RESEARCH ARTICLE

DRUG SCREENING

Using Functional Signature Ontology (FUSION) to Identify Mechanisms of Action for Natural Products

Malia B. Potts,[1]* Hyun Seok Kim,[1]* Kurt W. Fisher,[2]* Youcai Hu,[3]* Yazmin P. Carrasco,[3] Gamze Betul Bulut,[1] Yi-Hung Ou,[1] Mireya L. Herrera-Herrera,[1] Federico Cubillos,[1] Saurabh Mendiratta,[1] Guanghua Xiao,[4] Matan Hofree,[5] Trey Ideker,[5] Yang Xie,[4,6] Lily Jun-shen Huang,[1] Robert E. Lewis,[2]† John B. MacMillan,[3]† Michael A. White[1]†

A challenge for biomedical research is the development of pharmaceuticals that appropriately target disease mechanisms. Natural products can be a rich source of bioactive chemicals for medicinal applications but can act through unknown mechanisms and can be difficult to produce or obtain. To address these challenges, we developed a new marine-derived, renewable natural products resource and a method for linking bioactive derivatives of this library to the proteins and biological processes that they target in cells. We used cell-based screening and computational analysis to match gene expression signatures produced by natural products to those produced by small interfering RNA (siRNA) and synthetic microRNA (miRNA) libraries. With this strategy, we matched proteins and miRNAs with diverse biological processes and also identified putative protein targets and mechanisms of action for several previously undescribed marine-derived natural products. We confirmed mechanistic relationships for selected siRNAs, miRNAs, and compounds with functional roles in autophagy, chemotaxis mediated by discoidin domain receptor 2, or activation of the kinase AKT. Thus, this approach may be an effective method for screening new drugs while simultaneously identifying their targets.

INTRODUCTION

The diverse chemistry found in nature provides a valuable resource for medicinal drugs and has been the source of many antibiotics and cancer chemotherapeutics over the past 6 decades (*1*). The chemical structures of natural products have evolved for interaction with biological systems, resulting in submicromolar potency for many compounds. A proven drug screening paradigm for natural product collections is the use of cell-based phenotypic screens for the selection of molecules with reasonable physiochemical properties (*2*, *3*). However, critical barriers to rapid development of drug-like molecules include the challenges of structural determination, purification or synthesis, and elucidation of molecular targets and mechanisms of action. To address these issues, we combined a renewable natural product library with a method of screening and functional annotation to enable rapid identification, characterization, and production of drugs with known mechanisms of action. We generated a library of metabolites from laboratory-cultured, marine-derived bacterial species and marine invertebrates. The bacterial collection was isolated from sea-floor sediments, and the organisms were cultivated using specialized strategies such as quorum-sensing molecules and habitat-specific media, yielding a large phylogenetic diversity of microorganisms.

The resulting natural product fractions are typically mixtures of two to six compounds. To classify the biological activities associated with the natural products in this collection, we used a high-throughput gene expression platform (*4–9*) that enables quantitative discrimination of concordant cellular responses to genetic [microRNA (miRNA)– and small interfering RNA (siRNA)–induced] and chemical perturbations in human cultured cells. The genetic perturbations included a combination of oligonucleotides that mimic miRNAs and a collection of siRNA pools targeting the kinome, including most known kinases, phosphatases, and kinase signaling accessory proteins. We chose these collections as a practical means of interrogating a broad landscape of genetic relationships with a minimal number of tests. For example, endogenous miRNAs presumably represent a limited sequence space that evolved to regulate distinct cellular processes through the combinatorial inhibition of translation of groups of transcribed genes (*10*). In addition, the entire protein-coding genome has undergone selective pressure to avoid detrimental targeting by miRNAs (*11*). Likewise, the kinome is heavily enriched for proteins that participate in the molecular signaling networks that specify dynamic cell regulatory events (*12*), inhibition of which can also modulate many cellular processes.

We assembled signatures of the expression of a specific set of genes that resulted from screening of siRNA, miRNA mimic, or natural product libraries into a similarity matrix to identify biologically related genetic and chemical perturbations. Using this approach, we produced functional signature ontology (FUSION) maps that linked bioactive molecules to the proteins and biological processes that they engage in cells. Experimental evaluation of hypotheses generated by FUSION mapping confirmed biological roles for previously uncharacterized miRNAs, kinases, and natural product compounds. To facilitate the use of this resource beyond the specific mechanistic relationships reported here, we developed an open access search engine that allows users to query for chemical and genetic "functionalogs" of a gene of interest and provides statistical evaluations of the strength of each match (http://whitelab.swmed.edu/fmap/fusion1.php).

[1]Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [2]Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE 68198, USA. [3]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [4]Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [5]Department of Medicine and Bioengineering and Program in Bioinformatics, University of California at San Diego, La Jolla, CA 92093, USA. [6]Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA.
*These authors contributed equally to this work.
†Corresponding author. E-mail: michael.white@utsouthwestern.edu (M.A.W.); john.macmillan@utsouthwestern.edu (J.B.M.); rlewis@unmc.edu (R.E.L.)

RESULTS

Creating FUSION maps from gene expression signatures

To build a molecular functional ontology for annotation of the activity of molecules, such as natural products or small molecules, or the effect of genetically targeted perturbations, such as siRNAs or miRNA mimics, in cell-based phenotypic screens, we measured the expression of a specific set of genes representing a signature that could serve as a proxy for the physiological state of the cell (4–9). To select a small set of high-performance endogenous mRNAs that could serve as reporters of diverse cell biological states, we queried a collection of publicly available whole-genome expression arrays (~3700 arrays) for gene products displaying high variation and low covariation across multiple genetic or pharmacological perturbations in human cells (13). Six genes (ACSL5, BNIP3L, ALDOC, LOXL2, BNIP3, and NDRG1) fit these criteria and also displayed concordant responses to genetic or chemical perturbation of the mitogen-activated protein kinase (MAPK) pathway in HCT116 cells. In addition, two genes (PPIB and HPRT) that showed limited variation in expression were used as internal normalization controls [http://www.bcgsc.ca/project/bomge/coexpression/ (13) (fig. S1A)]. We made high-throughput measurements of this "endogenous reporter gene set" using multianalyte capture beads and branched-DNA signal amplification to enable direct simultaneous quantitation of multiple mRNAs in each well of a 384-well plate. We used this approach to screen a collection of 780 siRNAs targeting human kinases and related proteins and 344 nonredundant miRNA mimics. Normalized reporter gene expression was measured 72 hours after transfection of miRNAs or siRNAs into HCT116 colorectal adenocarcinoma cells (fig. S1B and table S1). Assembly of the endogenous reporter gene signatures resulting from each perturbation into a similarity matrix, using Euclidean distance distributions, linked biologically similar genetic perturbations. For intuitive data visualization, we used two phylogenetic tree construction algorithms. These included complete-linkage clustering, which is commonly used for unsupervised hierarchical clustering of transcript array data, and neighbor joining, which finds the near-optimum tree topology with minimal branch length (14, 15). In addition, we estimated all the pairwise distances between genetic perturbations and evaluated the statistical significance per pair by permutation (see Materials and Methods). This analysis produced FUSION maps (Fig. 1A), linking bioactive compounds to molecular entities, biological process, or both (table S1; http://whitelab.swmed.edu/fmap/fusion1.php).

Validating the FUSION approach on the basis of known relationships among proteins targeted by miRNA mimics or siRNAs

To evaluate the performance of the reporter gene set, we tested the prediction that, when expressed as miRNA mimics, miRNA families that share the same seed region should produce more similar gene reporter signatures than those produced by random pairs of miRNAs. The seed sequence of a miRNA, which anneals to complementary sequences on target mRNAs, is a major determinant of miRNA-mediated suppression of gene expression (11). The miRNA mimic library that we used corresponds to endogenous human miRNAs from miRBase version 8.0 (16). This miRNA mimic library has an internal redundancy of 20% because multiple genetic loci produce the same mature miRNA. Thus, the library contains 426 miRNA mimics composed of 344 nonredundant sequences. Reporter gene expression was largely uncoupled from endogenous miRNA abundance (fig. S1, C and D, and tables S1 and S2). Moreover, as a set, the expression of the selected reporter genes responded to transfection of most of the miRNA mimics; 73.5% of the miRNA mimics elicited a twofold or greater change in expression of one or more reporter genes (fig. S2 and table S1, "genetic_all" tab). From the similarity matrix computed for the miRNA library, we found concordance of reporter gene signatures among miRNA seed families (Fig. 1A and fig. S2). Among miRNA mimics with matching seed sequences (nucleotides 2 to 7), reporter gene signatures correlated to an almost equivalent degree as observed for miRNA mimics that were completely identical (Fig. 1, B to D). Furthermore, the concordance of reporter gene signatures among miRNA mimics with identical seed sequences was not influenced by the remaining nonseed sequence (Fig. 1, E and F). Thus, the correlation between the seed sequence and the gene signatures suggests that the equivalence of miRNA targets drives the miRNA mimic clustering, which, by extrapolation, suggests that the chosen reporter gene set appropriately reports the effects of these reagents across diverse biological activities.

To evaluate the performance of the reporter gene set in cells expressing the kinome siRNA library, we tested the prediction that pairs of kinases in the kinome that physically interact (15) should have more similar reporter gene signatures than random pairs of kinases targeted by the siRNA library. From the FUSION map of the siRNAs clustered by reporter gene signature similarity (fig. S3 and table S1), we plotted a cumulative distribution of reporter gene signature similarity scores for the kinome siRNA data set (Fig. 1G). This analysis indicated a significant enrichment of similar reporter gene signatures among siRNAs targeting kinases that physically interact, further indicating that FUSION mapping can successfully parse siRNA-mediated perturbations across diverse biological responses.

We also computed a combined miRNA mimic and siRNA FUSION matrix (fig. S4A and table S1, "genetic_all" tab) to evaluate performance of the reporter gene set across different classes of genetic perturbations. To minimize the contribution of off-target effects of the siRNAs due to sequence overlap between siRNAs and the seed region of miRNAs, we screened pools of four individual siRNAs per target gene (17–20). Of the 311 possible pairwise comparisons between a specific miRNA mimic and a corresponding siRNA pool in which the miRNA's seed sequence was also present in an siRNA oligo within the pool, we detected 94 siRNA-miRNA gene expression correlations (with Pearson correlation >0.5) that were potentially caused by the siRNA recognizing the same sequence as that of the miRNA mimic seed sequence (table S3). That is, a subset of the siRNAs with sequences that included miRNA seed sequences showed a slightly larger average correlation than those with different seed sequence regions (fig. S4B), supporting the notion that off-target effects of siRNA oligos made a minor but identifiable contribution to the FUSION matrix.

