Proteomics and Phosphoproteomic Measurements Enhance Ability to Predict Ex Vivo Drug Response in AML

Sara Gosline

Pacific Northwest National Laboratory

Cristina Tognon

Oregon Health & Science University

Michael Nestor

Pacific Northwest National Laboratory

Sunil Joshi

Oregon Health & Science University

Rucha Modak

Oregon Health & Science University

Alisa Damnemsawad

Oregon Health & Science University

Camilo Posso

Pacific Northwest National Laboratory

Jamie Moon

Pacific Northwest National Laboratory

Joshua R. Hansen

Pacific Northwest National Laboratory

Chelsea Hutchinson-Bunch

Pacific Northwest National Laboratory

James C Pino

Pacific Northwest National Laboratory

Marina A. Gritsenko

Pacific Northwest National Laboratory

Karl K. Weitz

Pacific Northwest National Laboratory

Elie Traer

Oregon Health & Science University

Jeffrey Tyner

Pacific Northwest National Laboratory

Brian Druker

Oregon Health & Science University

Anupriya Agarwal

Oregon Health & Science University

Paul Piehowski

Pacific Northwest National Laboratory

Jason E. McDermott

Pacific Northwest National Laboratory

Karin Rodland (✉ karin.rodland@pnnl.gov)

Pacific Northwest National Laboratory

Research Article

Keywords:

Posted Date: January 26th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1277906/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Acute Myeloid Leukemia (AML) affects 20,000 patients in the US annually with a five-year survival rate of approximately 25%. One reason for the low survival rate is the high prevalence of clonal evolution that gives rise to heterogeneous sub-populations of leukemic cells with diverse mutation spectra, which eventually leads to disease relapse. This genetic heterogeneity drives the activation of complex signaling pathways that is reflected at the protein level. This diversity makes it difficult to treat AML with targeted therapy, requiring custom patient treatment protocols tailored to each individual's leukemia. Toward this end, the Beat AML research program prospectively collected genomic and transcriptomic data from over 1000 AML patients and carried out *ex vivo* drug sensitivity assays to identify genomic signatures that could predict patient-specific drug responses. However, there are inherent weaknesses in using only genetic and transcriptomic measurements as surrogates of drug response, particularly the absence of direct information about phosphorylation-mediated signal transduction. As a member of the Clinical Proteomic Tumor Analysis Consortium, we have extended the molecular characterization of this cohort by collecting proteomic and phosphoproteomic measurements from a subset of these patient samples to evaluate the hypothesis that proteomic signatures can improve the ability to predict drug response in AML patients. In this work we describe our systematic, multi-omic approach to evaluate proteomic signatures of drug response and compare protein levels to other markers of drug response such as mutational patterns. We explore the nuances of this approach using two drugs that target key pathways activated in AML: quizartinib (FLT3) and trametinib (Ras/MEK), and show how patient-derived signatures can be interpreted biologically and validated in cell lines. In conclusion, this pilot study demonstrates strong promise for proteomics-based patient stratification to assess drug sensitivity in AML.

Background

Acute myeloid leukemia (AML) is characterized by the incomplete maturation of myeloblasts and their expansion in blood and bone marrow, which impacts healthy blood cell formation resulting in decreased numbers of granulocytes, platelets, and red blood cells¹. Though the number of FDA-approved treatments for AML has increased significantly over the past five years, prognosis remains poor with a 5-year survival rate of 25% for individuals over the age of 20². Targeted agents have shown promise in mutationally defined subsets of patients, but due to the genetic evolution of this highly heterogeneous disease, drug response is often lost and patients relapse. Proper selection of personalized drugs and drug combinations over the course of a patient's disease will be required to provide more durable clinical responses, and will require a comprehensive mechanistic evaluation of each patient's leukemia.

The goal of the Beat AML program was to improve drug selection by collecting large quantities of molecular data together with *ex vivo* small molecule inhibitor assays performed on freshly isolated patient leukemia cells. In these studies, peripheral blood and bone marrow mononuclear cells from AML patients are isolated and exposed to a panel of approximately 145 drugs over a three-day period and cell viability is used as the primary readout for drug efficacy. Patient genomics and transcriptomics, as well as extensive clinical annotation, enable the stratification of patients by these measures which are more

effective than predictions of drug response based on genetics alone³. This functional genomic and transcriptomic dataset uncovered numerous novel genetic, transcriptomic, and microenvironmental drivers of AML pathogenesis and drug resistance⁴⁻⁸.

Proteomic measurements, including measurements of global protein levels and specific phosphosites, have been shown to better identify clinically relevant patterns in patient tumors compared to transcriptomics or genetics alone⁹. This has motivated significant investment by the National Cancer Institute through the Clinical Proteomic Tumor Analysis Consortium (CPTAC) in which patient-derived samples have been assayed using state-of-the-art mass spectrometry (MS) pipelines to produce proteomic and phosphoproteomic measurements of hundreds of tumors in breast, ovary, kidney, head and neck, endometrium, brain and other tissues¹⁰⁻¹⁵. In each study, these proteomic measurements reveal patterns that are not evident at the genomic or transcriptomic level⁹. Efforts to study the impact of the proteomics signature on drug response have been previously evaluated in cell lines¹⁶ and AML patient samples for a handful of drugs¹⁷. The integration of proteomic, phosphoproteomic, transcriptomic, and genomic data with drug response has not yet been evaluated in AML samples.

There exist numerous computational modeling and machine learning approaches to predict the response of cancer cell lines to drug perturbation using baseline genomics or transcriptomics^{18,19}. These approaches have been widely successful using genomic data together with subsequent dose response measurements to identify specific signatures capable of predicting which drugs affect cell lines from basal genomic and transcriptomic data of those same cell lines^{20,21}. These datasets have been further supplemented by global proteomic analysis of the same cell line library²² that have also be used to predict drug response. However, cell line-derived computational models have their flaws, as they sample a limited subset of patient genetics and have been shown to correlate poorly with patient-derived xenograft data of the same tumor type²³. There are still ongoing innovations in the computational space that predict drug response from the underlying genomic phenotype²⁴ including Bayesian approaches²⁵, variational auto-encoders²⁶, and deep learning²⁷. To date, however, most of these predictive models are based on cancer cell lines, which are limited in their ability to recapitulate the diversity of genetic backgrounds found in patients and lack potential contributions from the tumor microenvironment.

