Bioinformatics, 31(23), 2015, 3799–3806
doi: 10.1093/bioinformatics/btv427
Advance Access Publication Date: 23 July 2015
Original Paper

OXFORD

Systems biology

Identifying kinase dependency in cancer cells by integrating high-throughput drug screening and kinase inhibition data

Karen A. Ryall[1], Jimin Shin[1], Minjae Yoo[1], Trista K. Hinz[2], Jihye Kim[1], Jaewoo Kang[3], Lynn E. Heasley[2] and Aik Choon Tan[1,3,4]*

[1]Translational Bioinformatics and Cancer Systems Biology Laboratory, Division of Medical Oncology, Department of Medicine, [2]Department of Craniofacial Biology, School of Dental Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA, [3]Department of Computer Science and Engineering, Korea University, Seoul, Korea and [4]Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

*To whom correspondence should be addressed.
Associate Editor: Jonathan Wren

Received on May 20, 2015; revised on June 26, 2015; accepted on July 16, 2015

Abstract

Motivation: Targeted kinase inhibitors have dramatically improved cancer treatment, but kinase dependency for an individual patient or cancer cell can be challenging to predict. Kinase dependency does not always correspond with gene expression and mutation status. High-throughput drug screens are powerful tools for determining kinase dependency, but drug polypharmacology can make results difficult to interpret.

Results: We developed Kinase Addiction Ranker (KAR), an algorithm that integrates high-throughput drug screening data, comprehensive kinase inhibition data and gene expression profiles to identify kinase dependency in cancer cells. We applied KAR to predict kinase dependency of 21 lung cancer cell lines and 151 leukemia patient samples using published datasets. We experimentally validated KAR predictions of FGFR and MTOR dependence in lung cancer cell line H1581, showing synergistic reduction in proliferation after combining ponatinib and AZD8055.

Availability and implementation: KAR can be downloaded as a Python function or a MATLAB script along with example inputs and outputs at: http://tanlab.ucdenver.edu/KAR/.

Contact: aikchoon.tan@ucdenver.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Kinases play essential roles in cell survival, growth and proliferation and are currently the largest protein class in clinical trials (Rask-Andersen *et al.*, 2014). Kinases are frequently mutated in cancer and acquire oncogenic properties to drive tumorgenesis. These cancer cells are often 'addicted' to the mutated oncogenes (e.g. kinases). Targeted cancer therapies have exploited this 'oncogene addiction' concept, and deployed small molecules that could inhibit these oncogenic kinases (Sawyers, 2004). While kinases are predominantly targeted for cancer therapy, they are also implicated in immunological, neurological, metabolic and infectious diseases (Zhang *et al.*, 2009). Induction of cell death through inhibition of a specific essential kinase creates selective pressure for cancer cells to develop resistance mechanisms. Cancer cells often acquire resistance through mutations that interfere with drug binding (Azam *et al.*, 2008). Other resistance mechanisms include target amplification, upregulation of alternative kinase pathways, and intrinsic resistance of a subset of cells in the larger population (Glickman and Sawyers, 2012; Sun and Bernards, 2014). Combination of kinase inhibitors could limit development of these resistance pathways and dramatically improve

cancer therapy (Al-Lazikani *et al.*, 2012). In order for combination therapy to be more widely adopted, new systems approaches are needed to prioritize target combinations for experimental validation (Ryall and Tan, 2015).

Before targeted kinase inhibitor therapies can be applied, kinase dependency within a cancer cell needs to be established. High-throughput pharmacological screening is a powerful method for determining kinase dependency, (Garnett *et al.*, 2012; Barretina *et al.*, 2012). However, due to unexpected drug-kinase interactions (polypharmacology), target deconvolution for drug screening data remains a challenge in chemical systems biology. Moreover, highly expressed kinases are not always effective molecular targets in cancer (Wei *et al.*, 2006). Unfortunately, the large number off-target interactions of most kinase inhibitors can lead to misinterpretation of drug screening results. For example, the commonly reported targets of FDA-approved drug bosutinib are SRC and ABL; however, bosutinib also inhibits another 40 kinases by more than 85% inhibition at 500 nM (Anastassiadis *et al.*, 2011). While this poses a challenge for target deconvolution, it also provides a unique opportunity to study the effects of a more comprehensive set of kinases as well as combinations of kinases in a given screen. As quantitative kinase inhibition data is becoming increasingly available (Davis *et al.*, 2011; Anastassiadis *et al.*, 2011), it can be used to better identify critical kinases following drug screens.

Here, we developed Kinase Addiction Ranker (KAR), an algorithm that integrates high-throughput drug screening data, comprehensive kinase inhibition data and gene expression profiles to determine kinase dependency in cancer cells. This algorithm was inspired by previous work using kinase inhibition profiles and drug sensitivity data to predict kinase targets for leukemia patients (Tyner *et al.*, 2013). Using publicly available data, we demonstrated the utility of KAR in ranking kinase targets for 21 lung cancer cell lines and used statistical clustering to group cell lines by kinase dependency. We experimentally validated KAR predictions for non-small cell lung cancer cell line H1581. We also applied this approach to previously published data from 151 leukemia patient samples.

2 Methods

2.1 Kinase addiction ranker (KAR) algorithm

We have developed KAR, a computational algorithm that integrates high-throughput drug screening data, comprehensive quantitative drug-kinase binding data, and transcriptomics data to predict kinase dependence in cancer cells (Fig. 1). KAR generates lists of kinases with high correlation with a phenotypic output such as cell proliferation or survival. Kinases are scored based on the sensitivity of each drug that inhibits the kinase. Since the ultimate goal of KAR is to generate lists of kinase targets for therapeutic application, KAR first filters low expressing kinases from subsequent analysis (Fig. 1). This shortens computation time and ensures that each high scoring kinase is expressed in the cell above a user-defined threshold.

