

DIFFERENTIAL PATHWAY DEPENDENCY DISCOVERY ASSOCIATED WITH DRUG RESPONSE ACROSS CANCER CELL LINES^{*}

GIL SPEYER, DIVYA MAHENDRA[†], HAI J. TRAN[†], JEFF KIEFER

*The Translational Genomics Research Institute
Phoenix, AZ 85004, U.S.A.*

Email: gspeyer@tgen.org, mahendradivya@gmail.com, hjtran@brown.edu, jkiefer@tgen.org

STUART L. SCHREIBER, PAUL A. CLEMONS

*Broad Institute of Harvard and MIT
Cambridge MA 02142, U.S.A.*

Email: stuart_schreiber@harvard.edu, pclemons@broadinstitute.org

HARSHIL DHRUV, MICHAEL BERENS, SEUNGCHAN KIM

*The Translational Genomics Research Institute
Phoenix, AZ 85004, U.S.A.*

Email: hdhruv@tgen.org, mberens@tgen.org, skim@tgen.org

The effort to personalize treatment plans for cancer patients involves the identification of drug treatments that can effectively target the disease while minimizing the likelihood of adverse reactions. In this study, the gene-expression profile of 810 cancer cell lines and their response data to 368 small molecules from the Cancer Therapeutics Research Portal (CTRP) are analyzed to identify pathways with significant rewiring between genes, or differential gene dependency, between sensitive and non-sensitive cell lines. Identified pathways and their corresponding differential dependency networks are further analyzed to discover essentiality and specificity mediators of cell line response to drugs/compounds. For analysis we use the previously published method EDDY (Evaluation of Differential DependENCY). EDDY first constructs likelihood distributions of gene-dependency networks, aided by known gene-gene interaction, for two given conditions, for example, sensitive cell lines vs. non-sensitive cell lines. These sets of networks yield a divergence value between two distributions of network likelihoods that can be assessed for significance using permutation tests. Resulting differential dependency networks are then further analyzed to identify genes, termed *mediators*, which may play important roles in biological signaling in certain cell lines that are sensitive or non-sensitive to the drugs. Establishing statistical correspondence between compounds and mediators can improve understanding of known gene dependencies associated with drug response while also discovering new dependencies. Millions of compute hours resulted in thousands of these statistical discoveries. EDDY identified 8,811 statistically significant pathways leading to 26,822 compound-pathway-mediator triplets. By incorporating STITCH and STRING databases, we could construct evidence networks for 14,415 compound-pathway-mediator triplets for support. The results of this analysis are presented in a searchable website to aid researchers in studying potential molecular mechanisms underlying cells' drug response as well as in designing experiments for the purpose of personalized treatment regimens.

^{*} This work was supported in part by the National Cancer Institute, National Institutes of Health [1U01CA168397] and a grant from Dell, Inc. via its *Legacy of Good* program that seeks to put technology and expertise to work where it can do the most for people and the planet.

[†] D. Mahendra and H. Tran were supported by the Helios Education Foundation through the Helios Scholars at TGen summer internship program in biomedical research at the Translational Genomics Research Institute in Phoenix, AZ.

1. Introduction

The effort to personalize treatment plans for patients involves the identification of drug treatments that can effectively target the disease while minimizing the likelihood of adverse reactions. The advent of high-throughput –omics and drug-screening data has given rise to the development of complex analytical approaches to identify biomarkers and drug-targets) [1]. Considering complex molecular mechanisms underlying complex diseases such as cancer, the discovery of such biomarkers and subtype-specific drug targets must be based on activities of multiple genes rather than individual genes. Gene Set Enrichment Analysis (GSEA) [2] is one popular method of testing for differential expression of gene sets between conditions. As pathways are capable of complex rewiring between conditions, network-based analyses have become increasingly attractive for extraction of biological hypotheses from big data [3]. For example, the approaches to identify individual differential dependencies[‡] [4-8] or condition-specific sub-networks from genome-wide dependency networks such as a protein-protein interaction networks have gained much interest [9-11] for the determination of biomarkers and subtype-specific therapeutic vulnerabilities.

Recently, we developed a novel computational method *Evaluation of Differential Dependency* (EDDY) that identifies pathways enriched with differential dependencies and that discovers mediators as potential therapeutic targets. The method has been further improved by incorporating known gene interactions as prior knowledge. The method has been successfully applied to the study of glioblastoma (GBM) [12, 13] and adrenocortical carcinoma (ACC) [14].

In this study, we present results from an integrated analysis of large-scale transcriptomic data of 810 cancer cell lines and large-scale high-throughput screening data of the same cancer cell lines across 368 compounds using EDDY algorithm. The analysis not only identified the pathways enriched with differential dependencies between sensitive and non-sensitive cancer cell lines to each compound, but also discovered mediators as potential novel targets of the compound via graphical analysis of differential dependency networks. Identified compound-pathway-mediator triplets were further queried across known drug-gene database as well as a known gene-gene interaction database to identify corroborating evidence to support newly discovered compound-pathway-mediator triplets. We also developed a searchable website to aid researchers in studying potential molecular mechanisms underlying cells' drug response and in designing experiments for the purpose of personalized treatment regimens, publicly available at <http://biocomputing.tgen.org/software/EDDY/CTRP>.