In this work, we combine the rigorous pre-clinical drug testing and genomic profiling of the Beat AML dataset with patient-derived proteomic and phosphoproteomic measurements, to determine the potential for protein-level data to produce robust molecular biomarkers of drug response. Using a small pilot proteomic dataset, we focus on two drugs that target the FLT3 and Ras/MEK pathways in AML (quizartinib and trametinib respectively) and evaluate how the genes, transcripts and proteins measured in each patient sample correlate with each other and drug sensitivity. We expand our analysis to 24 additional drugs to determine how well basal proteomic and phosphoproteomic measurements can predict drug response compared to genomic or transcriptomic measurements. We then explore the signatures that result from our analysis to determine how best to interpret these results biologically, by both evaluating their role in signaling networks and also assessing their expression in drug-resistant cell

lines. Together this work represents a robust toolkit by which protein-derived signatures can be used to predict drug response and understand the biological pathways these signatures represent.

Methods

Experimental design

Our overall experimental design is depicted in **Figure S1**. It entails subjecting patient AML samples to genomic and proteomic analysis and *ex-vivo* drug screening followed by the construction of predictive models of drug response for each type of data collected. We then use the signatures determined by the model to assess their performance in cross-validation experiments, explore their role in biological networks, and then validate them in cell lines.

Sample collection

Samples were collected and processed as described in detail previously³. Briefly, all patients gave informed consent to participate in the Beat AML study, which had the approval and guidance of the Institutional Review Boards (IRB) from participating institutions. All samples used in this manuscript were collected at Oregon Health & Science University with a broad 'research use' clause. Mononuclear cells (MNCs) were isolated from freshly obtained bone marrow or peripheral blood samples from AML patients via Ficoll gradient centrifugation. Isolated MNCs were utilized for genomic (500x WES; RNA-seq) and *ex vivo* functional drug screens. WES and RNA-seq were performed using standard methods and data analysis was performed as previously described³. Clinical, prognostic, genetic, cytogenetic and pathologic laboratory values as well as treatment and outcome data were manually curated from the patient electronic medical records (EMR). Patients were assigned a specific diagnosis based on the prioritization of genetic and clinical factors as determined by WHO guidelines. We selected 38 unique patients from our ongoing study that had complete proteomic and phosphoproteomic measurements.

Ex vivo drug screening analysis

For drug sensitivity assays, 10,000 viable cells were dispensed into each well of a 384-well plate containing 7 point, 3 fold dilution, drug concentration series from a library of small molecule inhibitors. Cells were incubated with the drugs in RPMI media containing 10% FBS without supplementary cytokines. After 3 days of culture at 37 °C in 5% CO₂, MTS reagent (CellTiter96 AQueous One; Promega) was added, the optical density was measured at 490 nm, and raw absorbance values were adjusted to a reference blank value and then used to determine cell viability (normalized to untreated control wells). *Ex vivo* functional drug screen data processing was performed as described, and drug fitting was carried out using the probit regression on quality-controlled data as in our previous work³.

Protein digestion and tandem mass tag (TMT) labeling

Sample preparation for proteomics was based on the protocol developed under the CPTAC consortium with minimal modifications²⁸. Patient cell pellets were lysed with 500 mL fresh lysis buffer, containing 8 M urea (Sigma-Aldrich), 50 mM Tris pH 8.0, 75 mM sodium chloride, 1 mM ethylenediamine tetra-acetic acid, 2 mg/mL Aprotinin (Sigma-Aldrich), 10 mg/mL Leupeptin (Roche), 1 mM PMSF in EtOH, 10 mM sodium fluoride, 1% of phosphatase inhibitor cocktail 2 and 3 (Sigma-Aldrich), 20 mM PUGNAc, and 0.01 U/ mL Benzonase. The samples were then vortexed for 10 seconds and placed in a thermomixer for 15 minutes at 4°C and 800 RPM. Vortexing was repeated and the samples incubated again for 15 minutes utilizing the same settings. After incubation, the samples were centrifuged for 10 minutes at 4°C and 18000 rcf to remove cell debris. The supernatant was then transferred to a fresh tube. A BCA (ThermoFisher) assay was performed on the supernatant to determine protein yield.

Protein concentrations were normalized to the same concentration prior to beginning digestion. The sample was reduced with 5 mM dithiothreitol (DTT) (Sigma-Aldrich) for 1 hour at 37°C and 800 rpm. Reduced cystines were alkylated with 10 mM iodacetamide (IAA) (Sigma-Aldrich) for 45 minutes at 25°C and 800 rpm in the dark. The sample was diluted fourfold with 50 mM Tris HCL pH 8.0 and then Lys-C (Wako) was added at a 1:20 enzyme:substrate ratio, followed by incubation for 2 hours at 25°C, shaking at 800 rpm. Trypsin (Promega) was then added at a 1:20 enzyme:substrate ratio, followed by a 14-hour incubation at 25°C and 800 rpm. The sample was quenched by adding formic acid to 1% by volume, and centrifuged for 15 minutes at 1500 rcf to remove any remaining cell debris. Peptides samples were desalted using a C18 solid phase extraction (SPE) cartridge (Waters Sep-Pak).

After drying down SPE eluates, each sample was reconstituted with 50 mM HEPES, pH 8.5 to a concentration of 5 mg/ mL. Each isobaric tag aliquot was dissolved in 250 mL anhydrous acetonitrile to a final concentration of 20 mg/ mL. The tag was added to the sample at a 1:1 peptide:label ratio and incubated for 1 hour at 25°C and 400 rpm and then diluted to 2.5 mg/mL with 50 mM HEPES pH 8.5, 20% acetonitrile (ACN). Finally, the reaction was quenched with 5% hydroxylamine and incubated for 15 minutes at 25°C and 400 rpm. The samples were then combined per each plex set and concentrated in a speed-vac before a final C18 SPE cleanup. Each 11-plex experiment was fractionated into 96 fractions by high pH reversed phase separation, followed by concatenation into 24 or 12 global fractions for MS analysis.