Predicting cellular functions of understudied kinases and linking miRNAs to signaling pathways with FUSION mapping

To test whether FUSION mapping could be used to predict functional relationships among genetic perturbations, we focused on two sets of miRNA mimics and siRNAs that clustered together in the miRNA-siRNA similarity matrix (fig. S4A) and that, on the basis of some of the targeted proteins, would be predicted to have opposing effects on autophagy (Fig. 2, A and B). One cluster included Unc-51–like autophagy-activating kinase 1 (ULK1), which is required for autophagosome formation and maturation (21), whereas the other cluster included mechanistic target of rapamycin (mTOR), which functions in the mTOR complex 1 to inactivate ULK1 and suppress autophagy (22). Reporter gene signatures produced by siRNA knockdown of ULK1 or mTOR produced distinct functional signatures that correlated with those produced by miRNA mimics and siRNAs within the same cluster (present in the same clade) at a false discovery rate (FDR) of <20% (Fig. 2, A and B). These clusters contained several known modulators of autophagy and their first-order protein interaction partners (23) (Fig. 2C).

Fig. 1. Functional signature-based ontology. (**A**) Left: The neighbor joining tree of the Euclidean distance among reporter gene signatures for all miRNA mimics. Coclustering of seed-based miRNA families is indicated and color-coded. The let-7 and miR-30 branches are circled. (See table S1 for details.) Right: Unsupervised hierarchical clustering of expression data from miRNAs in these two branches. Identical seed sequences are indicated by colorcoding. # indicates that the seed sequence shown and present in the miRNA library differs from the annotated seed sequence in miRBase. (**B**) Density distributions (standard probability density plot) for redundant miRNA mimics (Same sequence, $n = 105$ pairs) and for nonredundant miRNA mimics (Different sequence, $n = 90,420$ pairs). (**C**) Density distributions for the nonredundant miRNA mimics that share an identical seed sequence (Same seed, Different sequence, $n = 330$ pairs) and for miRNA mimics with different seeds (Different seed, $n = 90,090$ pairs). (**D**) Density distributions of reporter gene signatures for miRNA mimic pairs with the same sequence or different sequence and that did not reduce the viability of HCT116 cells below 20% of the median ($n = 405$ miRNA mimics). (**E**) Overall sequence similarity was calculated between every miRNA mimic pair: nonredundant miRNA mimics sharing identical seed sequences, $n = 330$ pairs (red); miRNA mimics with nonidentical seeds, $n = 90,090$ pairs (black). (**F**) Euclidean distance between reporter gene signatures of each miRNA pair in the indicated categories. To determine whether miRNA mimic pairs with identical seeds but dissimilar sequences outside the seed region produce expression signatures that are more highly correlated than those produced by miRNA pairs with similar overall sequence similarity but nonidentical seeds, we performed statistical analysis on the subset of miRNA mimic pairs with sequence similarity <0.5. (C to F) $P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test. (**G**) Cumulative distribution of Euclidean distances between all possible pairs of siRNAs targeting kinases (solid line) compared to siRNA pairs targeting kinases known to physically interact (dotted line) (47). $P = 0.0057$, Kolmogorov-Smirnov statistic.
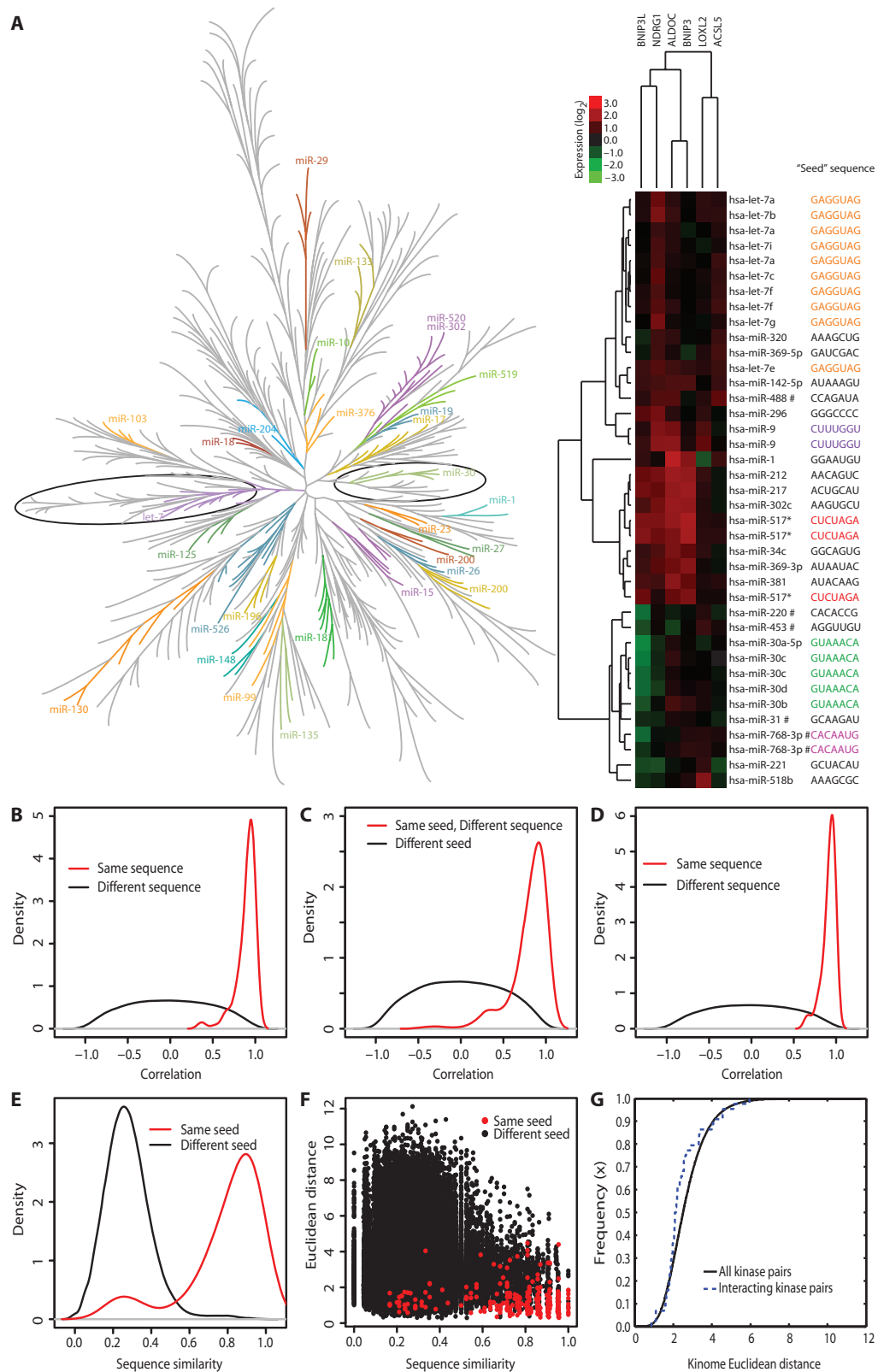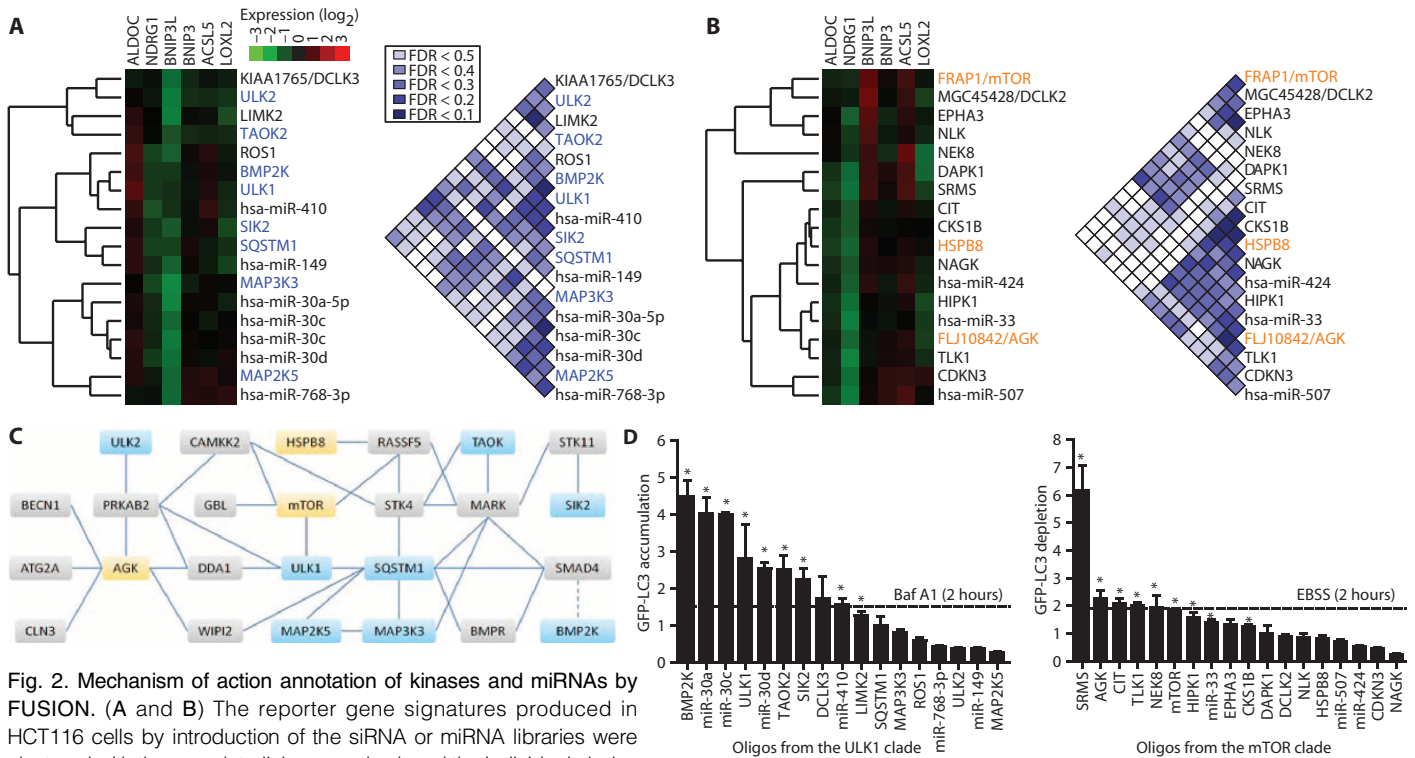
Fig. 2. Mechanism of action annotation of kinases and miRNAs by FUSION. (**A** and **B**) The reporter gene signatures produced in HCT116 cells by introduction of the siRNA or miRNA libraries were clustered with the complete-linkage method, and the individual clades containing siRNAs targeting ULK1 (A) or mTOR (B) are presented as gene expression heat maps. FDRs are depicted for pairwise comparisons within each group. (**C**) A map of the autophagy regulatory network (*23*) including physical interactions (solid lines) and regulatory relationships (dotted line) among kinases in the ULK1 (blue) and mTOR groups (yellow) and between these kinases and other proteins (gray). (**D**) Quantification of GFP-LC3 by flow cytometry after siRNA pool or miRNA mimic introduction into U2OS cells. Median GFP fluorescence values were normalized to negative control oligonucleotides (LONRF1 for siRNAs or MC1 for miRNA mimics). Means ± SEM; $n \geq 3$. *$P < 0.05$, one-tailed t test. Baf A1, bafilomycin A; EBSS, Earle's balanced salt solution nutrient deprivation media.

