(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 June 2012 (21.06.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/083231 A1

(51) International Patent Classification:
*C12Q 1/68* (2006.01)

(21) International Application Number:
PCT/US2011/065611

(22) International Filing Date:
16 December 2011 (16.12.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/423,975 16 December 2010 (16.12.2010) US

(71) Applicant *(for all designated States except US)*: **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139-4307 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **JIANG, Hai. PRITCHARD, Justin. LAUFFENBURGER, Douglas, A.. HEMANN, Michael.**

(74) Agents: **COLLINS, Anne, J.** et al.; Hamilson, Brook, Smith & Reynolds, P.C., 530 Virginia Rd., P.O. Box 9133, Concord, MA 01742-9133 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published**:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2012/083231 A1

(54) **Title**: RNAI BASED METHOD OF DRUG SCREENING AND CHARACTERIZATION

(57) **Abstract**: The invention is directed to a method of characterizing a mechanism of action of an agent (*e.g.*, a chemotherapeutic agent, a genotoxic agent). The method comprises contacting a plurality of populations of cells with an agent to be assessed, wherein each population of cells have one gene of interest targeted by a small hairpin RNA (shRNA) and wherein said gene of interest regulates cell death and a plurality of genes that regulate cell death are targeted in the plurality of populations of cells. A responsiveness of each population of cells to the agent is determined, thereby obtaining an shRNA signature of the agent, so as to identify one or more genes that mediate a response to the agent, thereby characterizing the mechanism of action of the agent. The invention is also directed an article of manufacture for characterizing a mechanism of action of a chemotherapeutic or genotoxic agent.

# RNAI BASED METHOD OF DRUG SCREENING AND CHARACTERIZATION

## RELATED APPLICATION(S)

This application claims the benefit of U.S. Provisional Application No.61/423,975, filed on December 16, 2010.

The entire teachings of the above application are incorporated herein by reference.

## GOVERNMENT SUPPORT

This invention was made with government support under R01 CA128803-03 awarded by the National Institutes of Health. The government has certain rights in the invention.

## BACKGROUND OF THE INVENTION

Chemotherapy remains the frontline therapy for systemic malignancies. However, drug development has been severely hampered by an inability to efficiently elucidate mechanisms of drug action. This limits both the development of modified compounds with improved efficacy and the capability to predict mechanisms of drug resistance and select optimal patient populations for a given agent. Although drug-target interactions have traditionally been examined using biochemical approaches (Sato, S., et al., Chem. Biol. 17,616-623 (2010)), a number of genetic strategies have been developed to identify pathways targeted by uncharacterized small molecules. A well-established genetic approach to drug classification is chemogenomic profiling in yeast (Giaever, G, et al. Nat. Genet, 21,278-283 (1999); Giaever, G, et al. Proc, Natl. Acad, Sci. USA 101,793-798 (2004); Lum, P.Y. et al, Cell 116, 121-137 (2004); Parsons, A,B, et al. Nat. Bioteehnol. 22,62-69 (2004); Hillenmeyer, M.E. et al. Science 320, 362-365 (2008)). In this approach, bar-coded yeast deletion strains are exposed to select agents, and genotype-dependent drug sensitivity is used to identify genes and pathways affected by a given drug, as well as to develop a response signature that can be compared with other chemical or genetic perturbations (Parsons, A,B, et al. Nat. Bioteehnol. 22,62-69 (2004); Parsons, A.B. et al. Cell 126, 611-625 (2006);

Hillenmeyer, M.E. et al. Genome Biol. 11, R30 (2010)). This approach has proven quite powerful and has been broadly disseminated; however, its efficacy in interrogating cancer chemotherapeutics is limited by the lack of conservation of certain drug targets from yeast to mammals. This is a particular problem in the context of targeted therapeutics, which are frequently directed toward alterations that are specific to mammalian tumors.

More recently, genetic approaches have been developed to examine drug action in mammalian settings. One such approach is to examine drug response in a diverse panel of tumor cell lines (Shoemaller, R.H. Nat. Rev. Cancer 6, 813-823 (2006)). In this case, the pattern of cell line sensitivity and resistance can serve as a signature that defines drug mechanism. Additionally, drug response can be correlated with the presence of specific cancer related alterations, although this analysis can be confounded by the large diversity of alterations present in a given tumor. An alternative approach is to compare the global transcriptional changes induced by test compounds to those induced by known drugs or defined genetic alterations (Hughes. T.R. et al. Cell 102, 109-126 (2000); Gardner, T.S et al., Science 301,102-105 (2003); Lamb, I. et al. Science 313, 1929-1935 (2006); Hieronymus, H. et al. Cancer Ce11,10, 321-330 (2006)). Gene expression changes are used as signatures that are characteristic of exposure to a given agent or the presence of a specific cellular state, and common expression changes can be used to cluster similar small molecules. Although each of these approaches have yielded important new insights into drug action, these strategies retain a level of technical variability and resource requirement that limits both disseminated use and overall efficacy.

Thus, a need exists for improved methods that screen and characterize drugs.

## SUMMARY OF THE INVENTION

Described herein is a tractable ribonucleic acid interference-based (RNAi-based) approach that represents a simple yet powerful platform for drug screening and characterization.

Specifically, the invention is directed to a method of characterizing a mechanism of action of an agent (*e.g.*, a chemotherapeutic agent, a genotoxic agent).

The method comprises contacting a plurality of populations of cells with an agent to be assessed, wherein each population of cells have one gene of interest targeted by a small hairpin RNA (shRNA) and wherein said gene of interest regulates cell death and a plurality of genes that regulate cell death are targeted in the plurality of populations of cells. A responsiveness of each population of cells to the agent is determined, thereby obtaining an shRNA signature of the agent, so as to identify one or more genes that mediate a response to the agent, thereby characterizing the mechanism of action of the agent.

In another aspect, the invention is directed an article of manufacture for characterizing a mechanism of action of a chemotherapeutic or genotoxic agent. In one aspect, the article of manufacture comprises a plurality of populations of cells, each population having an shRNA that targets a gene of interest that mediates a response to a chemotherapeutic or genotoxic agent, and an algorithm for clustering plurality of chemotherapeutic or genotoxic agents into one or more groups based on a responsiveness of each population of cells to each agent. In particular aspect, the ATM, Chk2 and p53 genes are targeted. In other aspects, the p53, ATR, Chk1, Chk2, Smg-1, DNA-PKcs, Bok and Bim genes are targeted.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1a-1d: Functional characterization of chemotherapeutic drugs according to patterns of shRNA-conferred drug resistance or sensitivity. (1a) A diagram showing the principle of GFP-based competition assays. Suppression of genes that alter drug sensitivity leads to changes in the percentage of GFP-positive cells after treatment, which can be used to calculate the RI (see Methods). (1b) Unsupervised clustering of RI values of 15 reference compounds. Agglomerative hierarchical clustering was performed on log-transformed RI values for the initial 15 reference drugs, using a correlation metric and centroid linkage. Bootstrapping data is shown to indicate clustering robustness. 'Approximately unbiased' (AU) values from the PVclust function are indicated next to the relevant branches in the clustergram. (1c) The branching pattern for SAHA, DAC and Rosco and the 15 reference chemodrugs. Numbers below the dendogram demarcate drug categories. (1d) A heat map showing the response of cells expressing shRNAs targeting the Bim

transcriptional regulator Chop and Foxo3a to SAHA and DAC. Log-transformed RI values are shown.

Figures 2a-2d: RNAi-based characterization of a compound derivative of bendamustine. (2a) The chemical structures of bendamustine and a chemical derivative, CY190602. (2b) Dose response curves comparing the viability of the multiple myeloma cell lines RPMI-8226 (top) and MM1S (bottom) following treatment with bendamustine or CY190602. (2c) RI patterns for bendamustine, CY190602 and a related compound, chlorambucil (CBL). Bendamustine and CY190602 were used at LD80-90 of 110 μM and 1.4 μM, respectively. (2d) The branching pattern for the 18 reference drugs plus bendamustine and CY190602.

Figures 3a-3e: Identification and functional characterization of ill-defined genotoxic drugs. (3a) A heat map showing the response of cells expressing shATM, shChk2 or shp53 to 16 genotoxic (upper panel) and 15 nongenotoxic (lower panel) chemotherapeutics (see Supplementary Table 2 for drug abbreviations). (3b) The shATM-Chk2-p53 response signature for apigenin (APG) and NSC3852 (NSC). (3c) The branching pattern for the 18 reference compounds plus APG and NSC. APG clusters with the TopoII poisons Dox and VP-16, whereas NSC clusters with the Topol poison CPT. (3d) A comparison of the shTopoI and shTopoII response signatures for APG and NSC3852 with response signatures derived from established Topol (CPT and CPT11) and TopoII poisons (Dox, Mito and VP-16). Although NSC3852 and APG show response patterns characteristic of Topol and TopoII poisons, respectively, none of the other genotoxic drugs showed either of these resistance and sensitivity patterns. (3e) A graph showing the number of surviving shTopoII, shTopol or vector control-expressing cells 12 days after drug treatment with APG or NSC3852. In each case, one million cells were plated before treatment. Data shown are mean ± s.e.m. from three Independent experiments.

Figures 4a-4c: A feature reduction identifies a reduced eight-shRNA set. (4a) Analysis of the dataset used for Figure 1c, using a randomized search strategy. The graph shows the relative efficacy of drug prediction as a function of increasing shRNA subset size, The maximum predictability for 2,000 iterations at each shRNA subset size is shown. (4b) A graph showing the correlation between enrichment or depletion of cells expressing shRNAs from the original eight-shRNA set and cells expressing shRNAs from the additional eight-shRNA set after drug treatment. Each square represents the $\log_2$RI values following single-drug treatment of cells

expressing an original shRNA (x axis) or the second shRNA targeting the same gene (y axis). The slope of the best fit line is 0.64, indicating that the absolute RI values are consistently lower in cells expressing hairpins from the second eight-shRNA set. (4c) A heat map showing the relative enrichment or depletion of a second set of shRNAs (labeled with asterisks) targeting each of the genes in the eight-shRNA signature, The associated dendograms show clustering between shRNA pairs, as well as clustering of small molecules into the same seven categories predicted from the 29-shRNA signature. (d) A graph showing the correlation between enrichment or depletion of cells expressing shRNAs from the original eight-shRNA set and cells expressing shRNAs from the additional eight-shRNA set after drug treatment. Each square represents the log2RI values following single-drug treatment of cells expressing an original shRNA (x axis) or the second shRNA targeting the same gene (y axis). The slope of the best-fit line is 0.64, indicating that the absolute RI values are consistently lower in cells expressing hairpins from the second eight-shRNA set.

Figures 5a-5c: A reduced shRNA signature can accurately predict drug mechanism of action. (5a) A diagram of the possible outcomes for a test compound when it is compared to the training set. A test compound could be interpolated within the definition of a drug category that is provided by the training set (left). Alternatively, a test compound could be outside of the drug category (right). Our probabilistic nearest-neighbors algorithm attempts to define an "acceptable" category extension. (5b) A schematic depicting the methodology behind probabilistic nearest-neighbors predictions. An initial training set with empirically validated drug categories is used to calculate the drug category-specific cluster sizes. This same methodology is used for compounds whose known mechanism of action is distinct from a particular drug category. (5c) The increase in the drug category definition that is observed by forcing these empirically derived negative controls to cluster in an erroneous category is used to build a null distribution and an empirical cumulative distribution function.

Figure 6: Adaptation of the eight-shRNA signature to a distinct cell line. A heat map comparing the eight-shRNA response signatures of Myc $p19^{Afr-/-}$ lymphoma cells and p185+ BCR-Abl leukemia cells following treatment with alkylating agents in a model of acute lymphoblastic leukemia. The eight-shRNA signature from p185+ BCR-Abl $p19^{Afr-/-}$ leukemia cells can identify CDDP as an alkylating agent when CBL and MMC are used as a reference set.

Figures 7a-7b: shRNAmir-mediated stable suppression of drug response genes. 7a, Western blot image showing knockdown of p53-activating kinases. Underlined lanes demarcate shRNAs used in subsequent studies. Starred lines demarcate shRNAs used in Figure 4c as additional shRNAs. 7b, QPCR data showing knockdown of Bcl2 family genes. Data represent the results from two independent experiments. Bcl2 and Mcl1 were omitted from this study, as Bcl2 and Mcl1 stable knockdown cells could not be established.

Figures 8a-8b: Significance analysis of the 18 drug and DNA damage subcategory clustering. 8a, PCA Monte Carlo analysis comparing the percent variance explained in the actual 7 category (7C) decomposition of the 18-drug set versus 7C decomposition of 1000 randomized data sets. 8b, PCA Monte Carlo analysis comparing the percent variance explained in the actual 3 category (3C) decomposition of the DNA damage set versus 3C decomposition of 1000 randomized data sets.

Figures 9a-9b: 9a, Unsupervised clustering of RI values of the 15 reference compounds, SAHA, decitabine (DAC), and roscovitine (Rosco). Agglomerative hierarchichal clustering was performed on log transformed RI values for these 18 drugs, using a correlation metric and centroid linkage. Their cluster position is underlined in red. 9b, Lymphoma cells were treated with SAHA or DAC for 6 or 9 hours. Bim expression level was analyzed by western blot.

Figure 10: The effect of TopoI and TopoII deficiency on NSC3852 and APG induced DNA damage. Vector control, TopoI and TopoII knockdown cells were treated with NSC3852 and APG, respectively. Cells were monitored by western blot for activation of a DNA damage response, as indicated by γ-H2AX phosphorylation.

Figures 11a-11c: 11a, "8-shRNA signatures" that exhibit a 100% cross-validation rate. The columns show the composition of each 8-shRNA set that cross validates at 100%. Grey boxes indicate the presence of an shRNA in a particular 8-shRNA set. 11b, A scatter plot of the correlation between the pairwise distances in the reference drug set for the original 29 shRNA set versus the reduced 8 shRNA set ($r^2=0.81$). 11c, A clustergram of 17 references drug plus APG, NSC3852, bendamustine and CY190602 using 8 shRNAs.

Figure 12: A heat map showing the 8 shRNA resistance and sensitivity patterns of all chemotherapeutics examined.

Figure 13: A diagram showing the relative correlation distances within and between clusters. The bubble size represents the intracategory average linkage distance between drugs. The line thickness represents the average intercategory distances between the drugs of distinct categories. P values show how significantly HER2 inhibitors and Sunitib cluster with EGFR inhibitors.

Figure 14: Identifying the cellular targets of small molecules remains a central challenge of chemical biology. The application of an RNAi-based functional genomics approach permitted the clustering of drugs with related targets by "shRNA signatures" which served as a basis set to assign models of action to compounds with unknown targets.

Figure 15: Illustration of a computer network or similar digital processing environment in which the present invention may be implemented.

Figure16: Diagram of the internal structure of a computer (e.g., client processor/device 50 or server computers 60) in the computer system of Fig. 15.

## DETAILED DESCRIPTION OF THE INVENTION

A description of example embodiments of the invention follows.

Identifying mechanisms of drug action remains a fundamental impediment to the development and effective use of chemotherapeutics. An RNA interference (RNAi)-based strategy to characterize small-molecule function in mammalian cells is described herein. By examining the response of cells expressing short hairpin RNAs (shRNAs) to a diverse selection of chemotherapeutics, a functional shRNA signature that accurately grouped drugs into established biochemical modes of action was generated. This, in turn, provided a diversely sampled reference set for high-resolution prediction of mechanisms of action for poorly characterized small molecules. The predictive shRNA target set was further reduced to as few as eight genes and, using a newly derived probability-based nearest-neighbors approach, the predictive power of this shRNA set was extended to characterize additional drug categories. Thus, the focused shRNA phenotypic signature described herein provided a highly sensitive and tractable approach for characterizing new anticancer drugs (see Jiang, H., et al., Nature Chemical Biology 7, 92-100 (2011) which is incorporated herein by reference).

Accordingly, the invention is directed to a method of characterizing a mechanism of action of an agent. The method comprises contacting a plurality of populations of cells with an agent to be assessed, wherein each population of cells have one gene of interest targeted by a small hairpin RNA (shRNA) and wherein said gene of interest regulates cell death and a plurality of genes that regulate cell death are targeted in the plurality of populations of cells. A responsiveness of each population of cells to the agent is determined, thereby obtaining an shRNA signature of the agent, so as to identify one or more genes that mediate a response to the agent, thereby characterizing the mechanism of action of the agent.

As will be appreciated by those of skill in the art, the mechanism of action of a variety of agents can be characterized using the methods described herein. For example, the agent can be a chemical compound, a nucleic acid, a peptide (a protein), a lipid, a sugar (*e.g.,* polysaccharide), a lipopolysaccharide and the like. In one aspect, the agent is a chemotherapeutic agent. In another aspect, the agent is a genotoxic agent. In yet other aspects, the agent is a derivative of a chemotherapeutic or genotoxic agent.

There are a variety of mechanisms of action by which agents (*e.g.,* chemotherapeutic agents) exert their effects. Examples of mechanisms of action of a chemotherapeutic agent include inhibition of a topoisomerase, cross linking of DNA, inducement of single stand break of DNA, inhibition of nucleic acid synthesis, inhibition of mitosis, inhibition of RNA transcription, inhibition of histone modification enzymes, inhibition of heat shock proteins (*e.g.,* Hsp90), alkylation of DNA, inhibition of proteasomes inducement of apoptosis or the like. The methods described herein can further comprise classifying the agent within a group of agents having in common one or more mechanisms of action.

As described herein, the method of determining a mechanism of action of an agent involves contacting a plurality of populations of cells with an agent to be assessed wherein each population of cells has one gene of interest that is not functional (*e.g.,* not expressed). In a particular aspect, the method of determining a mechanism of action of an agent involves contacting a plurality of populations of cells with an agent to be assessed wherein each population of cells have one gene of interest targeted by a small hairpin (shRNA). As is known in the art, shRNA is a ribonucleic acid (RNA) polymer that is designed based on the study of naturally-occurring hairpin RNAs involved in RNA interference (RNAi). shRNA function in

the cell is to drive the degradation of messenger RNAs (mRNAs) in a sequence-specific manner. More specifically, shRNA is a short sequence of RNA which makes a tight hairpin turn and can be used to silence gene expression via RNA interference (*e.g.,* Paddison, P., et al., Genes Dev. 16 (8): 948–958 (2002)). That is, in one aspect of the method described herein, each shRNA acts to knock down one gene.