Phosphopeptide enrichment using IMAC

The global samples were further concatenated to create 6 samples per plex for further enrichment. Fe³⁺-NTA-agarose beads were freshly prepared using Ni-NTA-agarose beads (Qiagen). Sample peptides were reconstituted to a 0.5 mg/mL concentration with 80% ACN, 0.1% TFA and incubated with 40 mL of the bead suspension for 30 minutes at RT in a thermomixer set at 800 rpm. After incubation the beads were washed with 100 mL 80% ACN, 0.1% TFA and 50 mL 1% FA to remove any non-specific binding. Phosphopeptides were eluted off beads with 210 mL 500 mM K₂HPO₄, pH 7.0 directly onto C18 stage tips

and eluted from C18 material with 60 mL 50% ACN, 0.1% FA. Samples were dried in speed-vac concentrator for storage and reconstituted with 12 mL 3% ACN, 0.1% FA immediately prior to MS analysis.

LC-MS/MS analysis

Proteomic fractions were separated using a Waters nano-Aquity UPLC system (Waters) equipped with a 75 μ m I.D. x 25 cm length C18 column packed in-house with 1.9 μ m ReproSil-Pur 120 C18-AQ (Dr. Maisch GmbH). A 120-minute gradient of 95% mobile phase A (0.1% (v/v) formic acid in water) to 19% mobile phase B (0.1% (v/v) FA in acetonitrile) was applied to each fraction. The separation was coupled to either a Thermo Orbitrap™ Fusion Lumos™ (patient samples) or Q Exactive™ HF (cell lines) Hybrid Quadrupole-Orbitrap™ mass spectrometer for MS/MS analysis. MS Spectra were collected from 350 to 1800 m/z at a mass resolution setting of 60,000. A top speed method was used for the collection of MS2 spectra at a mass resolution of 50K. An isolation window of 0.7 m/z was used for higher energy collision dissociation (HCD), singly charged species were excluded, and the dynamic exclusion window was 45 seconds. For the Fusion Lumos™, a top speed method was used for the collection of MS2 spectra at a mass resolution of 50K. For the Q Exactive™ HF experiments, a top 16 method was used for the collection of MS² spectra at a mass resolution of 30K.

TMT global proteomics data processing

All Thermo RAW files were processed using mzRefinery to correct for mass calibration errors, and then spectra were searched with MS-GF+ v9881²⁹⁻³¹ to match against the human reference protein sequence database downloaded in April of 2018 (71,599 proteins), combined with common contaminants (e.g., trypsin, keratin). A partially tryptic search was used with a ± 10 parts per million (ppm) parent ion mass tolerance. A reversed sequence decoy database approach was used for false discovery rate calculation. MS-GF+ considered static carbamidomethylation (+57.0215 Da) on Cys residues and TMT modification (+229.1629 Da) on the peptide N terminus and Lys residues, and dynamic oxidation (+15.9949 Da) on Met residues. The resulting peptide identifications were filtered to a 1% false discovery rate at the unique peptide level. A sequence coverage minimum of 6 per 1000 amino acids was used to maintain a 1% FDR at the protein level after assembly by parsimonious inference.

The intensities of TMT 11 reporter ions were extracted using MASIC software³². Extracted intensities were then linked to PSMs passing the confidence thresholds described above. Relative protein abundance was calculated as the ratio of sample abundance to reference channel abundance, using the summed reporter ion intensities from peptides that could be uniquely mapped to a gene. The relative abundances were log2 transformed and zero-centered for each gene to obtain final relative abundance values.

TMT phosphoproteomics data processing

IMAC enriched fraction datasets were searched as described above with the addition of a dynamic phosphorylation (+79.9663 Da) modification on Ser, Thr, or Tyr residues. The phosphoproteomic data were further processed with the Ascore algorithm³³ for phosphorylation site localization, and the top-scoring assignments were reported. To account for sample loading biases in the phosphoproteome analysis, we applied the same correction factors derived from median-centering of the global proteomic dataset for normalization.

All proteomic data can be found on our [synapse site](#). The cohort is spread across three tranches, as described in **Table 1** below.

Patients	Data type	File	Table
Primary patient cohort	Proteomics	syn22130778	syn22172602
Patients with Sorafenib treatment	Proteomics	syn22313435	syn22314121
Patients with drug combination	Proteomics	syn25672089	syn22156810
Primary patient cohort	Phosphoproteomics	syn24610481	syn24227903
Patients with Sorafenib treatment	Phosphoproteomics	syn24227680	syn24228075
Patients with drug combination	Phosphoproteomics	syn24240156	syn24240355

Table 1: Location of processed proteomics files on Synapse.

Identifying drugs and samples for analysis

The list of available data for each patient is in **Table S1**. Although ~145 total compounds were tested in the drug panels, we filtered the drugs in this study to collect those that exhibited a range of responses across the 38 patients as determined by area under the curve (AUC) of the dose response. AUC represents the amount of drug required to reduce cell viability, so higher AUC values mean the samples are less sensitive to the drug, and lower AUC values indicate the samples are more sensitive. Specifically, we selected drugs for which at least 10% or 2 (whichever was greater) samples exhibited an AUC less than 100 (determined to be sensitive in previous work³). This selection produced a “balanced” distribution of AUC scores to enable our downstream analysis. We also added Gilteritinib (ASP-2215) to the panel as it is currently being evaluated in numerous clinical trials. The full range of drug responses across patients is shown in **Figure 2A**. While some patient samples lacked data on all 26 drugs (indicated in blue in **Figure 2A**), we were still able to use these samples to compare the efficacy of genomics, transcriptomics, proteomics and phosphoproteomics to model drug sensitivity based on the available data.

Linear models of proteomics and drug response

We constructed linear models for each of the 26 different drugs across up to 38 patients (depending on how many patient samples were evaluated with that drug) by regressing the AUC values (which ranged between 0 and ~300, as depicted in **Figure 2A**) against the molecular data as shown in **Table S1**. The input data for each model were each scaled slightly differently: the genetic mutations were represented as a binary matrix in which a 1 represented the presence of a somatic mutation and a 0 represented no mutation, the transcriptomics was represented by Counts per Million (CPM) of gene expression values, while proteomics and phosphoproteomics were represented as the log ratio of gene/phosphosites compared to the reference sample described above.

