SCIENTIFIC REPORTS

OPEN

Characterizing Cancer Drug Response and Biological Correlates: A Geometric Network Approach

Received: 11 December 2017
Accepted: 9 April 2018
Published online: 23 April 2018

Maryam Pouryahya[1,2], Jung Hun Oh[2], James C. Mathews[2], Joseph O. Deasy[2] & Allen R. Tannenbaum[1,3]

In the present work, we apply a geometric network approach to study common biological features of anticancer drug response. We use for this purpose the panel of 60 human cell lines (NCI-60) provided by the National Cancer Institute. Our study suggests that mathematical tools for network-based analysis can provide novel insights into drug response and cancer biology. We adopted a discrete notion of Ricci curvature to measure, via a link between Ricci curvature and network robustness established by the theory of optimal mass transport, the robustness of biological networks constructed with a pre-treatment gene expression dataset and coupled the results with the GI50 response of the cell lines to the drugs. Based on the resulting drug response ranking, we assessed the impact of genes that are likely associated with individual drug response. For genes identified as important, we performed a gene ontology enrichment analysis using a curated bioinformatics database which resulted in biological processes associated with drug response across cell lines and tissue types which are plausible from the point of view of the biological literature. These results demonstrate the potential of using the mathematical network analysis in assessing drug response and in identifying relevant genomic biomarkers and biological processes for precision medicine.

In this paper, we propose the use of certain tools from discrete geometry to gain new insights into cancer drug response. For this purpose, we tested our methodology on a panel of 60 human cancer cell lines (NCI-60). It has been more than 30 years since the U.S. National Cancer Institute (NCI) established a human cell line panel for the purpose of discovering novel cancer drugs. The NCI-60 panel was designed to recast the previous murine-based drugs from leukemia treatment to the treatment of more diverse human solid tumors. This departure was due to the difference and diversity of the biology of human tumors from murine leukemia[1]. This panel was developed as part of the NCI's Developmental Therapeutics Program (DTP, http://dtp.nci.nih.gov) to screen *in vitro* response to over 100,000 chemical compounds and natural products including FDA-approved anti-cancer drugs and those currently undergoing clinical trial. This ongoing service is accepting global submissions and continues screening up to 3,000 small molecules per year as potential anti-cancer therapies. The NCI-60 panel represents nine tumor types: leukemia, breast, central nervous system, colon, skin, lung, ovarian, prostate, and renal cancers. The NCI-60 panel is thus an established tool for *in vitro* drug screening and has significantly improved the philosophy and research of human cancer drugs[2]. This panel has led to many important discoveries, including a general advance in the understanding of the mechanism of cancer and the action of drugs[3,4]. Moreover, comprehensive genomic data including transcript expression data, protein expression data, sequencing (mutation) data, DNA copy number, and methylation as well as drug screening data on the 60 cell lines make it a unique resource for system pharmacogenomics and systems biology[5]. Most importantly for our work, this data resource enables us to explore both pre-treatment genomic data and drug responses of a notable number of FDA approved anticancer agents (~130) which is unmatched by any other cancer databases[1].

We are interested in analyzing the gene interaction network for these cell lines via mathematical tools. In recent years, there have been tremendous efforts to elucidate the complex mechanisms of biological networks by investigating the interactions of different genetic and epigenetic factors. Given that gene/protein interactions inherently form a mathematical network, it is reasonable to expect that mathematical tools can facilitate a better

[1]Department of Applied Mathematics & Statistics, Stony Brook University, Stony Brook, 11794, USA. [2]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, 10064, USA. [3]Department of Computer Science, Stony Brook University, Stony Brook, 11794, USA. Correspondence and requests for materials should be addressed to A.R.T. (email: allen.tannenbaum@stonybrook.edu)

understanding of the complexities of such networks[6]. The methods and tools employed in network analyses are quite diverse and heterogeneous, ranging from graph theory for abstract representation of pairwise interactions to complicated systems of partial differential equations that try to capture all details of biological interactions. The ability to sequence and analyze genomes has revolutionized the diagnosis and treatment of diseases. With the exponential growth of genomic data, the need for improved mathematical methods to analyze the data is becoming even more prevalent. Here, we adopt a discrete mathematical notion of curvature defined on networks to study the robustness of gene interaction networks in response to drugs. We rank the effectiveness of anticancer drugs and find the biological processes in which the important genes are involved. The network based analysis of 'omics' allows identification of new disease genes, pathways and rational drug targets that were not easily detectable by isolated gene analysis. This study illustrates the use of a novel mathematical approach to networks to identify pertinent biological processes as well as effective drugs for the treatment of cancer.

The mathematical notions can be summarized as follows. Ricci curvature is a fundamental concept in Riemannian geometry; see[7,8] for all of the details. Here, we use an analogous notion on discrete spaces, namely, Olivier-Ricci curvature[9,10]. The concept of curvature was initially introduced to express the deviation of a geometric object from being flat. The Riemann curvature tensor of such a manifold encodes key geometric properties and expresses the deviation from Euclidean (flat) space. The sectional curvature is defined on two-dimensional subspaces of the tangent planes, and Ricci curvature is the average of sectional curvatures of all tangent planes containing some given direction[7]. Interestingly, Ricci curvature also appears in optimal mass transport theory[11,12], and serves as the motivation for certain discrete analogues. Indeed, on a Riemannian manifold, one can endow the space of probability densities with a natural Riemannian structure[13,14] employing the 2-Wasserstein distance from optimal mass transport[15]. Thus, given the Riemannian-type metric, one can define a notion of geodesics on the space of probability densities. As noted by Lott-Sturm-Villani[16–18], considering this Riemannian structure, one can relate the Ricci curvature of the underlying manifold, the entropy of densities along a given geodesic path, and the 2-Wasserstein distance in one remarkable formula (see our discussion below for the details). In conjunction with the Fluctuation Theorem[19], we can conclude that increases in the Ricci curvature are positively correlated with increases in the robustness, herein expressed as $\Delta \mathrm{Ric} \times \Delta R \geq 0$. Following our previous work[20–22], we are interested in finding important nodes (genes) within the network in terms of robustness.

Coupling the results of the network analysis with the drug growth inhibition values provides us with a network-based guide to the sensitivity/resistance of the tumor cell lines to these drugs. The hypothesis behind this idea is that robustness in the biological network contributes to tumor drug resistance, thereby enabling us to predict the effectiveness and sensitivity of drugs in the cell lines. Here, due to some missing values, we focus on a subset of 58 cell lines. The transcription expression data provided for these cell lines along with the gene-to-gene relationships enables us to construct a weighted network. Investigating this network and relating the information it provides to the drug response gives a novel insight into the NCI-60 database which has not been studied using a network mathematical approach before.

Our main results are based on the application of the aforementioned discrete notion of Ricci curvature to the network generated from the pre-treatment gene expression for all 58 cell lines. This notion allows us to identify possible targets for the anti-cancer drugs. For a given drug, we find the average Ricci curvature of the genes whose expressions are significantly correlated to the response of the drug. Using this we identify which part of the network is most correlated to a specific drug's action. The average Ricci curvature for this subnetwork can act as a guide to the sensitivity/resistance of cell lines to the drug. Specifically, a higher degree of robustness for the subnetwork identifies resistance to the drug along the tested cell lines. We are also interested in the biological processes in which the significant genes of effective drugs are involved. This can help us to detect key biological processes associated with the drug response.

The results in the present work are all derived from the network analysis of the NCI-60 genomic information. In our work, we utilized geometric tools in discrete mathematics to better understand these complex networks. This point of view can help to elucidate important drug stratification and biomarkers, which in turn may allow researchers to glean new clinical information from the NCI-60 database.

Methods

Background on curvature.

In the present study, we employed an analogue of Ricci curvature to analyze cancer protein expression networks. In the classical continuous setting, the Ricci curvature tensor provides a way of measuring the degree to which the geometry determined by a given Riemannian metric differs from that of ordinary Euclidean space; see[7] for all the details. We briefly sketch the main ideas to motivate the discrete definition applicable to the networks of interest.

Assume that M is a complete connected Riemannian manifold equipped with metric g. On such spaces, one has a notion of *geodesic*, namely curves which locally travel the shortest distance between points. Geodesics generalize the notion of straight lines in Euclidean space. Let $x, y \in M$ be two very close points defining tangent vector (xy). Moreover, let ω be a tangent vector at x and ω' be the tangent vector at y obtained by parallel transport of w along (xy).

Positivity for the Ricci curvature near x or y is characterized by the fact that the trajectories of the geodesics corresponding to ω and ω' will approach each other. This can be compared to the situation of the traditional flat geometry of Euclidean space, where such geodesics are always equidistant from each other with their "direction" being unchanged by parallel transport. Equivalently, this positivity may be formulated by the fact that the average distance between two small geodesic balls is less than the distance of their centers. Ricci curvature along the direction (xy) quantifies this, averaged on all directions w at x. On the other hand, when the curvature is negative, the geodesics diverge. Lower bounds on the Ricci curvature prevent geodesics from diverging too fast and geodesic balls from growing too quickly in volume[7]. In other words, lower Ricci curvature bounds estimate the tendency of geodesics to converge.

Interestingly, optimal transport offers a lower bound on the Ricci curvature in terms of entropy[17,18]. *This fact will be exploited below to show that on a weighted graph one may use curvature as a proxy for functional robustness.* Optimal mass transport theory is concerned with the problem of finding an optimal transport plan (relative to some cost function) for moving a given initial mass distribution (or gene expression levels in our case) μ into a final configuration ν in a mass preserving manner[11,12,16,23]. We will assume that μ and ν are normalized to be probability measures. Let $(\mathscr{X}, d)$ denote a metric measure space. Then the p-Wasserstein distance between μ and ν is defined as

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathscr{X} \times \mathscr{X}} d(x, y)^p d\pi(x, y) \right)^{1/p},$$

where the latter infimum is taken over all joint probability measures π on $\mathscr{X} \times \mathscr{X}$ whose marginals are μ and ν, i.e.:

$$\forall\, U, V \in X \quad \mu(U) = \pi(U \times X), \quad \nu(V) = \pi(X \times V).$$

Consider the case $\mathscr{X}$ is a Riemannian manifold. The Wasserstein distance defines a Riemannian distance function on the space of the probability measures on $\mathscr{X}$[13,14]. We denote this space by $P_2(\mathscr{X}) := (P(\mathscr{X}), W_2)$. Using the theory of optimal transport, Lott, Sturm and Villani[17,18] derived an elegant connection between Ricci curvature, Ric, and the Boltzmann entropy, Ent. Namely, Ric $\geq k$ if and only if the entropy functional is displacement k-concave along the 2-Wasserstein geodesics, i.e. for all $\mu_0, \mu_1 \in P_2(\mathscr{X})$ and $t \in [0, 1]$ we have:

$$\mathrm{Ent}(\mu_t) \geq (1 - t)\,\mathrm{Ent}(\mu_0) + t\,\mathrm{Ent}(\mu_1) + k\frac{t(1 - t)}{2} W_2(\mu_0, \mu_1)^2. \tag{1}$$

Note that by definition,

$$\mathrm{Ent}(\mu) := -\int_X \rho \log \rho \ \mathrm{dvol},$$

where $\rho = d\mu/\mathrm{dvol}$.

Robustness defined on networks.

The relation (1) indicates a positive correlation between changes in entropy and changes in Ricci curvature that we express as

$$\Delta\mathrm{Ent} \times \Delta\mathrm{Ric} \geq 0. \tag{2}$$

We will describe notions of Ricci curvature and entropy on graphs below. Here, we show that changes in *robustness*, i.e., the ability of a system to functionally adapt to changes in the environment (denoted as ΔR) is also positively correlated with entropy[24], and thus with network curvature (2):

$$\Delta R \times \Delta\mathrm{Ric} \geq 0. \tag{3}$$

More precisely, the measure of robustness employed in[19] is the *rate function*, R, from the theory of large deviations[25]. One considers random perturbations of a given network that result in deviations of some observable. We let $p_\varepsilon(t)$ denote the probability that the mean of the observable deviates by more than ε from the original (unperturbed) value at time t. Since $p_\varepsilon(t) \to 0$, we want to measure its relative rate, that is, we set

$$R := \lim_{t\to\infty} \left(-\frac{1}{t} \log p_\varepsilon(t) \right).$$

Therefore, large R means not much deviation and small R corresponds to a large deviation. In thermodynamics, it is well-known that entropy and rate functions from large deviations are very closely related[19]. The Fluctuation Theorem is an expression of this fact for networks, and may be written as

$$\Delta\mathrm{Ent} \times \Delta R \geq 0. \tag{4}$$

In classical thermodynamics, the entropy controls the asymptotics of thermodynamical limits, and states of highest probability have maximum entropy.

The Fluctuation Theorem has consequences for just about any type of network: biological, communication, social, or neural[19]. In rough terms, it means that the ability of a network to maintain its functionality in the face of perturbations (internal or external) can be quantified by the correlation of activities of various elements that comprise the network. In the standard statement, this correlation is given via entropy. This has been reformulated geometrically in terms of curvature[21,22]. In fact, the correlations (2) and (4) yield (3). Therefore, by calculating the Ricci curvature, we can measure the robustness of the networks. We provide some examples of known graph models in the Supplementary Information. In all cases, the graphs that are expected to be more robust possess higher average Ricci curvatures (Supplementary Figs S2, S3). We will now give the precise definition for networks modeled as weighted graphs.

Curvature on weighted graphs.

In discrete settings, we assume that our network is represented by an undirected and positively weighted graph, $G = (V, E)$, where V is the set of n vertices (nodes) in the network and E is the set of edges. We set

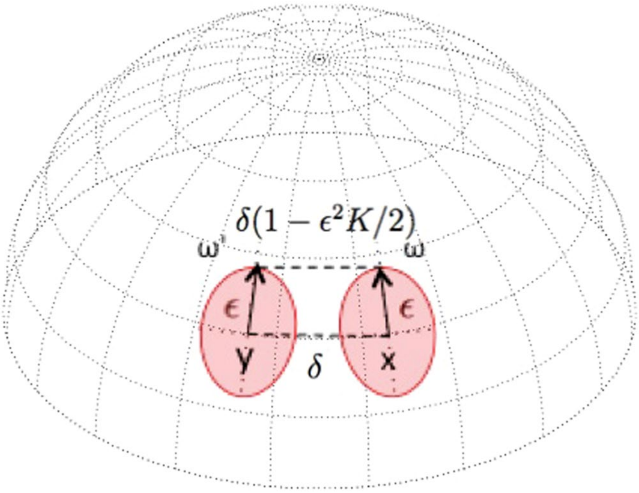

Figure 1. In a positively curved space the distance between the end points of tangent vectors ω and ω' is less than δ. Curvature (K) quantifies this difference.

$$\begin{aligned} d_x &:= \sum_z w_{xz}, \\ \mu_x(y) &:= \frac{w_{xy}}{d_x}, \end{aligned} \tag{5}$$

where the sum is taken over all neighbors z of x, and w_{xy} denotes the weight of an edge connecting x and y (it is taken as zero if there is no connecting edge between x and y).

One of the key notions of Ricci curvature on a discrete metric measure space is the *Ollivier-Ricci curvature*. As discussed by Ollivier[10] and indicated in Fig. 1, the Ricci curvature of a Riemannian manifold can be characterized by comparing the average distance between small geodesic spheres and the distance between their centers. Ollivier then extended this idea from the geodesic sphere to an associated probability measure near a point on a metric space $\mathcal{X}$.

Consider any metric $d: V \times V \to \mathbb{R}^+$ on the set of vertices V. For example, $d(x, y)$ may denote the number of edges in the shortest path connecting x and y. For any two distinct points $x, y \in V$, the *Ollivier-Ricci (OR) curvature* is defined as follows:

$$k(x, y) := 1 - \frac{W_1(\mu_x, \mu_y)}{d(x, y)},$$

where μ_x, μ_y are defined in (5). In the present study, we used the Hungarian algorithm[26] to compute the Earth Mover Distance on our reference networks. This discrete notion of Ricci curvature has been already used to investigate the robustness of cancer networks[21,22].

Using this edge based notion of curvature, we can also define the *scalar curvature* of a given node in the graph as follows:

$$S_{OR}(x) := \sum_y k(x, y),$$

where the sum is taken over all neighbors of x.

Gene expression and drug activity data.

The summary of our methodology is shown in Fig. 2. The mRNA expression data for the NCI-60 human tumor cell lines were retrieved from the CellMiner web application (http://discover.nci.nih.gov/cellminer). Cellminer, written by the Genomics & Bioinformatics Group, (LMP, CCR,NCI)[5], provides freely accessible analysis tools and downloadable data sets for exploring NCI-60 data. The database contains transcript expression values for several assays of the NCI-60 cell lines. This study utilizes Affymetrix HG-U133 (A-B) with GeneChip RMA (GC-RMA) normalization from this website. There was no expression information of Affymetrix HG-U133 (A-B) for LC:NCI-H23 in non-small cell lung cancer (NSCLC), and there were many missing GI50 values for the drug responses of ME:MDA-N of the melanoma cell line. Excluding these two cell lines resulted in 58 complete data sets (see Fig. 3(a)).

Using the gene expressions arrays, we found the gene-to-gene correlations to build our weighted networks. The underlying topology has been derived from Human Protein Reference Database (HPRD, http://www.hprd.org)[27]. Specifically, we took the intersection of the genes that appear in both HPRD data and the gene expression data, and then retained the largest connected component. The weights of the edges, however, come from NCI-60 gene expressions. Even though our method is robust to network topologies, we chose the HPRD due to its reliability (a manually

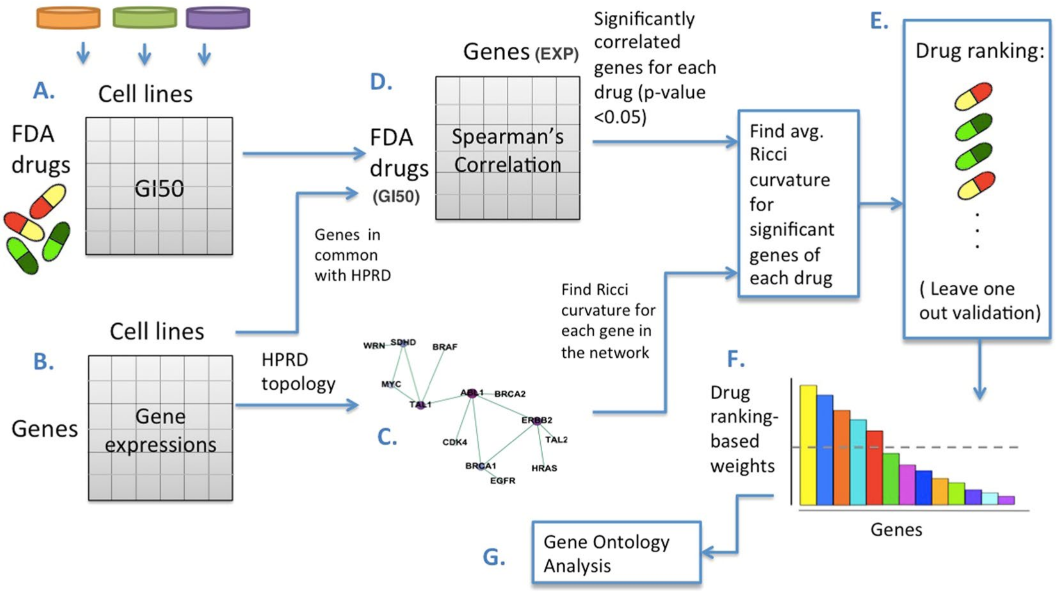
A.
Cell lines
FDA drugs
GI50
D.
Genes (EXP)
FDA drugs (GI50)
Spearman's Correlation
Significantly correlated genes for each drug (p-value <0.05)
Find avg. Ricci curvature for significant genes of each drug
E.
Drug ranking:
(Leave one out validation)
Genes in common with HPRD
B.
Cell lines
Genes
Gene expressions
HPRD topology
C.
WRN
SDHD
BRAF
MYC
TAL1
ABL1
BRCA2
ERBB2
TAL2
CDK4
BRCA1
HRAS
EGFR
Find Ricci curvature for each gene in the network
F.
Drug ranking-based weights
Genes
G.
Gene Ontology Analysis

Figure 2. Methodology for establishing a network-robustness ranking of genes across cancer drugs and cancer cell lines: (**A**) GI50 drug activity matrix of 129 drugs for 58 cell lines. (**B**) Matrix of 8240 gene expressions for 58 cell lines. (**C**) Pre-treatment network made by the gene expression correlation along 58 cell lines as the weights, and the underlying topology of gene-to-gene interactions is derived from HPRD. (**D**) Matrix of Spearman's correlations between each drug's activity (rows of matrix A.) and gene expression (rows of matrix B.) along 58 cell lines. (**E**) Drug ranking in ascending order of average Ricci curvature values of significant genes. (**F**) Drug ranking used to score the significant genes correlated to the drugs. (**G**) Top 200 genes selected for gene ontology enrichment analysis.

curated database) compared to other databases[28]. Affymetrix HG-U133 (A-B) data contains 34,899 probes. For the repeated gene IDs, we used the average RNA expression of the corresponding probes. We used only probes with known gene names, resulting in 16,821 genes with RNA expression. We chose the intersection of these genes with the HPRD database as the nodes of the pre-treatment network. Overall, the network consists of 8240 genes. The weights of the edges are defined by the Pearson correlation between the gene expressions along the 58 cell lines. We further used the transformation of $\frac{(1 + corr(i,\ j))}{2}$ for the genes i and j to make a positively weighted network. The main advantage of using this affine transformation as compared to taking the absolute values of the correlations is that the transformation is *invertible*, and thus does not result in any loss of information of the weighted network. This is not true of the absolute value. In practice, however, using either absolute value or the affine transformation seems to give similar results[21,24]. We then found the Ollivier-Ricci curvature for all the edges, and the scalar curvature for all the nodes (genes) within the pre-treatment network. We further identified significant genes (nodes) for each drug in this network. These important genes were selected based on significant Spearman's correlations (*Rs*) values (*p*-value <0.05) between drugs' activity, 50% growth inhibition (GI50), and gene expressions across the cell lines. The correlation values can be either positive or negative depending on whether the given gene acts as an oncogene or tumor suppressor. The observed range for absolute value of the significant correlations was $0.26 < |Rs| < 0.68$. Finally, we computed the average of Ricci curvature values for the significant genes associated with each drug. The programming was done primarily in Matlab.

The GI50 is the drug concentration resulting in a 50% reduction in the net protein increase during drug incubation as compared with the same increase in control cells[29,30]. Normalized ($-\log_{10}$) GI50 was retrieved using the R package *rcellminer*[31]. This package complements the functionality of CellMiner by providing programmatic data access as well as analysis and visualization tool. By computing the Spearman's correlation between the expression of all 8240 genes and the GI50 response of 129 FDA approved drugs along the 58 cell lines we constructed the correlation matrix (D) shown in Fig. 2D. This is a matrix of 8240 × 129 elements which includes the correlations for all the pairs of genes and drugs. As an example, Fig. 3(b) derived from the rcellminer package illustrates the Spearman's correlation between the gene expression of MYC (z-score) and methotrexate activity (z-score). In this case, the correlation is very significant ($Rs = 0.41$) with a very low *p*-value of 0.001, therefore, MYC would be one of the selected genes for methotrexate. We provide another example, the gene MAPK8 and drug Salinomycin with a significant negative correlation ($Rs = -0.38$, $p\text{-value} = 0.004$) in Supplementary Fig. S4.

Ranking drugs and gene ontology enrichment analysis. The FDA approved drugs correspond to 161 NSC drug numbers (numeric identifiers for substances submitted to the National Cancer Institute). Among them we analyzed those with less than 2 missing GI50 values across all the cell lines, resulting in 129 drugs. For each drug, we identified genes whose expressions were significantly correlated ($p\text{-value} < 0.05$) to GI50. We compared the Ricci curvature distribution of these significant genes across the top and bottom ranked drugs and also compared them to the calculated Spearman's correlation in Supplementary Figs S5, S6. The median number of the

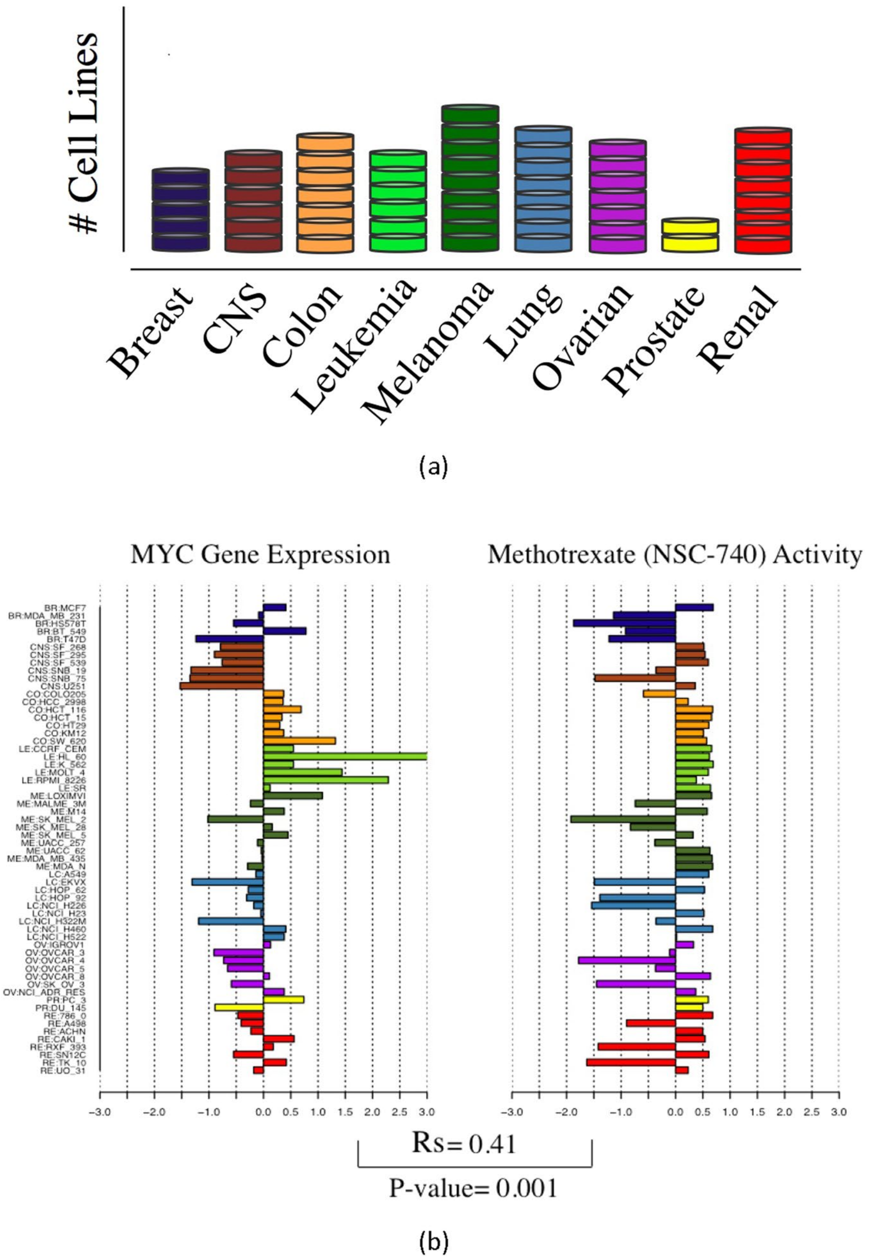

Figure 3. (**a**) Distribution of the 58 cell lines by type; (**b**) The Spearman's correlation along cell lines between each drug's GI50 activity and each gene's expression (only Methotrexate and MYC are shown) was calculated to create the correlation matrix (D) shown in Fig. 2.

genes selected for each drug was ~500 with the minimum number of ~300. The average Ollivier-Ricci curvature of these genes was calculated for each drug.

We then sorted the 129 FDA approved drugs in ascending order according to the average Ricci curvature (Fig. 2E) arguing that since the effective drugs should be able to perturb the subnetwork of its significant genes, this subnetwork should possess a lower average Ricci curvature. We used this ranking of 129 drugs to assess the importance of 8240 genes in our network across cell lines. More precisely, we gave a linear weight of (129−r) + 1 to all the selected genes associated with the r-th ranked drug. Then, for each gene, we computed the sum of all these weights and ranked the genes in descending order of the total weights. Thus, the final gene score increases if a gene (a) is important for many drugs, and (b) strongly contributes to robustness across multiple drugs. For a biological analysis, the top 200 genes were selected where the histogram has an apparent sharp decline; see Fig. 4. We performed a gene ontology (GO) enrichment analysis on the top 200 ranked genes using the MetaCore software (Thomson Reuters). MetaCore is an integrated software system based on a manually-curated database of molecular interactions, molecular pathways, gene-disease associations, chemical metabolism and toxicity information.

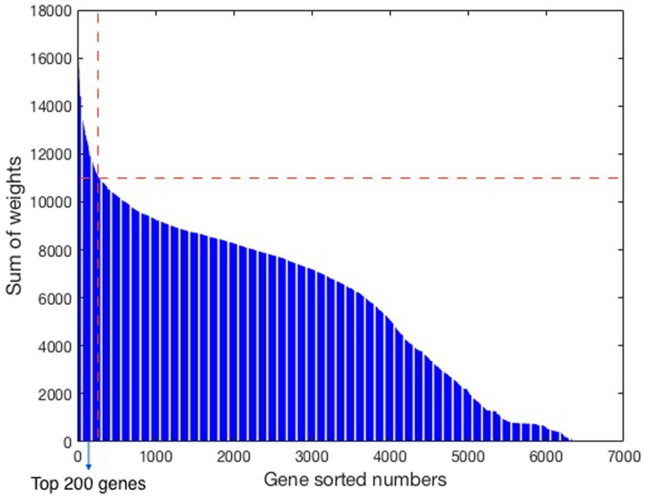

Figure 4. Top 200 genes selected for the gene ontology enrichment analysis.

Drug ranking	Drug name	Drug ranking	Drug name
1	Salinomycin	16	Doxorubicin
2	Gefitinib	17	Simvastatin
3	Homoharringtonine	18	Batracylin
4	Mitomycin	19	Daunorubicin
5	Idarubicin	20	Azacitidine
6	Geldanamycin Analog	21	Itraconazole
7	Cabozantinib	22	Dasatinib
8	Vinblastine	23	Arsenic trioxide
9	PX-316	24	Ibrutinib
10	Raloxifene	25	Tyrothricin
11	Pipamperone	26	Crizotinib
12	Erlotinib	27	Paclitaxel
13	Fluorouracil	28	Trametinib
14	Matinib	29	Fenretinide
15	Irinotecan	30	Tamoxifen

Table 1. Top 30 drugs ranked by average Ricci curvature of significantly correlated genes.

The list of 129 drugs includes 23 repeated drugs. For these repeated drugs, the GI50 data were different for the different NSC numbers, and therefore, they have different rankings. Although the rankings were close for these drugs, we chose the highest ranking as a representative for the repeated drug. Consequently, our final ranking consists of 106 drugs.

Results

We found the scalar Ollivier-Ricci curvature of all the 8240 genes in the pre-treatment interaction network discussed previously. The scalar curvatures range between −210.4 and 3.6 with an average of −5.2. Supplementary Table S1 presents the top 40 genes with the highest absolute value of Ricci curvature. The top two genes, TP53 and YWHAG, stand out with regards to their Ricci curvatures. A visualization of the pre-treatment network is provided in Supplementary Fig. S7. We then ranked the drugs based on the average Ricci curvature of significant genes for each drug. The top 30 ranked drugs are presented in Table 1. There are a number of very effective drugs that are ranked highly in this table; we elaborate on these drugs further in the discussion section below. The ranking of all 106 drugs are presented in Supplementary Table S2.

We validated this ranking by running the entire algorithmic pipeline nine times; see Fig. 2. Each time we excluded all the cell lines of one of the nine cancer tissues. We then computed the Spearman's correlation between drug rankings resulting from all the cell lines and those based on leave-one-out rankings. Interestingly, correlations are very high with low *p*-values, showing relative consistency of the results across cell lines. The color map in Fig. 5 illustrates these correlations. The rows (columns) of this symmetric map are numbered by excluding all the cell lines of one cancer type (Fig. 3(a)). For example, 'w/o Breast' corresponds to exclusion of all the five cell lines of breast cancer. The 10th row (column) corresponds to the case of considering all the cell lines. The

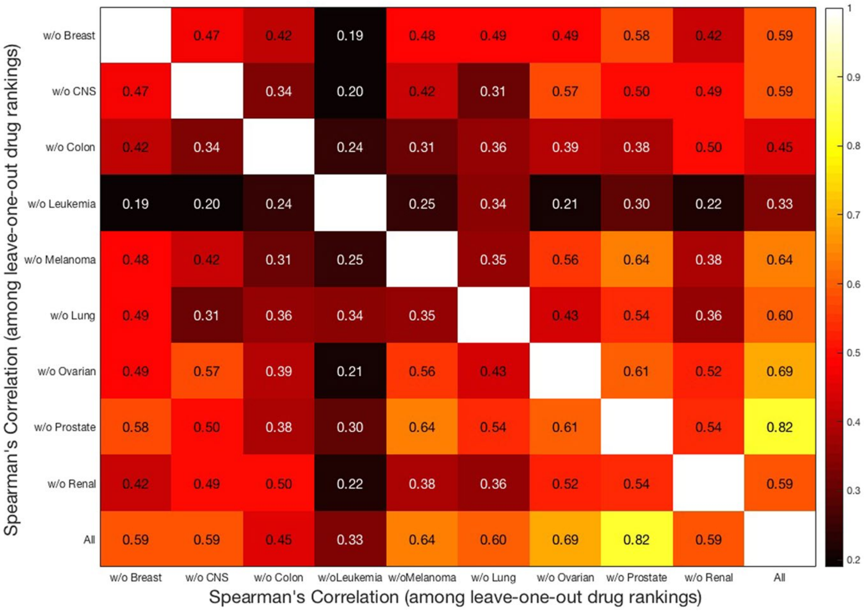

Figure 5. Color map of Spearman's correlation between drug rankings after excluding all the cell lines of one cancer type. "All" corresponds to the drug ranking with inclusion of all cell lines. Correlation values are also shown in the color map and are mostly very high between these drug rankings (with the average p-value = 0.002).

correlation values are also included in the color map. As we see in Fig. 5, the correlation is very significant among the rankings ($0.19 < Rs < 0.82$, p-value $\ll 0.05$), yet it is less significant after excluding the cell lines of leukemia (4th column or row). In other words, ranking of the drugs after exclusion of cell lines of leukemia is different from excluding any of the other eight cancer cell lines. This highlights that the effect of leukemia in the drug ranking is different than other cancer tissues. Of note, this is not a computational effect alone, given that leukemia is not even the cancer type with the greatest number of cell lines (colon, melanoma, lung and renal have more cell lines; see Fig. 3(a)). In order to determine which subset of leukemia accounts for this observation, we repeated the pipeline for the 6 cell lines of leukemia. We found two clusters of cell lines for leukemia according to their drug rankings (Supplementary Fig. S8). We then compared the drug rankings after adding these two subsets to the other cell lines. We discovered that a subset consisting of two cell lines most likely caused this deviation of leukemia from the other cell lines. Interestingly, these two cell lines belong to the same type of leukemia, namely, ALL. Of note, even though leukemia is different from other cancer types, there still exists a strong correlation with the others. The highest p-value ($Rs = 0.19$) that is between 'w/o Leukemia' and 'w/o Breast' is 0.03, which is still less than 0.05. The correlation between all cell lines and w4 (leukemia) is 0.33 with a p-value of 0.00012.

The top 200 ranked genes are presented in Supplementary Table S4. The results of the gene ontology enrichment analysis of this 200 gene set are shown in Fig. 6(a). The top ten biological processes presented with very small p-values ($<10^{-8}$). These p-values correspond to the hypergeometric test performed by MetaCore using the number of input genes, ontology related genes, and the total number of genes in the database. The top three biological processes are all involved with cellular localization. Also, the protein-protein interaction network of this analysis is presented in Fig. 6(b). The network contains two hubs associated with the gene product of CUX1 and PRKACA, which will be elaborated upon in the discussion section below. We further investigated the connectivity of these 200 genes in our pre-treatment network to ensure that these genes are not simply spread across distant parts of the network. We found that a subset of 146 genes of these 200 genes are mutually connected by paths of length two (i.e., connected with two edges), which is shown in Supplementary Fig. S9. The gene ontology enrichment analysis of these 146 genes resulted in very similar top biological processes (Supplementary Fig. S10), and included a branch of cellular localization, namely intracellular transport, with a p-value of 7.8×10^{-6}. The table of the top 50 biological processes with their corresponding p-values is provided in Supplementary Table S3. Also, Supplementary Table S4 includes these 146 genes.

In addition to the analysis of all cancer cell lines, we performed our algorithmic pipeline separately for four specific cancer tissues with the greatest number of cell lines: melanoma (9 cell lines), lung and renal (both 8 cell lines) and colon (7 cell lines) cancers. We performed the gene ontology enrichment analysis using the top 200 genes (Supplementary Table S4) of these specific cancer tissues, and compared the results to the biological processes of all the cell lines. We present the results of the gene ontology enrichment analysis of the top 200 genes of renal, lung, melanoma and colon cancer tumors in Supplementary Figs S11, S12, S13 and S14. Interestingly, these cancer types share some similar biological processes to those resulting from all 58 cell lines, which we will discuss further in the next section. However, the cancer specific analysis of this database is limited due to the small number of cell lines in any given specific cancer.

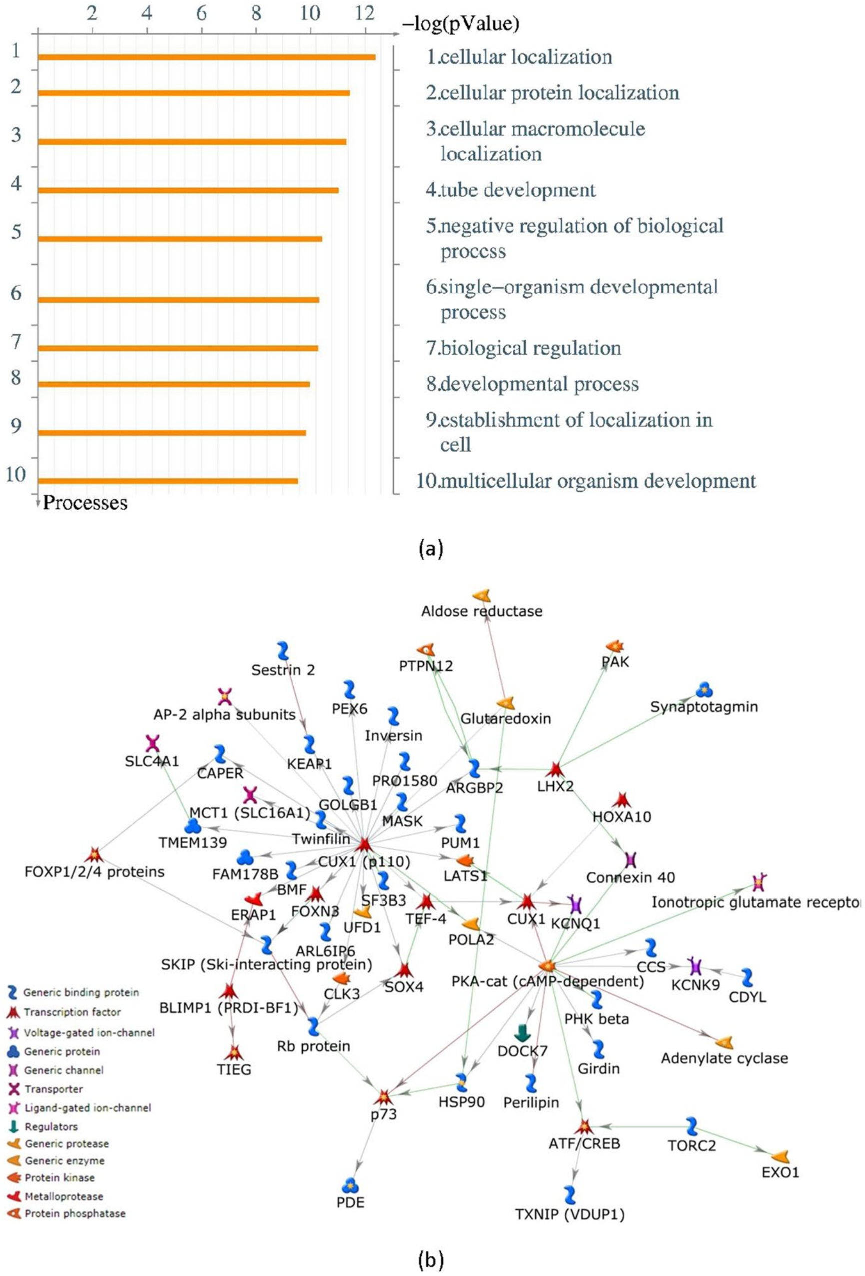

Figure 6. Gene ontology enrichment analysis (MetaCore) of the significant genes correlated with the top ranked drugs: (**a**) Top ten biological processes; top three biological processes are involved with the cellular localization. The corresponding *p*-value (hypergeometric test) of the top biological processes are also included in the bar plot. (**b**) Protein-protein interaction network has two hubs: CUX1 and cAMP-dependent protein kinase.

Discussion

In the present study, we considered the NCI-60 panel comprised of 58 individual cancer cell lines derived from nine different tissues (breast, brain, colon, blood, skin, lung, ovarian, prostate and kidney). The gene expressions of the cell lines were used to construct the network consisting of 8240 genes. This is a weighted graph where the underlying interactions are derived from Human Protein Reference Database. In this graph, the weights of the edges are the gene-to-gene (Pearson) correlations. The network was analyzed by calculating the discrete Ricci curvature. Based on our arguments in the Methods section, we claim that this is a guide to the robustness of the network, i.e., the ability to withstand perturbations in the system. In our case, these perturbations are induced by the drugs.

The scalar Ricci curvature helps us to identify the important targets (genes) for the drugs within the network. We present the top 40 genes with the least negative value (highest absolute value) of Ricci curvature in the pre-treatment network in Supplementary Table S1. The top two genes, TP53 and YWHAG have very low scalar Ricci curvatures compared to other genes (See Supplementary Table S1). Of note, mutations of TP53 are present in more than 50% of human cancers, making it the most common genetic event in human cancer[32,33]. This gene has many connections to other genes within the network which also contributes to its extreme value of scalar Ricci curvature (See Supplementary Fig. S7). Even though our primary focus in this study is to identify the important genes based on drug response (Fig. 4) and the biological processes they are involved in (Fig. 6), we briefly discuss the roles of TP53 and YWHAG (top 2 ranked genes in the pre-treatment network) in cancer pathogenesis in the Supplementary Information.

Furthermore, we measure the effectiveness of the drugs by the average Ricci curvature of the nodes (genes) it affects. For each drug, we found the significantly (positive/ negative) correlated genes by computing the correlation between GI50 and gene expression along the 58 cell lines. The significant genes for each drug were chosen based on p-values less than 0.05. On average ~500 genes were selected for each drug. These genes identify a subnetwork within our pre-treatment network which is affected by the drug. We provide the average discrete Ricci curvature of these genes for each drug. Since robustness and Ricci curvature are positively correlated, we can study the efficacy of the drug on the network by considering the average Ricci curvature of the subnetwork affected by that drug. If these subnetworks possess a higher average Ricci curvature, we expect them to show more resistance to the drug. In other words, drugs cannot effectively perturb the subnetworks with high average curvature. Therefore, the ranking of the drugs in ascending order (of average Ricci curvature) is a guide to the efficacy of the drugs for the cell lines. This network based view of the drug's effectiveness considers the gene interactions of the cell lines as well as the drug response. The table of top 30 drugs is presented in Table 1. We also provide the table of all 106 drug ranking in Supplementary Table S2.

Salinomycin, the first ranked drug, has recently been considered as a promising novel anti-cancer agent for targeting human cancer stem cells despite its not well-known mechanism of action[34–36]. The chemotherapeutic property of Salinomycin can overcome the resistance of tumor cells toward multiple drugs while selectively targeting the cancer stem cells. This antibiotic drug has been shown to kill breast cancer stem cells in mice at least 100 times more effectively than the known anti-cancer drug Paclitaxel. The study screened 16,000 different chemical compounds and found that only a few drugs, including Salinomycin, targeted cancer stem cells responsible for metastasis[37].

Gefitinib, our second ranked drug, is a molecular targeted drug in the treatment of non-small cell lung cancer. Approximately 85–90% of lung cancer cases, the most deadly cancer in the US, are NSCLC tumors. Mutations in the EGFR (epidermal growth factor receptor) gene are present in about 10 percent of NSCLC tumors (https://www.iressa-usa.com). EGFR overexpression leads to inappropriate activation of the anti-apoptotic Ras signalling cascade, thereby leading to uncontrolled cell proliferation[38]. Gefitinib competes with adenosine triphosphate at the ATP binding site in epithelial cells, blocking its tyrosine kinase activity, and consequently inhibiting EGFR signaling pathway, which can induce tumor cell apoptosis[39]. In 2015, gefitinib was FDA approved as a first line treatment in patients with metastatic NSCLC who harbor the most common types of EGFR mutations in NSCLC (exon 19 deletions or exon 21 L858R substitution gene mutations)[40].

Omacetaxine mepesuccinate, also known as homoharringtonine, the 3rd ranked drug, was originally identified over 35 years ago as a novel plant alkaloid with antitumor properties. Its mechanism of action is thought to be inhibition of protein translation by preventing the initial elongation step of protein synthesis via an interaction with the ribosomal A-site[41]. It was approved by the FDA in October 2012 for the treatment of adult patients with chronic myeloid leukemia (CML) with resistance and/or intolerance to two or more tyroskine kinase inhibitors as the current first-line treatment[42]. Furthermore, clinical studies have shown activity of omacetaxine in other malignancies as a single agent or in combination with other therapies in acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS), and studies are ongoing in this regard[43–45]. Also, a number common antitumor agents were ranked highly: doxorubicin, paclitaxel, fluorouracil (5-FU) and tamoxifen are commonly used in breast cancer treatment. Paclitaxel, vinblastine and irinotecan are often used in NSCLC. Homoharringtonine, azacitidine, and arsenic trioxide are common anticancer agents against leukemia.

The leave-one-out validation of drug rankings suggests that the rankings are not highly dependent on specific cancer tissue. As in clinical practice, this supports the use of anticancer drugs for the treatment of different cancer types. Overall, the Spearman's correlations are higher among solid tumors as compared to the liquid tumor, Leukemia. Network-based analysis can also help to understand important biological processes in which the most correlated genes of the top ranked drugs are involved. To this end, we ranked the genes by using our drug ranking. The genes that are correlated to the top ranked drugs are given higher weights. The genes were then ranked in the descending order of the sum of the weighted values. Therefore, the top ranked genes are more correlated with the drug sensitivity in the cell lines and are better targets in the network. The gene ontology analysis, performed via MetaCore, was employed to find the biological processes with which the top 200 genes are involved.

The top three biological processes are concerned with cellular localization. Cells consist of many different compartments that are specialized to carry out various tasks. Based on the gene ontology directory, cellular localization of a protein is involved with the process whereby a protein complex is transported to, and/or maintained in, a specific location within a cell, including the localization of substances or cellular entities to the cell membrane. The number of proteins that have reliable subcellular location annotations is approximately 20% of all known proteins to date[46]. It is of particular interest to determine if a potential target is a cell surface or secreted molecule which would be more easily accessible for the targeted drug approach[47]. Therefore, knowledge of the subcellular localization of a protein can significantly improve target identification during the drug development process[48]. We further repeated the gene ontology enrichment analysis of the subset of 146 among the top 200 genes, which are connected with a path of length two. A more specific ontology term which is a branch of cellular localization, namely intracellular transport, is included in the top biological processes of these genes (Supplementary Table S3). In this regard, disruption of normal intracellular protein transport is critically involved in the pathophysiology of a broad range of human malignancies. Aberrant localization of oncoproteins and tumor suppressors may result in their over-activation or inactivation, respectively, and this can allow evasion of anti-neoplastic therapy as well[49,50]. Therefore, the therapeutic targeting of the nucleocytoplasmic shuttling of macromolecules has emerged as a promising novel therapeutic approach to treat human diseases including cancer. More specifically, targets involving the nuclear envelope and nuclear intracellular transport machinery including cargo proteins, transport receptors, Ran regulators, and the nuclear pore complex have been proposed for therapeutic intervention. Several agents have been developed against these targets, some of them with promising therapeutic windows[51].

We also present in Fig. 6(b) the protein-protein interaction network identified by MetaCore corresponding to the top 200 genes. There are two notable hubs in this network: CUX1 and cAMP-dependent protein kinase which is a gene product of PRKACA. CUX1 is specifically important since it is a transcription factor to a number of our top ranked genes associated with the drug sensitivity. CUX1 (also known as CUTL1) is a homeobox transcription factor highly evolutionarily conserved and plays a known role in embryonic development, cell growth and differentiation in mammals[52]. Moreover, the role of CUX1 in drug resistance/sensitivity, has been explored. Specifically, gain-of-function as well as loss-of-function studies have shown that increased CUX1 activity significantly enhanced cell sensitivity and cancer tissue response to chemotherapy drugs and resulted in increased apoptosis and growth inhibition. In contrast, decreased CUX1 expression reduced cell sensitivity to chemotherapy drugs with fewer apoptoses and resultant drug resistance[53]. These studies, which have been in the context of gastric cancer, suggest an inverse association between CUX1 and drug resistance, implying that CUX1 is an attractive therapeutic target. Whether this phenomenon applies to cancers other than gastric cancer remains to be elucidated. The human PRKACA gene encodes the PKA catalytic subunit alpha (C_α) isoform. With regards to drug sensitivity/resistance, PRKACA is over expressed in invasive and anti-HER2 therapy (trastuzumab/ lapatinib)-resistant breast cancers. In addition to PRKACA conferring resistance to anti-HER2 therapy, it also impairs apoptosis[54]. Consequently, inhibition of PRKACA and/or its downstream anti-apoptotic effectors in combination with anti-HER2 therapy may increase the drug sensitivity. Its role in drug sensitivity/resistance for other cancers needs to be studied further.

Finally, we were interested in comparing the results of the gene ontology enrichment analysis of all 58 cancer type cell lines to some specific cancer tissues with the largest number of cell lines. The repeated algorithmic process for lung, renal, melanoma and colon cancers yields some consistency. Similar to all cancer type genes, the gene ontology enrichment analysis for the top genes involved with renal and lung cancers results in cellular localization and cellular component organization. However, most of the biological processes from melanoma and colon cancer gene ontology enrichment analysis are quite broad such as negative regulation of cellular process, which is also a top biological process of all cancer types as well as renal cancer. Of note, colon cancer has fewer cell lines (7 cell lines) than melanoma (9 cell lines) and lung and renal cancer (8 cell lines). However, as mentioned earlier, the cancer specific analysis of this database is limited due to the small number of cell lines in each specific cancer.

The need for network based techniques is becoming more and more prevalent as a result of the exponential growth of data in the genomic era. In this work, we utilized mathematical techniques to drive a network based analysis in order to explore the genomic and pharmacogenomics information of NCI-60. The framework of this study can be extended to find possible optimal combinations of drugs. Combining anti-cancer agents, whether cytotoxic or molecularly targeted, with different mechanisms of action is the most practical approach to overcome single drug resistance and can produce sustained clinical remissions. Also, the study described in the present work may be extended to tissue-specific drugs employing a more appropriate tissue-specific cell line database. This analysis on the NCI-60 is promising and supports efforts to analyze larger datasets with advanced network mathematics[55,56]

References

1. Shoemaker, R. H. The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer.* **6**(10), 813–823 (2006).
2. Chabner, B. A. NCI-60 Cell Line Screening: A Radical Departure in its Time. *J. Natl. Cancer Inst.* **108**(5), djv388 (2016).
3. Boyd, M. R. & Paull, K. D. Some practical considerations and applications of the National Cancer Institute *in vitro* anticancer drug discovery screen. *Drug Dev. Res.* **34**(2), 91–109 (1995).
4. Weinstein, J. N. Spotlight on molecular profiling: "Integromic" analysis of the NCI-60 cancer cell lines. *Mol. Cancer Ther.* **5**(11), 2601–2605 (2006).
5. Reinhold, W. C. *et al.* CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res.* **72**(14), 3499–3511 (2012).
6. Wodarz, D. & Komarova N. L. *Dynamics of Cancer: Mathematical Foundations of Oncology* (World Scientific, 2014).
7. DoCarmo, M. *Riemannian Geometry* (Birkhäuser, 1992).
8. Jost, J. *Riemannian Geometry and Geometric Analysis* (Springer-Verlag, 2011).
9. Ollivier, Y. Ricci curvature of metric spaces. *C. R. Math Acad. Sci. Paris.* **345**(11), 643–646 (2007).
10. Ollivier, Y. Ricci curvature of Markov chains on metric spaces. *J. of Funct. Anal.* **256**(3), 810–864 (2009).
11. Rachev, S. T. & Rüschendorf, L. *Mass Transportation Problems. Vol. I: Theory*, *Vol. II: Applications* (Springer-Verlag, 1998).
12. Villani, C. *Topics in Optimal Transportation* (American Mathematical Society, 2003).
13. Jordan, R., Kinderlehrer, D. & Otto, F. The variational formulation of the Fokker-Planck equation. *SIAM J. Math Anal.* **29**(1), 1–17 (1998).
14. Otto, F. The geometry of dissipative evolution equation: the porous medium equation. *Comm. Partial Differential Equations*. **26**(1–2), 101–174 (2001).
15. Benamou, J.-D. & Brenier, Y. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik* **84**(3), 375–393 (2000).
16. Villani, C. *Optimal Transport, Old and New*, (Springer, 2008).
17. Lott, J. & Villani, C. Ricci curvature for metric-measure spaces via optimal transport. *Anal. of Mathematics.* **169**(3), 903–991 (2009).
18. Sturm K.-T. On the geometry of metric measure spaces. I and II *Acta. Mathematica*. **196**(**1**), 65–131 & 133–177 (2006).
19. Demetrius, L. & Manke, T. Robustness and network evolution - an entropic principle. *Physica A.* **364**(3), 682–696 (2005).
20. Pouryahya M, *et al.* Bakry-Émery Ricci Curvature on Weighted Graphs with Applications to Biological Networks. *Int. Symp. on Math. Theory of Net. and Sys.* **22** (2016).
21. Sandhu, R. *et al.* Graph curvature for differentiating cancer networks. *Sci. Rep.* **5**, 12323 (2015).
22. Tannenbaum A. *et al.* Graph curvature and the robustness of cancer networks. Preprint at http://arxiv.org/abs/1502.04512 (2015).
23. Evans, L. C. Partial differential equations and Monge-Kantorovich mass transfer. *Current Dev. in Math*. 65–126 (1999).
24. Teschendorff, A., Sollich, P. & Kuehn, R. Signalling entropy: A novel network-theoretical framework for systems analysis and interpretation of functional omic data. *Methods* **67**, 282–293 (2014).
25. Varadhan S. R. S. *Large Deviations and Applications*, *CBMS-NSF Regional Conference Series in Applied Mathematics* **46** (SIAM, 1984).

26. Rubner, Y., Tomasi, C. & Guibas, L. J. The earth mover's distance as a metric for image retrieval. *Int. J. of Computer Vision* **40**(2), 99–121 (2000).
27. Prasad, T. S. K. *et al*. Human Protein Reference Database–2009 update. *Nucleic Acids Res.* **37**, D767–D772 (2009). (Database issue:).
28. Nguyen, C. D., Gardiner, K. J. & Cios, K. J. Protein annotation from protein interaction networks and Gene Ontology. *J. Biomed. Inform.* **44**(5), 824–829 (2011).
29. Monks, A. *et al*. Feasibility of a high-flux anticancer screen using a diverse panel of cultured human tumor lines. *J. Natl. Cancer Inst.* **83**(11), 757–766 (1991).
30. Boyd, M. R., Paull, K. D. & Rubinstein, L. R. Data Display and Analysis Strategies for the NCI Disease-Oriented *in Vitro* Antitumor Drug Screen. *Models and Concepts for Drug Discovery and Development. Developments in Oncology*, *vol 68*, (eds Valeriote F. A., Corbett T. H., Baker L. H.) 11–34 (Springer, 1992).
31. Luna, A. *et al*. rcellminer: exploring molecular profiles and drug response of the NCI-60 Cell Lines in R. *Bioinformatics.* **32**(8), 1272–1274 (2016).
32. Levine, A. J. & Oren, M. The first 30 years ofp53: growing ever more complex. *Nat. Rev. Cancer.* **9**(10), 749–758 (2009).
33. Freed-Pastor, W. A. & Prives, C. Mutantp53: one name, many proteins. *Genes Dev.* **26**(12), 1268–1286 (2012).
34. Naujokat, C. & Steinhart, R. Salinomycin as a drug for targeting human cancer stem cells. *J. Biomed. Biotechnol.* **950658** (2012).
35. Zhou, S. *et al*. Salinomycin: a novel anti-cancer agent with known anti-coccidial activities. *Curr. Med. Chem.* **20**(33), 4095–4101 (2013).
36. Dewangan, J., Srivastava, S., Rath, S. K. Salinomycin: a new paradigm in cancer therapy. *Tumour Biol*. **39**(**3**) (2017).
37. Gupta, P. B. *et al*. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell.* **138**(4), 645–659 (2009).
38. Sordella, R., Bell, D. W., Haber, D. A. & Settleman, J. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science.* **305**(5687), 1163–1167 (2004).
39. Dhillon, S. Gefitinib: a review of its use in adults with advanced non-small cell lung cancer. *Target Oncol.* **10**(1), 153–170 (2015).
40. Dangi-Garimella, S. Gefitinib approved as frontline in EGFR-positive NSCLC. . *Evidenced-Based Oncology.* **21**(12), SP417–419 (2015).
41. Wetzler, M. & Segal, D. Omacetaxine as an anticancer therapeutic: what is old is new again. *Curr. Pharm. Des.* **17**(1), 59–64 (2011).
42. FDA approval for Omacetaxine Mepesuccinate. https://www.cancer.gov/about-cancer/treatment/drugs/fda-omacetaxinemepesuccinate (2012).
43. Zhang, W. G. *et al*. Combination chemotherapy with low-dose cytarabine, homoharringtonine, and granulocyte colony-stimulating factor priming in patients with relapsed or refractory acute myeloid leukemia. *Am. J. Hematol.* **83**(3), 185–188 (2008).
44. Wang, J. *et al*. A homoharringtonine-based induction regimen for the treatment of elderly patients with acute myeloid leukemia: a single center experience from China. *J. Hematol. Oncol.* **2**, 32 (2009).
45. Jin, J. Homoharringtonine in combination with cytarabine and aclarubicin resulted in high complete remission rate after the first induction therapy in patients with de novo acute myeloid leukemia. *Leukemia.* **20**(8), 1361–1367 (2006).
46. Chou, K. C. & Shen, H. B. Recent progresses in protein subcellular location prediction. *Anal. Biochem.* **370**(1), 1–16 (2007).
47. Simon, I. *et al*. Determining subcellular localization of novel drug targets by transient transfection in COS cells. *Cytotechnology.* **35**(3), 189–196 (2001).
48. Nakai, K. Protein sorting signals and prediction of subcellular localization. *Adv. Protein Chem.* **54**, 277–344 (2000).
49. Turner, J. G., Dawson, J. & Sullivan, D. M. Nuclear export of proteins and drug resistance in cancer. *Biochem. Pharmacol.* **83**(8), 1021–1032 (2012).
50. Hung, M. C. & Link, W. Protein localization in disease and therapy. *J. Cell Sci.* **24**(20), 3381–3392 (2011).
51. Hill, R., Cautain, B., de Pedro, N., Link, W. Targeting nucleocytoplasmic transport in cancer therapy. *Oncotarget*. **5**(**1**) (2013)
52. Wang, J. *et al*. CUTL1 induces epithelial-mesenchymal transition in non-small cell lung cancer. *Oncol. Rep.* **37**(5), 3068–3074 (2017).
53. Li, T. *et al*. Transcription factor CUTL1 is a negative regulator of drug resistance in gastric cancer. *J. Biol. Chem.* **288**(6), 4135–4147 (2013).
54. Moody, S. E. *et al*. PRKACA mediates resistance to HER2-targeted therapy in breast cancer cells and restores anti-apoptotic signaling. *Oncogene.* **34**(16), 2061–2071 (2015).
55. Zhang, B. & Horvath, S. A General Framework for Weighted Gene Co-expression Network Analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article 17 (2005).
56. Oh, J. H. & Deasy, J. O. A literature mining-based approach for identification of cellular pathways associated with chemoresistance in cancer. *Brief Bioinform.* **17**(3), 468–478 (2016).

Acknowledgements

This project was supported by AFOSR grants (FA9550-15-1-0045 and FA9550-17-1-0435), ARO grant (W911NF-17-1-049), grants from the National Center for Research Resources (P41-RR-013218) and the National Institute of Biomedical Imaging and Bioengineering (P41-EB-015902), NCI grant (1U24CA18092401A1), NIA grant (R01 AG053991), MSK Cancer Center Support Grant/Core Grant (P30 CA008748), and a grant from the Breast Cancer Research Foundation.

Author Contributions

M.P., J.-H.O., J.D. and A.T. designed research; M.P. performed research; M.P. and J.-H.O. analyzed data; M.P. wrote the paper; J.-H.O., J.D., J.M. and A.T. edited the paper; J.D. and A.T. directed research.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-24679-3.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.