In particular aspects, the method comprises introducing the plurality of shRNAs targeting the plurality of genes of interest into the plurality of populations of cells, wherein each shRNA targets one gene of interest that regulates cell death, wherein each population of cells have one gene of interest targeted. In other aspects, the method can comprise introducing the plurality of shRNAs which suppresses expression of the plurality of genes into the plurality of cells, wherein each shRNA suppresses expression of one gene that regulates cell death, and one gene is suppressed in each cell.

As will be appreciated by those of skill in the art there are a number of genes that regulate cell death. In a particular aspect, the gene that regulates cell death is a gene in the Bcl2 family of genes, a p53 gene, or a p53-activating kinase gene. Examples of a gene in the Bcl2 family of genes includes a Bax gene, Bak gene, a Bok gene, a Bim gene, a Bid gene, a Puma gene, a Noxa gene, a Bad gene, a Bmf gene, a Bik gene, a Hrk gene, a Bclx gene, a Bclb gene, a Bclw gene, an A1 gene, a Bclg gene, a Mil1 gene, a Mule gene, a BPR gene, a BNIP gene, a Bcld gene, a Bcl2 gene, or a Mcl 1 gene. Examples of a p53 activating kinase gene include an ATM gene, an ATR gene, a Chkl gene, a Chk2 gene, a DNAPKcs gene, a Smg-1 gene, a JNK1 gene or a p38 gene.

In the methods of the invention a plurality of genes that regulate cell death are targeted in the plurality of cell populations. In particular aspects, three genes, four genes, five genes, six genes, seven genes, eight genes, nine genes, ten genes, eleven genes, twelve genes, thirteen genes, fourteen genes, fifteen genes, sixteen genes, seventeen genes, eighteen genes, nineteen genes, twenty genes, twenty one genes, twenty two genes, twenty three genes, twenty four genes, twenty five genes, twenty six genes, twenty seven genes, twenty eight genes, twenty nine genes, thirty genes or more are targeted by the corresponding shRNAs. As will be appreciated by those of skill in the art, libraries encompassing hundreds and thousands of such genes can be used in the methods described herein.

The particular genes chosen for targeting can thereby provide a particular shRNA signature of the agent when assessed using the methods provided herein. For example, in a particular aspect, the plurality of genes targeted by the corresponding shRNAs are ATM, Chk2 and p53 genes, thereby allowing one to characterize the mechanism of action of an agent as a *shATM-Chk2-p53* 'resistance signature'. In another aspect, the plurality of genes targeted by the corresponding shRNAs are p53, ATR, Chk1, Chk2, Smg-l, DNA-PKcs, Bok and Bim genes, thereby allowing one to characterize the mechanism of action of an agent as a *shp53, ATR, Chk1, Chk2, Smg-l,,DNA-PKcs,Bok, Bim* 'resistance signature'. As will be appreciated by those of skill in the art, other resistant signatures can be determined as described herein.

In the methods of the invention, the shRNAs can be introduced into the cells using a variety of methods. For example, as described herein a viral vector is used. Numerous viral vectors that can be used in the methods are known to those of skill in the art. Specific examples include a retroviral vector, an adenoviral vector and the like.

As will be appreciated by those of skill in the art, the vector can include other components. In a particular aspect, the viral vector further expresses a marker gene. Any variety of marker genes can be incorporated into the viral vector. In one aspect, the marker gene is a fluorescent marker gene. In a particular aspect, the marker gene is green fluorescent protein (GFP) gene.

Marker genes and the expression thereof can be measured in the cell populations using a variety of techniques known in the art. Thus, the methods described herein can further comprise measuring the marker gene (*e.g.,* a fluorescent marker gene or GFP gene) expression level in each population of cells. In one aspect, flow cytometry is used to measure the marker gene or expression thereof.

As described herein a responsiveness of each population of cells to the agent is determined, thereby obtaining the shRNA signature of the agent so as to identify one or more genes that mediate a response to the agent. Examples of a type of responsiveness that can be determined include resistance or sensitivity to the agent. In one aspect, the responsiveness of each population of cells to the agent is a relative level of chemo-resistance and sensitization conferred by each shRNA. In a particular aspect, the responsiveness is a relative survival rate of each population of cells compared to control cells that do not contain said shRNA targeting the gene of interest.

The determination of responsiveness can be determined using a variety of methods. In one aspect, the determination of the responsiveness is accomplished using cell flow cytometry, hybridization techniques or sequencing techniques.

In the methods of the invention the plurality of populations of cells can be contacted with the chemotherapeutic agent for any suitable amount of time. In some aspects, the plurality of populations of cells are contacted once with the agent. In other aspects, the plurality of populations of cells are contacted repeatedly (more than once) with the agent. In addition, the plurality of populations of cells can be contacted with the agent for about 1 hour, 4 hours, 8 hours, 12 hours, 16 hours, 20 hours, 24 hours, 28 hours, 32 hours, 36 hours, 40 hours, 44 hours, 48 hours, 52 hours, 56 hours, 60 hours, 64 hours, 68 hours, 72 hours, 76 hours, 80 hours, 84 hours, 88 hours, 92 hours, 96 hours, 100 hours or longer.

In the methods described herein, the amount of agent that is contacted with the plurality of populations of cells will vary and will deoend on a variety of factors (*e.g.,* the type of agent being assessed; the type of response being sought, etc). For example, the amount (*e.g.,* concentration) of agent that is contacted with the populations of cells can be based on an agent's lethal dose (LD), if known. The LD of the agent that can be used in the methods includes the lethal dose that is sufficient to kill 50% of a cell population (LD50), 60% of a cell population (LD60), 70% of a cell population (LD70), 80% of a cell population (LD80), 90% of a cell population (LD90), or 100% of a cell population (LD100). In particular aspects, the agent is used in an effective amount to induce a response in cells that do not contain said shRNA targeting said gene of interest.

Any of a variety of cells can be used in the methods of the invention. In one aspect, the cells are mammalian cells. Examples of mammalian cells include primate cells (*e.g.,* human cells), murine cells (*e.g.,* mouse cells, rat cells), feline cells, canine cells, bovine cells and the like. In a particular aspect, the cells are from a pathological or diseased source. For example, the cells can be tumor cells. Examples of tumor cells include lymphoma cells, acute lymphocytic leukemia cells and the like.

As will be appreciated by those of skill in the art, the methods described herein can further comprise comparing the responsiveness of each population of cells to the agent to a control. As will be apparent to those of skill in the art, a

variety of suitable controls can be used. In one aspect, the control is a population of cells into which the shRNA targeting the gene of interest has not been introduced.

As will also be appreciated by those of skill in the art, the methods of the invention can be performed in vitro, as described herein. Alternatively, or additionally, the methods described herein can be performed in vivo. An example of an in vivo method involves the use of a pooled shRNA format. In this aspect, shRNAs are pooled and transduced into a target cell population and the population is then engrafted into a recipient non-human mammal such as a rodent (*e.g.,* a mouse or a rat). A pretreatment baseline is established by sequencing or hybridization. The non-human mammals are dosed with drugs and following treatment, reassessed for the shRNA pool composition.

As shown herein, the methods of the invention can also be automated. In one aspect, the methods can further comprise using an algorithm to cluster a plurality of agents into groups based on the responsiveness of each population of cells to each agent.

Figure 15 illustrates a computer network or similar digital processing environment in which the present invention may be implemented. For example, a computer(s)/devices 50 and server computer(s) 60 provide processing, storage, and input/output devices executing application programs and the like. The computer(s)/devices 50 can also be linked through communications network 70 to other computing devices, including other client devices/processes 50 and server computer(s) 60. A communications network 70 can be part of a remote access network, a global network (e.g., the Internet), a worldwide collection of computers, Local area or Wide area networks, and gateways that currently use respective protocols (TCP/IP, Bluetooth, etc.) to communicate with one another. Other electronic device/computer network architectures are suitable.

Fig. 16 is a diagram of the internal structure of a computer (e.g., processor/device 50 or server computers 60) in the computer system of Fig. 15. Each computer 50, 60 contains system bus 79, where a bus is a set of hardware lines used for data transfer among the components of a computer or processing system. Bus 79 is essentially a shared conduit that connects different elements of a computer system (e.g., processor, disk storage, memory, input/output ports, network ports, etc.) that enables the transfer of information between the elements. Attached to system bus 79 is I/O device interface 82 for connecting various input and output

devices (e.g., keyboard, mouse, displays, printers, speakers, etc.) to the computer 50, 60. Network interface 86 allows the computer to connect to various other devices attached to a network (e.g., network 70 of Fig. 4). Memory 90 provides volatile storage for computer software instructions 92 and data 94 used to implement an embodiment of the present invention. Disk storage 95 provides non-volatile storage for computer software instructions 92 and data 94 used to implement an embodiment of the present invention. Central processor unit 84 is also attached to system bus 79 and provides for the execution of computer instructions.

In one embodiment, the processor routines 92 and data 94 are a computer program product (generally referenced 92), including a computer readable medium (e.g., a removable storage medium such as one or more DVD-ROM's, CD-ROM's, diskettes, tapes, etc.) that provides at least a portion of the software instructions for the invention system. Computer program product 92 can be installed by any suitable software installation procedure, as is well known in the art. In another embodiment, at least a portion of the software instructions may also be downloaded over a cable, communication and/or wireless connection. In other embodiments, the invention programs are a computer program propagated signal product 107 embodied on a propagated signal on a propagation medium (e.g., a radio wave, an infrared wave, a laser wave, a sound wave, or an electrical wave propagated over a global network such as the Internet, or other network(s)). Such carrier medium or signals provide at least a portion of the software instructions for the present invention routines/program 92.

In alternate embodiments, the propagated signal is an analog carrier wave or digital signal carried on the propagated medium. For example, the propagated signal may be a digitized signal propagated over a global network (e.g., the Internet), a telecommunications network, or other network. In one embodiment, the propagated signal is a signal that is transmitted over the propagation medium over a period of time, such as the instructions for a software application sent in packets over a network over a period of milliseconds, seconds, minutes, or longer. In another embodiment, the computer readable medium of computer program product 92 is a propagation medium that the computer system 50 may receive and read, such as by receiving the propagation medium and identifying a propagated signal embodied in the propagation medium, as described above for computer program propagated signal product.

Generally speaking, the term "carrier medium" or transient carrier encompasses the foregoing transient signals, propagated signals, propagated medium, storage medium and the like.

The invention is also directed an article of manufacture for characterizing a mechanism of action of a chemotherapeutic or genotoxic agent. In one aspect, the article of manufacture comprises a plurality of populations of cells, each population having an shRNA that targets a gene of interest that mediates a response to a chemotherapeutic or genotoxic agent, and an algorithm for clustering plurality of chemotherapeutic or genotoxic agents into one or more groups based on a responsiveness of each population of cells to each agent. In particular aspect, the ATM, Chk2 and p53 genes are targeted. In other aspects, the p53, ATR, Chk1, Chk2, Smg-1, DNA-PKcs, Bok and Bim genes are targeted.

As will be appreciated by those of skill in the art, the article of manufacture can be used, for example, to screen a library of agents for an agent having a chemotherapeutic or genotoxic effect.

Exemplification

A Mammalian Functional-Genetics Approach to Characterizing Cancer Therapeutics

METHODS

Cell lines and drugs

Eμ-Mycp19$^{Arf-/-}$ mouse lymphoma cells were cultured in B cell medium as described (Schmitt, C.A., *et al., Genes Dev, 13:*2670-2677 (1999)). MM1S and RPMIB226 cells were cultured in RPMI medium supplemented with glutamate and 10% (v/v) FBS. Drugs were obtained from Sigma, Tocris, Calbiochem, VWR, LC Laboratories and other suppliers. shRNA vectors were generated as described (Dickins, R.A. et al. Nat. Genet, 37, 1289-1295 (2005); Jiang, H. et al. Genes Dev. 23, 1895-1909 (2009)). P185+p19$^{Arf-/-}$ acute lymphoblastic leukemia cells were derived and cultured according to the procedures outlined in ref. 35.

Drug treatment and flow cytometry

Eμ-Mycp19$^{Arf-/-}$ cells were counted and seeded at 1 million cells per ml in 48-well plates and treated with various concentrations of drugs. To approximate therapeutic situations in which drug dose decreases over time, half of the volume

from each experiment was removed and replenished with fresh medium every 24 h. Cells were analyzed by fluorescence-activated cell sorting (FACS), with propidium iodide as a viability marker. LD80-90 of drugs are defined as concentrations at which the lowest viability reading out of three FACS time points (24,48 and 72 h) is between 10% and 20%. After drug dose was determined, Eμ-Mycp19$^{Arf-/-}$ cells were infected with retroviruses encoding shRNAs targeting particular genes. Individual infected cell populations were counted and seeded at 1 million cells per ml in 48-well plates and treated with drugs using the aforementioned protocol. At 72 h, treated and untreated cells were analyzed by flow cytometry. GFP percentages of live (PI-negative) cells were recorded and used to calculate relative resistance index. To avoid outgrowth of untreated control cells, they were typically seeded them at 0.25 million per ml, and 75% of medium was replaced at 24 and 48 h.

Calculation of relative resistance index (X)

To compare the relative level of chemoresistance and sensitization conferred by each gene knockdown, the concept of RI (see definition above) was introduced, to more accurately analyze the GFP competition results. The value of RI is defined as X. The biological meaning of this factor X is that in a mixture of uninfected and infected (knockdown) cells, the infected (knockdown) cells will be X-fold as likely to survive drug treatment when compared to uninfected cells. By this definition of X, if one out of *n* uninfected cells survives a drug treatment, then X our of *n* infected cells should survive. If the total number of uninfected and infected cells are defined as T and the GFP percentage of untreated population are defined as G1, then the number of surviving, uninfected cells (un) can be defined as *n* - un = T x (1 - Gl) x l/*n*, and the number of surviving, infected cells (in) can be defined as *n* - in = Tx Gl x X/*n*. Hence, the GFP percentage of the treated, surviving population (G2) can be calculated as G2 = (*n* - in)/((*n* - un) +(*n* - in)). From this equation, it can be derived that X = (G2 - Gl x G2)/(Gl- G1 x G2). This equation was used in the studies to compute RI values.

Enhanced K-nearest-neighbors methods

K-nearest-neighbors modeling is a weighted-voting methodology in which the proximity to the training set is used to predict drug class membership. This analysis is included for four reasons. (i) It provides independent validation of the

clustering result. (ii) It allows quantification of the predictive power of the reference set through leave-one-out-cross-validation. (iii) Leave-one-out cross-validation allows performance of a feature reduction to discover smaller gene sets. (iv) It provides an objective prediction of classes for new compounds.

K-nearest-neighbors predictions were performed using a correlation-based metric and a consensus voting scheme. The MATLAB knnclassify.m function was used as a basis for the feature reduction search, as well as cross-validation and predictions. The cross-validation for the K-nearest-neighbors approach was done by systematically leaving out one of the 18 drugs at a time in the final dataset (Fig. 1c) and using the remaining 17 to predict the left-out drugs' identities.

To reduce the size of the feature set to a smaller group of key shRNAs, a subset of 2,000 unique shRNA sets of increasing size were randomly searched. Sampled subsets were scored on the basis of their ability to cross-validate. As much more extensive search (> 50,000 subsets) of eight shRNA signatures that would be able to correctly classify all of the drugs in the reference set was then performed. The shRNA subsets that cross-validated at 100% were then ranked by their least-squares correlation with the distances between drugs in the 29-shRNA signature, and the eight-shRNA set with the highest correlation score was chosen for later experiments.

A K-nearest-neighbors-based approach will always yield a prediction of drug class on the basis of proximity. Therefore, to evaluate the similarity of a new drug to its predicted class a linkage ratio *p*-value test was developed. Briefly, the initial cluster size of each of the seven drug groups was calculated (Fig. 1c) by evaluating the average of all pairwise linkage distances amongst all members of a drug group. When a test compound was predicted to belong to a drug group on the basis of proximity, then the cluster size of that particular drug group was calculated again with the new test drug included. A linkage ratio was then calculated by comparing the cluster size with and without the tested compound. A linkage ratio of less than one indicated that the addition of the drug to a cluster made the average distance between drugs in that category smaller, whereas a linkage ratio greater than one indicated that the cluster expanded. An obvious tradeoff exists between cluster expansion to accommodate modestly distinct compounds with highly homologous mechanisms and expanding the definition to a point where one masks the existence of a completely new compound. This tradeoff varies among drug classes as a

function of the inter-class distances. To estimate the significance of a K-nearest-neighbors prediction, as well as to determine whether a compound had a mechanism of action different from those of the original seven drug groups, the negative control distributions of drug classifications were sampled. This was done on a class-by-class basis by taking the previously studied compounds and forcing them to erroneously classify. A linkage ratio for all of these erroneous classifications was then calculated. On a class-by-class basis a normal distribution was fit to the range of misclassified linkage ratios. The value of the cumulative distribution function was used to calculate the *p*-value of the new classifications (Fig. 5c), using the null hypothesis that the linkage ratio for a prediction is identical to the linkage ratios of the negative control distribution. The complete MATLAB algorithm used to perform this analysis is provided as the "Drug Prediction Score.M" file found at web.mit.edu/icbp/data/index.htmI.

Clustering: Agglomerative Hierarchical Clustering was performed in Matlab v7.0.

All RI values were Log2 transformed to represent depletion and enrichment data on the same scale. To measure the distance between clusters, an inverse correlation based metric was used : After all drug pairwise distances were calculated, centroid linkage was used to compute the distance between cluster groups. Several forms of significance calculations were performed. To estimate the overall number of significant underlying drug groups in the data set, the number of latent variables that could explain the majority of the variance in the data set via a principal components analysis was analyzed. However, Random Matrix Theory for small data sets suggests that small noisy data sets may have large eigenvalues based upon chance. Therefore, in order to estimate the significance of the categorization of underlying drugs, a Monte Carlo analysis on our dataset was performed. Briefly, 1000 data matrices from our drug-gene data were sampled. Then the distribution of the cumulative variance explained by our 7 -component model relative to randomized matrices was plotted. This Monte Carlo analysis estimated the significance of the number of components that one uses to interpret the PCA model.

