Bioinformatics-driven discovery of rational combination for overcoming EGFR-mutant lung cancer resistance to EGFR therapy

Jihye Kim[1], Vihas T. Vasu[2], Rangnath Mishra[2], Katherine R. Singleton[3], Minjae Yoo[1], Sonia M. Leach[2], Eveline Farias-Hesson[2], Robert J. Mason[2,4], Jaewoo Kang[5], Preveen Ramamoorthy[2], Jeffrey A. Kern[2,4], Lynn E. Heasley[3], James H. Finigan[2,4,*,†] and Aik Choon Tan[1,5,*,†]

[1]Division of Medical Oncology, Department of Medicine, Translational Bioinformatics and Cancer Systems Biology Laboratory, University of Colorado Anschutz Medical Campus, 80045 Aurora, [2]Department of Medicine, National Jewish Health, 80206 Denver, [3]Department of Craniofacial Biology, School of Dental Medicine, [4]Division of Pulmonary Sciences and Critical Care Medicine, University of Colorado Anschutz Medical Campus, 80045 Aurora, CO, USA and [5]Department of Computer Science and Engineering, Korea University, Seoul 136-713, Korea

Associate Editor: Inanc Birol

ABSTRACT

Motivation: Non–small-cell lung cancer (NSCLC) is the leading cause of cancer death in the United States. Targeted tyrosine kinase inhibitors (TKIs) directed against the epidermal growth factor receptor (EGFR) have been widely and successfully used in treating NSCLC patients with activating EGFR mutations. Unfortunately, the duration of response is short-lived, and all patients eventually relapse by acquiring resistance mechanisms.

Result: We performed an integrative systems biology approach to determine essential kinases that drive EGFR-TKI resistance in cancer cell lines. We used a series of bioinformatics methods to analyze and integrate the functional genetics screen and RNA-seq data to identify a set of kinases that are critical in survival and proliferation in these TKI-resistant lines. By connecting the essential kinases to compounds using a novel kinase connectivity map (K-Map), we identified and validated bosutinib as an effective compound that could inhibit proliferation and induce apoptosis in TKI-resistant lines. A rational combination of bosutinib and gefitinib showed additive and synergistic effects in cancer cell lines resistant to EGFR TKI alone.

Conclusions: We have demonstrated a bioinformatics-driven discovery roadmap for drug repurposing and development in overcoming resistance in EGFR-mutant NSCLC, which could be generalized to other cancer types in the era of personalized medicine.

Availability and implementation: K-Map can be accessible at: http://tanlab.ucdenver.edu/kMap.

Contact: aikchoon.tan@ucdenver.edu or finiganj@njhealth.org

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

Received on October 13, 2013; revised and accepted on May 1, 2014

1 INTRODUCTION

Genomic medicine has dramatically increased our knowledge of the molecular changes that underpin disease states. Understanding alterations in gene expression can identify proteins and signaling pathways, which might serve as therapeutic targets. Moreover, recent technologic advances in next-generation sequencing facilitate the rapid assessment of gene expression changes in specific patients, allowing for individualized treatment. Although personalized genomics heralds the age of precision medicine, the comprehensive datasets obtained through sequencing require sophisticated bioinformatics tools to identify the critical genes that influence disease. Once these critical genes are identified, the next challenge is to predict what drugs would be useful in reversing the disease states.

The connectivity map represents the first attempt to provide a computational framework to connect genes, drugs and diseases based on gene expression signatures (Lamb *et al.*, 2006). This method assumes gene expression changes could be used as a 'universal language' to connect distinct biological states (e.g. diseases), allowing for the successful repurposing of compounds (Hieronymus *et al.*, 2006; Wei *et al.*, 2006). In short, drugs known to be effective in one disease can serve as candidates for use in other diseases marked by similar gene expression changes. The power of this method has inspired other related work (Chung *et al.*, 2014; Li *et al.*, 2009; Zhang and Gant, 2008, 2009) with the goal of improving the utility of the connectivity map in drug repurposing and development.