After filtering kinases with low gene expression, the drugs used in the screen are sorted into one of five bins based on drug sensitivity (Fig. 1). Drugs meeting the highest sensitivity threshold (e.g. $IC_{50} \leq 1\,\mu M$) are placed into Bin 1. Kinase targets of drugs in Bin 1 receive the highest point value (20 points) by the algorithm. Bin 2 and 3 contain drugs with high (e.g. $IC_{50} \leq 2\,\mu M$) and intermediate (e.g. $IC_{50} \leq 5\,\mu M$) sensitivity values. Kinase targets of drugs in these bins receive fewer points than targets in Bin 1 (10 points and 5 points, respectively). Finally, Bin 4 and 5 contain drugs that do not meet the threshold for sensitivity (e.g. $IC_{50} \geq 5\,\mu M$). Targets of

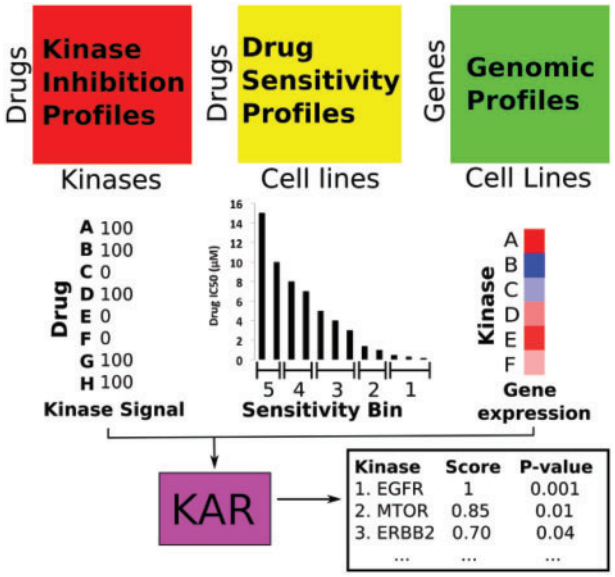

Fig. 1. Kinase Addiction Ranker (KAR) algorithm overview. KAR integrates drug sensitivity, kinase inhibition, and gene expression data to generate a ranked list of kinase targets associated with drug sensitivity. Kinase targets for each drug screened in a cancer sample are scored based on the sensitivity of the drug

drugs in Bin 4 receive no points and targets of drugs in Bin 5 receive negative 10 points by the algorithm. Four thresholds are used to define the bins: Bin 1: $IC_{50} <$ Threshold 1, Bin 2: Threshold $1 \leq IC_{50} <$ Threshold 2, Bin 3: Threshold $2 \leq IC_{50} <$ Threshold 3, Bin 4: Threshold $3 \leq IC_{50} \leq$ Threshold 4, and Bin 5: $IC_{50} >$ Threshold 4. For our lung cancer cell line data we used IC_{50} thresholds of 1, 2, 5 and 10 μM to define the five bins and for the leukemia patient data we used IC_{50} thresholds of 0.5, 1, 2.5 and 5 μM. Lower thresholds were used for the leukemia data since this dataset had smaller ranges of IC_{50} values (max $IC_{50} = 10\,\mu M$). The highest ranking kinases were consistent over a variety of different threshold sets. See Supplementary Tables S1 and S2 for examples of KAR ranking by varying thresholds for two different lung cancer cell lines. The threshold for tiering drug sensitivity in KAR could be tuned for different experiments as deemed appropriate by the user. We used thresholds such that Bin 1 contained the top ~15% most sensitive drugs, Bin 2: top ~15–20%, Bin 3: top ~20–30%, Bin 4: top ~30–40% and Bin 5: bottom ~60%. While we used IC_{50} measurements of sensitivity, the thresholds can easily be tuned to accommodate other measures of sensitivity such as K_i (inhibition constant) or percent of control measurements at a single concentration. Since K_i measurements are less sensitive to assay type, they could be useful for combining data from different experimental sources.

After binning the drugs, a score is calculated for each kinase in the dataset. Each time the kinase is inhibited above the threshold, the appropriate amount of points are added or subtracted based on the bin of the drug that inhibited the kinase (Table 1). If the kinase is not inhibited by the drug, no points are added or subtracted for that drug. We sum over all drugs and kinases to get a final raw score for each kinase. A kinase is considered inhibited based on previously published competitive binding data of hundreds of kinases for each drug. Data using different measurements of inhibition strength (e.g. percent of control, K_d, IC_{50}) were combined to ensure the most comprehensive list of drugs was available for analysis. We defined a kinase target as inhibited if its percent of control measurement at a

Table 1. Example calculation of raw KAR score for Kinase X

Drug	Sensitivity bin	Kinase X inhibited?	Δ score
1	1	Yes	20
2	2	No	0
3	5	Yes	−10
4	1	Yes	20
5	3	No	0
6	4	No	0

Note: If a drug inhibits Kinase X, points are added based on the sensitivity bin of the drug. For example, Drug 1 inhibits kinase X and is in sensitivity bin 1 (highest sensitivity), therefore 20 points are added to the score. Drugs that do not inhibit Kinase X do not affect the score (Drugs 2, 5 and 6). The total Raw score for Kinase X is calculated by summing over all the drugs in the panel.

Raw KAR Score = 20-10 + 20 = 30.

single drug concentration was less than 15% (>85% inhibition of the target) or if its IC_{50} or K_d measurement was <1μM.

Similarly, KAR also calculates scores for each pair of kinases. Here, points are only added or subtracted if two kinases are inhibited above threshold by the same drug. Then the appropriate number of points are given based on the bin of the drug that inhibited the pair of targets. While points are calculated for every possible combination of kinases in the dataset, certain pairs of targets are unlikely to ever have a high score since the number of drugs that inhibit a particular pair may be low or nonexistent. An example is the pair FGFR1 and MTOR, which is only inhibited at the same time by one compound in our dataset, AZD-7762.

