

Data Min Knowl Disc (2016) 30:226–242
DOI 10.1007/s10618-015-0410-5



A computational approach inspired by simulated annealing to study the stability of protein interaction networks in cancer and neurological disorders

Kristina Ibáñez · María Guijarro ·
Gonzalo Pajares · Alfonso Valencia

Received: 25 April 2014 / Accepted: 26 February 2015 / Published online: 21 March 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Molecular networks provide a powerful tool for the study of biomedical systems, in particular several studies have detected alterations of the network structure associated to disease states. Here we propose that diseases cannot only alter the structure of the network but also its stability. To evaluate network stability we have developed a new methodological framework. Our approach is an adaptation of the classical Deterministic Simulated Annealing algorithm to work with discrete states. Adjusted energy values are used to compare the network stability in disease and control states. The results show that cancer networks are less stable than the Alzheimer's disease (AD) ones. These results can be interpreted in terms of our previous observations on cancer and AD inverse comorbidity, i.e. AD patients have lower than expected risk to suffer cancer.

Keywords Simulated annealing · Systems biology · Gene expression · Protein interaction networks · Cancer · Neurological disorders · Inverse comorbidity

Responsible editor: Pierre Baldi.

Electronic supplementary material The online version of this article (doi:[10.1007/s10618-015-0410-5](https://doi.org/10.1007/s10618-015-0410-5)) contains supplementary material, which is available to authorized users.

K. Ibáñez (✉) · M. Guijarro · G. Pajares
Departamento Ingeniería del Software e Inteligencia Artificial, Facultad Informática,
Universidad Complutense, 28040 Madrid, Spain
e-mail: kibagar@ucm.es

K. Ibáñez · A. Valencia (✉)
Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO),
28029 Madrid, Spain
e-mail: avalencia@cnio.es

1 Introduction

Neurological disorders and cancer are two current global health priorities. Interestingly, epidemiological evidence is mounting that patients with certain neurological disorders, including those suffering from Schizophrenia (SCZ) and Alzheimer's disease (AD), have a lower than expected tendency to develop some forms of cancer (Behrens et al. 2009, 2012; Tabarés-Seisdedos and Rubenstein 2013; Tabarés-Seisdedos et al. 2011). Hence, we performed a systematic meta-analysis of gene expression in order to investigate the molecular mechanisms that might underlie such inverse comorbidity, identifying genes and pathways differentially expressed in neurological disorders and some types of cancer (Ibáñez et al. 2014). Interestingly, we found a common set of genes and biological processes that were apparently deregulated in opposing directions in cancers and neurological disorders.

Here, we set out to broaden our understanding of the molecular basis underlying the differences between cancers and neurological conditions. As such, and given that the central dogma of molecular biology dictates that information flows from genes to proteins via RNA (Crick 1970), we integrated gene expression data with protein-protein interaction networks (PPINs) in order to study these differences in terms of network organization rather than at the level of individual genes. Gene expression data informs whether a gene that encodes a given protein is active or not. Yet proteins function in the context of their interactions with other proteins, interactions that are described in PPINs in which each protein represents a node in the network.

In PPINs, it is assumed that proteins corresponding to genes that are not active (i.e.: unexpressed) will not interact with their potential partners. Therefore, the production of RNA by genes is commonly used as a proxy of the activity of the gene, and this is correlated with the activation of molecular systems within PPINs that underlie physiological and developmental processes. Indeed, in many cases, deregulation of gene expression provokes dramatic phenotypic changes, as occurs in several diseases (Kaern et al. 2005).

Protein interaction maps have been used to study the molecular organization of cellular systems and the perturbations in them created by disease. PPINs reflect the functionality of interacting proteins and for example, the consequence of a single gene deletion in the yeast *Saccharomyces cerevisiae* would appear to depend on the position of the gene product within the PPIN (Jeong et al. 2001). Thus, the proteins most important for a cell's survival are highly connected (Jeong et al. 2001; Wuchty and Almaas 2005) and altering them has profound effects on the PPIN. In terms of cancer, it is thought that cancer related proteins correspond to central hubs and that they are highly connected within networks (Jonsson and Bates 2006). Indeed, the genomic and network characteristics of genes mutated in cancer seem to confirm that these genes tend to encode central hubs within PPINs (Rambaldi et al. 2008). In addition, PPINs have been used as background layers when mapping gene expression data in order to gain information about the state of the nodes and their possible dynamics (Börnigen et al. 2013; Chuang et al. 2007; Hudson et al. 2009; Komurov and Ram 2010; Liu et al. 2013; Milanesi et al. 2009; Pujana et al. 2007; Schramm et al. 2010; Teschendorff and Severini 2010; Pel et al. 2013; West et al. 2012). For example, genes

that are over expressed in lung cancer are more strongly connected than those that are suppressed or selected at random (Wachi et al. 2005).

We hypothesize that PPINs related to cancer are more unstable than those based on neurological data. This may be because there are more active interactions between cancer related proteins and thus, a mutation or change in any of these would cause an important destabilization of the network. By contrast, proteins corresponding to genes affected in neurological disorders have less active connections and consequently, they are less susceptible to destabilization. In this context, we present an approach based on the combination of gene expression data and PPINs to study the relationship between cancers and neurological disorders. To achieve this we associate each protein (or node) in the network with a state that is directly related to the level of expression of the corresponding gene. The expression data used is derived from a large series of experiments carried out on cancer and neurological disorders in humans, information that makes the PPINs disease specific and that allows comparative studies to be performed.

In terms of the computational methodology to study the differences between disease specific networks, we have found an appropriate equivalence in the Deterministic Simulated Annealing (DSA) algorithm proposed previously (Duda et al. 2007). The DSA algorithm was designed to find the optimal solution inspired by different biological or physical phenomena. The DSA is based on the shifting of metals from an unstable state as a liquid to a stable solid state, a process mediated by a decrease in the temperature of the material. These transformations can be simulated by the evolution of the states of interconnected network nodes that evolve until an optimal solution with minimal energy is reached. This evolution is controlled through an energy minimization process that determines the network's stability as the energy decreases. Therefore, lower energies correspond to greater stability.

Inspired by the DSA algorithm, we designed and implemented a new method to measure the stability of PPINS based on a defined energy function. In this approach the concept of stability differs from that in the original DSA, in which the network evolves towards states with different stabilities via temporal transitions or another equivalent value (Cruz García et al. 2011, 2002; Pajares and Cruz 2004; Sánchez-Lladó et al. 2011). The proposed approach used in this study computes energy based on existing interactions and it computes the energy difference between two states, such as disease and control samples. In this manner, the temporal aspect of the original DSA is reduced to the comparison between a reference and a new model. The reference state can be considered to be equivalent to the initial state and the new model as a single progressive step. Furthermore, any simulated annealing process (DSA or probabilistic) is driven by an optimization process in order to achieve stable states (minimum energy values). By contrast, since only one transition is considered in our approach, there is no optimization process involved and local minima energy are avoided. These substantial differences from the original DSA are introduced to make it possible to perform a large scale systematic comparison of networks associated to cancer, neurological disorders and normal controls for which the information available comes from experiments carried out at only one time point, representing a single state of these conditions.