In an idealized scenario where the distances between and within drug clusters are similar across drug types, a uniform cutoff at a single branch length should guide interpretation of the clusters. However, in stratified datasets described herein, where considerable variation exists within and between clusters, a more

stratified approach becomes appropriate. The DNA damage drug set contained extraordinarily close correlations between distinct drugs relative to the rest of the dataset. To determine whether sub-categories of drugs within this cluster could be confirmed, the PCA sampling approach was extended to this subset of data. Utilizing this stratified approach to cluster interpretation, a hypothesis of three distinct DNA damage sub-clusters was supported. This variegated approach to cluster interpretation was also evaluated by doing Bootstrapping analysis in R using the PVClust function (Suzuki, R & Shimodaira, H., Bioinformatics, 22, 1540-1542 (2006)). This approach was used to complement the PCA data. The PCA data indicate how many significant underlying drug variables one can interpret from the data, and the bootstrapping indicated whether particular branches were significant.

```
function [class, normpvalue, predlr] = DrugPredictionScore(training, test, groups,
numbern)
% This function takes as an argument a (training) set that has
% observations(drugs)as rows and variables as columns, (test) is a test vector to
classify using
% training and in this version is limited to n=1. (groups) is a column vector
% of group memberships(that are sequential and numeric starting at n=1).  Note if you
have 7 groups they must be ordered 1-7 for
% the observations in (training). Numbern is the
% number of nearest neighbors to include, the default is majority rule.  The class is
the predicted group
% mebership in test, and pvalue is the predicted significance of the
% inclusion of the test compound in the predicted class
% Currently a KNN classifier is hardcoded into the algorithm, but the
% algorithm can be used with any matlab multi-class classifier if the code
% is changed.  In our shRNA dataset we have noted that Naive-Bayes
% Classifiers perform similarly well.  In either case a euclidean metric of
% cluster size is used to assess significance.

%Step1 creating the reference distances and null distributions
%_____________________________________________________________________
%_____________________________________________________________________
DistanceMatrix=squareform(pdist(training,'euclidean'));
M=max(groups)
i=1
nulldist={}
ALtot=[]
while i< M+1
    j=1%makes an index that terminates when the max group number is reached
    AL=[]
    gpos=[]
    nongpos=[]
    while j< length(groups)+1  %looks through the groups vector sequentially to
determine which entries match the group label i
        if groups(j,1)==i %sequentially checks the equality
            gpos=horzcat(gpos,j)%horizontally concatenates the reference row for a given
group.
        elseif groups(j,1)~=i
            nongpos=horzcat(nongpos,j)
        end
        j=j+1
    end
    S=length(gpos)
    Sn=length(nongpos)
    submati=[DistanceMatrix(gpos(1,1):gpos(1,S),gpos(1,1):gpos(1,S))]%builds submatrix
of group specific distances and then calculates the averages on the basis of this
    AL=sum(sum(submati))/(S*(S-1))
    ALtot=horzcat(ALtot,AL)
    k=1
    newalvec=[]
    while k < Sn+1 %this step build null distributions based on the empircal prediction
using false negatives
        alldistances=DistanceMatrix(nongpos(1,k),:)
        l=1
        nullpredictionsdist=[]
        while l < S+1
```

```
            nullpredictionsdist=vertcat(nullpredictionsdist,alldistances(1,gpos(1,l)))
            l=l+1
        end
        newal=(sum(sum(submati))+sum(nullpredictionsdist))/((S*(S-1))+S)
        newalvec=vertcat(newalvec,newal)
        k=k+1
    end
    storednull{1,i}=newalvec
    i=i+1
end

%_____________________________________________________________________
%step2 is the prediction of drug class and updating a new training set to
%take into account that prediction, and then calculating the average
%linkage following the inclusion of the test drug
%_____________________________________________________________________
training2=training
test2=test
groups2=groups
class=knnclassify(test, training, groups, numbern, 'euclidean')
m=1
while m < length(class)+1
    n=1
    while n < length(groups)+1
        if groups(n,1)==class(m,1)
            if n<length(groups)
                trainingnew=vertcat(training2(1:n,:),test2(m,:),training2(n+1,:))
                groupsnew=vertcat(groups2(1:n,:),class(m,:),groups2(n+1,:))
            elseif n==length(groups)
                trainingnew=vertcat(training2(1:n,:),test2(m,:))
                groupsnew=vertcat(groups2(1:n,:),class(m,:))
            end
        end
        n=n+1
    end
    m=m+1
end
DistanceMatrixnew=squareform(pdist(trainingnew,'euclidean'))
Mnew=max(groupsnew)
i=1
ALnew=[]
while i< Mnew+1
    j=1%makes an index that terminates when the max group number is reached
    AL2=[]
    gposnew=[]
    nongposnew=[]
    while j< length(groupsnew)+1  %looks through the groups vector sequentially to
determine which entries match the group label i
        if groupsnew(j,1)==i %sequentially checks the equality
            gposnew=horzcat(gposnew,j)%horizontally concatenates the reference row for a
given group.
        elseif groupsnew(j,1)~=i
            nongposnew=horzcat(nongposnew,j)
        end
        j=j+1
    end
    Snew=length(gposnew)
```

```
    Snnew=length(nongposnew)

submatinew=[DistanceMatrixnew(gposnew(1,1):gposnew(1,Snew),gposnew(1,1):gposnew(1,Snew)
]%builds submatrix of group specific distances and then calculates the averages on the
basis of this
    AL2=sum(sum(submatinew))/(Snew*(Snew-1))
    ALnew=horzcat(ALnew,AL2)
    i=i+1
end

%_____________________________________________________________________________
%_Step3 is comparing to the background distribution
%1st calculate for the group predicted in class, what is the distribution of
%linkage ratios for the given group, calculate from altot(yourclass) and
%the classes entry in storednull what the distribution of increases for
%null(false) predictions are.  Then calculate the linkage ratio alnew/altot
%then get pvalue
%_____________________________________________________________________________
lrdist=[]
p=1
while p < length(storednull{1,class})+1
    lrdist=vertcat(lrdist,(storednull{1,class}(p))/ALtot(1,class))
    p=p+1
end
predlr=ALnew(1,class)/(ALtot(1,class))
lr=lrdist
lr = lr(:);

% Set up figure to receive datasets and fits
f_ = clf;
figure(f_);
set(f_,'Units','Pixels','Position',[547.333 181 680 475.45]);
legh_ = []; legt_ = {};   % handles and text for legend
ax_ = newplot;
set(ax_,'Box','on');
hold on;

% --- Plot data originally in dataset "lr data"
t_ = ~isnan(lr);
Data_ = lr(t_);
[Y_,X_] = ecdf(Data_,'Function','cdf'...
               );  % compute empirical function
h_ = stairs(X_,Y_);
set(h_,'Color',[0.333333 0 0.666667],'LineStyle','-', 'LineWidth',1);
xlabel('Data');
ylabel('Cumulative probability')
legh_(end+1) = h_;
legt_{end+1} = 'lr data';

% Nudge axis limits beyond data limits
xlim_ = get(ax_,'XLim');
if all(isfinite(xlim_))
   xlim_ = xlim_ + [-1 1] * 0.01 * diff(xlim_);
   set(ax_,'XLim',xlim_)
end

x_ = linspace(xlim_(1),xlim_(2),100);
```

```
%_____________________________________

% --- Create fit "fit 1"
%_____________________________________

% Fit this distribution to get parameter values
t_ = ~isnan(lr);
Data_ = lr(t_);
% To use parameter estimates from the original fit:
%     p_ = [ 5.5, 3.027650354097];
pargs_ = cell(1,2);
[pargs_{:}] = normfit(Data_, 0.05);
p_ = [pargs_{:}];
y_ = normcdf(x_,p_(1), p_(2)); % compute cdf
h_ = plot(x_,y_,'Color',[1 0 0],...
          'LineStyle','-', 'LineWidth',2,...
          'Marker','none', 'MarkerSize',6);
legh_(end+1) = h_;
legt_{end+1} = 'fit 1';

hold off;
leginfo_ = {'Orientation', 'vertical', 'Location', 'NorthWest'};
h_ = legend(ax_,legh_,legt_,leginfo_{:});  % create legend
set(h_,'Interpreter','none');
normpvalue=normcdf(predlr,p_(1), p_(2))
%this works
```

Comparison of shRNA's to miRNA's

Local sequence alignments were performed in matlab using the localalign.m function. Briefly, each shRNA in the 8 shRNA signature was pairwise aligned to every miRNA in the Mus musculus genome.

RESULTS

Clustering drugs via shRNA-mediated phenotypes

As described herein, it was hypothesized that RNAi-mediated suppression of cell death regulators in mammalian cells would uniquely affect the cellular response to certain types of drugs and that drugs with similar mechanisms of action would elicit similar shRNA-dependent responses. To test this strategy, a cell line derived from tumors from a well-established mouse model of Burkitt's lymphoma was used (Adams, J.M. et al. Nature 318, 533-538 (1985); Schmitt, C.A., et al., Genes Dev.13, 2670-2677 (1999)). This cell line was chosen as an experimental system for two reasons. First, these cells are highly sensitive to a diverse set of chemotherapeutics, allowing small molecules to be used at pharmacologically relevant doses. Second, like many high-grade lymphomas, these cells undergo rapid apoptosis, as opposed to prolonged cell cycle arrest, following treatment. This common biological outcome after treatment allows for a systematic comparison of drugs.

In determining which genes to knock down for the studies, two classes of genes known to be critical for cell fate decisions after drug treatment were chosen. The Bc12 family of genes includes both central mediators and inhibitors of cell death, and different members of this gene family are involved in the response to distinct cell death stimuli (Schmitt, C.A., et al., Genes Dev.13, 2670-2677 (1999); Youle, R.I. & Strasser, Nat. Rev. Mol. Cell Bioi. 9, 47-59 (2008)). The transcription factor p53 functions upstream of components of the Bcl2 family and is another important cell death regulator (Lu, C. & EI-Deiry, W.S. Apoptosis 14, 597-606 (2009)). Mutation or deletion of p53 has been shown to affect the cellular response to many types of chemotherapeutic drugs (Lowe, S.W. et al., Cell 74, 957-967 (1993); Lowe, S.W. et al. Science 266, 807-810 (1994)). As the stabilization and activity of p53 is strongly regulated by phosphorylation, a panel of p53-activating kinases, including ATM, ATR, Chkl, Chk2, DNAPKcs, Smg-l, JNKl and p38 was also targeted (Bode, A.M. & Dong. Z. Nat, Rev. Cancer 4, 793-805 (2004); Brumbaugh, K.M. et al. Mol. Cell 14, 585-59B (2004)). Importantly, aside from

their roles as regulators of p53, these kinases are also involved in additional cellular responses to chemotherapy, such as DNA replication and repair, the activation' of cell cycle checkpoints, regulation of RNA stability and stress signaling (Lavin, M.P. Nat. Rev. Mol. Cell Bioi. 9, 759-769 (2008); Cimprich, KA. & Cortez, D, Nat. Rev. Mol. Cell Biol. 9, 616-627 (2008); Bartek, J. & Lukas, J. Cancer Cell 3, 421-429 (2003); Reinhardt, H.C , et al., Cancer Cell 11, 175-189 (2007); Pearce, A.I. & Humphrey, T.C. Trends Cell Biol. 11, 426-433 (2001)). Thus, shRNA vectors targeting the Bc12 family, p53 and its activating kinases were generated (Supplementary Results, Figs. 7a-7b and Table 2).

To enable a quick and accurate analysis of how the suppression of a given gene affects drug-induced cell death, a single-cell flow cytometry-based GFP competition assay was used. Lymphoma cells were infected with retroviruses coexpressing a given shRNA and green fluorescent protein (GFP) and subjected to 72 h of drug treatment (Fig. la). In this assay, GFP-negative cells in the same population serve as an internal control. Using this approach, how suppression of individual genes affected drug-induced cell death was systematically investigated. As an initial proof of principle, 15 chemotherapeutics representing major categories of anticancer drugs in clinical use were chosen. To compare different drugs using an objective criterion, all drugs were used at their LD80-90-a concentration at which 80-90% of uninfected lymphoma cells were killed (Table 3). A control retrovirus lacking an shRNA or retroviruses expressing shRNAs targeting 29 genes were individually used to infect lymphoma cells. Each infected population was separately treated with 15 chemotherapeutic drugs, and the effect of a particular gene knockdown on therapeutic response was compiled as values of the GFP-determined 'resistance index' (RI) (Fig. 1b). Drugs with similar mechanism of action were expected to have similar patterns of genetic dependence on these 29 genes, which would manifest as similar patterns of RI values. To test this hypothesis in an unbiased manner, an unsupervised agglomerative hierarchical clustering approach was used to compare the RI values of different drugs (Fig. 1b). The significance of this hypothesis was then evaluated using a Monte Carlo principal components analysis-based method (Pritchard, J.R. et al. Mol. Cancer Ther. 8, 2183-2192 (2009)). Notably, all 15 drugs tested in this initial experiment formed six distinct clusters that were consistent with their molecular mechanisms of action (Figs. 8a-8b). Specifically, clear groupings were seen between topoisomerase II (TopoII)

poisons doxorubicin (Dox) and etoposide (VP-16), DNA cross-linking agents cisplatin (CDDP), mitomycin C (MMC) and chlorambucil (CBL), single-strand break (SSB)-inducing agents camptothecin (CPT), 6-thioguanine (6-TG) and temozolomide (TMZ) (Swann, P.P. et al. Science 273, 1109-1111 (1996); Mojas, N., et al., Genes Dev. 21, 3342-3355 (2007)), nucleic acid synthesis inhibitors methotrexate (MTX), 5-flurouracil (5-FU) and hydroxyurea (HU), and spindle poisons vincristine (VCR) and paclitaxel (Taxol). Taken together, these data showed that a simple comparison of drug response in cells expressing a small set of shRNAs could effectively categorize established chemotherapeutic drugs into subgroups that demarcate common target proteins and pathways.

To investigate whether this platform could be used to characterize mechanisms of drug action, several recently developed chemotherapeutics were examined: suberoylanilide hydroxamic acid (SARA), decitabine and roscovitine. Although the immediate biochemical targets of these new chemotherapeutics are known, the mechanisms of cell death induced by these drugs are less well defined. Using the RNAi-based approach, RI values for each of these three drugs were compiled and compared with the 15 reference drugs mentioned earlier. It ws observed that the CDK inhibitor roscovitine (Rosco) was most similar to the RNA polyinerase inhibitor actinomycin D (ActD) (Fig. 1c and Figs. 9a). This is consistent with the findings of several studies showing that roscovitine inhibits CDK7, a component of the general transcription factor TFIIH, to inhibit RNA transcription (Akhtar, M.S. et al. Mol Cell 34, 387-393 (2009); Ljungman, M. & Paulsen, M.T. Mol Pharmacal. 60, 785-789 (2001); MacCallum, D.E. et al. Cancer Res. 65, 5399-5407 (2005)). Notably, the HDAC inhibitor SAHA and the DNA methyltransferase inhibitor decitabine (DAC) formed a distinct cluster outside of the 15 reference drugs (Fig. lc), indicating that these two drugs may share a similar mechanism of cell death. To extract the most relevant genes for distinguishing the SABA-DAC cluster, shRNAs were ranked upon their ability to classify this cluster relative to the rest of the dataset. The most unique aspects of the new SAHA-DAC cell death signature were the (i) p53-independence ($\log_2$RI≈0) and (ii) Bim-dependence ($\log_2$RI≈2) of cell death, consistent with previous studies of SABA treatment in mouse lymphoma models (Lindemann, R.K et al. Proc. Natl. Acad. Sci. USA 104, 8071-8076 (2007)). Indeed, both SABA and DAC treatment resulted in an increase in the levels of the proapoptotic BH3-only protein Bim (Supplementary Fig. 3b).

Furthermore, suppression of the Bim transcription regulator Chop, but not Foxo3a, resulted in resistance to both SAHA and DAC (Fig. 1d). Thus, the RI patterns of these newly established drugs could effectively identify their mechanism of action.

Functional characterization of derivatized compounds

A significant challenge in drug development is determining whether lead compound derivatives with enhanced efficacy share the same mechanism of action as the original small molecule. Theoretically, derivatized compounds could show enhanced efficacy, owing to either the activation of additional cell death pathways or, alternatively, through altered pharmacodynamic properties. To examine whether our approach could be used to differentiate between these possibilities, an shRNA-based functional analysis of CY190602, a chemical derivative of the nitrogen mustard bendamustine was performed (Fig. 2a). Compared to the parental drug, CY190602 showed approximately 20-100-fold enhanced toxicity toward cells from patients with multiple myeloma (Fig. 2b), an indication for which bendamustine is currently in clinical use. However, the mechanism underlying this increase in cytotoxicity remains unclear. Notably, CY190602's modification on bendamustine occurs on a side chain well away from the nitrogen mustard functional group. To address whether CY190602's toxicity could still be attributed to the nitrogen mustard or whether it was a result of altered target specificity caused by the side chain modification moieties, the RI values of bendamustine and CY190602 were compiled and compared to those of the 18 reference drugs. Notably, bendamustine and CY190602 showed highly similar patterns of RI values (Fig. 2c), despite a 100-fold-lower dose of CY190602. Additionally, both drugs clustered together with chlorambucil, another nitrogen mustard (Fig. 2d), and a supervised K-nearest-neighbors approach (see Methods for a detailed rationale) predicted a chlorambucil-like mechanism for both drugs. This indicated that the primary mode of action of CY190602 is nitrogen mustard-mediated DNA damage rather than an off-target effect conferred during drug optimization.

Screening for compounds on the basis of shRNA signatures

Next, it was asked whether this approach could be adapted to phenotype-based screens for new drug candidates without well-established mechanisms of action. Suppression of ATM, Chk2 and p53 all led to significant resistance to

genotoxic drugs such as Dox, VP-16, CPT, TMZ, 6TG, CDDP, MMC and CBL (Fig. 1b). This indicated that the *shATM-Chk2-p53* 'resistance signature' might be used to identify genotoxic drugs. To test this hypothesis quantitatively, whether a supervised K-nearest-neighbors approach could accurately characterize all of the drugs in the dataset as either genotoxic or nongenotoxic was examined. Indeed, when a broad panel of chemotherapeutic drugs was tested, all 16 genotoxic chemotherapeutics, but none of 15 nongenotoxic chemotherapeutics, showed a distinct *shATM-Chk2-p53* resistance signature (Fig. 3a). This three-gene resistance signature was subsequently used to screen a chemical library for genotoxic compounds. Two compounds, apigenin and NSC38S2, were identified on the basis of their strong *shATM-Chk2-p53* resistance signature (Fig. 3b). The full 29-gene Rl values for these two compounds were compiled and compared with reference drugs (Fig. 3c). Notably, the K-nearest-neighbors approach predicted apigenin to be most similar to the TopoII poisons doxorubicin and etoposide and NSC3852 to be most like the SSB-inducing agents. Subsequent clustering showed NSC38S2 to be most similar to the topoisomerase I (Topol) poison camptothecin. Previous studies demonstrated that TopoII poisons are ineffective in killing TopoII-deficient cells, while showing enhanced toxicity for cells lacking Topol (Burgess, D., et al. Proc. Natl. Acad. Sci. USA 105, 9053-9058 (2008)). Consistent with the clustering-based functional predictions, apigenin showed a pattern of shTopoII resistance and shTopoI sensitivity similar to the established TopoII poisons doxorubicin, etoposide and mitoxantrone (Fig. 3d). Conversely, NSC3852 showed a characteristic pattern of resistance, similar to established TopoI poisons camptothecin and irinotecan (CPTll). Notably, none of the other genotoxic drugs showed these resistance and sensitivity patterns with shTopoI and shTopoII (Fig. 3d). It was also found that apigenin and NSC3852 failed to induce DNA damage in TopoII- and TopoI-deficient cells, respectively (Fig. 10). Moreover, in a long-term survival assay, TopoII deficiency resulted in significant protection from apigenin, whereas TopoI deficiency significantly protected cells from NSC3852 (Fig. 3e). Taken together, these assays confirmed the classification of apigenin and NSC3852 as TopoII and TopoI poisons, respectively. Thus, small shRNA signatures can be used to screen chemical libraries to identify and characterize new compounds with particular target specificities.