KAR scales the raw scores such that the kinase or pair with the highest raw score has a scaled score of 1. This allows for comparison of KAR scores between samples. KAR also calculates a percent effective score by computing the percentage of times the kinase is inhibited and is included in one of the sensitivity bins receiving positive points (sensitivity Bins 1–3). KAR also calculates chi-square and Fisher's exact test *P*-values for the kinases and combinations using the contingency table for two variables: (i) kinase inhibited (>85% inhibition or $IC_{50}/K_d < 1\mu M$) and (ii) drug sensitivity (sensitivity bins 1–3). The *P*-values compute if there is a significant association between a kinase being inhibited and the inhibiting drug being sensitive (sensitivity bins 1–3) and are not related to the KAR score. We sort the table of kinase scores by *P*-value to ensure that kinases that are not inhibited by as many compounds are not overlooked by the algorithm since kinases that are inhibited by more compounds have greater potential to receive higher scores. Each of these scores together result in a more complete picture of the importance of a given kinase or combination to the sample. The key kinases to test in follow-up experiments will likely have significant *P*-values, high scores, and high percent effective values. An example of all the outputs of KAR for one of the cell lines we tested is given in Supplementary Table S3.

2.2 Implementation of KAR

We implemented KAR in MATLAB (version 2015a) and Python Scripting Language (version 2.7.8). We tested KAR in OSX Version 10. 9.5. KAR code is freely available for download at http://tanlab.ucdenver.edu/KAR.

2.3 Drug sensitivity data

We obtained high-throughput pharmacological profiling data for 21 lung cancer cell lines from the Genomics of Drug Sensitivity in Cancer (GDSC) database (Yang *et al.*, 2012) (Supplementary Table S4). Screening data from 151 leukemia patient samples was obtained from a recent publication (Tyner *et al.*, 2013) (Supplementary Table S5).

2.4 Microarray gene expression data

We obtained microarray gene expression data for the 21 lung cancer cell lines from the Cancer Cell Line Encyclopedia (GSE36133). Raw CEL files for these cell lines were normalized using Robust Multiarray Average (RMA) (Irizarry *et al.*, 2003) approach in Affymetrix Power Tools (APT).

2.5 Quantitative kinase inhibition data

For lung cancer cell lines study, we obtained comprehensive quantitative kinase inhibition data for 49 kinase inhibitors used in the GDSC database. References for publications and databases used to acquire the kinase profiles are in Supplementary Table S6. For the leukemia patient study, kinase inhibition data for 66 kinase inhibitors were collected from published papers (Tyner *et al.*, 2013). References for publications and databases used to acquire the kinase profiles are in Supplementary Table S7. Quantitative kinase binding data was dichotomized as inhibited or not inhibited using thresholds of $IC_{50}/K_d < 1\mu M$ or percent inhibition >85%. Databases such as ChEMBL (Bento *et al.*, 2014), PubChem (Wang *et al.*, 2014) and DSigDB (Yoo *et al.*, 2015) are useful resources for finding published quantitative kinase target information.

2.6 Cluster analysis

We performed hierarchical clustering of the data using the MATLAB bioinformatics toolbox with Euclidean distance metric and average linkage to generate the hierarchical tree. Data columns were normalized so that the mean was 0 and the standard deviation was 1.

2.7 Cell proliferation assay

H1581 cells were plated at 100 cells per well in 96-well tissue culture plates and treated with inhibitors at various doses. When the DMSO-treated control wells became confluent (10 days) cell numbers were assessed using a CYQUANT Direct Cell Proliferation Assay (Invitrogen) according to the manufacturer's instructions.

2.8 Quantifying combination effects

To quantify the combination effect of drugs used in this study, we used the Bliss independence model, that predicts the combined response C for two single compounds with effects A and B using the following equation: $C = A \times B$, where each effect is expressed as fractional activity compared to control between 0 (maximal effect, 100% inhibition) and 1 (no effect, 0% inhibition). The combination is synergistic if the %inhibition of the combination is greater than the predicted C.

2.9 Immunoblot analysis

For immunoblot analysis cells were plated at 1.5×10^6 cells per plate in 4–10 cm plates. Twenty-four hours later, cells were switched to HITES media for 2 h and subsequently treated with either DMSO, 100 nM ponatinib, 100 nM AZD8055, or the combination of ponatinib + AZD8055 for 2 h. Cells were collected in PBS, centrifuged (3 min at 3000 rpm) and suspended in lysis buffer. Aliquots of the cell lysates containing 60 μg of protein were submitted to SDS-PAGE and immunoblotted for phospho-mTOR (#5536), mTOR