2 Materials and methods

2.1 Materials

The protein interaction and gene expression data used in this study were obtained from PPIN and Gene expression data sets.

2.1.1 *The protein–protein interaction network (PPIN)*

We used the human PPIN from the protein interaction network analysis database (PINA, <http://cbg.garvan.unsw.edu.au/pina/interactome.stat.do>, version October 2011. Online Resource 1: Wu et al. 2009). PINA is an integrated platform of PPIN data that has been extracted from six different public databases: IntAct, MINT, BioGRID, DIP, HPRD, and MIPS/MPact. It includes self-interactions, interactions predicted by computational methods, and interactions between human proteins and proteins from other species. Moreover, it has recently been used in other similar studies (Xia et al. 2011; Laakso and Hautaniemi 2010).

Besides the PINA network, we also used two additional PPINs in order to guarantee that a similar outcome was obtained: The Human Protein Reference Database (HPRD, <http://www.hprd.org/download>, version April 2010) that contains pairs of human protein interactions based on experimental evidence from the literature and that has been used in several studies (Teschendorff and Severini 2010; West et al. 2012); and the Human Integrated protein–protein interaction rEference (HIPPIE, <http://cbdm.mdc-berlin.de/tools/hippie/download.php>, version September 2014) that incorporates a human PPI dataset with a normalized scoring scheme, integrating data from HPRD, BioGRID, IntAct, MINT, Rual05, Lim06, Bell09, Stelzl05, DIP, BIND, Coland04, Lehner04, Albers05, MIPS, Venkatesan09, Kaltenbach07 and Nakayama02. We selected the interactions from these PPINs with a curated score above 0.73 in order to be confident that the pairs of proteins interact (Schaefer et al. 2012).

2.1.2 *Gene expression data sets*

Measuring gene expression with microarrays is now a common molecular biology approach in biomedicine, making it possible to simultaneously measure the relative expression of thousands of genes under different experimental conditions (*Current Topics in Computational Molecular Biology*, 2002). Thousands of gene expression data sets are available in public databases, each containing a description of the corresponding biomedical origin of the sample, the analytic procedures followed and the experimental results in terms of expression (i.e.: the amount of RNA produced for each gene in the genome).

Raw experimental gene expression data (CEL files) for Ovarian, Colon, Liver and Kidney datasets were downloaded from the Barcode human transcriptome repository (Gene Expression Barcode, <http://barcode.luhs.org/>), and for the SCZ and AD datasets they were downloaded from the NCBI GEO omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>) and Stanley Medical Research Institute Online Genomics Database (SMRI, <https://www.stanleygenomics.org>; Online Resource 2). Importantly, each dataset cor-

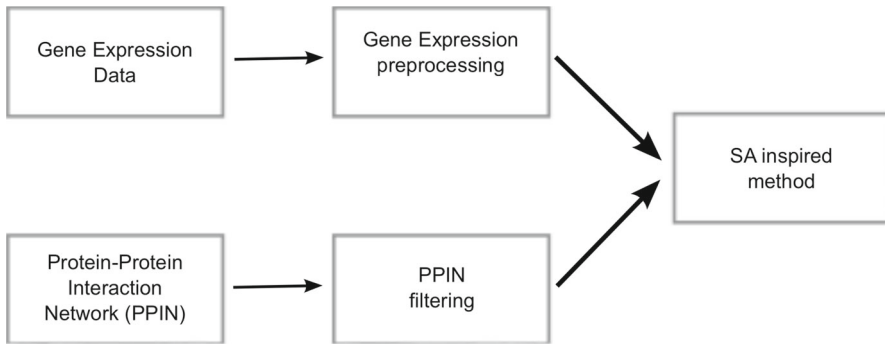


Fig. 1 Flow chart of the overall study

responds to a collection of disease and control samples. For the analysis we filtered out the cases with too few disease/control cases (less than 9) and we only used those produced in the same platform (Affymetrix array GeneChip Human Genome U133 Plus 2.0), rendering information on 23,945 human genes. This technical platform has been widely used, and using the same platform on all data sets facilitates comparative studies and ameliorates potential experimental errors.

2.2 Methods

In order to study the stability of the PPIN in cancer, neurological and normal samples, we implemented an original method inspired by the well-known DSA approach that was customized to study *neighbor-energy* (nE). In this case, stability describes a network state that is not significantly altered, even when fundamental properties have changed or perturbations have been introduced. From the biological point of view, network instability could reflect a situation where mutations in a key protein involved in many interactions will alter several associated biological processes.

A filtered PPIN (Sect. 2.2.1), and preprocessed and normalized gene expression data (Sect. 2.2.3) for three different conditions (cancer, normal and neurological disorders), were the inputs for our approach (Sect. 2.2.4). A scheme of the workflow is presented in Fig. 1, where preprocessing and filtering are clearly represented as two separate modules.

2.2.1 Protein–protein interaction network filtering

Data from the PINA network were filtered by requiring experimental evidence for PPIs, removing redundancy and self-interactions, as well as interactions involving proteins that were not from *Homo sapiens*. Thus, we only considered those interactions between proteins that were also detected in the Human Genome U133 Plus 2.0 microarray platform. The resulting filtered PINA network contains 10,650 proteins with 63,119 interactions. Each node denotes a protein encoded by a gene and each edge denotes an interaction between two proteins (Fig. 2a).

