



A network approach to quantifying radiotherapy effect on cancer: Radiosensitive gene group centrality



Yu-Xiang Yao^a, Zhi-Tong Bing^{b,c,d}, Liang Huang^{a,*}, Zi-Gang Huang^{e,*}, Ying-Cheng Lai^{f,g}

^aSchool of Physical Science and Technology, Lanzhou University, Lanzhou 730000, China

^bEvidence Based Medicine Center, School of Basic Medical Science of Lanzhou University, Lanzhou 730000, China

^cKey Laboratory of Evidence Based Medicine and Knowledge Translation of Gansu Province, Lanzhou 730000, China

^dDepartment of Computational Physics, Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou 730000, China

^eThe Key Laboratory of Biomedical Information Engineering of Ministry of Education, National Engineering Research Center of Health Care and Medical Devices, The Key Laboratory of Neuro-informatics & Rehabilitation Engineering of Ministry of Civil Affairs, and Institute of Health and Rehabilitation Science, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China.

^fSchool of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA

^gDepartment of Physics, Arizona State University, Tempe, AZ 85287, USA

ARTICLE INFO

Article history:

Received 4 October 2018

Accepted 1 December 2018

Available online 3 December 2018

Keywords:

Radiotherapy

Radiosensitive

Network science

Network centrality

Cancer

ABSTRACT

Radiotherapy plays a vital role in cancer treatment, for which accurate prognosis is important for guiding sequential treatment and improving the curative effect for patients. An issue of great significance in radiotherapy is to assess tumor radiosensitivity for devising the optimal treatment strategy. Previous studies focused on gene expression in cells closely associated with radiosensitivity, but factors such as the response of a cancer patient to irradiation and the patient survival time are largely ignored. For clinical cancer treatment, a specific pre-treatment indicator taking into account cancer cell type and patient radiosensitivity is of great value but it has been missing. Here, we propose an effective indicator for radiosensitivity: radiosensitive gene group centrality (RSGGC), which characterizes the importance of the group of genes that are radiosensitive in the whole gene correlation network. We demonstrate, using both clinical patient data and experimental cancer cell lines, which RSGGC can provide a quantitative estimate of the effect of radiotherapy, with factors such as the patient survival time and the survived fraction of cancer cell lines under radiotherapy fully taken into account. Our main finding is that, for patients with a higher RSGGC score before radiotherapy, cancer treatment tends to be more effective. The RSGGC can have significant applications in clinical prognosis, serving as a key measure to classifying radiosensitive and radioresistant patients.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Radiotherapy has been an indispensable tool for treating cancer and controlling its growth, which is received by nearly 50% of the cancer patients (Moding et al., 2013). A fundamental issue in radiotherapy is to assess the radiosensitivity (RS) of a cancer patient to enable decision making toward optimal treatment strategy (Eschrich et al., 2009). In the modern era of precise medicine, gene signatures as a response predictor for radiotherapy and chemotherapy have been effective in the treatment of various cancers (Zhu et al., 2010; Van De Vijver et al., 2002; Lee et al., 2015; Bing et al., 2016). Previous studies established that a group of 31 genes

from NCI-60 cancer cells are closely associated with radiotherapy (Kim et al., 2012). However, due to the high complexity of the gene regulatory system, the intrinsic relationship between these genes has been unclear and it remains difficult to predict the effectiveness of radiotherapy for cancer patients.

Network science has the potential to provide powerful tools for analyzing a variety of natural and artificial complex systems (Palla et al., 2005; Barthe, 2011; Holme, 2012; Wang et al., 2016), including those in life science and medicine. For example, metabolic pathways can be identified through analyzing the network structure of genes/proteins (Kauffman et al., 2003; Barabasi and Oltvai, 2004), epidemic outbreak can be explained based on spreading dynamics on networks (Pastor-Satorras and Vespignani, 2001; Brockmann and Helbing, 2013), ecological stability associated with environment can be assessed using the network approach (Proulx et al., 2005), designing genetic circuits can benefit from the principles

* Corresponding authors.

E-mail addresses: huangl@lzu.edu.cn (L. Huang), huangzg@xjtu.edu.cn (Z.-G. Huang).

Table 1
Datasets used in this article.

| Dataset | Brief description | Platform ¹ |
|----------|--|-----------------------|
| GSE59 | cDNA microarrays of NCI-60 cancer cell lines | GPL167, 169 |
| GSE3578 | mRNA expression of cervical cancer patients during therapy | GPL2895 |
| GSE5949 | mRNA expression of NCI-60 cancer cell lines | GPL91~95 |
| GSE7505 | cDNA microarrays of NCI-60 cancer cell lines after radiation | GPL5080 |
| GSE7696 | mRNA expression of GBM patients | GPL570 |
| GSE16011 | mRNA expression of GBM patients | GPL8542 |
| GSE32474 | mRNA expression of NCI-60 cancer cell lines | GPL570 |
| TCGA-GBM | GBM patient samples from TCGA | GPL570 |

(1) GPL91~95, Affymetrix Human Genome U95A U95E Array.

(2) GPL167, 169, 10kPrint3, 10kPrint2 of spotted DNA/cDNA (non-commercial).

(3) GPL570, Affymetrix Human Genome U133 Plus 2.0 Array.

(4) GPL2895, GE Healthcare/Amersham Biosciences CodeLink Human Whole Genome Bioarray.

(5) GPL5080, NHGRI Homo sapiens 6K of spotted DNA/cDNA.

(6) GPL8542, Affymetrix GeneChip Human Genome U133 Plus 2.0 Array.

¹ We adopt GEO accession for data sets to simplify the description of the GEO Platform (GPL). The detailed information, which can be found from the website of National Center for Biotechnology Information (NCBI), is.

from network science (Alon, 2007), and tipping points in complex mutualistic networks can be analyzed and predicted (Jiang et al., 2018). Recently, network science has been employed to advance research in cancer and oncology, providing unprecedented insights into physiological phenomena related to tumor growth (Yarden and Pines, 2012; Pujana et al., 2007) with clinical applications (Chuang et al., 2007; Gevaert et al., 2006), clarifying the biochemical factors and signaling pathways during primary tumor development (Ivliev et al., 2010; Balkwill, 2004), and leading to pathway-directed drug discovery (Altieri, 2008). The purpose of this paper is to exploit principles of complex network science to propose, analyze, and validate a geometric indicator to effectively and quantitatively characterize the impact of radiotherapy on cancer patients. This is the radiosensitive gene group centrality (RSGGC), which can be calculated through identifying both the relations among gene signatures and the efficiency of the radiotherapy treatment.