An eight-shRNA set for accurate drug mechanism predication

Given that a three-gene signature could effectively predict and classify genotoxic drugs, it was hypothesized that the combined resistance and sensitivity pattern of a small number of genes may be sufficient to accurately characterize most of our chemotherapeutic drugs in this cell line. To test this hypothesis, the seven drug clusters demarcated in our secondary analysis was examined (Fig. 1c) and it was asked which smaller sets of shRNAs could similarly define these groupings. Here, a K-nearest-neighbors cross-validation-based approach and a randomized search through 50,000 potential gene subsets was used. Although most smaller shRNA sets showed a significant loss in resolution relative to the reference set, it was found that a set of eight shRNAs, targeting p53, ATR, Chk1, Chk2, Smg-1, DNAPKcs, Bok and Bim, was able to classify the reference dataset with 100% accuracy and was highly correlated ($r^2 = 0.81$) with the original 29 shRNA signature (Fig. 4a and Fig. 11a, 11b). Although several other sets of eight shRNAs could also classify chemotherapeutics with 100% accuracy, this eight-shRNA signature had the highest range of measurement across all drugs. Notably, this eight-shRNA signature could also correctly classify bendamustine, CY190602, apigenin and NSC3852-drugs that were not included in the feature reduction and cross-validation of the eight-shRNA signature (Fig. 11c).

Given the known off-target potential of RNAi, it was next determined whether the functional signature derived from these eight shRNAs was attributable to the specific effect of shRNA target gene suppression on therapeutic response. To do this, a second set of shRNAs targeting the same eight genes was used to generate an independent drug response signature. Comparison of shRNA pairs revealed a high correlation between drug response signatures ($r^2 = 0.86$) in cells transduced with distinct shRNAs targeting the same gene; indicating that the major effects of these shRNAs are 'on target' (Fig. 4b). Additionally, unsupervised hierarchical clustering of the first eight-shRNA response signature or the combined response signatures generated using the first and second eight-shRNA sets revealed the same seven drug classes identified with the original 29-shRNA signature (Fig. 4c). Notably, however, the second set of eight shRNAs could independently predict only five out of seven drug classes. This loss of resolution in the second shRNA set may represent trace 'off-target' shRNA activity in either eight shRNA set. Alternatively, these differences may be attributable to small differences in the degree of target gene

knockdown conferred by distinct hairpins. Consistent with the latter argument, shRNAs in the second set frequently showed reduced target gene suppression (Figs. 7a-7b and Table 2) and yielded more subtle biological effects, as evidenced by the relative RI values seen in shRNA pairwise comparisons (Fig. 4b).

To extend the eight-shRNA signature approach in a scalable and stringent manner, a common problem in machine learning was revisited. A nonparametric classification method like K-nearest neighbors will classify any test compound according to its closest neighbor(s), even if the two compounds are quite distinct. Thus, it becomes difficult to determine how distantly a given compound can reside from a reference category of drugs and still be considered to share a similar mechanism of action (Fig. 5a). To overcome this problem, the carefully selected mechanistic diversity of the training set was taken advantage of to create specific empirical cumulative distribution functions for each drug category (Fig. 5b). This allowed the determination of whether a test compound was likely to belong to either an existing or a new drug category-a process critical to the broader applicability of this approach.

To determine whether this methodology could correctly categorize chemotherapeutics absent from the initial reference set, a set of 16 additional anticancer drugs were examined (Table 1 and Fig. 12). In each case, the eight-shRNA approach successfully grouped drugs according to their mechanism of action. Importantly, when compounds that represent new drug categories were examined, they were not misclassified into the 'nearest' drug' category. Rather, they were identifiable as distinct agents that were significantly different from all other drug categories. Consequently although this eight-shRNA panel was assembled on the basis of responses to seven drug classes, it was also successful in predicting other classes of chemotherapeutics when the training set was updated with new reference compounds. For example, the eight shRNA signature accurately predicted that the proteasome inhibitor gliotoxin belonged to a drug category not represented by any of the existing reference drugs. However, when the proteasome inhibitors bortezomib (PS341) and MG132 were used to update the training set, the eight-shRNA signature was able to successfully classify gliotoxin and epoxomycin as proteasome inhibitors (Table 1). The eight-shRNA set could be similarly trained to identify two entirely distinct drug categories-Hsp90 inhibitors and EGFR inhibitors, neither of which was used to create the eight-shRNA reference set. Notably, the

eight-shRNA signature could also distinguish functional drug subclasses within larger targeted classes of therapeutics. For example, the HER2 inhibitors lapatinib and AEE788 and the multikinase inhibitor sunitinib clustered in distinct categories relative to EGFR inhibitors (Fig. 13), despite all of these drugs belonging to the broader category of tyrosine kinase inhibitors. Although the use of more optimized sets of shRNAs may be necessary to probe fine details of certain drug categories, these data indicate that this eight-shRNA set has resolution over a broad range of cytotoxic activities.

Although the cells used in this study are responsive to a number of targeted chemotherapeutics, such as EGFR inhibitors, a potential limitation of this approach is that it lacks resolution for certain compounds requiring cellular targets not present in lymphoma cells. To determine whether this approach could be adapted to cell lines expressing targetable genetic lesions, the performance of the eight-shRNA signature in cells derived from a BCR-Abl-driven model of acute B cell leukemia (B-ALL) was examined (Williams, RT., et al., Proc. Natl. Acad, Sci. USA 103, 6688-6693 (2006)). Strikingly, a robust functional signature for alkylating agents could be generated in these cells using the same eight-shRNA set (Fig. 6). Notably, however, the response signature in B-ALL cells differed from that in lymphoma cells. For example, leukemia cells showed distinct genetic dependencies on ATR, DNA-PKcs and Bok. Thus, informative signatures can be derived in distinct cell lines, even if the signatures differ between cell types. Notably, this eight-shRNA signature may not be optimal for B-ALL cells, as feature reduction from the 29-shRNA signature was not performed in this context. Additionally, this signature may not have the same resolution as in lymphoma cells. However, these data indicate that even suboptimal signatures may provide resolution sufficient to cluster classes of chemotherapeutics.

## DISCUSSION

The functional genetic approach described herein has similarities to well-characterized chemogenomic profiling strategies in lower organisms. However, this approach also has notable advantages over existing genetic approaches for examining drug mechanisms of action and identifying drug targets. First, this approach is sufficiently sensitive to differentiate drugs with distinct targets but common downstream signaling pathways. For example, TopoI and II poisons

produce distinct shRNA sensitivity profiles, yet both ultimately engage common transcriptional networks. Microarray approaches that focus on downstream changes in gene expression are, consequently, less able to distinguish between conventional anticancer agents. In fact, previous microarray studies have shown limited resolution over a number of frontline chemotherapeutics (Table 4). Second, this approach is unaffected by pharmacodynamic variability, such as distinctions in drug efflux or detoxification, that obscures comparisons between different cancer cell lines. Finally, and most importantly, this approach is both simple and tractable. Although microarray studies suffer from significant variability between experiments and laboratories, RNAi-based functional arrays are highly reproducible and can be widely disseminated.

Perhaps the most unanticipated aspect of this work lies in the quantity of information that can be derived from of a small set of mammalian loss-of-function phenotypes. This focused shRNA signature can characterize a diverse range of drug categories at high resolution and is extendable to completely new drug categories and distinct cell types, indicating that such signatures serve as a tractable approach to screen chemical libraries for diverse functional classes of small molecules in a high-throughput manner. Although this specific set of shRNAs may not provide optimal resolution for all cell types or small molecules, these data also indicate that alternative small sets of shRNAs may yield similar information content. For example, although this work focuses on cell viability, it is likely that given appropriate phenotypic resolution-bioactive compounds affecting diverse aspects of biology can similarly be interrogated with distinct targeted sets of shRNAs.

Table 1

(a) Table showing the predictive power of the eight-shRNA signature on a set of drugs that were not used to derive the signature. The Prediction column indicates the mechanism of action of the compound as predicted by a nearest-neighbors approach. The linkage ratio describes the proximity of a test compound to a particular class of compounds and defines the observed increase (or decrease) in drug category site. For example, a linkage ratio of 1:1 indicates that the addition of a new drug expands the drug category by 10%. The p-value describes whether the proximity of a compound to a given drug category is significant when compared to a negative control distribution for that drug category. Cantharidin (a protein phosphatase inhibitor), apoptosis activator 2 (AA2, a direct activator of the apoptosome), gliotoxin (a proteasome inhibitor) and AG1478 (an EGFR inhibitor) were used as negative controls and were predicted to be distinct from any of the existing reference drugs. (b) Category predictions and significance levels upon adding three new drug categories (proteasome, Hsp90 and EGFR inhibitors) that were not used to develop the initial eight-shRNA signature. PU-H71-Br is a chemical derivative of the benzyladenine-based Hsp90 inhibitor PU-H71. VER-50589 and neopentylamine-42 are hsp90 inhibitors. PD 173074 (a FGFR inhibitor) and GDC 0941 (a PI3K Inhibitor) were used as negative controls to test the stringency of predictions after the incorporation of these new drug categories.

| **Compounds** | **Prediction** | **Linkage Ratio** | ***P*-value** | **Correct Prediction** |
|---|---|---|---|---|
| Daunorubicin | Topoll poison | 0.83 | 0.0004 | Yes |
| Zebularine | HDAC/DNMT inhibitor | 0.82 | 0.002 | Yes |
| Busulfan | DNA cross-linking | 0.97 | 0.007 | Yes |
| 2-Methoxyestradiol | Antimicrotubule | 1.08 | 0.009 | Yes |
| Vinblastine | Antimicrotubule | 0.95 | 0.001 | Yes |

| Compounds | Prediction | Linkage Ratio | *P*-value | Correct Prediction |
|---|---|---|---|---|
| Scriptaid | HDAC/DNMT inhibitor | 0.85 | 0.004 | Yes |
| Carmustine | DNA cross-linking | 0.95 | 0.005 | Yes |
| Mitoxantrone | Topoll poison | 0.94 | 0.002 | Yes |
| Thiotepa | DNA cross-linking | 0.92 | 0.003 | Yes |
| Gemicitabine | Nucleotide synthesis inhibitor | 0.93 | 0.002 | Yes |
| Melphanlan | DNA cross-linking | 0.91 | 0.002 | Yes |
| Carboplatin | DNA cross-linking | 0.86 | 0.0008 | Yes |
| Streptozocin | SSB inducer | 1.03 | 0.02 | Yes |
| Mephosphamide | DNA cross-linking | 0.93 | 0.003 | Yes |
| Irinotecan | SSB inducer | 0.88 | 0.002 | Yes |
| Noscapine | Antimicrotubule | 0.83 | 0.0002 | Yes |
| Cantharidin | Novel mechanism | 1.27 | 0.10 | Negative control |
| AA2 | Novel mechanism | 1.39 | 0.30 | Negative control |
| Gliotoxin | Novel mechanism | 1.11 | 0.09 | Negative control |
| AG1478 | Novel mechanism | 1.08 | 0.06 | Negative control |

Table 2 - shRNA Target Sequences

| Gene | Gene Function | Gene ID | shRNA Target Sequences * |
|---|---|---|---|
| | sequence-specific transcription factor, pro- | | CCACTACAAGTACATGTGTAA (SEQ ID NO: 1) |
| p53 | apoptotic | 22059 | TGGAGAGTATTTCACCCTCAA (18%) (SEQ ID NO: 2) |
| ATM | DNA damage response, checkpoint signaling, DNA repair, phosphorylation of p53 | 11920 | CACGAAGTCCTCAATAATCTA (SEQ ID NO: 3) |
| | DNA damage response, checkpoint signaling, | | CAGAAACACATAATCATTAAA (SEQ ID NO: 4) |
| Chk2 | DNA repair, phosphorylation of p53 | 50883 | CACTTTCACTATGTAGAAATA (SEQ ID NO: 5) |
| | DNA repair, DNA replication, phosphorylation of | | ACCCATGTTCTTGACATTGAA (SEQ ID NO: 6) |
| ATR | p53 | 245000 | ACCTTTAATGAGTGTCTTAAA (SEQ ID NO: 7) |
| | DNA repair, DNA replication, phosphorylation of | | CAGGAATATTCTGATTGGAAA (SEQ ID NO: 8) |
| Chk1 | p53 | 12649 | AAGGGCTTGACCAATTATAAA (SEQ ID NO: 9) |
| Smg1 | nonsense-mediated mRNA decay, DNA damage response, checkpoint signaling, phosphorylation of p53 | 233789 | CAGGATAGCAATAAAGATGAA CAGGCTGCATTCAATAACTTA (SEQ ID NO: 10) |
| DNA-PKcs | DNA damage response, DNA repair phosphorylation of p53 | 19090 | CAGGCCTATACTTACAGTTAA CTCCAACATGTAGAGAACAAA (SEQ ID NO: 11) |
| JNK1 | DNA damage response, stress signaling, phosphorylation of p53 | 26419 | TCAGAGCATAACAAACTTAAA (SEQ ID NO: 12) |
| p38 | DNA damage response, checkpoint signaling, stress signaling, phosphorylation of p53 | 26416 | CAGGTCTTGTGTTTAGGTCAA (SEQ ID NO: 3) |
| A1 | Bcl-2 family gene, anti-apoptotic | 12044 | GGAAGATGGCTTCATAAAGAA (SEQ ID NO: 14) |
| Bclb | Bcl-2 family gene, anti-apoptotic | 12049 | AAGGAATCCCTTGAAACCTAA (SEQ ID NO: 15) |
| | | | |
| | | | |

| Gene | Gene Function | Gene ID | shRNA Target Sequences * |
|---|---|---|---|
| **Bclw** | Bcl-2 family gene, anti-apoptotic | 12050 | GGCTATAAGCTGAGGCAGAAG (SEQ ID NO: 16) |
| | Bcl-2 family gene, anti-apoptotic (long form), pro- | | GGAGAGCGTTCAGTGATCTAA (SEQ ID NO: 17) (targets |
| **Bclx** | apoptotic (short form) | 12048 | both long and short forms of Bclx) |
| **Bad** | Bcl-2 family gene, pro-apoptotic | 12015 | CGCGAGAAACGTGCTTTATAA (SEQ ID NO: 18) |
| **Bak** | Bcl-2 family gene, pro-apoptotic | 12018 | CCGGAACCTATGATTACTTGA (SEQ ID NO: 19) |
| **Bax** | Bcl-2 family gene, pro-apoptotic | 12028 | CCGCGTGGTTGCCCTCTTCTA (SEQ ID NO: 20) |
| **Bid** | Bcl-2 family gene, pro-apoptotic | 12122 | CACAGAAGATTCCATATCAAA (SEQ ID NO: 21) |
| **Bik** | Bcl-2 family gene, pro-apoptotic | 12124 | CCGGACAGGTGTCAGAGGTAT (SEQ ID NO: 22) |
| | | | TAGGAACAGAGAAATATGCAA (SEQ ID NO: 23) |
| **Bim** | Bcl-2 family gene, pro-apoptotic | 12125 | CACCCTCAAATGGTTATCTTA (22%) (SEQ ID NO: 24) |
| **Bmf** | Bcl-2 family gene, pro-apoptotic | 171543 | CGCAGAGCCCTGGCATCACAA (SEQ ID NO: 25) |
| **Bnip3l** | Bcl-2 family gene, pro-apoptotic | 12177 | GGTATCAGACTGGTCCAGTAG (SEQ ID NO: 26) |
| **Bclg** | Bcl-2 family gene, less defined | 66813 | TCCAAACAGCATAGAGTTCAA (SEQ ID NO: 27) |
| | | | CTGGCCTCTGTGACTGCTCTA (SEQ ID NO: 28) |
| **Bok** | Bcl-2 family gene, less defined | 51800 | TCGGTGTCCAGCCCTAGAGAA (25%) (SEQ ID NO: 29) |
| **BPR** | Bcl-2 family gene, less defined | 75736 | CCCAGCCTCTTCCGAGTTCTA (SEQ ID NO: 30) |
| **Hrk** | Bcl-2 family gene, pro-apoptotic | 12123 | CAGCAGGGAGTGTCTACTTTA (SEQ ID NO: 31) |
| **Mil1** | Bcl-2 family gene, pro-apoptotic | 94044 | CCTGAAGAAGTGAAGAGCTTA (SEQ ID NO: 32) |
| **Mule** | Bcl-2 family gene, E3 ligase for Mcl-1 and p53 | 59026 | CCACCTCAGCTACTTCAAGTT (SEQ ID NO: 33) |
| | | | |

| Gene | Gene Function | Gene ID | shRNA Target Sequences * |
|---|---|---|---|
| Noxa | Bcl-2 family gene, pro-apoptotic | 58801 | CAGATTGAATAGTATGTGATA (SEQ ID NO: 34) |
| | | | CTGTAGATATACTGGAATGAA (SEQ ID NO: 35) |
| Puma | Bcl-2 family gene, pro-apoptotic | 170770 | |