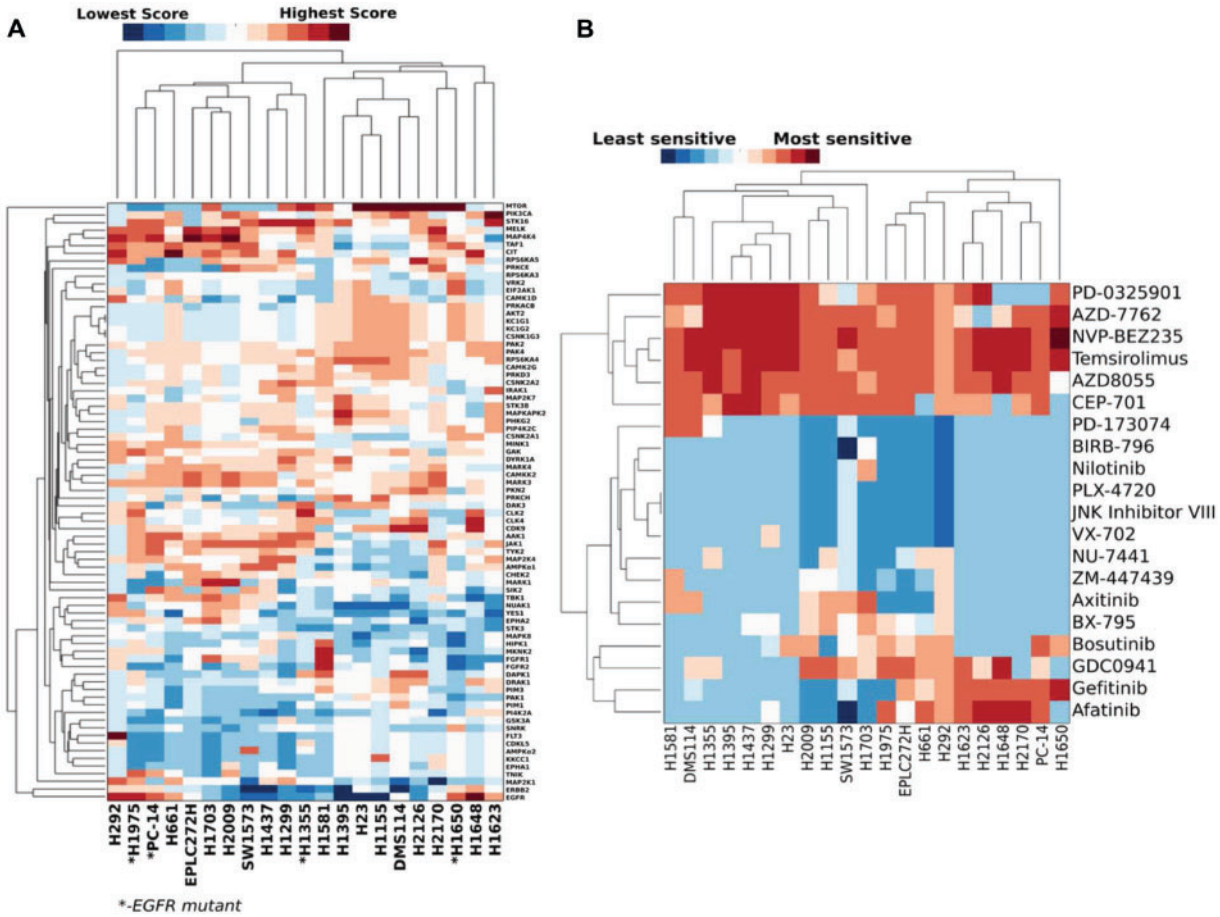

Fig. 2. KAR identifies relationships in kinase dependence among lung cancer cell lines. (**A**) Hierarchical clustering of scaled KAR scores for 21 lung cancer cell lines, showing relationships among cell lines and kinases. (**B**) Hierarchical clustering of drug IC_{50} data for the 20 overlapping drugs in each cell line's dataset. Each column was normalized before clustering to give a mean of 0 and a standard deviation of 1. Red indicates that a cell line has higher sensitivity to a given kinase/drug and blue indicates relatively lower sensitivity. Clustering results differ when clustering based on kinase score vs. drug IC_{50}

(#2983), phospho-Akt S473 (#9271), Akt (#9272), phospho-p70S6K (#9234), p70S6K (#9202), phospho-S6 (#4857), S6 (#2317), phospho-ERK p-p44/42 MAPK (#9101); Cell Signaling Technology and ERK1 (sc-93), ERK2 (SC-154), NaK-ATPase α-subunit (sc-21712); Santa Cruz Biotechnology.

3 Results

3.1 Determining kinase dependency in lung cancer

We initially applied KAR to a panel of 21 lung cancer cell lines using drug sensitivity data from the GDSC database (Yang *et al.*, 2012). We obtained microarray gene expression data from the Cancer Cell Line Encyclopedia (GSE36133). Raw CEL files for these cell lines were normalized using RMA approach (Irizarry *et al.*, 2003) in Affymetrix Power Tools (APT). Genes with expression level lower than seven (log2 signal from RMA) were deemed to be low expressed and filtered out in this study. Drugs were included in our analysis for each kinase inhibitor profiled in GDSC with published kinase inhibition profiles. Sensitivity data for each cell line contained data from between 21 and 49 kinase inhibitors with a median of 30 inhibitors. The list of the kinase inhibitors used in this study is available in Supplementary Table S4. Each pair of cell lines had at least 20 overlapping kinase inhibitors screened in the dataset. The top five ranking kinases and kinase pairs for each cell line are provided in Supplementary Tables S8 and S9.

3.2 Cluster analysis of the kinase dependency in lung cancer cell lines

Hierarchical clustering of the scaled KAR scores (Fig. 2A) reveal relationships in kinase dependence among the lung cancer cell lines and kinases with similar scoring patterns. For example, MELK, MAP4K4 and TAF1 group together and have high scores in the same cell lines. Cell lines H1703, EPLC272H and H2009 group together partially due to high scores in MELK, MAP4K4, TAF1 and CAMKK2. A subset of kinases was selected for clustering by identifying kinases with a significant association with drug sensitivity (Fisher's exact test, $P < 0.05$) in one of the 21 cell lines. MTOR was most frequently significantly associated with drug sensitivity among the lung cancer cell lines studied. This is relevant as MTOR is a key kinase that regulates the survival pathway and has been previously shown to be active in non-small cell cancer (Ekman *et al.*, 2012; Fumarola *et al.*, 2014). Clustering results reveal that MTOR, EGFR and ERBB2 are among the kinases with the most distinct scoring patterns, making the scores of these kinase more unique identifiers of the cell lines (Fig. 2A). This is interesting as EGFR is one of the targets with FDA-approved drugs (e.g. gefitinib and erlotinib) approved for non-small cell lung cancer.