Our work is motivated by the considerations that, in spite of the completeness of the Human Genome Project (Venter et al., 2001), the detailed functions and relationship among the genes remain unclear, and the existing protein-protein interaction (PPI) networks are unable to capture the overall features of the regulation process. It is thus useful to concentrate on the intrinsic properties of the genes with a network dynamic topology. In particular, we consider the intergenic correlations among data samples of the same cluster or classification to obtain a network through Pearson's correlation matrix, which is also known as the gene co-expression matrix (Stuart et al., 2003). The matrix quantifies the gene-gene relationship across tissue samples. In network science, the centrality of a node quantifies its importance in the network based on the local connection strength to its neighbors (Freeman, 1978; Gomez et al., 2003; Borgatti and Everett, 2006). Similarly, the centrality of a cluster of nodes is the sum of the centrality values over all the nodes in the cluster, which characterizes the importance of the focal cluster (Aguirre et al., 2013). Based on this concept, we propose RSGGC as a novel indicator to quantitatively assess malignant tumor development and to provide prognosis for patients. Combining with the finding of work (Kim et al., 2012), we group the 31 radiosensitive genes to be our focal cluster to calculate the RSGGC, as shown schematically in Fig. 1.

Through analysis based on different data sets from patients and cancer cell lines, we find a strong correlation between RSGGC and the patient's clinical or physiological indicators, suggesting that RSGGC can be potentially be applied to analyzing the therapeutic effect of radiotherapy. Specifically, for clinical data, we collect the previously public datasets from Gene Expression Omnibus (GEO) and the Cancer Genome Atlas (TCGA) to calculate the correlation between the survival time of a patient and his/her

RSGGC value from the clinical perspective. We then analyze the experimental data of cancer cell lines from the existing literature (Amundson et al., 2008). A typical data set includes the survival fraction (SF) of parallel cell lines with different radiation dosages (2 Gy, 5 Gy, 8 Gy) as labels of irradiation resistance for follow-up process. We calculate the RSGGC values for the corresponding NCI-60 cancer cell lines mRNA expression data. A statistical analysis shows that patients with a high RSGGC score have a longer survival time after radiotherapy and, consistently, cell lines with a higher RSGGC value have a smaller survival fraction after irradiation. Further, for different glioblastoma multiforme (GBM) cancer subtypes, our results indicate that the higher the RSGGC score is, the better therapeutic effect would be for patients receiving radiotherapy. Finally, based on the time sequencing data associated with irradiation, we compare the values of RSGGC before and after irradiation and find a sensitive response of RSGGC to irradiation, indicating the potential advantage of using RSGGC to assess the effect of radiotherapy in a quantitative manner. Our detailed and systematic analysis of RSGGC suggest that it can be used as a potential indicator of the effectiveness of radiotherapy to greatly facilitate decision making toward an optimal strategy for treating cancer.

2. Material and methods

2.1. Data collection

The required mRNA/cDNA expression profile of a patient or a cancer cell line involved in radiotherapy or in irradiation experiment, respectively, is collected from the previous published datasets from Gene Expression Omnibus (GEO) and the Cancer Genome Atlas (TCGA), with detailed information listed in Table 1.

Radiotherapy is the leading therapeutic strategy for patients suffering from GBM, as the surgical risk for this type of cancer is greater than others. GBM thus provides substantially more radiotherapy cases in the test datasets (GSE7696 (Murat et al., 2008), GSE16011 (Gravendeel et al., 2009), and TCGA-GBM). In order to investigate the effect of radiotherapy on patients, we use a number of cervical cancer patients in the dataset GSE3578 (Iwakawa et al., 2007) that contains the sequencing changes of mRNA expression during therapy. We use NCI-60 cancer cell lines in a more homogeneous environment as the validation datasets: GSE32474 (Pfister et al., 2009) and GSE5949 (Reinhold et al., 2010), which were widely used in exploring the underlying mechanisms of cancer and for developing drugs (Shoemaker, 2006; Gao et al., 2015; Lee et al., 2007). As an additional validation of the cancer cell lines, we study the integrated datasets: GSE59 (Ross et al., 2000) and GSE7505 (Amundson et al., 2008), which record the abundance of

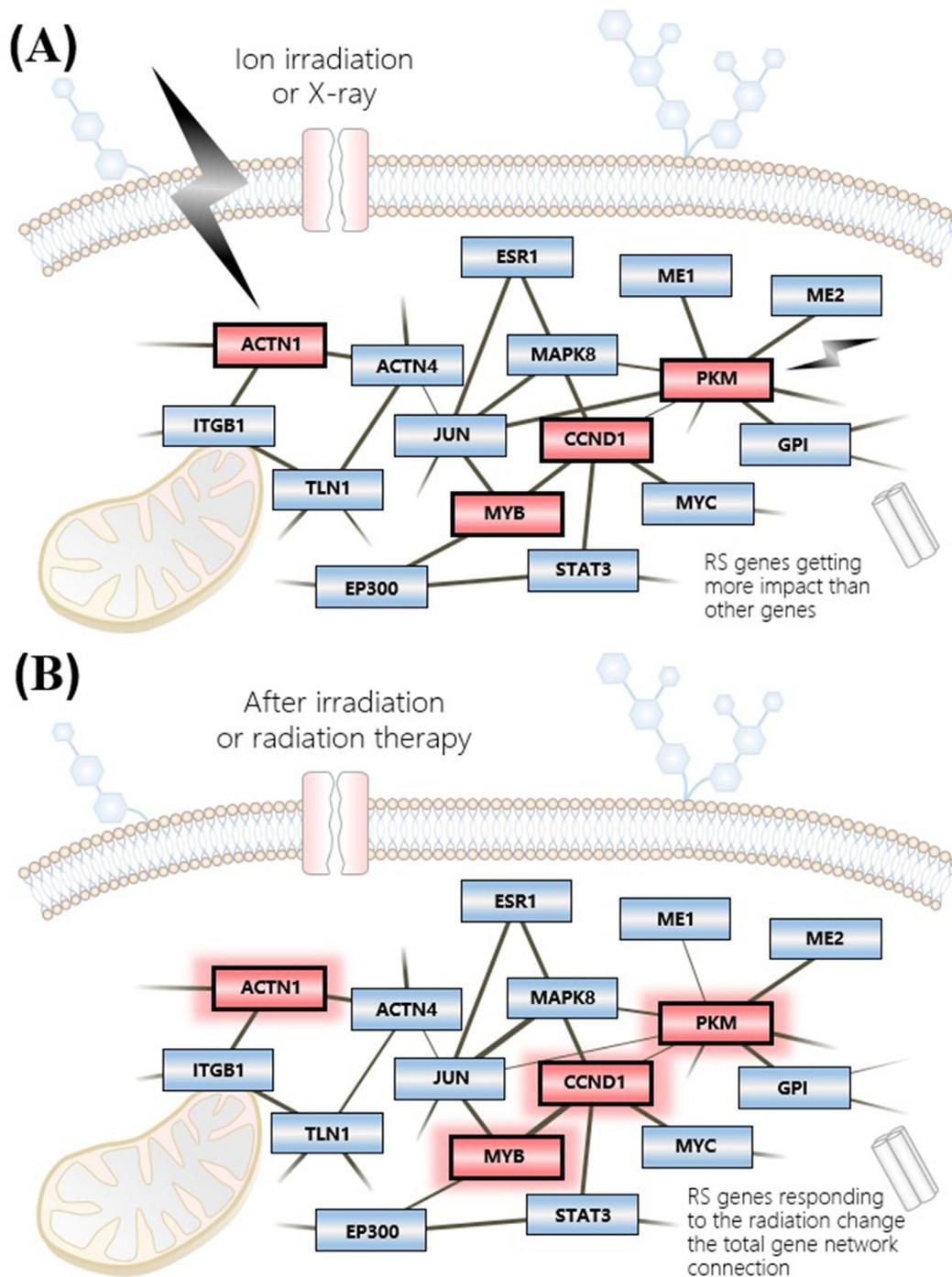


Fig. 1. Schematic illustration of the effects of radiation sensitive genes on the intercellular interaction or correlation of genes in cells: (A) before and (B) after irradiation. The red (thicker border) and blue (thinner border) blocks represent the radiation sensitive and the remaining genes, respectively. For simplicity and clarity, only a small fraction of the genes are demonstrated. The links in the network are drawn according to the open database STRING-DB. The thickness of a line between a pair of genes indicates the interaction strength between the genes. A comparison between (A) and (B) gives a glimpse of one possible change in the interaction strength due to irradiation.

cDNA of the NCI-60 cell lines before and after irradiation, respectively.

2.2. Data preprocessing

For reliable and meaningful statistical analysis, preprocessing of data is necessary. Briefly, we first select proper data through ranking and variance cutoff. We then apply either linear or logarithmic scales to the pertinent index, e.g., the survival time of patients (Table 2) or the survival fraction of cell lines (Table 3). In the fol-

lowing, we describe the preprocessing details for clinical data and cancer cell lines separately.

Preprocessing of clinical data. For patients, we screen out the cases GSE7696, GSE16011 and TCGA-GBM from GEO and TCGA primarily. All clinical data are filtered according to the following criteria: (1) null and void cases were removed (for example, lost record, abnormal expression, and rewritten data), and patients with both mRNA expression and the corresponding clinical information were selected; (2) one clinical index was used to classify patients (e.g., survival time or subtypes of patients). The GSE3578 (cervical can-

Table 2
Clinical information and groups of patients in different datasets.

| Dataset | Sample size | Grouping to the survival time (K days) | Mean survival time (K days) ¹ | Mean Age | Gender (M/F) |
|----------|-------------|--|--|----------|--------------|
| GSE7696 | 65 | 0–0.25–0.50–∞ | 0.438 | 51.178 | 48/17 |
| GSE16011 | 180 | 0–0.1–0.2–0.4–0.8–1.6–3.2–∞ | 0.836 | 49.986 | 122/58 |
| TCGA-GBM | 318 | 0–0.25–0.50–1.00–2.50–∞ | 0.563 | 56.138 | 195/123 |

¹ Only the patients with a clinical event (death) are taken into account.

Table 3
Method to group cancer cell lines based on the survival fractions under different radiation dosages (2, 5, and 8 Gy, respectively).

| Dataset | Sample size | Grouping to the survival fraction | | |
|----------|-------------|-----------------------------------|----------------------------|---|
| | | SF2 | SF5 | SF8 |
| GSE59 | 60 | quartiles | 0–0.01–0.05– | 0–0.001–0.005– |
| GSE5949 | 60 | quartiles | 0.10–0.25–1.00 | 0.010–0.25–1.00 |
| GSE7505 | 60 | quartiles | | |
| GSE32474 | 174 | deciles | 0–0.005–0.05–0.1–0.25–1.00 | 0–0.0005–0.005–0.010–0.025–0.05–0.1–1.0 |

cer patient set) has relatively adequate, complete clinical and gene expression information, so it was chosen as an additional validation of the ability of RSGGC to evaluate therapy, from which samples were classified according to the difference in therapeutic strategies and checking time. For convenience, samples were classified on a logarithmical scale to balance the survival time and the relevant order of magnitude of samples. **Table 2** shows the physiological or pathological brief summary and grouping results.

Preprocessing of cell lines data. We match the description and annotation of the experimental samples with the expression profile, and remove samples with missing information. We adopt only the cancer cell lines described in previous work (Amundson et al., 2008) and label them with the SF value. The radiosensitive and radioresistant lines can be represented by the replication rate, cell migration capacity, and SF under specific condition or environment. We use the SF value after receiving irradiation of different dosage level to characterize cell's radiosensitivity / radioresistance. All the cell line datasets are ranked according to the SF values of the cell lines after receiving 2 Gy, 5 Gy, and 8 Gy irradiation, denoted by SF2, SF5, and SF8, respectively (Amundson et al., 2008). **Table 3** shows the method of sampling groups for different radiation dosage, where the low dose irradiation dataset (2 Gy) is grouped linearly according to SF2, such as deciles or quartiles, while the high dose irradiation datasets (5 Gy and 8 Gy) have logarithmical bins according to SF5 and SF8. The different ways of data binning were adopted just for convenience.

2.3. Correlation matrix and RSGGC measure

Correlation matrix. A basic fact in systems biology and biomedical science is that genes are not isolated with each other but work collectively as an interacting network, regardless of whether the underlying process is intracellular or extracellular. To characterize the responses of patients or cancer cell lines to irradiation, we introduce the measure of RSGGC by considering the differences among sample groups and the corresponding inherent dynamic relationship from the point of view of a complex network.

A prerequisite to defining RSGGC is the correlation matrix. To begin, for a given dataset, we rank all the genes by the inter-sample variance with a proper cutoff to ensure computational efficiency, taking into account the balance of heterogeneity of the platforms as in previous work (Volinia and Croce, 2013). The resulting $M \times N$ expression matrix (for M genes and N samples) are ordered again by some index, e.g., the survival rate or subtypes. We then divide the matrix into different sections: $M \times n_1, M \times n_2,$

\dots , where $n_1 + n_2 + \dots = N$. For each sectional expression matrix, we calculate its Pearson's correlation matrix from all the gene pairs based on the available samples, which leads to an M -dimensional, fully connected, real symmetric matrix, with each element characterizing the similarity in the expression level of the two genes. To distinguish this $M \times M$ matrix from a gene co-expression network, we do not set any threshold so as to maintain the original correlation between the genes, and obtain an adjacency matrix without information loss, with the element of the matrix given by

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \quad (1)$$

where $\text{cov}(X, Y)$ stands the covariance between variables X and Y (genes), σ_X and σ_Y are the standard deviations (SD) of the two variables, respectively.

After the data is filtered and grouped, we calculate the intra-sample variance of a single gene for one total dataset. Some datasets provide gene's expression data, while others record the original expression profiles of RNA fragments. For the former, it is straightforward to calculate the variance but for the latter, we first merge the multiple probes that match the same genes via arithmetic averaging before calculating the variance. Probes without the corresponding gene names are dropped from the calculation. The detailed information of the remaining gene number in each step is presented in Table S1.

To treat the different datasets on an equal footing, it is necessary to determine a variance cutoff. In particular, the datasets are obtained from different platforms and are processed by probe merging, and each single dataset contains a different number of genes. One difficulty is that, after ranking by the variance size, many genes have values of variance that are close to each other. Moreover, invalid genes in the data lead to wasted computation. In addition, certain genes have almost the same expression among diverse cells or environment. To overcome these difficulties, we use the insights from previous work (Oldham et al., 2006; Zhang and Horvath, 2004) and choose the first 8000 genes with large variance to ensure that they are computationally distinct without loss of generality or universality of the results. Since not all 31 radiation sensitive genes are included in the top-8000 large variance genes, we generate the 31 genes contained in the gene list then complement it with the 8000th genes by means of variance. The variance distribution of the 31 genes is presented in Table S2).

RSGGC measure. In network science, a large number of centrality measures have been introduced in different contexts (Newman, 2010). For example, degree centrality (DC) represents

the number of edges of a node in the network, closeness centrality (CC) characterizes how close a pointed node from other nodes (Bavelas, 1950), betweenness centrality (BC) reflects a node's intermediary status of route or pathway in the network (Freeman, 1977; Park et al., 2004), and eigenvector centrality (EC) measures the relative influence of a node in the network (Bonacich, 1987). We focus on the EC of the correlation matrix as it is appropriate to our task. In general, the eigenvector \vec{x} associated with the maximum eigenvalue of the adjacency matrix is closely related to the asymptotic behavior of the collective dynamics on the network and control (Aguirre et al., 2013). Let x_i be the component of the eigenvector corresponding to node i . The EC of node i is given by

$$x_i = \frac{1}{\lambda} \sum_{j \in N_i} x_j = \frac{1}{\lambda} \sum_{j \in M} a_{ij} x_j \quad (2)$$

where N_i is the set of node i 's neighbors, M is the set of all nodes in the network, and a_{ij} is the element of the Pearson's correlation matrix \mathbf{A} . The vector form of Eq. 2 is

$$\mathbf{A} \cdot \vec{x} = \lambda \vec{x} \quad (3)$$

Our RSGGC measure is defined in terms of EC. In particular, for all the selected genes, RSGGC is the ratio of the EC of the focal group to that of the entire gene set:

$$\text{RSGGC} = \frac{\sum_n |x_i|}{\sum_N |x_i|} \quad (4)$$

where n is the number of genes in the radiation sensitivity gene group and $N=8000$ is the total number of genes in the whole set. For the available data set in our study, we have $n=31$ and $N=8000$. The absolute values in the sums indicate a focus on the importance of individual genes rather than distinguishing the detailed passive or active role of a specific gene in the system. The RSGGC value characterizes the topological and dynamical properties of a small group of genes in the whole gene network.

For a group, the correlation matrix represents its average level under various circumstances. The $8000 \times N$ matrices, with N being the sample size of each entire dataset, are treated as described in Tables 2 and 3 so that the within-group RSGGC values can be computed. The multi-step analysis is summarized as a workflow chart, as shown in Fig. 2.

3. Results

3.1. Power of RSGGC as a predictor of radiotherapy outcome for clinical patients

We calculate RSGGC for each classified clinical patient group, as shown in Fig. 3. The striking finding is the robust positive correlation between the RSGGC value and the survival time, indicating that the patients with a higher RSGGC score have longer expected survival time after radiotherapy. That is, radiotherapy is more effective for patients with a higher RSGGC score. More specifically, the Pearson's correlation coefficient for the clinical data in Fig. 3(A)–(C) are 0.96, 0.83, and 0.87, respectively, with the significance index values of 0.06, 0.02, and 0.19 (Table S3). After random grouping of the clinical data, the positive correlation is lost completely (Fig. S1), providing strong evidence for the reliability of RSGGC as a quantitative indicator of the effect of radiotherapy. The finding of the positive correlation between RSGGC and patient survival time is unprecedented and practically significant, as it can serve as the base for more reliable prediction of the outcome of radiotherapy for cancer patients.

3.2. Ability of RSGGC to predict radiation outcome in cell line experiments

Observation of the survival fraction of cancer cell lines in response to radiation in experiments is more straightforward and more controllable than clinical tests with patients. Our data analysis suggests that the results from experimental cell lines at the microscopic level strongly corroborate the role of RSGGC in clinical tests. Fig. 4 shows the downtrend relationship between the RSGGC value and the survival fraction (SF), which is completely consistent with the results from the clinical data. In particular, the group of cancer cell lines with higher RSGGC scores has lower survival fraction after radiation, corresponding to longer survival time of patients.

The values of Pearson's correlation coefficients of RSGGC and SF in Fig. 4 are -0.77 and -0.91 , respectively, with the significance index values of 0.01 and 0.09 (More statistics are presented in Table S3). Similar results of negative correlation between RSGGC and SF have been obtained from the two datasets with larger irradiation doses (5 Gy and 8 Gy, see Fig. S2). In clinical practice, the general protocol for radiotherapy consists of daily exposure to fractionized radiation of 2 Gy irradiation for 5–7 weeks. A dosage over 2 Gy is in fact harmful to patient's health. Thus, the relationship between the survival fraction with 2 Gy and RSGGC is practically significant for generating quantitative patient prognosis.

3.3. Further exploitation of RSGGC for GBM subtypes

The results in above focus on the relationship between RSGGC and the survival time for clinical cases of GBM, where information about the detailed subtypes of GBM is ignored. In a previous experimental study (Verhaak et al., 2010), it was found that the subtypes can have quite different radiation therapy effects. For example, radiation can have a significant effect on the subtypes "Classical", "Mesenchymal", and "Neural", but the effect is small for the subtype "Proneural". We ask whether RSGGC is capable of characterizing the radiation therapy effects at the subtypes level of GBM. To address this issue, we group cases according to the subtypes rather than the survival time and calculate the RSGGC values. Table 4 presents the detailed clinical index and RSGGC for the four subtypes: Classical, Mesenchymal, Neural, and Proneural. We see that the first three subtypes have similar values of RSGGC, which are larger than that of the fourth subtype (Proneural). This coincides well with the therapeutic effect observed from the experimental studies of these subtypes. The general result is that radiation therapy is more effective for subtypes with a larger value of RSGGC.