For each combination of drug and data type, we constructed a linear model $Y \sim X$ where Y represents the vector of AUC values and X represents the molecular measurements for that patient. We used three different linear modeling approaches to reduce the number of features selected by the model: LASSO regression³⁴, Elastic Net Regression³⁵, and logistic regression as implemented by the *glm* $\neq$ *t* package³⁶. For the logistic regression, we discretized the AUC by representing Y as a binary variable, where 1 represented an AUC greater than 100 (patient is resistant to drug) and 0 if the AUC is less than 100 (patient is sensitive to drug).

For each model, we employed K-fold cross validation with K=5 on each type of data (e.g. mutations, proteomics, etc.) to assess performance. Within each K, we used leave-one-out cross-validation for each combination of data to select the alpha parameter that minimized cross-validation error. The model performance scores in **Figure 2B** and **Table S2** represented the average correlation between predicted and actual values across all 5 models for each drug/data type. All of our analysis can be found in the *amlresistance* $\neq$ *tw* or *ks* package we built at <http://github.com/PNNL-CompBio/amlresistancenetworks> and implemented at <https://github.com/PNNL-CompBio/beatamlpilotproteomics>. Those models that failed to select any molecular features were not included in our final analysis. The results are depicted in **Figures 2B**.

Signature interpretation using pathway annotations and statistical enrichment

To identify patterns in the features selected by the LASSO, Elastic Net, and logistic models we employed three main approaches. For gene, transcript, and proteomic signatures, we first used the *clusterProfi* $\leq$ *r* package³⁷ to identify GO biological process tools that are enriched for the specific genes, transcripts, or proteins selected by the model. The results are listed in **Table S2**. In cases where there were no significant (corrected $p < 0.01$) terms, the column is blank. For phosphoproteomic features, we used the $\leq$ *apR* R package³⁸ to identify specific kinases that were over-represented among the selected substrates, though none were identified. We believe this is due to the sparsity of the signatures as well as the lack of information about the kinase-substrate interactions.

Supplementing sparse proteomic signatures with interaction networks

To provide further context for the phosphoproteomic features selected by the models, we mapped selected proteins and/or selected phosphosites to published protein-protein³⁹ and kinase-substrate^{40,41} interactions and then reduced this network to identify subnetworks using the Prize Collecting Steiner Forest (PCSF) R package^{42,43}. Specifically, we used the STRING database³⁹ together with networkKin⁴⁰ and PhosphoSitePlus⁴¹ predictions of kinase substrate interactions to build a graph that combined protein-protein interactions with kinase-substrate interactions. To do this we added each phosphosite as its own node in the underlying graph. We weighted each edge from the node representing the substrate gene to the phosphosite with a cost of $m/4$ where m represents the mean cost of all the edges in the graph. The weight of each edge between the phosphosite node and the kinase gene was weighted with a cost of $3m/2$ where m represents the mean cost of all edges in the graph. We then ran the PCSF algorithm^{42,43} over 100 randomizations using phosphosites, proteins, or genes from a single drug model. The results for the quizartinib proteomic and trametinib genomic logistic signatures are in **Figure 3**.

Using the proteins selected by the PCSF algorithm, which are a combination of those selected by the linear model as well as those selected by the PCSF algorithm, we used Cytoscape⁴⁴ and the BinGO⁴⁵ application to identify which GO biological process terms were enriched. The results are depicted in **Tables S4**.

Trametinib resistant cell line cultures

Human MOLM13 cells with FLT3-ITD mutation, were obtained from the Sanger Institute Cancer Cell Line Panel. Cell lines were maintained in RPMI 1640 (Gibco) supplemented with 20% Fetal Bovine Serum (HyClone), 2% L-glutamine, 1% penicillin/streptomycin (Life Technologies). Trametinib-resistant MOLM13 cell lines were generated by culturing MOLM13 cells in increasing concentrations of trametinib (Selleck). Cell viability was measured bi-weekly and cells were replenished with new media and trametinib. Resistance was assessed using the MTS assay for drug sensitivity. Once resistant, cell lines were maintained in 50nM trametinib added bi-weekly. Cell lines were screened for mycoplasma contamination on a monthly schedule.

For proteomic and phosphoproteomic profiling, 5 million parental MOLM13 (N=3) and resistant MOLM13 (N=3) cell lines were starved overnight in starvation media (RPMI supplemented with 0.1% BSA). Trametinib (50nM) was added to the starvation media of the resistant cell lines. Cells were washed three times in PBS, pelleted and flash frozen.

Quizartinib resistant cell line cultures

Human MOLM14 cells were generously provided by Dr. Yoshinobu Matsuo (Fujisaki Cell Center, Hayashibara Biochemical Labs, Okayama, Japan). Cells were grown in RPMI (Life Technologies Inc., Carlsbad, CA) supplemented with 10% FBS (Atlanta Biologicals, Flowery Branch, GA), 2% L-glutamine, 1% penicillin/streptomycin (Life Technologies Inc.), and 0.1% amphotericin B (HyClone, South Logan, UT). Cell line authentication was performed at the OHSU DNA Services Core facility.

To establish resistant cultures, 10 million MOLM14 cells were treated with 10 nM of quizartinib (Selleck Chemicals, Houston, TX) in media alone (N = 4) or in media supplemented with 10 ng/mL of FGF2 (N = 4) or FLT3 ligand (N = 4, FL; PeproTech Inc., Rocky Hill, NJ)⁴⁶. All cultures were maintained in 10 mL of media. Every 2 or 3 days, recombinant ligands and quizartinib were replaced and cell viability was evaluated using the Guava cell counter (Millipore Inc., Burlington, MA). Following ligand withdrawal, quizartinib and media were similarly replenished and viability was monitored every 2 to 3 days. All cell lines were tested for mycoplasma on a monthly schedule.

For proteomic and phosphoproteomic profiling, naïve MOLM14 (N = 4), quizartinib-resistant parental (N = 2, no ligand), early (N = 4/ligand) and late (N = 4/ligand) cultures were washed three times with PBS to remove any trace of fetal bovine serum, pelleted, and flash frozen.

Results

Multi-omic data highlights varied impact of drug response in AML patient samples

We first explored the relationship between the individual molecular (genetic, transcriptomics, proteomic, etc.) measurements in our matched patient cohort. Given the success of molecular profiling using RNA-seq in the Beat AML dataset³, and general knowledge that mRNA can often, but not always, be a proxy for protein expression, we wanted to ask if mRNA and protein levels are correlated in our specific dataset. The results, shown in **Figure S2** confirm previously published work⁹ that mRNA and protein levels are weakly correlated (Spearman R=0.25) across all patient samples. We also mapped phosphosites to their corresponding proteins and found that the overall abundance values were also weakly correlated (Spearman R=0.15), aligning with our previous work⁴⁷.