In contrast, clustering based on drug sensitivity (Fig. 2B) instead of kinase score resulted in different groupings of cell lines. For example, H1623 and H2126 clustered together by drug sensitivity partially due to shared sensitivity to EGFR inhibitors gefitinib and

Table 2. Subset of the KAR output for three lung cancer cell lines: H1975, H1299 and H1581

Cell Line	Top 5 kinases (Scaled score, P-value) ranked by KAR				
	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
H1975	MAP4K4 (0.85, 0.014)	TNIK (0.85, 0.014)	MELK (0.85, 0.028)	STK16 (0.77, 0.006)	GAK (0.77, 0.014)
H1299	STK16 (1.00, 0.0004)	CSNK2A2 (0.92, 0.001)	AAK1 (0.92, 0.003)	MTOR (0.92, 0.003)	DYRK1A (0.85, 0.003)
H1581	FGFR1 (1.00, 0.001)	FGFR2 (1.00, 0.001)	MKNK2 (0.86, 0.003)	MTOR (0.79, 0.011)	HIPK1 (0.79, 0.011)

Note: The top 5 ranking kinases and pairs of kinase for all 21 lung cancer cell lines tested are available in Supplementary Tables S8 and S9.

afatinib. However, when clustering based on kinase score (Fig. 2A), we see more separation between these cell lines, with H1623 having a higher EGFR score than H2126. Data from drugs with strong off-target effects on EGFR such as AZD-7762, bosutinib and ponatinib decrease the EGFR score for H2126. The KAR score clustering also highlights the high dependence of H2126 on MTOR and CDK9 compared to H1623. This is consistent with published data suggesting that H2126 acquired resistance to EGFR inhibition via activation of the AKT/MTOR pathway (Wu *et al.*, 2013). This highlights that clustering based on KAR could delineate the kinase dependency in individual cell lines, which is not possible to distinguish based on drug sensitivity data clusters.

Additionally, H292 grouped together with other cell lines inhibited by bosutinib like H661 (Fig. 2B), but when we cluster based on kinase score we more clearly see that H292 has a unique sensitivity to FLT3 inhibition. These examples further illustrate how incorporation of comprehensive kinase inhibition profiles for each drug allows for better kinase target deconvolution.

KAR results also demonstrate that effective kinase targets cannot be predicted based on gene expression alone. While some cell lines have high kinase scores for kinases with high gene expression (e.g. PC-14 - EGFR), many high expressing kinases have low associations with drug sensitivity. For example, AURKA is one of the kinases with the highest gene expression in cell line H1299, but has a scaled score of −0.14 and a chi-square *P*-value of 0.784, indicating low correlation with drug sensitivity. Another example is high MET expression in H1975, but a scaled KAR score of 0.38 and a chi-square *P*-value of 1.00.

3.3 Validation of KAR-predicted kinase dependency

To demonstrate that KAR could delineate kinase dependency in individual cell lines, we validated KAR algorithm predictions for three non-small cell lung cancer cell lines: H1975, H1299 and H1581 (Table 2) based on published literature and experimental results.

3.3.1 Validation of kinase dependency in H1975

H1975 is an EGFR double-mutant (L858R, T790M) cell line with high EGFR gene expression. The first mutation (L858R) correlates with sensitivity to EGFR kinase inhibitors (e.g. erlotinib and gefitinib). In contrast, the second mutation (T790M) is the gatekeeper mutation that generates resistance to the first-generation EGFR inhibitors (erlotinib and gefitinib). While EGFR was the highest scoring kinase by KAR, the association between EGFR and drug sensitivity was not significant (chi-square $P = 0.08$). This is due to the T790M mutation that confers resistance to the EGFR drug sensitivity for this cell line. EGFR, however, was present in significant inhibition pairs with TNIK and GAK (chi-square $P = 0.01$), which are inhibited by the dual SRC/ABL inhibitor bosutinib. High scoring single targets TNIK, MAP4K4, STK3, AAK1, GAK and EGFR are also inhibited by bosutinib. These findings are supported by a previous study demonstrating decreased proliferation in H1975 after combining EGFR inhibitor gefitinib with bosutinib compared to either drug alone (Kim *et al.*, 2014).

3.3.2 Validation of kinase dependency in H1299

KAR results for H1299 showed high ranking for casein kinase 2 alpha (Table 2) and several high scoring combinations of kinases containing casein kinase 2 (Supplementary Table S9). This target is supported by a previous study showing that CX-4945, a selective casein kinase 2 alpha inhibitor, induces dose-dependent decreases in cell proliferation in H1299 and has an IC_{50} of 1.8 μM (Zhang *et al.*, 2013). More recently, CX-4945 was shown to down-regulate AKT/mTOR signaling pathway in H1299 and induces apoptosis in this cell line (So *et al.*, 2015). This supports that H1299 is dependent on the CNSK2A1 and MTOR kinases as predicted by KAR.

3.3.3 Validation of kinase dependency in H1581

The highest scoring kinases for FGFR1 amplified NSCLC cell line H1581 were FGFR1, FGFR2, MKNK2 and MTOR (Table 2). H1581 was the only lung cancer cell line analyzed with a significant association between FGFR1 inhibition and drug sensitivity. Previously, we have demonstrated that H1581 is a cell line that has high FGFR1 gene and protein expressions, and this cell line is sensitive to ponatinib, a FGFR1 inhibitor (Wynes *et al.*, 2014). Moreover, MTOR was identified as a biomarker of resistance to targeted therapy in recent studies in breast cancer and melanoma(Corcoran *et al.*, 2013; Kelsey and Manning, 2013; Elkabets *et al.*, 2013). We experimentally tested this prediction by combining FGFR1 (ponatinib) and MTOR (AZD8055) inhibitors in this cell line. Experimental results show enhanced reduction in proliferation with the combination and the combination is synergistic by Bliss independence (Fig. 3A). Western blots confirm decreased ERK1/2 activation with ponatinib and decreased mTOR, AKT, p70S6K and S6 activation with AZD8055 (Fig. 3B-C). More importantly, the combination of ponatinib and AZD8055 shows inhibition of both ERK and MTOR pathways (Fig. 3B-C). This results is consistent with a recent kinome-wide RNAi screens that identified MTOR is synthetic lethal with FGFR1 in this cell line (Singleton *et al.*, 2015).

Taken together, we demonstrated that KAR-predicted kinase dependency in these lung cancer cell lines could be validated by experimental results and/or published literature. This supports the utility of KAR in delineating kinase dependency by integrating high-throughput drug screening data and *in vitro* kinase binding data.

