threonine, and tyrosine, and cyclization of peptide N-terminal glutamine and carbamidomethylated cysteine to pyroglutamic acid (pyroGlu) and pyro-carbamidomethyl cysteine were set as variable modifications. For phosphoproteome analysis, phosphorylation of serine, threonine, and tyrosine were allowed as additional variable modifications, while de-amidation of asparagine and hydroxylation of proline were disabled. Trypsin Allow P was specified as the proteolytic enzyme with up to 4 missed cleavage sites allowed. For proteome analysis, the allowed precursor mass shift range was −18 to 326 Da. For phosphoproteome analysis, the range was −18 to 272 Da, to allow for up to 3 phosphorylations and 2 Met-ox per peptide. Precursor and product mass tolerances were set to ±20 ppm. For core biopsy analysis, peptide FDR limits were set to 0.8% for charge states 2-4 and 0.4% for charge states 5-6, and for PDX analysis peptide FDR limits were set to 0.6% for 2-4 and 0.3% for 5-6, employing a target-decoy approach using reversed protein sequences (PMID: 20013364). For PDX analyses, the subgroup-specific (SGS) option in Spectrum Mill was enabled as previously described [7] . This allowed better dissection of proteins of human and mouse origin. If specific evidence for both human and mouse peptides from an orthologous protein were observed, then peptides that cannot distinguish the two (shared) were ignored. However, the

peptides shared between species were retained if there was specific evidence for only one of the species, thus yielding a protein group with a single subgroup attributed to only the single species consistent with the specific peptides.

G. Quantification, normalization and filtering of proteomics data

[0160] Before 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 PSMs 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.

H. PDX proteomics data generation and analysis

[0161] For the PDX experiment, cryopulverized PDX tumor tissues were lysed and digested as described above. 50ug peptides were dissolved in 200ul 50 mM HEPES, pH 8.5 and labeled with 400ug of TMT reagent. TMT sample generation, basic reverse fractionation and proteomic analysis was performed identical to that of clinical core biopsies. Raw files were searched against the human and mouse (PDX samples) UniProt protein databases complemented with 553 small-open reading frames (smORFs) and common contaminants (Human and mouse: UniProt.human.mouse.20171228.RIsnrNF.553smORFs.264contams.fasta) using Spectrum Mill . For PDX analyses, the subgroup-specific (SGS) option in Spectrum Mill was enabled as previously described [7]. This allowed better dissection of proteins of human and mouse origin. If specific evidence for both human and mouse peptides from an orthologous protein were observed, then peptides that cannot distinguish the two (shared) were ignored. However, the peptides shared between species were retained if there was specific evidence for only one of the species, thus yielding a protein group with a single subgroup attributed to only the single species consistent with the specific peptides.

I. Data QC and differential expression and pathway enrichment analysis

[0162] Samples with estimated tumor content below 45% were entirely removed from the dataset due to lack of RNA to protein correlation in these samples (FIG. 8B). 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 and between samples with LIG1 loss (GISTIC = -1) and those without loss (LIG1 WT/Gain ,GISTIC >= 0). WebGestaltR (PMID: 31114916) and PTM-SEA[20] were used to identify MSigDB Hallmark pathways (gene level data) and 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. Additionally, ssGSEA R package [47, 48] was applied to data from three "omes" and scores for Hallmark pathways were obtained for individual samples). 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 (PMID: 31114916) and PTM-SEA were 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 (normalized enrichment scores for kinase target PTM sets).

J. Functional prediction based on gene co-expression

[0163] Co-expression network construction using mRNA and protein expression data and network-based gene function prediction for KEGG pathways were performed as previously described in Wang et al, 2017 [12] using OmicsEV (on the bzhanglab webpage of github https://github.com/bzhanglab/OmicsEV).

K. Multi-gene proliferation and immune profiling scores

[0164] RNA-based multi-gene proliferation scores (MGPS) were calculated as described previously [11, 49] by averaging the gene-centered log2 RSEM data for all genes previously characterized as cycle-regulated [50] 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 Benjamini-Hochberg fdr correction). Immune profile and microenvironment scores were inferred from the FPKM version of the RNA-seq data using ESTIMATE (R package; PMID: 24113773), Cibersort (webtool (https://cibersort.stanford.edu/) run in absolute mode; PMID: 25822800), and xCell (webtool (https://xcell.ucsf.edu/); PMID: 29141660). Protein-based immune modulator scores were calculated as described previously [11] by averaging log2 TMT ratios for expert curated sets of immune modulators belonging to three categories: immune stimulatory, immune inhibitory, and human leukocyte antigen (HLA) [51].

L. Immunoblotting

[0165] Fresh frozen WHIM68, WHIM74, and WHIM75 tumors were cryopulverized (Covaris CP02) then lysed in RIPA buffer (10 mM Tris-Cl [pH=8.0], 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 140 mM NaCl, 10 mM PMSF, 1 ug/mL pepstatin, phosSTOP phosphatase inhibitor table [Roche], and cOmplete EDTA-free protease inhibitor tablet [Roche]) for 20 min. on ice. 30 µg of clarified lysates were heated at 90°C for 10 min. before loading on 4-12% SDS-PAGE gels (Invitrogen) and electroblotted onto PVDF membranes (Bio- Rad). Membranes were blocked in 5% milk for 1h hour at RT followed by incubation of primary antibodies at 4°C overnight: LIG1 (cat# 18051-1-AP, ProteinTech, 1:1000), POLD1 (cat# 15656- 1-AP, ProteinTeech, 1:1000), XRCC1 (cat# ab134056, Abcam, 1:1000). GAPDH (cat# sc-47724, Santa Cruz Biotechnology, 1:4000) was used as a loading control. Proteins were visualized by incubation with Cytiva Amersham ECL Select Western Blot Detection reagent (Fisher Scientific) and images were captured with a ChemiDoc Imaging System (Bio-Rad).

M. Validation using DepMap

[0166] Global-TMT measurements for RB1 and response profiles to approved drugs from the Cancer Response Therapeutics Response Portal (CTRP), Genomics of Drug Sensitivity in Cancer (GDSC), and Profiling Relative Inhibition Simultaneously in Mixtures (PRISM) drug response datasets for cancer cell lines were retrieved from the DepMap resource (www.depmap.org). TNBC cell lines were selected based on ERneg_HER2neg lineage_sub_subtype for breast lineages from

sample information provided by DepMap. For TNBC cell lines, Pearson's correlation was calculated between RB1 protein abundance (log2 TMT ratio) and drug responses (AUC). P-values < 0.05 were considered significant.

N. Data availability

[0167] The genomics and transcriptomics data has been deposited in the dbGAP database under the accession code phs002505.v1. All genomics and proteomic raw data associated with this study will be available via dbGAP and CPTAC portal (https://proteomics.cancer.gov/data-portal) respectively upon publication.

* * *

[0168] All of the methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

1. Dent R, Trudeau M, Pritchard KI, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 2007;13(15 Pt 1):4429-34.
2. Staaf J, Glodzik D, Bosch A, et al. Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. Nat Med 2019;25(10):1526-1533.
3. Symmans WF, Wei C, Gould R, et al. Long-Term Prognostic Risk After Neoadjuvant Chemotherapy Associated With Residual Cancer Burden and Breast Cancer Subtype. J Clin Oncol 2017;35(10):1049-1060.

4. Masuda N, Lee SJ, Ohtani S, et al. Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy. N Engl J Med 2017;376(22):2147-2159.
5. Schmid P, Cortes J, Pusztai L, et al. Pembrolizumab for Early Triple-Negative Breast Cancer. N Engl J Med 2020;382(9):810-821.
6. Fountzila E, Ignatiadis M. Neoadjuvant immunotherapy in breast cancer: a paradigm shift? Ecancermedicalscience 2020;14:1147.
7. Satpathy S, Jaehnig EJ, Krug K, et al. Microscaled proteogenomic methods for precision oncology. Nat Commun 2020;11(1):532.
8. Ademuyiwa FO, Chen I, Luo J, et al. Immunogenomic profiling and pathological response results from a clinical trial of docetaxel and carboplatin in triple-negative breast cancer. Breast Cancer Res Treat 2021;189(1):187-202.
9. Chia SK, Bramwell VH, Tu D, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. Clin Cancer Res 2012;18(16):4465-72.
10. Lehmann BD, Bauer JA, Chen X, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011;121(7):2750-67.
11. Krug K, Jaehnig EJ, Satpathy S, et al. Proteogenomic Landscape of Breast Cancer Tumorigenesis and Targeted Therapy. Cell 2020;183(5):1436-1456 e31.
12. Wang J, Ma Z, Carr SA, et al. Proteome Profiling Outperforms Transcriptome Profiling for Coexpression Based Gene Function Prediction. Mol Cell Proteomics 2017;16(1):121-134.
13. Vasaikar S, Huang C, Wang X, et al. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell 2019;177(4):1035-1049 e19.
14. Huang C, Chen L, Savage SR, et al. Proteogenomic insights into the biology and treatment of HPV- negative head and neck squamous cell carcinoma. Cancer Cell 2021;39(3):361-379 e16.
15. Christmann M, Kaina B. Transcriptional regulation of human DNA repair genes following genotoxic stress: trigger mechanisms, inducible responses and genotoxic adaptation. Nucleic Acids Res 2013;41(18):8403-20.
16. Symmans WF, Peintinger F, Hatzis C, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. J Clin Oncol 2007;25(28):4414-22.

17. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature 2013;500(7463):415-21.
18. Nik-Zainal S, Alexandrov LB, Wedge DC, et al. Mutational processes molding the genomes of 21 breast cancers. Cell 2012;149(5):979-93.
19. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Deciphering signatures of mutational processes operative in human cancer. Cell Rep 2013;3(1):246-59.
20. Krug K, Mertins P, Zhang B, et al. A Curated Resource for Phosphosite-specific Signature Analysis. Mol Cell Proteomics 2019;18(3):576-593.
21. Hubaux R, Thu KL, Vucic EA, et al. Microtubule affinity-regulating kinase 2 is associated with DNA damage response and cisplatin resistance in non-small cell lung cancer. Int J Cancer 2015;137(9):2072-82.
22. Wei X, Xu L, Jeddo SF, et al. MARK2 enhances cisplatin resistance via PI3K/AKT/NF-kappaB signaling pathway in osteosarcoma cells. Am J Transl Res 2020;12(5):1807-1823.
23. O'Meara T, Safonov A, Casadevall D, et al. Immune microenvironment of triple-negative breast cancer in African-American and Caucasian women. Breast Cancer Res Treat 2019;175(1):247-259.
24. Iorio F, Knijnenburg TA, Vis DJ, et al. A Landscape of Pharmacogenomic Interactions in Cancer. Cell 2016;166(3):740-754.
25. Picco G, Chen ED, Alonso LG, et al. Functional linkage of gene fusions to cancer cell fitness assessed by pharmacological and CRISPR-Cas9 screening. Nat Commun 2019;10(1):2198.
26. Loibl S, O'Shaughnessy J, Untch M, et al. Addition of the PARP inhibitor veliparib plus carboplatin or carboplatin alone to standard neoadjuvant chemotherapy in triple-negative breast cancer (BrighTNess): a randomised, phase 3 trial. Lancet Oncol 2018;19(4):497-509.
27. Smid M, Hoes M, Sieuwerts AM, et al. Patterns and incidence of chromosomal instability and their prognostic relevance in breast cancer subtypes. Breast Cancer Res Treat 2011;128(1):23-30.
28. Hatzis C, Pusztai L, Valero V, et al. A genomic predictor of response and survival following taxane- anthracycline chemotherapy for invasive breast cancer. JAMA 2011;305(18):1873-81.

29. Alifrangis C, Sharma A, Chowdhury S, et al. Single-agent carboplatin AUC10 in metastatic seminoma: A multi-centre UK study of 216 patients. Eur J Cancer 2020; 10.1016/j.ejca.2020.08.031.
30. Howes TR, Tomkinson AE. DNA ligase I, the replicative DNA ligase. Subcell Biochem 2012;62:327-41.
31. Waga S, Bauer G, Stillman B. Reconstitution of complete SV40 DNA replication with purified replication factors. J Biol Chem 1994;269(14):10923-34.
32. Soderhall S, Lindahl T. DNA ligases of eukaryotes. FEBS Lett 1976;67(1):1-8.
33. Petrini JH, Xiao Y, Weaver DT. DNA ligase I mediates essential functions in mammalian cells. Mol Cell Biol 1995;15(8):4303-8.
34. Paul-Konietzko K, Thomale J, Arakawa H, et al. DNA Ligases I and III Support Nucleotide Excision Repair in DT40 Cells with Similar Efficiency. Photochem Photobiol 2015;91(5):1173-80.
35. Pascucci B, Stucki M, Jonsson ZO, et al. Long patch base excision repair with purified human proteins. DNA ligase I as patch size mediator for DNA polymerases delta and epsilon. J Biol Chem 1999;274(47):33696-702.
36. Barnes DE, Tomkinson AE, Lehmann AR, et al. Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. Cell 1992;69(3):495-503.
37. Soza S, Leva V, Vago R, et al. DNA ligase I deficiency leads to replication-dependent DNA damage and impacts cell morphology without blocking cell cycle progression. Mol Cell Biol 2009;29(8):2032-41.
38. Harrison C, Ketchen AM, Redhead NJ, et al. Replication failure, genome instability, and increased cancer susceptibility in mice with a point mutation in the DNA ligase I gene. Cancer Res 2002;62(14):4065- 74.
39. Bentley D, Selfridge J, Millar JK, et al. DNA ligase I is required for fetal liver erythropoiesis but is not essential for mammalian cell viability. Nat Genet 1996;13(4):489-91.
40. Le Chalony C, Hoffschir F, Gauthier LR, et al. Partial complementation of a DNA ligase I deficiency by DNA ligase III and its impact on cell survival and telomere stability in mammalian cells. Cell Mol Life Sci 2012;69(17):2933-49.

41. Rokita JL, Rathi KS, Cardenas MF, et al. Genomic Profiling of Childhood Tumor Patient-Derived Xenograft Models to Enable Rational Clinical Trial Design. Cell Rep 2019;29(6):1675-1689 e9.
42. Bainbridge MN, Wang M, Wu Y, et al. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biol 2011;12(7):R68.
43. Rosenthal R, McGranahan N, Herrero J, et al. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol 2016;17:31.
44. Kuilman T, Velds A, Kemper K, et al. CopywriteR: DNA copy number detection from off-target sequence data. Genome Biol 2015;16:49.
45. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 2007;23(6):657-63.
46. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 2011;12(4):R41.
47. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009;462(7269):108-12.
48. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102(43):15545- 50.
49. Ellis MJ, Suman VJ, Hoog J, et al. Ki67 Proliferation Index as a Tool for Chemotherapy Decisions During and After Neoadjuvant Aromatase Inhibitor Treatment of Breast Cancer: Results From the American College of Surgeons Oncology Group Z1031 Trial (Alliance). J Clin Oncol 2017;35(10):1061- 1069.
50. Whitfield ML, Sherlock G, Saldanha AJ, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 2002;13(6):1977-2000.
51. Thorsson V, Gibbs DL, Brown SD, et al. The Immune Landscape of Cancer. Immunity 2019;51(2):411-412.

WHAT IS CLAIMED IS:

1. A method of determining a treatment therapy for an individual with cancer, comprising the step of administering a therapy other than a platinum-based drug and/or other than an inhibitor of EGFR or PI3K, or comprising the step of administering one or more CDK2 inhibitors and/or one or more PARP inhibitors, to an individual having gene copy number gain, high mRNA expression, and/or high protein level from one or more genes on cytoband 8q21.2 and/or has gene copy number deletion, low mRNA expression and/or low protein level from one or more genes on cytoband 19q13.3.

2. The method of claim 1, further defined as when a sample from the individual identifies gene deletion and/or low mRNA expression and/or low protein level of DNA ligase 1 (LIG1); DNA Polymerase Delta 1, Catalytic Subunit (POLD1); and/or X-ray repair cross-complementing protein 1 (XRCC1) with or without other genes on cytoband 19q13.3, the individual is administered a therapeutically effective amount of one or more therapies that is not a platinum-based drug.

3. The method of claim 1 or 2, further defined as when a sample from the individual identifies gene gain and/or high mRNA expression and/or high protein level of RIPK2, RMDN1, CPNE3, DECR1, and OTUB6B with or without other genes on cytoband 8q21.3, the individual is administered a therapeutically effective amount of one or more therapies that is not a platinum-based drug.

4. The method of claim 1, 2, or 3, wherein the platinum-based drug is carboplatin alone or in combination with docetaxel.

5. The method of any one of claims 1-4, wherein the individual has triple negative breast cancer, endometrial carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, colon adenocarcinoma and renal papillary cell carcinoma.

6. The method of any one of claims 1-5, wherein the sample is analyzed for LIG1 copy number, LIG1 mRNA expression and/or LIG1 protein level.

7. A method of prognosticating a metastasis outcome for an individual, comprising the step of identifying that an individual will have poor metastasis-free survival when the individual

has gene deletion and/or low mRNA expression and/or low protein level from one or more genes on cytoband 19q13.3.

8. The method of claim 7, wherein the one or more genes is LIG1, POLD1, and/or XRCC1.

9. A method of treating an individual for cancer, comprising the step of administering a therapy other than a platinum-based drug and/or other than an inhibitor of EGFR or PI3K, or comprising the step of administering one or more CDK2 inhibitors and/or a PARP inhibitors,to an individual having gene deletion and/or low mRNA expression and/or low protein level of DNA ligase 1 (LIG1).

10. The method of claim 9, wherein a sample from the individual is analyzed for gene deletion and/or low mRNA expression and/or low protein level of LIG1.

11. A method of detecting LIG1 deletion or reduced expression in an individual, comprising the steps of:

    a. combining into multiple vessels a plurality of sections from different regions of the biological sample, said biological sample optionally being a biopsy;

    b. isolating DNA from the plurality of sections in a first vessel, isolating RNA from the plurality of sections in a second vessel, and isolating protein from the plurality of sections in a third vessel, alternatively wherein DNA and protein are isolated from the same vessel; and

    c. analyzing the isolated DNA, analyzing the isolated RNA, and/or analyzing the isolated protein to determine the status of at least one marker present in cells in the biopsy, wherein the marker comprises LIG1,

    wherein the combining step (a) provides a mixture of cells and/or tissues from different regions of the biopsy for the plurality of sections.

12. The method of claim 11, further comprising the step of sectioning the biopsy, and wherein the sectioning and combining step (a) are performed concurrently or separately.

13. The method of claim 11 or 12, wherein a plurality of sections from a fourth vessel are used for sample quality analysis.

14. The method of any one of claims 11-13, wherein the plurality of section for at least one of the 1st, 2nd, and/or 3rd vessels comprises nonadjacent sections from the different regions.

15. The method of any one of claims 11-14, wherein the protein is denatured protein.

16. The method of any one of claims 11-15, wherein analyzing the isolated protein further comprises determining the phosphorylation status of LIG1.

17. The method of claim 16, wherein the phosphorylation status of the cancer markers is determined by mass spectrometry.

18. The method of any one of claims 11-17, wherein analyzing the isolated DNA, analyzing the isolated RNA, and analyzing the isolated protein determines the status of LIG1 in the biopsy.

19. The method of claim 18, wherein when a status determination that LIG1 is deleted or reduced expression, the method further comprises the step of administering to the individual one or more therapies, wherein the therapies do not comprise a platinum-based drug and/or other than an inhibitor of EGFR or PI3K, or comprising the step of administering one or more CDK2 inhibitors, Chk2 inhibitors, and/or a PARP inhibitors,.

20. The method of claim 19, wherein the platinum-based drug comprises carboplatin.

21. The method of any one of claims 11-20, wherein the individual has, or is suspected of having, a cancer.

22. A method of determining the susceptibility of an individual having, or suspected of having, a cancer to a cancer treatment, comprising the step of subjecting a biological sample from the individual to the method of any one of claims 9-21, wherein *LIG1* status is not depleted and the cancer treatment comprises a platinum-based drug and/or an inhibitor of EGFR or PI3K.

23. The method of claim 22, further comprising administering or not administering carboplatin to the individual.

24. The method of claim 22 or 23, wherein the platinum-based drug comprises carboplatin.2qa

25. The method of any one of claims 22-24, wherein the cancer is a breast cancer.

26. A method of treating an individual having, or suspected of having, a cancer using a biological sample from the individual comprising the steps of:

    a. combining into multiple vessels a plurality of sections from different regions of the biological sample, said biological sample optionally being a biopsy;

    b. isolating DNA from the plurality of sections in a first vessel, isolating RNA from the plurality of sections in a second vessel, and isolating protein from the plurality of sections in a third vessel;

    c. analyzing the isolated DNA, analyzing the isolated RNA, and/or analyzing the isolated protein to determine the status of at least one marker present in cells in the biopsy, wherein the marker comprises LIG1;

    d. determining the LIG1 status in the biological sample; and

    e. administering one or more CDK2 inhibitors, Chk2 inhibitors, and/or PARP inhibitors, or not administering a platinum-based drug, inhibitor of EGFR, and/or inhibitor of PI3K, to the individual based on the LIG1 status,

    wherein the combining step (a) provides a mixture of cells and/or tissues from different regions of the biopsy for the plurality of sections.

27. The method of claim 26, wherein step (e) comprises administering the platinum-based drug when the LIG1 status is determined to not be deleted in the sample.

28. The method of claim 26, wherein step (e) comprises not administering the platinum-based drug when the LIG1 status is determined to be deleted in the sample.

29. The method of any one of claims 26-28, wherein the platinum-based drug comprises carboplatin.

30. The method of any one of claims 26-29, further comprising the step of sectioning the biopsy, and wherein the sectioning and combining step (a) are performed concurrently or separately.

31. The method of any one of claims 26-30, wherein a plurality of sections from a fourth vessel are used for sample quality analysis.

32. The method of any one of claims 26-31, wherein the plurality of section for at least one of the 1st, 2nd, and/or 3rd vessels comprises nonadjacent sections from the different regions.

33. The method of any one of claims 26-32, wherein the protein is denatured protein.

34. The method of any one of claims 26-33, wherein analyzing the isolated protein further comprises determining the phosphorylation status of LIG1.

35. The method of claim 34, wherein the phosphorylation status of the cancer markers is determined by mass spectrometry.

36. A method of detecting LIG1 deletion in a sample comprising the steps of analyzing genomic DNA from the sample at a locus comprising LIG1 and/or measuring expression of LIG1 mRNA and/or protein in the sample.

37. A method of determining an individual's susceptibility to a platinum-based drug, inhibitor of EGFR, inhibitor of PI3K, CDK2 inhibitor, Chk2 inhibitor, or PARP inhibitor, comprising the steps of analyzing genomic DNA from the sample at a locus comprising LIG1 and/or measuring expression of LIG1 mRNA and/or protein in the sample.

38. The method of claim 37, wherein the individual has, or is suspected of having a cancer.

39. The method of claim 38, wherein the cancer comprises a breast cancer.

40. The method of claim 39, wherein the breast cancer comprises a triple negative breast cancer.

41. The method of any one of claims 37-40, further comprising the step of administering or not administering the platinum-based drug,

42. The method of any one of claims 37-41, wherein the platinum-based drug comprises carboplatin.

43. A method of detecting LIG1 deletion or reduced expression in an individual, comprising the steps of:

(a) combining into multiple vessels a plurality of sections from different regions of a biological sample, said biological sample optionally being a biopsy;

(b) isolating DNA from the plurality of sections in a first vessel, isolating RNA from the plurality of sections in a second vessel, and isolating protein from the plurality of sections in a third vessel or from the plurality of sections in the first vessel;

wherein the combining step (a) provides a mixture of cells and/or tissues from different regions of the biopsy for the plurality of sections.

44. The method of claim 43, wherein the plurality of section for at least one of the 1st, 2nd, and/or 3rd vessels comprises nonadjacent sections from the different regions.

FIG. 1A

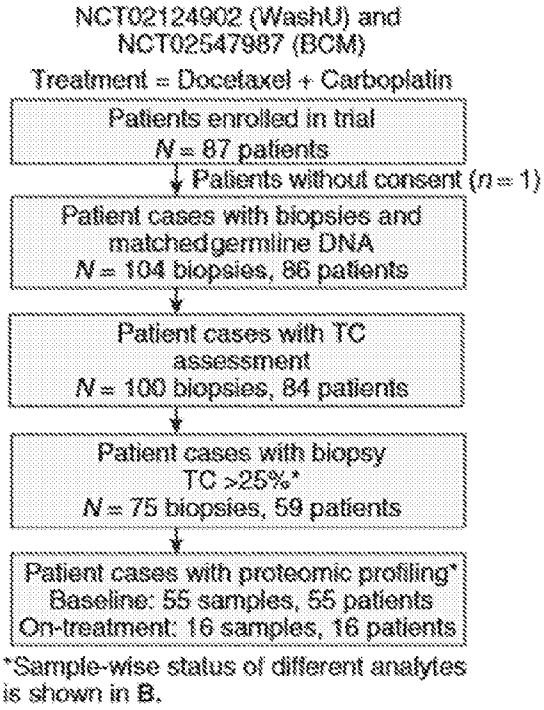
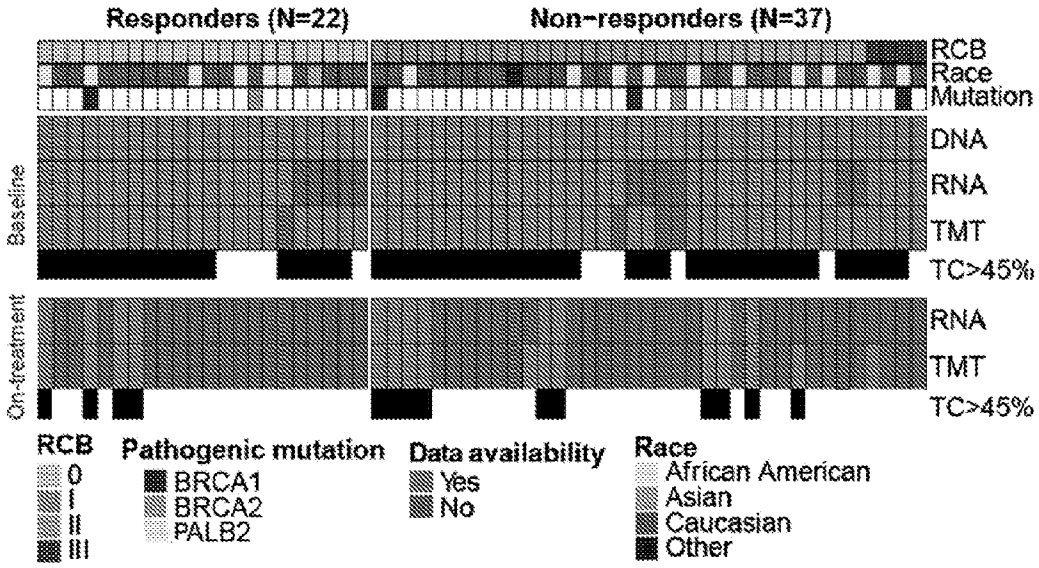


FIG. 1B



FIGs. 1A-1B

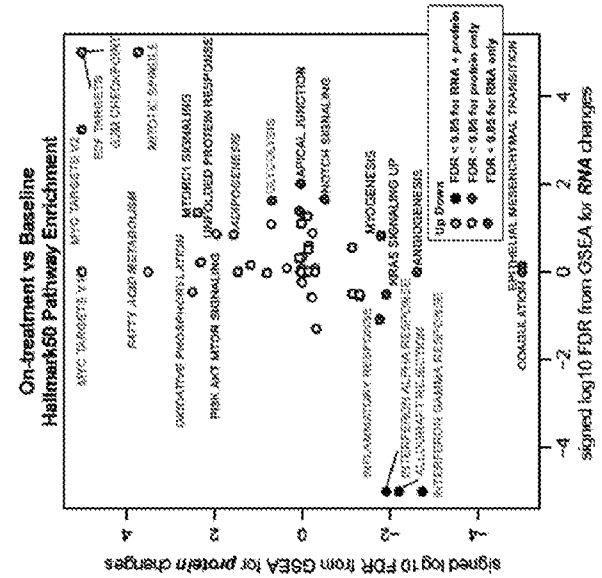


FIG. 1D

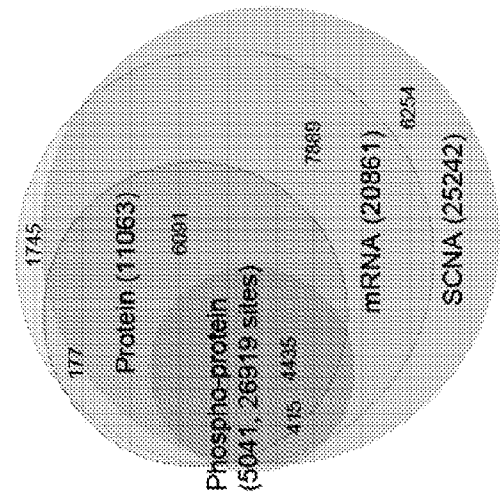


FIG. 1C

FIGs. 1C-1D

FIG. 1F

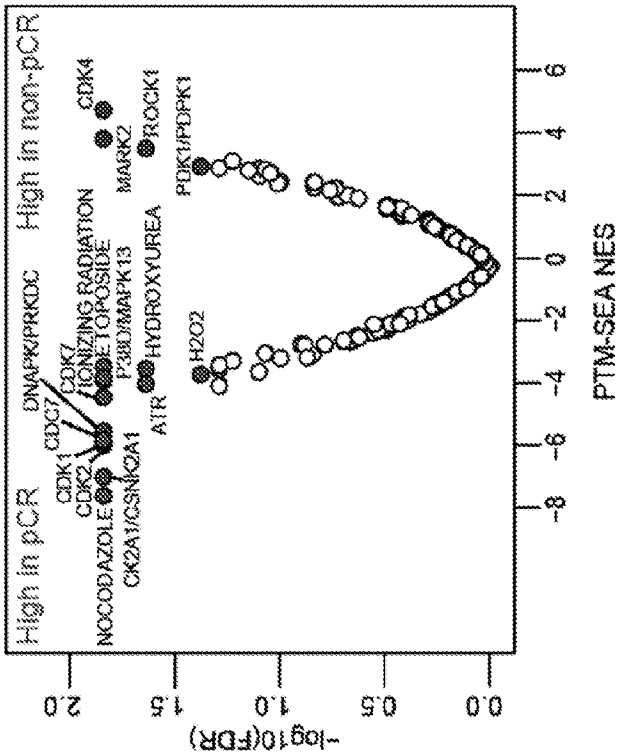
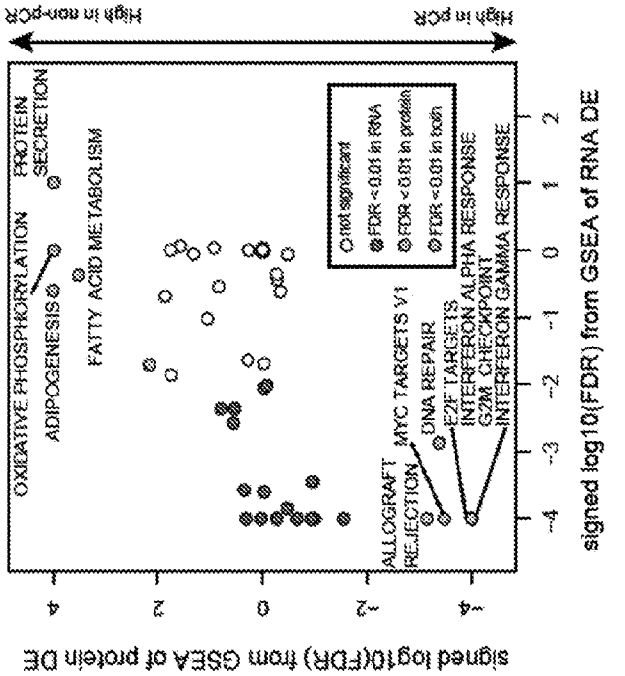


FIG. 1E



FIGs. 1E-1F

FIG. 2A

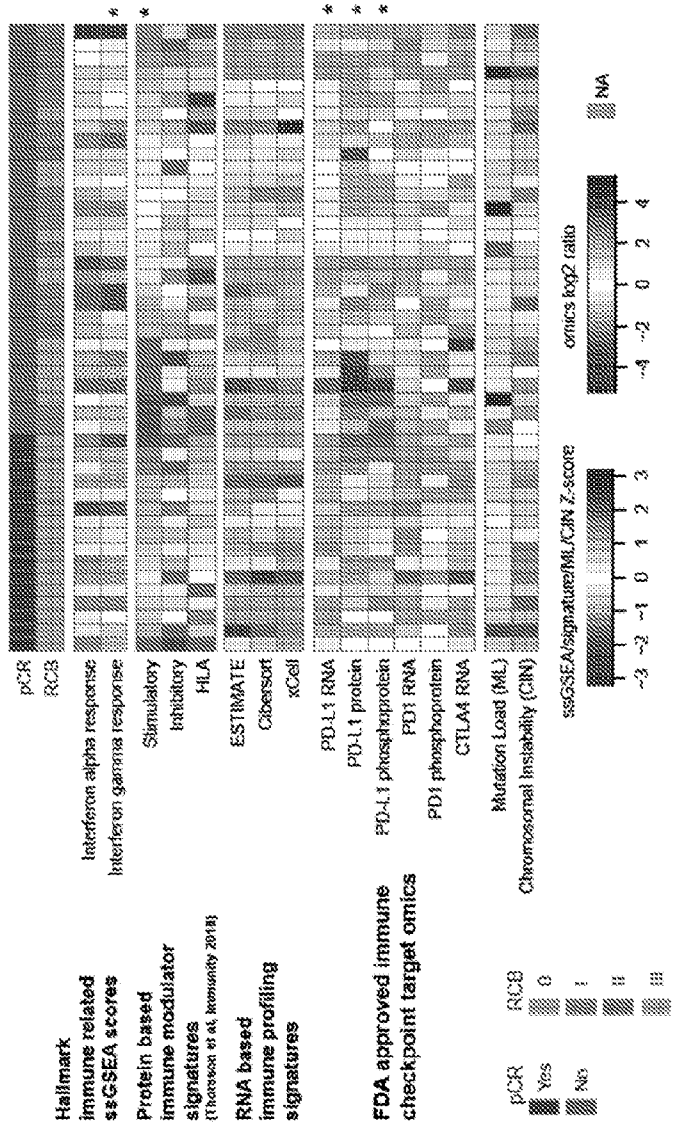


FIG. 2B

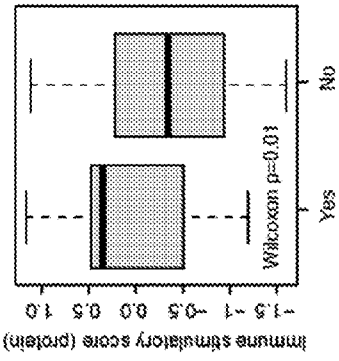
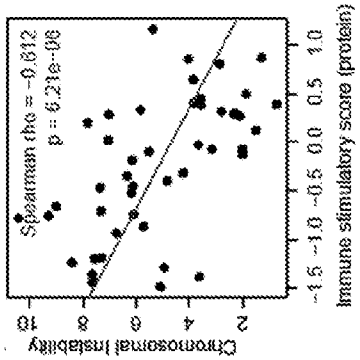


FIG. 2C



FIGs. 2A-2C

FIG. 2D

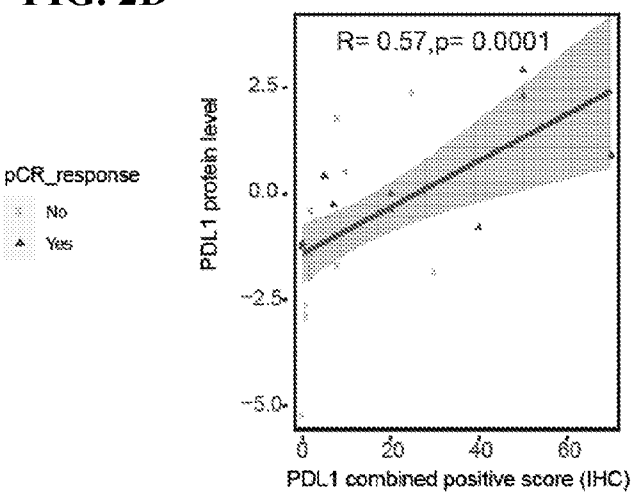


FIG. 2E

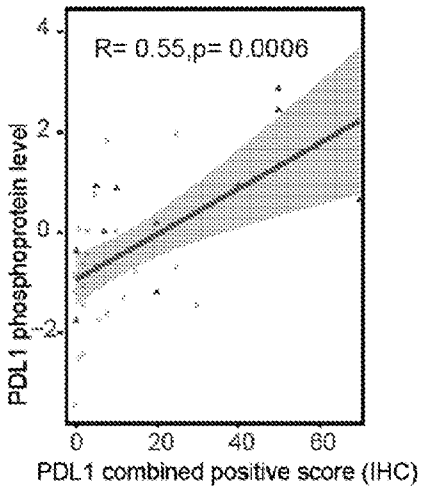
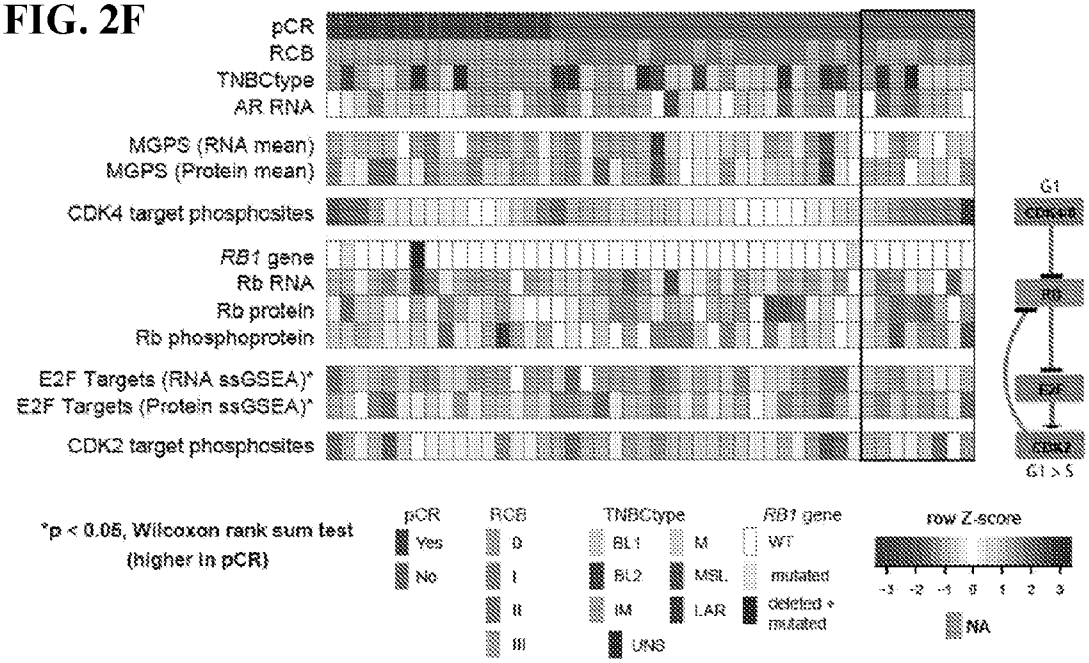


FIG. 2F



FIGs. 2D-2F

FIG. 3A

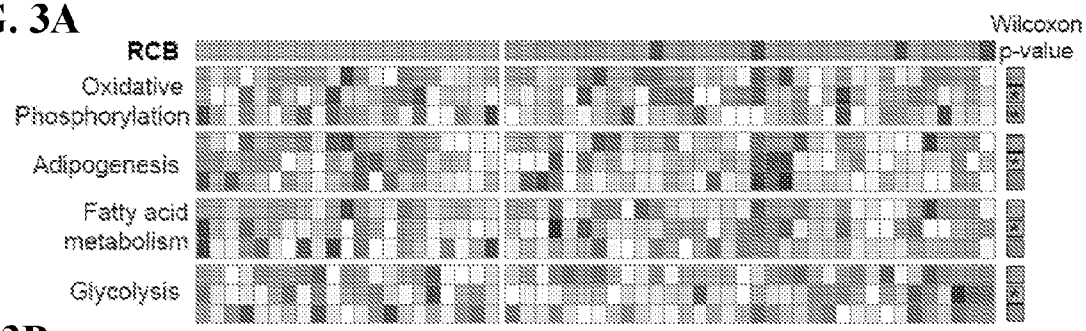


FIG. 3B

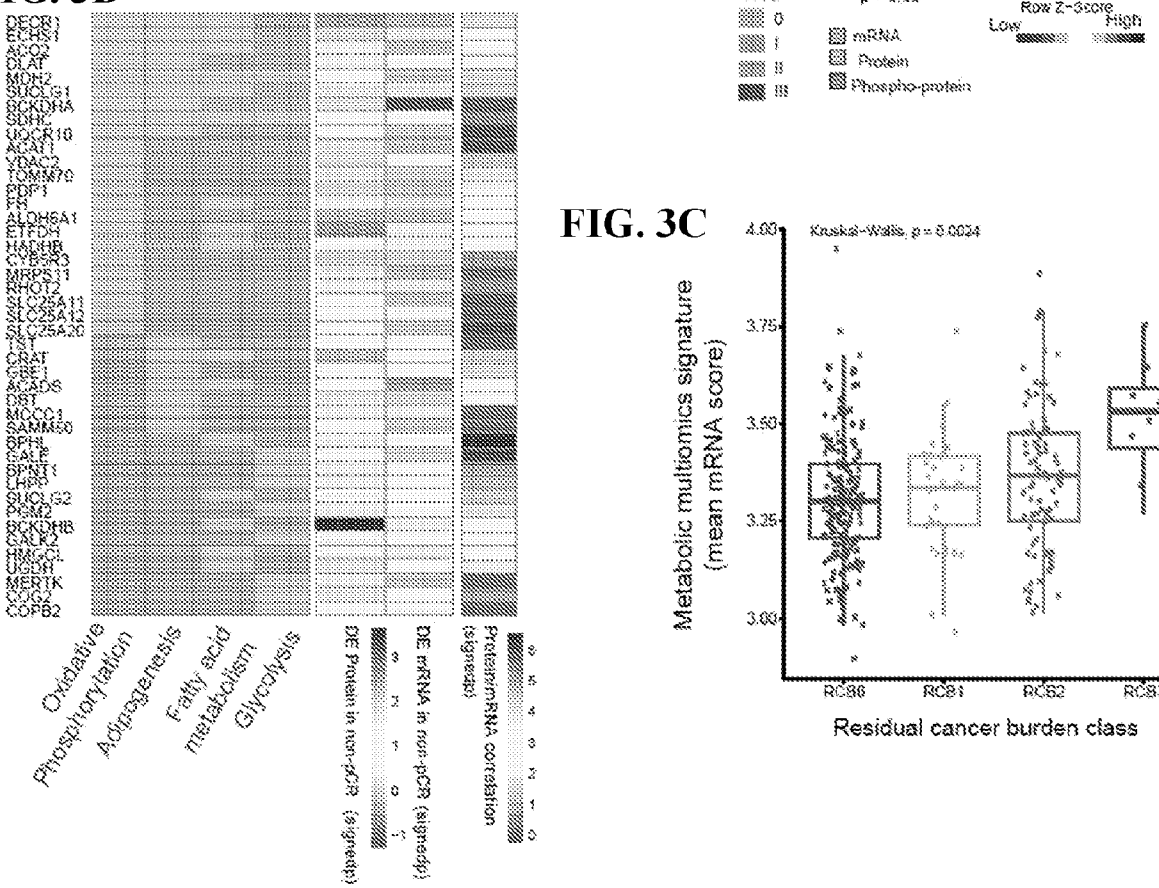


FIG. 3C

FIGs. 3A-3C

FIG. 4A

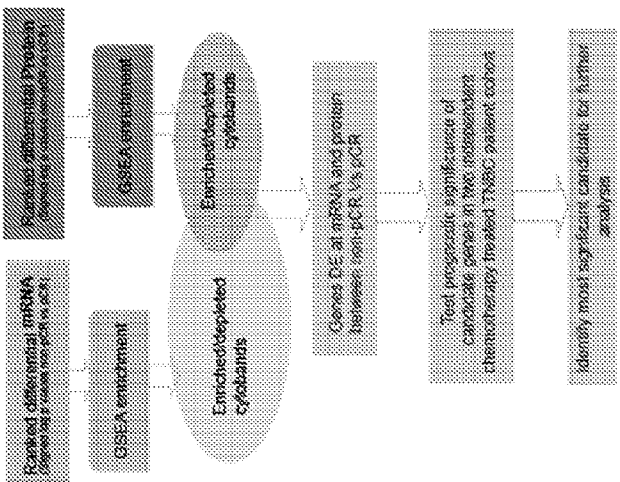
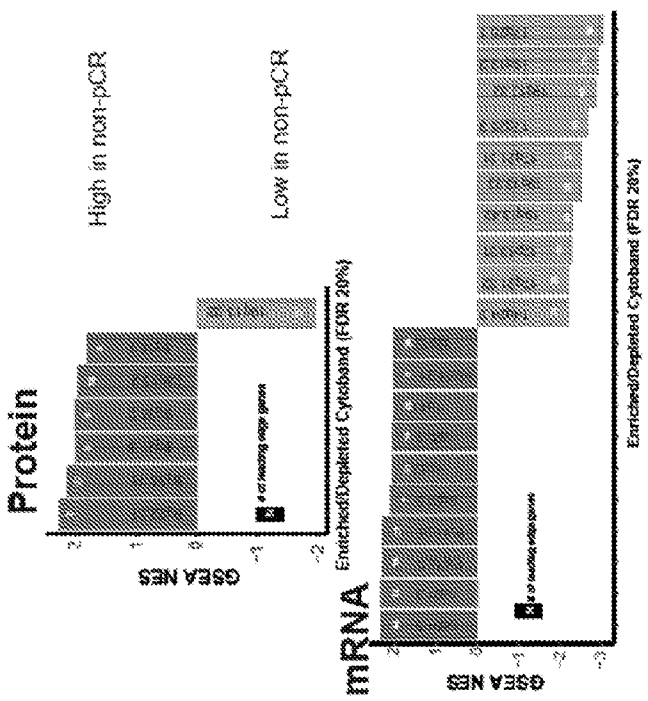


FIG. 4B



FIGs. 4A-4B

FIG. 4C

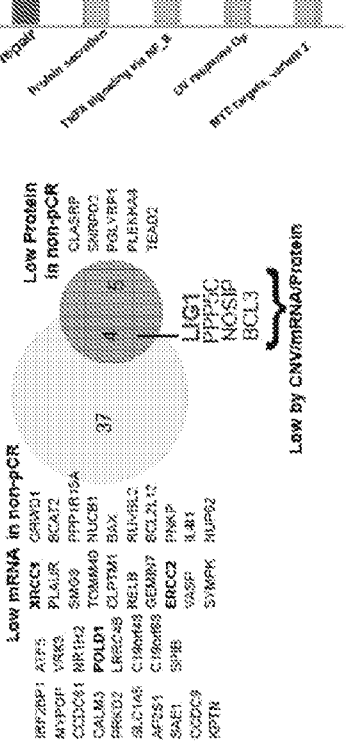


FIG. 4D

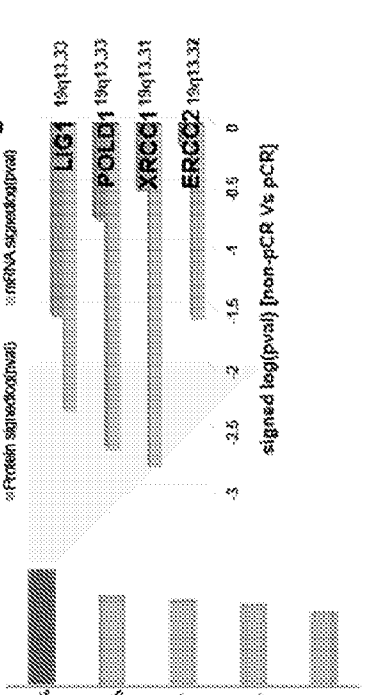


FIG. 4E

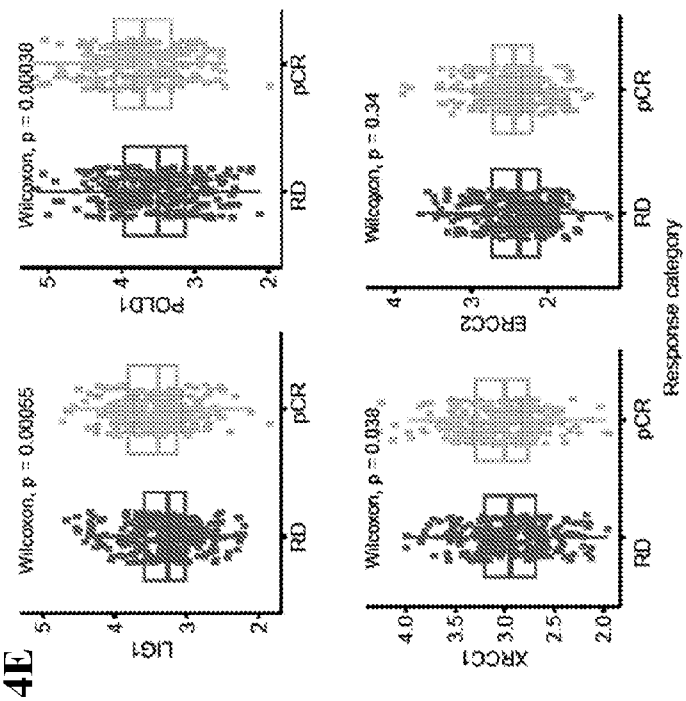
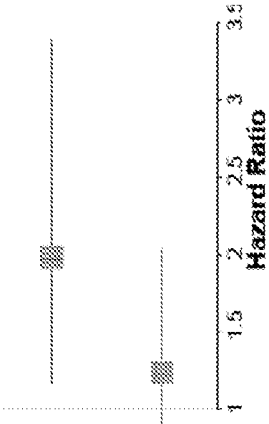


FIG. 4F



FIGs. 4B-4F

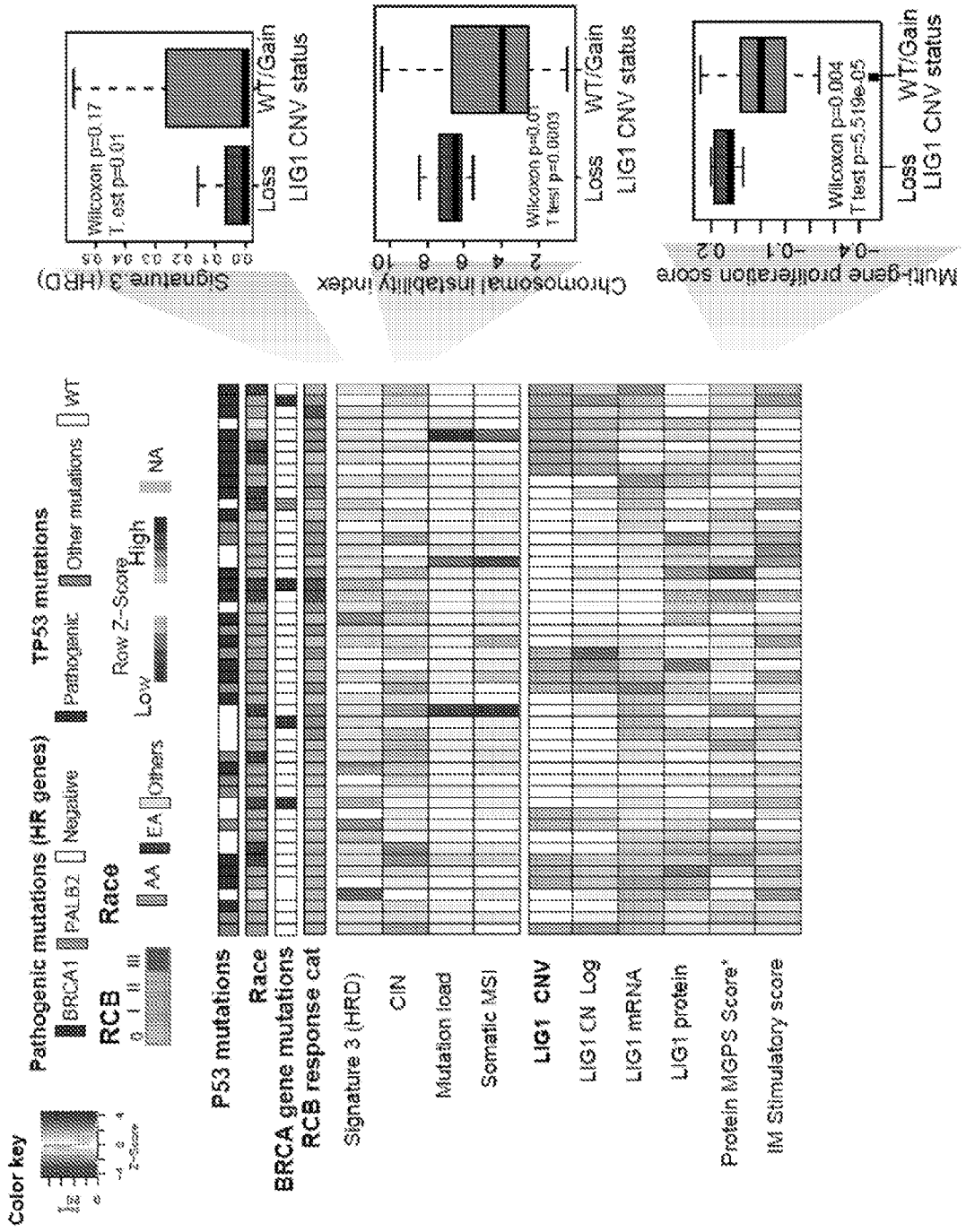

Color key
Pathogenic mutations (HR genes)
BRCA1
PALB2
Negative
RCB
Race
AA
EA
Others
TP53 mutations
Pathogenic
Other mutations
WT
Row Z-Score
Low
High
NA
P53 mutations
Race
BRCA gene mutations
RCB response cat
Signature 3 (HRD)
CIN
Mutation load
Somatic MSI
LIG1 CNV
LIG1 CN Log
LIG1 mRNA
LIG1 protein
Protein MGPS Score'
IM Stimulatory score
Signature 3 (HRD)
Wilcoxon p=0.17
T. est p=0.01
Loss
WT/Gain
LIG1 CNV status
Chromosomal Instability Index
Wilcoxon p=0.07
T test p=0.0003
Loss
WT/Gain
LIG1 CNV status
Multi-gene proliferation score
Wilcoxon p=0.004
T test p=5.519e-05
Loss
WT/Gain
LIG1 CNV status

FIG. 5

FIG. 6A

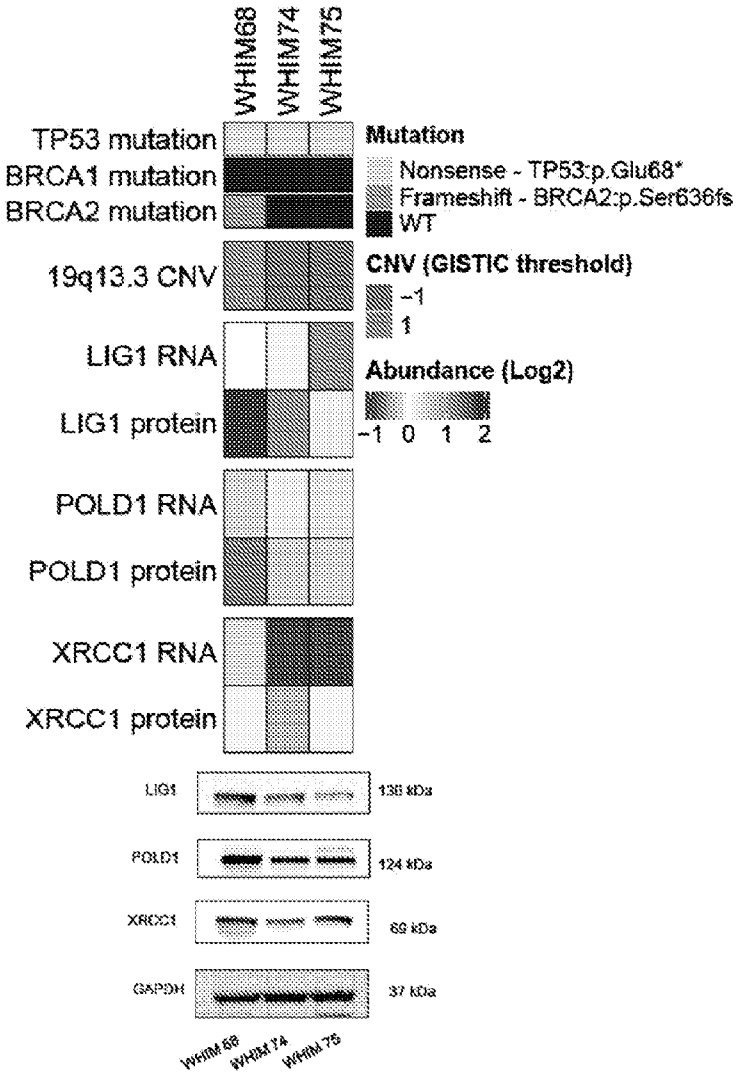
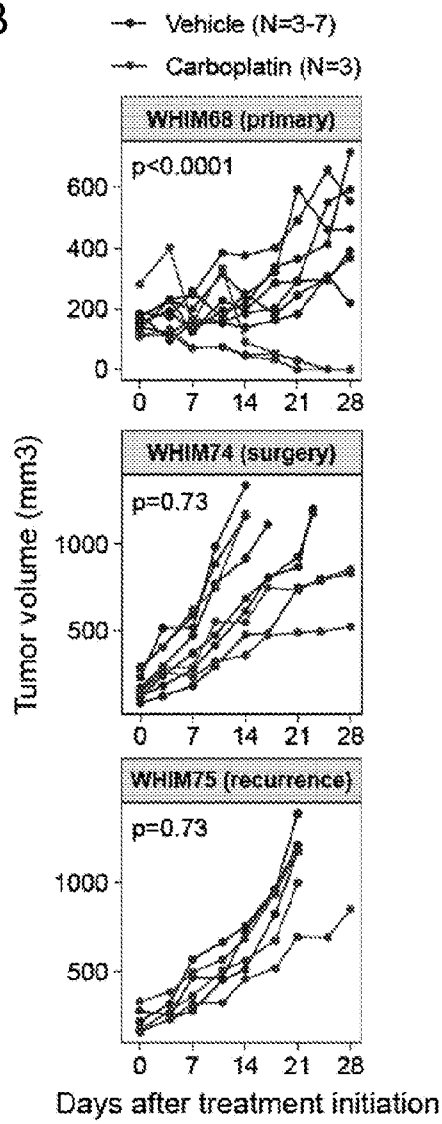


FIG. 6B



FIGs. 6A-6B

FIG. 6C

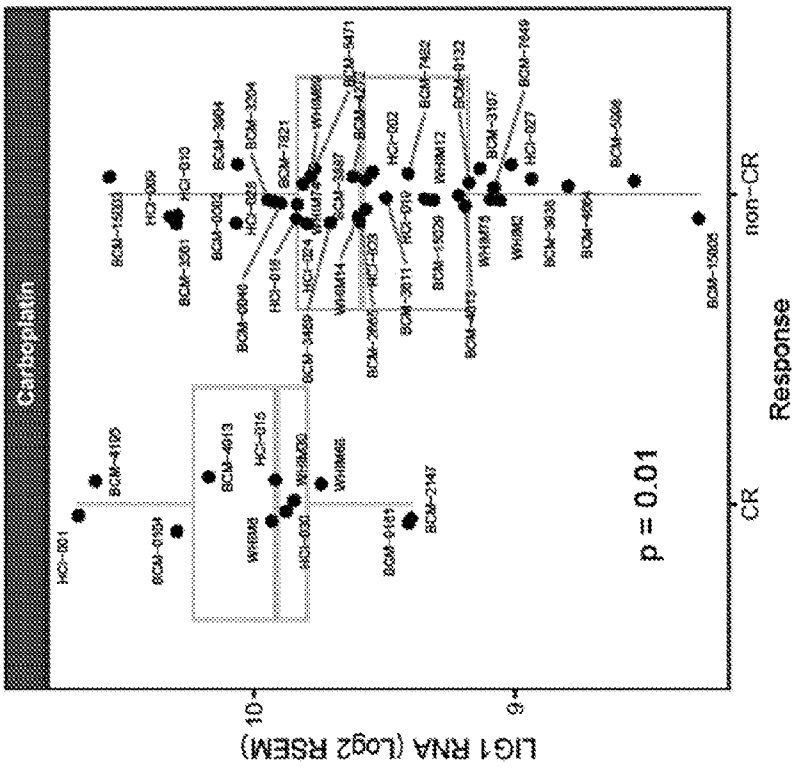
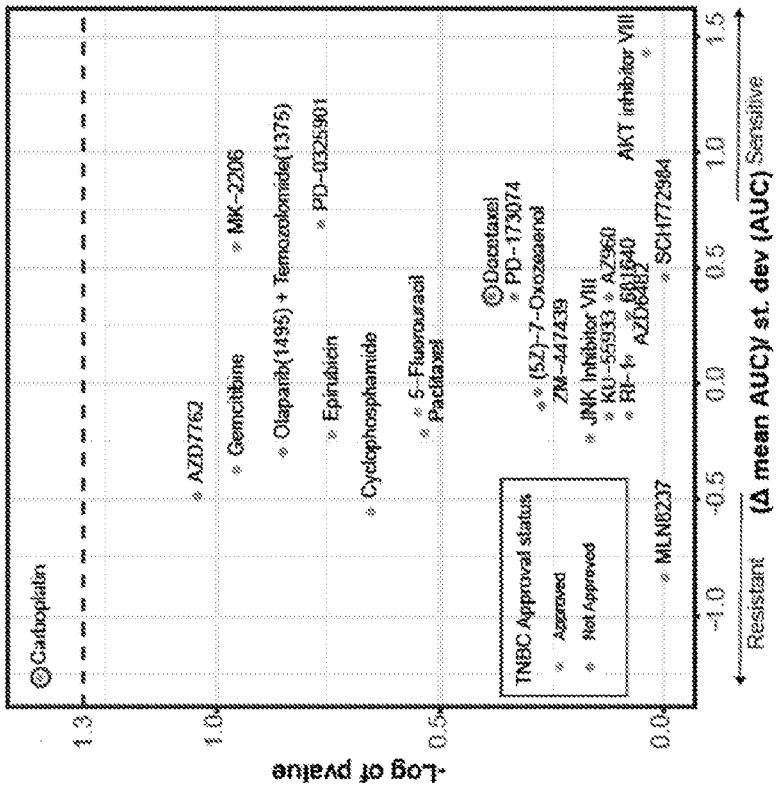


FIG. 6D



FIGs. 6C-6D

FIG. 7A

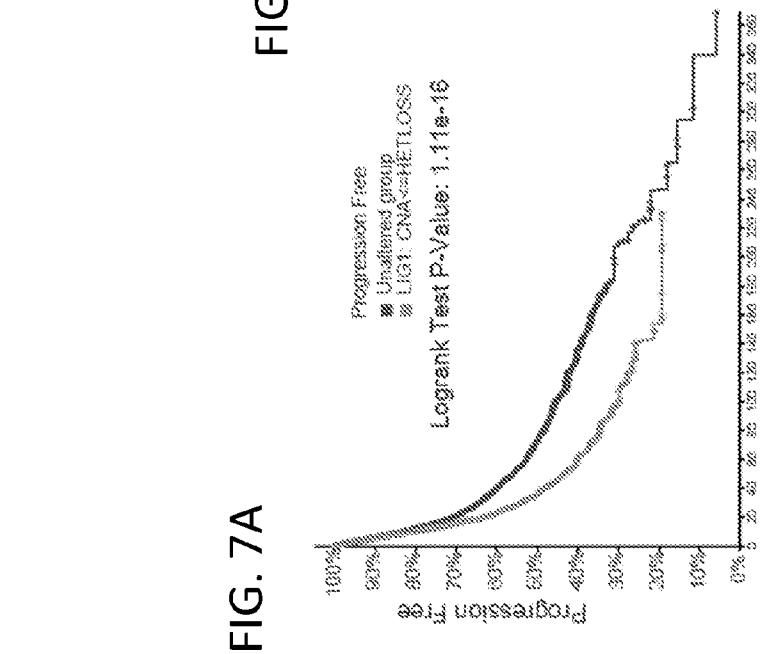


FIG. 7B

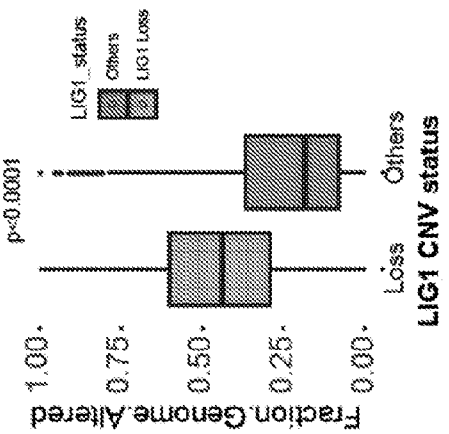
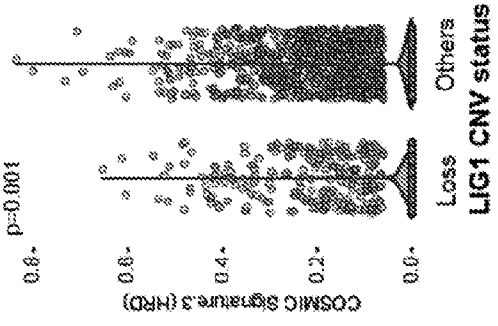


FIG. 7C

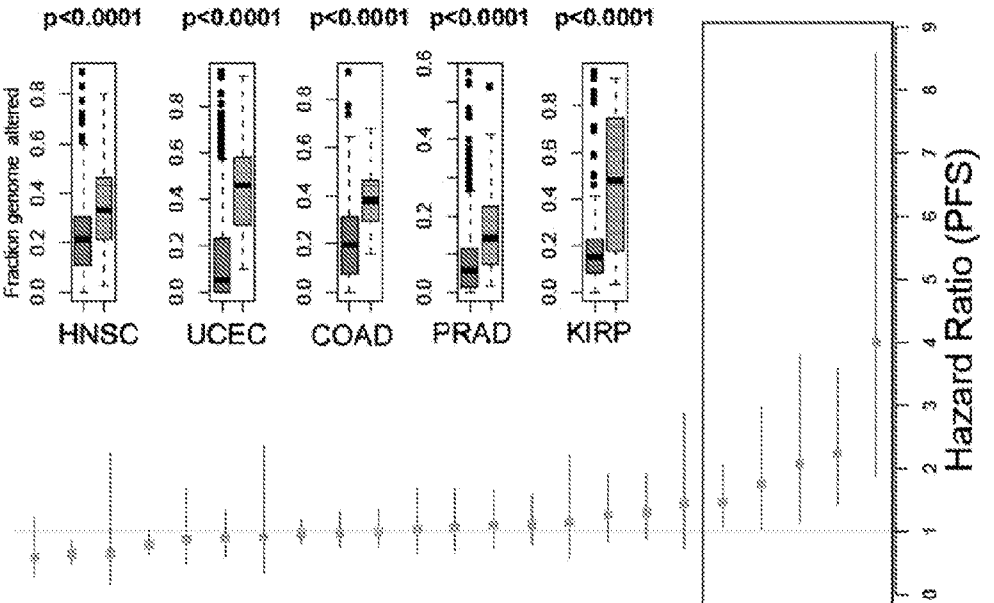


FIGs. 7A-7C

FIG. 7D

Cancer Type	LIG1 WT/gain	LIG1 loss	LIG1 loss %	HR	p-value
UCS	39	17	30.36	0.58	0.1
LGG	255	256	50.1	0.64	0.002
MESO	72	15	17.24	0.84	0.5
GBM	461	94	16.35	0.79	0.08
PAAD	161	22	12.02	0.87	0.7
LUSC	350	137	28.13	0.89	0.5
READ	133	22	14.19	0.9	0.8
OV	278	294	51.4	0.98	0.8
LUAD	321	190	37.18	0.98	0.9
SKCM	288	79	21.53	1	1
ESCA	131	51	28.02	1.03	0.9
LIHC	324	43	11.72	1.06	0.8
BRCA	867	203	18.97	1.09	0.7
SARC	171	82	32.41	1.11	0.5
CESC	257	36	12.29	1.13	0.7
STAD	364	74	16.89	1.26	0.3
BLCA	344	64	15.69	1.29	0.2
TGCT	91	58	38.93	1.44	0.3
HNSC	421	96	18.57	1.46	0.03
COAD	397	40	9.15	1.75	0.03
PRAD	452	37	7.57	2.07	0.02
UCEC	467	58	10.71	2.23	6e−04
KIRP	265	18	6.36	4	1e−04

FIG. 7E



FIGs. 7D-7E

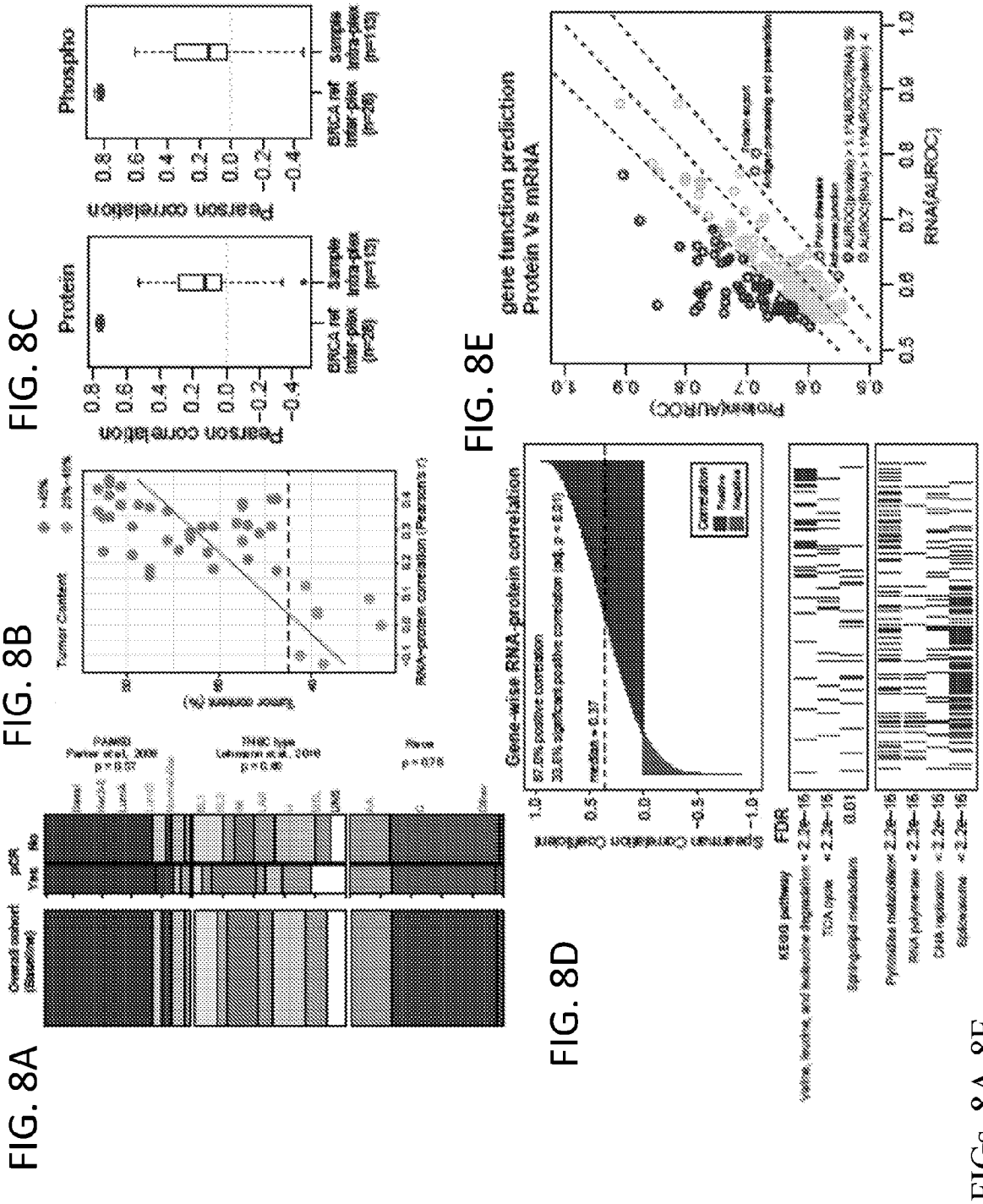


FIG. 8A FIG. 8B FIG. 8C FIG. 8D FIG. 8E

FIGs. 8A-8E

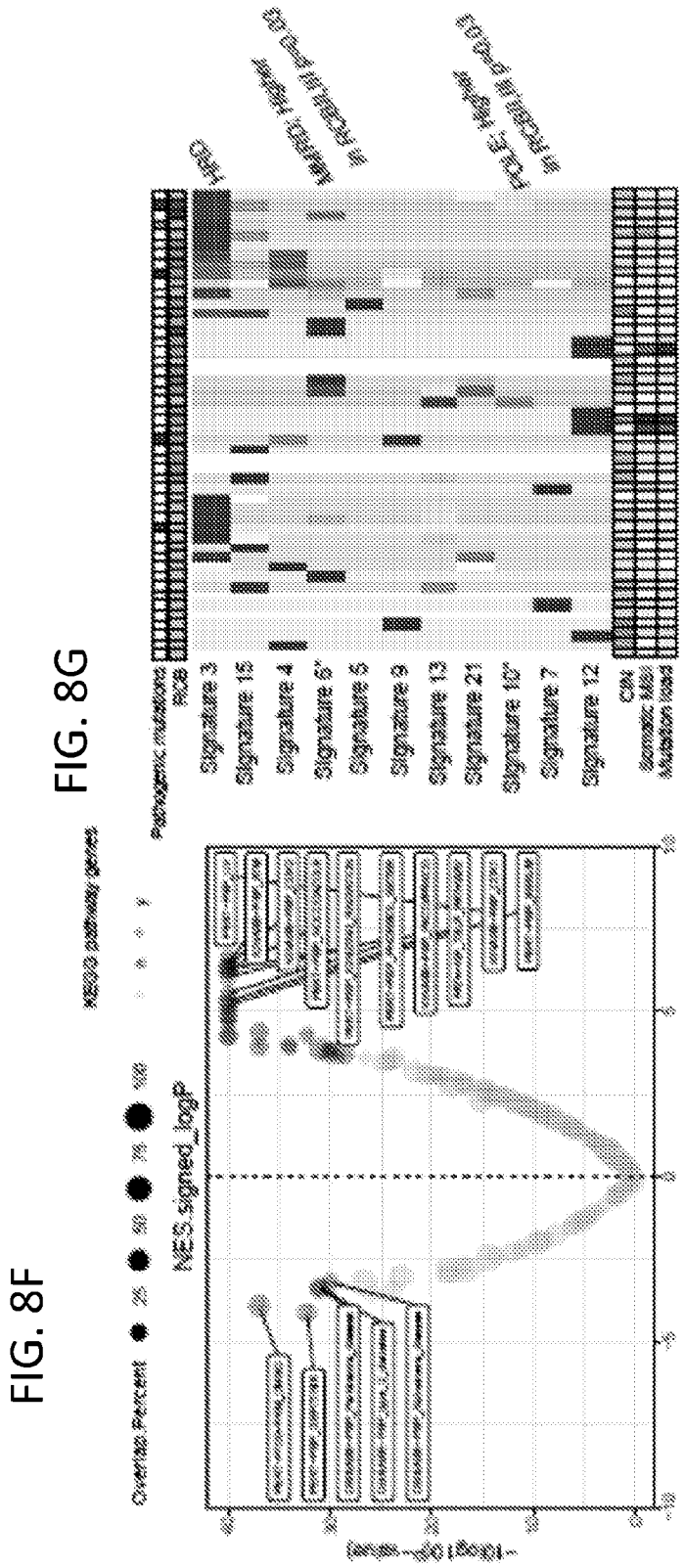


FIG. 8F

FIG. 8G

FIGs. 8F-8G

FIG. 9A

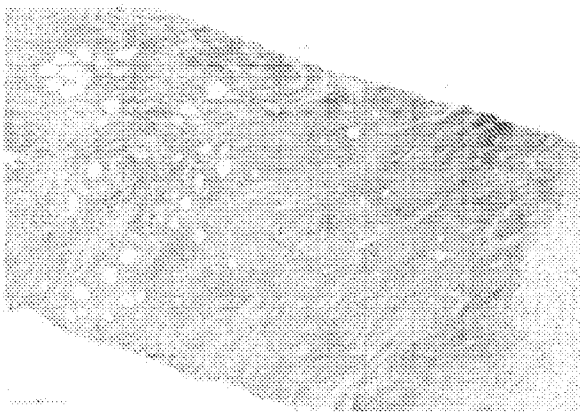


FIG. 9B

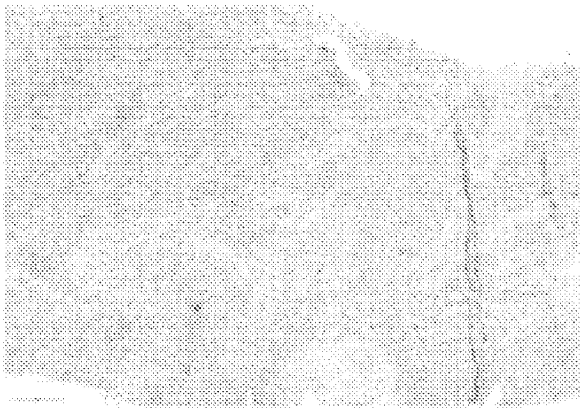


FIG. 9C

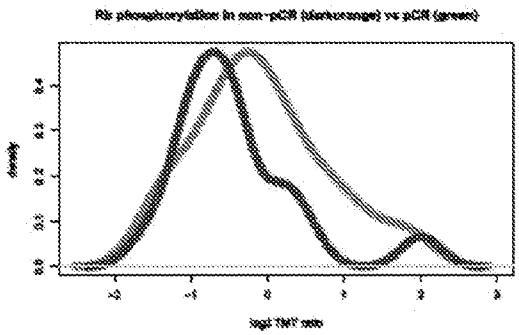
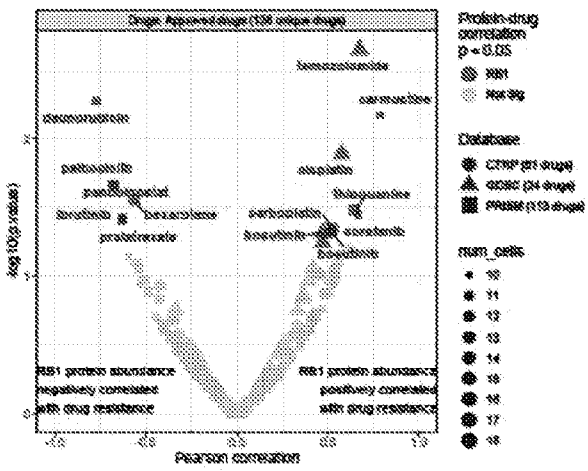


FIG. 9D



FIGs. 9A-9D

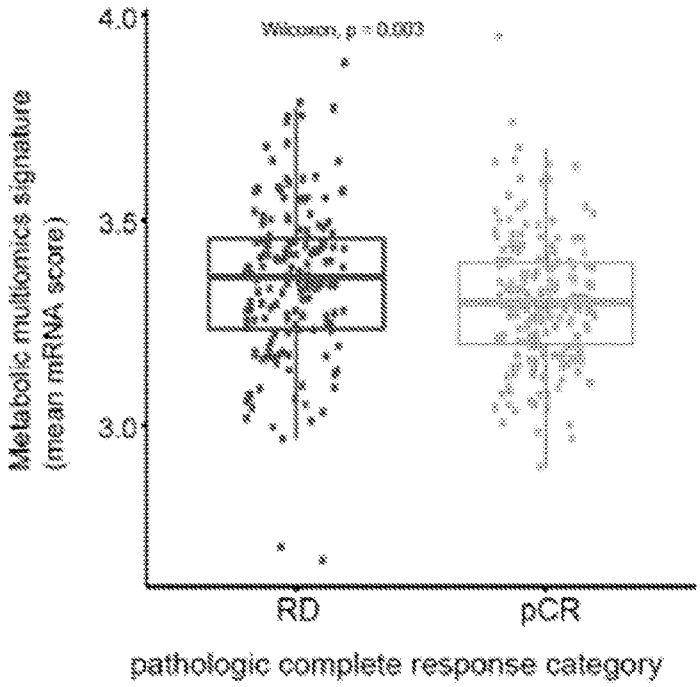

Wilcoxon, p = 0.003
4.0
3.5
3.0
Metabolic multiomics signature
(mean mRNA score)
RD
pCR
pathologic complete response category

FIG. 10

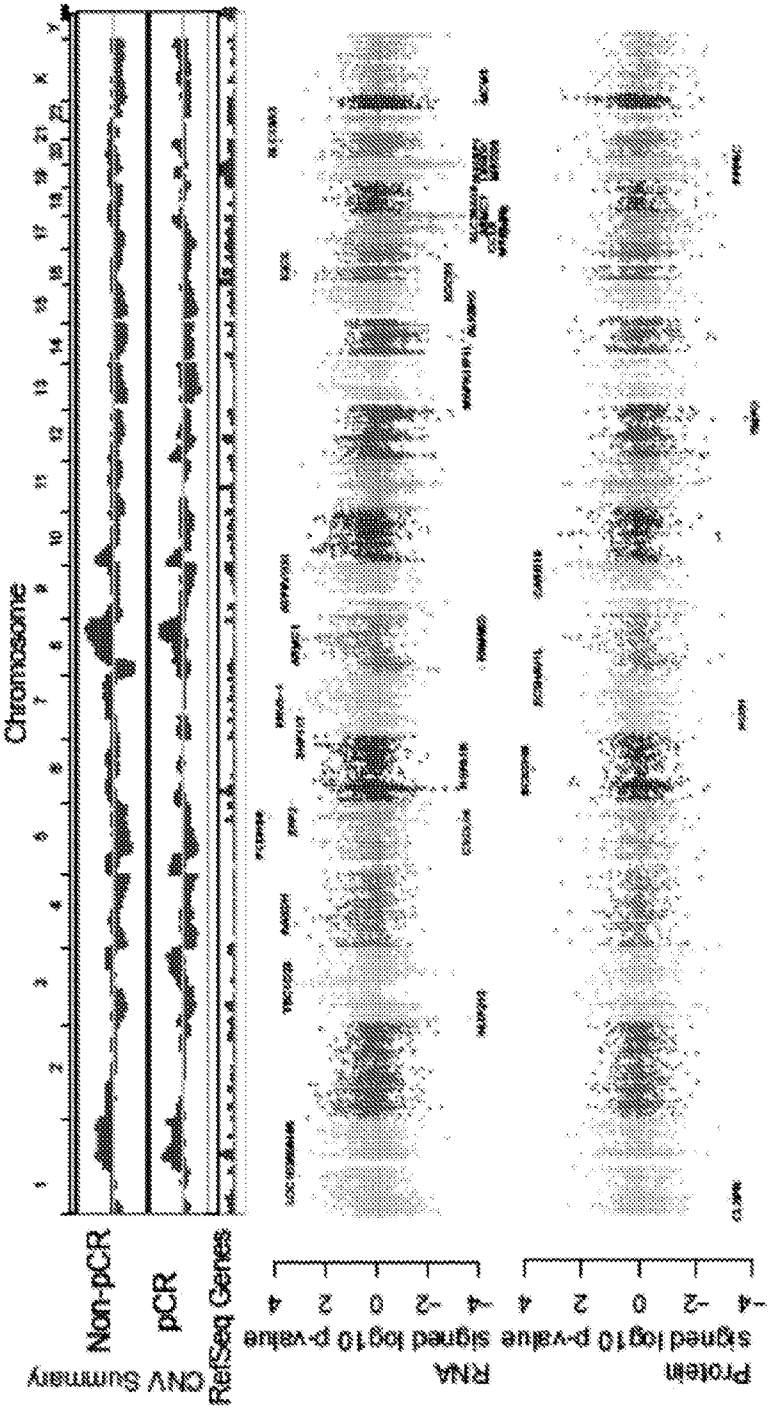

Chromosome
CNV Summary
Non-pCR
pCR
RefSeq Genes
RNA
signed log10 p-value
Protein
signed log10 p-value

FIG. 11A

FIG. 11B

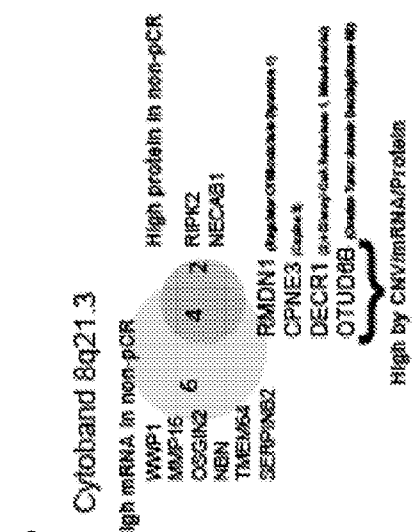


FIG. 11C

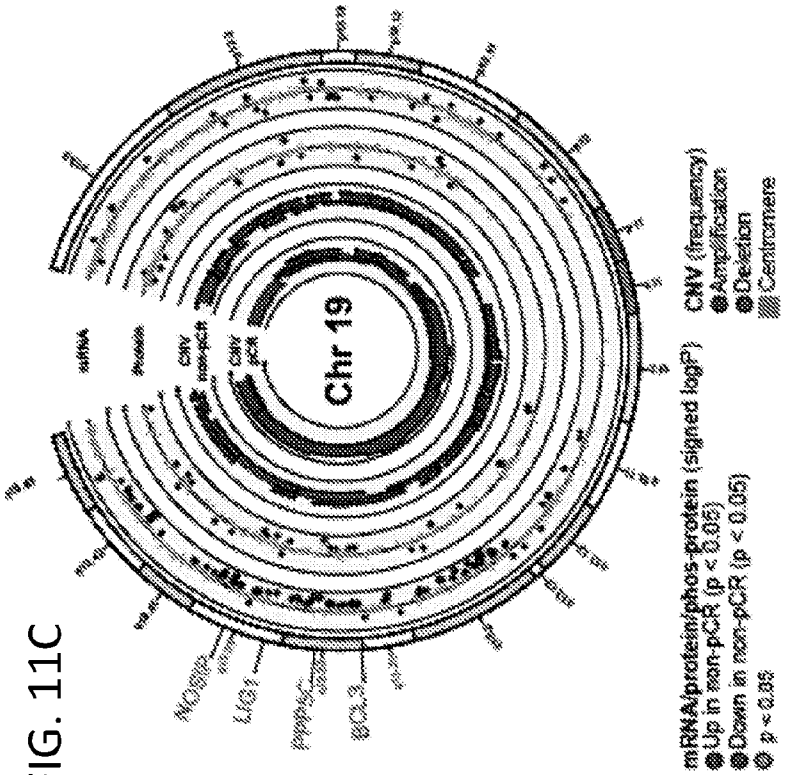
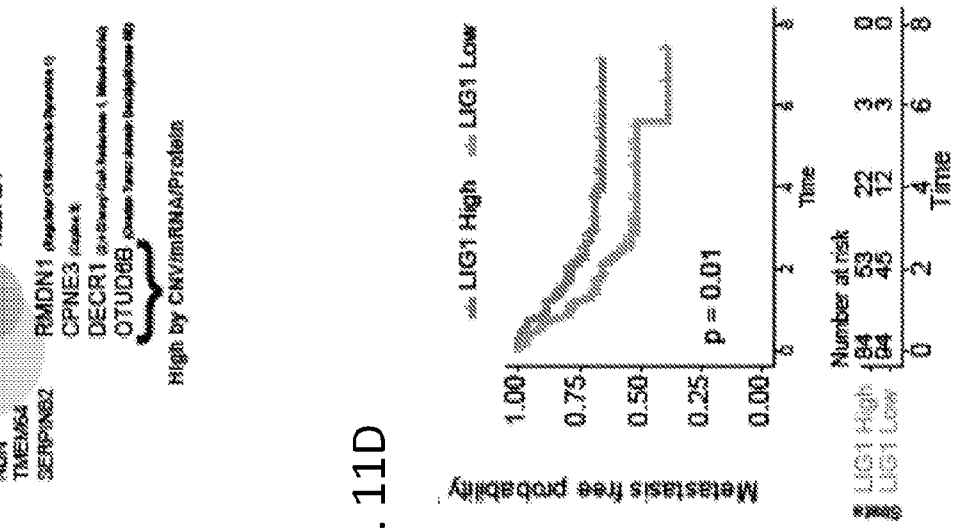


FIG. 11D



FIGs. 11B-11D

FIG. 12A

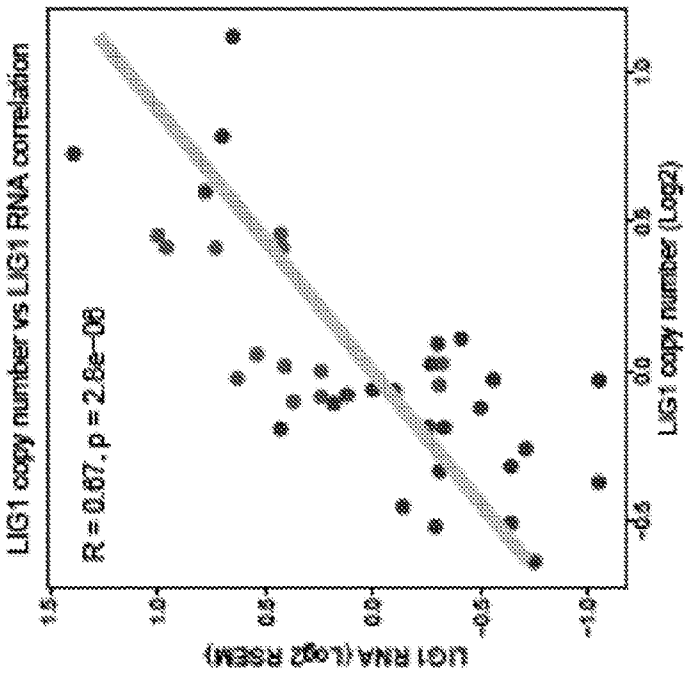
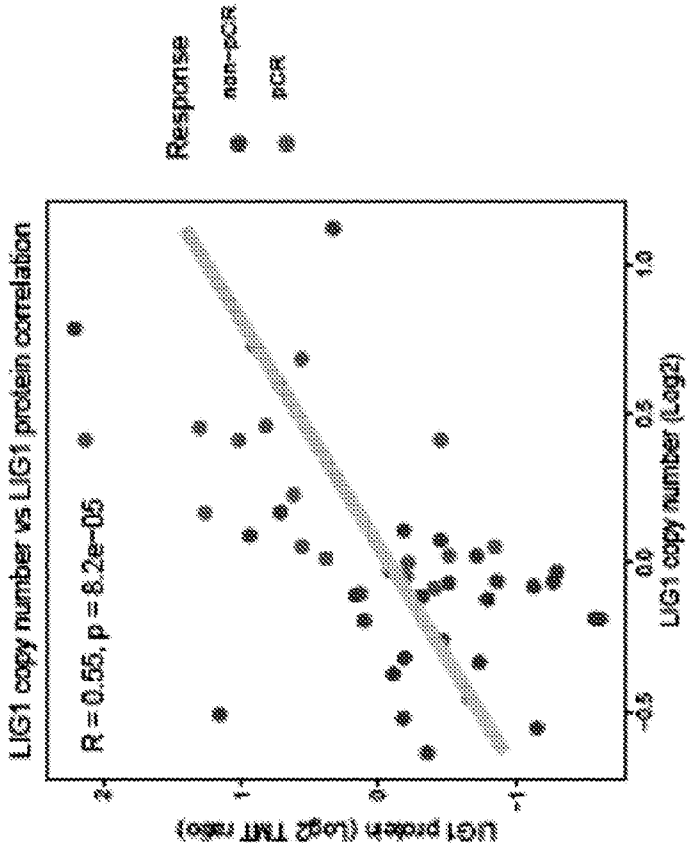


FIG. 12B



FIGs. 12A-12B

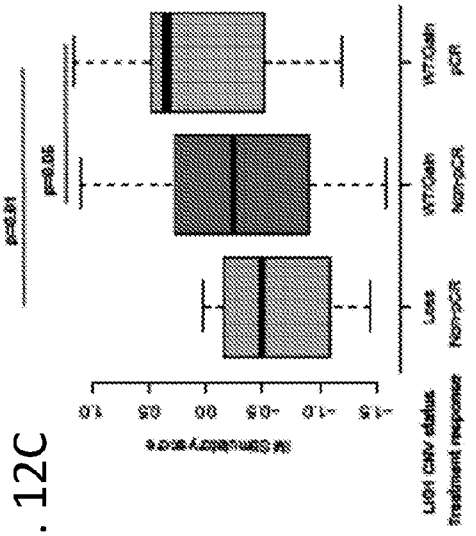


FIG. 12C

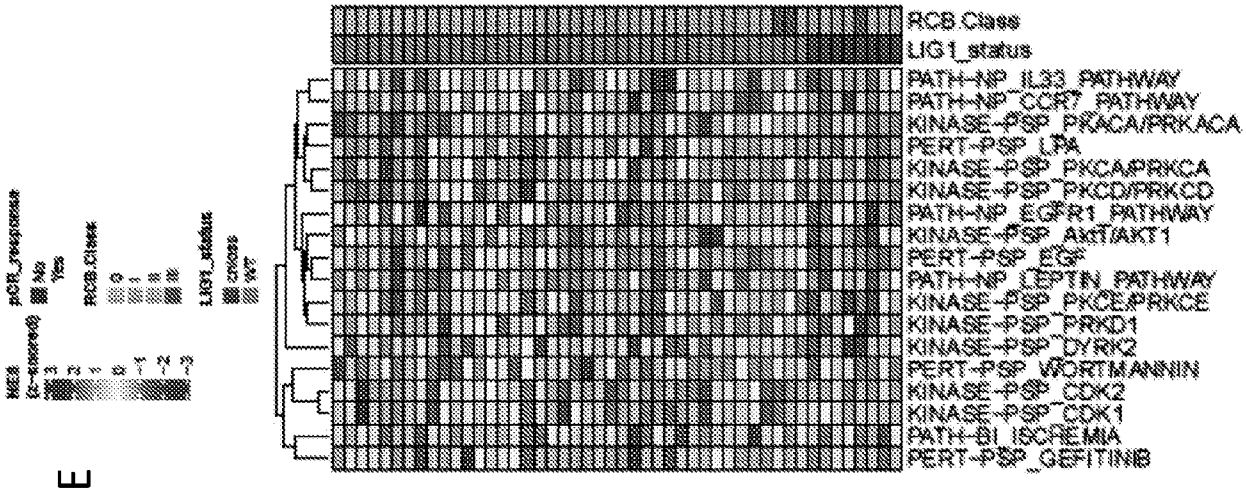


FIG. 12E

FIG. 12D

FIGs. 12C-12E

FIG. 13A

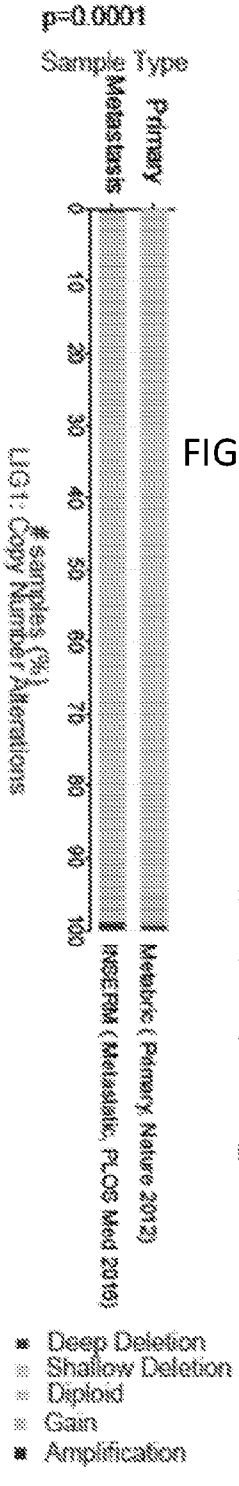


FIG. 13B

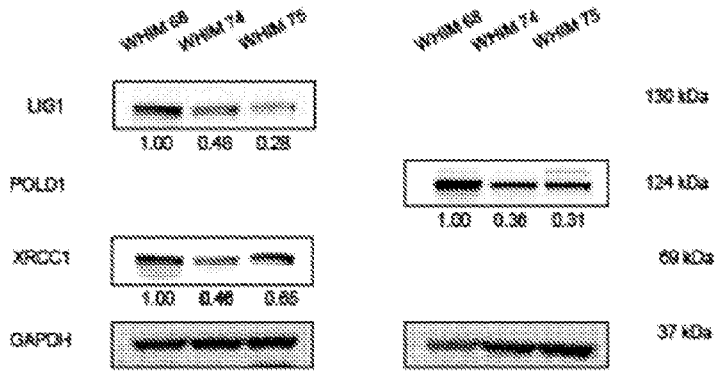


FIG. 13C

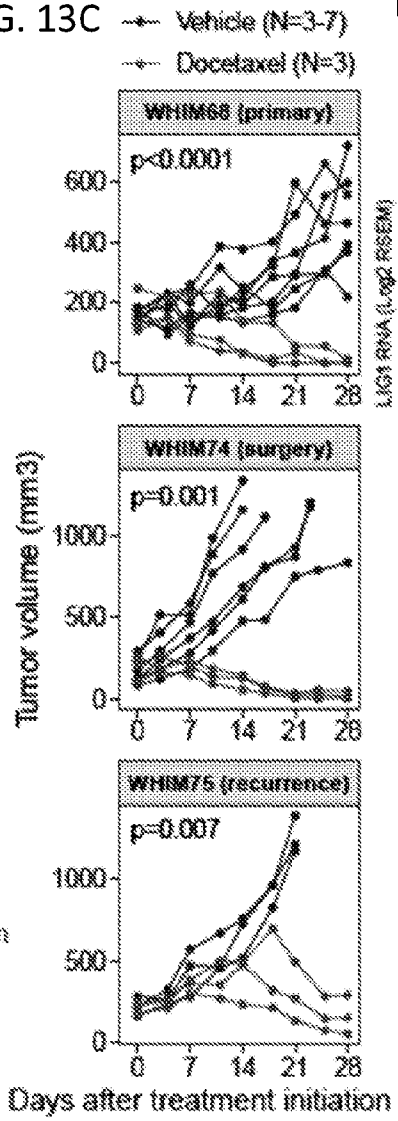
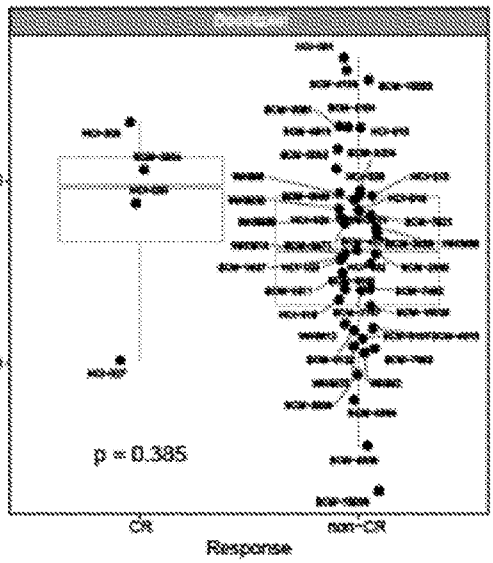


FIG. 13D



FIGs. 13A-13D

FIG. 14A

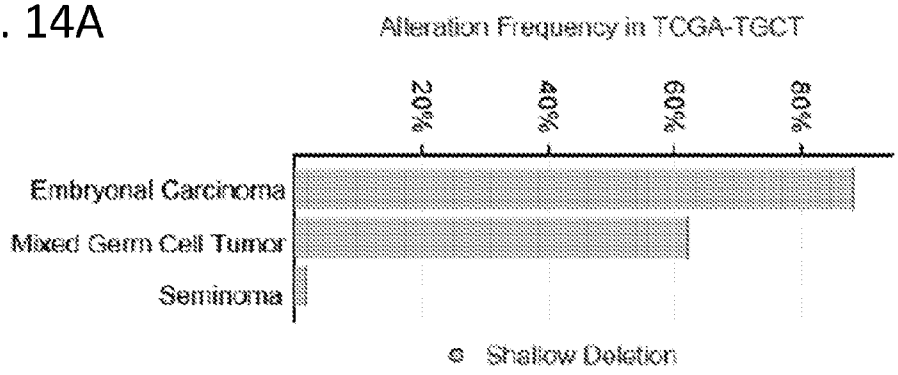
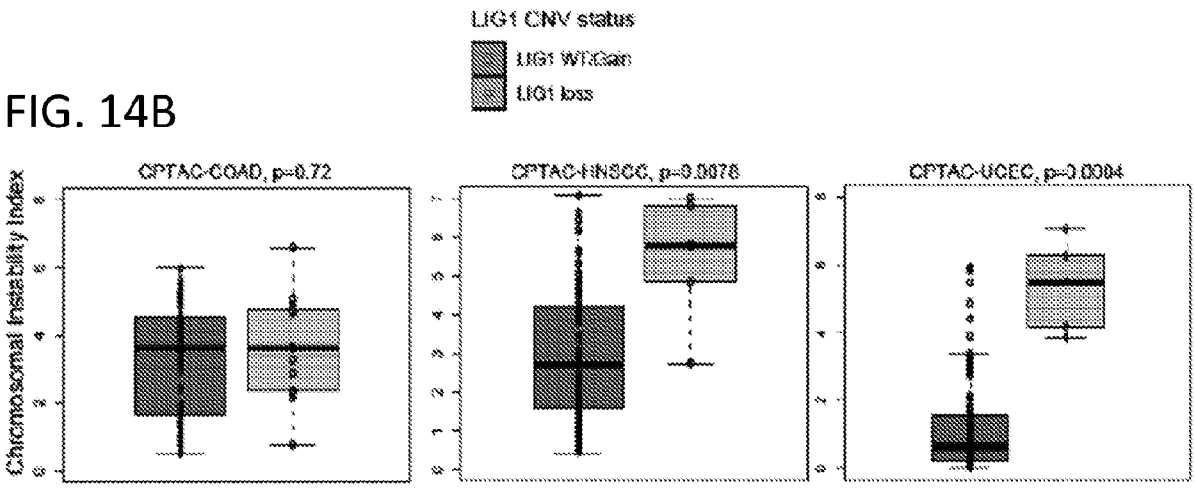


FIG. 14B



FIGs. 14A-14B

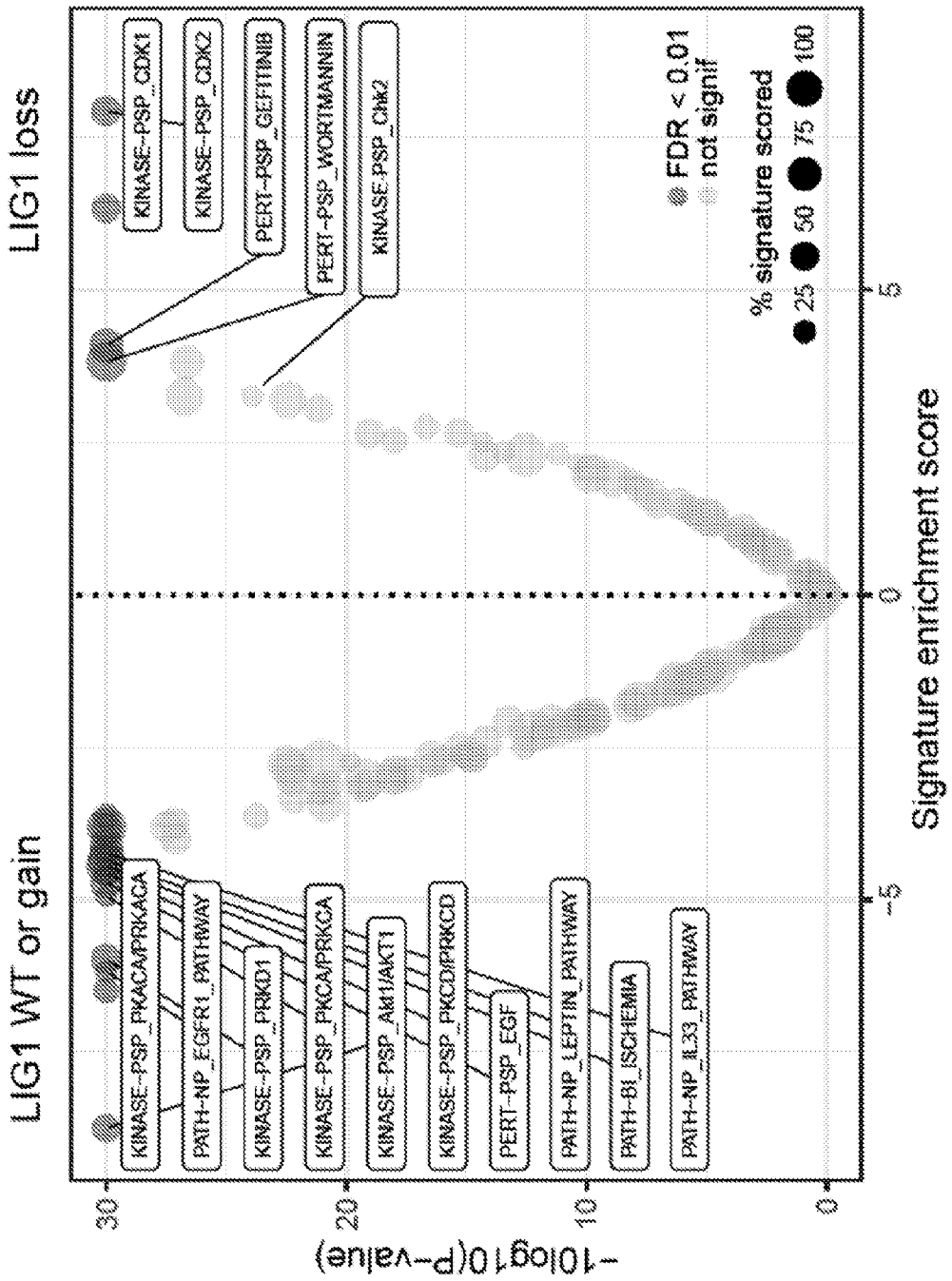

LIG1 loss
LIG1 WT or gain
KINASE-PSP_CDK1
KINASE-PSP_CDK2
PERT-PSP_GEFITINIB
PERT-PSP_WORTMANNIN
KINASE-PSP_Chk2
KINASE-PSP_PKACA/PRKACA
PATH-NP_EGFR1_PATHWAY
KINASE-PSP_PRKD1
KINASE-PSP_PKCA/PRKCA
KINASE-PSP_Akt1/AKT1
KINASE-PSP_PKCD/PRKCD
PERT-PSP_EGF
PATH-NP_LEPTIN_PATHWAY
PATH-BI_ISCHEMIA
PATH-NP_IL33_PATHWAY
FDR < 0.01
not signif
% signature scored
25
50
75
100
−10log10(P−value)
0
10
20
30
−5
0
5
Signature enrichment score

FIG. 15

FIG. 16A

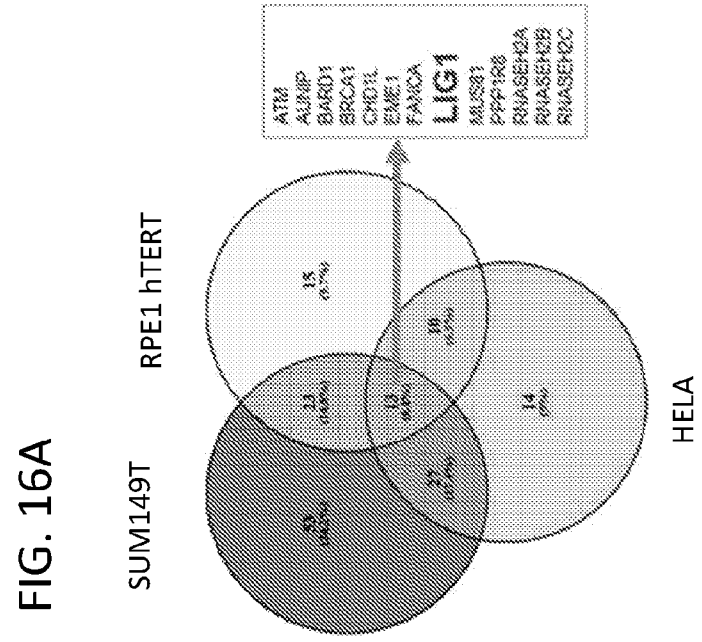
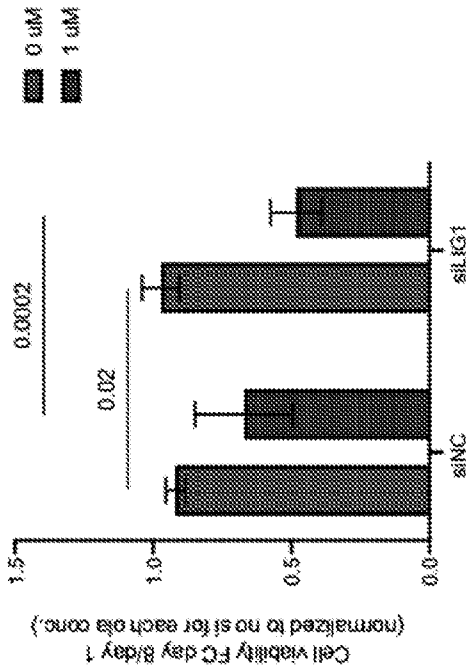


FIG. 16B



FIGS. 16A and 16B

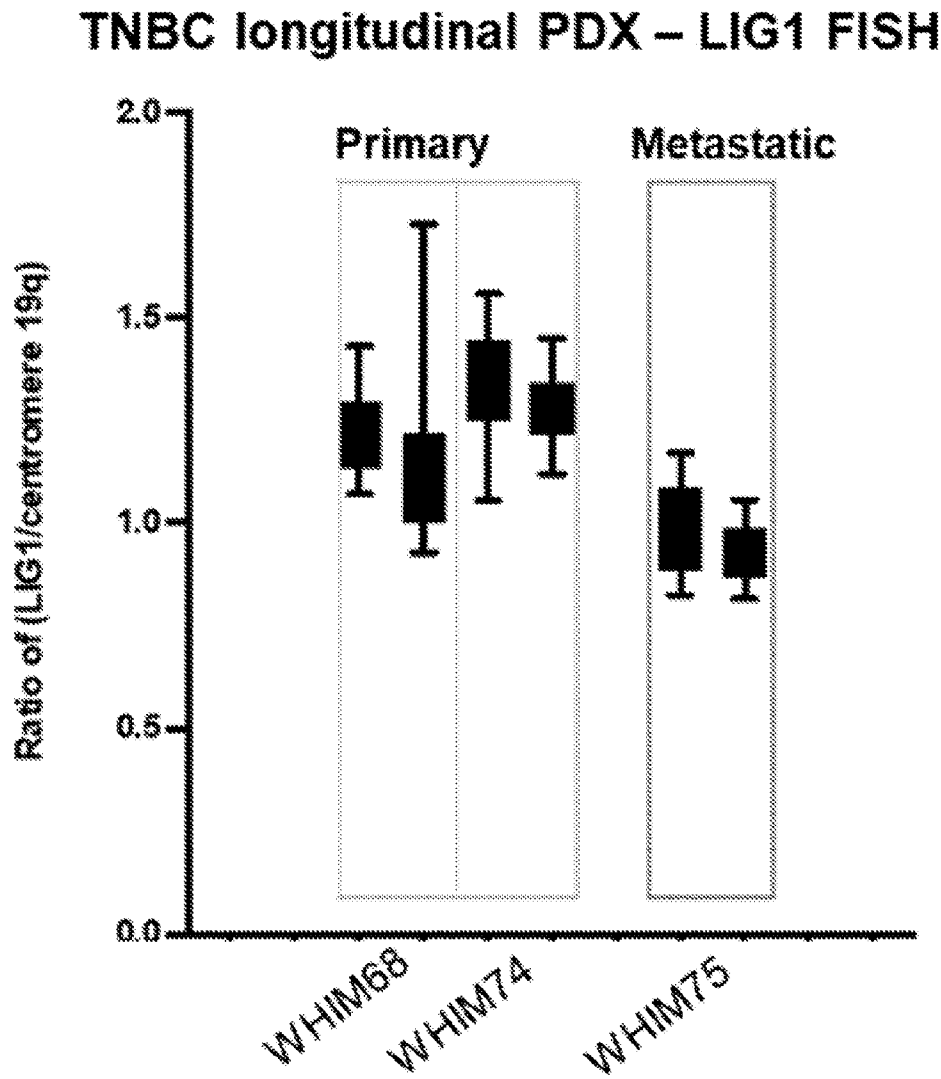
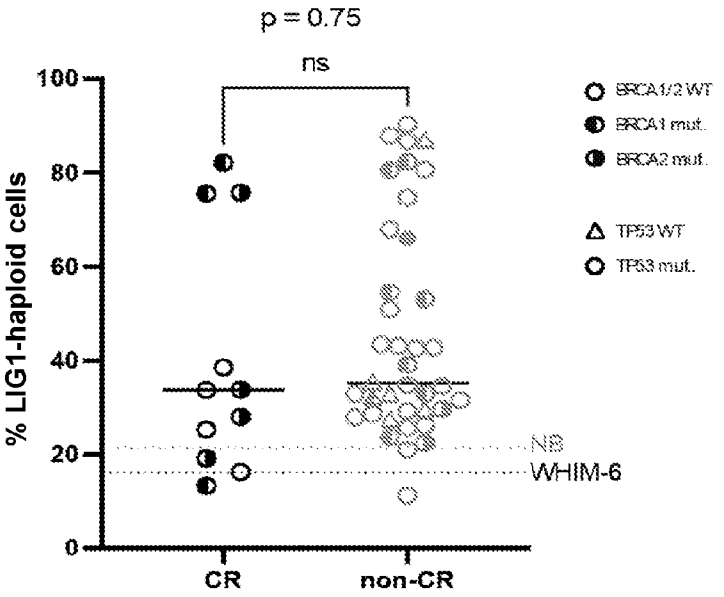

TNBC longitudinal PDX – LIG1 FISH
Primary
Metastatic
Ratio of (LIG1/centromere 19q)
2.0
1.5
1.0
0.5
0.0
WHIM68
WHIM74
WHIM75

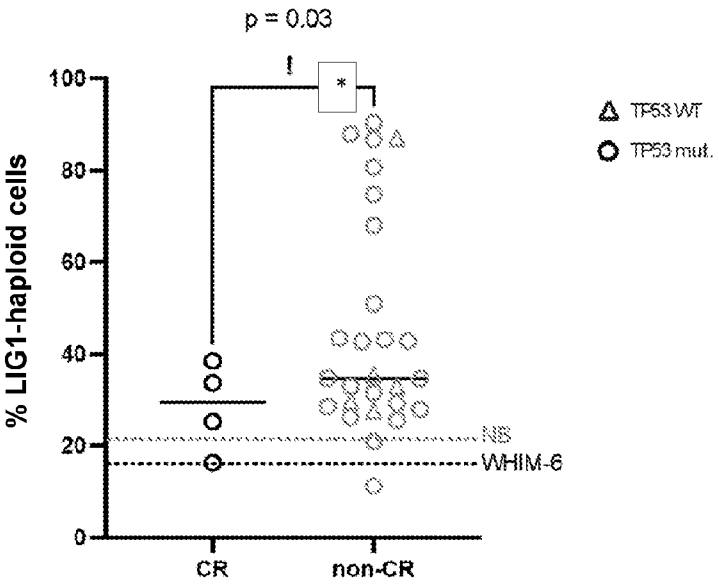
FIG. 17A



With BRCA1/2 mutations

FIG. 17B

excluding confounding factors



* Indicates p<0.05

FIG. 17C

FIGs. 17B-17C