To determine if clade membership revealed undiscovered regulators of autophagy, we tested each siRNA or miRNA mimic in each clade for an effect on autophagy. We used flow cytometry to monitor the fluorescence intensity of LC3 (microtubule-associated protein 1 light chain 3α) fused to green fluorescent protein (GFP-LC3) expressed in U2OS cells as a marker of autophagic flux (fig. S4C). Lipidated LC3 coats the inner and outer membranes of nascent autophagosomes and is proteolytically degraded in mature autophagosomes (*24*); thus, accumulation of LC3 indicates a reduction in autophagic flux, and depletion of LC3 indicates an increase in autophagic flux. Pharmacological inhibition of autophagosome acidification with bafilomycin A1 or siRNA knockdown of ULK1 increased GFP-LC3 fluorescence intensity, whereas activation of autophagic flux by nutrient deprivation or siRNA knockdown of mTOR decreased GFP-LC3 fluorescence intensity (Fig. 2D and fig. S4, D to F). We found that 9 of 17 siRNAs and miRNA mimics from the ULK1 cluster significantly blocked autophagic flux, triggering an increase in GFP-LC3 intensity, and 9 of 18 siRNAs and miRNA mimics from the mTOR cluster enhanced autophagy, triggering a decrease in GFP-LC3 fluorescence intensity similar to or exceeding that of incubation with nutrient deprivation medium (Fig. 2D and fig. S5A).

In particular, knockdown of BMP2K (BMP2-inducible kinase), ULK1, TAOK2 (Tao kinase 2), or SIK2 (salt-inducible kinase 2) induced GFP-LC3 accumulation to an extent equal to or greater than that observed with treatment with bafilomycin A1, as did introduction of the miRNA mimics from the miR-30 family or miR-410. Conversely, knockdown of SRMS

(Src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristoylation sites), AGK (acylglycerol kinase), CIT (citron), TLK1 (tousled-like kinase 1), NEK8 (NIMA-related kinase 8), or mTOR depleted GFP-LC3 to an extent equal to or greater than that observed with nutrient deprivation.

We selected BMP2K for further investigation because siRNA knockdown of BMP2K produced the strongest inhibition of autophagy on the basis of flow cytometric analysis of GFP-LC3 abundance (Fig. 2D). Multiple independent siRNAs targeting BMP2K increased GFP-LC3 fluorescence intensity, consistent with inhibition of autophagy (fig. S5B). Moreover, siRNA knockdown of BMP2K reduced the average number of LC3-positive puncta per cell (Fig. 3A), whereas overexpression of BMP2K increased the number of LC3-positive puncta per cell (Fig. 3A).

Erythrocyte maturation is a cellular differentiation process that depends on mitochondrial autophagy (*25*). Tissue-specific gene expression profiling indicates a 3- to 10-fold enrichment of *BMP2K* expression in human CD71$^+$ early erythroid cells compared with other human tissues and cells (*26*). We found that knockdown of BMP2K or ULK1 in mouse erythroid progenitors through retrovirally delivered short hairpin RNA (shRNA) impaired ex vivo differentiation and blocked mitochondrial autophagy (Fig. 3, B and C, and fig. S5, C and D). Measurement of the abundance of mitochondria with a mitochondrially targeted fluorescent protein (*27*) showed that knockdown of BMP2K in U2OS cells increased mitochondrial mass (fluorescence intensity) (Fig. 3D). Thus, FUSION analysis identified BMP2K
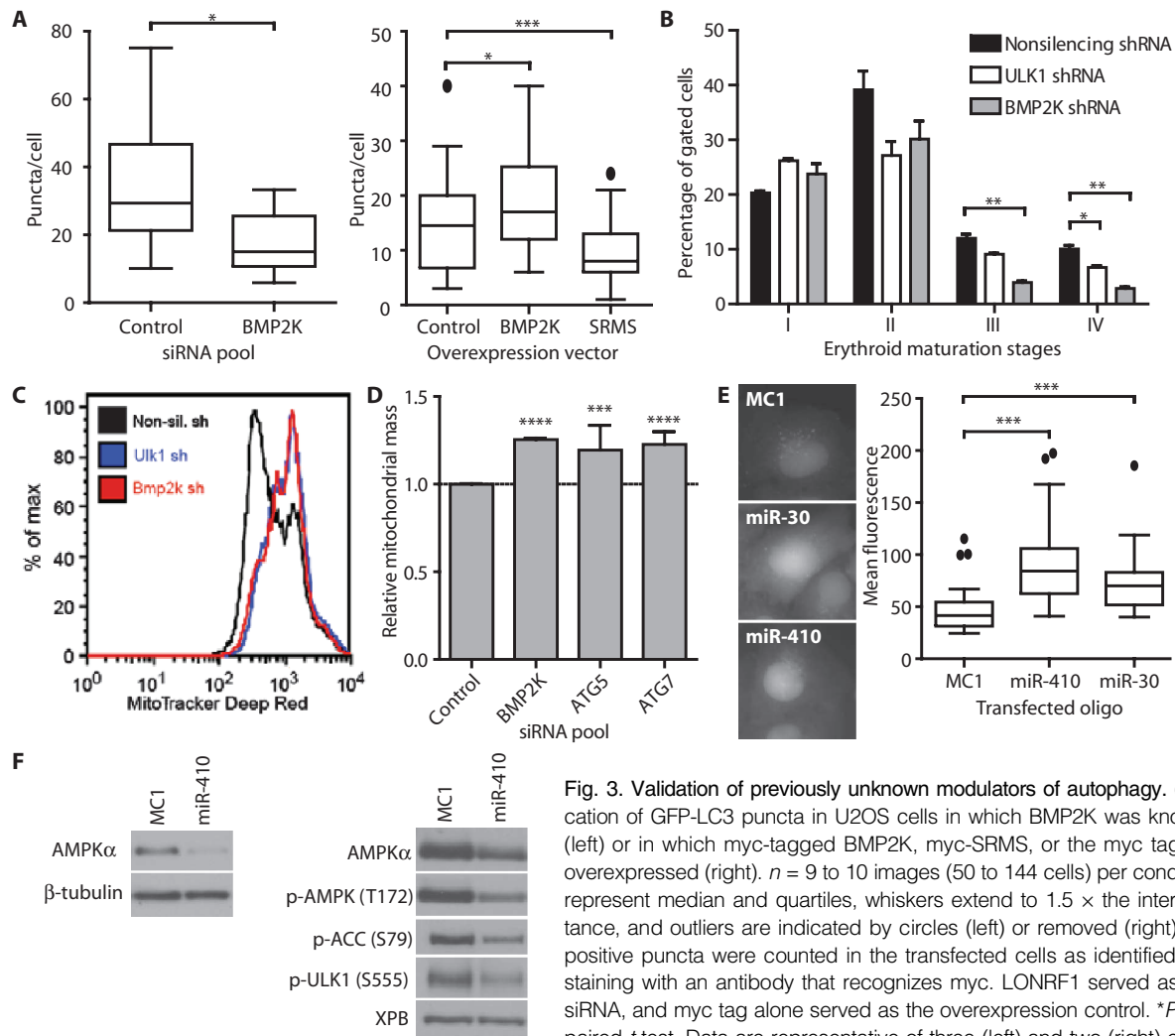
Fig. 3. Validation of previously unknown modulators of autophagy. (A) Quantification of GFP-LC3 puncta in U2OS cells in which BMP2K was knocked down (left) or in which myc-tagged BMP2K, myc-SRMS, or the myc tag alone was overexpressed (right). n = 9 to 10 images (50 to 144 cells) per condition. Boxes represent median and quartiles, whiskers extend to 1.5 × the interquartile distance, and outliers are indicated by circles (left) or removed (right). GFP-LC3–positive puncta were counted in the transfected cells as identified by positive staining with an antibody that recognizes myc. LONRF1 served as the control siRNA, and myc tag alone served as the overexpression control. *P < 0.05, unpaired t test. Data are representative of three (left) and two (right) experiments. (B) Mouse erythroid progenitors with shRNAs targeting ULK1, BMP2K, or nonsilencing control. Stages of differentiation were identified by flow cytometry for CD71 and Ter119: proerythroblasts (I), basophilic erythroblasts (II), late basophilic and chromatophilic erythroblasts (III), and orthochromatic erythroblasts (IV). Means ± SEM; n = 3 experiments. See fig. S5D for representative FACS (fluorescence-activated cell sorting) plots. (C) Histograms of the fluorescence intensity of MitoTracker Deep Red FM used to reveal mitochondrial content of erythroid progenitor cells infected with the indicated shRNAs. Data are representative of three experiments. (D) Quantification of flow cytometry of stable mito-mCherry in U2OS cells transfected with the indicated siRNAs. Means ± SEM; n = 3 experiments. ***P < 0.001, ****P < 0.0001, unpaired t test. (E) Representative images of GFP-LC3 in U2OS cells transfected with the indicated miRNA mimics (left). Image-based quantification GFP-LC3 in U2OS cells transfected with miRNAs (right). n = 36 to 46 cells per condition. Boxes represent median and quartiles, whiskers extend to 1.5 × the interquartile distance, and outliers are indicated by circles. ***P < 0.0001, unpaired t test. (F) Western blot of HCT116 cells (left) or U2OS GFP-LC3 cells (right) transfected with miR-410 or negative control (MC1). Data are representative of five experiments.

as required for basal autophagy in U2OS cells and differentiation-induced mitochondrial autophagy during erythrocyte development.