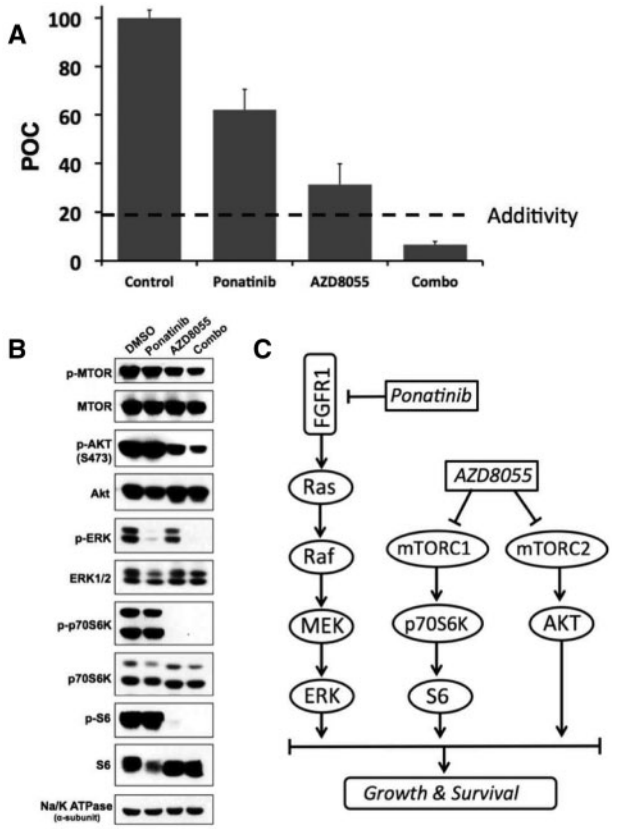

Fig. 3. Experimental validation of KAR prediction of FGFR1 and mTOR dependence for lung cancer cell line H1581. (**A**) 10 nm ponatinib (FGFR inhibitor) and AZD8055 (mTOR inhibitor) were applied to H1581 cells and cell proliferation was measured with the CYQUANT assay kit. The combination of ponatinib and AZD8055 was synergistic by Bliss Independence (Additive prediction = 19.4%, Actual = 6.7%). Bar graphs display mean percent of control (POC) +/− SEM (**B**) Western blots showing changes in signaling with ponatinib, AZD8055, and the combo. Ponatinib decreases ERK1/2 activation and AZD8055 decreases signaling downstream of mTOR. (**C**) Signaling network diagram for pathways targeted by ponatinib and AZD8055 in H1581

3.4 Deciphering kinase dependency in leukemia patient samples

Next, we applied KAR to a dataset of 151 leukemia patient samples screened with 66 kinase inhibitors (Tyner *et al.*, 2013). The list of the kinase inhibitors used in this study is available in Supplementary Table S5. Since gene expression data was not available, no kinases were filtered prior to scoring. As with the lung cancer data, we applied hierarchical clustering to the scaled KAR scores for a subset of the kinases for each patient sample to observe relationships in scoring patterns (Fig. 4). We selected the 50 kinases with the highest variance in score and the 10 kinases most commonly significantly associated with drug sensitivity (chi-square $P < 0.05$). Even with patients grouped by disease type, we see a large variance in kinase dependence among the patients with no set of kinases uniquely identifying a disease type. This variance further illustrates the need for pharmacological screens to help plan targeted patient therapy.

FLT3, a kinase with mutations in up to 30% of acute myeloid leukemia (AML) patients (Zarrinkar *et al.*, 2009), had the highest variance in score among the kinases analyzed. EPHA5, EPHA3 and BTK were most commonly significantly associated with drug sensitivity. These kinases had significant associations in 72, 58 and 54% of the patient samples, respectively (Supplementary Table S10). Eph receptors have been shown to affect cancer growth, migration and invasion *in vitro* and *in vivo* (Pasquale, 2010). Consistently, a RNAi screen identified EPHA5 sensitivity in a subset of the 30 patient leukemia samples studied (Tyner *et al.*, 2009). Interestingly, Eph receptors use bidirectional signaling mechanisms to induce both tumor promotion and suppression (Pasquale, 2010). The frequency of BTK dependence is interesting given a phase IB/II clinical trial of BTK inhibitor ibrutinib resulting in a high frequency of durable remissions in patients with chronic lymphocytic leukemia (CLL) (Byrd *et al.*, 2013). The progression-free survival rate at 26 months was 75%. This is consistent with our data showing 70% of CLL patient data had significant association between BTK inhibition and drug sensitivity (Supplementary Table S10).

4 Discussion

We developed and validated KAR, a novel algorithm to improve interpretation of high-throughput drug screens by incorporating comprehensive drug-kinase binding profiles and transcriptomics data. KAR takes advantage of drug polypharmacology to study a larger variety of kinases as well as combinations of kinases. Two major factors that could influence KAR data analysis are the (i) number of effective drugs and (ii) diversity of drug targets. Influential kinases cannot be rapidly identified without multiple inhibitors in your drug set that target a given kinase. The KAR percent effective scores from a preliminary screen can be used to identify kinases with potential associations with drug sensitivity for further analysis with other drugs. Moreover, many kinase pairs are uncommonly inhibited together (e.g. FGFR1 and MTOR), and must be hypothesized based on the single kinase scores. While our algorithm most directly applies to studying kinase dependence, it could be easily modified to study other targets if inhibition data for these targets is available.