Non–small-cell lung cancer (NSCLC) serves as an ideal disease for a connectivity map-based approach. NSCLC accounts for ∼85% of lung cancers, and is the leading cause of cancer-related death in the United States (Siegel *et al.*, 2013) and worldwide. Comprehensive characterization of cancer genomes have increased our understanding of cancer biology and moved NSCLC beyond standard clinico-pathologic classifications and staging to include molecular characterization based on newly identified oncogenic drivers, such as the mutant epidermal growth factor receptor gene (*EGFR*) (Pao and Chmielecki, 2010). Although therapies directed against EGFR, such as gefitinib (Hirsch *et al.*, 2003), have revolutionized NSCLC care, the duration of response is short-lived, and all patients eventually relapse through acquired resistance mechanisms
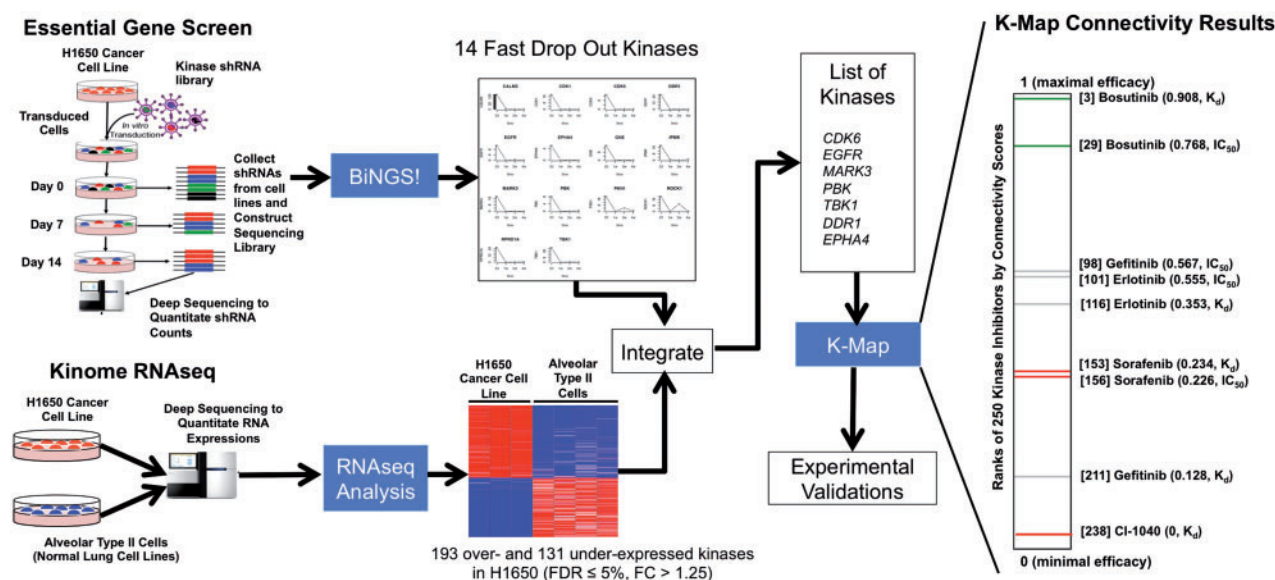
*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the last two authors should be regarded as Joint Last Authors.

Fig. 1. Workflow of the experimental and bioinformatics analyses of this study. Blue boxes represent bioinformatics analyses. The K-Map connectivity results for validation (right)

(Casás-Selves *et al.*, 2012; Engelman and Jänne, 2008; Ohashi *et al.*, 2013; Ware *et al.*, 2013). Identification of new targeted therapeutics is a priority, yet the ability to translate genomic datasets into new drugs has been limited.

We devised an integrative systems biology approach using a functional screen combined with RNA-seq expression data to determine essential kinases other than EGFR that drive proliferation and survival in EGFR-mutant NSCLC. We performed an initial focused, functional genetic screen using short hairpin RNAs (shRNAs) that target the kinome (~600 kinases) in the H1650 lung cancer cell line. This line harbors an EGFR mutation, yet is intrinsically resistant to EGFR TKIs. We used Bioinformatics for Next-Generation Sequencing! (BiNGS!) (Kim and Tan, 2012), a novel bioinformatics pipeline to analyze and interpret functional genomic screening data by next-generation sequencing to identify essential kinases that drive the survival and proliferation signaling pathways of this EGFR-mutant line. Next, we determined the differentially expressed kinases in this cancer cell line by next-generation sequencing (RNA-seq). Integrative analysis was performed to identify kinases that were essential and dysregulated, and the list of essential and functional kinases was connected to explore therapeutic opportunities using the kinase connectivity map (K-Map) (Kim *et al.*, 2013). The efficacy of the K-map prediction of kinase inhibitors was then validated *in vitro* (Fig. 1).

2 MATERIALS AND METHODS

Cell culture. H1650 and H1975 cells were obtained from the University of Colorado Cancer Center (UCCC) Tissue Culture Core. All cell lines were routinely cultured in RPMI-1640 growth medium supplemented with 10% fetal bovine serum (Sigma, St. Louis, MO, USA) at 37°C in a humidified 5% CO_2 incubator. Alveolar type II (ATII) cells are primary, normal cells that were isolated and cultured in an air/liquid interface, as previously described (Wang *et al.*, 2007).