We identified two miRNA mimics that inhibited autophagic flux—the miR-30 family and miR-410. The miRNA mimic corresponding to miR-30a, which targets the mRNA encoding Beclin-1 and thereby inhibits autophagy, including that induced by chemotherapy (28, 29), enhanced GFP-LC3 fluorescence intensity in U2OS cells (Figs. 2D and 3E). An miRNA mimic corresponding to miR-410, which is predicted to target adenosine monophosphate–activated protein kinase α 1 and 2 (AMPKα) (http://www.targetscan.org/), also promoted GFP-LC3 accumulation to a similar extent as that observed with bafilomycin A1 treatment (Figs. 2D and 3E and fig. S5A). Introduction of the miR-410 mimic into HCT116 or U2OS cells reduced the abundance of AMPKα and inhibited phosphorylation of AMPKα downstream targets, including ULK1 (Fig. 3F). AMPKα directly activates ULK1 to induce autophagosome formation in response to nutrient deprivation (22). Thus, our results suggest the mechanism by which miR-410 inhibits autophagy.

To a greater extent than did siRNA knockdown of mTOR, nutrient deprivation, or any other treatment tested, siRNA knockdown of SRMS (30) increased autophagic flux, as indicated by the decreased abundance of GFP-LC3 and p62 (Fig. 2D and figs. S5A and S6, A to E). Knockdown of SRMS significantly increased the number of LC3-positive puncta per

cell and decreased the abundance of GFP-LC3 and p62 in a ULK1- and ATG5 (autophagy-related 5)–dependent manner (Fig. 4, A to C), whereas overexpression of SRMS decreased the number of LC3-postive puncta per cell (Fig. 3A). This suggests that SRMS inhibits autophagy upstream of autophagosome formation. SRMS-depleted cells maintained a robust response to nutrient deprivation despite increased basal autophagy, suggesting that loss of SRMS does not mimic nutrient deprivation (Fig. 4D). SRMS is unlikely to function through mTOR because knockdown of SRMS did not inhibit mTOR pathway activity, as measured by the phosphorylation of the kinase p70S6K (Fig. 4C and fig. S6E). These observations indicate a role for SRMS in nutrient-independent inhibition of autophagy.

Autophagy has tumor-suppressive activity in luminal breast epithelial cells (31). *SRMS* expression is increased in 11% of luminal breast tumors, with amplification of the *SRMS* locus observed in about one-third of cases (http://www.cbioportal.org/) (32, 33). Within this disease setting of estrogen receptor–positive breast cancer, which is typically associated with a good prognosis, tumors with increased expression of *SRMS* correlated with poor prognosis (Fig. 4E). Knockdown of SRMS inhibited the growth of the luminal breast cancer cell lines MCF7 and T47D (Fig. 4F). These

data indicate that SRMS may regulate autophagy in breast cancer and could be a therapeutic target in this disease.

Using FUSION to match natural products to biological mechanisms of action

We generated a renewable resource of natural products by harvesting Actinomycetes (65), Firmicutes (20), and α-Proteobacteria (7) from previously undersampled marine environments. The organisms were cultivated in the laboratory under defined conditions designed to enhance the success of clonal populations and reproducibility of fermentation conditions. Large-scale fermentation of each bacterial strain followed by metabolite extraction and chromatic separation produced 20 natural product fractions per strain, totaling 1186 fractions with an estimated two to six compounds per fraction (see Supplementary Materials and Methods). We screened these natural product fractions for the ability to modulate reporter gene expression and computed pairwise distances between all natural products, siRNAs, and miRNAs to create a FUSION map. Given our observations that several genetic perturbations modulated autophagy, we used FUSION mapping to predict which natural product fractions could also modulate
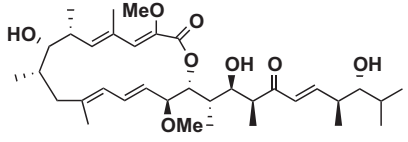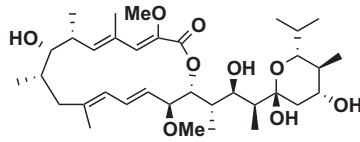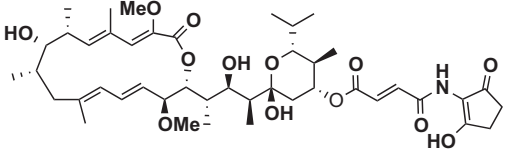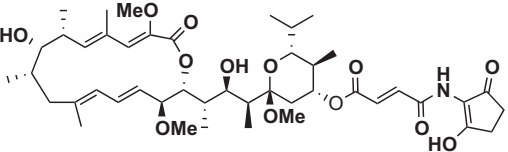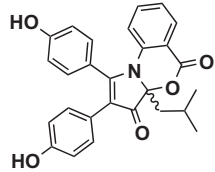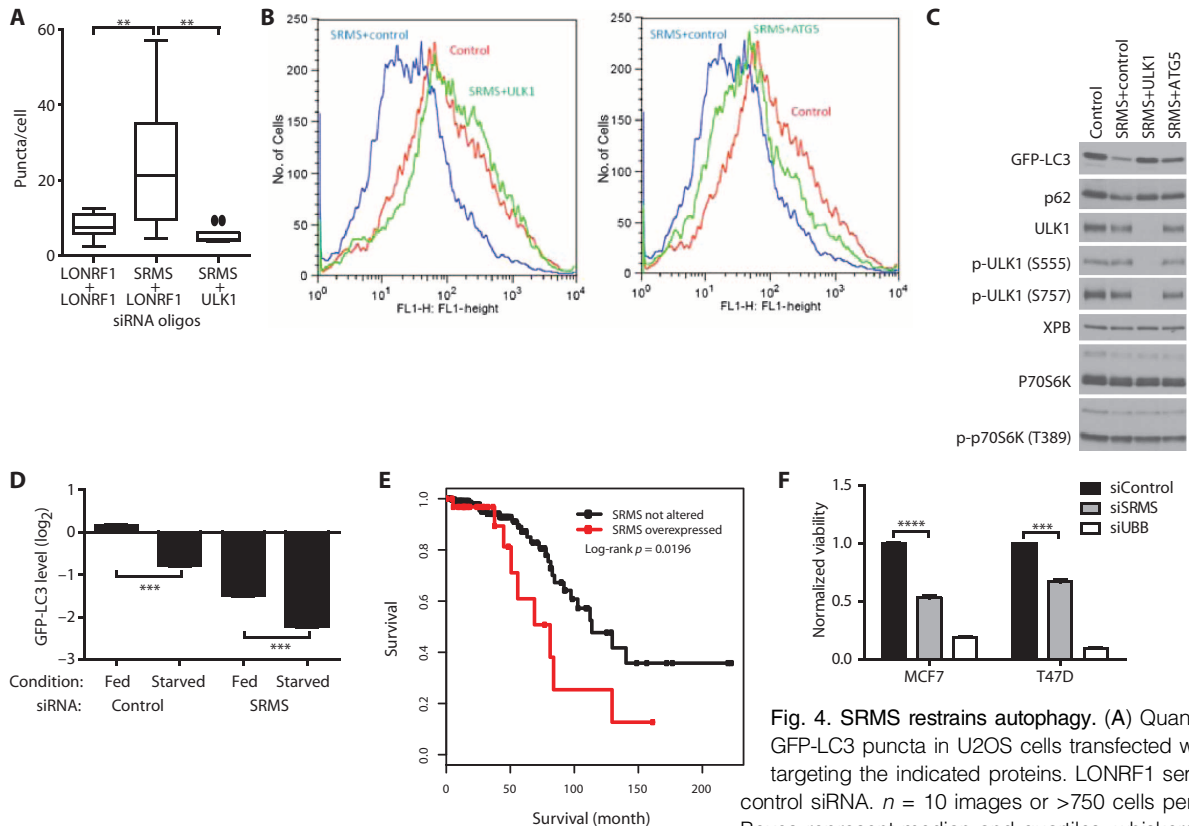

Fig. 4. SRMS restrains autophagy. (A) Quantification of GFP-LC3 puncta in U2OS cells transfected with siRNAs targeting the indicated proteins. LONRF1 served as the control siRNA. n = 10 images or >750 cells per condition. Boxes represent median and quartiles, whiskers extend to 1.5 × interquartile distance, and outliers indicated by circles. **$P < 0.01$, unpaired t test. Data are representative of five experiments. (B) GFP-LC3 fluorescence intensity in U2OS cells transfected with the indicated siRNAs. Data are representative of five experiments. (C) Western blot of U2OS cells transfected with the indicated siRNAs. Data are representative of three experiments. (D) Quantification of flow cytometry of GFP-LC3 in U2OS cells transfected with siRNA pools and then incubated in normal growth medium (fed) or EBSS nutrient deprivation medium (starved) for 2 hours. Means ± SEM; n = 3 experiments. ***$P < 0.001$, unpaired t test. (E) Fraction of patients surviving from the luminal (A/B) case set of the Breast Invasive Carcinoma TCGA Study stratified by SRMS mRNA abundance. SRMS overexpression (red) was defined as "reads per kilobase per million reads"; z score > 2.0, $P < 0.02$, log-rank test. (F) Normalized cell viability in MCF7 and T47D luminal breast cancer cell lines measured with CellTiter-Glo 96 hours after transfection with siRNAs against LONRF1 (negative control), SRMS, or UBB (positive control) in triplicate. Means ± SEM; n = 3 biological replicates. ***$P < 0.001$, ****$P < 0.0001$, unpaired t test.

autophagy. Metabolites isolated from the *Streptomyces bacillaris* strain SN-B-019 induced reporter gene signatures that positively correlated with knockdown of ULK1 (fig. S7A). Of the five pure compounds isolated from the SN-B-019 fraction, four (Table 1) inhibited autophagy as measured by accumulation of LC3-positive puncta (fig. S7B). Of these, SN-B-019-cmp1 also induced accumulation of GFP-LC3 and p62 and phosphorylation of AMPKα (Thr[172]) in a dose-dependent manner [median inhibitory concentration (IC_{50}), 150 nM; Fig. 5, A and B]. Structural determination revealed this compound to be identical to bafilomycin D (Fig. 5B and Table 1), which blocks autophagosome maturation through direct inhibition of the vacuolar adenosine triphosphatase (34, 35). These data show that FUSION mapping can successfully predict matching chemical and genetic perturbations.

Next, we asked whether FUSION mapping could be used to identify natural products that perturb cellular processes other than autophagy. The discoidin domain receptor tyrosine kinase 2 (DDR2) is encoded by a proto-oncogene in lung cancer, and activating mutations occur in ~4% of human squamous cell lung cancer (36). Moreover, in nontransformed mesenchymal cells, DDR2 supports growth factor–induced migration through collagen matrices (37). By FUSION mapping, we identified three natural product fractions with reporter gene signatures that were similar to those produced by knockdown of DDR2 with siRNA (FDR <16%; fig. S7C) and that inhibited PDGF-induced migration of primary human fibroblasts through collagen disks (Fig. 5C). One fraction, SN-A-048-7, dose-dependently increased the apparent molecular weight of DDR2, suggesting that compounds in this fraction may bind DDR2 or alter its

Table 1. Structures of the purified natural products characterized in this study.