Other approaches have been developed that use overlap in drug kinase profiles to identify key targets. Gujral *et al.* used principal component analysis to identify an optimal set of 32 kinase inhibitors for profiling and then used elastic net regularization to identify key kinases influencing cell migration(Gujral *et al.*, 2014). Similar to KAR, they also used gene expression data to filter kinases from analysis. Tran *et al.* also applied elastic net regression to identify important kinases for cancer cell lines following an *in vitro* screen (Tran *et al.*, 2014). Another algorithm based on set theory uses drug screen data and kinase inhibitor profiles to predict the most influential kinases and produce circuit diagrams illustrating if the kinase is effective inhibited alone or if it needs to be inhibited with other kinases (Berlow *et al.*, 2013). Compared to these approaches, KAR benefits from producing straightforward scores and *P*-values that could be readily interpreted by scientists without computational backgrounds. Moreover, the drug lists do not need preliminary optimization, as chi-square and fisher's exact test *P*-values take differences in the number of inhibitors that target each kinases and the total number of sensitive drugs into account.

We applied KAR to leukemia patient samples profiled with 66 kinase inhibitors (Tyner *et al.*, 2013), demonstrating the applicability of the tool for predicting patient therapy. Given resource limitations when working with patient samples, it may not be possible to screen patient biopsies with large numbers of compounds. Therefore, future studies could benefit greatly from prior optimization of the set of drugs used for profiling. One recent example of this is Gujral *et al.*'s (2014) use of principal component analysis to reduce the number of kinase inhibitors profiled to an optimal set of 32 from 178. Moreover, instead of using IC_{50} measurements of drug sensitivity, which requires measurements at multiple concentrations,

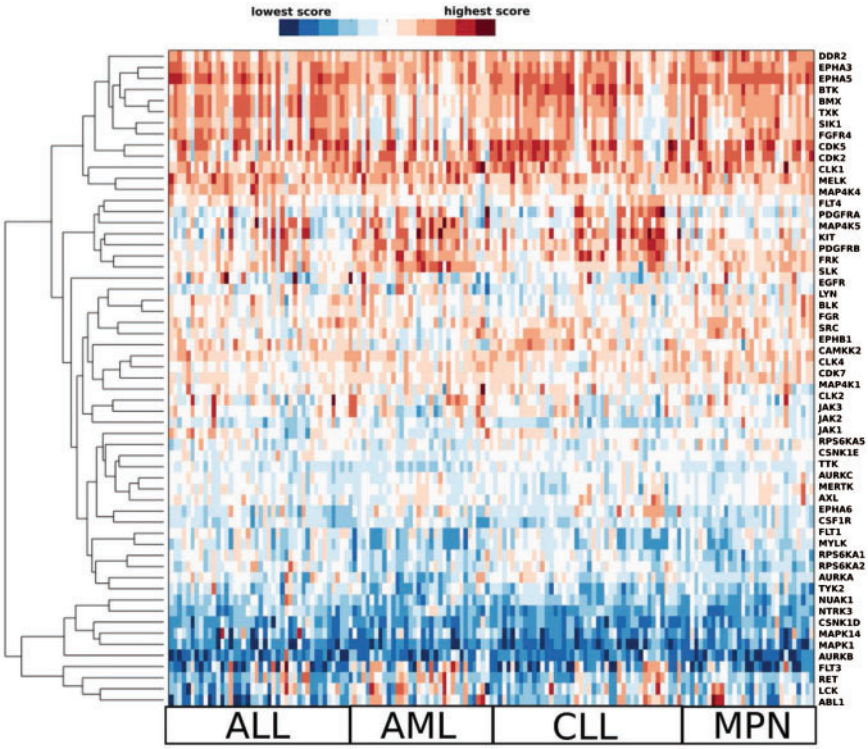

Fig. 4. KAR identifies relationships in kinase dependence in leukemia patient samples. Hierarchical clustering of scaled KAR scores for a subset of kinases for 151 leukemia patient samples (Tyner *et al.*, 2013). Samples (columns) are organized by disease type. Each column was normalized before clustering to give a mean of 0 and a standard deviation of 1. Red indicates high dependence on a given kinase and blue indicates less dependence. The kinases with the highest scores appear highly varied among the patients, even when organized by disease type (ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; CLL, chronic lymphocytic leukemia; MPN, other myeloproliferative neoplasms)

cell viability measurements at single concentrations can be considered.

KAR was inspired by a previous algorithm implemented using Excel Visual Basic and macro code (Tyner *et al.*, 2013). Compared to the Tyner algorithm, we introduced tiering score values based on drug sensitivity instead of target inhibition strength, percent effective scores, optional incorporation of gene expression data, stronger penalties for insensitive drugs, and calculation of *P*-values. We also made the algorithm more accessible by providing MATLAB and python functions. Calculation of *P*-values helps decrease potential for false positives, as kinases targeted by more drugs have the potential for higher raw scores. Incorporation of gene expression data helps ensure that highly ranked kinases are translationally meaningful. Moreover, we found that tiering kinase scores based on target inhibition strength instead of drug sensitivity resulted in much lower percent effective scores, indicating that weaker targets of the inhibitors may not be accurate indicators of kinase dependency in the samples. Moreover, a single threshold for kinase inhibition allows for easier incorporation of kinase inhibition data from multiple platforms with different inhibition measurement types (e.g. percent inhibition compared to control, K_d, and IC_{50}), allowing for more drugs to be included in analysis. We believe that integrating high-throughput drug screening data and *in vitro* kinome inhibition data as demonstrated in this study could be a useful systems approach to identify novel targets and kinase dependency in cancer cells.

5 Conclusions

In summary, KAR integrates drug sensitivity, comprehensive kinase inhibition data and gene expression profiles to identify kinases dependency in cancer cells. We applied KAR to published drug screen data from lung cancer cell lines and leukemia patient samples. Clustering analysis revealed lung cancer cell lines with similarities in kinase dependence. We experimentally validated KAR predictions of FGFR1 and MTOR dependence in lung cancer cell line H1581. Our analysis revealed candidate kinases as potential targets in lung cancer and leukemia for further pharmacological and biological studies. We believe that the research reported in this study provides a new research strategy to delineate kinase dependency in cancer cells. This approach can be applied to other cancer cell lines and patient tumor samples to discover effective kinase targets for personalized medicine.