Table 3

| Drug Name | Abbreviation | Mechanism of Action | Drug Category | Concentration used* |
|---|---|---|---|---|
| Chlorambucil | CBL | DNA crosslinking | Alkylating agent | 6.6 uM |
| Bendamustine | | DNA crosslinking | Alkylating agent | 110 uM |
| CY-190602 | CY-B | DNA crosslinking | Alkylating agent | 1.4 uM |
| Melphalan | L-PAM | DNA crosslinking | Alkylating agent | 3.9 uM |
| Maphosphamide | MAF | DNA crosslinking | Alkylating agent | 6.9 uM |
| Carmustine | BCNU | DNA crosslinking | Alkylating agent | 26.7 uM |
| Busulfan | BSF | DNA crosslinking | Alkylating agent | 40.6 uM |
| ThioTEPA | TSPA | DNA crosslinking | Alkylating agent | 2.1 uM |
| Carboplatin | CDDPC | DNA crosslinking | Platinum | 16.4 uM |
| Cisplatin | CDDP | DNA crosslinking | Platinum | 3.3 uM |
| Mitomycin C | MMC | DNA crosslinking | Anti-tumor antibiotic | 90 nM |
| Camptothecin | CPT | Induction of SSBs | Topo I poison | 0.4 nM |
| Irinotecan | CPT11 | Induction of SSBs | Topo I poison | 2.1 uM |
| NSC3852 | NSC3852 | Induction of SSBs | Topo I poison | 1.7 uM |
| Doxorubicin | Dox | Induction of DSBs | Topo II poison | 16.2 nM |
| Etoposide | VP16 | Induction of DSBs | Topo II poison | 34 nM |
| Mitoxantrone | MITO | Induction of DSBs | Topo II poison | 0.4 nM |
| Apigenin | APG | Induction of DSBs | Topo II poison | 100 nM |
| Daunorubicin | Dau | Induction of DSBs | Topo II poison | 22.5 nM |
| 6-Thioguanine | 6TG | DNA methylation memetic | Antimetabolite | 54 nM |
| Temozolomide | TMZ | DNA methylation | Alkylating agent | 23.2 uM |
| Streptozocin | STZ | DNA methylation | Alkylating agent | 83 uM |
| 17 AAG | | HSP90 inhibition | HSP90 inhibitor | 2.6 uM |
| BIIB021 | | HSP90 inhibition | HSP90 inhibitor | 0.79 uM |
| PU-H71-Br | | HSP90 inhibition | HSP90 inhibitor | 0.53 uM |
| Neopentylamine 42 | | HSP90 inhibition | HSP90 inhibitor | 1.1 uM |
| VER-50589 | | HSP90 inhibition | HSP90 inhibitor | 0.16 uM |
| Vorinostat | SAHA | HDAC inhibition | HDAC inhibitor | 0.2 uM |
| Scriptaid | | HDAC inhibition | HDAC inhibitor | 0.5 uM |
| Decitabine | DAC | DNMT inhibition | Antimetabolite | 0.6 uM |
| Zebularine | | DNMT inhibition | Antimetabolite | 30 uM |
| PD 173074 | | FGFR inhibition | FGFR inhibitor | 4.2 uM |
| | | | | |
| | | | | |
| | | | | |

| Drug Name | Abbreviation | Mechanism of Action | Drug Category | Concentration used* |
|---|---|---|---|---|
| AG1478 | | EGFR inhibition | EGFR inhibitor | 19 uM |
| Gefitinib | | EGFR inhibition | EGFR inhibitor | 12.5uM |
| Erlotinib | | EGFR inhibition | EGFR inhibitor | 6.5 uM |
| GDC 0941 | | PI3K inhibition | PI3K inhibitor | 5.4 uM |
| Actinomycin D | ActD | RNA syn. inhibition | Anti-tumor antibiotic | 1.2 nM |
| Roscovitine | Rosco | CDK inhibition | Kinase inhibitor | 26.7 uM |
| Rapamycin | RAPA | Protein syn. inhibition | Anti-tumor antibiotic | 15 uM |
| Bortezomib | PS341 | Proteasome inhibition | Proteasome inhibitor | 13 nM |
| Epoxomicin | | Proteasome inhibition | Proteosome inhibitor | 45 nM |
| Gliotoxin | | Proteasome inhibition | Proteosome inhibitor | 0.3 uM |
| Vinblastine | VLB | Disruption of mitosis | Antimicrotuble agent | 2.5 nM |
| Paclitaxel | Taxol | Disruption of mitosis | Antimicrotuble agent | 8 nM |
| Vincristine | VCR | Disruption of mitosis | Antimicrotuble agent | 1.5 nM |
| Noscapine | | Disruption of mitosis | Antimicrotuble agent | 58.1 uM |
| 2-Methoxyestradiol | | Disruption of mitosis | Antimicrotuble agent | 1.8 uM |
| Methotrexate | MTX | nucleic acid syn. inhibition | Antimetabolite | 33 nM |
| 5-Fluorouracil | 5FU | nucleic acid syn. inhibition | Antimetabolite | 30.5 nM |
| Hydroxyurea | HU | nucleic acid syn. inhibition | Antimetabolite | 78 uM |
| Fludarabine | dFdC | nucleic acid syn. inhibition | Antimetabolite | 3.8 uM |
| Gemcitabine | GEM | nucleic acid syn. inhibition | Antimetabolite | 100 nM |
| Cantharidin | | Phophatase inhibition | Phophatase inhibitor | 10.7 uM |
| Apoptosis Activator 2 | AA2 | Direct activation of apoptosome | Activator of apoptosis | 1.9 uM |

Table 4

a

Vorinostat query

| Rank | cmap name | Mean | n | Enrichment | p | Specificity | Percent non-null | Known Mechanism/Target |
|---|---|---|---|---|---|---|---|---|
| 1 | vorinostat | 0.743 | 12 | 0.976 | 0 | 0.0151 | 100 | HDACi |
| 2 | trichostatin A | 0.631 | 182 | 0.898 | 0 | 0.0095 | 97 | HDACi |
| 3 | resveratrol | 0.275 | 9 | 0.757 | 0 | 0.0343 | 100 | HDACi |
| 4 | 15-delta prostaglanc | 0.16 | 15 | 0.589 | 0 | 0.1173 | 86 | PPAR-gamma agonist |
| 5 | thioridazine | 0.169 | 20 | 0.588 | 0 | 0.1324 | 75 | Seratonin antagonist |
| 6 | geldanamycin | 0.203 | 15 | 0.587 | 0 | 0.0924 | 86 | HSP90i |
| 7 | LY-294002 | 0.202 | 61 | 0.578 | 0 | 0.0134 | 85 | PI3Ki |
| 8 | sirolimus | 0.143 | 44 | 0.515 | 0 | 0.0361 | 68 | mTOR |
| 9 | valproic acid | 0.206 | 57 | 0.512 | 0 | 0 | 70 | HDACi |
| 10 | tanespimycin | 0.172 | 62 | 0.503 | 0 | 0.0725 | 80 | HSP90i |

Geldanamycin query

| Rank | cmap name | Mean | n | Enrichment | p | Specificity | Percent non-null | Known Mechanism/Target |
|---|---|---|---|---|---|---|---|---|
| 1 | withaferin A | 0.345 | 4 | 0.969 | 0 | 0.0211 | 100 | Heat shock response inducer |
| 2 | lomustine | 0.311 | 4 | 0.941 | 0 | 0.0118 | 100 | Alkylating agent |
| 3 | geldanamycin | 0.522 | 15 | 0.925 | 0 | 0 | 100 | HSP90i |
| 4 | tanespimycin | 0.438 | 62 | 0.827 | 0 | 0.0052 | 98 | HSP90i |
| 5 | alvespimycin | 0.394 | 12 | 0.802 | 0 | 0.0058 | 100 | HSP90i |
| 6 | vorinostat | 0.263 | 12 | 0.742 | 0 | 0.1759 | 91 | HDACi |
| 7 | monorden | 0.302 | 22 | 0.719 | 0 | 0.0054 | 100 | Poorly characterized in mammals |
| 8 | 15-delta prostaglanc | 0.282 | 15 | 0.71 | 0 | 0.0335 | 93 | PPAR-gamma agonist |
| 9 | trifluoperazine | 0.177 | 16 | 0.631 | 0 | 0.0529 | 68 | Dopamine D1, D2 receptor |
| 10 | thioridazine | 0.123 | 20 | 0.599 | 0 | 0.1279 | 50 | Anti-cholinergic |

b

Chlorambucil query

| Rank | cmap name | Mean | n | Enrichment | p | Specificity | Percent non-null | Known Mechanism/Target |
|---|---|---|---|---|---|---|---|---|
| 1 | monensin | -0.337 | 6 | -0.831 | 0.00006 | 0 | 50 | Ionophore |
| 2 | nialamide | 0.223 | 4 | 0.909 | 0.00008 | 0 | 100 | MAOI |
| 3 | quinethazone | 0.206 | 4 | 0.877 | 0.00034 | 0 | 100 | Anti-hypertensive |
| 4 | capsaicin | 0.161 | 4 | 0.86 | 0.0005 | 0 | 100 | VR1 agonist |
| 5 | fluocinonide | 0.14 | 5 | 0.774 | 0.00134 | 0.0075 | 80 | Glucocorticoid |
| 6 | Prestwick-983 | -0.732 | 3 | -0.905 | 0.00162 | 0.0069 | 100 | Poorly characterized |
| 7 | biperiden | -0.55 | 5 | -0.751 | 0.00174 | 0.0748 | 80 | Anti-cholinergic |
| 8 | bendroflumethiazide | 0.157 | 6 | 0.705 | 0.00175 | 0 | 83 | Anti-hypertensive |
| 9 | cefamandole | -0.345 | 4 | -0.805 | 0.00282 | 0.0083 | 50 | Beta-lactam antibiotic |
| 10 | ikarugamycin | -0.467 | 3 | -0.885 | 0.00296 | 0.0152 | 66 | Endocytosis inhibitor |
| | | | | | | | | |
| 90 | semustine | 0.103 | 4 | 0.568 | 0.09445 | 0.3051 | 75 | Alkylating agent |

Vinblastine query

| Rank | cmap name | Mean | n | Enrichment | p | Specificity | Percent non-null | Known Mechanism/Target |
|---|---|---|---|---|---|---|---|---|
| 1 | methotrexate | -0.473 | 8 | -0.679 | 0.00044 | 0.0069 | 87 | Nucleotide depletion |
| 2 | nifedipine | 0.061 | 7 | 0.696 | 0.00056 | 0 | 71 | Calcium channels |
| 3 | doxylamine | -0.273 | 5 | -0.796 | 0.00074 | 0.0267 | 60 | Anti-histamine |
| 4 | Prestwick-642 | 0.148 | 4 | 0.844 | 0.00097 | 0 | 100 | Poorly characterized |
| 5 | meptazinol | -0.635 | 4 | -0.826 | 0.00165 | 0.0054 | 100 | Opioid receptor agonist |
| 6 | pirenperone | 0.089 | 5 | 0.758 | 0.00194 | 0 | 60 | 5-HT2A receptor |
| 7 | valinomycin | 0.119 | 4 | 0.811 | 0.00241 | 0.0349 | 100 | K+ Carrier |
| 8 | cinnarizine | 0.12 | 4 | 0.805 | 0.00269 | 0 | 100 | Anti-histamine |
| 9 | (-)-atenolol | 0.12 | 4 | 0.8 | 0.00302 | 0 | 100 | Beta blocker |
| 10 | urapidil | 0.114 | 4 | 0.798 | 0.0032 | 0 | 100 | Alpha-1 adrenergic receptor |
| | | | | | | | | |
| 271 | podophyllotoxin | 0.079 | 4 | 0.434 | 0.32889 | 0.7286 | 50 | Microtubule assembly |

c

| Connectivity Map Comparison Criteria | Out of 14 |
|---|---|
| Top hit is either a replicate of the same compound or a compound with the same mechanism | 6 |
| At least one of the top 5 is the same mechanism | 10 |
| At least one of the top 10 is the same mechanism | 10 |

A summary of small molecule queries using the Connectivity Map. a, Tables showing results in which compounds - Vorinostat (above), and Geldanamycin (below) were queried against the connectivity map. These compounds show clear mechanistic signatures characteristic of their molecular drug class. Analogous compounds present in the top 10 search results are shown in red. b, Tables showing results in which the queried compounds are vinblastine and chlorambucil. The red text indicates the first compound with a known mechanistic relationship. Notably, while the 8-shRNA can effectively classify these compounds, the Connectivity Map lacks resolution for either agent. All data was obtained at: www.broadinstitute.org/cmap/. c, Connectivity Map analysis of 14 compounds categorized in this study. The drugs examined were Mitoxantrone, Doxorubicin, Daunorubicin, Camptothecin, Irinotecan, Carmustine, Vinblastine, Paclitaxel, Methotrexate, Vorinostat, Geldanamycin, Lomustine, MG132, and Rapamycin.

The teachings of all patents, published applications and references cited herein are incorporated by reference in their entirety.

While this invention has been particularly shown and described with references to example embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

# CLAIMS

What is claimed is:

1. A method of characterizing a mechanism of action of an agent comprising

a) contacting a plurality of populations of cells with an agent to be assessed, wherein each population of cells have one gene of interest targeted by a small hairpin RNA (shRNA) and wherein said gene of interest regulates cell death and a plurality of genes that regulate cell death are targeted in the plurality of populations of cells; and

b) determining a responsiveness of each population of cells to the agent, thereby obtaining an shRNA signature of the agent so as to identify one or more genes that mediate a response to the agent,

thereby characterizing the mechanism of action of the agent.

2. The method of Claim 1 wherein the mechanism of action of the agent comprises inhibition of a topoisomerase, cross linking of DNA, inducement of single stand break of DNA, inhibition of nucleic acid synthesis, inhibition of mitosis, inhibition of RNA transcription, inhibition of histone modification enzymes, inhibition of heat shock proteins, alkylation of DNA, or inhibition of proteasomes inducement of apoptosis.

3. The method of Claim 1 further comprising classifying the agent within a group of agents having in common one or more mechanisms of action.

4. The method of Claim 1 wherein each shRNA acts to knock down one gene.

5. The method of Claim 1 wherein the agent is used in an effective amount to induce a response in cells that do not contain said shRNA targeting said gene of interest.

6. The method of Claim 1 wherein the agent is a derivative of a chemotherapeutic agent.

7. The method of any of Claims 1-6 wherein the gene of interest that regulates cell death is a gene in the Bcl2 family of genes, a p53 gene, a p53-activating kinase gene.

8. The method of Claim 7 wherein the gene in the Bcl2 family of genes is a Bax gene, Bak gene, a Bok gene, a Bim gene, a Bid gene, a Puma gene, a Noxa gene, a Bad gene, a Bmf gene, a Bik gene, a Hrk gene, a Bclx gene, a Bclb gene, a Bclw gene, an A1 gene, a Bclg gene, a Mil1 gene, a Mule gene, a BPR gene, a BNIP gene, a Bcld gene, a Bcl2 gene, or a Mcl 1 gene.

9. The method of Claim 7 wherein the p53 activating kinase gene is an ATM gene, an ATR gene, a Chkl gene, a Chk2 gene, a DNAPKcs gene, a Smg-1 gene, a JNK1 gene or a p38 gene.

10. The method of any one of Claims 1-9 wherein the responsiveness of each population of cells to the agent is a relative level of chemo-resistance and sensitization conferred by each shRNA.

11. The method of any one of Claims 1-10 wherein the responsiveness is a relative survival rate of each population of cells compared to control cells that do not contain said shRNA targeting the gene of interest.

12. The method of any one of Claims 1-11 wherein each population of cells further express a marker gene.

13. The method of any one of Claims 1-12 wherein the shRNAs are introduced into the cells using a viral vector.

14. The method of Claim 13 wherein the viral vector is a retroviral vector.

15. The method of Claim 13 or 14 wherein the vector further expresses a marker gene.

16. The method of Claim 12 or 15 wherein the marker gene is a fluorescent marker gene.

17. The method of Claim 12 or 15 wherein the marker gene is green fluorescent protein (GFP) gene.

18. The method of Claim 16 or 17 further comprising measuring the fluorescent marker gene or GFP gene expression level in each population of cells using flow cytometry.

19. The method of any one of Claims 1-18 further comprising comparing the responsiveness of each population of cells to the agent to a control.

20. The method of Claim 19 wherein the control is a population of cells into which the shRNA targeting the gene of interest has not been introduced.

21. The method of any one of Claims 1-20 wherein the determination of the responsiveness is accomplished using cell flow cytometry, hybridization techniques or sequencing techniques.

22. The method of any one of Claims 1-21 wherein the plurality of genes are three genes, four genes, five genes, six genes, seven genes, eight genes, nine genes, ten genes, eleven genes, twelve genes, thirteen genes, fourteen genes, fifteen genes, sixteen genes, seventeen genes, eighteen genes, nineteen genes, twenty genes, twenty one genes, twenty two genes, twenty three genes, twenty four genes, twenty five genes, twenty six genes, twenty seven genes, twenty eight genes, twenty nine genes, thirty genes or more which are targeted by the corresponding shRNAs.

23. The method of any one of Claims 1-22 wherein the plurality of genes are ATM, Chk2 and p53 genes.

24. The method of any one of Claims 1-22 wherein the plurality of genes are p53, ATR, Chk1, Chk2, Smg-l, DNA-PKcs, Bok and Bim genes.

25. The method of any one of Claims 1-24 wherein the plurality of populations of cells are contacted repeatedly with the agent.

26. The method of any of Claims 1-25 further comprising introducing the plurality of shRNAs targeting the plurality of genes of interest into the plurality of populations of cells, wherein each shRNA targets one gene of interest that regulates cell death, wherein each population of cells have one gene of interest targeted.

27. The method of any one of Claims 1-26 further comprising using an algorithm to cluster a plurality of agents into groups based on the responsiveness of each population of cells to each agent.