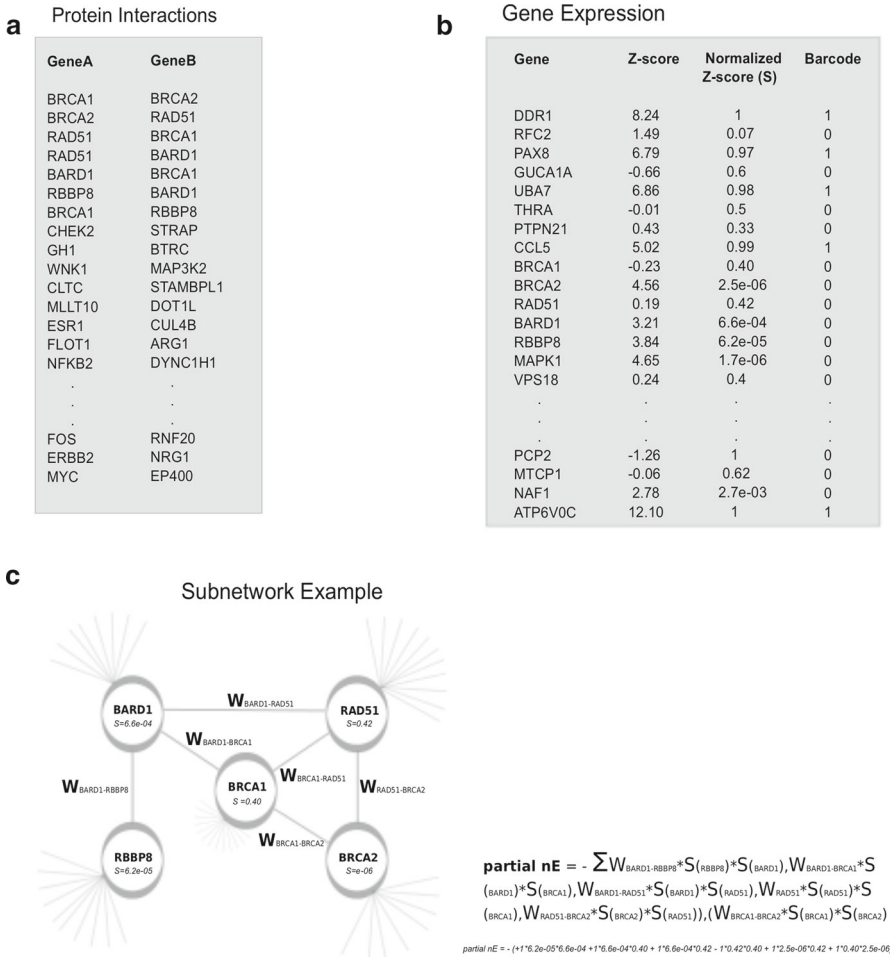


Fig. 2 **a** Filtered PPIN. **b** Preprocessed and normalized gene expression data. **c** Representation of the nodes in a PPIN sub-network and application of the algorithm to a PPIN sub-network

2.2.2 Sub-network related to the synaptic vesicle cycle

A sub-network of proteins encoded by genes related to the synaptic vesicle cycle was analyzed, retrieving proteins in the synaptic vesicle (SV) cycle from the KEGG pathway (http://www.genome.jp/dbget-bin/www_bget?pathway:hsa04721, version September 2014). The number of genes involved in the SV cycle pathway are 63, and 50 out of 63 genes were detected in microarrays. The resulting sub-network contains 50 proteins and 3815 interactions.

2.2.3 Microarray gene expression preprocessing

Handling microarrays requires the preprocessing of each individual microarray to estimate the expression of each gene in the array. Gene expression data from Ovarian,

Colon, Liver and Kidney cancers, and from SCZ and AD samples, were normalized by frozen Robust Multiarray Analysis (fRMA: (McCall et al. 2012) from the R Affy package (Gautier et al. 2004). Background-corrected gene intensities were obtained by applying fRMA processes to each array individually, and accounting for probe variability, batch effects, probe effects, array-to-array variability and background noise. The samples were then processed using Barcode (McCall et al. 2012) in order to convert gene intensities into estimates of gene expression (Z-score, Fig. 2b). Additionally, gene intensities were mapped into a binary vector of “ones” and “zeros” that denote whether a gene was expressed (1, when the Z-score is higher than a threshold value: 4.98 by default) or not (0) in each sample (Fig. 2b and Supplementary Material: McCall et al. 2011; Zilliox and Irizarry 2007). These values were used in Eq. 1, in which it is not necessary to specify whether a gene is expressed or not.

To compare the Z-score between these diseases, we normalized them using the `pnorm` function of the R stats package to calculate the normal distribution function of each Z-score. This normalization step is commonly employed to avoid values in a given range dominating other values. High Z-scores indicate intense gene expression, while small Z-scores correspond to weak expression. For expressed genes, defining S as the normalized Z-score, $S = \text{pnorm}(Z\text{-score})$, represents the probability of the gene being expressed. When the gene is not expressed, $S = 1 - \text{pnorm}(Z\text{-score})$ indicates the probability of the gene not being expressed. These S values were used in Eq. 2. Hence, each state in the system would represent the significance (S) of the expression of each gene (Fig. 2b). In summary, for each disease we associated a binary value reflecting whether or the gene is expressed (one or zero, respectively), attributing a value and a significance to the expression each of the 10,650 genes in the network (Fig. 2b).

2.2.4 Adapted simulated annealing approach

To study network stability we adopted an approach based on the SA concept, a probabilistic method that allows the global minimum of a generic cost function to be found (Kirkpatrick et al. 1983; Cerny 1985). This procedure reproduces the way the structure of a solid reaches its minimum energy configuration through cooling, becoming “frozen” at this minimum energy.

A full description of the DSA is included in Online Resource 3 (Duda et al. 2007; Haykin 1994), which also follows a physical analogy based on a set of interconnected nodes, each one with its associated state. During the cooling process forces between interconnected nodes act on the structure, which evolves until each node reaches a stable state. Thus, the nodes interacting with other nodes within the system influence one another with a defined weight.

Our algorithm is inspired on the definition of a nE function that measures the stability of the network, as well as on the general deterministic approach whereby a lower nE is related to greater stability. In our case, using a nE function that decreases in function of the interactions or over time does not make sense given the characteristics of the biological problem. Indeed, our approach does not evolve through iterations or time and thus, this part of the algorithm was not considered.

Our system is represented by a PPIN in which nodes represent proteins associated to the expression of the corresponding gene (S_i describes the significance (S) of a gene i being expressed or not). The DSA approach is applied to estimate the dynamic structures in the PPIN (Fig. 2c), where S_i represents the state of the node in the original DSA approach and the edges reflect the interactions existing between proteins. Each W_{ij} represents the weight required (Eq. 1), where W_{ij} is inversely associated to the existence of the interaction between two proteins. If the two genes i and j are both expressed, then the two corresponding proteins can interact (W_{ij} value -1). The value of W_{ij} will be $+1$ if the interaction is not possible because one of the two genes is not expressed.

$$W_{ij} = \begin{cases} -1 & \text{if } i \text{ expressed, } j \text{ expressed} \\ +1 & \text{if } i \text{ or } j \text{ not expressed} \\ +1 & \text{if } i \text{ not expressed, } j \text{ not expressed} \end{cases} \quad (1)$$

Consistent with the main idea of the SA algorithm, the $local_nE$ is defined as the sum of the energy from all the nodes connected to a given node i . This influence is calculated by multiplying the expression of each gene (normalized value of expression, S) by the associated weights of the connected nodes (W_{ij}), as summarized in (2).

$$local_nE(i) = - \sum_j W_{ij} * S_i * S_j \quad (2)$$

According to the definition in Eq. 2, the $local_nE$ is maximal when $W_{ij} * S_i * S_j$ is at its minimum, representing active connections between nodes of expressed genes (Eq. 1, case 1) and indicating that any alteration in this node will destabilize the network.

The value of the $local_nE$ decreases for those node connections that involve at least one gene that is not expressed in that condition, reflecting the fact that the interactions cannot take place (Eq. 1, cases 2 and 3). In these situations, the $local_nE$ achieves its minimum values indicating network stability.

The $local_nE$ function measures the stability of a single protein or node i in function of its neighborhood, i.e. only with respect to the directly interacting partners and not within the entire network. The global nE value (Eq. 3), and therefore the stability of the entire network, will be a consequence of the equilibrium between interactions among active (corresponding to the expressed genes) and inactive nodes (corresponding to non-expressed genes).

$$nE = \sum_i local_nE(i) \quad (3)$$

2.2.5 Computation of network robustness

To assess the robustness of the system, we analyzed how the network structure changes as nodes are removed in accordance with previously defined procedures (Iyer et al. 2013). Changes in the network structure are evaluated in terms of the size of the largest connected component of the network. Networks in which the largest component decreases faster than that of the original network are considered to be less robust to

perturbations. Thus, nodes were removed in decreasing order of their *local_nE* scores (Eq. 2), removing the proteins (or nodes) with higher *local_nE* values first (i.e.: those with more active connections) and those with the lowest *local_nE* scores last (i.e.: those less connected to their neighbors).

Network robustness was measured through the *R-index* in Eq. 4, where α corresponds to the size of the largest connected component within the network after a node is removed.

$$R = \frac{1}{N} \sum_{i=1}^N \alpha(i/N) \quad (4)$$

We computed the *R-index* for cancer and normal control samples at each step after the removal of nodes in function of the order of *local_nE* scores.

3 Results

Using this new approach, we have analyzed four gene expression datasets for cancer (Ovarian, Colon, Liver and Kidney), four data sets for SCZ and five for AD (Online Resource 1), each having sufficient disease and control samples, and fulfilling our quality control criteria (see Sect. 2). For each disease and data set, PPIN stability was assessed in both the disease and control samples. In other words, we simulated a weighted interaction network for each sample, mapping S into the PPIN, directly applying the proposed algorithm and obtaining a *nE* value. The distribution of the *nE* values for the normal (N) and disease (C) conditions were then studied (Fig. 3) and a global *nE* was obtained for each disease.

3.1 Increased neighbor-energy in cancer tissue

The cancer PPINs present characteristic instability, reflected by higher *nE* values than their normal control samples (Fig. 3a). A Mann–Whitney (Wilcoxon-rank) test was used to evaluate whether the medians of a test variable differed significantly between the normal and cancer samples, which proved to be the case for each tissue (represented below the x-axis). Indeed, very significant Wilcoxon test *p* values were obtained for the Ovarian, Colon, Liver and Kidney data sets (3.11e−04, 2.62e−03, 2.10e−05 and 2.33e−08, respectively), indicative of meaningful and important differences between the *nE* distributions in cancer and normal samples, with cancer samples being considerably less stable than their normal counterparts.

3.2 Decreased neighbor-energy in tissues from neurological disorders

Significant differences in *nE* distributions were evident when AD (C) and normal (N) samples were compared (Fig. 3b), and significant Wilcoxon *p* values were obtained for the *nE* distribution in virtually all of the AD studies. AD samples had smaller *nE* values than the normal samples, reflecting increased stability (decreased instability)

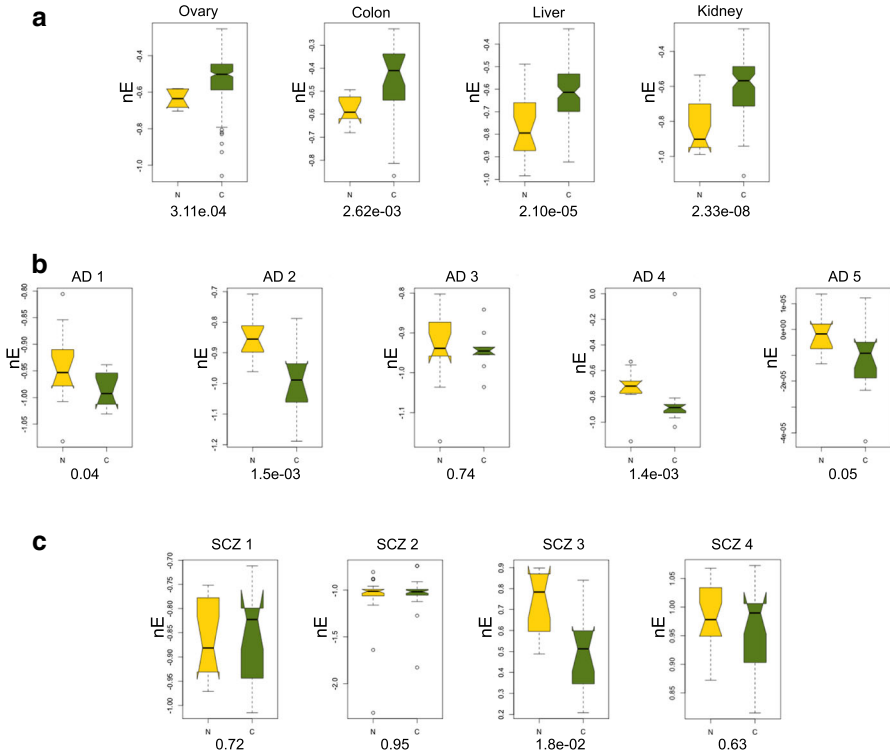


Fig. 3 The nE distribution that maps all the genes in the PINA network in the: **a** normal (N) and cancer (C) states (Ovarian, Colon, Liver and Kidney); **b** normal (N) and AD (C); **c** normal (N) and SCZ disease (C) state. The Wilcoxon-rank p value is presented below the x -axis

in the AD network. By contrast, we only observed relevant differences between the nE distributions of the normal and disease states for one of the four SCZ data sets available (Fig. 3c). This discrepancy between the different SCZ networks suggests that further studies are required for this condition and the underlying cause is unlikely to be revealed until new, high quality experimental datasets become available.

Similar results were obtained when networks other than PINA networks were used, including a smaller HPRD network (Online Resource 4) and a larger HIPPIE one (Online Resource 5). It is important to clarify whether these differences are the product of general differences in expression between cancer, normal and neurological disease tissues. However, the normalized expression data (Fig. 4) indicated that there was no difference between the global levels of normalized expression in this study.

3.3 Consistency of the results

In order to assess the consistency of the results we analyzed sub-networks obtained by randomly sampling the complete network. Accordingly, 86 % of the sub-networks containing 10 % of the proteins of the original PINA network produced similar results

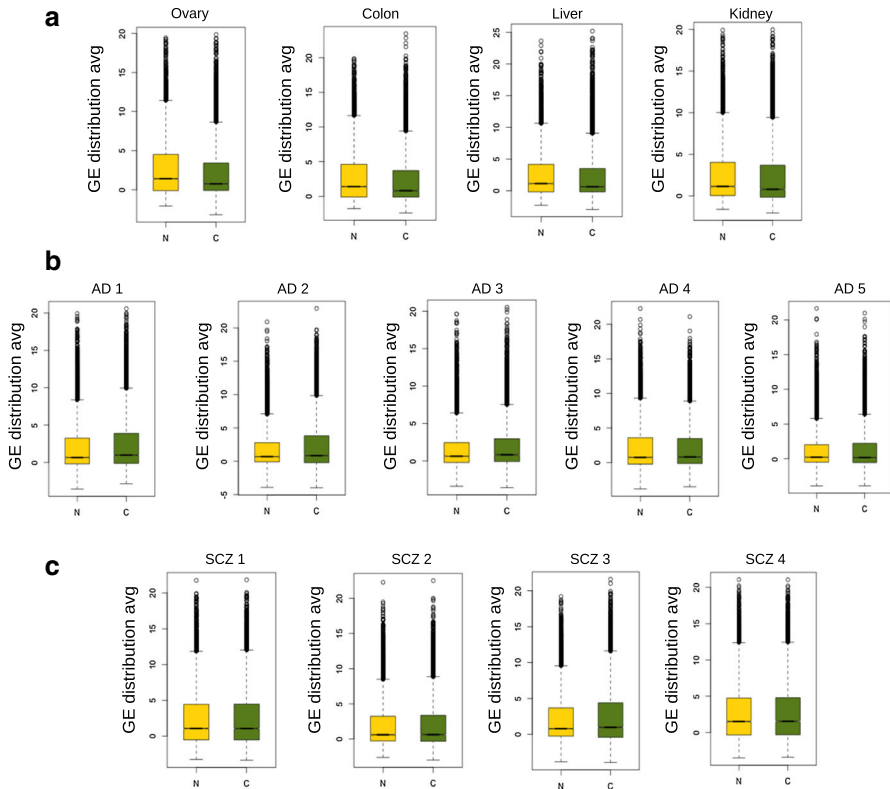


Fig. 4 Gene expression distribution in: **a** normal (N) and cancer (C) states (Ovarian, Colon, Liver and Kidney); **b** normal (N) and AD (C) conditions; and **c** normal (N) and SCZ (C) conditions

to the complete network. In other words, not only was there significant instability in the overall network but most of the regions of the network conformed to this behavior, with only a few of them behaving distinctly (Online Resource 6 which includes the nE scores for the first one hundred random sub-samples).

3.4 Increased neighbor-energy in cancer evolution

To further study the network instability in cancer, we assessed whether tumor progression might be related with increased instability. Indeed, the initial results showed a significant increase in network instability when the datasets obtained at different stages of tumor progression were compared (Fig. 5).

3.5 Network stability towards perturbations

Stability has previously been described as a relatively invariant network state when perturbations are introduced. Thus, it is necessary to perform additional experiments to show that our definition of network stability measured through the nE score corre-

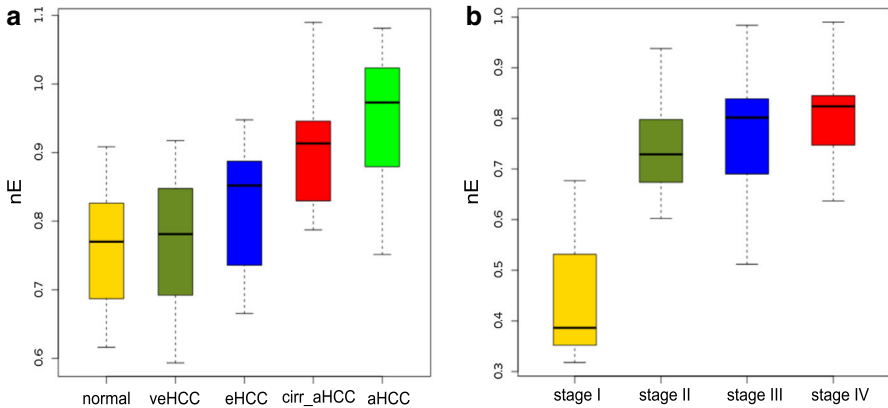


Fig. 5 The nE distributions mapping all the genes at: **a** evolving stages of hepatocellular carcinoma (HCC)—normal, very early HCC (veHCC), early HCC (eHCC), cirrhosis with HCC (cirr_aHCC) and advanced HCC (aHCC); and **b** progressive stages of colon cancer stages I, II, III and IV

lates well with this classical definition of robustness. Removing nodes from a network and then studying the evolution of the network's connectivity provides a natural model to study the robustness of networked systems (Iyer et al. 2013; Callaway et al. 2000; Cohen et al. 2000). Accordingly, the R -index can be used to quantify network robustness (see Sect. 2.2.5).

The successive removal of nodes according to their $local_nE$ score produced a significant difference between the perturbation robustness in cancer and normal samples (Fig. 6a), and in AD and normal samples (Fig. 6b). When nodes were removed in a descending order of $local_nE$ scores, greater robustness was evident in normal control networks (R -index = 0.52) than in cancer networks (R -index = 0.33; Fig. 6a). By contrast, AD networks are more robust (R -index = 0.50) than their corresponding normal control networks (R -index = 0.39; Fig. 6b). Hence, the definition of the nE score appears to be closely associated to network stability and as such, with the network's robustness to perturbation.

3.6 Decreased instability in biological pathways implicated in Alzheimer's disease

We analyzed the decreased network instability observed for AD samples in more detail and in particular, we investigated the possible role of the proteins implicated in vesicle trafficking at synapses. Communication between neurons is mediated by the release of neurotransmitter from SVs and the expression of a group of genes involved in SV trafficking is reduced in brain tissues from AD cases. Indeed, the loss of synapses has been correlated with cognitive decline in AD and a malfunction of SV trafficking could be implicated in disrupting neuronal circuits in AD (Yao et al. 2003).

As for the complete PPIN, there was a consistent decrease in instability in the SV related sub-network of proteins from AD samples (Online Resource 7a). The difference in the nE score suggests that important hubs within the network are expressed and regulated in opposite directions in AD and normal samples. Indeed, nine genes related

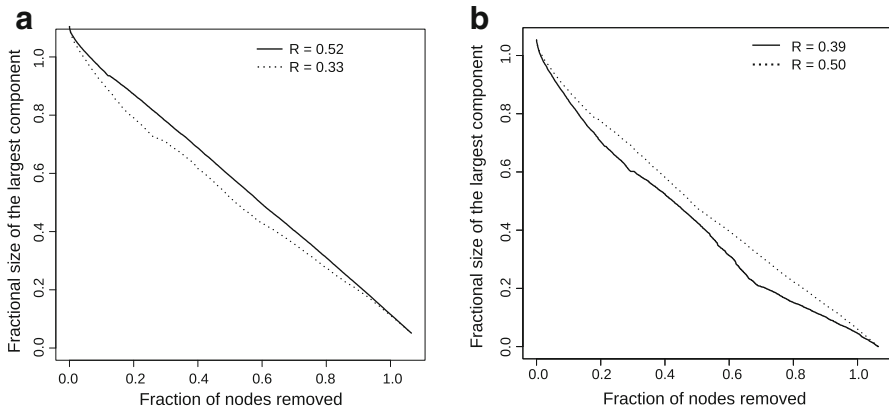


Fig. 6 Perturbation robustness against *local_nE* sorted score in: a cancer samples (*dotted line*) and normal controls (*solid line*); and in b AD samples (*dotted line*) and normal controls (*solid line*)

to endocytosis were expressed in opposite manners in normal and AD samples: *KIT*, *CLTA*, *CLTB*, *AP2M1*, *AP2S1*, *AP2B1*, *HLA-B1*, *AP2A2*, and *RAB11FIB2*. Three genes associated with SV trafficking (*SYP*, *STX1A* and *UNC13B*) were inversely expressed in both conditions and they were highly connected in the protein network (hubs). In particular syntaxin 1A (*STX1A*) is known to regulate the exocytosis of SVs and neurotransmitter release (Bennett and Scheller 1993; Greengard et al. 1993; Hosaka et al. 1999). There was a clear trend towards reduced *STX1A* expression in all AD samples, which had a lower *nE* score than in normal control samples. Indeed, when the *STX1A* gene was not expressed (in blue) nor were its neighbors and conversely, when the *STX1A* gene was expressed (in red) so were most of its neighbors (Online Resources 7b and 7c). Accordingly, the stability of a particular sub-network relevant to a neurological disease under study is affected in the same way as the stability of the entire network.

4 Discussion

In this work we have designed an approach inspired on SA, representing PPINs as systems of nodes that are dynamically updated towards a global state of stability. Our strategy is based on the definition of a neighbor-energy function that measures the stability of the network in the general deterministic approach, where *nE* indicates network stability, and it can be interpreted in terms of resistance to alterations and perturbations. In this study, we analyzed a large set of experimental data on gene expression and various PPINs.

The first significant finding of this study is that networks containing information about expression in four human cancers (Ovarian, Colon, Kidney and Liver) are less stable than the control networks of normal samples. Moreover, this instability in the network seems to increase as these cancers evolve, at least in the tumor progression data sets analyzed. The approach employed is based on the analyses of samples in different conditions and it does not include temporal evolution per se. Thus, the results obtained

by analyzing the temporal progression of tumors can be taken as an indication of network evolution towards a less stable state and a way of reconciling our methodology with the standard SA applications.

The randomness or disorder in the local flux distribution surrounding any given node in the network i has been quantified (West et al. 2012), showing that cancer is characterized by an increase in network entropy. This observation could be considered as independent confirmation of our general conclusion. Indeed, when gene expression data was previously integrated with a PPIN for six cancer tissues (Teschendorff and Severini 2010), an increase in network entropy was again seen to be associated to cancer based on a fluctuation theorem of dynamic systems theory. At the biological level cancer has been associated with a general destabilization of cellular processes related to the organization of the genome, its replication and repair (Murga and Fernández-Capetillo 2007). A conceptual framework explains how mutations in genes that control genetic stability are selected during tumor progression (Loeb 2011; Negrini et al. 2010; Solé et al. 2014; Wadhwa et al. 2013). Therefore, our observation of network instability in cancers fits well with current ideas in this field.

Technically, our approach offers important advantages. First, raw gene expression data sets are divergent and independent, which represents an important difference. Additionally, we use a high quality filtered and curated PPIN, which while having practically the same number of total nodes it is less connected than those used in earlier studies. To deal with our biological problem we need to consider both the state of the nodes as well as the strength of the connections between them. This is possible with methods where these two important issues are considered, such as DSA, one of the generic means to resolve the optimization problem (Kirkpatrick et al. 1983).

Our second important finding is that the AD network is more stable than its control normal network, with a significant increase in the nE of the corresponding networks. This is an interesting behavior that contrasts with that of cancers, and as far as we know is detected here for the first time. One possible interpretation of these results would be that cancer implies a general deregulation of cell growth through the hyperactivation of certain pathways, resulting in a destabilization of their interactions, while AD and other neurological disorders imply the stabilization of biological processes and network interactions, and their general slowing down. The striking contrast in the behavior of cancer and AD networks, from less to more stable networks, should be considered in the context of the observed “inverse comorbidity” of these two groups of diseases. A substantial number of epidemiological studies have shown that there is an inverse relationship between cancer and several central nervous system diseases, including AD. In other words, patients with AD tend to less frequently suffer some types of cancer (Tabarés-Seisdedos and Rubenstein 2013; for a complete meta-study of the available epidemiological studies see Catalá-López et al. 2013).

Finally, given the importance of the diseases discussed in this work, it is necessary to make these results accessible for future experimental analysis. In this sense, an initial study of the molecular basis of this inverse comorbidity identified sets of genes expressed weakly in AD and strongly in cancers (Ibáñez et al. 2014). The new methodological approach developed here represents a further advance with respect to that initial approximation, where genes are not considered as independent units but rather as part of a connected network. This approach could be used as a classi-

fier to distinguish cancer and normal samples. Another possibility will be to cluster the results of this procedure in order to extract specific proteins for which additional experimental information could be available, or could be tracked in direct experiments. Furthermore, this scheme could be also applied to any network system where the elements are characterized by a state S_i and their interactions associated to a weight W_{ij} . In a biological context there are numerous systems with these characteristics, such as protein interaction and gene control networks. In the future, the application of clustering techniques to disease networks, such as Self Organizing Maps (SOM), will render information not on single genes but on clusters of collaborating genes, moving towards the study of the molecular causes of comorbidity to the level of systems biology.

Acknowledgments This work was supported by an Obra Social la Caixa grant (to K. I.) and Grant BIO2012-40205. We thank Clara Higuera, Anaïs Baudot, Daniel Rico and the reviewers for helpful reading, advice and comments.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Behrens MI, Lendon C, Roe CM (2009) A common biological mechanism in cancer and Alzheimer's disease? *Curr Alzheimer Res* 6(3):196–204
- Behrens MI, Silva M, Salech F, Ponce DP, Merino D, Sinning M, Xiong C, Roe CM, Quest AFG (2012) Inverse susceptibility to oxidative death of lymphocytes obtained from Alzheimer's patients and skin cancer survivors: increased apoptosis in Alzheimer's and reduced necrosis in cancer. *J Gerontol* 67(10):1036–1040. doi:[10.1093/gerona/glr258](https://doi.org/10.1093/gerona/glr258)
- Bennett MK, Scheller RH (1993) The molecular machinery for secretion is conserved from yeast to neurons. *Proc Natl Acad Sci USA* 90(7):2559–2563
- Börnigen D, Pers TH, Thorrez L, Huttenhower C, Moreau Y, Brunak S (2013) Concordance of gene expression in human protein complexes reveals tissue specificity and pathology. *Nucleic Acids Res* 41(18):e171. doi:[10.1093/nar/gkt661](https://doi.org/10.1093/nar/gkt661)
- Callaway DS, Newman ME, Strogatz SH, Watts DJ (2000) Network robustness and fragility: percolation on random graphs. *Phys Rev Lett* 85(25):5468–5471
- Catalá-López F, Gènova-Maleras R, Vieta E, Tabarés-Seisdedos R (2013) The increasing burden of mental and neurological disorders. *Eur Neuropsychopharmacol* 23(11):1337–1339. doi:[10.1016/j.euroneuro.2013.04.001](https://doi.org/10.1016/j.euroneuro.2013.04.001)
- Cerny V (1985) Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm I. *J Optim Theory Appl* 45(1):41–51
- Chuang HY, Lee E, Liu YT, Lee D, Ideker T (2007) Network-based classification of breast cancer metastasis. *Mol Syst Biol* 3:140. doi:[10.1038/msb4100180](https://doi.org/10.1038/msb4100180)
- Cohen R, Erez K, ben-Avraham D, Havlin S (2000) Resilience of the internet to random breakdowns. *Phys Rev Lett* 85(21):4626–4628
- Crick F (1970) Central dogma of molecular biology. *Nature* 227(5258):561–563
- de la Cruz García JM, Herrera Caro PJ, Pajares Martinsanz G, Guijarro Mata-García M (2011) Combining support vector machines and simulated annealing for stereovision matching with fish eye lenses in forest environments. *Expert Syst Appl* 38(7):8622–8631
- de la Cruz García JM, Herrera Caro PJ, Pajares Martinsanz G, Guijarro Mata-García M (2002) Current topics in computational molecular biology. MIT Press, Cambridge, p 542
- Duda RO, Hart PE, Stork DG (2007) Pattern classification, New York: John Wiley & Sons, 2001, pp. xx + 654, ISBN: 0-471-05669-3. *J Classif* 24(2):305–307. doi:[10.1007/s00357-007-0015-9](https://doi.org/10.1007/s00357-007-0015-9)
- Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxf, Engl)* 20(3):307–315. doi:[10.1093/bioinformatics/btg405](https://doi.org/10.1093/bioinformatics/btg405)

- Greengard P, Valtorta F, Czernik AJ, Benfenati F (1993) Synaptic vesicle phosphoproteins and regulation of synaptic function. *Science* 259(5096):780–785
- Haykin S (1994) *Neural networks: a comprehensive foundation*. Macmillan, New York
- Hosaka M, Hammer RE, Südhof TC (1999) A phospho-switch controls the dynamic association of synapsins with synaptic vesicles. *Neuron* 24(2):377–387
- Hudson NJ, Reverter A, Dalrymple BP (2009) A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput Biol* 5(5):e1000382. doi:[10.1371/journal.pcbi.1000382](https://doi.org/10.1371/journal.pcbi.1000382)
- Ibáñez K, Boullosa C, Tabarés-Seisdedos R, Baudot A, Valencia A (2014) Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses. *PLoS Genet* 10(2):e1004173. doi:[10.1371/journal.pgen.1004173](https://doi.org/10.1371/journal.pgen.1004173)
- Iyer S, Killingback T, Sundaram B, Wang Z (2013) Attack robustness and centrality of complex networks. *PLoS One* 8(4):e59613. doi:[10.1371/journal.pone.0059613](https://doi.org/10.1371/journal.pone.0059613)
- Jeong H, Mason SP, Barabási AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411(6833):41–42. doi:[10.1038/35075138](https://doi.org/10.1038/35075138)
- Jonsson PF, Bates PA (2006) Global topological features of cancer proteins in the human interactome. *Bioinformatics (Oxf, Engl)* 22(18):2291–2297. doi:[10.1093/bioinformatics/btl390](https://doi.org/10.1093/bioinformatics/btl390)
- Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 6(6):451–464. doi:[10.1038/nrg1615](https://doi.org/10.1038/nrg1615)
- Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. *Science* 220(4598):671–680. doi:[10.1126/science.220.4598.671](https://doi.org/10.1126/science.220.4598.671)
- Komurov K, Ram PT (2010) Patterns of human gene expression variance show strong associations with signaling network hierarchy. *BMC Syst Biol* 4:154. doi:[10.1186/1752-0509-4-154](https://doi.org/10.1186/1752-0509-4-154)
- Laakso M, Hautaniemi S (2010) Integrative platform to translate gene sets to networks. *Bioinformatics (Oxf, Engl)* 26(14):1802–1803. doi:[10.1093/bioinformatics/btq277](https://doi.org/10.1093/bioinformatics/btq277)
- Liu CH, Chen TC, Chau GY, Jan YH, Chen CH, Hsu CN, Lin KT, Juang YL, Lu PJ, Cheng HC, Chen MH, Chang CF, Ting YS, Kao CY, Hsiao M, Huang CYF (2013) An analysis of protein-protein interactions in cross-talk pathways reveals CRKL as a novel prognostic marker in hepatocellular carcinoma. *Mol Cell Proteomics* 12(5):1335–1349. doi:[10.1074/mcp.O112.020404](https://doi.org/10.1074/mcp.O112.020404)
- Loeb LA (2011) Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat Rev Cancer* 11(6):450–457. doi:[10.1038/nrc3063](https://doi.org/10.1038/nrc3063)
- McCall MN, Jaffee HA, Irizarry RA (2012) rRNA ST: frozen robust multiarray analysis for Affymetrix Exon and Gene ST arrays. *Bioinformatics (Oxf, Engl)* 28(23):3153–3154. doi:[10.1093/bioinformatics/bts588](https://doi.org/10.1093/bioinformatics/bts588)
- McCall MN, Uppal K, Jaffee HA, Zilliox MJ, Irizarry RA (2011) The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes. *Nucleic Acids Res* 39(Database issue):D1011–D1015. doi:[10.1093/nar/gkq1259](https://doi.org/10.1093/nar/gkq1259)
- Milanesi L, Romano P, Castellani G, Remondini D, Liò P (2009) Trends in modeling biomedical complex systems. *BMC Bioinform* 10(Suppl 1):I1. doi:[10.1186/1471-2105-10-S12-11](https://doi.org/10.1186/1471-2105-10-S12-11)
- Murga M, Fernández-Capetillo O (2007) Genomic instability: on the birth and death of cancer. *Clin Transl Oncol* 9(4):216–220
- Negrini S, Gorgoulis VG, Halazonetis TD (2010) Genomic instability—an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* 11(3):220–228. doi:[10.1038/nrm2858](https://doi.org/10.1038/nrm2858)
- Pajares G, de la Cruz JM (2004) On combining support vector machines and simulated annealing in stereovision matching. *IEEE Trans Syst Man Cybern B Cybern* 34(4):1646–1647
- Pujana MA, Han J-DJ, Starita LM, Stevens KN, Tewari M, Ahn JS et al (2007) Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat Genet* 39(11):1338–1349
- Rambaldi D, Giorgi FM, Capuani F, Ciliberto A, Ciccarelli FD (2008) Low duplicability and network fragility of cancer genes. *Trends Genet* 24(9):427–430. doi:[10.1016/j.tig.2008.06.003](https://doi.org/10.1016/j.tig.2008.06.003)
- Sánchez-Lladó FJ, Pajares G, López-Martínez C (2011) Improving the wishart synthetic aperture radar image classifications through deterministic simulated annealing. *ISPRS J Photogramm Remote Sens* 66(6):845–857. doi:[10.1016/j.isprsjprs.2011.09.007](https://doi.org/10.1016/j.isprsjprs.2011.09.007)
- Schaefer MH, Fontaine J-F, Vinayagam A, Porras P, Wanker EE, Andrade-Navarro MA (2012) HIPPIE: integrating protein interaction networks with experiment based quality scores. *PLoS One* 7(2):e31826. doi:[10.1371/journal.pone.0031826](https://doi.org/10.1371/journal.pone.0031826)
- Schramm G, Kannabiran N, König R (2010) Regulation patterns in signaling networks of cancer. *BMC Syst Biol* 4:162. doi:[10.1186/1752-0509-4-162](https://doi.org/10.1186/1752-0509-4-162)