Source (UTSW designation)	Compound designation	Compound name	Structure
Streptomyces bacillaris (SNB-019)	SNB-019-cmp1	bafilomycin D	
Streptomyces bacillaris (SNB-019)	SNB-019-cmp2	bafilomycin A1	
Streptomyces bacillaris (SNB-019)	SNB-019-cmp3	bafilomycin B1	
Streptomyces bacillaris (SNB-019)	SNB-019-cmp4	bafilomycin B2	
Bacillus hunanensis (SNA-048)	SNA-048-7-2	discoipyrrole A	
Salinispora arenicola (SNB-003)	SNB-003-cmp1	5SNB-003-11-3-2	structure not shown

posttranslational modifications (Fig. 5D). Knockdown of DDR2 is selectively toxic to lung cancer cell lines with activating mutations in *DDR2* (*36*). Compared to the effect of SN-A-048-7 on the viability of A549 cells with wild-type DDR2, SN-A-048-7 reduced the viability of the *DDR2*-mutant cell line HCC366 (Fig. 5E and fig. S7D). Subfractionation of SN-A-048-7 yielded the pure compound (Table 1) discoipyrrole A, which inhibited PDGF-induced migration of fibroblasts in a dose-dependent man-
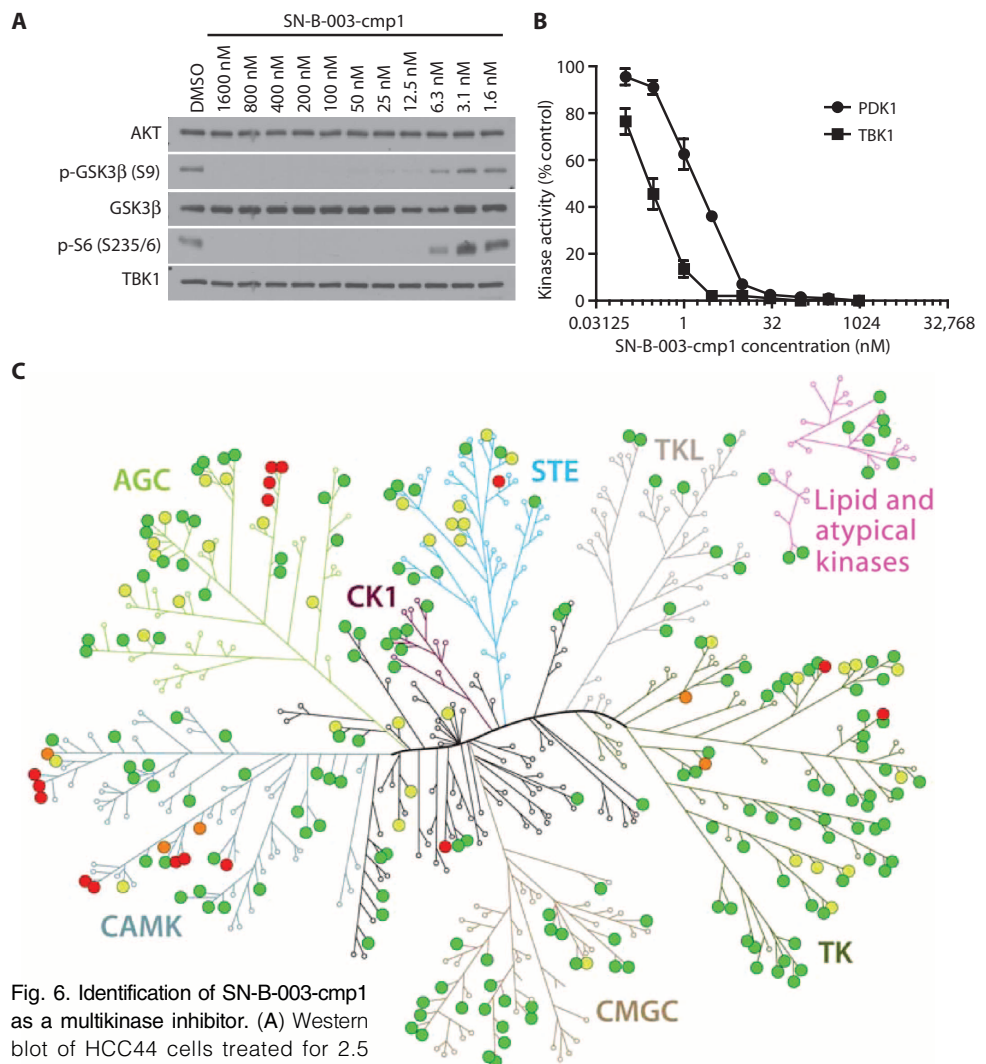
ner (Fig. 5F and fig. S7E). A combination of nuclear magnetic resonance, mass spectroscopy, and x-ray crystallization analyses of discoipyrrole A revealed a previously unreported structure composed of a core dihydro-2*H*-pyrrolooxazine-2,8(8a*H*)-dione ring system substituted at multiple positions with phenyl rings (Fig. 5F). Thus, functional validation of discoipyrrole A derived from natural products supports the conclusion that FUSION mapping is an effective means to identify biologically active, clinically relevant compounds.
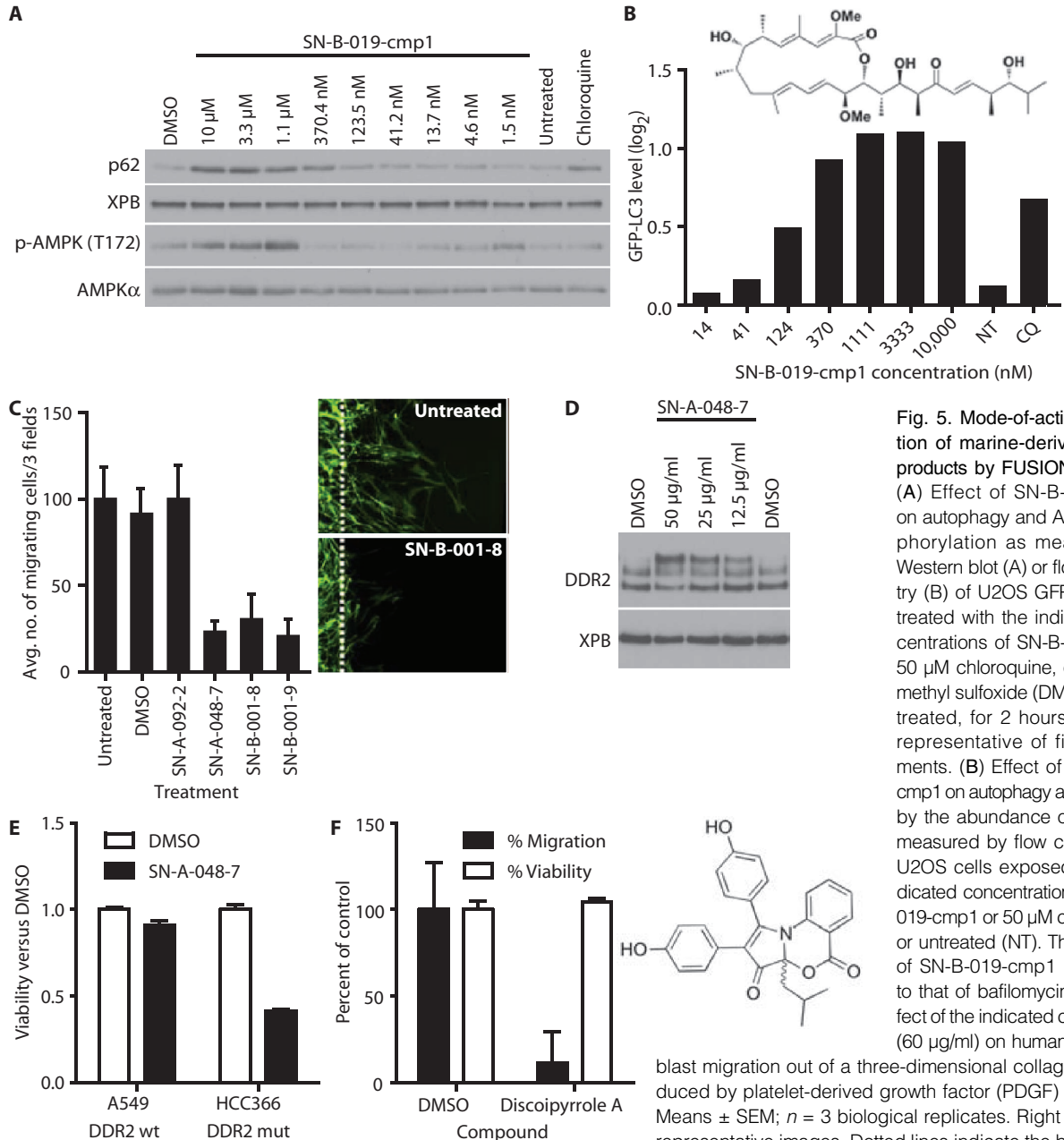
Fig. 5. Mode-of-action annotation of marine-derived natural products by FUSION mapping. (**A**) Effect of SN-B-019-cmp1 on autophagy and AMPK phosphorylation as measured by Western blot (A) or flow cytometry (B) of U2OS GFP-LC3 cells treated with the indicated concentrations of SN-B-019-cmp1, 50 μM chloroquine, or 0.1% dimethyl sulfoxide (DMSO), or untreated, for 2 hours. Data are representative of five experiments. (**B**) Effect of SN-B-019-cmp1 on autophagy as measured by the abundance of GFP-LC3 measured by flow cytometry in U2OS cells exposed to the indicated concentrations of SN-B-019-cmp1 or 50 μM chloroquine, or untreated (NT). The structure of SN-B-019-cmp1 is identical to that of bafilomycin D. (**C**) Effect of the indicated compounds (60 μg/ml) on human BR5 fibroblast migration out of a three-dimensional collagen plug induced by platelet-derived growth factor (PDGF) (50 ng/ml). Means ± SEM; *n* = 3 biological replicates. Right panels are representative images. Dotted lines indicate the boundary of the collagen plug. Cells were stained for filamentous actin. (**D**) Western blot for DDR2 in BR5 cells exposed to the indicated concentrations off SN-A-048-7. Data are representative of three experiments. (**E**) Normalized cell viability of A549 and *DDR2*-mutant HCC366 cells. Viability was measured with CellTiter-Glo. Data are means ± SEM of triplicate wells. These results are representative of five experiments. wt, wild type. (**F**) Effect of discoipyrrole A (135 μM) or DMSO (0.1%) on the migration or viability of BR5 cells. Results were normalized to DMSO-treated cells. Means ± SEM; *n* = 3 experiments. The structure of SN-A-048-7 is identical to that of discoipyrrole A.

In contrast to DDR2, aberrant activation of the AKT survival pathway is widespread throughout cancers of various tissues. Several drugs targeting AKT or other enzymes in the AKT pathway are currently in late-stage clinical trials as chemotherapies for multiple cancers (38). The protein kinases TANK-binding kinase 1 (TBK1) and 3-phosphoinositide–dependent protein kinase 1 (PDPK1; also called PDK1) directly and coordinately activate AKT and its family members by phosphorylation in many cells, including HCT116 cells (39). Therefore, we used FUSION mapping to identify 12 natural product fractions that produced reporter gene signatures similar to those produced by knockdown of TBK1 and those produced by knockdown of PDPK1 (fig. S8A), predicting that these would also reduce AKT pathway activity. Of these 12 fractions, seven inhibited AKT signaling as assessed by Western blot for Thr^{389}-phosphorylated 70-kD ribosomal protein S6 kinase polypeptide 1 (p70S6K), a target of the AKT pathway (fig. S8A). Sequential bioassay-guided subfractionation of SN-B-003-5 (fig. S8B) revealed the pure compound SN-B-003-cmp1 (Table 1), which inhibited p70S6K-mediated phosphorylation of its target ribosomal S6 protein (S6) at low nanomolar concentrations (Fig. 6A). In vitro assays with purified TBK1 or PDK1 revealed direct inhibition of kinase activity with an IC_{50} of ~300 pM for TBK1 (Fig. 6B). To determine specificity, we tested SN-B-003-cmp1 against a panel of 235 human kinases at 1× and 10× the IC_{50} for TBK1. At 300 pM, we observed 50% or more inhibition of TBK1 and calcium/calmodulin-dependent protein kinase IIγ (CAMKIIγ), whereas at 3 nM, we also found inhibition of PDPK1; AMPKα; glucokinase; MAP/microtubule affinity–regulating kinase 2; 90-kD ribosomal protein S6 kinase polypeptides 1, 2, 3, and 6; and testis-specific serine kinase 1 (Fig. 6, B and C, and table S4). These results suggest that SN-B-003-cmp1 is a multikinase inhibitor with moderate specificity for TBK1 and CAMKIIγ and further support the conclusion that FUSION mapping can be used to identify biologically active molecules from natural product libraries.

Fig. 6. Identification of SN-B-003-cmp1 as a multikinase inhibitor. (A) Western blot of HCC44 cells treated for 2.5 hours with the indicated concentrations of SN-B-003-cmp1. Data are representative of five experiments. (B) Dose-response curves for SN-B-003-cmp1 on recombinant PDK1 (IC_{50}, 2 nM) and TBK1 (IC_{50}, 0.3 nM). Means ± range of $n = 2$ experiments. (C) Effects of 3 nM SN-B-003-cmp1 on the activity of 235 kinases tested in vitro. Red circles, >75% inhibition; orange, >50% inhibition; yellow, >25% inhibition; green, <25% inhibition. See table S4 for further details.

DISCUSSION

Assigning mechanism of action to bioactive molecules is a rate-limiting step in biomedical research and pharmaceutical development. We constructed a similarity matrix of reporter gene signatures among miRNAs, siRNAs, and natural products using a simple distance metric. The resulting "guilt by association" relationships were used to generate hypotheses about the cellular pathways and processes that these molecules can perturb. Experimental testing of these hypotheses confirmed the identification of underappreciated roles of several proteins and miRNAs in cell biological processes that are important for normal physiology and disease. Moreover, we were able to assign modes of action and suggest potential therapeutic uses for novel marine-derived natural products. To leverage this resource beyond the chemical-genetic relationships described in this study, we have provided an open-source Web-based tool to access our FUSION map for users to identify natural product fractions with statistically similar reporter gene signatures to miRNAs and genes of interest (http://whitelab.swmed.edu/fmap/fusion1.php). Moreover, the FUSION map is amenable to expansion with data from screening of additional genetic and chemical perturbations. Therefore, this resource has the potential to enhance the pace of cell-based chemical lead discovery.

MATERIALS AND METHODS

Cell culture and siRNA, miRNA, and plasmid delivery

Cell lines were gifts from F. Grinnell (BR5 human fibroblasts), X. Wang (U2OS GFP-LC3), R. Potts (U2OS), G. Pearson (T47D and MCF7), and J. Minna (A549, HCC366, and HCC44) or were purchased from the American Type Culture Collection (HCT116). Parameters for cell culture and transfection of each line are listed in table S5. U2OS mito-mCherry cells stably harbor a mitochondrially localized mCherry and were made by transfecting parental U2OS cells with pcDNA3 containing the 5′ COX VIII leader sequence fused to mCherry (27) (a gift from C. Junker) and selecting a single stable clone.

miRNA and siRNA reagents and antibodies

We used the miRIDIAN miRNA library and SMARTpool Kinome siRNA library for cell-based screening (Dharmacon, Thermo Scientific). Additional siRNAs were purchased from Dharmacon, Sigma, or Ambion and used as pools of four targeting a specific gene except where use of individual siRNAs was indicated. Antibodies against AMPKα, glycogen synthase kinase 3β (GSK3β), p70S6K, phosphorylated acetyl–coenzyme A carboxylase α (Ser^{79}), phosphorylated AKT (Ser^{473}), phosphorylated AMPK (Thr^{172}), pan-AKT, phosphorylated GSK3β (Ser^9), phospho-p70S6K (Thr^{389}), phosphorylated Rictor (RPTOR-independent companion of mTOR) (Thr^{1135}), phosphorylated ribosomal protein S6 (Ser$^{235/6}$), phosphorylated tuberous sclerosis 2 (TSC2) (Thr^{1462}), phosphorylated ULK1 (Ser^{555}), phosphorylated ULK1 (Ser^{757}), Rictor, TBK1, TSC2, and ULK1 were from Cell Signaling Technology. The p62 antibody was from Progen Biotechnik, the GFP and XPB (excision repair cross-complementing rodent repair deficiency, complementation group 3) antibodies were from Santa Cruz Biotechnology, the β-tubulin antibody was from Sigma, and the DDR2 antibody was from R&D Systems.

Natural product library

The natural product library used for this study was composed of extracts from 92 marine-derived bacterial strains and 20 marine invertebrates (19 sponges, 1 tunicate). The library of microbial natural product fractions was derived from marine-derived Actinomycetes (65), Firmicutes (20), and α-Proteobacteria (7). These bacteria were cultivated from marine sediment samples collected in the Gulf of Mexico (Texas and Louisiana), estuaries in South Carolina, and the Bahamas. Various techniques were used to isolate strains, including the use of small-molecule signaling compounds (N-acylhomoserine lactones, siderophores, and surfactins) that mimic the natural environment of the bacteria of interest. Nutrient-limited isolation media, such as those composed of only humic or fulvic acid, were used for isolation of actinomycetes and α-proteobacteria from mangrove and estuary sediments. Most of the bacterial isolates were characterized by 16S ribosomal RNA (rRNA) phylogenetic analysis following previously described methods (40). Universal 16S rRNA primers FC27 and RC 1492 were used. The sequences of 16S rRNA were compared to sequences in available databases, using the Basic Local Alignment Search Tool. Specific details of the strain isolation and phylogenetics for bacteria discussed in the text are included below. For bacterial isolates not characterized by 16S rRNA, we used morphology to ensure that strains in the collection were not duplicated.

To generate the library, bacterial strains were fermented in 5×2.8-liter Fernbach flasks each containing 1 liter of a seawater-based medium (10 g of starch, 4 g of yeast extract, 2 g of peptone, 1 g of $CaCO_3$, 40 mg of $Fe_2(SO_4)_3 \cdot 4H_2O$, 100 mg of KBr) and shaken at 200 rpm for 7 days at 27°C. After 7 days of cultivation, XAD7-HP resin (20 g/liter) was added to adsorb the organic products, and the culture and resin were shaken at 200 rpm for 2 hours. The resin was filtered through a cheesecloth, washed with deionized water, and eluted with acetone to give a crude extract. Fractionation of the bacterial crude extracts was accomplished using an Isco medium-pressure automatic purification system with reversed-phase C_{18} chromatography (gradient from 90:10 H_2O/CH_3CN to 0:100 H_2O/CH_3CN over 25 min). Each bacterial strain gave rise to 20 fractions for further testing. All natural product fractions in the library were standardized to a concentration of 10 mg/ml in DMSO.

The following generic code was used for the natural product fraction nomenclature, SN-X-###-##. SN-A designates strains for 2007 collection; SN-B designates strains for 2008 collection. As an example, SN-A-048 would be the 48th strain isolated in the 2007 collection. Further, numbers SN-A-048-# refer to the active fraction in each of the purification steps leading to the pure compound. For example, SN-A-048-7 would designate fraction number 7 from the first purification of the crude extract of SN-A-048. Identification of pure bioactive compounds requires regrowth of the bacterial culture followed by bioassay-guided fractionation. Intermediate purification fractions were labeled SN-X-###-int#, and resulting pure compounds were labeled SN-X-###-cmp#. Complete details of the isolation and purification of natural products can be found in Supplementary Materials and Methods. Details of the analysis of the purified compounds are provided in fig. S9.