2. Methods

2.1. High-Throughput Drug Screening of Cancer Cell Lines

The Cancer Cell Line Encyclopedia (CCLE) project is an effort to conduct detailed genetic characterization of a large panel of human cancer cell lines. The CCLE provides public access to DNA copy number, mRNA expression, and mutation data for 1,000 cancer cell lines,

[‡] In this manuscript, we use ‘dependency’ to denote statistical dependencies derived from data such as co-expression, conditional dependencies, and ‘interaction’ to denote *known* relationships between genes or related molecules.

encompassing 36 different tumor types [15].

The Center for the Science of Therapeutics at Broad Institute performed analysis of sensitivity of CCLE cell lines using ~500 small molecules as perturbagens, and made the data available at the Cancer Therapeutics Response Portal (CTRP; <http://www.broadinstitute.org/ctrp/>). The “Informer Set” consists of 481 small compounds, including 70 FDA approved drugs, 100 clinical candidates and 311 small-molecule probes. In this study, we used the transcriptomic profile and CTRP drug-response data to identify pathways with condition-specific rewiring of gene dependencies in the context of drug sensitivity [16, 17]. All of these aforementioned processed data is publicly available on the CTD² data portal (<https://ctd2.nci.nih.gov/dataPortal/>).

2.2. EDDY: Evaluation of Differential Dependency

EDDY is a statistical approach that combines pathway-guided and differential dependency analyses in a probabilistic framework [12, 13]. The algorithm queries each pathway (gene set) in a database such as BioCarta (http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways) or REACTOME [18] to test for differential dependencies across the set of genes between two or more conditions, by comparing gene-dependency networks constructed for each condition. In evaluating differential dependency, EDDY uses a network likelihood distribution over multiple networks constructed via resampling for each condition and compares the distributions between the conditions, instead of just using the single, most probable network from each condition. The statistical significance of

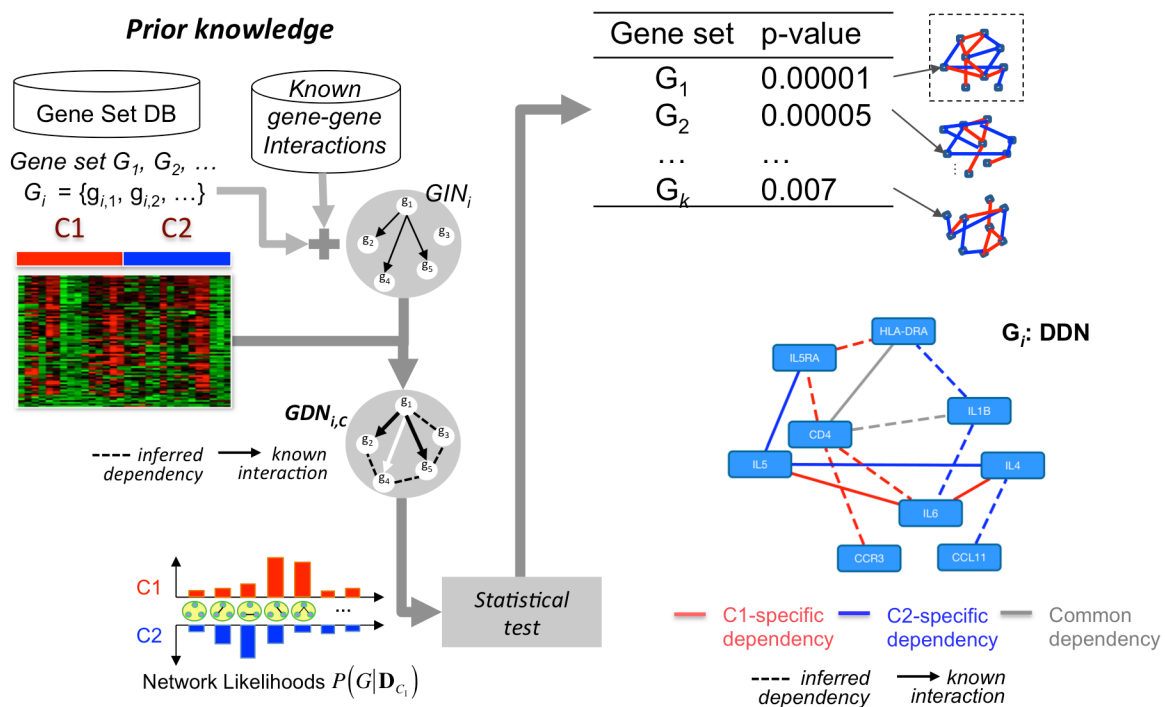


Figure 1. Knowledge-assisted EDDY Workflow. $GDN_{i,C}$ is a gene-dependency network constructed for a gene set G_i , for condition C , aided by gene interaction network GIN_i . A network likelihood distribution over multiple networks is constructed via resampling for each condition and the network score distributions between the conditions are compared. Permutation testing assesses the significance of the divergence between the distributions of scores. Differential dependency networks can then be constructed for statistically significant gene sets.

the divergence is then estimated using asymptotic approximation of Jensen-Shannon divergence based on a beta distribution whose parameters are estimated using a permutation test. Probabilistic and gene-set assisted approaches together contribute to significantly higher sensitivity and specificity of EDDY, compared to other methods, such as GSEA and Gene Set Co-expression Analysis (GSCA) [12].