- Solé RV, Valverde S, Rodriguez-Caso C, Sardanyés J (2014) Can a minimal replicating construct be identified as the embodiment of cancer? *BioEssays: News Rev Mol Cell Dev Biol* 36(5):503–512. doi:[10.1002/bies.201300098](https://doi.org/10.1002/bies.201300098)
- Tabarés-Seisdedos R, Dumont N, Baudot A, Valderas JM, Climent J, Valencia A, Crespo-Facorro B, Vieta B, Gómez-Beneyto M, Martínez S, Rubenstein JL et al (2011) No paradox, no progress: inverse cancer comorbidity in people with other complex diseases. *Lancet Oncol* 12(6):604–608. doi:[10.1016/S1470-2045\(11\)70041-9](https://doi.org/10.1016/S1470-2045(11)70041-9)
- Tabarés-Seisdedos R, Rubenstein JL (2013) Inverse cancer comorbidity: a serendipitous opportunity to gain insight into CNS disorders. *Nat Rev Neurosci* 14(April):293–304. doi:[10.1038/nrn3464](https://doi.org/10.1038/nrn3464)
- Teschendorff AE, Severini S (2010) Increased entropy of signal transduction in the cancer metastasis phenotype. *BMC Syst Biol* 4(1):104. doi:[10.1186/1752-0509-4-104](https://doi.org/10.1186/1752-0509-4-104)
- Van Pel DM, Barrett IJ, Shimizu Y, Sajesh BV, Guppy BJ, Pfeifer T, McManus KJ, Hieter P (2013) An evolutionarily conserved synthetic lethal interaction network identifies FEN1 as a broad-spectrum target for anticancer therapeutic development. *PLoS Genet* 9(1):e1003254. doi:[10.1371/journal.pgen.1003254](https://doi.org/10.1371/journal.pgen.1003254)
- Wachi S, Yoneda K, Wu R (2005) Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics (Oxf, Engl)* 21(23):4205–4208. doi:[10.1093/bioinformatics/bti688](https://doi.org/10.1093/bioinformatics/bti688)
- Wadhwa N, Mathew BB, Jatawa SK, Tiwari A (2013) Genetic instability in urinary bladder cancer: an evolving hallmark. *J Postgrad Med* 59(4):284–288. doi:[10.4103/0022-3859.123156](https://doi.org/10.4103/0022-3859.123156)
- West J, Bianconi G, Severini S, Teschendorff AE, Genomics SC (2012) Differential network entropy reveals cancer system hallmarks. *Sci Rep* 2:802. doi:[10.1038/srep00802](https://doi.org/10.1038/srep00802)
- Wu J, Vallenius T, Ovaska K, Westermarck J, Mäkelä TP, Hautaniemi S (2009) Integrated network analysis platform for protein–protein interactions. *Nat Methods* 6(1):75–77. doi:[10.1038/nmeth.1282](https://doi.org/10.1038/nmeth.1282)
- Wuchty S, Almaas E (2005) Peeling the yeast protein network. *Proteomics* 5(2):444–449. doi:[10.1002/pmic.200400962](https://doi.org/10.1002/pmic.200400962)
- Xia J, Sun J, Jia P, Zhao Z (2011) Do cancer proteins really interact strongly in the human protein–protein interaction network? *Comput Biol Chem* 35(3):121–125. doi:[10.1016/j.compbiolchem.2011.04.005](https://doi.org/10.1016/j.compbiolchem.2011.04.005)
- Yao PJ, Zhu M, Pyun EI, Brooks AI, Therianos S, Meyers VE, Coleman PD (2003) Defects in expression of genes related to synaptic vesicle trafficking in frontal cortex of Alzheimer’s disease. *Neurobiol Dis* 12(2):97–109
- Zilliox MJ, Irizarry RA (2007) A gene expression bar code for microarray data. *Nat Methods* 4(11):911–913. doi:[10.1038/nmeth1102](https://doi.org/10.1038/nmeth1102)