3.4. Mechanism of RSGGC as an estimator and predictor of radiotherapy effect

Our computations and analysis taking into account the survival time of clinical patients, the survival fraction of cancer cell lines, or different subtypes, give strong evidence that RSGGC is effective for assessing and predicting the radiotherapy effect against cancer. A plausible reason for the power of RSGGC is that the corresponding selected genes may participate in the key pathways associated with repairing DNA damage, activating cell cycle checkpoints or maintaining signal transduction pathways after the irradiation, either directly or indirectly. To verify this conjecture, we employ clinical data (Iwakawa et al., 2007) from cervical cancer patients prior to and during radiotherapy to test how the RSGGC scores calculated from the group of genes change as radiotherapy treatment is being implemented. Specifically, we classify the expression profile into two groups: prior to or during treatment, as shown in Fig. 5. We find that the RSGGC scores of patients receiving radiotherapy increase dramatically in comparison with those prior to

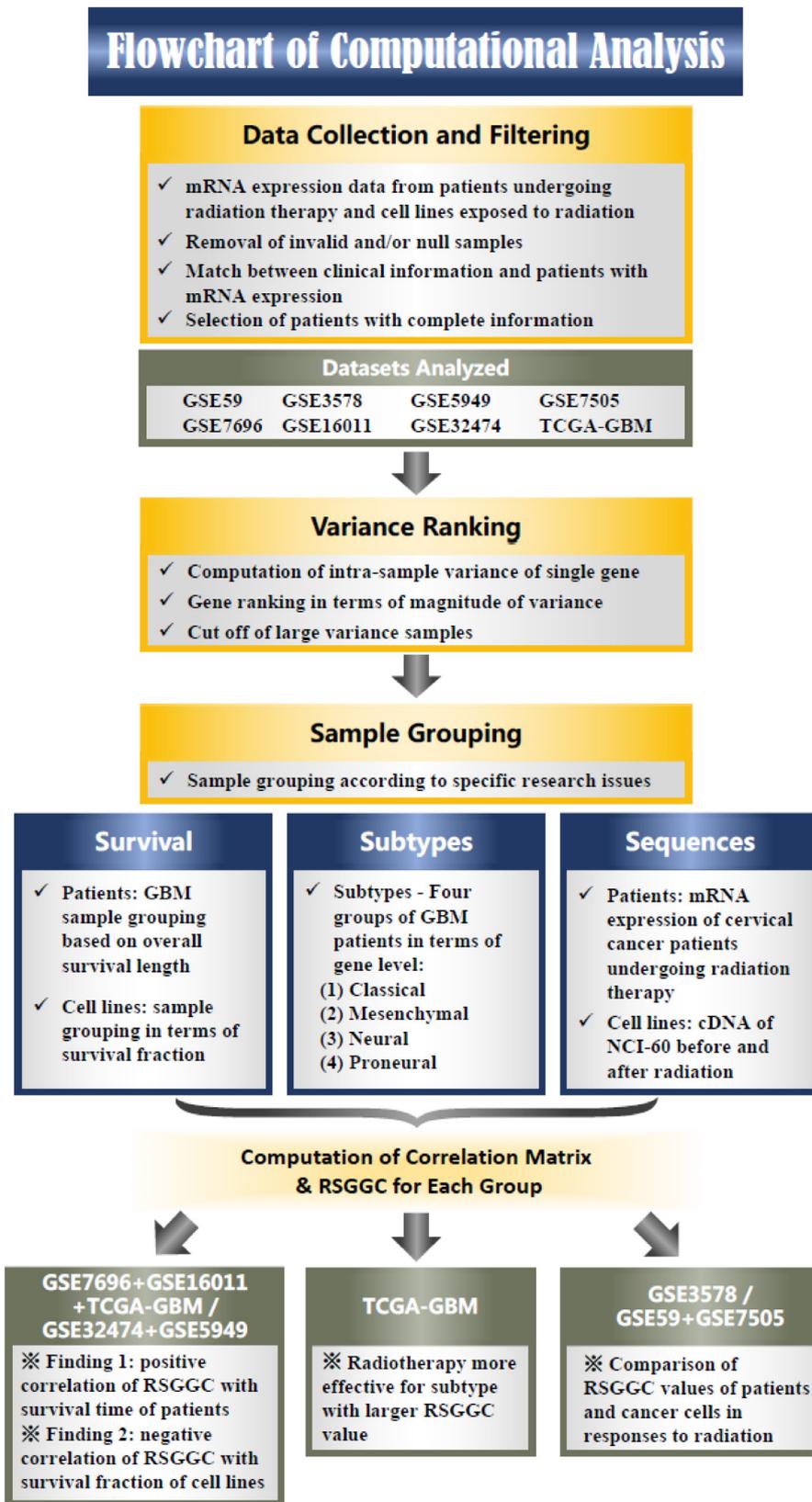


Fig. 2. Workflow of multi-step calculation of RSGGC.

the therapy. We also utilize the time sequences of samples with irradiation process obtained from two independent datasets: GSE59 and GSE7505, the cDNA microarrays of NCI-60 cell lines before and after irradiation (Amundson et al., 2008; Ross et al., 2000). Since the datasets are from different experimental platforms, we con-

sider only the common genes of the cell lines. After classifying the lines as describe in Materials and Methods, we calculate the RSGGC values before and after irradiation, as shown in Fig. 5. We find that, for the three different irradiation dosages, the RSGGC values of the groups with low SF decrease drastically while those with high SF

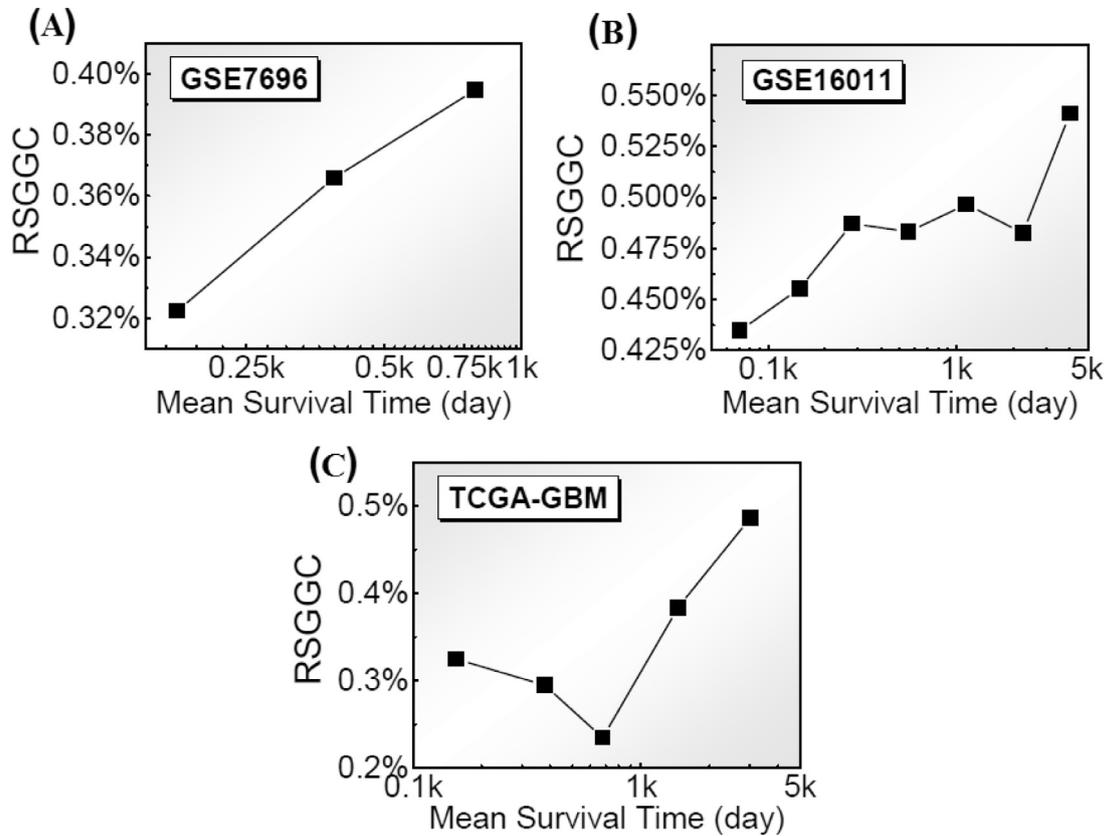


Fig. 3. Correlation between RSGGC score and survival time of GBM patients. Datasets are (A) GSE7696, (B) GSE16011, and (C) TCGA-GBM. The horizontal axis denotes the mean survival time for each sample group (black squares). A semi-logarithmic scale is used. The positive correlation between patients' RSGGC value and the mean survival time has been identified in all the datasets. The dip in the RSGGC score in the region of small survival time in (C) is due to the unusually short-living samples in the dataset.

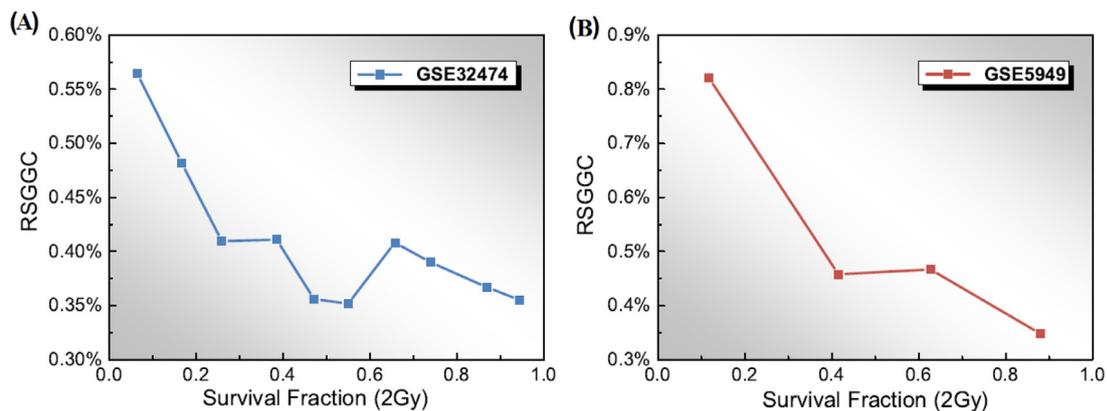


Fig. 4. Negative correlation between RSGGC and survival fraction of cancer cell lines. The datasets are (A) GSE32474 and (B) GSE5949 with 2 Gy radiation. The horizontal axis is the mean SF value of the samples within the group. RSGGC exhibits a negative relationship with the survival fraction, which is completely consistent with the results in Fig. 3.

Table 4

Characteristic description of clinical subtypes of GBM and the corresponding RSGGC values.

| Subtypes | Sample size ¹ | Age (mean) | Gender (M/F) | Therapeutic effect ² | HR ² | P-value ² | RSGGC(%) |
|-------------|--------------------------|-------------------|--------------|---------------------------------|-----------------|----------------------|----------|
| Classical | 141 | 10.9–86.59 (58.9) | 83/58 | significant | 0.45 | 0.02 | 0.321 |
| Mesenchymal | 151 | 24.4–84.8 (59.6) | 91/60 | significant | 0.54 | 0.02 | 0.357 |
| Neural | 82 | 23.1–88.6 (59.9) | 54/28 | effective | 0.56 | 0.1 | 0.334 |
| Proneural | 129 | 17.7–89.3 (53.9) | 77/52 | less effective | 0.80 | 0.4 | 0.223 |

¹ The total sample size is 503, which includes patients with clinical events (death).

² The therapeutic effect, the values of hazard ratio (HR), and P-values are from Ref. (Verhaak et al., 2010).

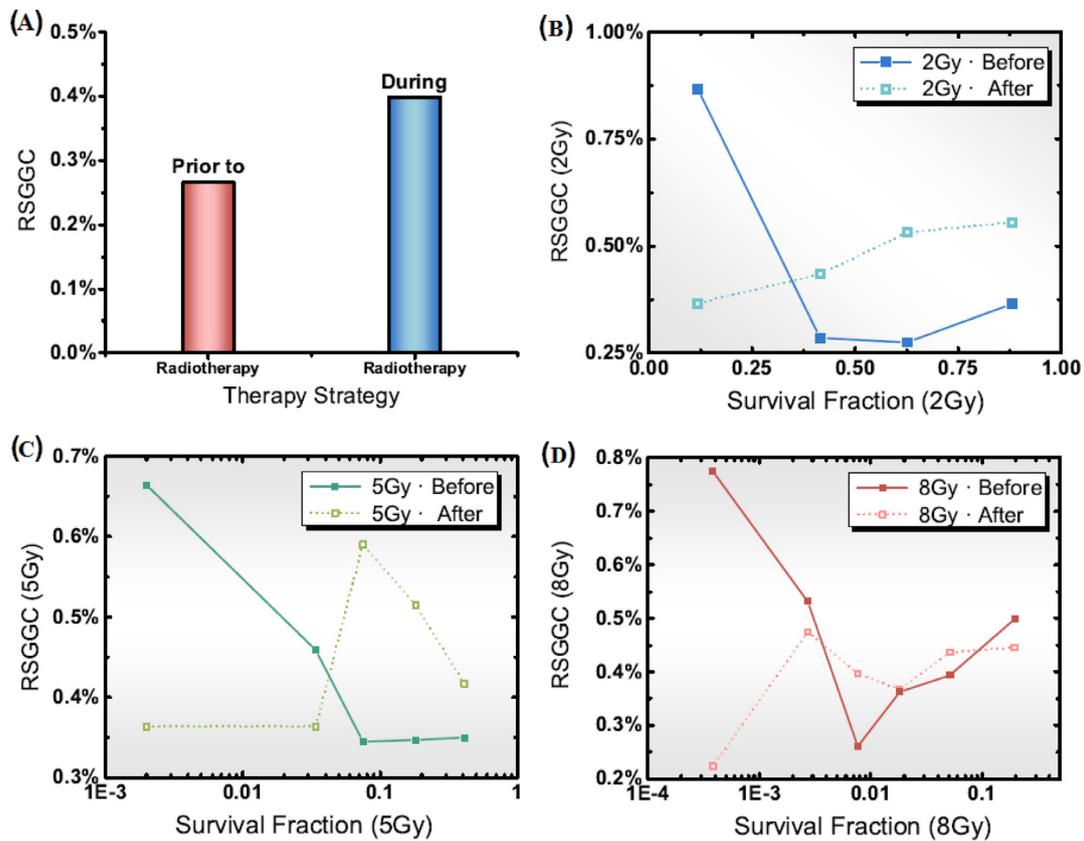


Fig. 5. Applicability of RSGGC to cervical cancer patients and cell lines. (A) RSGGC values of cervical cancer patients prior to or during irradiation treatment. There is a dramatic increase in the RSGGC value as a result of the irradiation treatment. (B–D) RSGGC values of NCI-60 cancer cell lines before and after irradiation, with dosages of 2, 5, and 8 Gy, respectively. The solid and open squares represent the RSGGC values of each group of cell lines before (from data in GSE59) and after receiving the irradiation (from data in GSE7505).

increase, indicating that the effects of irradiation at the molecular level vary for different cancer cell lines. The observable response of RSGGC to irradiation implies again its potential power in predicting the effect of irradiation therapy. In a general sense, RSGGC can effectively be regarded as a geometrical indicator of the activity of the radiosensitive gene group during irradiation. Since the centrality of the whole set of genes is normalized, an increase in the RSGGC score for a subset of genes implies that the centrality values of the remaining genes must decrease.

4. Discussion

A standard and widely used method to treat cancer patients is radiotherapy. An outstanding problem in medical science is to predict the survival time of a patient who has undergone radiotherapy. A major deficiency of previous work is the focus on gene expression in cells directly pertinent to radiosensitivity with factors such as the response of a cancer patient to irradiation and the patient survival time totally ignored. To overcome the deficiency and to devise a more accurate and reliable predictor of the patient survival time, we exploit modern complex network science to articulate a geometric approach to estimating and predicting the effect of radiotherapy on cancer. In particular, we propose a measure, the radiosensitive gene group centrality (RSGGC), that can be used to predict the survival time of a patient undergoing radiotherapy. We validate the predictive power of RSGGC by using data from both clinical patients and experimental cancer cell lines. Results from clinical data reveal a positive correlation between RSGGC and the survival time of the patients going through radiotherapy. Since, in clinical practice, a patient's prognosis is influenced by multiple

factors (Berchuck et al., 2005; Wang et al., 2015) that can introduce fluctuations in the outcomes (Table S1), we also systematically analyze data from cancer cell lines, which are more reliable due to the homogeneous microenvironment and the controllability of external conditions in experiments. Results from cancer cell lines support our finding from the clinical data in a completely consistent way: the cell lines with higher RSGGC values are more sensitive to irradiation and thus have smaller values of the survival fraction.

RSGGC as a novel indicator/predictor for characterizing radiosensitivity from a geometric viewpoint has potential advantages over the traditional clinical indicators. RSGGC can lead to new insights into understanding the relationship among the known radiosensitive gene signatures and can be used for data based analysis of extensive risk gene sets, intracellular pathways regulation and control.

Conflicts of interest

All authors have read the journal's policy on disclosure of potential conflicts of interest and declare no competing financial interests. All authors have read the journal's authorship agreement. The manuscript has been reviewed and approved by all named authors.

Funding

This work was supported by NNSF of China under Grants Nos. 11775101, and 61431012, the Fundamental Research Funds for the Central Universities Nos. lzujbky-2016-123 and xjtujbky-2018-125.

ZGH gratefully acknowledges the support of K. C. Wong Education Foundation. YCL would like to acknowledge support from the Vannevar Bush Faculty Fellowship program sponsored by the Basic Research Office of the Assistant Secretary of Defense for Research and Engineering and funded by the Office of Naval Research through Grant No. N00014-16-1-2828.

Author contributions

Z.-G.H. and L.H. designed research; Y.-X.Y. performed research; all analyzed data; Z.-T. B. gave suggestions from the viewpoint of biological research; Z.-G.H., L.H., and Y.-C.L. wrote the paper.

Data and materials availability

All data utilized in this paper are from public datasets. The access numbers are presented in main text and Supplementary Materials.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jtbi.2018.12.001.

References

- Aguirre, J., Papo, D., Buld'v, J.M., 2013. Successful strategies for competing networks. *Nat. Phys.* 9, 230–234.
- Alon, U., 2007. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 450–461 8%6.
- Altieri, D.C., 2008. Survivin, cancer networks and pathway-directed drug discovery. *Nat. Rev. Cancer* 8, 61–70.
- Amundson, S.A., Do, K.T., Vinikoor, L.C., Lee, R.A., Koch-Paiz, C.A., Ahn, J., et al., 2008. Integrating global gene expression and radiation survival parameters across the 60 cell lines of the National Cancer Institute Anticancer Drug Screen. *Cancer Res.* 68, 415–424.
- Balkwill, F., 2004. Cancer and the chemokine network. *Nat. Rev. Cancer* 4, 540–550.
- Barabasi, A.-L., Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113.
- Barthélemy, M., 2011. Spatial networks. *Phys. Rep.* 499, 1–101.
- Bavelas, A., 1950. Communication patterns in task-oriented groups. *J. Acoust. Soc. Am.* 22, 725–730.
- Berchuck, A., Iversen, E.S., Lancaster, J.M., Pittman, J., Luo, J., Lee, P., et al., 2005. Patterns of gene expression that characterize long-term survival in advanced stage serous ovarian cancers. *Clin. Cancer Res.* 11, 3686.
- Bing, Z., Tian, J., Zhang, J., Li, X., Wang, X., Yang, K., 2016. An integrative model of miRNA and mRNA expression signature for patients of breast invasive carcinoma with radiotherapy prognosis. *Cancer Biother. Radiopharma.* 31, 253–260.
- Bonacich, P., 1987. Power and centrality: a family of measures. *Am. J. Sociol.* 92, 1170–1182.
- Borgatti, S.P., Everett, M.G., 2006. A graph-theoretic perspective on centrality. *Soc. Net.* 28, 466–484.
- Brockmann, D., Helbing, D., 2013. The hidden geometry of complex, network-driven contagion phenomena. *Science* 342, 1337–1342.
- Chuang, H.-Y., Lee, E., Liu, Y.-T., Lee, D., Ideker, T., 2007. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* 3, 140.
- Eschrich, S., Zhang, H., Zhao, H., Boulware, D., Lee, J.-H., Bloom, G., et al., 2009. Systems biology modeling of the radiation sensitivity network: a biomarker discovery platform. *Int. J. Radiat. Oncol. Biol. Phys.* 75, 497–505.
- Freeman, L.C., 1977. A set of measures of centrality based on betweenness. *Sociometry* 35–41.
- Freeman, L.C., 1978. Centrality in social networks conceptual clarification. *Soc. Net.* 1, 215–239.
- Gao, H., Korn, J.M., Ferretti, S., Monahan, J.E., Wang, Y., Singh, M., et al., 2015. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* 21, 1318–1325.
- Gevaert, O., De Smet, F., Timmerman, D., Moreau, Y., De Moor, B., 2006. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics* 22, e184–e190.
- Gomez, D., González-Arangüena, E., Manuel, C., Owen, G., del Pozo, M., Tejada, J., 2003. Centrality and power in social networks: a game theoretic approach. *Math. Soc. Sci.* 46, 27–54.
- Gravendeel, L.A.M., Kouwenhoven, M.C.M., Gevaert, O., de Rooij, J.J., Stubbs, A.P., Duijm, J.E., et al., 2009. Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res.* 69, 9065–9072.
- Holme, P., Saramäki, J., 2012. Temporal networks. *Phys. Rep.* 519, 97–125.
- Ivliev, A.E., 't Hoen, P.A., Sergeeva, M.G., 2010. Coexpression network analysis identifies transcriptional modules related to proastrocytic differentiation and sprouty signaling in glioma. *Cancer Res.* 70, 10060–10070.
- Iwakawa, M., Ohno, T., Imadome, K., Nakawatari, M., Ishikawa, K.-i., Sakai, M., et al., 2007. The radiation-induced cell-death signaling pathway is activated by concurrent use of cisplatin in sequential biopsy specimens from patients with cervical cancer. *Cancer Biol. Ther.* 6, 905–911.
- Jiang, J., Huang, Z.-G., Seager, T.P., Lin, W., Grebogi, C., Hastings, A., et al., 2018. Predicting tipping points in mutualistic networks through dimension reduction. *Proc. Natl. Acad. Sci.*, 201714958.
- Kauffman, S., Peterson, C., Bor, S., Troein, C., 2003. Random Boolean network models and the yeast transcriptional network. *Proc. Natl. Acad. Sci. USA* 100, 14796–14799.
- Kim, H.S., Kim, S.C., Kim, S.J., Park, C.H., Jeung, H.-C., Kim, Y.B., et al., 2012. Identification of a radiosensitivity signature using integrative metaanalysis of published microarray data for NCI-60 cancer cells. *BMC Genomics* 13, 1.
- Lee, H.J., Lee, J.-J., Song, I.H., Park, I.A., Kang, J., Yu, J.H., et al., 2015. Prognostic and predictive value of NanoString-based immune-related gene signatures in a neoadjuvant setting of triple-negative breast cancer: relationship to tumor-infiltrating lymphocytes. *Breast Cancer Res. Treat.* 151, 619–627.
- Lee, J.K., Havaleshko, D.M., Cho, H., Weinstein, J.N., Kaldjian, E.P., Karpovich, J., et al., 2007. A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery. *Proc. Natl. Acad. Sci. USA* 104, 13086–13091.
- Moding, E.J., Kastan, M.B., Kirsch, D.G., 2013. Strategies for optimizing the response of cancer and normal tissues to radiation. *Nat. Rev. Drug Disc.* 12, 526–542.
- Murat, A., Migliavacca, E., Gorlia, T., Lambiv, W.L., Shay, T., Hamou, M.-F., et al., 2008. Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. *J. Clin. Oncol.* 26, 3015–3024.
- Newman, M.E.J., 2010. *Networks: An Introduction*. Oxford University Press.
- Oldham, M.C., Horvath, S., Geschwind, D.H., 2006. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc. Natl. Acad. Sci. USA* 103, 17973–17978.
- Palla, G., Derényi, I., Ies, F., Tas, V., 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 814–818.
- Park, K.H., Lai, Y.C., Ye, N., 2004. Characterization of weighted complex networks. *Phys. Rev. E* 70, 026109.
- Pastor-Satorras, R., Vespignani, A., 2001. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86, 3200.
- Pfister, T.D., Reinhold, W.C., Agama, K., Gupta, S., Khin, S.A., Kinders, R.J., et al., 2009. Topoisomerase I levels in the NCI-60 cancer cell line panel determined by validated ELISA and microarray analysis and correlation with indenoisoquinoline sensitivity. *Mol. Cancer Therap.* 8, 1878–1884.
- Proulx, S.R., Promislow, D.E.L., Phillips, P.C., 2005. Network thinking in ecology and evolution. *Trends Ecol. Evol.* 20, 345–353.
- Pujana, M.A., Han, J.-D.J., Starita, L.M., Stevens, K.N., Tewari, M., Ahn, J.S., et al., 2007. Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat. Genet.* 39, 1338–1349.
- Reinhold, W.C., Reimers, M.A., Lorenzi, P., Ho, J., Shankavaram, U.T., Ziegler, M.S., et al., 2010. Multifactorial regulation of E-cadherin expression: an integrative study. *Mol. Cancer Therap.* 1535–7163.
- Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Spellman, P., et al., 2000. Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 24, 227–235.
- Shoemaker, R.H., 2006. The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer* 6, 813–823.
- Stuart, J.M., Segal, E., Koller, D., Kim, S.K., 2003. A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302, 249–255.
- Van De Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A.M., Voskuil, D.W., et al., 2002. A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 347, 1999–2009.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., et al., 2001. The sequence of the human genome. *Science* 291, 1304–1351.
- Verhaak, R.G.W., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., et al., 2010. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17, 98–110.
- Volinia, S., Croce, C.M., 2013. Prognostic microRNA/mRNA signature from the integrated analysis of patients with invasive breast cancer. *Proc. Natl. Acad. Sci. USA* 110, 7413–7417.
- Wang, M.J., Ping, J., Li, Y., Adell, G., Arbman, G., Nodin, B., et al., 2015. The prognostic factors and multiple biomarkers in young patients with colorectal cancer. *Sci. Rep.* 5, 10645.
- Wang, W.-X., Lai, Y.-C., Grebogi, C., 2016. Data based identification and prediction of nonlinear and complex dynamical systems. *Phys. Rep.* 644, 1–76.
- Yarden, Y., Pines, G., 2012. The ERBB network: at last, cancer therapy meets systems biology. *Nat. Rev. Cancer* 12, 553–563.
- Zhang, B., Horvath, S., 2004. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4 Article17.
- Zhu, C.-Q., Ding, K., Strumpf, D., Weir, B.A., Meyerson, M., Pennell, N., et al., 2010. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J. Clin. Oncol.* 28, 4417–4424.