To examine the correlation of molecular values with drug response, we first sampled sensitivity to quizartinib and the genes, transcripts, and proteins within the pathway quizartinib was designed to target. Specifically, we looked at these molecules in the FLT3/MAPK pathway (**Figure 1A**) and compared them with the *ex vivo* sensitivity to quizartinib (**Figure 1B-C**). The proteins and transcripts in the pathway itself are variably correlated. Specifically, we found that some proteins, e.g. NRAS and FLT3, are positively correlated with the mRNA levels for the same gene (R=0.41, R=0.34, respectively), while proteins such as SOS1 and PTPN11 are more negatively correlated (R=-0.22, R=-0.11 respectively). We then compared transcript (**Figure 1B**) and protein (**Figure 1C**) levels with the AUC for quizartinib by plotting a heatmap of

the molecular values ranked by drug response. The results illustrate that, even in the case of targeted drugs, there are few robust biomarkers to predict drug response.

We expanded our correlative analysis to study the Ras/MEK pathway, which is downstream of Ras and targeted by trametinib. The correlation of the mRNA and protein levels of the pathway trametinib targets (**Figure 1D**) was again modestly positive in some genes such as JAK1, JAK2, and MAP2K1 but negative in others such as PTPN11. We also measured the correlation of mRNA levels (**Figure 1E**) and protein levels (**Figure 1F**) with trametinib response in the patients. Here we found that the three patients with NRAS mutations were sensitive to trametinib, but that few other mRNA or protein levels seemed to correlate with drug response.

Linear modeling enables broad sweep of data space to identify multi-omic signatures of drug response

Our findings in **Figure 1** show that molecular measurements across a pathway targeted by a specific drug may fail to adequately summarize the drug response in patient samples. As such, we turned to a basic statistical approach to identify such groups of genes, transcripts, proteins, or phosphosites that predict drug response.

We examined a panel of 26 drugs measured in the Beat AML *ex vivo* drug sensitivity functional assay described above, specifically selecting drugs that exhibited a variable response in the pilot samples as described in the experimental procedures and shown in **Figure 2A**. We constructed three types of linear models as described above for each drug and data modality (genomics, transcriptomics, proteomics, phosphosites, proteomics + phosphosites combined, and all four data types combined) for a total of 24 possible models for each drug. We measured the performance of each model using 5-fold cross validation and measured the correlation between the predicted response on the held-out data and the actual value. The correlation values of each of the five models is shown in **Figure 2B** and summarized in **Table S2**. In numerous cases, the models were unable to select any features and therefore were not counted. This was particularly noticeable in the case of logistic regression, where the division of test data into sensitive/resistant samples left fewer data points for model construction.

This framework enabled us to compare modeling approach and data type. While all three flavors of regression performed similarly, the logistic regression created fewer models overall, and was not very accurate using mutation data or transcriptomic data alone (median correlation <0.1 , **Figure 2B**). As expected, the models built with all four types of data generally perform well (yellow, **Figure 2B**). Interestingly, these models also perform well on just proteomics data (blue). However, despite the general good performance of models, there was a high degree of variability between drugs and drug families. **Figure S3** shows the performance of each model across

individual drugs (**Figure S3A**) and drug classes (**Figure S3B**). This diversity shows that individual model selection requires a robust cross validation approach to avoid generalizing with only one type of model or data modality.

Model selection via cross-validation and network analysis provides robust interpretations of molecular signatures

To show how the cross validation framework can be used in practice, we selected the top performing model by drug from **Figure S3A** by correlation between training and test data and then examined the features to determine if they aligned with the known mechanism of action of each of the drugs. The selected models are depicted in **Table S3**. We highlight the results for quizartinib and trametinib, which we examined in more detail.

Specifically we focused on the proteomic logistic model of quizartinib response. To identify the features that drove the models, we re-ran it on all the data (instead of on just the training data subsets) and plotted the features in **Figure 3A**. Here we noticed INPP5D, which is identified in both the LASSO and logistic regression models and highly down-regulated in quizartinib sensitive samples. This gene encodes the inositol 5-phosphatase known as SHIP-1 which acts as a negative regulator of the PI3K/AKT pathway. SHIP-1 affects cell proliferation in AML, due to mutations in the nuclear localization signature or phosphorylation site⁴⁸. It has also been shown to act as an adaptor protein linked to wild type FLT3 signaling^{49,50}.

Given the sparsity constraints inherent in any regression-based approach, we found that the proteins selected by the model represented an incomplete picture of the underlying signaling networks that were altered (**Table S3**). Therefore we used the OmicsIntegrator⁴² tool to supplement the proteins selected by the model (yellow nodes in **Figure 3B**) with proteins from the local protein-protein interaction network (blue nodes in **Figure 3B**). Here we identified additional proteins, such as COMMD3-BMI1, with a large number of partners in the underlying protein interaction network, that has been predicted in other models to regulate a number of genes in AML⁵¹. Other examples of potential mediators of quizartinib response include MSH2, a protein involved in DNA mismatch repair⁵² and KAT7, which has been shown to alter histone regulation in AML⁵³. By linking the proteins in the signature together through other protein interactions, we can examine how the proteomic signature connects various proteins involved in signaling, histone regulation, and DNA damage to show how alterations in diverse pathways give rise to drug sensitivity.

We also examined the top model that predicted trametinib response from **Table S3**. Here we found that genetic mutations in 8 genes nearly perfectly clustered response in patient samples, depicted in **Figure 3C** (though the mean correlation in cross-validation was still less than 0.5). The network inferred from these genes is much smaller, due to the reduced number of genes and also the fact that TP53 has many

interactors. However, it shows a direct link between KRAS, which is typically the mutation targeted by trametinib, and the other mutations that give rise to drug sensitivity.