Incorporation of Prior Knowledge into EDDY: Known interactions from the Pathway Commons 2 (<http://www.pathwaycommons.org>) database are integrated into EDDY as prior knowledge (Figure 1). This integration has been shown to improve the interpretability of results from EDDY. Prior weight (W_p) is specified to determine the degree of weight that is given to the prior knowledge in evaluating new edges to be included in the proposed dependency structure. Since prior knowledge is not condition-specific, large prior weight could decrease EDDY's sensitivity to detect differential dependency while reducing discovery of false-positive dependencies. For this analysis, a prior weight of $W_p = 0.5$ was used, meaning that any edges with half the support from data were included in the dependency network. The choice was based on extensive analysis of various data sets where $W_p = 0.5$ seemed to give the best compromise between sensitivity and false discovery rate when varying prior weight, as reported in Speyer et. al. [13].

2.3. Input Data

Transcriptomic data: BAM files of 935 CCLE cell lines downloaded from the Cancer Genomics Hub (<https://cghub.ucsc.edu>) were converted to a FASTQ format and transcript quantification was performed using Salmon [19] to obtain quantitative estimate of mRNA expression in TPM (transcripts per million). These mRNA expression values were \log_2 transformed and quantized to values -1 (under-expressed), 0 (intermediate), and 1 (over-expressed). For each gene, median average deviation (MAD) was computed and used to determine under-expression ($MAD < -1$), over-expression ($MAD > 1$), and intermediate.

Drug sensitivity: The cell lines were grouped into sensitive and non-sensitive classes using the Small-Molecule Cancer Cell Line Sensitivity Profiling CTRP 2.0 2015 Dataset, acquired from CTD² (Cancer Target Discovery and Development). CTRP summarizes drug sensitivity between each cell line and drug pair using the area-under-percent-viability-curve (AUC) values [16, 17]. We used the 'extremevalues' R package to identify outliers in AUC values and group the cell lines into sensitive (-1; lower-end outliers), non-sensitive (1; upper-end outliers), and intermediate (0; non-outliers) groups for each compound.

In order to conduct a statistically meaningful analysis using EDDY, only those drugs that had at least 50 samples in each sensitive and non-sensitive class were analyzed. This reduced the number of drugs that could be analyzed to 368 drugs.

2.4. Identification of Mediators

For each compound, the results from EDDY analysis (Figure 2) are summarized into 1) a list of pathways enriched with differential dependency of statistical significance, and 2) a differential dependency network (DDN) that captures how gene dependency changes between sensitive and

non-sensitive cell lines. We identified those genes that seemed to play a significantly different role (based on statistical dependencies) between cell lines that were sensitive to a drug and cell lines that were non-sensitive, and termed them as *mediators*.

Essentiality mediators: Each DDN is split into condition-specific dependency networks (CDNs) where each CDN is composed of dependencies manifested in each condition. We then compute between-ness centrality for each gene in both CDNs and compute the difference of the betweenness centrality. The genes with the most differential betweenness centrality are termed *essentiality mediators*, as the genes with highest betweenness centrality in gene regulatory network are often interpreted as essential genes [20].

Specificity mediators: We also analyzed how many dependencies for each gene change between the CDN from sensitive cell lines and the CDN from non-sensitive cell lines. Formally, Let $P_C = E_C / (E_C + E_S)$, a proportion of condition-specific edges (E_C) across the overall number of edges ($E_C + E_S$), and E_{C_i} be the number of condition-specific edges and E_{S_i} be number of shared edges, of a gene i . Note $E_C = \sum_i E_{C_i}$ and $E_S = \sum_i E_{S_i}$. We can then compute the probability, $\Pr(k \geq E_{C_i})$, that a gene i can have E_{C_i} or more condition-specific edges by random chance, via binomial probability $B(k, E_{C_i} + E_{S_i}, P_C)$. If this probability, $\Pr(k \geq E_{C_i}) < 0.05$, we termed gene i as *specificity mediator*.

2.5. Evidence Networks

However, uncertainty in interpreting these drug-pathway-mediator triplets hinders prioritization of hypotheses or experimental design to explore these potentially valuable results. We address this challenge by constructing evidence networks built with protein and drug interactions from the STRING and STITCH interaction databases. STRING and STITCH are sister knowledge-bases that store scored drug-protein interactions and protein-protein interactions, respectively [21, 22]. As compounds can have multiple names, from commercial and generic labels to chemical formula