28. The method of any one of Claims 1-27 wherein the agent is a chemotherapeutic or genotoxic agent.

29. The method of any one of Claims 1-28 wherein the agent is a chemical compound.

30. An article of manufacture for characterizing a mechanism of action of a chemotherapeutic or genotoxic agent comprising:

a plurality of populations of cells, each population having an shRNA that targets a gene of interest that mediates a response to a chemotherapeutic or genotoxic agent, and

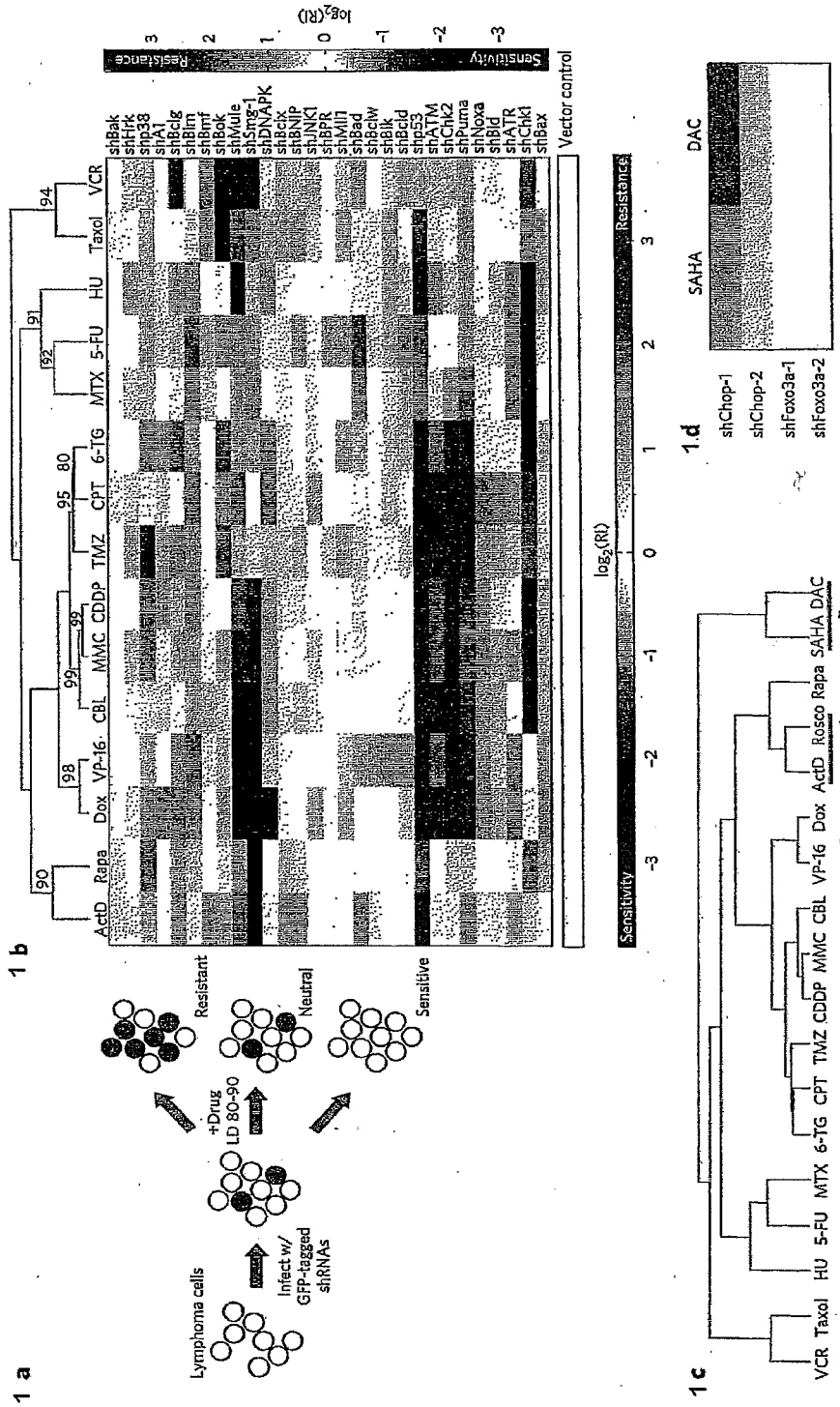
an algorithm for clustering plurality of chemotherapeutic or genotoxic agents into one or more groups based on a responsiveness of each population of cells to each agent.

31. The article of manufacture of Claim 30 wherein ATM, Chk2 and p53 genes are targeted.

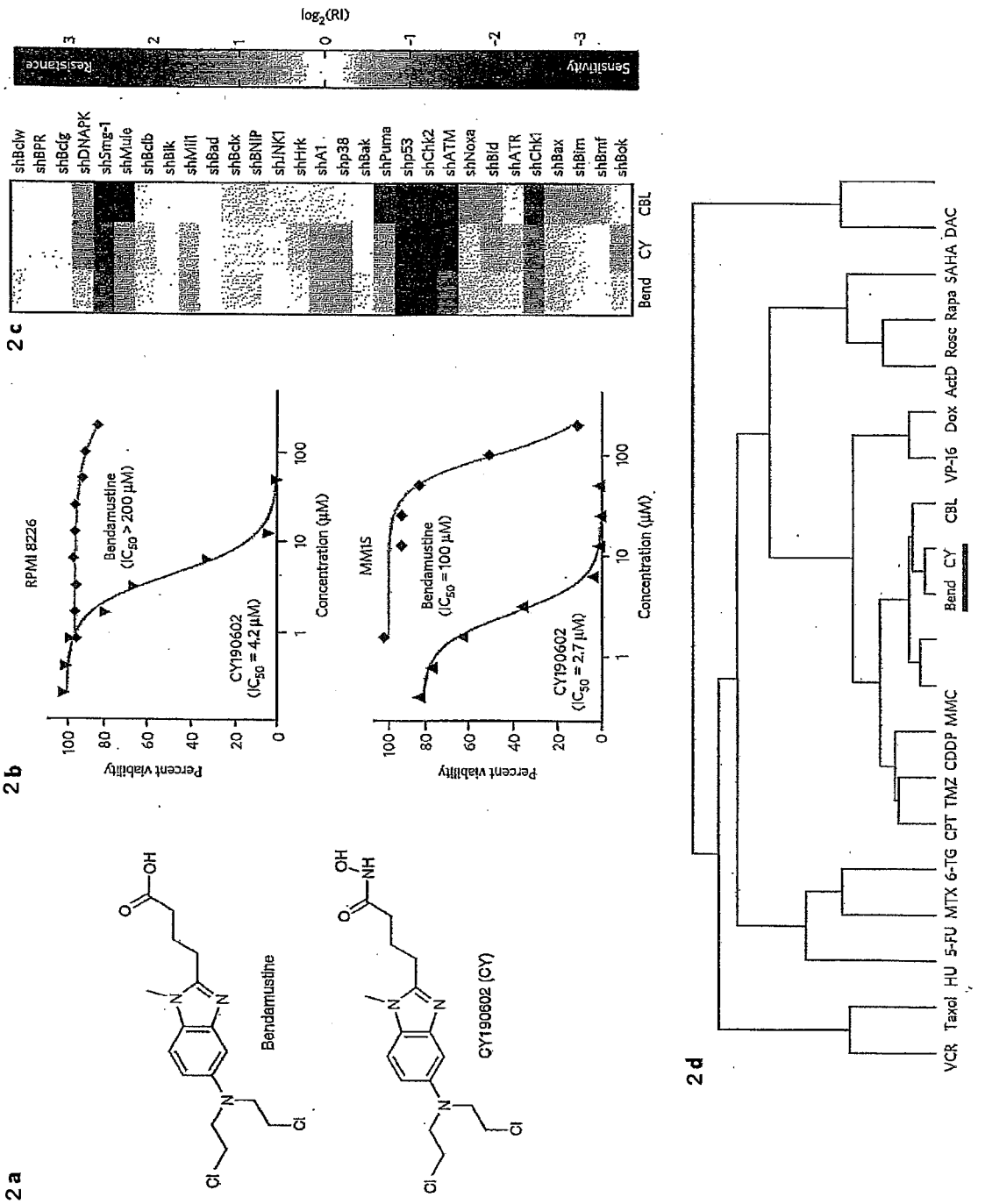
32. The article of manufacture of Claim 30 wherein p53, ATR, Chk1, Chk2, Smg-1, DNA-PKcs, Bok and Bim genes are targeted.

33. Use of the article of manufacture of any one of Claims 30, 31 or 32 for screening a library of agents for an agent having a chemotherapeutic or genotoxic effect.

Figures 1a-1d



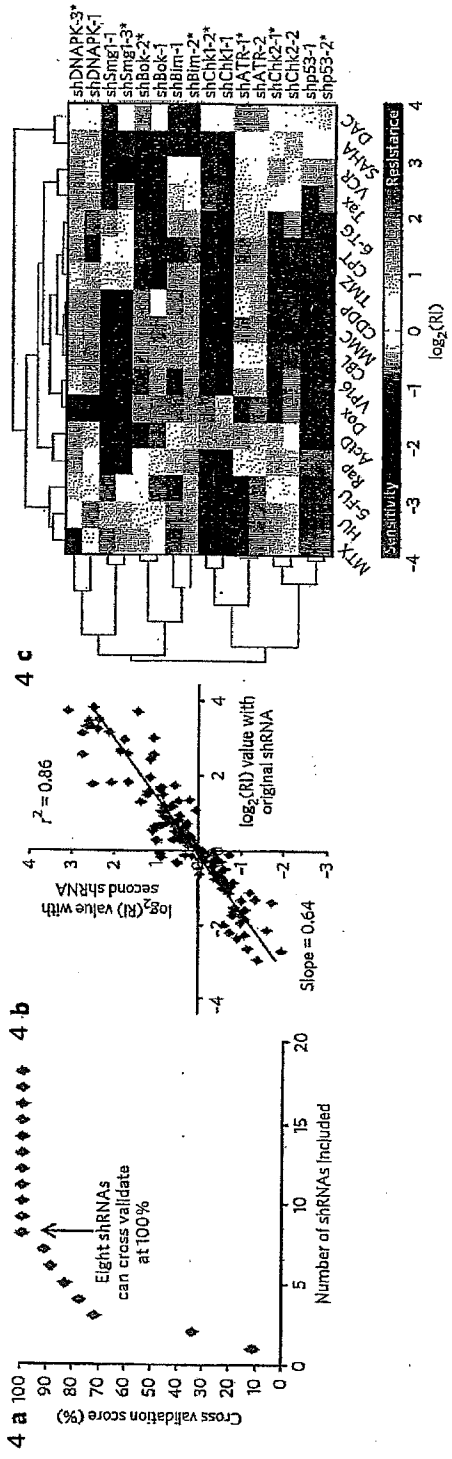
Figures 2a-2d



Figures 3a-3e



Figures 4a-4c



Figures 5a-5c

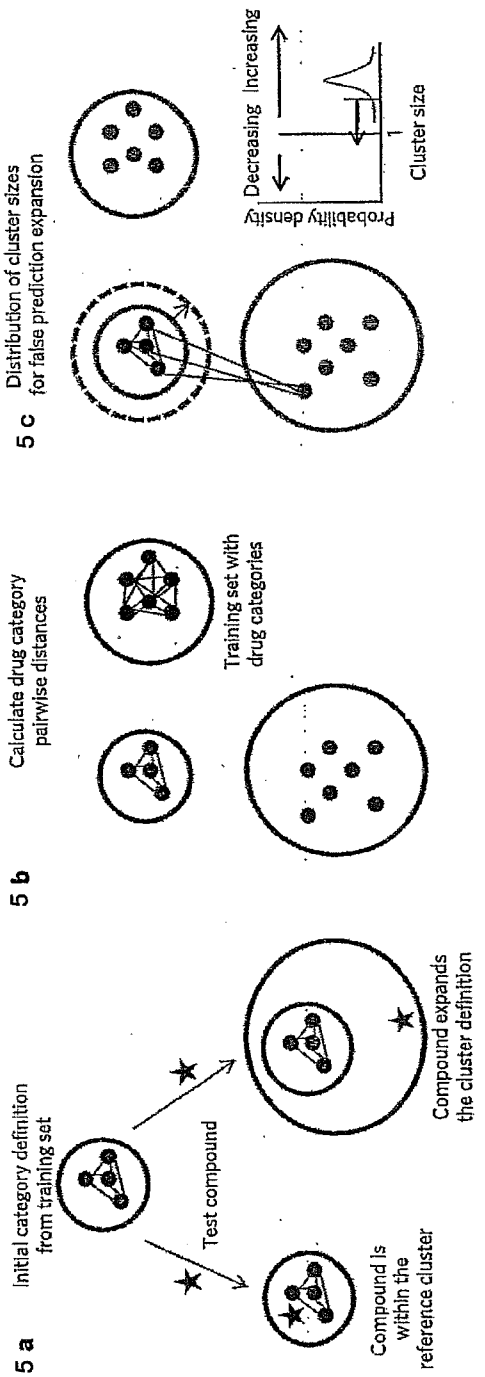
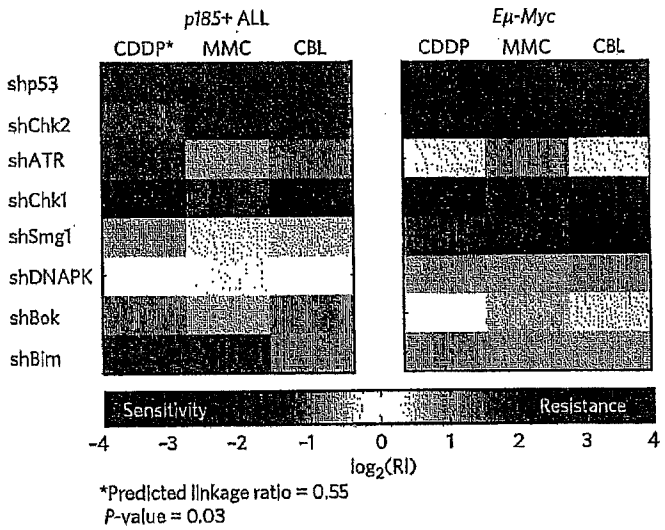
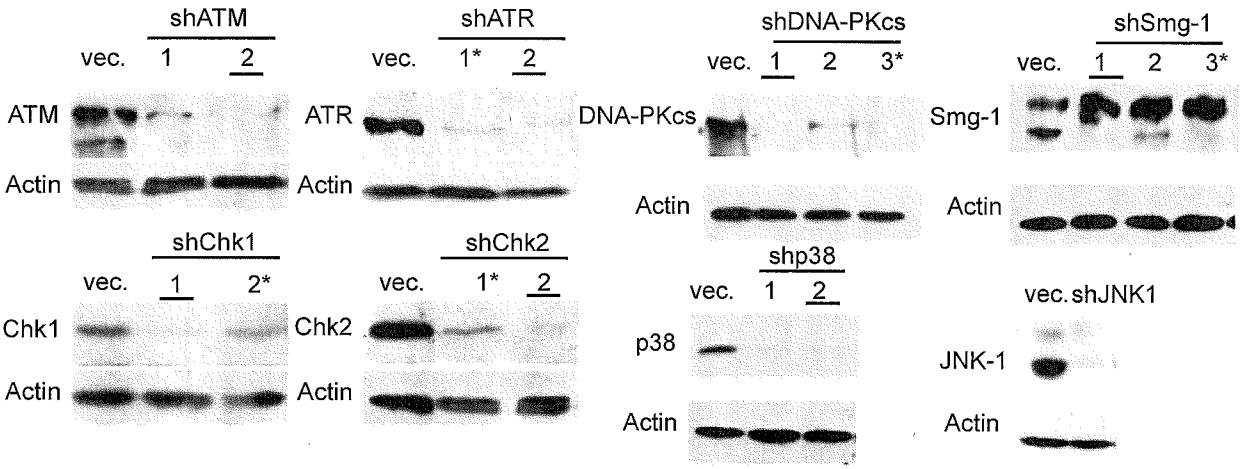


Figure 6

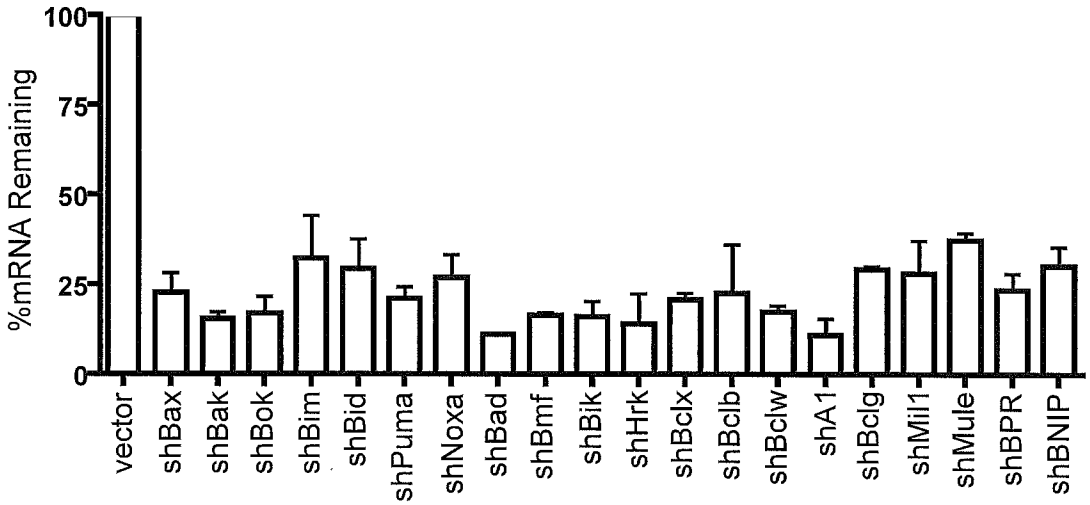

p185+ ALL
Eμ-Myc
CDDP*
MMC
CBL
CDDP
MMC
CBL
shp53
shChk2
shATR
shChk1
shSmg1
shDNAPK
shBok
shBim
Sensitivity
Resistance
-4
-3
-2
-1
0
1
2
3
4
log2(RI)
*Predicted linkage ratio = 0.55
P-value = 0.03