Kinome essential screen by next-generation sequencing. An essential screen identifies genes critical for cell survival by knocking out individual genes in cells (one gene deleted per individual cell). The essential screen is a high-throughput assessment of gene function as opposed to expression. A total of 3×10^6 cells of H1650 were transduced with the short-hairpin loop lentiviral kinome library (~3700 shRNAs targeting ~600 kinases) developed by the RNAi Consortium (TRC 1.0/1.5) and obtained from the UCCC Functional Genomics Shared Resource Core. Cells were transduced with the lentiviral shRNA library using conditions to result in a single shRNA expressed in each infected cell. Cells were cultured and harvested after 2, 7, 14 and 28 days of transduction. shRNAs from surviving cells were extracted, reverse transcribed and barcoded for individual replicates (four replicates per time point). The DNA was purified and amplified, and Illumina adapters were added. The relative abundance of the unique shRNA tags was quantified by the Illumina Genome Analyzer, as previously described (Casás-Selves *et al.*, 2012; Singleton *et al.*, 2013; Spreafico *et al.*, 2013; Sullivan *et al.*, 2012). Loss of shRNAs represents essential kinases. See Supplementary Methods for details.

BiNGS! analysis. We used BiNGS! for analyzing and interpreting the essential screen (Kim and Tan, 2012). In brief, a preprocessing step filtered out erroneous and low-quality reads. Filtered reads were mapped against the shRNA reference library using Bowtie (Langmead *et al.*, 2009). Output from this step is a P × N matrix, where P and N represent shRNA counts and samples, respectively. We also filtered out shRNAs where the median raw count in the time-point group 1 is greater than the maximum raw count in the time-point group 2 if the shRNA is enriched in the time-point group 1, and vice versa. We then used a negative binomial to model the count distribution in the sequencing data using edgeR (Robinson *et al.*, 2010). We computed the q-value of false discovery rate (FDR) for multiple comparisons for these shRNAs, and carried out a meta-analysis by combining adjusted *P*-values for all shRNAs representing the same gene using weighted Z-transformation (Whitlock, 2005). For each gene, we computed a *P*-value P(wZ), and we used this P(wZ) to sort the kinase list, as previously described (Casás-Selves *et al.*, 2012; Singleton *et al.*, 2013). We performed pair-wise comparisons for each of the time points, and we grouped the kinases into three classes using the following rules based on the P(wZ) obtained from each gene (similar
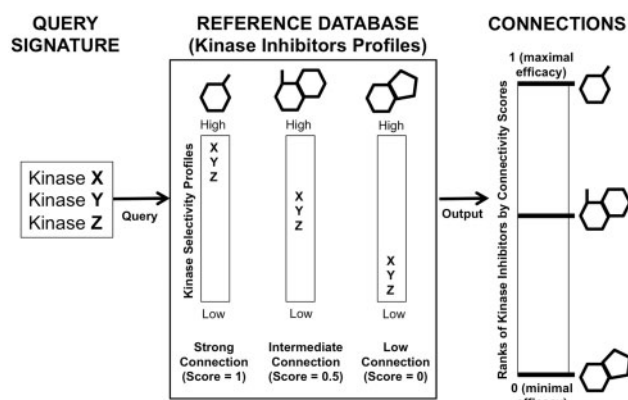
Fig. 2. The K-Map concept. Query signature is derived from experimental study (left). Reference database of K-Map is based on kinase inhibitor selectivity profiles (center). Pattern-matching algorithm provides a score for each reference profile based on its enrichment of the query signature (center). Kinase inhibitors are ranked by the 'connectivity score'; those at the top ('strong connection') and bottom ('low connection') are predicted to have maximal and minimal efficacy against the query signature (right)

to the classification rules in Marcotte *et al.*, 2012). In this study, we considered kinases that, when deleted by shRNAs, induced cell death in the first time point (day 7) as candidate kinases. These kinases were 'essential' for cell survival and they were never recovered in the later time points (day 14 and day 28). See Supplementary Methods for details.

RNA-seq of kinome and bioinformatics analysis. Transcriptome libraries were prepared following Applied BioSystems SOLiD Total RNA-Seq protocol. The libraries were sequenced using the Applied BioSystems SOLiD 5500 platform, using 75 base pair by 35 base paired-end reads. Mapping of sequencing reads and quantification of known RefSeq transcripts were performed using LifeScope v2.1 (ABI). Expression values for each transcript were calculated as reads per kilobase of exon model per million mapped reads (RPKM). RPKM is a method of quantifying gene expression from RNA-seq data by normalizing for total read length and the number of sequencing reads obtained within each sample (Mortazavi *et al.*, 2008). See Supplementary Methods and Supplementary Table S3 for details. To determine the differentially expressed kinases, we computed a FDR on the P-values obtained by t-test. We used FDR $\leq 5\%$ and fold change >1.25 as thresholds for determining differentially expressed kinases between cancer and normal samples.

K-Map analysis. We have recently developed and implemented a K-Map that systematically connects a kinase profile with a reference kinase inhibitor database and predicts the most effective inhibitor for a queried kinase profile (Kim *et al.*, 2013) (Fig. 2). The K-Map is inspired by the connectivity map concept (Lamb *et al.*, 2006), where the main assumption of this concept is that gene expression profiles could be used as 'universal language' to connect between biological states, genes and drugs. The connectivity map has three key components: (i) a reference database that contains a set of predefined gene expression profiles; (ii) a query gene signature; and (iii) a pattern matching algorithm or similarity metric defined between a query gene signature and a reference gene expression profile to quantify the connection (or similarity) between the two biological states. Instead of gene expression signatures, we used the kinase activity profiles as the 'language' for connecting kinases with small molecule kinase inhibitors in K-Map to reveal the interactions of kinases and inhibitors (Kim *et al.*, 2013). Figure 2 illustrates the concept of K-Map. *Reference Database*. We built the reference database based on two recently published comprehensive analyses of kinase inhibitor

selectivity (Anastassiadis *et al.*, 2011; Davis *et al.*, 2011). The first study systematically interrogated 178 commercially available inhibitors against a panel of 300 protein kinases using a radiometric phospho-transfer method to assess the percent kinase inhibition (IC$_{50}$) (Anastassiadis *et al.*, 2011). The second study measured the selectivity and potency of 72 inhibitors against 442 kinases using direct binding affinities between inhibitors and kinases (K$_d$) (Davis *et al.*, 2011). These datasets were converted into rank-ordered lists according to the inhibitors' potencies against the kinases and used as the K-Map reference profiles for matching query kinases. *Query Signature*. Kinases found to be differentially expressed and essential were used as the query kinase profile and connected through the K-Map in this study. *Pattern Matching Algorithm*. We implemented the K-Map pattern matching strategy based on the Kolmogorov–Smirnov (KS) statistics. The KS-test is a non-parametric, rank-based pattern-matching approach implemented in the connectivity map (Lamb *et al.*, 2006). The goal of the algorithm is to correlate kinase inhibitors, based on kinase inhibition profiles in the reference database, with a given query (i.e. a list of kinases). For every inhibitor in the reference database, the KS statistic is computed and a 'connectivity score' is defined (see Supplementary Methods for details). K-Map will return a ranked list of kinase inhibitors that best inhibit the list of queried kinases based on their 'connectivity score'. We used K-Map to connect the differentially expressed and essential kinases with drugs in this study.

Drug Cytotoxicity/Proliferation assays. Bosutinib, gefitinib, sorafenib and CI-1040 were obtained commercially (LC Laboratories, Woburn, MA). Cytotoxic/proliferation effects were determined using WST-1 (Roche) assay. Water-soluble tetrazolium salt (WST-1) assay is a colorimetric assay for the non-radioactive quantification of cellular proliferation, viability and cytotoxicity. WST-1 is added into cultured cells, and the reading of absorbance correlates with the number of proliferating cells. This assay was used to measure the inhibitory effect of drugs used in this study on cell proliferation. The half maximal inhibitory concentration (IC$_{50}$) of individual drugs was determined from the cell proliferation curves. See Supplementary Methods for details.

Immunoblot assays. Apoptosis markers, cleaved and total caspase-3 protein levels, in cells treated with drugs were measured by immunoblotting. See Supplementary Methods for details.

Quantifying combination effects. To quantify the combination effects of drugs used in this study, we used two standard models of additivism (Borisy *et al.*, 2002). The first approach is the highest single agent (HSA) model that measures the larger of the effects produced by each of the combination's single agents at the same concentrations as in the mixture. The combination effect can be classified as: (i) additive (if the difference between combination and HSA is 0); (ii) synergistic (if the difference between combination and HSA is >0); or (iii) antagonistic (if the difference between combination and HSA is <0). The second approach is the Bliss additivism model, which predicts the combined response C for two single compounds, with effects A and B using the following equation: $C = (A + B) - (A \times B)$, where each effect is expressed as fractional inhibition between 0 (no effect, 0% inhibition) and 1 (maximal effect = 100% inhibition). By taking the difference between the observed combined effects and the predicted C, we can classify the combination effect as: (i) additive (if the difference is 0); (ii) synergistic (if the difference is >0); or (iii) antagonistic (if the difference is <0).

3 RESULTS

Essential kinases identified by the BiNGS! analysis. To determine essential kinases other than EGFR that drive the oncogenic signaling pathways in NSCLC, we performed a functional kinome genetic screen using H1650, an NSCLC line with an EGFR-activating mutation (deletion exon 19), yet for unknown reasons is relatively resistant to EGFR TKIs. From this kinome essential

screen, we identified 14 candidate kinases (*CALM3*, *CDK1*, *CDK6*, *DDR1*, *EGFR*, *EPHA4*, *GNE*, *IPMK*, *MARK3*, *PBK*, *PKN1*, *ROCK1*, *RPRD1A* and *TBK1*) as essential for survival of this cell line (see Supplementary Fig. S1 and Supplementary Methods for details). Importantly, *EGFR* was identified from this functional genetic screen as essential in H1650. To determine whether these kinases were mutated in this cell line, we queried the comprehensive curated Catalogue Of Somatic Mutations In Cancer (COSMIC) database (Forbes *et al*., 2011). According to the COSMIC database, besides EGFR, no other known somatic mutations have been reported in the 13 other essential kinases in H1650.

Differentially expressed kinases identified by the RNA-seq analysis. To determine the differentially expressed kinase genes in H1650, we compared the expressed kinome of H1650 defined by targeted RNA-seq with four normal, human ATII cells, the putative cell of origin for lung adenocarcinomas, a histologic subtype of NSCLC (Xu *et al*., 2012). From the RNA-seq data, we focused on the 611 kinases that were the same kinases targeted in the functional genetic screens. We identified 193 kinases overexpressed and 131 underexpressed comparing the H1650 data with the normal ATII RNA-seq data (FDR $\leq$0.05, fold change >1.25). Again, *EGFR* was the top overexpressed gene in the list, supporting H1650's dependence on this oncogene. The heat map of this analysis is illustrated in Supplementary Figure S2 and Supplementary Table S1.

Integrating a functional kinome genetic screen and RNA-seq analysis. To refine the panel of identified key kinases driving the survival and proliferation of H1650 cells, we sought to integrate the candidate kinase genes identified from the essential kinome screen with differential kinase transcriptional profiling, two different yet complementary approaches. Transcriptional profiling analysis identifies kinases that are differentially expressed compared with normal lung ATII cells, yet provides no insight on kinase function. Conversely, the functional genetic screen allows discovery of kinases required for cell survival, yet provides no information specific to transformation. Therefore, integrating these two analyses will provide a list of functionally essential, potentially transformative kinases in H1650. By comparing the candidate gene lists obtained from the two approaches (Fig. 1), seven kinases (*CDK6, EGFR, MARK3, PBK, TBK1, DDR1* and *EPHA4*) were found to be common to both lists.

Connecting the essential, potentially transformative kinases to therapeutics using the K-Map. We next asked what compounds could inhibit these essential and possibly transformative kinases and serve as a potential therapy to inhibit proliferation of this cell line. We queried the K-Map using the seven essential and potentially transformative kinases to connect them to drugs based on two different kinase activity assays (IC_{50} and K_d). The top connection in both assays was staurosporine, a natural product isolated from the bacterium *Streptomyces staurosporeus*. Staurosporine is a potent, general ATP-binding site inhibitor across kinases, lacking any specificity (Anastassiadis *et al*., 2011; Davis *et al*., 2011). Interestingly, bosutinib, a Src and Abl dual inhibitor, was also positively connected and ranked #3 and #29 in the K_d and IC_{50} assays, respectively (Fig. 1 and Supplementary Table S2). As expected, the two FDA-approved EGFR-specific TKIs for treating NSCLC patients, gefitinib and
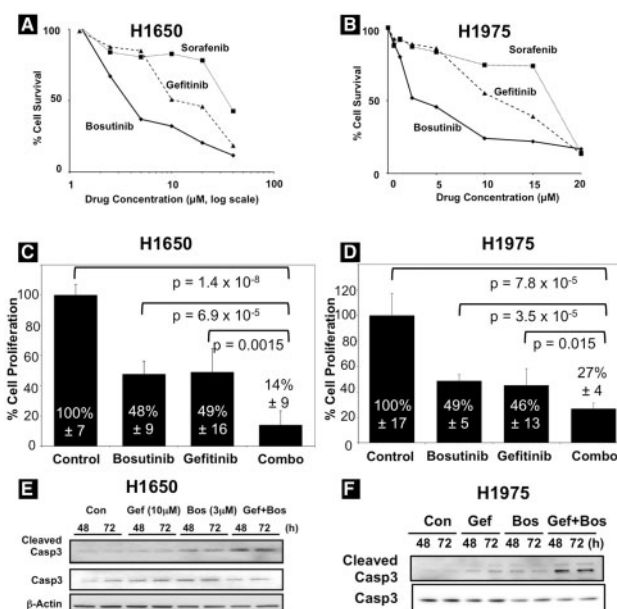

Fig. 3. Experimental validation of bosutinib in H1650 and H1975. Proliferation curves for bosutinib, gefitinib and sorafenib in (**A**) H1650 and (**B**) H1975. Combination of bosutinib and genfitinib in (**C**) H1650 and (**D**) H1975. Bosutinib and the combination induce apoptosis in (**E**) H1650 and (**F**) H1975 cell lines

erlotinib, were not connected by the K-Map as effective therapy for the queried essential kinases (Fig. 1).

Experimental validation of the compounds identified by K-map analysis. To test whether compounds identified by the K-Map could provide better growth inhibition of H1650 cells compared with the current standard of care with EGFR-specific TKIs, we selected bosutinib (Fig. 1) as the candidate compound prediction of the K-map and validated its effect on cell survival. In comparison, we selected gefitinib, a drug approved for use in EGFR-mutant NSCLC, which was not identified using our strategy and therefore, we predicted, would be less efficacious in H1650 cells (Fig. 1). To further test our algorithm and K-Map, we identified sorafenib (connectivity scores of 0.234 and 0.226 in both kinase assays, Fig. 1) as a compound which was predicted to be ineffective against the H1650 cell line.

To evaluate the sensitivity of bosutinib, gefitinib and sorafenib, H1650 NSCLC cells were exposed to increasing concentrations of these individual drugs and assessed for proliferation. As depicted in Figure 3A, bosutinib had the lowest IC_{50} among the three tested compounds. Conversely, gefitinib and sorafenib had higher IC_{50} values (IC_{50} > 10 μM), indicating both drugs were less effective in inhibiting the growth of H1650 cells. As an additional negative control, we validated CI-1040, a compound that has a connectivity score of 0 (minimal effect and ranked #238 of 250 drugs) and was the least effective of the four drugs tested (IC_{50} ~34 μM) in H1650 (Supplementary Fig. S3).

To estimate the statistical significance of the connection, we used the connection testing proposed by Zhang and Gant (2008, 2009), where the two-tailed *P*-value associated with the observed connection score is the number of times the connection score obtained by a random gene signature with the same number of

genes when queried to the database. Given a reference database and a query signature S with m kinases (here, $m = 7$), the K-Map connectivity score is KS_S, and the two-tailed P-value is estimated as: $p = Prob \{|KS_r| \geq |KS_S|\}$, where KS_r is the connection score obtained from the random query signature with m kinases. For this particular study, we performed 500 permutations, and computed the P-value of the bosutinib connection as $p = 0.052$ (see Supplementary Methods for details). As demonstrated by the experimental validation, this indicates that the connection is not a false positive identified by the method.

Generalizing the K-map prediction to acquired EGFR-TKI resistance. Acquired resistance to EGFR TKIs after therapy is universal in lung cancer patients limiting their usefulness as a treatment of NSCLC. Drugs that can overcome this resistance would dramatically impact lung cancer care. To explore whether bosutinib can be extrapolated as an effective therapy for an acquired mutation conferring resistance to EGFR TKIs, we tested H1975, a NSCLC line that harbors EGFR T790M, a gatekeeper mutation that is commonly acquired during treatment with EGFR TKIs and renders standard EGFR TKIs ineffective. As illustrated in Figure 3B, the IC_{50} value of bosutinib on H1975 was $\sim 3.7\,\mu M$. As a negative control, we also determined the IC_{50} value of gefitinib, sorafenib and CI-1040 for this cell line as ~ 12, ~ 17 and $\sim 31\,\mu M$, respectively (Fig. 3B and Supplementary Fig. S3). Thus, bosutinib is effective in inhibiting the proliferation of a NSCLC line with an acquired EGFR resistance mutation.

Rational combination of bosutinib and gefitinib shows synergistic effects in EGFR-mutant NSCLC cells. We further hypothesized that the addition of bosutinib to gefitinib would be synergistic in EGFR-mutant NSCLC cells resistant to single agent EGFR-TKI treatment. Indeed, the combination of bosutinib and gefitinib demonstrated additive or synergistic effects in both EGFR-mutant NSCLC lines (Fig. 3C and D). Combination therapy was significantly improved in H1650 and H1975 ($P < 0.05$, Welch two sample t-test) in inhibiting cell proliferation when compared with individual drug alone (Fig. 3C and D). Based on the additivism models, in H1650 combination therapy was synergistic (HSA = 0.34 and Bliss = 0.10), whereas in H1975 was additive (HSA = 0.19 and Bliss = –0.04) (see Supplementary Methods for details). Immunoblotting of H1650 and H1975 revealed an increase of cleaved caspase-3 (Casp3), an apoptotic marker, in the bosutinib and combination treated cells as compared with gefitinib alone and vehicle as early as 48 h after treatment (Fig. 3E and F).

4 DISCUSSION

Functional genetic screens have the potential to identify genes essential for cancer cell survival and proliferation, providing a 'functional' map of human cancer. Complementing the functional genetic screen with comprehensive genomics studies such as transcriptional profiling could reveal the 'vulnerable targets' for targeted treatment of cancer cells. Here, we performed a systematic and unbiased approach to determine essential kinases driving survival mechanisms of EGFR-mutant NSCLC lines resistant to EGFR TKIs. Using a series of novel bioinformatics analyses, specifically connecting the essential kinases with small molecules based on inhibition activities, we have identified that

bosutinib effectively inhibits the essential kinases in H1650 resulting in cell death. We validated bosutinib in H1650 (intrinsic EGFR-TKI resistance) and H1975 (acquired gatekeeper T790M mutation) and demonstrated that this compound inhibited cell proliferation and induced apoptosis in these cancer cell lines that are resistant to EGFR-specific TKIs better than standard care.

Our strategy combines high-throughput genetic screens with a new computational technique (BiNGS!) to rapidly identify kinases that are essential for cancer cell survival. As described earlier, the H1650 cell line is driven by an activating *EGFR* mutation; therefore, it was expected that *EGFR* would be one of the seven essential kinases found in the overlapping gene lists. Other differentially expressed and essential kinases identified in this study have been previously reported to play an important role in cancer cells. For example, *MARK3* has been shown to play a role in regulating cell cycle progression in cancer cells (Sha *et al.*, 2007). *EPHA4* was recently identified as an inhibitor of cell migration and invasion in lung cancer (Saintigny *et al.*, 2012), supporting our finding that the expression of this gene is lower than that found in normal ATII cells, yet functionally important in driving lung oncogenesis. *DDR1* was identified as an essential kinase across three different tumor types from a recent large-scale functional genetic screen (Marcotte *et al.*, 2012). Despite these data, prior to our analyses there was no method for identifying these seven kinases as the functional vulnerabilities of H1650, which could be targeted for therapeutic intervention.

By connecting the H1650 essential kinase profile through the K-Map to existing kinase inhibitors, we could repurpose bosutinib to treat EGFR-mutant NSCLC cell line resistant to EGFR TKIs. Bosutinib is a Src/Abl dual TKI, which has recently been approved by the FDA to treat CML patients. Our data suggest that by using the connectivity map concept, we could 'connect' essential kinases to therapeutics, facilitating the translation from *in silico* discovery to clinical trials. Recently, bosutinib has been tested in a Phase I clinical trial of advanced solid tumors (Daud *et al.*, 2012). Among the 16 NSCLC patients treated with the optimal dose of bosutinib, seven of the patients observed tumor shrinkage and nine patients had stable disease. As concluded by the authors (Daud *et al.*, 2012), bosutinib might provide benefit in combination with other drugs to result in better treatment responses. In this study, we provide a biological rationale for the effectiveness of bosutinib in NSCLC. Moreover, we confirmed that the combination of bosutinib with gefitinib has additive or synergistic effects in two gefitinib-resistant NSCLC cell lines. Future experiments for the combination of bosutinib and gefitinib are warranted to investigate this bioinformatics-driven discovery in more cell lines and in animal models.

In summary, we have demonstrated a proof-of-concept, bioinformatics-driven discovery roadmap for drug repurposing and development in cancer research, which could be generalized to other diseases in the era of personalized and precision medicine.

REFERENCES

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

Borisy,A.A. *et al.* (2002) Systematic discovery of multicomponent therapeutics. *Proc. Natl Acad. Sci. USA*, **100**, 7977–7982.

Casás-Selves,M. *et al.* (2012) Tankyrase and the canonical Wnt pathway protect lung cancer cells from EGFR inhibition. *Cancer Res.*, **72**, 4154–4164.