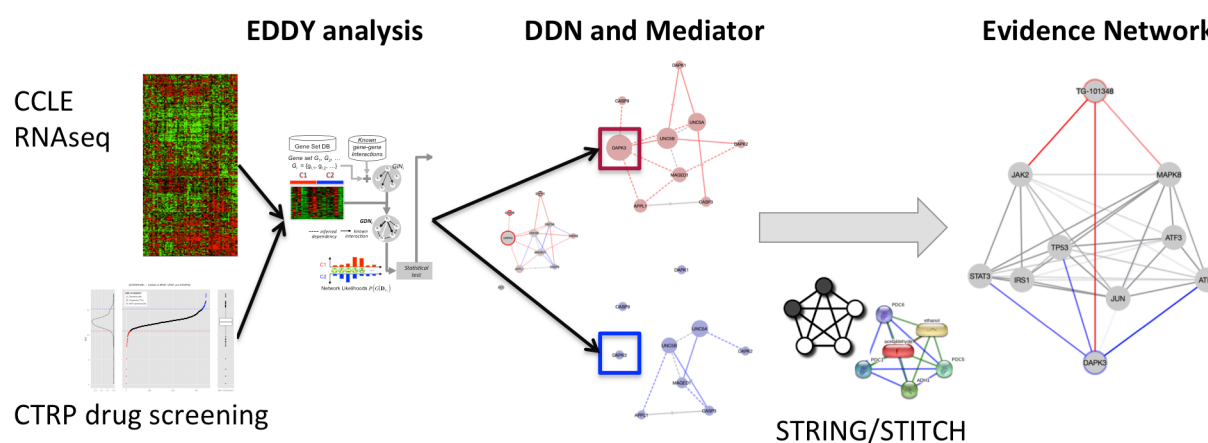


Figure 2. Overall workflow of EDDY analysis of CCLE and CTRP data. EDDY identifies significant pathways from RNA expression and compound-response categorization of cancer cell lines. Graphical analysis of output networks (edge color indicating condition) identifies important genes, termed mediators. Mining knowledge bases yields evidence networks for compound-mediator pairings (edge color here indicating evidence type).

and IUPAC ID, the database employed a unifying InChIKey to maximize comprehensiveness and to avoid false negatives.

Evidence networks were generated using a modified Yen's K -shortest paths algorithm [23] with a weight function of $W(\text{EDGE}) = 1 - \text{EDGE.SCORE}$, so that edges with higher scores would be preferred over edges with lower scores (all scores are within the interval $[0,1]$ and are based on how compelling the supporting evidence is). To generate the evidence networks, shortest paths were continually found and added to the network until there were no more paths from the drug to the gene or there were at least N distinct nodes in the sub-network, where N is some arbitrary threshold. N was not a strict floor as sometimes the last path added to the sub-network would add two or more distinct nodes pushing the total number of distinct nodes over the threshold. Instead, N was used simply as a stopping condition and was chosen in order to prevent generation of evidence networks that would be too overwhelming for users to interpret. Choosing $N = 5$ yielded abundant evidence nets without excessive density. Dijkstra's shortest-path algorithm with a Fibonacci heap was used as the supporting shortest-path algorithm in the modified Yen's K -shortest-paths algorithm [24, 25].

3. Results

3.1. Pathway and Mediator Analysis

EDDY analysis identified a total of 8,811 statistically significant pathways and 26,822 compound-pathway-mediator triplets. Of these, 534 pathways out of 685 BIOCARTA and REACTOME pathways were identified for at least one compound, and 2,401 genes out of 4,298 unique BIOCARTA and REACTOME genes were identified as mediators for at least one compound. On average each compound identified about 24 pathways and 73 mediators. We found that for 125 compounds, EDDY identified pathways that had the compound's intended target in their DDN, and 29 mediators were identified as intended targets. Only 248 out of the 368 compounds had intended targets that EDDY could potentially identify within the REACTOME and BIOCARTA pathways. Hence, EDDY identified pathways that included the intended target for 125 out of 248 compounds (50.4%). We tabulated (Table 1 & Table 2) the top 10 statistically significant pathways and mediators, respectively, which were identified by the largest number of compounds. We can see that the top two pathways that were statistically significant were ERYTH (erythrocyte

Table 1. Top 10 most commonly identified statistically significant pathways that were statistically significant

Pathway	# Compounds	Database
Erythrocyte differentiation (ERYTH)	78	BIOCARTA
Cells and molecules involved in local acute inflammatory response (LAIR)	61	BIOCARTA
CBL mediated ligand-induced downregulation of EGF receptors (CBL)	55	BIOCARTA
TERMINATION OF O GLYCAN BIOSYNTHESIS	52	REACTOME
SIGNALING BY HIPPO	49	REACTOME
NUCLEOTIDE LIKE PURINERGIC RECEPTORS	48	REACTOME
ZINC TRANSPORTERS	46	REACTOME
GRANULOCYTES	45	BIOCARTA
SYNTHESIS OF SUBSTRATES IN N GLYCAN BIOSYNTHESIS	44	REACTOME
PURINE CATABOLISM	43	REACTOME

differentiation pathway) and LAIR (pathway for cells and molecules involved in local acute inflammatory response) from BIOCARTA. The erythrocyte differentiation pathway is the pathway responsible for the formation of red blood cells from the bone marrow. It is expected that this pathway would be altered in hematopoietic cancers and that its alteration would be involved in immune responses. The genes found in this pathway include TGFB2 and cytokines IL1A, IL3, IL6, IL9, and IL11. Cytokines are involved in various immune responses and inflammatory processes. The LAIR pathway includes mechanisms associated with the releases of cytokines IL1A and IL6. The genes IL1A and IL6 are among the top fourteen mediators identified by compounds in EDDY and they are also intended targets for the ERYTH and LAIR pathways. IL1A gene is a cytokine involved in various immune responses, inflammatory processes, and hematopoiesis. This protein is released in response to cell injury. IL6 is also a cytokine that functions in inflammation and maturation of B cells [26]. Indeed, upon further examination of the response data for the compounds differentially dependent for the ERYTH and LAIR pathways, hematopoietic cell lines were on average six times more prevalent in the sensitive versus the non-sensitive groups.

The MAPK signaling pathway is an important signaling pathway in cancer studies because it is altered in many different cancer types and regulates processes such as cell proliferation, cell differentiation, and cell death. MAPK1, MAPK3 and MAPK14 are mitogen-activated protein kinases and are members of the MAP kinase family. These genes act in signaling pathways (MAPK signaling, immune response) and various other cellular processes such as proliferation, differentiation, and cell cycle progression. MAPK14 is activated by environmental stresses and cytokines associated to inflammatory responses. MAP kinases play important roles in cascades of cellular responses and lead to direct activation of transcription factors [27].

3.2. Evidence Network Analysis

EDDY-CTRP analysis identified 26,822 drug-pathway-mediator triplets. Among these pairs, 19,222 of them consisted of a drug or a gene that is contained within the STRING and STITCH databases. Mining STITCH and STRING for each of 19,222 unique compound-pathway-mediator triplets yielded 14,415 evidence networks (~75%) of a path with 3 or fewer intermediate genes. These evidence networks are integrated into the main EDDY-CTRP portal as searchable tables (Table 3).

We note that 102 evidence networks indeed were direct compound and mediator relations, among which only 34 were intended targets defined in the CTRP data and annotation. This indicates

Table 2. The top 10 most commonly identified mediators

Pathway	# Compounds
MAPK1	185
MAPK3	171
GRB2	168
NUP210	158
HRAS	136
NUP37	125
AKT1	120
ORC4	114
MAPK14	114
CDK1	114

Table 3. Distribution of the number of intermediate genes in shortest path between drug and mediator pair.

	Direct targets	Indirect targets		
		# of intermediate genes in shortest path		
		1	2	3
# of pairs	102	988	3,410	9,915

STITCH/STRING contain drug-target relations that were not included in the CTRP database, but EDDY-CTRP analysis was able to discover those relations. Most of these evidence networks were for drug-pathway-mediator triplets where mediators were not direct targets of drug but had some known functional association to the drug (based on STITCH/STRING database). Note that known “hub” genes such as TP53 turned out to have high prevalence in the constructed evidence networks. In future development, the algorithm will introduce weighting to counter this bias.

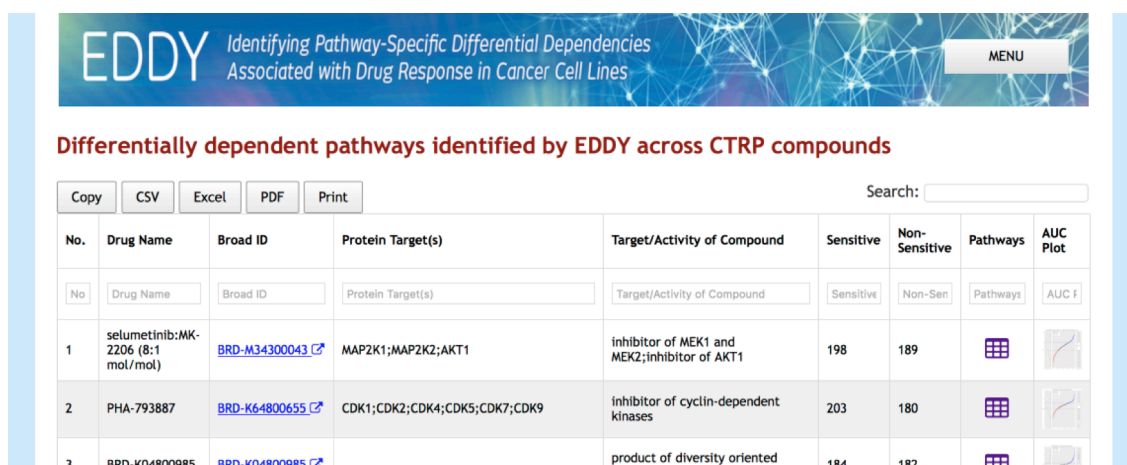
3.3. Interactive and Searchable Web-Portal for EDDY-CTRP Results

The web-portal of the CTRP analysis (<http://biocomputing.tgen.org/software/EDDY/CTRP>) consists of two main views: CTRP compound-centric and mediator-centric. These views provide alternate perspectives on hypothesis-testing data from the EDDY analysis. CTRP compound-centric view (Figure 3) provides pathways enriched with differential dependencies for each of 368 compounds uncovered by EDDY. For each compound, a user can explore each identified pathway, corresponding DDNs, and mediators. Mediator-centric view (Figure 4) lists all compound-pathway-mediator triplets uncovered across all compounds and all identified pathways. For each triplet, a user can also explore evidence networks as well as corresponding DDNs and pathways.