Cell-based high-throughput screens

All gene expression–based functional signature screens were performed in triplicate in the human colon tumor cell line HCT116 in 384-well microtiter plate format. Cells were propagated in Dulbecco's modified Eagle's medium (DMEM) (Cellgro) supplemented with 10% fetal bovine serum (serum lot M0017, Atlanta Biologicals), L-glutamine, and nonessential amino acids (propagation medium). Transfection of Thermo Fisher Scientific's (Dharmacon) human siARRAY library SMARTpools and miRNA mimics was based on the Wet Reverse Transfection Version 2.0 protocol with modifications developed for use with equipment and layout of the Eppley Cancer Center HTS Screening Facility. siRNA pools and miRNA mimics (final concentrations of 50 nM) were complexed with Dharmafect 4 (0.05 μl per well) reagent before plating with HCT116 cells (2500 cells per well) in a final volume of 30 μl of propagation medium. After 72 hours, cells in each individual well were lysed in 20 μl of QuantiGene Plex 2.0 assay (Panomics) lysis buffer containing 10 μl of proteinase K per milliliter of lysis buffer and stored at −80°C until further assayed for gene expression signatures. Natural product library screens were performed in HCT116 cells (2500 cells per well) plated in 50 μl of propagation medium. Forty-eight hours after plating, cells were treated with natural product extract (60 μg/ml; 0.3 μl) or the indicated concentration of known drug for 21 hours. Cell lysates were generated and frozen as described above. Although optimal timing for individual reagents within each library varied, this scheme was chosen to maximize the ability to detect changes in gene expression induced by most reagents within each class. Endogenous expression of *ALDOC*, *NDRG1*, *BNIP3*, *BNIP3L*, *ACSL5*, *LOXL2*, *HPRT*, and *PPIB* mRNA in each well was measured using QuantiGene Plex 2.0 assays (Panomics) with a Luminex 200 machine (Luminex) following the manufacturer's protocols. The protocol was based on branched DNA signal amplification with eight fluorescent magnetic bead panels, with each bead panel composed of eight different fluorescent magnetic beads (8×8 plex, with a total number of individual fluorescent bead classifications = 64). Each cell lysate (40 μl) was added directly to a distinct bead and reagent mix and hybridized overnight at 54°C in a Cytomat 2 shaking incubator (1500 Hz). After hybridization, branched signal amplification steps with preamplifier, amplifier, biotin label, and streptavidin-phycoerythrin reagents were performed according to the manufacturer's protocol and measured with Luminex 200. A background correction was applied by subtracting the mean of the vehicle-only control well values from the raw values

for each of the eight probes. To normalize well-to-well signal variation, the six background-corrected probe values per well were divided by the geometric mean of the two internal control probes, *HPRT* and *PPIB*. To normalize probe-to-probe variation, each probe was further median-normalized and $\log_2$-transformed. The distribution of the normalized expression values for each probe is shown in fig. S10; complete data are given in table S1.

Informatics and statistics
To quantify the similarity between expression profiles of different genetic and chemical perturbagens, we measured ordinary distance by the Euclidean and Mahalanobis methods, and nonordinary distance by rank conversion of the Euclidean distance and by Pearson correlation coefficient. These strategies were pursued in parallel, given the distinct strengths and caveats of each. Euclidean distance was used for most of the analyses here except where otherwise indicated. Although the accuracy of this distance measure can be affected by any covariance structure present in the expression data set, the six reporter genes used here responded relatively independently to each other (fig. S11A). However, given the presence of minor but statistically significant correlations, such as that observed between *NDRG1* and *ACSL5* ($r = 0.42$), Mahalanobis distance, which accounts for covariance, was also calculated. The Mahalanobis distance is reduced to Euclidean distance when the covariance matrix is the identity matrix. A correlation plot of corresponding Euclidean and Mahalanobis distances, for all pairwise comparisons within the siRNA and miRNA mimic data sets, suggests that the covariance structure makes, at most, a minor contribution to the distance measures (fig. S11B). To detect coherent expression signatures with different magnitudes, as might result from differential knockdown efficiency or chemical potency, we also used correlation coefficients. In all cases, to systematically assess the significance of short distance pairs, we calculated test statistics for distance-sum over triplicate measures from permuted data. Gene labels for each of the six reporter genes were permuted 100 times for each of the triplicates, and the background distance-sum distribution was obtained. The empirical P value for each pair of perturbations was obtained from the background density distribution. FDRs were defined by the fraction of expected false positives among all significant hypotheses, which was estimated by fitting a beta-uniform mixture model to the estimated P values using the Bum class of functions in the ClassComparison package for R (http://bioinformatics.mdanderson.org/Software/OOMPA; fig. S12). The natural product library screen was processed in the same way, and resulting empirical P values were corrected for multiplicity of testing by the Benjamini-Hochberg method. All data processing, permutation, and P value estimation were carried out using the statistical platform R (http://www.R-project.org). Hierarchical and neighbor joining clustering of Euclidean distance were carried out with the "hclust" and "ape" packages, respectively. "SplitsTree4" was used to draw neighbor joining trees (*41*). High-resolution PDFs of each analysis are provided (Fig. 1A and figs. S2 to S4A). A Web-based search and display tool for FUSION matrix is available at http://whitelab.swmed.edu/fmap/fusion1.php.

MiRNA mimic sequence alignments and functional similarity density distributions were assessed as follows. Among the 90,525 possible miRNA pairs present in the miRBase version 8.0 miRIDIAN library, 105 pairs have identical sequences, 330 pairs have identical seeds (as defined by nucleotides 2 to 7) but different sequences, and 90,090 pairs have different seeds. Pearson correlations, from the means of triplicate reporter gene expression values, were calculated for each pair of miRNA mimics, and density plots were drawn for all correlations within each group. Wilcoxon tests were used to evaluate significant differences between groups. To calculate sequence similarities between miRNA pairs, the sequence information for each miRNA was FASTA-formatted for input to the CLUSTALW

(*42*) multiple sequence alignment using BLOSUM as the substitution matrix of preference. The output alignment file was used to compute a similarity table using PHYLIP DNADIST (*43*) [Pasteur Institute of France at http://mobyle.pasteur.fr (*44*)], with values indicating the fraction of identical sites between sequences. The diagonal values were 1.0000.

Endogenous miRNA expression arrays
Total RNA was collected from 3 million cells with the RNeasy Plus Mini Kit (Qiagen) following the manufacturer's protocols for collection of small RNA and mRNA simultaneously. The Human Illumina miRNA Array platform was used to measure the abundance of mature miRNAs in HCT116 cells in biological quadruplicate. Data normalization was performed using quantile normalization and the Partek Batch Effect Remover algorithm. Results are given in table S2. In fig. S1D, negative values were replaced with the value "1."

GFP-LC3 and mito-mCherry quantitation
Total GFP or mCherry fluorescence per cell was measured by flow cytometry of at least 20,000 cells per condition on a BD FACSCalibur using CellQuest and FlowJo software. For visualization of GFP-LC3–positive puncta, cells were cultured on glass coverslips, rinsed in phosphate-buffered saline, fixed with 3.7% paraformaldehyde, permeabilized with cold acetone, and mounted with Vectashield containing 4′,6-diamidino-2-phenylindole. Images were acquired using a Zeiss Axioplan 2E microscope with a Hamamatsu monochrome digital camera with OpenLab software.

Erythroblast maturation assays
Ter119-negative erythroid progenitors were isolated from embryonic day 13.5 Balb/c murine fetal livers as described (*45*). The purity of the cells was confirmed by staining cells with fluorescein isothiocyanate–CD71 and allophycocyanin (APC)–Ter119 antibodies (all from BD Biosciences) by flow cytometry. Ter119-negative erythroid progenitors were transduced with retroviruses expressing shRNA against ULK1, BMP2K, or a nonsilencing shRNA control in the pWH99 vector that also expresses GFP (provided by H. Lodish). Twenty-four hours later, Ter119-negative cells were repurified and differentiated. Erythroid differentiation was monitored on days 1 and 3 with phycoerythrin-CD71 and APC-Ter119 by flow cytometry (*46*). Propidium iodide staining was used to exclude dead cells, and GFP-positive cells were gated for analysis. The data were acquired on a BD FACSCalibur and analyzed by FlowJo 8.8.4. To measure mitochondrial content, cells were stained with 500 nM MitoTracker Deep Red FM (Molecular Probes, Invitrogen) for 30 min at 37°C on day 3 of in vitro differentiation. GFP-positive cells were gated, and histograms were overlaid using FlowJo 8.8.4.

Migration through collagen
BR5 human fibroblasts were suspended in rat tail collagen (4 mg/ml; Collagen I, High Concentration from BD Bioscience, catalog no. 354249) at a concentration of 30×10^6 cells/ml. Two microliters was placed in the center of each well in a 96-well plate and incubated for 10 min to allow collagen polymerization. The remainder of the well was then coated with rat tail collagen (3 mg/ml) without cells for 10 min, followed by two washes with DMEM and the addition of growth medium containing PDGF (50 ng/ml) to stimulate migration with or without DMSO or the indicated natural products in triplicate wells per condition. After overnight incubation, cells were fixed, stained for actin and DNA, and imaged. Migration was quantified as the number of nuclei outside the rim of the initial collagen plug. Triplicates were averaged, and values were normalized to treatment with DMSO alone to obtain percent migration compared to control.

Cell viability assays

A549 or HCC366 cells were seeded at 5000 cells per 96-well and allowed to adhere overnight. SN-A-048-7 (6.7 µg/ml) or DMSO alone was added. Five days later, viability was measured with CellTiter-Glo.

Protein kinase activity assays

Kinase assays for TBK1 were performed as previously described (39). SN-B-003-cmp1 at the indicated concentrations was assayed with the KinaseProfiler (Millipore) in vitro kinase activity assays for TBK1, PDK1, and a panel of 235 additional human kinases (table S4).

SUPPLEMENTARY MATERIALS

www.sciencesignaling.org/cgi/content/full/6/297/ra90/DC1
Materials and Methods: Natural Products and Producing Organisms
Fig. S1. Characteristics of the reporter genes used for FUSION mapping.
Fig. S2. Mapping functional similarity among miRNAs.
Fig. S3. Mapping functional similarity among kinases and kinase accessory proteins.
Fig. S4. Using FUSION mapping to identify miRNAs and kinases that regulate autophagy.
Fig. S5. Representative flow cytometry data and characterization of the role of BMP2K in autophagy and erythrocyte development.
Fig. S6. SRMS as an inhibitor of autophagy.
Fig. S7. Using FUSION mapping to identify natural product inhibitors of autophagy and migration.
Fig. S8. Using FUSION mapping to identify inhibitors of AKT signaling.
Fig. S9. Natural product structure determination.
Fig. S10. Profiles of reporter gene expression changes in response to each of 1124 genetic perturbations.
Fig. S11. Analysis of correlation among reporter gene expression and between distance metrics.
Fig. S12. Estimation of FDR from P value density distribution, rank sum as an example.
Table S1. Reporter gene expression in response to individual genetic and chemical perturbations in HCT116 cells.
Table S2. Endogenous miRNA expression.
Table S3. MiRNA-siRNA oligonucleotide seed identities.
Table S4. Kinase inhibition screen (KinaseProfiler).
Table S5. Maintenance and transfection conditions for cell lines used in this study.

REFERENCES AND NOTES

1. D. J. Newman, G. M. Cragg, Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **75**, 311–335 (2012).
2. R. H. Shoemaker, The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer* **6**, 813–823 (2006).
3. G. T. Carter, Natural products and Pharma 2011: Strategic changes spur new opportunities. *Nat. Prod. Rep.* **28**, 1783–1789 (2011).
4. H. Hieronymus, J. Lamb, K. N. Ross, X. P. Peng, C. Clement, A. Rodina, M. Nieto, J. Du, K. Stegmaier, S. M. Raj, K. N. Maloney, J. Clardy, W. C. Hahn, G. Chiosis, T. R. Golub, Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* **10**, 321–330 (2006).
5. T. R. Hughes, M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade, P. Y. Lum, S. B. Stepaniants, D. D. Shoemaker, D. Gachotte, K. Chakraburtty, J. Simon, M. Bard, S. H. Friend, Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
6. J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, T. R. Golub, The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
7. K. Stegmaier, Genomic approaches in acute leukemia. *Best Pract. Res. Clin. Haematol.* **19**, 263–268 (2006).
8. K. Stegmaier, K. N. Ross, S. A. Colavito, S. O'Malley, B. R. Stockwell, T. R. Golub, Gene expression–based high-throughput screening (GE-HTS) and application to leukemia differentiation. *Nat. Genet.* **36**, 257–263 (2004).
9. M. Zimmer, J. Lamb, B. L. Ebert, M. Lynch, C. Neil, E. Schmidt, T. R. Golub, O. Iliopoulos, The connectivity map links iron regulatory protein-1–mediated inhibition of hypoxia-inducible factor-2a translation to the anti-inflammatory 15-deoxy-$\Delta^{12,14}$-prostaglandin J$_2$. *Cancer Res.* **70**, 3071–3079 (2010).
10. K. K. Farh, A. Grimson, C. Jan, B. P. Lewis, W. K. Johnston, L. P. Lim, C. B. Burge, D. P. Bartel, The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science* **310**, 1817–1821 (2005).
11. D. P. Bartel, MicroRNAs: Target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
12. J. A. Papin, T. Hunter, B. O. Palsson, S. Subramaniam, Reconstruction of cellular signalling networks and analysis of their properties. *Nat. Rev. Mol. Cell Biol.* **6**, 99–111 (2005).
13. O. L. Griffith, E. D. Pleasance, D. L. Fulton, M. Oveisi, M. Ester, A. S. Siddiqui, S. J. Jones, Assessment and integration of publicly available SAGE, cDNA microarray, and oligo-nucleotide microarray expression data for global coexpression analyses. *Genomics* **86**, 476–488 (2005).
14. M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863–14868 (1998).
15. N. Saitou, M. Nei, The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
16. S. Griffiths-Jones, R. J. Grocock, S. van Dongen, A. Bateman, A. J. Enright, miRBase: MicroRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34**, D140–D144 (2006).
17. D. R. Caffrey, J. Zhao, Z. Song, M. E. Schaffer, S. A. Haney, R. R. Subramanian, A. B. Seymour, J. D. Hughes, siRNA off-target effects can be reduced at concentrations that match their individual potency. *PLOS One* **6**, e21503 (2011).
18. B. D. Parsons, A. Schindler, D. H. Evans, E. Foley, A direct phenotypic comparison of siRNA pools and multiple individual duplexes in a functional assay. *PLOS One* **4**, e8471 (2009).
19. A. Birmingham, E. M. Anderson, A. Reynolds, D. Ilsley-Tyree, D. Leake, Y. Fedorov, S. Baskerville, E. Maksimova, K. Robinson, J. Karpilow, W. S. Marshall, A. Khvorova, 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods* **3**, 199–204 (2006).
20. A. L. Jackson, J. Burchard, J. Schelter, B. N. Chau, M. Cleary, L. Lim, P. S. Linsley, Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* **12**, 1179–1187 (2006).
21. E. Y. Chan, S. Kir, S. A. Tooze, siRNA screening of the kinome identifies ULK1 as a multidomain modulator of autophagy. *J. Biol. Chem.* **282**, 25464–25474 (2007).
22. J. Kim, M. Kundu, B. Viollet, K. L. Guan, AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1. *Nat. Cell Biol.* **13**, 132–141 (2011).
23. C. Behrends, M. E. Sowa, S. P. Gygi, J. W. Harper, Network organization of the human autophagy system. *Nature* **466**, 68–76 (2010).
24. H. M. Ni, A. Bockus, A. L. Wozniak, K. Jones, S. Weinman, X. M. Yin, W. X. Ding, Dissecting the dynamic turnover of GFP-LC3 in the autolysosome. *Autophagy* **7**, 188–204 (2011).
25. M. Kundu, T. Lindsten, C. Y. Yang, J. Wu, F. Zhao, J. Zhang, M. A. Selak, P. A. Ney, C. B. Thompson, Ulk1 plays a critical role in the autophagic clearance of mitochondria and ribosomes during reticulocyte maturation. *Blood* **112**, 1493–1502 (2008).
26. A. I. Su, T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M. P. Cooke, J. R. Walker, J. B. Hogenesch, A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6062–6067 (2004).
27. A. Quintana, C. Kummerow, C. Junker, U. Becherer, M. Hoth, Morphological changes of T cells following formation of the immunological synapse modulate intracellular calcium signals. *Cell Calcium* **45**, 109–122 (2009).
28. H. Zhu, H. Wu, X. Liu, B. Li, Y. Chen, X. Ren, C. G. Liu, J. M. Yang, Regulation of autophagy by a *beclin 1*-targeted microRNA, miR-30a, in cancer cells. *Autophagy* **5**, 816–823 (2009).
29. Z. Zou, L. Wu, H. Ding, Y. Wang, Y. Zhang, X. Chen, C. Y. Zhang, Q. Zhang, K. Zen, MicroRNA-30a sensitizes tumor cells to cis-platinum via suppressing beclin 1-mediated autophagy. *J. Biol. Chem.* **287**, 4148–4156 (2012).
30. N. Kohmura, T. Yagi, Y. Tomooka, M. Oyanagi, R. Kominami, N. Takeda, J. Chiba, Y. Ikawa, S. Aizawa, A novel nonreceptor tyrosine kinase, Srm: Cloning and targeted disruption. *Mol. Cell. Biol.* **14**, 6915–6925 (1994).
31. V. M. Aita, X. H. Liang, V. V. Murty, D. L. Pincus, W. Yu, E. Cayanis, S. Kalachikov, T. C. Gilliam, B. Levine, Cloning and genomic organization of *beclin 1*, a candidate tumor suppressor gene on chromosome 17q21. *Genomics* **59**, 59–65 (1999).
32. E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: An open platform for exploring multi-dimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
33. J. Gao, B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
34. G. Carr, D. E. Williams, A. R. Díaz-Marrero, B. O. Patrick, H. Bottriell, A. D. Balgi, E. Donohue, M. Roberge, R. J. Andersen, Bafilomycins produced in culture by *Streptomyces* spp. isolated from marine habitats are potent inhibitors of autophagy. *J. Nat. Prod.* **73**, 422–427 (2010).

35. A. Yamamoto, Y. Tagawa, T. Yoshimori, Y. Moriyama, R. Masaki, Y. Tashiro, Bafilomycin A1 prevents maturation of autophagic vacuoles by inhibiting fusion between autophagosomes and lysosomes in rat hepatoma cell line, H-4-II-E cells. *Cell Struct. Funct.* **23**, 33–42 (1998).

36. P. S. Hammerman, M. L. Sos, A. H. Ramos, C. Xu, A. Dutt, W. Zhou, L. E. Brace, B. A. Woods, W. Lin, J. Zhang, X. Deng, S. M. Lim, S. Heynck, M. Peifer, J. R. Simard, M. S. Lawrence, R. C. Onofrio, H. B. Salvesen, D. Seidel, T. Zander, J. M. Heuckmann, A. Soltermann, H. Moch, M. Koker, F. Leenders, F. Gabler, S. Querings, S. Ansén, E. Brambilla, C. Brambilla, P. Lorimier, O. T. Brustugun, A. Helland, I. Petersen, J. H. Clement, H. Groen, W. Timens, H. Sietsma, E. Stoelben, J. Wolf, D. G. Beer, M. S. Tsao, M. Hanna, C. Hatton, M. J. Eck, P. A. Janne, B. E. Johnson, W. Winckler, H. Greulich, A. J. Bass, J. Cho, D. Rauh, N. S. Gray, K. K. Wong, E. B. Haura, R. K. Thomas, M. Meyerson, Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer. *Cancer Discov.* **1**, 78–89 (2011).

37. P. A. Ruiz, G. Jarai, Discoidin domain receptors regulate the migration of primary human lung fibroblasts through collagen matrices. *Fibrogenesis Tissue Repair* **5**, 3 (2012).

38. P. Wu, Y. Z. Hu, PI3K/Akt/mTOR pathway inhibitors in cancer: A perspective on clinical progress. *Curr. Med. Chem.* **17**, 4326–4341 (2010).

39. Y. H. Ou, M. Torres, R. Ram, E. Formstecher, C. Roland, T. Cheng, R. Brekken, R. Wurz, A. Tasker, T. Polverino, S. L. Tan, M. A. White, TBK1 directly engages Akt/PKB survival signaling to support oncogenic transformation. *Mol. Cell* **41**, 458–470 (2011).

40. T. J. Mincer, P. R. Jensen, C. A. Kauffman, W. Fenical, Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. *Appl. Environ. Microbiol.* **68**, 5005–5011 (2002).

41. D. H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).

42. J. D. Thompson, D. G. Higgins, T. J. Gibson, CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

43. A. Lim, L. Zhang, WebPHYLIP: A web interface to PHYLIP. *Bioinformatics* **15**, 1068–1069 (1999).

44. B. Néron, H. Ménager, C. Maufrais, N. Joly, J. Maupetit, S. Letort, S. Carrere, P. Tuffery, C. Letondal, Mobyle: A new full web bioinformatics framework. *Bioinformatics* **25**, 3005–3011 (2009).

45. R. Sulahian, O. Cleaver, L. J. Huang, Ligand-induced EpoR internalization is mediated by JAK2 and p85 and is impaired by mutations responsible for primary familial and congenital polycythemia. *Blood* **113**, 5287–5297 (2009).

46. M. Socolovsky, A. E. Fallon, S. Wang, C. Brugnara, H. F. Lodish, Fetal anemia and apoptosis of red cell progenitors in Stat5a$^{-/-}$5b$^{-/-}$ mice: A direct role for Stat5 in Bcl-X$_L$ induction. *Cell* **98**, 181–191 (1999).

47. S. Bandyopadhyay, C. Y. Chiang, J. Srivastava, M. Gersten, S. White, R. Bell, C. Kurschner, C. Martin, M. Smoot, S. Sahasrabudhe, D. L. Barber, S. K. Chanda, T. Ideker, A human MAP kinase interactome. *Nat. Methods* **7**, 801–805 (2010).