Chung,F.H. *et al.* (2014) Functional module connectivity map (FMCM): a framework for searching repurposed drug compounds for systems treatment of cancer and an application to colorectal adenocarcinoma. *PLoS One*, **9**, e86299.

Daud,A.I. *et al.* (2012) Phase I study of bosutinib, a Src/Abl tyrosine kinase inhibitor, administered to patients with advanced solid tumors. *Clin. Cancer Res.*, **18**, 1092–1100.

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

Engelman,J.A. and Jänne,P.A. (2008) Mechanisms of acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in non-small cell lung cancer. *Clin. Cancer Res.*, **14**, 2895–2899.

Forbes,S.A. *et al.* (2011) COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, **39**, D945–D950.

Hieronymus,H. *et al.* (2006) Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell*, **10**, 321–330.

Hirsch,F.R. *et al.* (2003) Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J. Clin. Oncol.*, **21**, 3798–3807.

Kim,J. and Tan,A.C. (2012) BiNGS!SL-seq: a bioinformatics pipeline for the analysis and interpretation of deep sequencing genome-wide synthetic lethal screen. *Methods Mol. Biol.*, **802**, 389–398.

Kim,J. *et al.* (2013) K-map: connecting kinases with therapeutics for drug repurposing and development. *Hum. Genomics*, **7**, 20.

Lamb,J. *et al.* (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes and disease. *Science*, **131**, 1929–1935.

Langmead,B. *et al.* (2009) Ultrafast and memory efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.

Li,J. *et al.* (2009) Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts. *PLoS Comput. Biol.*, **5**, e1000450.

Marcotte,R. *et al.* (2012) Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discov.*, **2**, 172–189.

Mortazavi,A. *et al.* (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.

Ohashi,K. *et al.* (2013) Epidermal growth factor receptor tyrosine kinase inhibitor-resistant disease. *J. Clin. Oncol.*, **31**, 1070–1080.

Pao,W. and Chmielecki,J. (2010) Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. *Nat. Rev. Cancer*, **10**, 760–774.

Robinson,M.D. *et al.* (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Saintigny,P. *et al.* (2012) Global evaluation of Eph receptors and ephrins in lung adenocarcinomas identifies EphA4 as an inhibitor of cell migration and invasion. *Mol. Cancer Ther.*, **11**, 2021–2032.

Sha,S.K. *et al.* (2007) Cell cycle phenotype-based optimization of G2-abrogating peptides yields CBP501 with a unique mechanism of action at the G2 checkpoint. *Mol. Cancer Ther.*, **6**, 147–153.

Siegel,R. *et al.* (2013) Cancer statistics, 2013. *CA Cancer J. Clin.*, **63**, 11–30.

Singleton,K.R. *et al.* (2013) A receptor tyrosine kinase network composed of fibroblast growth factor receptors, epidermal growth factor receptor, v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, and hepatocyte growth factor receptor drives growth and survival of head and neck squamous carcinoma cell lines. *Mol. Pharmacol.*, **83**, 882–893.

Spreafico,A. *et al.* (2013) Rational combination of a MEK inhibitor, selumetinib, and the Wnt/calcium pathway modulator, cyclosporin A, in preclinical models of colorectal cancer. *Clin. Cancer Res.*, **19**, 4149–4162.

Sullivan,K.D. *et al.* (2012) ATM and MET kinases are synthetic lethal with non-genotoxic activation of p53. *Nat. Chem. Biol.*, **8**, 646–654.

Wang,J. *et al.* (2007) Differentiated human alveolar epithelial cells and reversibility of their phenotype *in vitro*. *Am. J. Respir. Cell Mol. Biol.*, **36**, 661–668.

Ware,K. *et al.* (2013) A mechanism of resistance to gefitinib mediated by cellular reprogramming and the acquisition of an FGF2-FGFR1 autocrine growth loop. *Oncogenesis*, **2**, e39.

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.

Whitlock,M.C. (2005) Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J. Evol. Biol.*, **18**, 1368–1373.

Xu,X. *et al.* (2012) Evidence of type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma. *Proc. Natl Acad. Sci. USA*, **109**, 4910–4915.

Zhang,S.-D. and Gant,T.W. (2008) A simple and robust method for connecting small-molecule drugs using gene-expression signatures. *BMC Bioinformatics*, **9**, 258.

Zhang,S.-D. and Gant,T.W. (2009) sscMap: an extensible Java application for connecting small-molecule drugs using gene-expression signatures. *BMC Bioinformatics*, **10**, 236.