4. Case Studies: Potential Alternative Drug Targets

4.1. DAPK3 as an Alternative Target for TG-101348

TG-101348 was developed as a selective inhibitor of JAK2 kinase for the treatment of myeloproliferative disorder [28]. EDDY identified 29 pathways significantly enriched with differential dependency, and 66 mediators. One of the pathways is the EPONFKB pathway, which has JAK2 as an identified mediator, and, examining this DDN, JAK2 has exclusively sensitive-specific edges. We obtained the evidence networks for 59 of 66 mediators, and one of those mediators with evidence network is DAPK3 which is identified as a direct target of TG-101348, based on STITCH database. DAPK3 was identified as a mediator for the "ROLE OF DCC IN REGULATING APOPTOSIS" pathway which has an altered differential dependency



EDDY Identifying Pathway-Specific Differential Dependencies Associated with Drug Response in Cancer Cell Lines

Differentially dependent pathways identified by EDDY across CTRP compounds

Copy CSV Excel PDF Print Search:

No.	Drug Name	Broad ID	Protein Target(s)	Target/Activity of Compound	Sensitive	Non-Sensitive	Pathways	AUC Plot
1	selumetinib:MK-2206 (8:1 mol/mol)	BRD-M34300043	MAP2K1;MAP2K2;AKT1	inhibitor of MEK1 and MEK2;inhibitor of AKT1	198	189		
2	PHA-793887	BRD-K64800655	CDK1;CDK2;CDK4;CDK5;CDK7;CDK9	inhibitor of cyclin-dependent kinases	203	180		
3	BRD-K04800985	BRD-K04800985		product of diversity oriented	184	182		

Figure 3. CTRP compound-centric view

Mediators identified by EDDY for CTRP compounds

Copy CSV Excel PDF Print Search:

No.	Drug Name	Broad ID	Intended Target	Protein Target	Pathway	P-Value	Mediator	Type	DDN	Ev Net
1	selumetinib:MK-2206 (8:1 mol/mol)	BRD-M34300043	inhibitor of MEK1 and MEK2;inhibitor of AKT1	MAP2K1;MAP2K2;AKT1	CHYLOMICRON MEDIATED LIPID TRANSPORT	3.94e-05	SDC1	Essentiality		
2	selumetinib:MK-2206 (8:1 mol/mol)	BRD-M34300043	inhibitor of MEK1 and MEK2;inhibitor of AKT1	MAP2K1;MAP2K2;AKT1	CHYLOMICRON MEDIATED LIPID TRANSPORT	3.94e-05	HSPG2	Essentiality		
3	selumetinib:MK-2206 (8:1 mol/mol)	BRD-M34300043	inhibitor of MEK1 and MEK2;inhibitor of AKT1	MAP2K1;MAP2K2;AKT1	TRANSPORT OF ORGANIC ANIONS	0.0024	SLCO2A1	Essentiality		

Figure 4. CTRP mediator-centric view

network for TG-101348. The gene product of DAPK3 was a mediator in this pathway due to high change of essentiality (betweenness centrality) between the condition-specific dependency networks TG-101348 sensitive cancer cell lines and non-sensitive cancer cell lines. In TG-101348-sensitive cell lines, DAPK3 is highly connected in the network (Figure 5a), consistent with DAPK3 playing a central role in a functioning apoptotic network. In the non-sensitive cell lines, however, DAPK3 is not connected to the rest of the network (Figure 5b), corroborating the indication that disconnected DAPK3 may confer insensitivity to TG-101348 sensitivity.

The evidence network built for TG-101348 - DAPK3 supports this hypothesis by showing a direct association between TG-101348 and DAPK3, discovered from the STITCH database (Figure 5c). Indeed, the evidence link was from a study that showed TG-101348 can inhibit the kinase activity of DAPK3, indicating that TG-101348 actually does target DAPK3 in addition to JAK2. Additionally, an association between the downstream JAK2 modulator and DAPK3 was revealed suggesting further signaling interactions targeted by TG-101348 [29]. So, while this target was not annotated in CTRP annotation for known targets of TG-101348, EDDY-CTRP

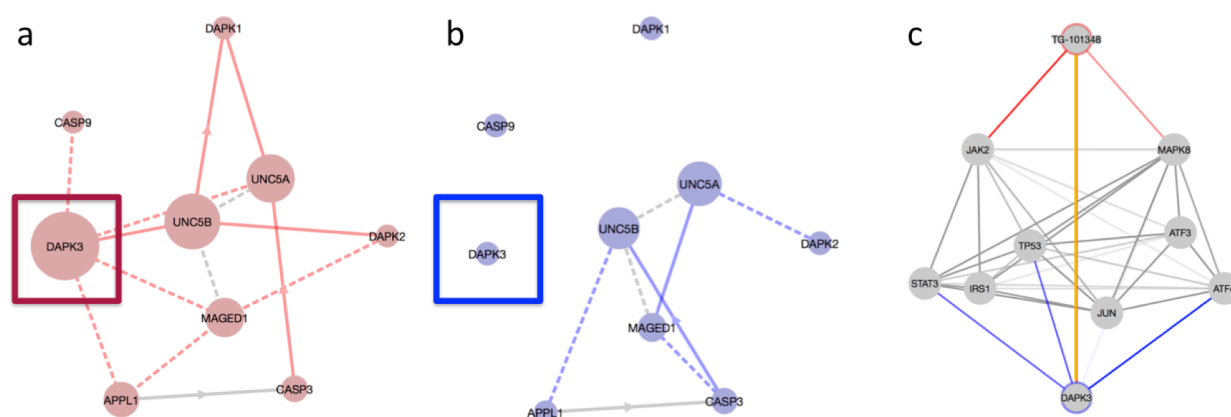


Figure 5. (a) Condition-specific dependency network (CDN) for TG-101348-sensitive cell lines. Dashed lines represent statistical dependencies while solid lines known interactions. Size of nodes represents node essentiality. (b) CDN for TG-101348-insensitive cell lines. (c) Evidence network for the TG-101348 – DAPK3 drug-mediator pair. All edges represent a known association based from the STRING/STITCH databases. Blue edges represent mediator-gene associations, red edges drug-gene associations, and yellow edge a direct drug-mediator association.

analysis was able to detect this relationship. This example illustrates EDDY can discover potentially novel targets of a compound and how the evidence network provides further contextual information regarding the possible mechanisms of how mediators selected in the EDDY analysis function to alter individual drug responses.

4.2. *HIF1A as an Alternative Target for Indisulam*

Indisulam is a carbonic anhydrase IX (CA9) inhibitor [30]. CA9 activity in cancer is associated with an acidic microenvironment that favors tumor cell survival and growth [31]. EDDY identified the HIF pathway as a DDN associated with indisulam response. The HIF pathway is important for cancer-cell survival in hypoxic conditions often seen in tumors [32]. In the non-responsive HIF pathway DDN two genes, HIF1A and JUN exhibit high essentiality compared to the responsive HIF DDN (Figure 6a). HIF1A is a major gene that signals for cell survival in hypoxic conditions [32]. The evidence network for indisulam and HIF1A reveals a direct link between CA9 and HIF1A (Figure 6c). This would not be evident if investigator had only HIF pathway DDN evidence. Inspection of the evidence from STRING shows that HIF1A positively regulates CA9 expression. Cancer cells may be non-responsive to indisulam because HIF1A increases CA9 levels such that the drug is not effective at tested concentration in fully inhibiting CA9. This example shows how the evidence network is able to mechanistically link EDDY DDNs to drug targets and expand understanding of signaling events associated with drug response.

5. Conclusions

While the current CTRP dataset allows the study of the correlations between genetic features with sensitivity to compounds, and while there are previous studies associating genes with compound sensitivity [33], this paper presents an unprecedented identification of pathways with differential dependency networks across a large number of cancer cell lines with drug-screening data. Additionally we have created a web repository to allow clinicians and researchers to view the results of our analysis. The web repository provides an interactive method to view the results for

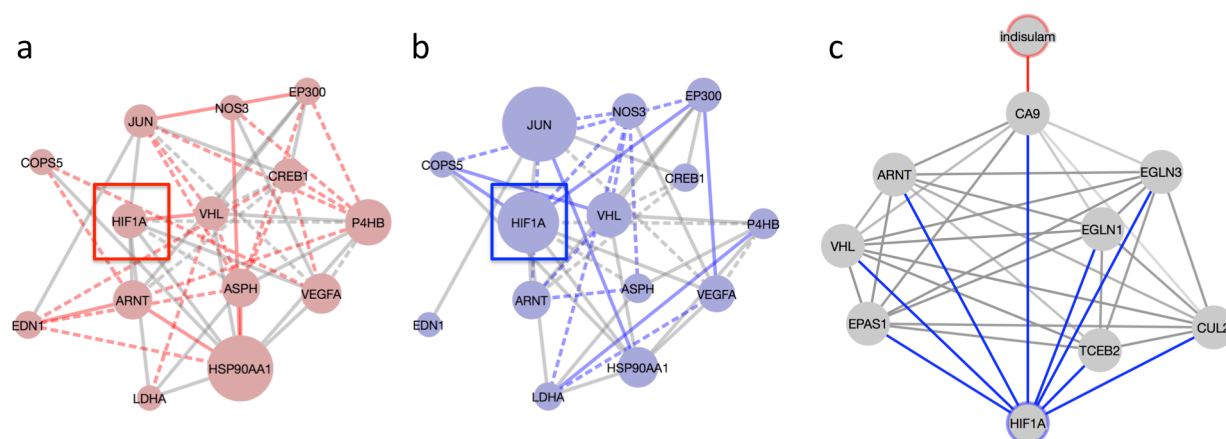


Figure 6. (a) Condition-specific dependency network (CDN) for indisulam for drug-sensitive cell lines. (b) CDN for indisulam for drug-insensitive cell lines. (c) Evidence network for the indisulam – HIF1A drug-mediator pair.

specific drugs. Researchers can query the intended targets, genes, or pathways to identify types of drugs, known targets, and to discover hitherto unknown mediators. We integrated quick unique links to the CTRP database, MSigDB Database, and Gene Cards, for each of the compounds, pathways, and genes. These links allow users to view the analysis and information about the drug, pathway, or gene seamlessly. We also provide links to the interactive DDN and condition-specific CDNs so that users can move around the nodes and edges to better analyze the results. In addition we provide links to generate the Oncoprints for the sensitive and non-sensitive cell lines for each DDN. These links allow the users to look at the mutation data used to generate the DDN.

This resource can be valuable for researchers to explore potential targets of their interest and allow them to look at differential dependencies across a large number of cell lines and compounds. It may aid in studying potential molecular mechanisms underlying cells' response to drug as well as designing experiments for the purpose of personalized treatment regimens.

Computational methods that can efficiently predict the effectiveness of drugs based on the genetic makeup of tumors would provide a major breakthrough towards personalized therapy for cancer patients based on their tumor's molecular markers. To strengthen the validity of our analysis and resource, experimental validation of the pathways identified by EDDY is warranted. We anticipate that this web repository will be a living resource for clinicians and researchers to use for designing experiments and identifying potential personalized treatment regimens.

References

1. Roden, D.M. and A.L. George, Jr., *The genetic basis of variability in drug responses*. Nat Rev Drug Discov, 2002. **1**(1): p. 37-44.
2. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15545-50.
3. Califano, A., *Rewiring makes the difference*. Mol Syst Biol, 2011. **7**: p. 463.
4. Lai, Y., et al., *A statistical method for identifying differential gene-gene co-expression patterns*. Bioinformatics, 2004. **20**(17): p. 3146-3155.
5. Hu, R., et al., *Detecting intergene correlation changes in microarray analysis: a new approach to gene selection*. BMC Bioinformatics, 2009. **10**(1): p. 20.
6. Mentzen, W., M. Floris, and A. de la Fuente, *Dissecting the dynamics of dysregulation of cellular processes in mouse mammary gland tumor*. BMC Genomics, 2009. **10**(1): p. 601.
7. Zhang, B., et al., *Differential dependency network analysis to identify condition-specific topological changes in biological networks*. Bioinformatics, 2009. **25**(4): p. 526-32.
8. Zhang, B., et al., *DDN: a caBIG(R) analytical tool for differential network analysis*. Bioinformatics, 2011. **27**(7): p. 1036-8.
9. Hwang, T. and T. Park, *Identification of differentially expressed subnetworks based on multivariate ANOVA*. BMC Bioinformatics, 2009. **10**(1): p. 128.
10. Kim, Y., et al., *Principal network analysis: Identification of subnetworks representing major dynamics using gene expression data*. Bioinformatics, 2010.
11. Ma, H., et al., *COSINE: COndition-Specific sub-NEtwork identification using a global optimization method*. Bioinformatics, 2011.
12. Jung, S. and S. Kim, *EDDY: a novel statistical gene set test method to detect differential genetic dependencies*. Nucleic Acids Res, 2014. **42**(7): p. e60.

13. Speyer, G., et al., *Knowledge-Assisted Approach to Identify Pathways with Differential Dependencies*. Pac Symp Biocomput, 2016. **21**: p. 33-44.
14. Zheng, S., et al., *Comprehensive Pan-Genomic Characterization of Adrenocortical Carcinoma*. Cancer Cell, 2016. **29**(5): p. 723-36.
15. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity*. Nature, 2012. **483**(7391): p. 603-7.
16. Seashore-Ludlow, B., et al., *Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset*. Cancer Discov, 2015. **5**(11): p. 1210-23.
17. Rees, M.G., et al., *Correlating chemical sensitivity and basal gene expression reveals mechanism of action*. Nat Chem Biol, 2016. **12**(2): p. 109-16.
18. Fabregat, A., et al., *The Reactome pathway Knowledgebase*. Nucleic Acids Res, 2016. **44**(D1): p. D481-7.
19. Patro, R., G. Duggal, and C. Kingsford, *Accurate, fast, and model-aware transcript expression quantification with Salmon*. biorxiv, 2015.
20. Khuri, S. and S. Wuchty, *Essentiality and centrality in protein interaction networks revisited*. BMC Bioinformatics, 2015. **16**: p. 109.
21. Kuhn, M., et al., *STITCH 4: integration of protein-chemical interactions with user data*. Nucleic Acids Res, 2014. **42**(Database issue): p. D401-7.
22. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life*. Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
23. Yen, J., *Finding the K Shortest Loopless Paths in a Network* Management Science, 1971. **17**(11): p. 712-716.
24. Dijkstra, E.W., *A note on two problems in connexion with graphs*. Numerische Mathematik, 1959. **1**(4): p. 269-271.
25. Fredman, M.L. and R.E. Tarjan, *Fibonacci heaps and their uses in improved network optimization algorithms*. Journal of the Association for Computing Machinery, 1987. **34**(3): p. 596-615.
26. Seruga, B., et al., *Cytokines and their relationship to the symptoms and outcome of cancer*. Nat Rev Cancer, 2008. **8**(11): p. 887-99.
27. Lei, Y.Y., et al., *Mitogen-activated protein kinase signal transduction in solid tumors*. Asian Pac J Cancer Prev, 2014. **15**(20): p. 8539-48.
28. Wernig, G., et al., *Efficacy of TG101348, a selective JAK2 inhibitor, in treatment of a murine model of JAK2V617F-induced polycythemia vera*. Cancer Cell, 2008. **13**(4): p. 311-20.
29. Sato, N., et al., *Physical and functional interactions between STAT3 and ZIP kinase*. Int Immunol, 2005. **17**(12): p. 1543-52.
30. Supuran, C.T., *Indisulam: an anticancer sulfonamide in clinical development*. Expert Opin Investig Drugs, 2003. **12**(2): p. 283-7.
31. Swietach, P., et al., *New insights into the physiological role of carbonic anhydrase IX in tumour pH regulation*. Oncogene, 2010. **29**(50): p. 6509-21.
32. Masson, N. and P.J. Ratcliffe, *Hypoxia signaling pathways in cancer metabolism: the importance of co-selecting interconnected physiological pathways*. Cancer Metab, 2014. **2**(1): p. 3.
33. Gottlieb, A. and R.B. Altman, *Integrating systems biology sources illuminates drug action*. Clin Pharmacol Ther, 2014. **95**(6): p. 663-9.