Proteins that predict drug response are dysregulated in resistant cell lines

To experimentally validate these signatures, we turned to cell culture models of AML. Here we explicitly focused on proteomic measurements to determine if proteomic signatures could predict resistance to drugs *in vitro*. We first examined MOLM13 cells that were grown in the presence of trametinib over 3-4 months to develop resistance, and measured protein expression in the resistant cells compared to the parental cells. We then evaluated all three proteomic signatures of trametinib response in these cell lines (**Figure 4**). Despite the fact that the proteomic signature was not as robust in our cross-validation, we found that each protein signature from the LASSO, logistic, and Elastic Net regressions cleanly clustered resistant and sensitive cells.

Proteomic signatures can distinguish between early and late models of drug resistance

To further confirm this role of regression-derived signatures in cell lines, we evaluated the proteins selected for the quizartinib signatures in MOLM14 cells grown in the presence of quizartinib to develop resistance. We then examined the proteins selected by all three regressions and found that regardless of which method, the proteins clustered the resistant and parent cells separately (**Figure 5A-C**).

We then wanted to explore if any of these signatures were related to temporal changes during development of resistance to quizartinib. We used resistant cell lines that were developed in two stages: early resistance, which is mediated by extrinsic ligands from the marrow microenvironment, and late resistance, which is mediated by the expansion of intrinsic resistance mutations – most commonly in the activation loop^{46,54}. Using this model we compared early resistance and late resistance, with the hypothesis that the patient-derived signature would more closely resemble the late resistance phenotype.

To test this hypothesis, we plotted the proteins selected by the LASSO model in these cell lines and clustered them in **Figures 5D**. We observed a similar split between sensitive and resistant cells as we did in **Figure 5A-C**, as the proteins that predict drug response cluster the MOLM14 parental cells (blue) distinctly from the fully resistant cells (beige). However, in this case, these proteins also separate those cells that represented early resistance (red) from those that represent late resistance (orange) in our previous work. This fits with our previous claim that the resistance to FLT3 inhibitors involves a two-step process, as cell lines exhibiting the early resistance phenotype cluster more closely with the parental cells than with the late resistance cells.

Since no GO terms appeared to be enriched among the proteins in the LASSO signature (**Table S2**), we also examined the network linked by the proteins and phosphosites using the Prize-Collecting Steiner Forest^{42,43} as described in our Experimental Procedures and depicted in **Figure 5E**. Here, the network enabled us to get a broader picture of how the proteins involved could participate in the same signaling pathways. Using this augmented network we were able to identify enriched GO terms (Table S4), specifically apoptotic related pathways. We also found enrichment in B cell activation, differentiation, and homeostasis, driven by network proteins INPP5D, FLT3, and CASP3.

Discussion

This study describes a straightforward framework to assess the role of protein-derived measurements in predicting *ex vivo* AML patient drug response. Given that proteins clearly capture a unique aspect in disease activity, we employed numerous types of regression analysis with cross-validation to determine how to identify which signature is best for each drug or drug family. We also showed how to interpret these signatures using data from external sources and also validated the signatures in cell culture models.

While we were able to compare different flavors of regression modeling, we believe that, for our data, there is no clear best choice. The logistic regression failed in many cases with low samples, so may not work for all drugs. However, the choice of data type does seem to have more of an impact, as genetic mutations are robust in cases of targeted therapy (e.g. trametinib and quizartinib for NRAS and FLT3 activating mutations respectively), but models involving proteins perform best when assessed over all drugs (**Figure 2B**). We are looking to expand this analysis using a larger patient cohort where we can further derive robust protein signatures that can be validated in the clinic.

We also underscore the need for interpretable models for drug response. While the regression approaches select the features that are numerically most valuable for predicting drug response, they fail to account for the biological context of the proteins or genes selected. As such, we believe that using the OmicsIntegrator or other tools to map selected proteins to the interaction network will provide better understanding of what causes drug resistance in some patients, and potentially assist in understanding the effects of drug combinations, which are becoming increasingly common in clinical trials^{55,56}.

In summary, this study presents an effective workflow for the future analysis of integrated genomic, transcriptomic, proteomic and phosphoproteomic data in larger cohorts, such as the larger Beat AML cohort (N=210). While the patient cohort used in this preliminary study is limited in size, the robust verification in cell line studies provides confidence in the scalability of these methods. Additionally, the performance of protein-based models compared to transcriptomic-based models opens up the possibility of developing antibody-based, CLIA-eligible assays for the rapid assessment of likely therapeutic targets at the time of biopsy, without the need for DNA sequencing. Lastly we believe that our network approaches could help identify potential novel drug synergies that could be tested in the clinic.

Declarations

Availability of data and materials

Data was uploaded to Synapse where it was used for subsequent analysis at <http://synapse.org/ptrc>. mRNA (counts per million) and genetic mutation measurements (variant allele frequency) can be found at <https://www.synapse.org/#!Synapse:syn22172602/tables/>.

All data used for this project is stored on Synapse at <http://synapse.org/ptrc>, where you can request access to the data specifically mentioned in this manuscript. All analysis and figures can be viewed at <https://github.com/PNNL-CompBio/beatamlpilotproteomics>.

Acknowledgements

The authors would like to acknowledge the patients that participated in the study.

Funding

This research was supported by awards from the National Cancer Institute, U01CA214116 (KDR, BD), U24CA210955 (TL, RD), R01CA229875-03 (Agarwal), R01CA229875-02S1 (Agarwal), U54CA224019-03 (Druker/Tyner), and the American Cancer Society, RSG-17-187-01-LIB (Agarwal).

Ethics approval and consent to participate

All patients gave informed consent to participate in the Beat AML study, which had the approval and guidance of the Institutional Review Boards (IRB) from participating institutions. All samples used in this manuscript were collected at Oregon Health & Science University with a broad 'research use' clause.

Consent for publication

All authors consent for publication

Competing interests

All authors declare no competing interests.

Authors' contributions

SG, CT, JM, and KR designed the study. CT, SJ, RM, AD, JM, JH, CH, MG, and KW collected samples and data for the project. SG, MN, CP, JP, and JM carried out the data processing. SG, CT, JM, PP, and KR wrote the manuscript. ET, AG, BD, JT, JM, and PP advised and edited the manuscript.