Funding

This work is partly supported by the National Institutes of Health under Ruth L. Kirschstein National Research Service Award T32CA17468 (K.A.R.), the National Institutes of Health P50CA058187, P30CA046934, Cancer League of Colorado, and the David F. and Margaret T. Grohne Family Foundation. Lynn Heasley is supported in part by a research grant from ARIAD Pharmaceuticals.

Conflict of Interest: none declared.

References

Al-Lazikani,B. *et al.* (2012) Combinatorial drug therapy for cancer in the post-genomic era. *Nat. Biotechnol.*, **30**, 679–692.

Anastassiadis,T. *et al.* (2011) Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat. Biotechnol.*, **29**, 1039–1045.

Azam,M. *et al.* (2008) Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nat. Struct. Mol. Biol.*, **15**, 1109–1118.

Barretina,J. *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.

Berlow,N. *et al.* (2013) A new approach for prediction of tumor sensitivity to targeted drugs based on functional data. *BMC Bioinformatics*, **14**, 239.

Bento,A.P.*et al.* (2014) The ChEMBL bioactivity database: an update. *Nucl Acids Res.*, **42**, D1083–D1090.

Byrd,J.C. *et al.* (2013) Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. *N. Engl. J. Med.*, **369**, 32–42.

Corcoran,R.B. *et al.* (2013) TORC1 suppression predicts responsiveness to RAF and MEK inhibition in BRAF-mutant melanoma. *Sci. Transl. Med.*, **5**, 196ra98.

Davis,M.I. *et al.* (2011) Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.*, **29**, 1046–1051.

Ekman,S. *et al.* (2012) The mTOR pathway in lung cancer and implications for therapy and biomarker analysis. *J. Thoracic Oncol.*, **7**, 947–953.

Elkabets,M. *et al.* (2013) mTORC1 inhibition is required for sensitivity to PI3K p110a inhibitors in PIK3CA-mutant breast cancer. *Sci. Transl. Med.*, **5**, 196ra99.

Fumarola,C. *et al.* (2014) Targeting PI3K/AKT/mTOR pathway in non small cell lung cancer. *Biochem. Pharmacol.*, **90**, 197–207.

Garnett,M.J. *et al.* (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, **483**, 570–575.

Glickman,M.S. and Sawyers,C.L. (2012) Converting cancer therapies into cures: lessons from infectious diseases. *Cell*, **148**, 1089–1098.

Gujral,T.S. *et al.* (2014) Exploiting polypharmacology for drug target deconvolution. *Proc. Natl Acad. Sci. USA*, **111**, 5048–5053.

Irizarry,R.A. *et al.* (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.

Kelsey,I. and Manning,B.D. (2013) mTORC1 status dictates tumor response to targeted therapeutics. *Sci. Signal.*, **6**, pe31.

Kim,J. *et al.* (2014) Bioinformatics-driven discovery of rational combination for over-coming EGFR-mutant lung cancer resistance to EGFR therapy. *Bioinformatics*, **30**, 2393–2398.

Pasquale,E.B. (2010) Eph receptors and ephrins in cancer: bidirectional signalling and beyond. *Nat. Rev. Cancer*, **10**, 165–180.

Rask-Andersen,M. *et al.* (2014) The druggable genome: evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. *Annu. Rev. Pharmacol. Toxicol.*, **54**, 9–26.

Ryall,K.A. and Tan,A.C. (2015) Systems biology approaches for advancing the discovery of effective drug combinations. *J. Cheminf.*, **7**, 7.

Sawyers,C. (2004) Targeted cancer therapy. *Nature*, **432**, 294–297.

Singleton,K.R. *et al.* (2015) Kinome RNAi screens identify MTOR for synergistic targetting with FGFR1 in lung cancer and HNSCC cell lines. *Cancer Res.*, In press.

So,K.S. *et al.* (2015) AKT/mTOR down-regulation by CX-4945, a CK2 inhibitor, promotes apoptosis in chemorefractory non-small cell lung cancer cells. *Anticancer Res.*, **35**, 1537–1542.

Sun,C. and Bernards,R. (2014) Feedback and redundancy in receptor tyrosine kinase signaling: relevance to cancer therapies. *Trends Biochem. Sci.*, **39**, 465–474.

Tran,T.P. *et al.* (2014) Prediction of kinase inhibitor response using activity profiling, in vitro screening, and elastic net regression. *BMC Syst. Biol.*, **8**, 74.

Tyner,J.W. *et al.* (2013) Kinase pathway dependence in primary human leukemias determined by rapid inhibitor screening. *Cancer Res.*, **73**, 285–296.

Tyner,J.W. *et al.* (2009) RNAi screen for rapid therapeutic target identification in leukemia patients. *Proc. Natl Acad. Sci. USA*, **106**, 8695–8700.

Wang,Y. *et al.* (2014) PubChem BioAssay: 2014 update. *Nucleic Acids Res.*, **42**, D1075–D1082.

Wei,G. *et al.* (2006) Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell*, **10**, 331–342.

Wu,K. *et al.* (2013) Gefitinib resistance resulted from STAT3-mediated Akt activation in lung cancer cells. *Oncotarget*, **4**, 2430–2438.

Wynes,M.W. *et al.* (2014) FGFR1 mRNA and protein expression, not gene copy number, predict FGFR TKI sensitivity across all lung cancer histologies. *Clin. Cancer Res.*, **20**, 3299–3309.

Yang,W. *et al.* (2012) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.*, **41**, D955–D961.

Yoo,M. *et al.* (2015) DSigDB: drug signatures database for gene set analysis. *Bioinformatics*, **31**, 3069–3071.

Zarrinkar,P.P. *et al.* (2009) AC220 is a uniquely potent and selective inhibitor of FLT3 for the treatment of acute myeloid leukemia (AML). *Blood*, **114**, 2984–2992.

Zhang,J. *et al.* (2009) Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer*, **9**, 28–39.

Zhang,S. *et al.* (2013) Inhibition of CK2α down-regulates Notch1 signalling in lung cancer cells. *J. Cell. Mol. Med.*, **17**, 854–862.