# Figures 7a-7b

7a

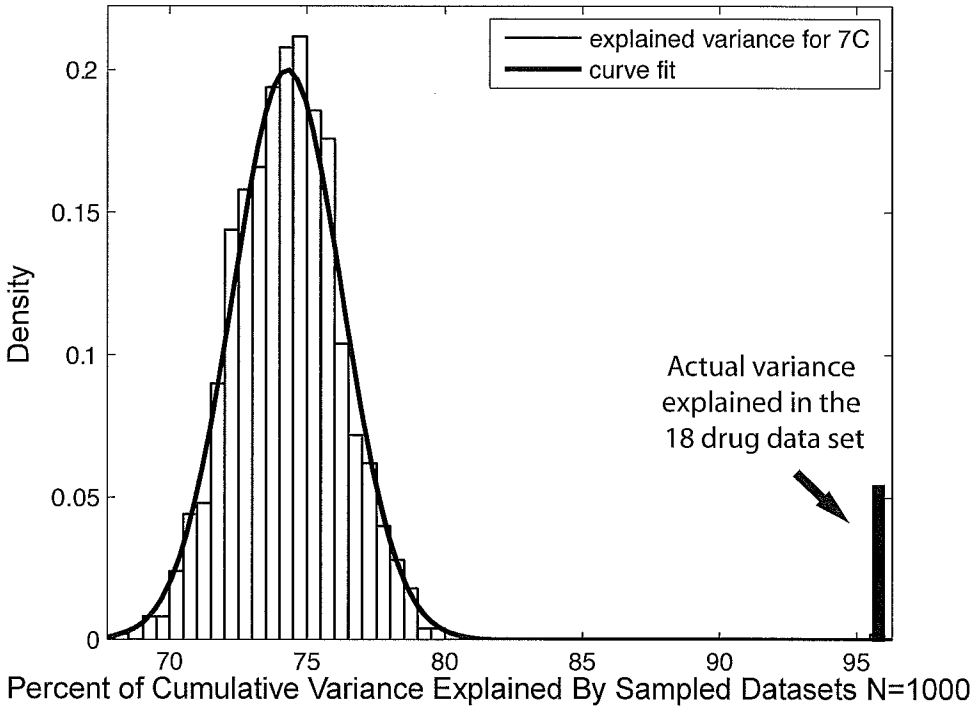



7b

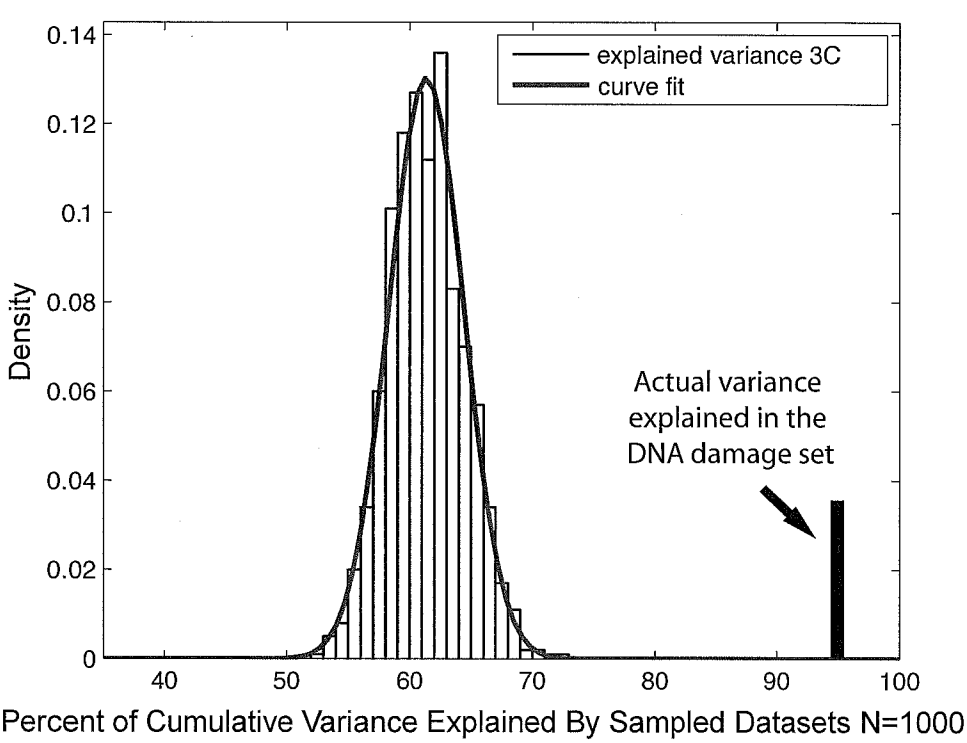


## Figures 8a-8b

8 a

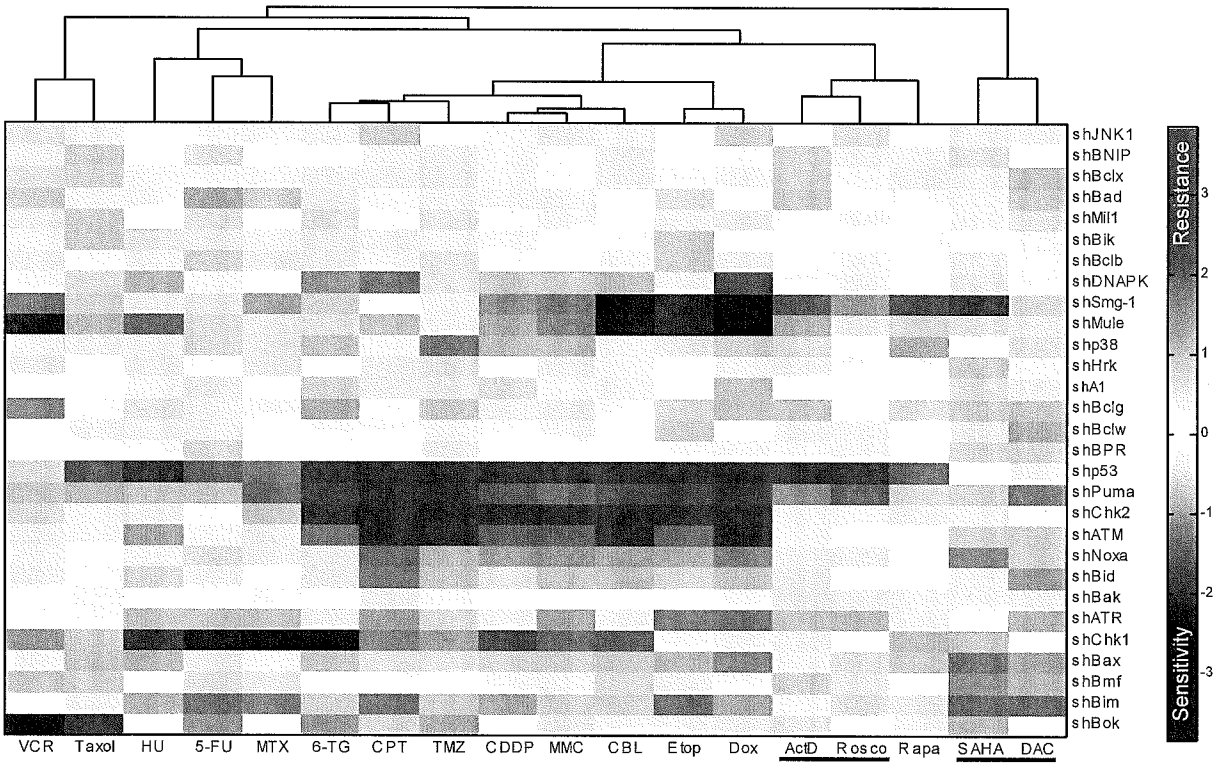



8 b



# Figures 9a-9b

9a




9b

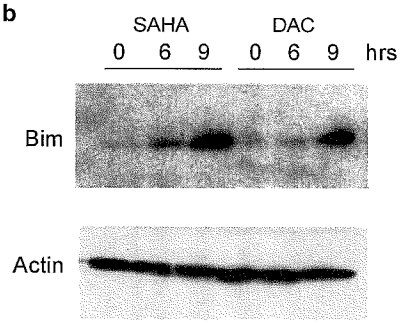


Figure 10

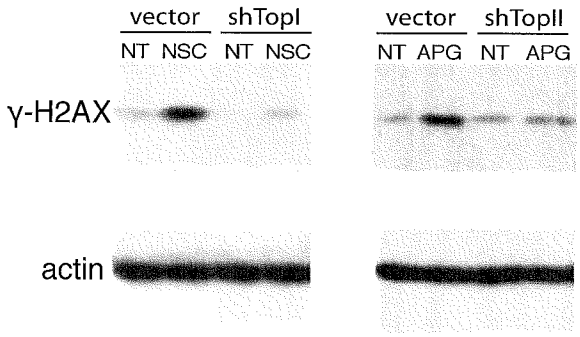

vector
shTopI
NT NSC NT NSC
vector
shTopII
NT APG NT APG
γ-H2AX
actin

Figure 11a

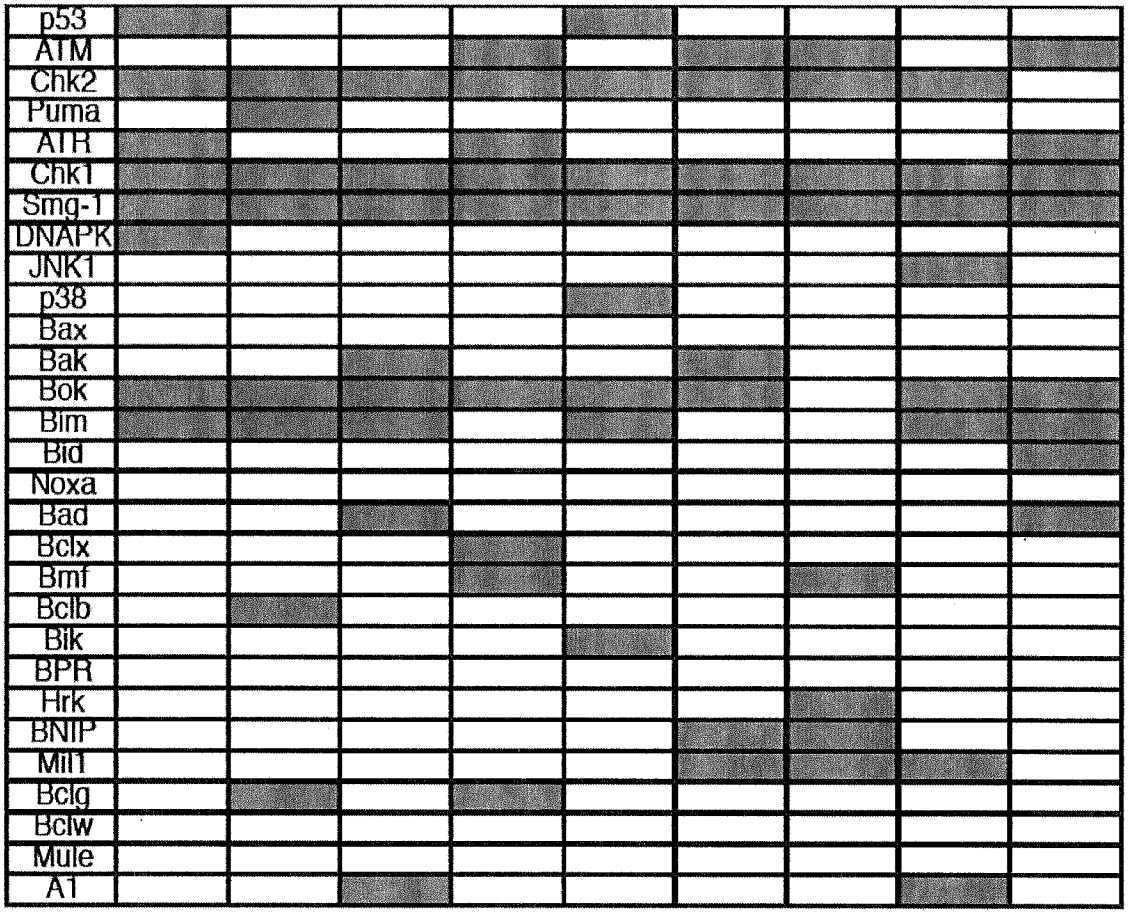

p53
ATM
Chk2
Puma
ATR
Chk1
Smg-1
DNAPK
JNK1
p38
Bax
Bak
Bok
Bim
Bid
Noxa
Bad
Bclx
Bmf
Bclb
Bik
BPR
Hrk
BNIP
Mil1
Bclg
Bclw
Mule
A1

Figure 11b

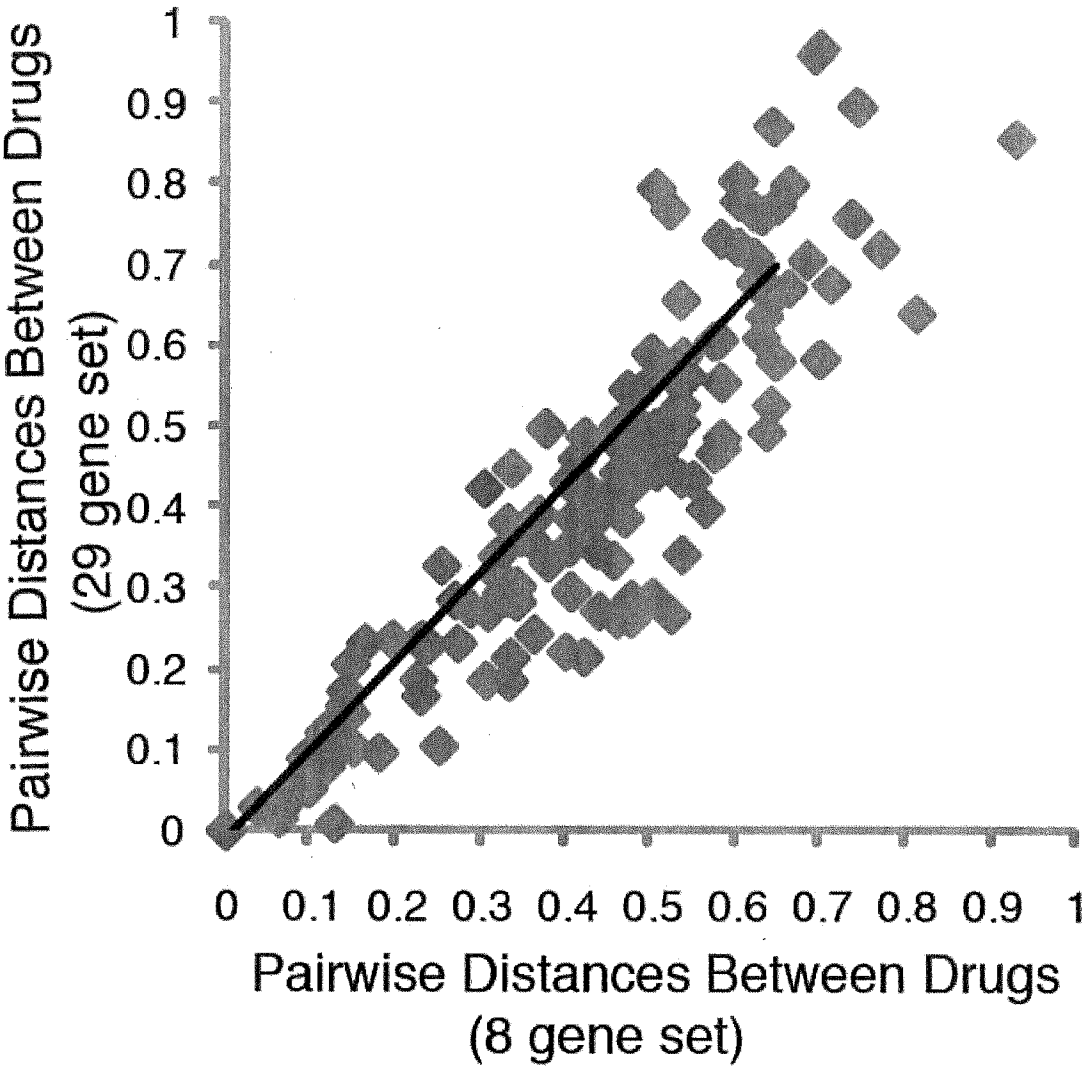

Pairwise Distances Between Drugs
(29 gene set)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Pairwise Distances Between Drugs
(8 gene set)

Figure 11c

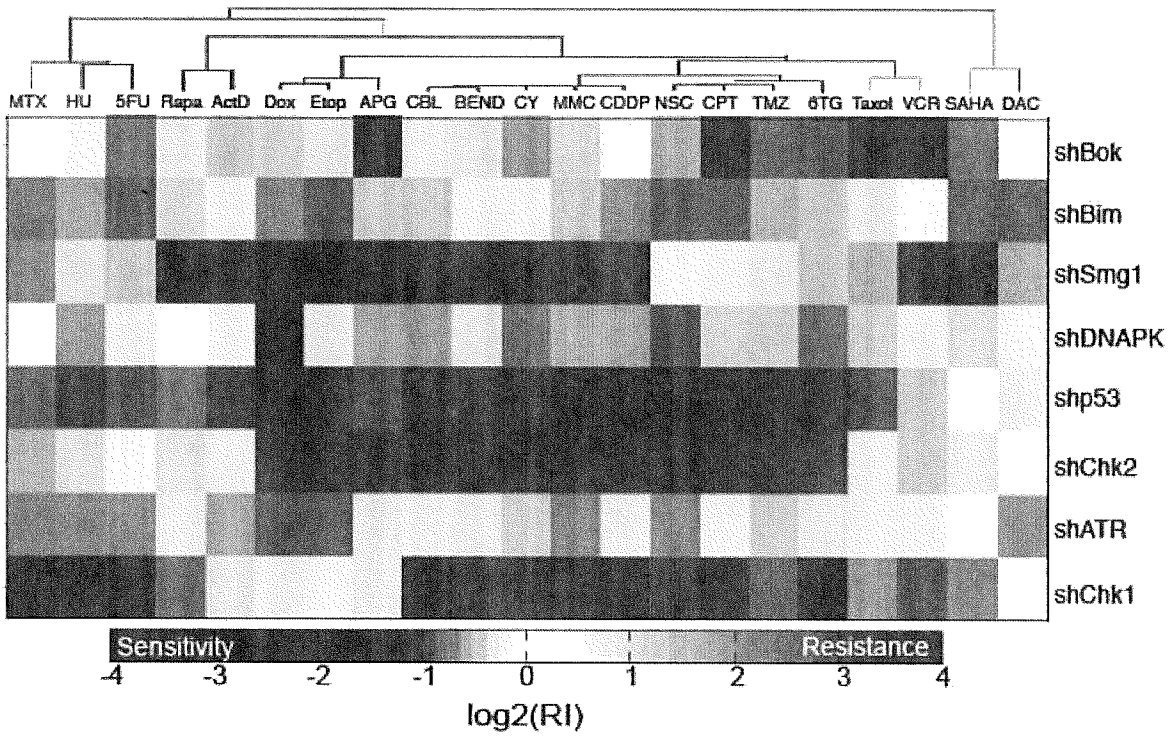

MTX
HU
5FU
Rapa
ActD
Dox
Etop
APG
CBL
BEND
CY
MMC
CDDP
NSC
CPT
TMZ
6TG
Taxol
VCR
SAHA
DAC
shBok
shBim
shSmg1
shDNAPK
shp53
shChk2
shATR
shChk1
Sensitivity
Resistance
-4
-3
-2
-1
0
1
2
3
4
log2(RI)

Figure 12

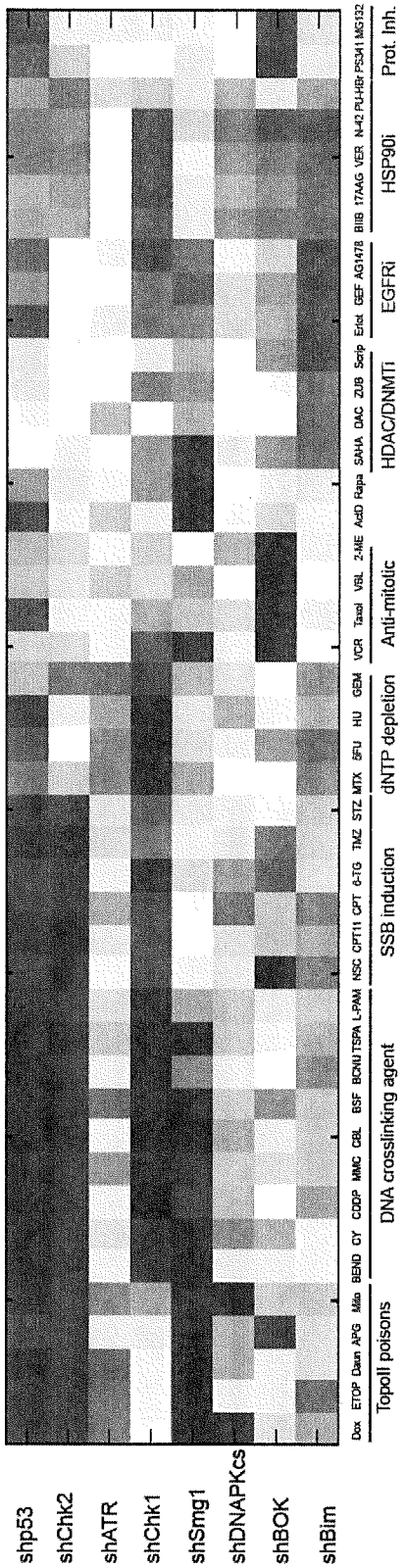

shp53
shChk2
shATR
shChk1
shSmg1
shDNAPKcs
shBOK
shBim
Dox ETOP Daun APG Mito BEND CY CDDP MMC CBL BSF BCNU TSPA L-PAM NSC CPT11 CPT 6-TG TMZ STZ MTX 5FU HU GEM VCR Taxol VBL 2-ME AdD Rapa SAHA DAC ZUB Scrip Erlot GEF AG1478 BIIB 17AAG VER N-42 PU-H8t PS341 MG132
TopoII poisons
DNA crosslinking agent
SSB induction
dNTP depletion
Anti-mitotic
HDAC/DNMTi
EGFRi
HSP90i
Prot. Inh.