References

1. Dong, Y. *et al.* Leukemia incidence trends at the global, regional, and national level between 1990 and 2017. *Exp Hematol Oncol***9**, 14, doi:10.1186/s40164-020-00170-6 (2020).
2. Board, C. N. E. *Leukemia - Acute Myeloid - AML: Statistics*, < <https://www.cancer.net/cancer-types/leukemia-acute-myeloid-aml/statistics>> (2021).
3. Tyner, J. W. *et al.* Functional genomic landscape of acute myeloid leukaemia. *Nature***562**, 526-531, doi:10.1038/s41586-018-0623-z (2018).
4. Nechiporuk, T. *et al.* The TP53 Apoptotic Network Is a Primary Mediator of Resistance to BCL2 Inhibition in AML Cells. *Cancer Discov***9**, 910-925, doi:10.1158/2159-8290.CD-19-0125 (2019).
5. Drusbosky, L. M. *et al.* Predicting response to BET inhibitors using computational modeling: A BEAT AML project study. *Leuk Res***77**, 42-50, doi:10.1016/j.leukres.2018.11.010 (2019).
6. Rosenberg, M. W. *et al.* Genomic markers of midostaurin drug sensitivity in FLT3 mutated and FLT3 wild-type acute myeloid leukemia patients. *Oncotarget***11**, 2807-2818, doi:10.18632/oncotarget.27656 (2020).
7. Kurtz, S. E. *et al.* Dual inhibition of JAK1/2 kinases and BCL2: a promising therapeutic strategy for acute myeloid leukemia. *Leukemia***32**, 2025-2028, doi:10.1038/s41375-018-0225-7 (2018).
8. Kurtz, S. E. *et al.* Molecularly targeted drug combinations demonstrate selective effectiveness for myeloid- and lymphoid-derived hematologic malignancies. *Proc Natl Acad Sci U S A***114**, E7554-E7563, doi:10.1073/pnas.1703094114 (2017).
9. Wang, J. *et al.* Proteome Profiling Outperforms Transcriptome Profiling for Coexpression Based Gene Function Prediction. *Mol Cell Proteomics***16**, 121-134, doi:10.1074/mcp.M116.060301 (2017).
10. Krug, K. *et al.* Proteogenomic Landscape of Breast Cancer Tumorigenesis and Targeted Therapy. *Cell***183**, 1436-1456 e1431, doi:10.1016/j.cell.2020.10.036 (2020).
11. Hu, Y. *et al.* Integrated Proteomic and Glycoproteomic Characterization of Human High-Grade Serous Ovarian Carcinoma. *Cell Rep***33**, 108276, doi:10.1016/j.celrep.2020.108276 (2020).
12. Clark, D. J. *et al.* Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell***180**, 207, doi:10.1016/j.cell.2019.12.026 (2020).
13. Huang, C. *et al.* Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. *Cancer Cell*, doi:10.1016/j.ccell.2020.12.007 (2021).
14. Dou, Y. *et al.* Proteogenomic Characterization of Endometrial Carcinoma. *Cell***180**, 729-748 e726, doi:10.1016/j.cell.2020.01.026 (2020).
15. Wang, L. B. *et al.* Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell*, doi:10.1016/j.ccell.2021.01.006 (2021).
16. Frejno, M. *et al.* Proteome activity landscapes of tumor cell lines determine drug responses. *Nat Commun***11**, 3639, doi:10.1038/s41467-020-17336-9 (2020).
17. Casado, P. *et al.* Proteomic and genomic integration identifies kinase and differentiation determinants of kinase inhibitor sensitivity in leukemia cells. *Leukemia***32**, 1818-1822, doi:10.1038/s41375-018-0032-1 (2018).

18. Harper, A. R. & Topol, E. J. Pharmacogenomics in clinical practice and drug development. *Nat Biotechnol***30**, 1117-1124, doi:10.1038/nbt.2424 (2012).
19. Ben-David, U. *et al.* Genetic and transcriptional evolution alters cancer cell line drug response. *Nature***560**, 325-330, doi:10.1038/s41586-018-0409-3 (2018).
20. Seashore-Ludlow, B. *et al.* Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov***5**, 1210-1223, doi:10.1158/2159-8290.CD-15-0235 (2015).
21. Iorio, F. *et al.* A Landscape of Pharmacogenomic Interactions in Cancer. *Cell***166**, 740-754, doi:10.1016/j.cell.2016.06.017 (2016).
22. Nusinow, D. P. *et al.* Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell***180**, 387-402 e316, doi:10.1016/j.cell.2019.12.023 (2020).
23. Gao, H. *et al.* High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med***21**, 1318-1325, doi:10.1038/nm.3954 (2015).
24. Costello, J. C. *et al.* A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol***32**, 1202-1212, doi:10.1038/nbt.2877 (2014).
25. Cortes-Ciriano, I. *et al.* Proteochemometric modeling in a Bayesian framework. *J Cheminform***6**, 35, doi:10.1186/1758-2946-6-35 (2014).
26. Rampasek, L., Hidru, D., Smirnov, P., Haibe-Kains, B. & Goldenberg, A. Dr.VAE: improving drug response prediction via modeling of drug perturbation effects. *Bioinformatics***35**, 3743-3751, doi:10.1093/bioinformatics/btz158 (2019).
27. Kuenzi, B. M. *et al.* Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. *Cancer Cell***38**, 672-684 e676, doi:10.1016/j.ccell.2020.09.014 (2020).
28. Mertins, P. *et al.* Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. *Nat Protoc***13**, 1632-1661, doi:10.1038/s41596-018-0006-9 (2018).
29. Gibbons, B. C., Chambers, M. C., Monroe, M. E., Tabb, D. L. & Payne, S. H. Correcting systematic bias and instrument measurement drift with mzRefinery. *Bioinformatics***31**, 3838-3840, doi:10.1093/bioinformatics/btv437 (2015).
30. Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun***5**, 5277, doi:10.1038/ncomms6277 (2014).
31. Kim, S., Gupta, N. & Pevzner, P. A. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J Proteome Res***7**, 3354-3363, doi:10.1021/pr8001244 (2008).
32. Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N. & Smith, R. D. MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput Biol Chem***32**, 215-217, doi:10.1016/j.compbiolchem.2008.02.006 (2008).
33. Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J. & Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol***24**, 1285-1292,

doi:10.1038/nbt1240 (2006).

34. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met***58**, 267-288, doi:DOI 10.1111/j.2517-6161.1996.tb02080.x (1996).
35. Zou, H. & Hastie, T. Regression shrinkage and selection via the elastic net, with applications to microarrays. *JR Stat Soc Ser B***67**, 301-320 (2003).
36. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw***33**, 1-22 (2010).
37. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS***16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
38. Danna, V. *et al.* leapR: An R Package for Multiomic Pathway Analysis. *J Proteome Res*, doi:10.1021/acs.jproteome.0c00963 (2021).
39. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res***49**, D605-D612, doi:10.1093/nar/gkaa1074 (2021).
40. Linding, R. *et al.* NetworkKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res***36**, D695-699, doi:10.1093/nar/gkm902 (2008).
41. Hornbeck, P. V. *et al.* 15 years of PhosphoSitePlus(R): integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res***47**, D433-D441, doi:10.1093/nar/gky1159 (2019).
42. Tuncbag, N. *et al.* Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package. *PLoS Comput Bio***12**, e1004879, doi:10.1371/journal.pcbi.1004879 (2016).
43. Akhmedov, M. *et al.* PCSF: An R-package for network-based interpretation of high-throughput data. *PLoS Comput Bio***13**, e1005694, doi:10.1371/journal.pcbi.1005694 (2017).
44. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res***13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
45. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics***21**, 3448-3449, doi:10.1093/bioinformatics/bti551 (2005).
46. Traer, E. *et al.* FGF2 from Marrow Microenvironment Promotes Resistance to FLT3 Inhibitors in Acute Myeloid Leukemia. *Cancer Res***76**, 6471-6482, doi:10.1158/0008-5472.CAN-15-3569 (2016).
47. Arshad, O. A. *et al.* An Integrative Analysis of Tumor Proteomic and Phosphoproteomic Profiles to Examine the Relationships Between Kinase Activity and Phosphorylation. *Mol Cell Proteomics***18**, S26-S36, doi:10.1074/mcp.RA119.001540 (2019).
48. Nalaskowski, M. M. *et al.* Nuclear accumulation of SHIP1 mutants derived from AML patients leads to increased proliferation of leukemic cells. *Cell Signa***49**, 87-94, doi:10.1016/j.cellsig.2018.05.006 (2018).

49. Zhang, S., Mantel, C. & Broxmeyer, H. E. Flt3 signaling involves tyrosyl-phosphorylation of SHP-2 and SHIP and their association with Grb2 and Shc in Baf3/Flt3 cells. *J Leukoc Biol***65**, 372-380, doi:10.1002/jlb.65.3.372 (1999).
50. Gu, T. L. *et al.* Survey of activated FLT3 signaling in leukemia. *PLoS One***6**, e19169, doi:10.1371/journal.pone.0019169 (2011).
51. Ye, J., Luo, D., Yu, J. & Zhu, S. Transcriptome analysis identifies key regulators and networks in Acute myeloid leukemia. *Hematology***24**, 487-491, doi:10.1080/16078454.2019.1631506 (2019).
52. Worrillow, L. J. & Allan, J. M. Deregulation of homologous recombination DNA repair in alkylating agent-treated stem cell clones: a possible role in the aetiology of chemotherapy-induced leukaemia. *Oncogene***25**, 1709-1720, doi:10.1038/sj.onc.1209208 (2006).
53. Au, Y. Z. *et al.* KAT7 is a genetic vulnerability of acute myeloid leukemias driven by MLL rearrangements. *Leukemia***35**, 1012-1022, doi:10.1038/s41375-020-1001-z (2021).
54. Joshi, S. K. *et al.* The AML microenvironment catalyzes a stepwise evolution to gilteritinib resistance. *Cancer Cell***39**, 999-1014 e1018, doi:10.1016/j.ccell.2021.06.003 (2021).
55. Kuusanmaki, H. *et al.* Phenotype-based drug screening reveals association between venetoclax response and differentiation stage in acute myeloid leukemia. *Haematologica***105**, 708-720, doi:10.3324/haematol.2018.214882 (2020).
56. Singh Mali, R. *et al.* Venetoclax combines synergistically with FLT3 inhibition to effectively target leukemic cells in FLT3-ITD+ acute myeloid leukemia models. *Haematologica***106**, 1034-1046, doi:10.3324/haematol.2019.244020 (2021).

Figures

Figure 1

Measuring correlation across data modalities. (A) Correlation between mRNA and protein levels for individual genes in the FLT3-MAPK signaling pathway. Correlation values map to legend inset. (B) Expression of transcripts in the FLT3-MAPK signaling pathway ordered by patient response to quizartinib. (C) Expression of proteins in the FLT3-MAPK signaling pathway ordered by patient response to quizartinib. (D) Correlation between mRNA and protein levels for individual genes in the MAPK signaling pathway. Legend is the same as A. (E) Expression of transcripts in the MEK1/2 signaling pathway ordered by patient response to trametinib (F) Expression of proteins in the MEK1/2 signaling pathway ordered by patient response to trametinib

Figure 2

Linear modeling description and performance. (A) Summary of drug response values across 26 drugs and 31 patient samples together with the data available for each patient sample (across top). (B) Cross-validation performance of Elastic Net, LASSO, and Logistic regression with various types of data or combinations of data types. Performance is measured by Spearman correlation with held out dataset.

Figure 3

Interpretation of signatures by heatmap and protein network integration. (A) Heatmap of proteins selected by the logistic regression depicts clustering of patients by sensitivity to quizartinib. (B) Interaction network links proteins selected by signature (yellow) to other proteins (blue) to illustrate how they relate to one another. (C) Heatmap of 8 somatic mutations selected by the logistic regression shows how they cluster patients by trametinib sensitivity. (D) Interaction network showing how 7 of those genes (yellow) are closely related via protein interactions with related proteins (blue).

Figure 4

All proteomic signatures selected from *ex vivo* modeling cluster MOLM13 parental cell lines from those that are resistant to trametinib. (A) Heatmap of signature selected from LASSO regression. (B) Heatmap of signature selected from Logistic regression. (C) Heatmap of signature selected from Elastic Net regression.

Figure 5

Proteomic signatures selected to predict *ex vivo* response to quizartinib using the (A) LASSO, (B) logistic and (C) Elastic Net regression cluster quizartinib resistant cells from parentals. (D) The same signature from (A) but with additional cell lines that were developed as models of late and early resistance. (E) Network construction of proteins from the LASSO signature (yellow) and intermediate proteins (blue).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaltables.xlsx](#)
- [SupFigures.docx](#)