Precision combination therapies from recurrent oncogenic co-alterations

[1]Xubin Li, Elisabeth Dowling[2], Gonghong Yan[1], Behnaz Bozorgui[1], Parisa Imarinad[3], Jacob H. Elnaggar[4], Augustin Luna[5,6], Chris Sander[5,6], Anil Korkut[1,*]

[1]Department of Bioinformatics and Computational Biology, UT MD Anderson Cancer Center, Houston, TX 77030, USA, [2]Department of Statistics, Rice University, Houston, TX, 77030, USA, [3]Department of Systems Biology, UT MD Anderson Cancer Center, Houston, TX 77030, USA. [4]Department of Microbiology, Immunology, and Parasitology, Louisiana State University Health Sciences Center, New Orleans, LA 70112, [5]cBio Center, Department of Data Sciences, Dana Farber Cancer Institute, Boston, MA 02215, USA, [6]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

*akorkut@mdanderson.org

Abstract

Cancer cells rely on co-activation of oncogenic processes that are driven by co-occurring molecular aberrations. Combination therapies that target co-activated oncogenic processes can induce stronger and more durable responses compared to mono-therapies. Here, we developed REFLECT (REcurrent Features Leveraged for Combination Therapies) to match precision combination therapies to co-alteration signatures that are recurrent within patient cohorts. Using REFLECT, we generated a comprehensive resource mapping the actionable co-alteration signatures at DNA, mRNA and protein levels across more than 10,000 cancer patients and 199 cohorts. The patient cohorts are generated by stratifying the patient meta-cohort based on the existence of key oncogenic alterations. In the cohort of patients carrying potential immunotherapy markers, we identified distinct patient sub-cohorts that are enriched by actionable DNA repair aberrations, inflammation markers and B-cell presence. Through REFLECT analysis and validations in molecularly matched cell lines, we nominated HER2 and Aurora kinase inhibitors as effective in patients with *AURKA* mutations and HER2 activation. The analyses of DNA repair aberrations combined with validation experiments have identified recurrent molecular signatures that match effective combinations involving PARP, AKT, MET and other DNA repair pathway inhibitors. We expect REFLECT will facilitate the selection of effective combination therapies and benefiting patient cohorts matched to aberration signatures in preclinical and data-driven clinical trials.

Introduction

Targeted therapies such as HER2 and PARP inhibitors have significantly improved clinical outcomes in cancer[1,2,3]. Despite the transformative advances in sequencing technologies and precision medicine, durable responses to targeted therapies in genetically-matched patient cohorts have remained limited[4,5]. The limited success can be at least partially attributed to treatment with mono-therapies targeting individual oncogenic aberrations. Oncogenesis, however, is driven by co-occurring aberrations in DNA, mRNA and protein levels that lead to co-activated oncogenic processes. A rich repertoire of preclinical and clinical studies has demonstrated that rational combination therapies blocking multiple oncogenic processes are more effective than mono-therapies[6,7,8]. Discovery of combination therapies that may benefit large patient cohorts, however, is a daunting task due to the complexity of the actionable oncogenic aberration landscapes[9].

There is growing evidence that co-targeting the co-occurring oncogenic driver aberrations may induce effective and durable therapeutic responses. For example, PI3K-AKT pathway co-alterations alongside BRAFV600E mutations impact response to BRAF inhibitors and establish a rationale for co-targeting the BRAF and PI3K-AKT pathways[10]. Similarly, co-alterations in KRAS, BRAF and PIK3CA modulate response to EGFR/HER2 targeted therapies in colorectal cancers[8]. Genomic studies focusing on co-driver events in KRAS mutated lung cancers[11] and recent combination therapy trials focusing on co-alterations (I-PREDICT[27], MyPathway[28]) also strongly support this argument[6,8]. Bioinformatics methods and resources that extract clinically useful knowledge (e.g., recurrent, co-occurring, co-actionable oncogenic events) from large multi-omic (genomic, transcriptomic and proteomic) data sets will accelerate the implementation of precision combination therapies.

Here, we developed the machine learning algorithm and precision oncology resource termed “Recurrent Features Leveraged for Combination Therapy" (REFLECT). First, REFLECT generates multi-omic signatures within a patient cohort (e.g., EGFR mutated patients) using a LASSO-penalty based sparse hierarchical clustering scheme[12]. Each signature is marked by the distribution of co-occurring and recurrent gene mutations or mRNA/protein expression aberrations that can discriminate patients into sub-cohorts. Next, REFLECT matches the co-occurring aberrations in multi-omic (mutation, mRNA, phosphoprotein) signatures to potential combination therapies through the identification of recurrent and co-actionable oncogenic alterations. We have determined comprehensive genomic, transcriptomic and proteomic REFLECT-signatures as well as combination therapies matching to patient sub-cohorts in 199 unique patient cohorts. Each patient cohort is constructed based on the existence of at least one actionable aberration with sampling from a PanCancer meta-cohort (N=10392) of patients from the TCGA resource[13]. Through REFLECT analysis of the patient cohorts and proof-of-concept experimental validations in molecularly matched cell lines, we identified a series of drug combinations that can benefit sub-cohorts of patients marked by clinically relevant events such as PD1/PDL1 expression/immune infiltration, DNA checkpoint/repair aberrations and increased HER2 activity. We provide a searchable online resource with interactive visualization tools and tutorials (https://bioinformatics.mdanderson.org/reflect/). We also provide an R package REFLECT for the generation of multi-omic signatures and matching combination therapies to patient cohorts with custom patient data (https://github.com/korkutlab/reflect). We expect the tool and the associated resource will guide and accelerate both the discovery and selection of novel combination therapies potentially benefiting patients for emerging combination therapy clinical trials.

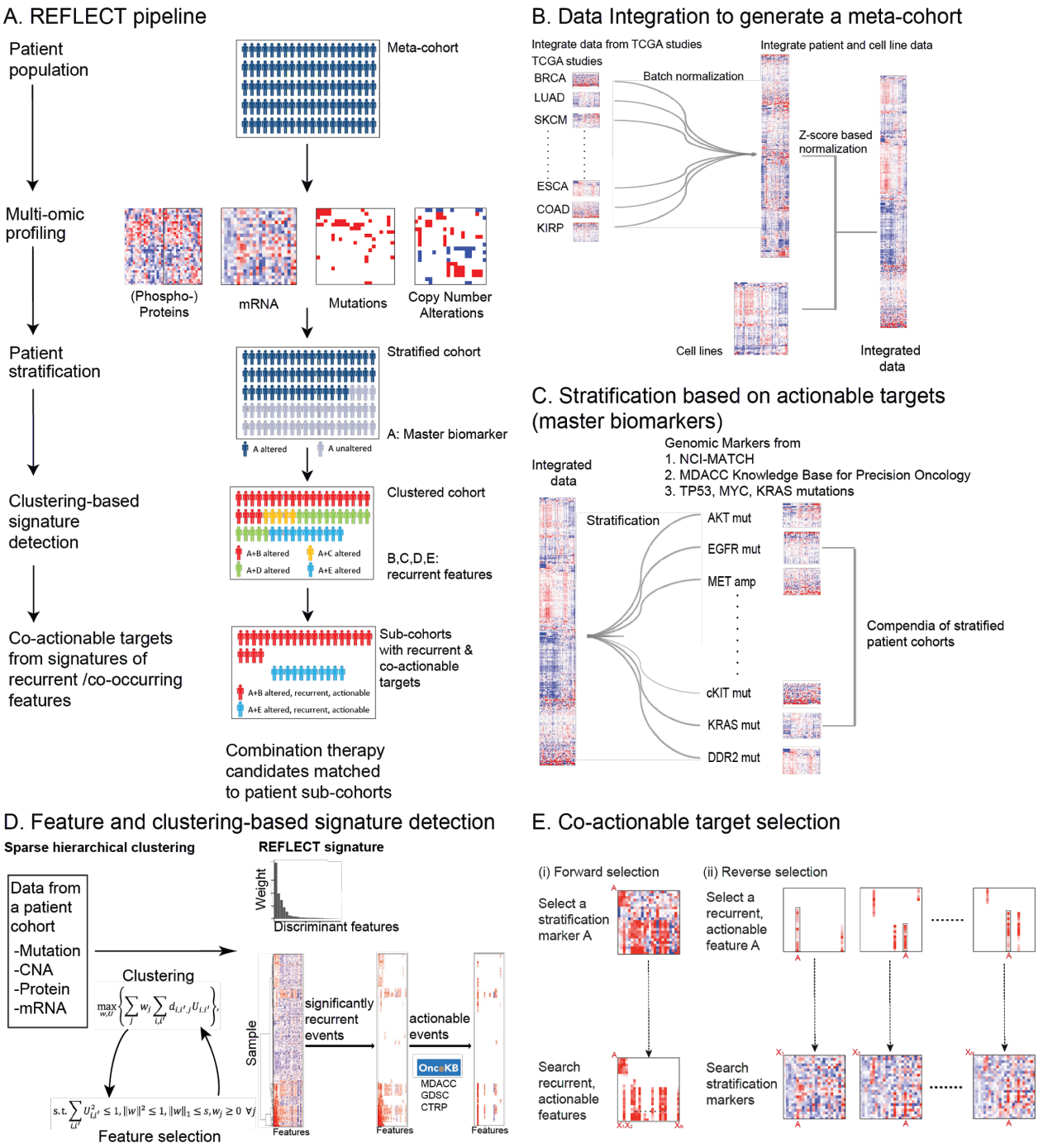

Figure 1. REFLECT. A. The pipeline. *REFLECT identifies recurrent, co-occurring and co-actionable aberrations within patient cohorts. The method uses mutation, copy number, mRNA, phosphoproteomics data. Patient cohort-specific multi-gene/protein signature detection with clustering/feature selection is followed by matching the signatures to combination therapies through searches in oncology knowledge bases.* ***B. Data integration.*** *Molecular datasets from diverse cancer types are integrated to generate a PanCancer set of patients and cell lines for each data type. The mRNA/phosphoproteomics data from patients and cell lines is integrated using a Z-score based normalization (see methods).* ***C. Actionable alteration based cohort stratification.*** *The integrated pan-cancer data is partitioned into overlapping cohorts based on*

the existence of actionable genomic aberrations (e.g., a cohort of EGFR mutant tumors). The mutational, copy number, mRNA and protein datasets from each partition are analyzed separately. ***D. Selection of significantly recurrent alterations.*** *To identify sub-cohorts of patients characterized by a small set of discriminant and recurrent features (i.e., genomic alterations), the algorithm couples feature selection with hierarchical clustering. Once the recurrent discriminant features are detected, co-actionable targets within each sub-cohort (sub-cluster) are selected through statistical assessment of significantly recurrent patterns and matching features to drugs using precision oncology knowledge bases* ***E. Selection of combinatorial drug targets.*** *The co-occurring combinatorial drug targets can be selected with either a forward or reverse selection protocol. In the forward selection, the primary target is the genomic aberration ("master biomarker", e.g., BRCA mutation) that defines the cohort. The secondary targets are statistically significant recurrent, actionable aberrations observed in fractions of the patients within the cohort. In the reverse selection protocol, the primary target is a recurrent aberration that is searched across diverse patient cohorts each defined by a genomic alteration. The secondary targets are the genomic alterations that define the cohorts, in which recurrent aberrations associated with the primary target are detected.*

Results

REFLECT-based signatures of oncogenic co-alterations

We generated a map of recurrent and co-actionable oncogenic alterations using the REFLECT pipeline (Figure 1). The analysis benefited from mutation, copy number, mRNA expression and phosphoproteomic analysis of > 10,000 patients from 33 cancer types (TCGA) and 1658 cell lines from 34 lineages (CCLE, COSMIC cell lines project (COSMIC-CLP) and MDACC cell line project (MCLP)) (Figure 2, S1 and Table S1). To generate a meta-cohort with no batch effects, we used the protein and mRNA expression data from diverse cancer types (TCGA) that had been integrated and batch normalized with a replica based normalization (RBN) and EB++ methods[14,15]. We also incorporated molecular profiling data from cell lines using a Z-score based normalization (see Methods). The integration of cell line data enables proof-of-principle experimental validations for matching drug combinations to specific signatures in molecularly matched lines. Next, we partitioned the cell line and patient pan-cancer meta-cohort into patient cohorts (stratified dataset), each defined by the existence of a "master" biomarker that is shared by all members of the cohort and potentially predictive of responses to targeted cancer therapy. The master biomarkers include therapeutically actionable genomic alterations such as those listed in the MDACC Precision Oncology Knowledge Base (N=180), MSKCC OncoKB resource (N=42) and NCI-MATCH precision therapy trials (N=15)[16,17,18]. We have defined and added an immunotherapy cohort based on PDL1 overexpression and immune cell infiltration, two potential markers of immunotherapy response in diverse cancer types. An additional three cohorts were constructed based on the existence of key oncogenic alterations, MYC amplifications, and TP53 and KRAS mutations. In total, we generated 199 biomarker-based cohorts. As provided in OncoKB, at least 25 of the "master" biomarkers are FDA-recognized and 14 biomarkers are NCCN-recommended as predictive of response to targeted therapy in clinical applications. The remaining biomarkers are under evaluation in precision oncology trials at MDACC and elsewhere.

For each biomarker-based cohort, we generated the multi-omic/multi-gene signatures, termed as REFLECT-signatures, using the sparse hierarchical clustering algorithm[12]. The algorithm iterates LASSO-based feature detection and hierarchical clustering until convergence to a final signature. The signature is defined by a small set of molecular features with a non-zero contribution to discrimination of sub-clusters (patient sub-cohorts). The algorithm quantifies the

SampleID	Type	PartitionMarkerStatus	Protein	Weight	RPPA_Zscore	Recur_pval
TCGA-FG-A4MW-01	LGG	EGFR(G312W)	HER2_pY1248	0.083	3.869	8.86e-14
TCGA-FG-A4MW-01	LGG	EGFR(G312W)	PKCDELTA_pS664	0.014	5.102	1.02e-13
TCGA-FG-A4MW-01	LGG	EGFR(G312W)	EGFR_pY1173	0.416	7.245	1.1e-12
TCGA-FG-A4MW-01	LGG	EGFR(G312W)	SRC_pY416	0.032	3.242	1.29e-12
TCGA-FG-A4MW-01	LGG	EGFR(G312W)	JNK_pT183Y185	0.016	2.604	2.77e-11
TCGA-FG-A4MW-01	LGG	EGFR(G312W)	EGFR	0.177	5.545	2.15e-09
TCGA-FG-A4MW-01	LGG	EGFR(G312W)	NOTCH1	0.020	2.605	8.5e-08
TCGA-FG-A70Z-01	LGG	EGFR(A289V,Q1143*)	EGFR	0.177	1.455	0.00106
TCGA-FG-A70Z-01	LGG	EGFR(A289V,Q1143*)	PKCDELTA_pS664	0.014	3.218	0.00114
TCGA-FG-A70Z-01	LGG	EGFR(A289V,Q1143*)	NOTCH1	0.020	1.695	0.00119
TCGA-FG-A70Z-01	LGG	EGFR(A289V,Q1143*)	JNK_pT183Y185	0.016	1.843	0.0064
TCGA-FG-A70Z-01	LGG	EGFR(A289V,Q1143*)	EGFR_pY1173	0.416	0.267	0.209

Figure 2. REFLECT resource and interface. A. The REFLECT resource *benefits from a pan-cancer meta-cohort of cancer patient samples and cell lines from CCLE, COSMIC-CLP and MCLP.* ***B. The pan-cancer meta-cohort*** *is stratified into 199 sub-cohorts based on the existence of actionable master biomarkers. The functional GO-term enrichment of the master biomarkers.* ***C. REFLECT resource outputs*** *the distributions of disease types in each cohort, the most discriminant features (DNA, RNA, protein) across each cohort, the REFLECT signature defined by the distribution of most discriminant features and selection of co-actionable targets based on the master biomarker and the REFLECT signature.*

contribution of each feature to patient clustering with a weight (w) term for the discrimination strength of the feature. The resulting signatures with weighted-features discriminate and cluster the patients into sub-cohorts identified by a combination of the master biomarker and REFLECT-generated combinatorial signature. To demonstrate the robustness of the REFLECT, we quantified the concordance between REFLECT signature features that were calculated with complete and subsampled partial data with varying coverage (60% to 95%) (Suppl Figure S4, S5). The concordance between partial and complete datasets increased monotonically with increased coverage such that concordance reaches an average of 83% for 70% coverage for the 8 patient cohorts tested (Figure S5). In summary, we have acquired a compendium of robust multi-omic REFLECT signatures spanning the patient cohorts identified by potentially co-actionable “master” biomarkers and sub-cohort specific features selected by our algorithm.

REFLECT signatures inform the selection of recurrent and co-actionable targets

Next, we extracted the recurrent and co-actionable alterations from the molecular signatures of patient cohorts. In each cohort, we first determined the significantly recurrent events from the REFLECT signatures with a permutation-based statistical test (see Methods). The statistical test filters out the features whose observed recurrence is likely by chance in a large patient cohort and preserves the ones that are significantly recurrent, and thus potentially functional. Next, we identified the patient sub-cohorts that carry recurrent and potentially co-actionable events. The co-actionable targets involve (i) the "master biomarker" that defines the cohort and (ii) co-occurring actionable biomarkers that are recurrent within a sub-cohort detected by REFLECT. To select actionable mutation and copy number events across sub-cohorts, we matched the oncogenic alterations in REFLECT-signatures to the MDACC and OnkoKB precision oncology knowledge bases. To match protein and mRNA expression aberrations to therapeutic agents, we developed a preclinical resource based on the drug target information from MDACC Precision Oncology Knowledge Base, OncoKB, GDSC and CTRP drug-target annotations as well as improvements through literature and expert curation (Suppl. Table S6 and S7)[19,20]. The core assumption in the mRNA/proteomics-based actionable event selection is that overexpression of a drug target predicts potential sensitivity to the respective targeted agent. For example, the increased phosphorylation and total levels of signaling molecules in key oncogenic pathways (e.g., AKT/PI3K/mTOR, RTK, MAPK, cell cycle, DNA repair pathways) are linked to their respective inhibitors (Table S6). Our resulting proteomics and transcriptomics based resource match 82 proteomic and 89 mRNA aberrations to targeted therapy options.

The REFLECT analysis led to 1356 and 614 unique drug combinations matched to patient sub-cohorts with recurrent features (P-recurrence < 0.05) based on mutation and copy number alterations, respectively. For protein and mRNA expression aberrations (Z-expression > 2 categorized as an aberration, P-recurrence < 0.05 categorized as recurrent), 1034 and 994 unique combination targets matched to FDA-approved drugs, respectively. With the current Z-expression and P-recurrence thresholds, 31% and 27% of patients with diverse cancers in the meta-cohort are matched to at least one combination therapy involving FDA-approved agents for protein and mRNA analyses, respectively. This is a subjective quantity that could be calibrated with adjusting the Z-threshold and requires extensive validation in preclinical and clinical settings for each aberration. The current resource, however, serves as an initial reference point for matching recurrent expression aberrations to potential therapies. The outcome is a REFLECT-based, comprehensive precision combination therapy resource that matches drug combinations to patient groups defined by the multi-omic, recurrent and co-actionable signatures.

The immunotherapy response signatures are enriched with inflammatory signals and actionable DNA repair pathway markers

A key PD1/PD-L1 inhibitor response predictor is high PDL1 expression in tumor cells and immune cell infiltration into the tumor microenvironment. To nominate precision combination therapy markers that may improve responses to PD1/PD-L1 inhibitors, we determined the multi-omic signatures and co-occurring actionable targets within a cohort that carry likely responders to immunotherapies (Figure 3). The potential immunotherapy responders are selected based on high PD-L1 levels as well as high tumor infiltrated leukocyte (TIL) fraction. To establish a robust PD-L1 high cohort, we included the tumors whose PD-L1 expressions are in the top quartile across all the pan-cancer meta-cohort for both mRNA and protein levels. Next, to determine tumors with high immune content, we included the samples in the top quartile of TIL fraction across all PD-L1 high tumors. The TIL fraction was quantified with a mixture model of tumor subpopulations based on methylation probes with the greatest differences between pure leukocyte cells and normal tissue. The resulting high PDL1/TIL cohort, that we call the

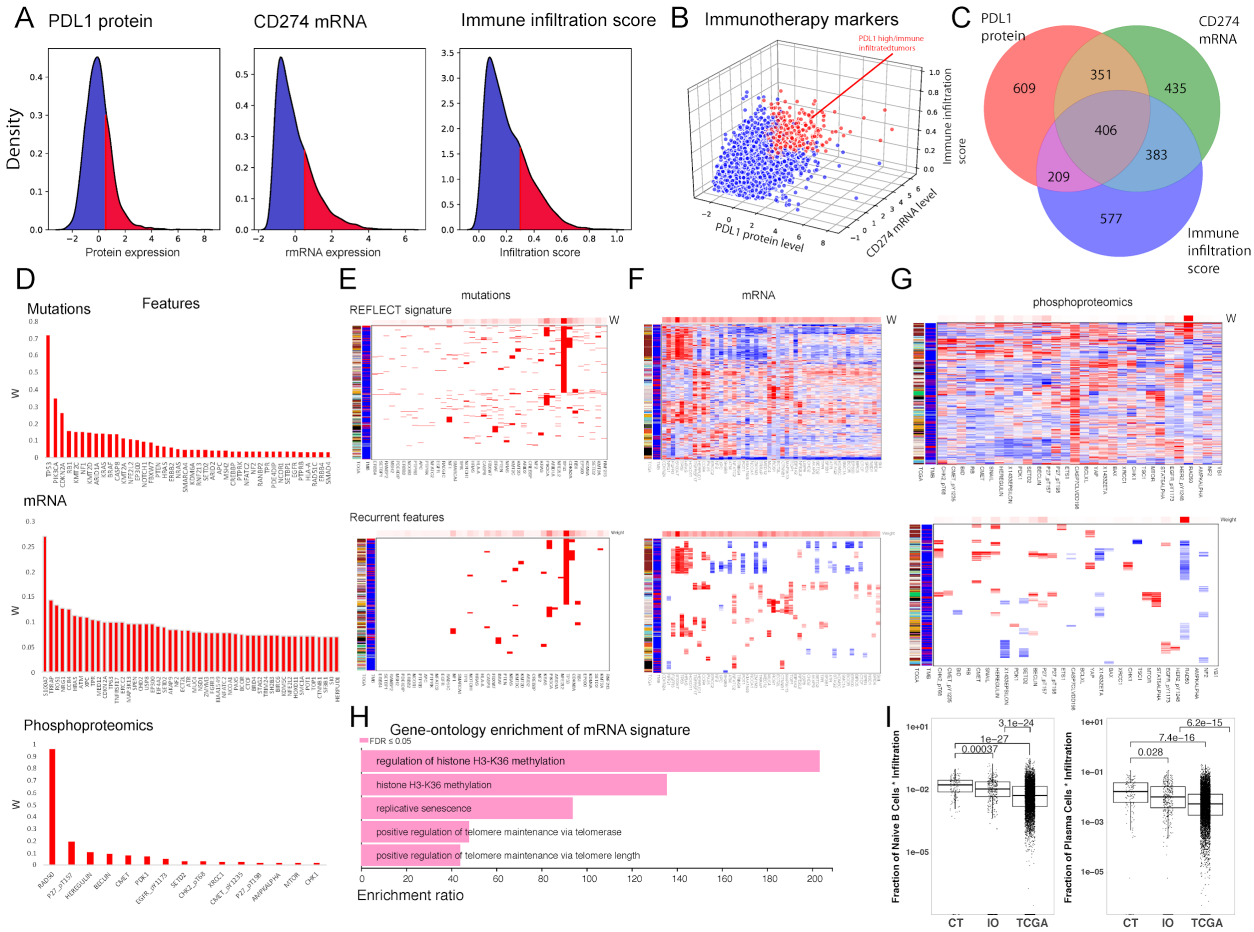

Figure 3. REFLECT analysis of the immunotherapy response marker (IO) cohort. A-B. *The patients carrying immunotherapy response markers are selected by the inclusion of cases within the intersection set of high (>μ+σ) PDL1 protein expression, CD274 (PDL1 gene) mRNA expression and immune infiltration.* ***C.*** *The intersection of IO markers led to a patient cohort of 406 patients.* ***D****. The most discriminant DNA, mRNA and protein features based on REFLECT-analysis in the IO cohort.* ***E.*** *The REFLECT signatures (top) and significantly recurrent features (bottom) for mutations.* ***F.*** *mRNA expression and* ***G.*** *phosphoproteomics.* ***H.*** *The cellular functions that are significantly enriched in the IO mRNA REFLECT signature as detected by functional enrichment analysis.* ***I.*** *The statistically significant enrichment of naive B-cells and plasma cells in the CCR4 high/TNFSFR17 (CT) high sub-cohorts within the IO cohort (all IO cases including CCR4 high/TNFSFR17 high sub-cohort).*

immunotherapy cohort, contains 394 patients from 23 disease types (Table S12). We then generated the REFLECT-based significantly recurrent mutational, mRNA and phosphoproteomic signatures co-occurring with high PDL1/TIL. An analysis of PD-L1 expression and infiltrating lymphocytes led to a patient cohort with >90% overlap and resulted in virtually identical signatures to those from infiltrating leukocytes.

The mutation analysis of the immunotherapy cohort identified a signature of eight genes (w > 0, P-recurrence < 0.05) covering the commonly mutated oncogenes in cancer. The most recurrent signature genes discriminating patient sub-cohorts are the TP53 (50%, P-recurrence < 0.001), PIK3CA (15%, P-recurrence < 0.001) and CDKN2A (10%, P-recurrence < 0.001) followed

by RB1, BRAF, KRAS. The mRNA REFLECT signature of the immunotherapy cohort is characterized by a combination of significantly recurrent events (w > 0, P-recurrence < 0.05) in immune-associated proteins as well as DNA damage response and repair pathway markers. The strongest contributing mRNA feature is the overexpression of S100A7 (w=0.27, P-recurrence < 0.0001). In cancer context, S100A7 protein functions in the tumor microenvironment with roles in inflammation-driven carcinogenesis and recruitment of immune cells including tumor-associated macrophages (TAM) and highly expressed in a series of cancers including head and neck cancers[21,22,23]. The S100A7 enriched sub-cohorts also carried upregulation of oncogenic genes including overexpression of TP63, HRAS, FGFR3 particularly in head and neck as well as bladder and squamous lung cancers. Another patient sub-cohort populated by lung squamous cancer cases was enriched with a co-occurring expression of the immune chemotactic receptor CCR4, which is associated with infiltration by diverse lymphocytes and TNFRSF17, a TNF receptor family member and marker of B-cells[24,25]. Indeed, we demonstrated, based on a CIBERSORT analysis, that the tumors in the high CCR4/TNFRSF17 expression sub-cohort are significantly enriched with naive and plasma B-cell markers (Figure 3I).

The transcriptomic signature is also enriched by chromosome organization, cell cycle checkpoint and DNA repair genes such as *TRRAP, CDKN2A, ATM, XPC, ERCC2, CHD2, SETD2, ATR, NSD1, BRD4 and ZYMIM3.* A functional enrichment analysis using the WEB-based GEne SeT AnaLysis Toolkit' (WebGestalt)[26] revealed highly significant enrichment of the gene sets involved in histone methylation (FDR-p<0.001) as well as DNA repair associated processes, replicative senescence (FDR-p=0.03) and telomere maintenance (FDR-p=0.01). Similarly, in the phosphoproteomics analysis, depletion of RAD50, a key enzyme for repairing the DNA double-strand breaks, has the highest contribution to discriminate the immunotherapy sub-cohorts with enrichment in two patient sub-cohorts. The first RAD50 depleted sub-cohort (N=111, P-recurrence < 0.0001) is populated by diverse cancer types while the second sub-cohort (N=48, P-recurrence < 0.0001) is enriched by bladder cancers. Increased CHK1 levels (P-recurrence < 0.0001, N=25) and CHK2 phosphorylation at T68 (P-recurrence < 0.0001, N=52), a marker of DNA damage response and checkpoint, also co-occurred within a fraction of the RAD50 depleted sub-cohort, suggesting aberrant DNA repair and damage in such tumors. The enrichment of DNA repair and checkpoint aberrations suggests potential benefit from inhibitors that target individual repair enzymes, particularly the PARP inhibitors that are synthetically lethal with such aberrations. In the context of emerging PARP and PD-L1 inhibitor trials, the sub-cohort specific co-occurrence patterns may guide the selection of patients that are likely to benefit. Such co-occurrence patterns warrant further preclinical testing for novel combination therapy trials that co-target the aberrant DNA repair proteins listed here and PD-L1/PD-1 immune checkpoints.

Recurrent co-alterations within DNA repair pathways nominate PARP-inhibitor based combination therapies

PARP inhibitors have provided significant therapeutic benefit particularly for a fraction of cancer patients based on synthetic lethality with DNA repair defects. Here, we used the REFLECT pipeline to nominate combination precision therapies, in which PARP and other DNA repair inhibitors can serve as the anchor compound. To identify the recurrent and co-actionable events that co-occur with DNA repair defects, we focused on patient cohorts with alterations (master biomarkers) in 22 DNA repair genes that are implicated in BRCAness and that are potentially synthetically lethal with PARP (or other DNA repair enzyme) inhibition as listed in the MDACC Precision Oncology Knowledge Base (Figure 2, Table S8). First, we generated the multi-omic signatures for each of the alterations in the master biomarkers of DNA repair, and next, selected potential combination therapies using the forward drug selection criterion (Figure 1E, Figure S3A, also see Methods).

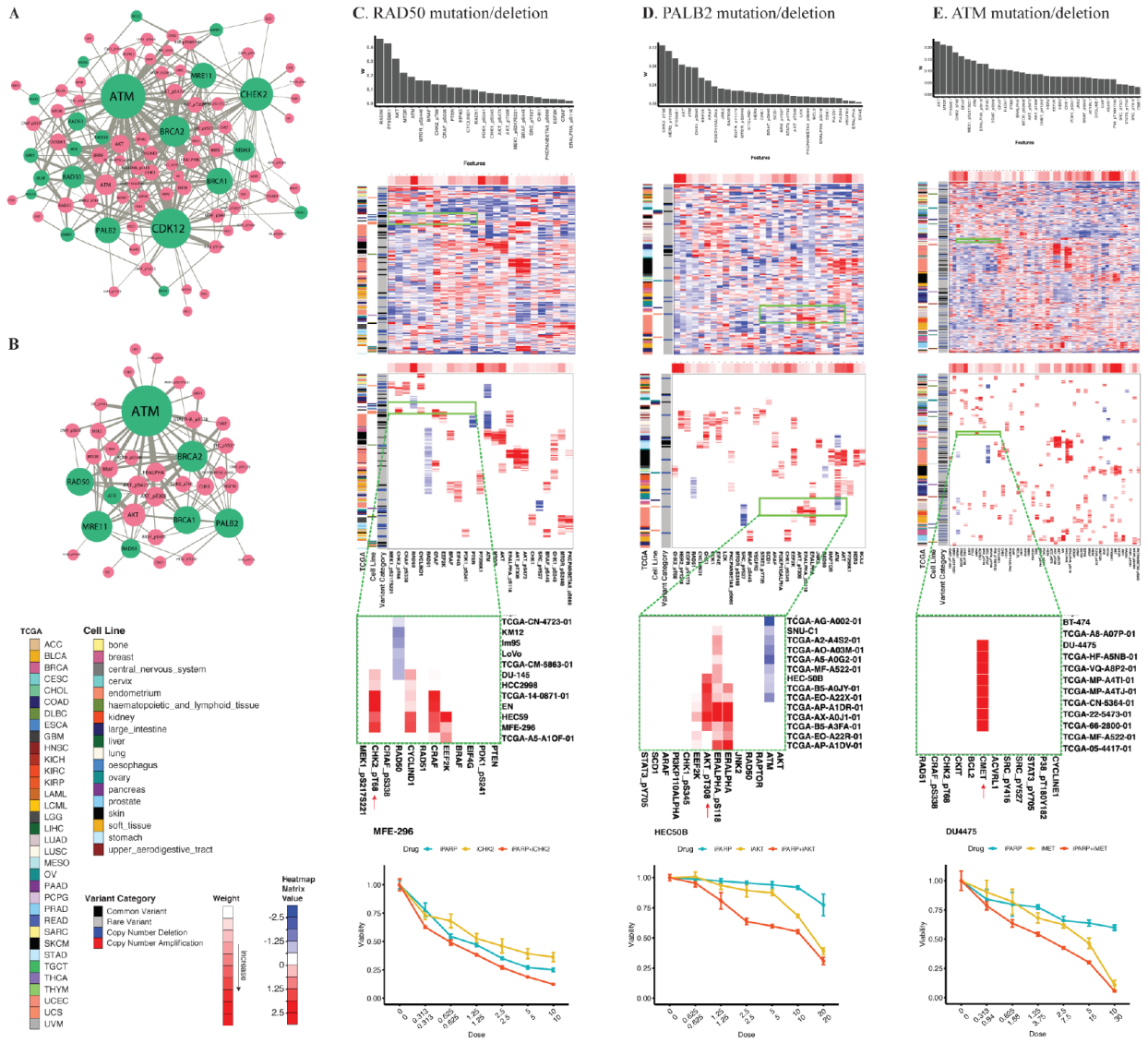

Figure 4. Combination targets for DNA repair deficiencies. A. *A bipartite graph showing the association between the DNA repair genes (master biomarker, green) and actionable protein targets (REFLECT detected, pink). In the graph, the size of a node is proportional to the degree of the node, and the width of an edge is proportional to the number of patients matching the combination pair.* ***B.*** *The association between 'core' repair genes and actionable proteins matched to FDA-approved drugs.* ***C - E*** *are representative combination targets of PARP and another drug, where the genomic stratification markers are mutation/deletions of RAD50, PALB2, and ATM, respectively. The top row shows weights of features during sparse hierarchical clustering, the second row displays heatmaps of protein expression (Z scores) after the clustering, the third row displays heatmaps of recurrently up/down-regulated protein expressions (Z scores) after the clustering (p-value < 0.25), the fourth row zooms in a specific region with recurrent patterns and matched cell line, and the bottom row shows proof-of-principle validations. The matched cell lines are treated with drugs corresponding to the combination targets and profiled for viability. In the bar charts and heatmaps, only non-zero actionable features are shown.*

We performed an integrated analysis of the actionable alterations paired with the master biomarkers across all DNA repair cohorts on a bipartite graph. We analyzed the significantly recurrent, co-occurring and actionable proteomic features with DNA repair aberrations on the bipartite graph, in which each recurrent/co-occurring feature is linked to the master biomarker (i.e., the DNA repair aberration that defines each of the cohorts) (Figure 4A-B). We observed an enrichment of aberrant expression in DNA repair proteins (RAD51, ATM, CHK1/2), RTKs (p-EGFR, p-HER2, p-MET, KIT, VEGFR) and downstream signaling proteins (p-AKT, mTOR, p70S6K, p-JNK, p-SRC, pP38/MAPK14, p-PDK1, p-RAF, p-MAPK) that co-occur with the master biomarkers in the DNA repair family.

Guided by the graph enrichment of DNA repair, RTK and signaling aberrations with the master biomarkers, we experimentally co-targeted the PARP based on master biomarkers and co-occurring actionable aberrations on REFLECT signatures (Figure 4D). In drug perturbation experiments, we used the cell lines molecularly matched to (co-cluster with the) specific patient sub-cohorts carrying the targeted co-alterations (Figure 4). We prioritized molecularly matched gynecological and breast cancer patient sub-cohorts and lines as PARP targeting has demonstrated strong but transient therapeutic benefit for a fraction of high grade serous ovarian and basal-like breast cancers. Using the forward drug selection (Figure 1E and S3A) in the RAD50, PALB2 and ATM altered cohorts, we identified sub-cohorts characterized by statistically significant recurrent Chk2 phosphorylation, AKT phosphorylation and MET expression, respectively (P-recurrence < 0.05). For each sub-cohort, we identified a cell line carrying the similar co-occurrence, namely MFE-296 (RAD50 mutation with recurrent Chk2 phosphorylation), HEC50B (PALB2 mutation with recurrent AKT phosphorylation), and DU4475 (ATM mutation with recurrent MET overexpression). The drug combinations involving PARP inhibitors and the corresponding compounds targeting the recurrent actionable aberrations in patient sub-cohorts produce additive to synergistic effects in each of the tests providing a proof-of-concept validation that such co-occurrences may have beneficial outcomes.

Combination precision therapies for HER2 activated tumors

Therapeutic targeting of HER2 has improved clinical outcomes particularly in HER2+ breast cancers yet resistance to therapy remains a major clinical problem. We used the REFLECT pipeline to identify patient cohorts that may benefit from combination therapies involving HER2 inhibitors (Figure 5). Based on a reverse search (Figure 1E, S3B) across REFLECT generated signatures, we identified patient sub-cohorts carrying actionable oncogenic aberrations (master biomarkers) and increased HER2 total or phosphorylation (p-Y1248) levels that may predict benefit from HER2 inhibitors. In the reverse search strategy, we scanned proteomic signatures within 199 cohorts, each defined by the actionable biomarkers. Within the cohorts, we detected the sub-cohorts that carry significantly recurrent (P-recurrence < 0.05) HER2 activity markers. The analysis led to the identification of 83 sub-cohorts that carry one of the markers of the HER2 activation (P-recurrence < 0.05) alongside the actionable master biomarker. As a result, 353 patients (5% of the TCGA meta-cohort) were matched to 1048 unique combinations involving HER2-targeting (Z>2, P-recurrence<0.05) (Table S10). For proof-of-principle validation experiments, we identified cell lines that map to sub-cohorts of increased HER2 activity. The molecularly matched cell lines were sampled within the cohorts of (i) AURKA mutation/amplification (cell line: BT474), (ii) MAP3K4 mutation/amplification (cell line: SKOV3), (iii) CDKN1B mutation/loss (cell line: HCC1954), (iv) SRC amplification (cell line: SKBR3), (v) PTEN loss/mutation (cell line: MDAMB468). The cell lines were perturbed with HER2/EGFR inhibitors and agents targeting the cohort biomarkers (Figure 5). The drug combinations targeting the co-alterations substantially improved responses over every single agent.

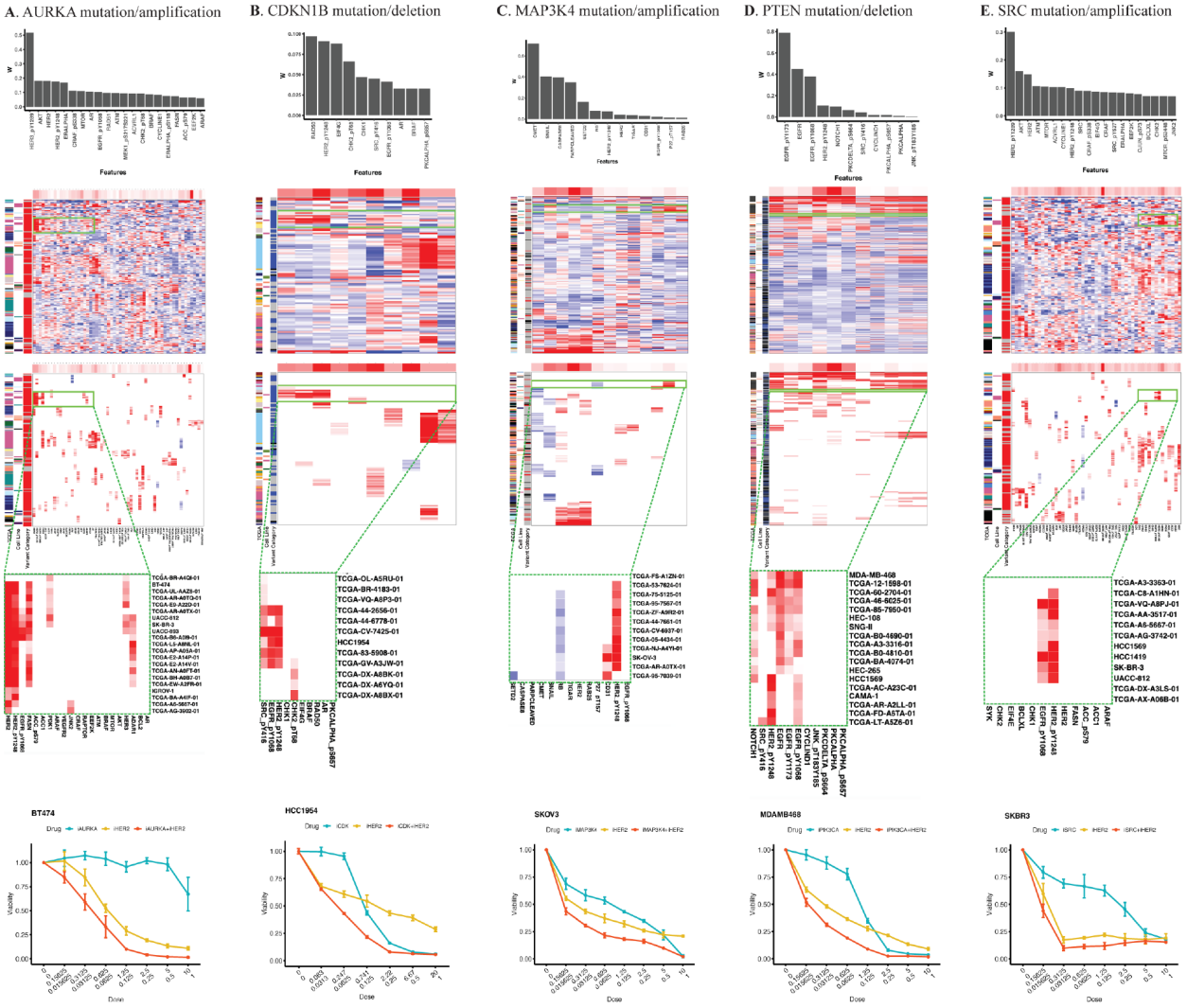

Figure 5. *Co-targeting in HER2 activated tumors. Combined targeting of HER2 activation with* ***A.*** *AURKA mutations,* ***B.*** *CDKN1B loss,* ***C.*** *MAP3K4 mutations,* ***D.*** *PTEN loss,* ***E.*** *SRC activation with a combination of HER2/EGFR inhibitor, neratinib and inhibitors of AURKA, CDK4, JNK1/2, AKT and SRC, respectively. The top row shows weights of features from sparse hierarchical clustering, the second-row displays REFLECT signatures, heatmaps of protein expression (Z scores) after the clustering, the third row displays heatmaps of recurrently up/down-regulated protein expressions (Z scores) (p-recurrence < 0.25), the fourth row zooms the region with the selected recurrent pattern (e.g., HER2/HER2_pY1248 upregulation) and matched cell line. The last row shows proof-of-principle cell viability tests. The matched cell lines are treated with REFLECT-selected drugs. In the bar charts and heatmaps, only non-zero actionable features are shown, except for column C (MAP3K4), where all non-zero features are displayed.*

Discussion

Here, we provide a bioinformatic analysis method, REFLECT to nominate combination therapies tailored to combinatorial aberration signatures in patient cohorts. The REFLECT includes a machine learning algorithm for the selection of recurrent co-alterations within distinct patient sub-cohorts and benefits from precision medicine knowledge bases to match the co-

alterations to therapies. With a REFLECT-based analysis, we have generated a resource that identifies combination therapies matched to > 10,000 patients across > 200 patient cohorts each defined by the existence of an actionable target. To demonstrate the capabilities of REFLECT, we focused on signatures of three clinically relevant events; potential immunotherapy response markers, DNA repair and damage response aberrations, and HER2 activation. The analysis of HER2 and DNA repair aberrations has led to the determination of patient sub-cohort specific combination therapy nominations that we validated in matched cell lines. The experimental validations revealed either novel drug synergies such as co-targeting Aurora kinase and HER2 in characterized patient groups or established the potentially benefiting patient cohorts for previously tested drug combinations such as those co-targeting PARP with Chk2[27], AKT[28,29] or MET[30] and HER2 with CDK4[31]and PI3K[32]. We have also determined enrichment of specific inflammation and immune cell (B-cells) markers as well as DNA repair alterations within the patient cohorts characterized by high PDL1 expression and immune infiltration, the two biomarkers of immunotherapy response[33]. The co-enrichment of DNA repair/damage response and immunotherapy response markers establishes a rationale for testing PD1/PDL1 inhibitors in combination with DNA checkpoint and repair inhibitors[34].

REFLECT benefits from a set of core principles in cancer genomics and therapeutics. First, oncogenesis and cancer progression are driven by multiple co-existing driver aberrations[11,35,36]. Such co-existing aberrations may be causally coupled to each other. Alternatively, co-alterations may drive parallel and relatively independent events with no trivial causal link in the same or separate tumor sub-clones, yet, may contribute to cancer progression in an additive manner. Second, oncogenic alterations that are significantly recurrent across patient cohorts are likely functional as such recurrence across samples suggests oncogenic selection to drive oncogenic hallmarks[37]. Finally, the success of the REFLECT algorithm relies on the observation that recurrent co-alterations that co-activate multiple oncogenic processes may prevent benefit from monotherapies[8]. The REFLECT nominates the drug combinations that may block such co-activated processes more effectively than monotherapies in a matched patient sub-cohort. Finally, precision therapies are usually matched to functional variants of an oncogene (e.g., BRAFV600E). Here, when selecting potentially actionable mutations, we have not filtered out the variants of unknown significance (VUS) and included all variants of a master biomarker if the variant exists in more than 2 patients in the meta-cohort. This approach involves the risk that some passenger mutations that recurred in more than 2 patients are included in patient stratification and subsequent analysis. Yet, we gain the overriding advantage that we detect the REFLECT signatures, particularly in mRNA protein analysis, shared between known functional/actionable variants and select VUS for master biomarkers. Such overlaps in the REFLECT signatures between known functional variants and VUS may guide therapeutic options to target potentially important VUS.

The contributions from bioinformatics tools have become more needed as multi-omic data-driven clinical studies such as the WINTHER trial are emerging[38]. Proteomics-based markers that indicate activation of actionable oncogenic pathways can also complement genomics-based patient selection and mutation-based drug sensitivity prediction. The proteomic modality is less explored and provides a major opportunity as most targeted agents impact protein activities to mediate the therapeutic effects. For example, increased EGFR and HER2 activating phosphorylation levels is a predictive marker for the EGFR/HER2 inhibitors while nearly half of HER2/EGFR-mutated tumors do not lead to high levels of EGFR/HER2 activating phosphorylations. Proteomic changes, however, may not serve as the sole criterion for drug selection, at least due to the current regulatory limitations around the use of proteomics for clinical tests but is an important complement to other data types. By assessing multiple data modalities at once, we expect REFLECT and tools will improve the fraction of benefiting patients while providing more reliable target matching in preclinical and data-guided clinical trials. In this respect,

the stratification into patient cohorts based on actionable biomarkers emulates the ongoing basket trials such as NCI-MATCH with the added value that it enables matching combination therapies to sub-cohorts with shared molecular compositions.

REFLECT is an informatics platform for therapy selection guided by multi-omics (mutation, transcriptomics, proteomics) profiling in preclinical and clinical trials. As lack of durable responses to genomically matched monotherapies significantly limits the population of benefiting cancer patients, REFLECT is potentially a critical addition to the current precision medicine paradigm. With expanding data from precision therapy trials and genomics studies, it will become increasingly feasible to match combination therapies to complex molecular signatures. We expect REFLECT will be a commonly used tool and resource to select precision combination therapies in both preclinical and clinical studies.

Acknowledgment.

We thank Scott Kopetz, Funda Meric-Bernstam, Scot Woodman, Gordon Mills for invaluable comments. This work is supported by grants from MDACC Support Grant P30 CA016672 (the Bioinformatics Shared Resource), OCRF Collaborative Research Award, ICI Fund, CPRIT High-Impact/High-Risk Award (RP170640).

Methods

Multi-omic Data

Phospho-proteomics data of TCGA patient cohorts and cancer cell lines are available from the website of The Cancer Proteome Atlas Portal (https://tcpaportal.org/). We obtain level 4 data (batch normalized across disease types) for both patient cohorts and cell lines. After filtering out samples having missing values, the TCGA patient dataset has 7544 samples from 32 cancer types, and the cell line dataset has 397 samples from 18 lineages (MCLP). Therefore, each dataset covers a large number of samples from diverse cancer types. Because of their diverse coverages, we assume that the TCGA patient dataset and the cell line dataset follow the same distribution. Considering they are measured from different batches, to remove the batch effect and merge them together, we standardize each dataset using Z-score transformation so that each has zero mean and unit variance. Then, the two standardized datasets are combined to form an integrated phospho-proteomics dataset. For mRNA expression data, we integrate TCGA patient RNA-seq data and CCLE cell line RNA-seq data. Similar integration strategy used in the phospho-proteomics data, we standardize the log2-transformed mRNA expressions in each dataset and then combine them together. The mutation and copy number data are from TCGA patient cohorts and the COSMIC-CLP. Mutations are discretized to binary values, either 1 indicating mutated or 0 denoting wide-type. Copy numbers are discretized to {-1, 0, 1}, representing deletion, no alteration, and amplification, respectively. Without the need of correcting batch effects for discretized mutations and copy numbers, TCGA patient data and cell line data are combined together directly. For each of the above four data types, the distributions of samples across cancer types or lineages are shown in Figure S1.

Genomic Stratification Markers

We select actionable genomic alterations as the stratification markers from two sources. One source is NCI-MATCH (Molecular Analysis for Therapy Choice), from which we select 15 genomic vulnerability markers. Another source is the MD Anderson Precision Oncology Decision Support (PODS). There are 207 potentially actionable alterations in the list for the PODS team at MD Anderson Cancer Center. From this list, we extract 180 genomic alteration markers

(belonging to 179 genes). The remaining alterations were already included in the NCI-MATCH. For each marker, the alterations include mutations, gene amplification/deletions, specified amino acid substitutions, e.g., BRAF V600 mutation. Depending on specific genes, activating mutations are combined with amplification, and inactivating mutations are combined with gene deletions. Additionally, we constructed three cohorts based on the existence of key oncogenic alterations - MYC amplifications, and TP53 and KRAS mutations. The current marker list also covers the actionable genes listed in OncoKB. All the genomic stratification markers are listed in Table S5. Annotations based on GO Slim terms are used to categorize the markers. The top three terms are used; terms are ranked by evidence types and counts as provided by Gene Ontology. Code is provided at https://github.com/cannin/extract_go_slim_annotation

Sparse Hierarchical Clustering

Sparse hierarchical clustering identifies clusters driven by the most discriminant features as follows. The input to the algorithm is X, an $n \times p$ data matrix with n observations (e.g., patients) and p features (e.g., genes). x_i is an observation that is a p-dimensional vector indexed by i, where $1 \leq i \leq n$, and $d(x_i, x_{i'})$ is a measure of the distance between observations x_i and $x_{i'}$. We assume that the distance measure is additive in features, that is, $d(x_i, x_{i'}) = \sum_j^p d_{i,i',j}$, where $d_{i,i',j}$ is the distance measure between observations x_i and $x_{i'}$ along feature j.

The distance is defined as either the squared Euclidean distance $d_{i,i',j} = \left(X_{ij} - X_{i'j}\right)^2$ or the absolute difference $d_{i,i',j} = \left|X_{ij} - X_{i'j}\right|$. The quadratic Euclidean distance, due to the square operation, assigns a relatively higher weight to larger values, while the absolute difference treats all values uniformly. In REFLECT, to smoothen small differences, the squared Euclidean distance is employed as the distance measure for the continuous data (i.e., RPPA and mRNA expression). For the discrete data (i.e., mutation and copy number alterations), only a small finite set of values exists, that is, {-1, 0, 1}, so we use the absolute difference as their distance measures.

The sparse hierarchical formulation is derived from the standard hierarchical clustering equations. In the standard hierarchical clustering, we solve the constrained maximization problem

$$\underset{U}{\text{maximize}} \left\{ \sum_j \sum_{i,i'} d_{i,i',j} U_{i,i'} \right\},$$

$$\text{subject to} \sum_{i,i'} U_{i,i'}^2 \leq 1,$$

which leads to the optimal solution $U_{i,i'}^* \propto \sum_j d_{i,i',j}$.

The iterative sparse hierarchical clustering is formulated as the solution to a weighted constrained optimization problem

$$\underset{w,U}{\text{maximize}} \left\{ \sum_j w_j \sum_{i,i'} d_{i,i',j} U_{i,i'} \right\},$$

$$\text{subject to} \sum_{i,i'} U_{i,i'}^2 \leq 1, \qquad \|w\|^2 \leq 1, \qquad \|w\|_1 \leq s, \qquad w_j \geq 0 \ \forall j,$$

where w_j is a weight for feature j, and s is a tuning parameter that controls sparsity of w. Let U^{**} optimize the above criterion, from which we have $U_{i,i'}^{**} \propto \sum_j w_j d_{i,i',j}$, the distance along each

feature is weighted according to w_j. And due to the constraints, a small value of s makes w sparse, so U^{**} depends only on a small subset of features. Once the w is optimized, hierarchical clustering is performed on U^{**} . The operation is iterated until convergence (w is fixed; U** is optimized and next U** is fixed; w is optimized), w and hence the sparse clustering. The output is a sparse clustering of observations defined by the distribution of a small set of most discriminant features. In REFLECT, we employ the sparcl R package (implemented by Witten and Tibshirani) to execute the sparse hierarchical clustering.

Selection of Tuning Parameter via the Gap Statistic

In the sparse hierarchical clustering, the tuning parameter s, which controls sparsity of w, needs to be determined prior to clustering. We apply a permutation approach called the gap statistic to select an appropriate value for s. For a specific tuning parameter s, the gap statistic measures the strength of the clustering on real data compared with that expected on a reference null distribution. The computational implementation is as follows.

Step 1: Generate B reference datasets $\{X_1, \cdots, X_B\}$ by permuting the observed dataset X independently within each feature.

Step 2: For each candidate of s, where $1 \leq s \leq \sqrt{p}$:

(i) Obtain the optimal objective $O(s) = \sum_j w_j \sum_{i,i'} d_{i,i',j} U_{i,i'}$ by performing the sparse hierarchical clustering on the data X.

(ii) For $b = 1, \cdots, B$, obtain the optimal objective $O_b(s)$ by performing the sparse hierarchical clustering on the data X_b.

(iii) Compute the gap statistic, the deviation between log objective function of clustering with actual data and mean of the log objective function from the null distribution $\mathrm{Gap}(s) = \log O(s) - 1/B \sum_{b=1}^{B} \log O_b(s)$.

Step 3: Use the following optimization algorithm to select the optimal s^* that maximizes $Gap(s)$:

(i) Search for an s_{min} such that when $s \geq s_{min}$, the number of non-zero features from sparse clustering is greater than N_{cutoff}. Here, we choose $N_{cutoff} = 10$ to prevent the selection of an extremely sparse set of features, which may not represent the complexity of the dataset.

(ii) Find s_{max} that corresponds to the maximum gap statistics.

(iii) Calculate the mean $E[Gap(s)]$, the standard deviation $SD[Gap(s)]$, and the standard error of the mean $SEM[Gap(s)]$ for all $s$values in $[s_{min}, s_{max}]$.
(iv) Find s^*_{min} such that $Gap(s^*_{min}) = E[Gap(s)] - SD[Gap(s)]$.

(v) Calculate the mean $E[Gap(s')]$ for all s' in $[s^*_{min}, s_{max}]$. If $E[Gap(s')] > E[Gap(s)] - SEM[Gap(s)]$, let $s^* = s^*_{min}$; otherwise, let $s_{min} = s^*_{min}$ and return to the step (iii).

Robustness of Feature Selection

To evaluate the robustness of feature selection in REFLECT analysis, we employ a subsampling approach to quantify to what degree the top features from a subsampled dataset can recapture features obtained from its original dataset. Since this calculation is costly, we perform it only for several selected phosphoproteomics datasets that corresponds to our experimental validations described below, including AURKA mutations/amplifications, CDKN1B mutations/deletions, MAP3K4 mutations/amplifications, NF1 mutations/deletions, RAD50 mutations/deletions, PALB2 mutations/deletions, and ATM mutations/deletions.

For each dataset, REFLECT is used to get an ordered list of the top N discriminant features $L_{whole} = [F_{(1)}, \cdots, F_{(N)}]$, where we chose $N = min(20, |\forall w > 0|$. Then, we randomly subsample the dataset to obtain a specified fraction, f, where $0.6 \leq f \leq 0.95$ in 0.05 increments, for Bsamplings where $B = 200$. The sparse hierarchical clustering is performed on each subsampled dataset to extract an ordered list of top N^* discriminant features $L_{sub} = [F^*_{(1)}, \cdots, F^*_{(N^*)}]$, where we chose $N^* = 10$. A score can be calculated as follows by comparing features in L_{sub} that recapture the ones in L_{whole}:

$$\text{Score} = \frac{\sum_{i=1}^{N^*} \sum_{j=1}^{N} w_j I\left(F^*_{(i)} = F_{(j)}\right)}{N^*},$$

where $I(\cdot)$ is an indicator function, w_j represents a rank scaled weight of feature j for $1 \leq j \leq N$. The rank-scaling is formulated as

$$w_j = \begin{cases} 1, & j \leq N^*, \\ \dfrac{2N^* - j}{N^*}, & N^* < j \leq 2N^*, \\ 0, & \text{otherwise.} \end{cases}$$

where the top-ranked (N=10) features contribute most to the score while the intermediate ranked (10<N<21) features have contributions that are scaled-down in proportion to their ranks. In this scheme a high weight, high-ranked feature has a higher contribution to the robustness score compared to a low-weight, low-ranked feature. Finally, the robustness scores are visualized using box plots, showing the distribution of $Score$ over B subsampled datasets as a function of the subsampling fraction f.

Statistical Significance of Recurrence

In order to assess a p-value for recurrence, we need to calculate the probability of having at least k consecutive altered samples $P_{val}(N, k)$, from a total of N samples. A simple way to calculate this probability is by calculating its complement, $P_{val}(N, k)$, the probability of having less than k consecutive samples. Here we derive the equation: The p-value of k consecutive altered samples by chance can be calculated in the form

$$P_{val}(N, k) = 1 - P(N, k),$$

where $P(N, k)$ is the probability that we do NOT have at least k consecutive altered samples.

N is the total number of samples, and N_a is the total number of altered samples, where $N_a < N$.

Here, we derive an approximate formula for the calculation of $P(N, k)$ by assuming the probability of the alteration p is a constant, which holds true when k is much smaller than the number of altered samples N_a. Then, the probability,p , of observing an alteration in a sample is calculated empirically by

$$p = \frac{N_a}{N}.$$

The probability of a sample being unaltered, denoted as q, is

$$q = 1 - p.$$

Given a sequence of samples that do not have k consecutive altered samples, the end of a sequence must be an unaltered sample followed by i altered samples, where $0 \leq i < k$. Let $P(n, k, i)$ denote the probability that a sequence of length n has less than k consecutive altered samples AND ends with i consecutive altered samples. Thus, we have

$$P(n,k) = \sum_{i=1}^{k-1} P(n,k,i).$$

Notice that if we remove i samples from one of the borders of i consecutive altered samples in the class (n,k,i), the sequence belongs to the class $(n-i,k,0)$. Thus, we have

$$P(n,k,i) = p^i P(n-i,k,0).$$

Moreover, if we add one unaltered sample to the end of i consecutive altered samples in the class (n,k,i), the sequence is in the class $(n+1,k,0)$, which gives us the relationship

$$P(n+1,k,0) = qP(n,k).$$

By combing the above three equations, we have

$$P(n+1,k,0) = q\sum_{i=1}^{k-1} p^i P(n-i,k,0),$$

which gives us a linear iterative relationship to calculate $P(n,k,0)$. The initial condition is $P(n,k) = 1$ if $n < k$. Finally, we can get $P(N,k)$ expressed in the form

$$P(N,k) = \frac{1}{q} P(N+1,k,0),$$

which allows us to get the p-value using $P_{\text{recurrence}}(N,k) = 1 - P(N,k)$.

Development of mRNA and RPPA based actionable lists

To get a comprehensive list of actionable targets and their inhibitors, we combined drug-target annotations from 4 databases: (i) Genomics of Drug Sensitivity in Cancer (GDSC), (ii) Cancer Therapeutics Response Portal v2 (CTRP v2), (iii) Precision Oncology Knowledge Base OncoKB, and (iv) MDACC Precision Oncology Knowledge Base (Kurnit, 2018, Table 1). The combined list includes actionable targets that can be inhibited by FDA-approved drugs, drugs used in clinical trials, and preclinical tests. We defined the mRNA targets based on the assumption that mRNA overexpression of the target predicts sensitivity to the respective agent. Similarly, we obtained proteomics-based actionable targets from the RPPA-assayed proteins for both total protein expression and activating phosphorylation levels. For example, high levels of activating AKT phosphorylation is a target for AKT inhibitors. Finally, we manually corrected the generated list based on literature and expert knowledge.

Selection of Nominated Combination Targets for Experimental Validation

REFLECT is a computational pipeline to nominate combination targets, where the primary target is its genomic stratification marker, and the secondary target comes from one of the recurrent features identified from the sparse hierarchical clustering. To validate the performance of this computational framework, in this paper, we perform *in vitro* experimental tests on some of the combination targets nominated from (phosphor-)proteomics-based REFLECT analyses. We employ two complementary selection strategies for validation experiments of combination targets X + Y, where X represents the primary target from one of the genomic stratification markers, and Y is the secondary target from a recurrent (phosphor-)proteomic feature. The first strategy is called a forward selection, in which X is fixed. We use REFLECT to analyze samples sharing X alterations, and then Y is selected to be one of the recurrent proteomic features. On the other hand, in a reverse selection strategy, X is varied while Y is fixed. In this study, we designed two series of validation experiments corresponding to both forward and reserve selection strategies. For the forward selection experiments, we tested combination targets having the primary targets

corresponding to genomic alterations that suggest PARP vulnerability. These genomic alterations are mutations or deletions of “BRCAness” genes, including ATM, BRCA1, BRCA2, PALB2, BRIP1, RAD50, and RAD51. The secondary targets are recurrent features after a REFLECT analysis of each dataset based on those stratified genomic markers. For the reverse selection experiments, we used recurrently upregulated HER2/HER2_pY1248 as the secondary target. We run REFLECT analyses on all datasets based on different genomic stratification markers and selected primary targets to co-occur with recurrent HER2/HER2_pY1248 overexpression.

Cell lines and reagents

The cell lines used to test the combined effect of candidate targets include BT-474, HCC1954, MDA-MB-468, SK-BR-3, and SK-OV-3. The cell lines were obtained from MD Anderson Characterized Core Cell Line Facility and were authenticated by fingerprinting using short tandem repeat testing. The absence of mycoplasma contamination was also verified. Cells were maintained in RPMI-1640 supplemented with 10% heat-inactivated FBS and 1% penicillin with streptomycin.

Cell culture, growth inhibition, and combinatorial inhibitor effect analysis

Depending on the cell line doubling times, 2000-5000 cells were seeded into a 96-well plate 24 hours before treatment. The cell lines were plated and treated with two individual inhibitors and their combination, in triplicates, over six twofold dilutions. The cell lines were treated in triplicates for different concentrations of the inhibitors from 0 to 10 or 20μM in 6 two-fold dilutions (in 5% FBS) to generate a dose-response curve. The selected doses of each drug were added using Tecan D300e Digital Dispenser. Viability was assessed 72 hours post-treatment using PrestoBlue cell viability reagent (Thermo Fisher Scientific) by measuring fluorescence on BioTek Synergy H1 Hybrid Multi-Mode Reader according to the manufacturer’s recommendation. Background values from empty wells were subtracted and the data was normalized to untreated control. The log of drug concentration was plotted against the growth inhibition.

References

1. Ashley, EA. (2016). Towards precision medicine. Nature Reviews Genetics, 17(9), 507-522. https://doi.org/10.1038/nrg.2016.86

2. Slamon, D.J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., & Norton, L. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. The New England Journal of Medicine, 344(11), 783-792. https://doi.org/10.1056/NEJM200103153441101

3. Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., Martin, N.M., Jackson, S.P., Smith, G.C., & Ashworth, A. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature, 434(7035), 917-921. https://doi.org/10.1038/nature03445

4. Garraway, L.A. & Janne, P.A. (2012). Circumventing cancer drug resistance in the era of personalized medicine. Cancer Discovery, 2(3), 214-226. https://doi.org/10.1158/2159-8290.CD-12-0012

5. Morash, M., Mitchell, H., Elemento, O., & Pathak, J. (2018). The role of next-generation sequencing in precision medicine: A review of outcomes in oncology. Journal of Personalized Medicine, 8(3). https://doi.org/10.3390/jpm8030030

6. Sicklick, J.K., Kato, S., Okamura, R., Schwaederle, M., Hahn, M.E., Williams, C.B., De, P., Krie, A., Piccioni, D.E., Miller, V.A., Ross, J.S., Benson, A., Webster, J., Stephens, P.J., Lee, J.J., Fanta, P.T., Lippman, S.M., Leyland-Jones, B., & Kurzrock, R. (2019). Molecular profiling of cancer patients enables personalized combination therapy: The I-PREDICT study. Nature Medicine, 25(5), 744-750. https://doi.org/10.1038/s41591-019-0407-5

7. Paller, C.J., Huang, E.P., Luechtefeld, T., Massett, H.A., Williams C.C., Zhao, J., Gravell A.E., Tamashiro, T., Reeves, S.A., Rosner, G.L., Carducci, M.A., Rubinstein, L, & Ivy, S.P. (2019). Factors affecting combinational trial success (FACTS): Investigator survey results on early-phase combination trials. Frontiers in Medicine, 6(122). https://doi.org/10.3389/fmed.2019.00122

8. Meric-Bernstram, F., Hurwitz, H., Raghav, K.P.S., McWilliams, R.R., Fakih, M., VanderWalde, A., Swanton, C., Kurzrock, R., Burris, H., Sweeney, C., Bose, R., Spigel, D.R., Beattie, M.S., Blotner, S., Stone, A., Schulze, K., Cuchelkar, V., & Hainsworth, J. (2019). Pertuzumab plus trastuzumab for HER2-amplified metastatic colorectal cancer (MyPathway): An updated report from a multicenter, open-label, phase 2a, multiple basket study. The Lancet Oncology¸ 20(4), 518-530. https://doi.org/10.1016/S1470-2045(18)30904-5

9. Lopez, J.S. & Banerji, U. (2017). Combine and conquer: challenges for targeted therapy combinations in early phase trials. Nature Reviews Clinical Oncology, 14(1), 57-66. https://doi.org/10.1038/nrclinonc.2016.96

10. Sen, S., Meric-Bernstram, F., Hong, D.S., Hess, K.R., & Subbiah, V. (2017). Co-occurring genomic alterations and association with progression-free survival in BRAFV600-mutated nonmelanoma tumors. Journal of the National Cancer Institute, 109(10). https://doi.org/10.1093/jnci/djx094

11. Skoulidis, F., Byers, L.A., Diao, L., Papadimitrakopoulou, V.A., Tong, P., Izzo, J., Behrens, C., Kadara, H., Parra, E.R., Canales, J.R., Zhang, J., Giri, U., Gudikote, J., Cortez, M.A., Yang, C., Fan, Y., Peyton, M., Girard, L, Coombes, K.R., ..., Heymach, J.V. (2016). Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities. Cancer Discovery, 5(8), 860-877. https://doi.org/10.1158/2159-8290.CD-14-1236

12. Witten, D.M. & Tibshirani, R. (2010). A framework for feature selection in clustering. Journal of the American Statistical Association, 105(490), 713-726. https://doi.org/10.1198/jasa.210.tm09415

13. Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., & Stuart, J.M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics, 45(10), 1113-1120. https://doi.org/10.1038/ng.2764

14. Akbani, R., Ng, P.K., Werner, H.M., Shahmoradgoli, M., Zhang, F., Ju, Z., Liu, W., Yang, J.Y., Yoshihara, K., Li, J., Ling, S., Seviour, E.G., Ram, P.T., Minna, J.D., Diao, L., Tong, P., Heymach, J.V., Hill, S.M., Dondelinger, F., …, Mills, G.B. (2014). A pan-cancer proteomic perspective on The Cancer Genome Atlas. Nature Communications. https://doi.org/10.1038/ncomms4887

15. Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Ronglai, S., Taylor, A.M., Cherniack, A.D.,Thorsson, V., Akbani, R., Bowlby, R., Wong, C.K., Wiznerowicz, M., Sanchez-Vega, F., Roberston, A.G., Schneider, B.G., Lawrence, M.S., Noushmehr, H., …, & Laird, P.W. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell, 173(2), 291-304. https://doi.org/10.1016/j.cell.2018.03.022

16. Kurnit, K.C., Dumbrava, E.E.I., Litzenburger, B., Khotskaya, Y.B., Johnson, A.M., Yap, T.A., Rodon, J., Zeng, J, Shufean, M.A., Bailey, A.M., Sanchez, N.S., Holla, V., Mendelsohn, J., Shaw, K.M., Bernstam E.V., Mills, G.B., Meric-Bernstam, F. (2018) Precision oncology decision support: Current approaches and strategies for the future. Clinical Cancer Research, 24(12), 2719-2731. https://doi.org/10.1158/1078-0432.CCR-17-2494

17. Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., Chang, M.T., Chandarlapaty, S., Traina, T.A., Paik, P.K., Ho, A.L., Hantash, F.M., Grupe, A., Baxi, S.S., Callahan, M.K., …, Schultz, N. (2017). OncoKB: A precision oncology knowledge base. JCO Precision Oncology. https://doi.org/10.1200/PO.17.00011

18. Flaherty, K.T., Gray, R., Chen, A., Li, S., Patton, D., Hamilton, S.R., Williams, P.M., Mitchell, E.P., Iafrate, A.J., Sklar, J., Harris, L.N., McShane, L.M., Rubinstein, L.V., Sims, D.J., Routbort, M., Coffey, B., Fu, T., Zwiebel, J.A., Little, R.F., …, NCI-MATCH Team. (2020). The molecular analysis for the therapy choice (NCI-MATCH) trial: Lessons for genomic trial design. Journal of the National Cancer Institute. https://doi.org/10.1093/jnci/djz248

19. Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Goncalves, E., Barthorpe, S., Lightfoot, H., Cokelaer, T., Greninger, P. van Dyke, E., Chang, H., de Silva, H., Heyn, H., Deng, X., Egan, R.K., Liu, Q., …, Garnett, M.J. (2016). A landscape of pharmacogenomic interactions in cancer. Cell, 166(3), 740-754. https://doi.org/10.1016/j.cell.2016.06.017

20. Ghandi, M., Huang, F.W., Jane-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R. 3rd, Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., Hu, K., Andreev-Drakhlin, A.Y., Kin, J., Hess, J.M., Haas, B.J., Aguet, F., Weir, B.A., Rothberg, M.V, Paolella, B.R., …, Sellers, W.R. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature, 569(7757), 503-508. https://doi.org/10.1038/s41586-019-1186-3

21. Nasser, M.W., Qamri, Z., Deol, Y.S., Ravi, J., Powell, C.A., Trikha, P., Schwendener, R.A., Bai, X.F., Shilo, K., Zou, X., Leone, G., Wolf, R., Yuspa, S.H., & Ganju, R.K. (2012). S100A7 enhances mammary tumorigenesis through upregulation of inflammatory pathways. Cancer Research, 72(3), 604-615. https://doi.org/10.1158/0009-5472.CAN-11-0669

22. Emberley, E.D., Murphy, L.C., & Watson, P.H. (2004). S100A7 and the progression of breast cancer. Breast Cancer Research, 6(4), 153-159. https://doi.org/10.1186/bcr816

23. Dey, K.K., Sarkar, S., Pal, I., Das, S., Dey, G., Bharti, R., Banik, P., Roy, J., Maity, S., Kulavi, I., Mandal, M. (2015). Mechanistic attributes of S100A7 (psoriasin) in resistance of anoikis resulting tumor progression in squamous cell carcinoma of the oral cavity. Cancer Cell International. https://doi.org/10.1186/s12935-015-0226-9

24. Corcione, A., Tortolina, G., Bonecchi R., Taborelli, G., Malavasi, F., Sozzani, S., Ottonello, L., Dallegri, F., & Pistoia, V. (2002). Chemotaxis of human tonsil B lymphocytes to CC chemokine receptor (CCR) 1, CCR2 and CCR4 ligands is restricted to non-germinal cancer cells. International Immunology, 14(8), 883-892. https://doi.org/10.1093/intimm/dxf054

25. Yoshie, O. & Matsushima, K. (2015). CCR4 and its ligands: From bench to bedside. International Immunology, 27(1), 11-20. https://doi.org/10.1093/intimm/dxu079

26. Wang, J., Duncan, D., Shi, Z., & Zhang, B. (2013). WEB-based Gene SeT AnaLysis Toolkit (WebGestalt): Update 2013. Nucleic Acids Research, 41, W77-83. https://doi.org/10.1093/nar/gkt439

27. Anderson, V.E., Walton, M.I., Eve, P.D., Boxall, K.J., Antoni, L., Caldwell, J.J., Aherne, W., Pearl, L.H., Oliver, A.W., Collins, I., & Garrett, M.D. (2011). CCT241533 is a potent and

selective inhibitor of CHK2 that potentiates the cytotoxicity of PARP inhibitors. Cancer Research, 71(2), 463-472. https://doi.org/10.1158/0008-5472.CAN-10-1252

28. Matulonis, U.A., Wulf, G.M., Barry, W.T., Birrer, M., Westin, S.N., Farooq, S., Bell-McGuinn, K.M., Obermayer, E., Whalen, C., Spagnoletti, T., Luo, W., Liu, H., Hok, R.C., Aghajanian, C., Solit, D.B., Mills, G.B., Taylor, B.S., Won, H., Berger, M.F., …, & Winer, E. (2017). Phase I dose escalation study of the PI3kinase pathway inhibitor BKM120 and the oral poly (ADP ribose) polymerase (PARP) inhibitor Olaparib for the treatment of high-grade serous ovarian and breast cancer. Annals of Oncology, 28(3), 512-518. https://doi.org/10.1093/annonc/mdw672

29. Ibrahim, Y.H., Garcia-Garcia, C., Serra, V., He, L., Torres-Lockhart, K., Prat, A., Anton, P., Cozar, P., Guzman, M., Grueso, J., Rodriguez, O., Calvo, M.T., Aura, C., Diez, O., Rubio, I.T., Perez, J., Rodon, J., Cortes, J., Ellisen, L.W., …, & Baselga, J. (2012). PI3K inhibition impairs BRCA1/2 expression and sensitizes BRCA-proficient triple-negative breast cancer to PARP inhibition. Cancer Discovery, 2(11), 1037-1047. https://doi.org/10.1158/2159-8290.CD-11-0348

30. Du, Y., Yamaguchi, H., Wei, Y., Hsu, J.L., Wang, H., Hsu, Y., Lin, W., Yu, W., Leonard, P.G., Lee, G.R., IV., Chen, M., Nakai, K., Hsu, M., Chen, C., Sun, Y., Wu, Y. Chang, W., Huang, W., Liu, C., …, & Hung, M. (2016). Blocking c-Met-meditated PARP1 phosphorylation enhances anti-tumor effects of PARP inhibitors. Nature Medicine, 22(2), 194-201. https://doi.org/10.1038/nm.4032

31. Goel, S., Wang, Q., Watt, A.C., Tolaney, S.M., Dillon, D.A., Li, W., Ramm, S., Palmer, A.C., Yuzugullu, H., Varadan, V., Tuck, D., Harris, L.N., Wong, K., Liu, X.S., Sicinski, P., Winer, E.P., Krop, I.E., & Zhao, J.J. (2016). Overcoming therapeutic resistance in HER2-positive breast cancers with CDK4/6 inhibitors. Cancer Cell, 29(3), 255-269. https://doi.org/10.1016/j.ccell.2016.02.006

32. Hanker, A.B., Pfefferle, A.D., Balko, J.M., Kuba, M.G., Young, C.D., Sanchez, V., Sutton, C.R., Cheng, H., Perou, C.M., Zhao, J.J., Cook, R.S., & Arteaga, C.L. (2013). Mutant PIK3CA accelerates HER2-driven transgenic mammary tumors and induces resistance to combinations of anti-HER2 therapies. Proceedings of the National Academy of Sciences of the United States of America, 110(35), 14372-14377. https://doi.org/10.1073/pnas.1303204110

33. Havel, J.J., Chowell, D., & Chan, T.A. (2019). The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. National Reviews Cancer, 19(3),133-150. https://doi.org/10.1038/s41568-019-01156-x

34. Jiao, S., Xia, W., Yamaguchi, H., Wei, Y., Chen, M., Hsu, J., Hsu, J.L., Yu, W., Du, Y., Lee, H., Li, C., Chou, C., Lim, S., Chang, S., Litton, J., Arun, B., Hortobagyi, G.N., & Hung, M. (2017). PARP inhibitor upregulates PD-L1 expression and enhances cancer-associated immunosuppression. Clinical Cancer Research, 23(14), 3711-3720. https://doi.org/10.1158/1078-0432.CCR-16-3215

35. Rich, T.A., Reckamp, K.L., Chae, Y.K., Doebele, R.C., Iams, W.T., Oh, M., Raymond, V.M., Lanman, R.B., Riess, J.W., Stinchcombe, T.E., Subbiah, V., Trevarthen, D.R., Fairclough, S., Yen, J., & Gautschi, O. (2019). Analysis of cell-free DNA from 32,989 advanced cancers reveals novel co-occurring activating RET alterations and oncogenic signaling pathway aberrations. Clinical Cancer Research, 25(19), 5832-5842. https://doi.org/10.1158/1078-0432.CCR-18-4049

36. Bao, L., Qian, Z., Lyng, M.B., Wang, L., Yu, Y.I, Wang, T., Zhang, X., Yang, H., Brunner, N., Wang, J., & Ditzel, H.J. (2018). Coexisting genomic aberrations associated with lymph node metastasis in breast cancer. Journal of Clinical Investigation, 128(6), 2310-2324. https://doi.org/10.1172/JCI97449

37. Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., & Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumor types. Nature, 505(7484), 495-501. https://doi.org/10.1038/nature12912

38. Rodon, J., Soria, J., Berger, R., Miller, W.H., Rubin, E., Kugel, A., Tsimberidou, A., Saintigny, P., Ackerstein, A., Brana, I., Loriot, Y., Afshar, M., Miller, V., Wunder, F., Bresson, C., Martini, J., Raynaud, J., Mendelsohn, J., Batist, G., …, & Kurzrock, R. (2019). Genomic and transcriptomic profiling expands precision cancer medicine: The WINTHER trial. Nature Medicine¸ 25(5), 751-758. https://doi.org/10.1038/s41591-019-0424-4