Figure 13

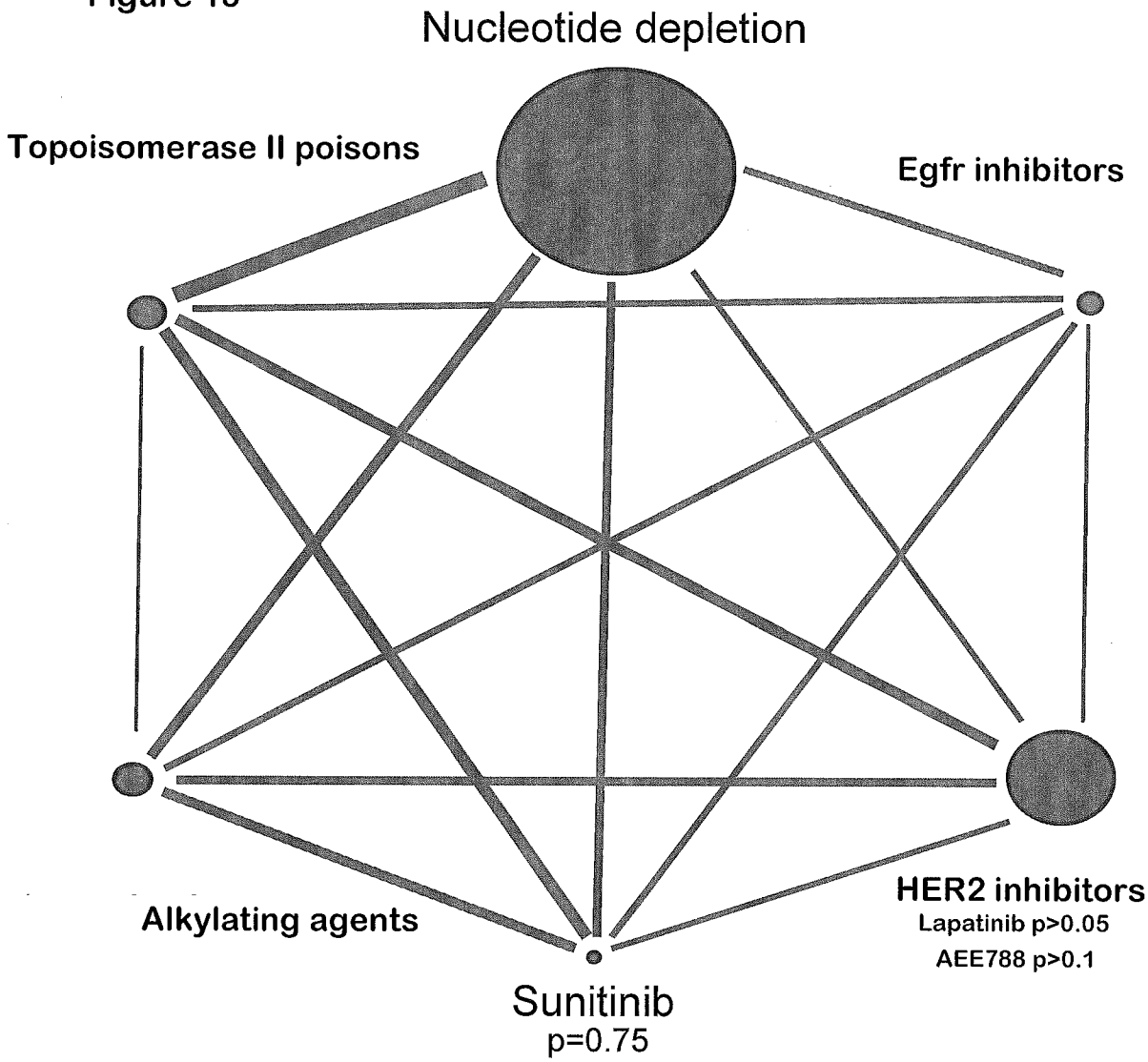
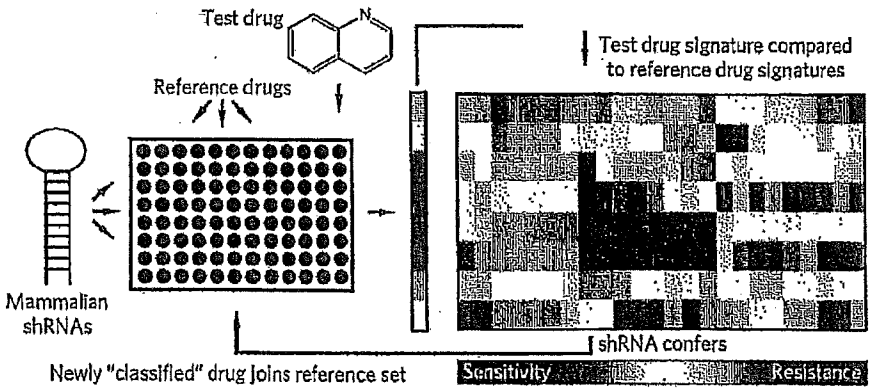

Nucleotide depletion
Topoisomerase II poisons
Egfr inhibitors
HER2 inhibitors
Lapatinib p>0.05
AEE788 p>0.1
Alkylating agents
Sunitinib
p=0.75

Figure 14


Test drug
N
Reference drugs
Test drug signature compared to reference drug signatures
Mammalian shRNAs
shRNA confers
Sensitivity
Resistance
Newly "classified" drug joins reference set

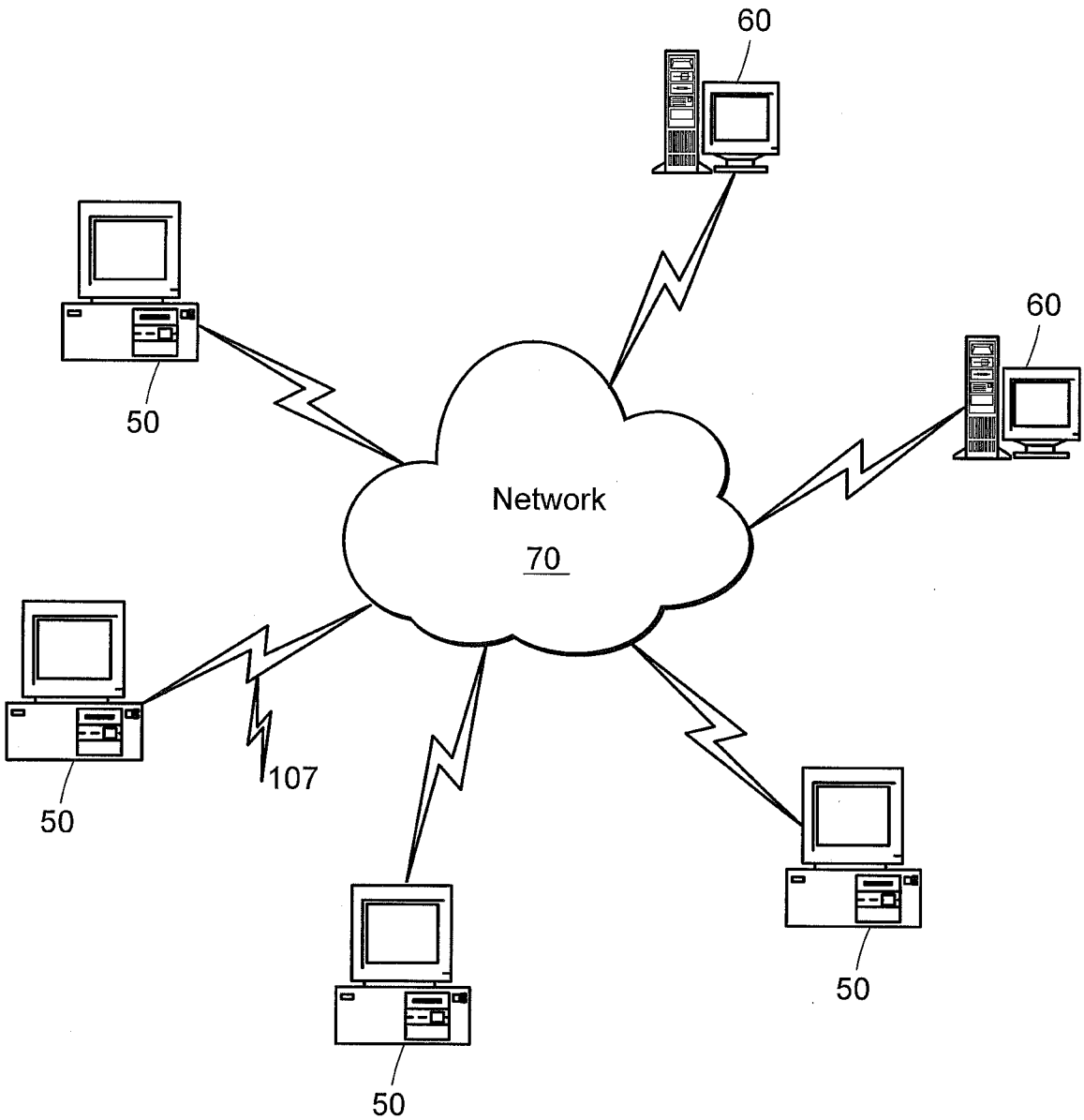
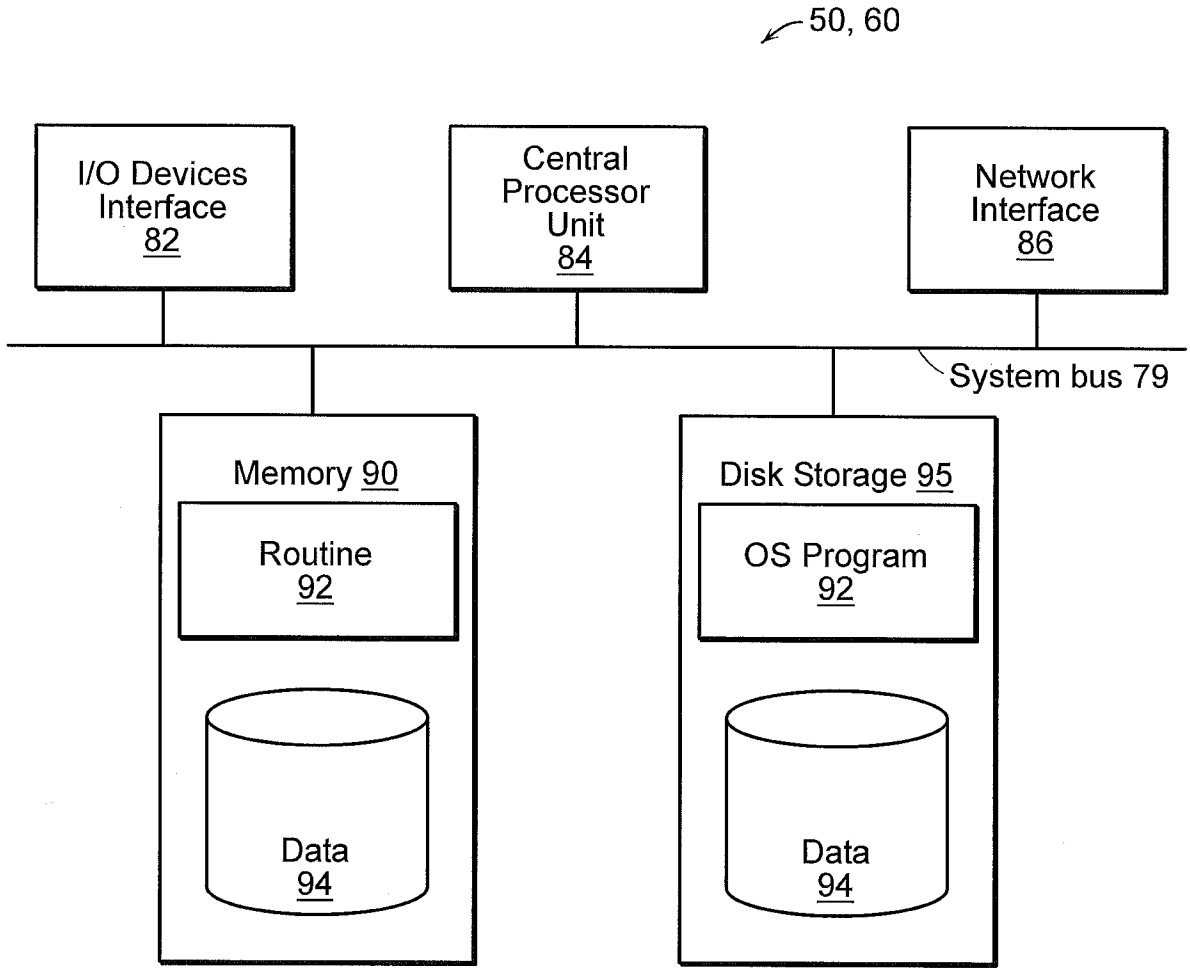

60
60
50
Network
70
107
50
50
50

Figure 16


50, 60
I/O Devices Interface 82
Central Processor Unit 84
Network Interface 86
System bus 79
Memory 90
Routine 92
Data 94
Disk Storage 95
OS Program 92
Data 94

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2011/065611

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, EMBASE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | JIANG HAI ET AL: "The combined status of ATM and p53 link tumor development with therapeutic response.", GENES & DEVELOPMENT, vol. 23, no. 16, 15 August 2009 (2009-08-15), pages 1895-1909, XP002675205, ISSN: 1549-5477 | 1-23, 25-31,33 |
| Y | abstract Materials and methods; page 1896, column 2 | 16-18, 20,24 |
| | ----- -/-- | |

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 3 May 2012 | 30/05/2012 |

| Name and mailing address of the ISA/ | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016 | Nurmi, Jussi |

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2011/065611

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | BRUMMELKAMP THIJN R ET AL: "An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors.", NATURE CHEMICAL BIOLOGY, vol. 2, no. 4, April 2006 (2006-04), pages 202-206, XP002675206, ISSN: 1552-4450 | 1-15,19, 21-32 |
| Y | abstract; figure 1 Methods | 16-18, 20,24 |
| | ----- | |
| X | BERNS KATRIEN ET AL: "A large-scale RNAi screen in human cells identifies new components of the p53 pathway.", NATURE, vol. 428, no. 6981, 25 March 2004 (2004-03-25), pages 431-437, XP002675207, ISSN: 1476-4687 | 30-32 |
| Y | abstract<br><br>-& BERNS KATRIEN ET AL: "A large-scale RNAi screen in human cells identifies new components of the p53 pathway:Supplementary Table 1 Genes in the NKI RNAi library", NATURE, vol. 428, no. 6981, 25 March 2004 (2004-03-25), page 7PP, XP002676276, Supplementary Table 1; gene nos: 13, 43, 81, 301, 1581, 1768, 3866, 5514 and 5863 | 7-9,23, 24,31,32 |
| | ----- | |
| X | MULLENDERS JASPER ET AL: "Candidate biomarkers of response to an experimental cancer drug identified through a large-scale RNA interference genetic screen.", CLINICAL CANCER RESEARCH : AN OFFICIAL JOURNAL OF THE AMERICAN ASSOCIATION FOR CANCER RESEARCH, vol. 15, no. 18, 15 September 2009 (2009-09-15), pages 5811-5819, XP002675208, ISSN: 1078-0432 | 1-15,19, 21-32 |
| Y | abstract Materials and methods | 16-18, 20,24 |
| | ----- | |
| | -/-- | |

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2011/065611

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | IORNS ELIZABETH ET AL: "Utilizing RNA interference to enhance cancer drug discovery.", NATURE REVIEWS. DRUG DISCOVERY, vol. 6, no. 7, July 2007 (2007-07), pages 556-568, XP002675209, ISSN: 1474-1776 | 1-6, 10-22, 25-30,33 |
| Y | the whole document | 7-9,23, 24,31,32 |
| | --- | |
| X,P | JIANG HAI ET AL: "A mammalian functional-genetic approach to characterizing cancer therapeutics", NATURE CHEMICAL BIOLOGY, vol. 7, no. 2, February 2011 (2011-02), pages 92-100, XP002675210, the whole document | 1-33 |
| | --- | |

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2011/065611

**Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)**

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   a. (means)

   [X] on paper

   [X] in electronic form

   b. (time)

   [ ] in the international application as filed

   [ ] together with the international application in electronic form

   [X] subsequently to this Authority for the purpose of search

2. [X] In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments: