nature biotechnology

Dynamic modularity in protein interaction networks predicts breast cancer outcome

Ian W Taylor[1,2], Rune Linding[1,3], David Warde-Farley[4,5], Yongmei Liu[1], Catia Pesquita[6], Daniel Faria[6], Shelley Bull[1,7], Tony Pawson[1,2], Quaid Morris[4,5] & Jeffrey L Wrana[1,2]

Changes in the biochemical wiring of oncogenic cells drives phenotypic transformations that directly affect disease outcome. Here we examine the dynamic structure of the human protein interaction network (interactome) to determine whether changes in the organization of the interactome can be used to predict patient outcome. An analysis of hub proteins identified intermodular hub proteins that are co-expressed with their interacting partners in a tissue-restricted manner and intramodular hub proteins that are co-expressed with their interacting partners in all or most tissues. Substantial differences in biochemical structure were observed between the two types of hubs. Signaling domains were found more often in intermodular hub proteins, which were also more frequently associated with oncogenesis. Analysis of two breast cancer patient cohorts revealed that altered modularity of the human interactome may be useful as an indicator of breast cancer prognosis.

Transcriptome analyses have been extensively applied as molecular diagnostic and prognostic tools in breast cancer. Recently, the prognostic predictive performance of gene expression signatures has been improved by incorporating interactome data[1], suggesting that altered gene expression in breast cancer might disturb the higher-level organization of the interactome and affect disease outcome.

To investigate this possibility, we first identified proteins that have many interacting partners (so called 'hubs') in a network of protein-protein interactions curated from the literature and high-throughput sources[2] (**Supplementary Fig. 1a** online). Next, we obtained genome-wide expression data measured in 79 human tissues[3], and quantified the extent to which a hub and its interacting partners were co-expressed in the same tissues (**Supplementary Methods** online). We used the average Pearson correlation coefficient (PCC) of co-expression of a hub protein and its partners to identify whether interactions are context specific (that is, interacting proteins are not always co-expressed) or constitutive (that is, interacting proteins are always co-expressed). This revealed a multi-modal distribution that appeared to be the superposition of distinct populations of hubs centered over increasing average PCC values (**Fig. 1a**, red asterisks). Randomly reassigning the expression data to different gene products in the same network resulted in an approximately normal distribution of PCC values (**Fig. 1a**, black dashed line). The shoulder (marked with a black asterisk) is largely due to strongly correlated gene products that have a high probability of reforming interactions with their true interactors when randomized (data not shown). We observed a similar multi-modal distribution using a literature-curated source alone[4] (**Supplementary Fig. 1b**) or a different high-confidence human PPI database[5] (**Supplementary Fig. 1c**).

The human interactome thus has two classes of hubs. One class displays low correlation of co-expression with its partners. We call these hubs intermodular hubs, as first proposed for the yeast interactome[6,7]. A second class, termed intramodular hubs, displays more highly correlated patterns of co-expression (**Fig. 1a**). These features reflect a modular architecture. Restricting the analysis to interactions conserved between yeast and humans revealed a single peak at high average PCC, suggestive of largely intramodular hubs (**Fig. 1b**). Previous analyses showed that the assembly of intramodular hubs into macromolecular complexes constrains intramodular hub evolution[6]. This is visualized as a cluster of highly correlated interactions interconnecting intramodular hubs in the human interactome (**Supplementary Fig. 1a**; green edges between blue nodes).

Modular structure can confer higher-order function to interactomes, such that intermodular hubs provide temporally and spatially restricted linkages to intramodular hubs that in turn fulfill specific functions, often as multi-subunit macromolecular machines[8,9]. For example, most components of the 26S proteasome show highly correlated expression and function together to mediate protein degradation (**Supplementary Fig. 2a** online). However, three hub components (PSMB1, PSMB2 and PSMD9) are intermodular, reflecting tissue-specific modulation of the proteasome[10,11]. Using the Gene Ontology (GO) molecular function database[12], we found that intramodular hubs shared more functional similarity with their partners than did intermodular hubs (Student's *t*-test, $P < 0.02$, **Supplementary Fig. 2b**).

[1]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Ave., Toronto, Ontario M5G 1X5, Canada. [2]Department of Molecular Genetics, University of Toronto, 1 King's College Circle, Room 4396, Toronto, Ontario M5S 1A8, Canada. [3]Cellular & Molecular Logic Team, Institute of Cancer Research (ICR), Section of Cell, Molecular & Systems Section, 237 Fulham Road, London, SW3 6JB, UK. [4]The Terrence Donnelly Centre for Cellular and Biomolecular Research, 160 College St., Toronto, Ontario M5S 3E1, Canada. [5]Department of Computer Science, University of Toronto, 10 King's College Road, Room 3303, Toronto, Ontario M5S 3G4, Canada. [6]Faculty of Sciences, University of Lisbon, Campo Grande, 1749-016, Lisbon, Portugal. [7]Dalla Lana, School of Public Health, University of Toronto, 155 College St., Toronto, ON M5T 3M7, Canada. Correspondence should be addressed to J.W. (wrana@lunenfeld.ca).

Received 25 September 2008; accepted 18 December 2008; published online 1 February 2009; doi:10.1038/nbt.1522

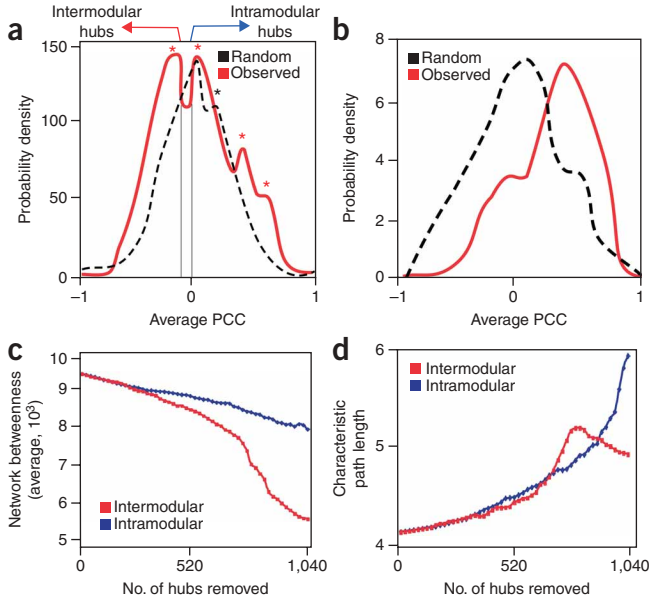

Figure 1 Evidence of dynamic network modularity in the human interactome. (**a**) The probability density of the average PCC of co-expression for human hub proteins with their interactors across 79 human tissues (red line) is compared to randomized data (dashed black line). (**b**) Same as (**a**) but only using human hub proteins conserved in yeast (red line) compared to randomized data (dashed black line). (**c**) Network betweenness as a function of removing equivalent numbers of intermodular or intramodular hubs. (**d**) Characteristic path length of the network as a function of removing equivalent numbers of intermodular or intramodular hubs.

Intermodular hubs have been proposed to be critical for global network connectivity[7]. We tested this by systematically removing either intermodular or intramodular hubs from the interaction network and analyzing the number of paths between nodes using a topological measure known as 'betweenness'[13]. Betweenness measures information flow through networks, with high betweenness reflecting multiple paths between nodes and low betweenness few paths. In a biological context, betweenness measures the ways in which signals can pass through the interaction network. Betweenness was more strongly affected by removing inter- rather than intramodular hubs (**Fig. 1c**). Another topological measure of global network connectivity is the characteristic path length (CPL), which is the average of the shortest path between all nodes in a network[14]. Systematic removal of intermodular hubs increased CPL to a threshold beyond which CPL rapidly collapsed due to splintering of the large network into small subnetworks (**Fig. 1d**). In contrast, intramodular hub removal only increased CPL. The greater sensitivity of both betweenness and CPL to removal of intermodular hubs is consistent with the notion that the human interactome is modular with intermodular hubs connecting functional modules that are comprised of intramodular hubs.

Next, we asked whether hub types display characteristic biochemical features. We found that intermodular hubs were larger than intramodular hub proteins (Mann-Whitney U-test, $P < 0.005$, **Supplementary Fig. 3a** online). Analysis of domain numbers (modularity) and size (globularity) revealed intermodular hubs have more domains compared to a randomized distribution, whereas intramodular hubs have fewer domains than expected by chance ($P < 0.05$ and $P < 0.01$ respectively, **Fig. 2a**). Conversely, intramodular hubs have greater globularity (domain size) and intermodular hubs less ($P < 0.05$ and $P < 0.01$, respectively, **Fig. 2b**). Linear motifs (that is, post-translational modifications and short binding motifs[15]) are over- and under-represented in intermodular and intramodular hubs, respectively ($P < 0.005$, **Fig. 2c**; **Supplementary Fig. 3b**).

We then explored domain types in the different hub classes. Cell signaling domains (as defined by the SMART database[16]) were enriched in intermodular hubs (sign test, $P < 0.001$), whereas

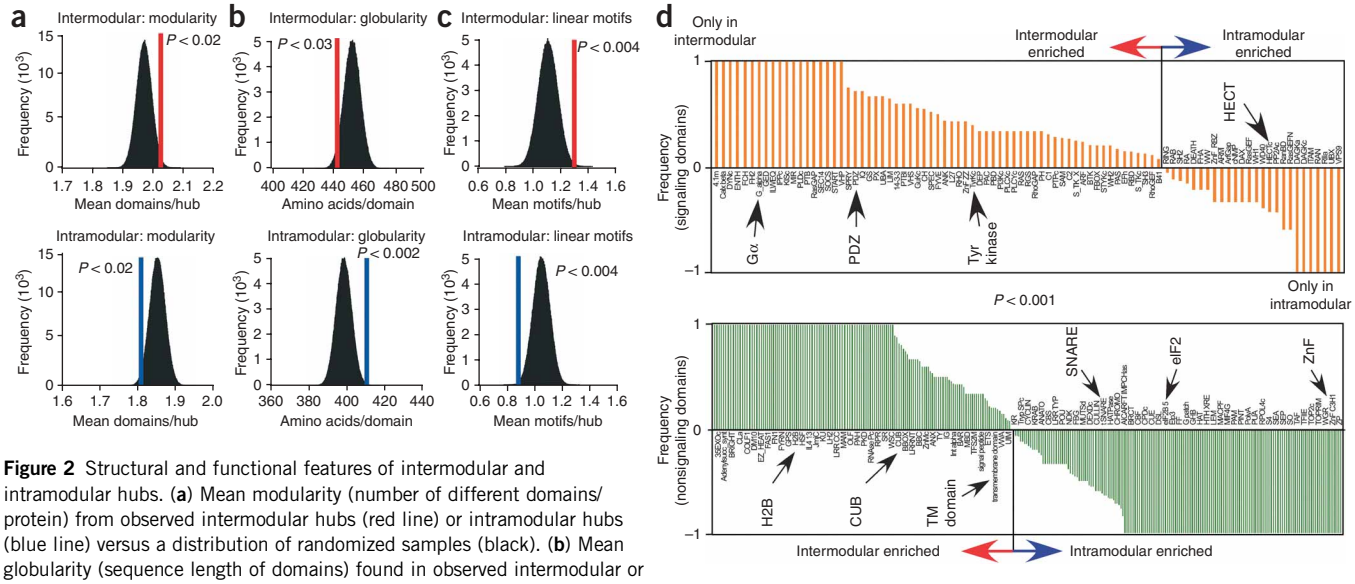

Figure 2 Structural and functional features of intermodular and intramodular hubs. (**a**) Mean modularity (number of different domains/protein) from observed intermodular hubs (red line) or intramodular hubs (blue line) versus a distribution of randomized samples (black). (**b**) Mean globularity (sequence length of domains) found in observed intermodular or intramodular hubs compared to randomized distributions. (**c**) Mean number of experimentally validated linear motifs and phosphosites from the ELM and Phospho-ELM database in intermodular or intramodular hubs compared to randomized distributions. (**d**) Domain distribution between intermodular hubs and intramodular hubs. The frequency of individual domains in intermodular hubs minus their frequency in intramodular hubs was plotted for each of the signaling domains (top panel, orange bars) or non-signaling domains (bottom panel, green bars), as indicated. A frequency of 1 indicates domains are found exclusively in intermodular hubs, whereas a frequency of –1 indicates exclusively intramodular hubs. Note that to retain legibility only a fraction of nonsignaling domains are labeled.

nonsignaling domains were evenly distributed between the hub types (**Fig. 2d**). For example, tyrosine kinase, PDZ and Gα domains were found predominantly or exclusively in intermodular hubs (**Fig. 2d**). The two hub types have similar degree distributions (that is, number of interactions per hub; **Supplementary Fig. 4** online), indicating that the biochemical attributes of hub proteins are an inherent property of the hub type and are not a function of the number of interacting partners. Taken together, these results indicate that intra- and intermodular hubs display distinctive structural characteristics consistent with their roles in organizing communication and function of dynamic protein networks.

To explore this in detail we examined the well-characterized RAS subnetwork. RAS behaves as an intramodular hub, with many highly correlated regulatory partners, such as RALGDS and SOS (**Supplementary Fig 5a** online). In contrast, partners that employ RAS as an effector (that is, Insulin receptor adaptor protein, IRS1 (ref. 17)) or a regulator (that is, BRAF[17]) tended to be intermodular. The latter is connected to a large cluster of intramodular transcription factors, such as NFκB and p53. Also notable is that connections between the RAS module and the downstream intramodular cluster occur almost exclusively via intermodular hubs. This suggests a modular assembly of signaling networks with intermodular hubs organizing the interconnectivity of functional modules such as RAS and the downstream RAS transcriptional effectors.

During tumor progression, rewiring of signaling networks drives phenotypic alterations while maintaining the robustness of the network[8], suggesting that there may be differences in hub-type association with cancer. We queried Online Mendelian Inheritance in Man (OMIM)[18], the census of cancer genes[19], and oncogenic translocations and found that mutations of intermodular hubs were associated with cancer phenotypes more frequently than those of intramodular hubs (Fisher's exact test, $P < 0.05$, **Supplementary Figs. 5b,c** and **6** online). As intermodular hubs regulate the global functions of modular networks, these results suggest that alterations in network modularity may occur in cancer.

To investigate this we analyzed a well-described cohort of sporadic, nonfamilial breast cancer patients[20]. We first looked for significant differences in the average PCC of hub proteins and their interacting partners in patients who were disease free after extended follow-up (hereafter referred to as 'good outcome') and those who died of disease ('poor outcome') (**Supplementary Fig. 7** online). This revealed 256 hubs that displayed altered PCC as a function of disease outcome. One such hub was BRCA1, a protein that is mutated in a subset of familial breast cancers. The expression of BRCA1 was strongly correlated with the expression of its partners in tumors from surviving patients, but not well correlated with their expression in tumors from poor-outcome patients (**Fig. 3a**). In contrast, the transcription factor Sp1, which shares some interacting partners with BRCA1, was not significantly changed. Of the BRCA1 partners highly correlated in good outcome tumors, both MRE11 and BRCA2 were notable as they are members of the BRCA1-associated genome surveillance complex (BASC) and are misregulated in poor prognosis breast cancer[21,22]. Our results suggest that disorganization of the BASC by loss of coordinated co-expression of components is associated with poor outcome.

Analysis of interactions between the 256 hub proteins revealed that they form an interconnected network (**Fig. 3b**). Notably, we did not identify hubs that were themselves significantly up or downregulated in the good versus poor outcome groups, but rather we identified hubs that had altered PCC of expression between outcome groups (**Supplementary Fig. 7**). Of the 256 hubs identified in our study, only 23% (59 hubs) showed significantly altered expression in our cohort when analyzed using 'significance analysis of microarrays'[23].

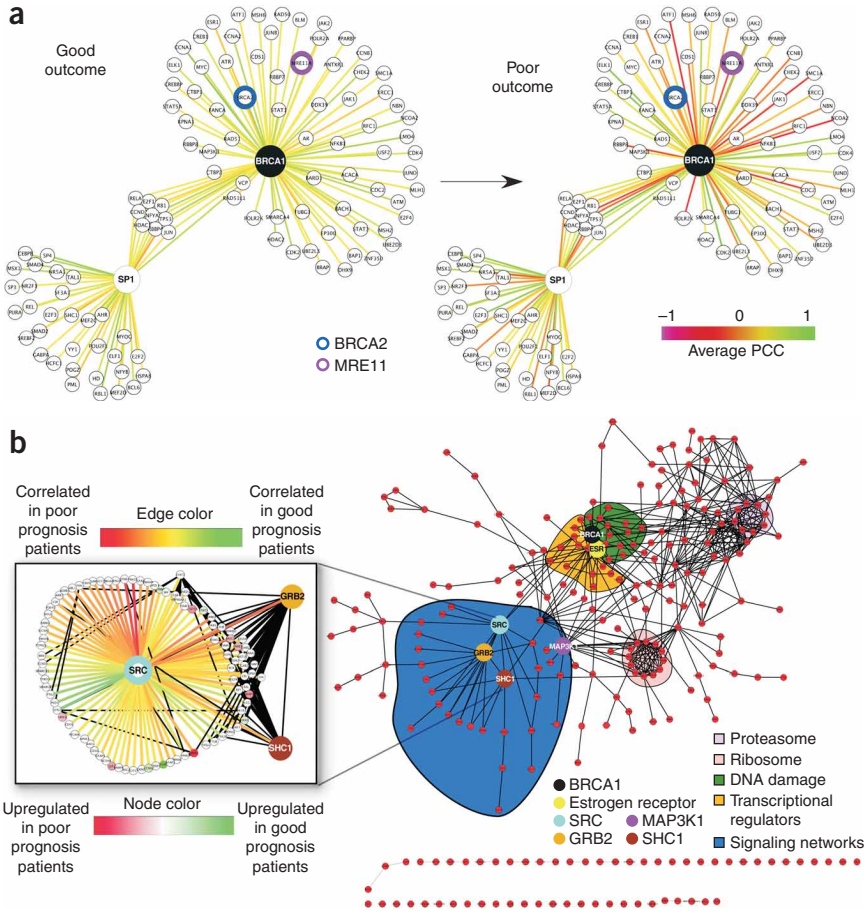

Figure 3 Differences in dynamic network properties in breast cancer tumors. (**a**) Network of the interacting partners of BRACA1 and SP1. BRCA1 and its interactors (e.g., BRCA2 and MRE11, as indicated) are highly ordered (green edges indicate correlated expression between protein pairs) in the surviving patients, whereas that organization is lost in patients who die of disease. Interactions involving Sp1 are not significantly altered. (**b**) Shown are all hubs (red nodes) that have, as a function of patient outcome, significantly different correlation of co-expression with their partners. Black edges connect hubs that have direct protein-protein interactions. Note that most hubs are components of a an interconnected network. The network includes many functional groups known to be misregulated in breast cancer pathogenesis (highlighted in legend). Inset shows a subnetwork focused on SRC and its interactors together with GRB2 and SHC1. Edge colors represent the correlation between SRC and each of its partners, while node colors represent changes in gene expression between outcome groups. Black edges indicate interactions not involving SRC. Note that while SRC is not significantly differently expressed between patient groups, it is a significant predictor hub because of differences in the coordinated co-expression of SRC and many of its partners.

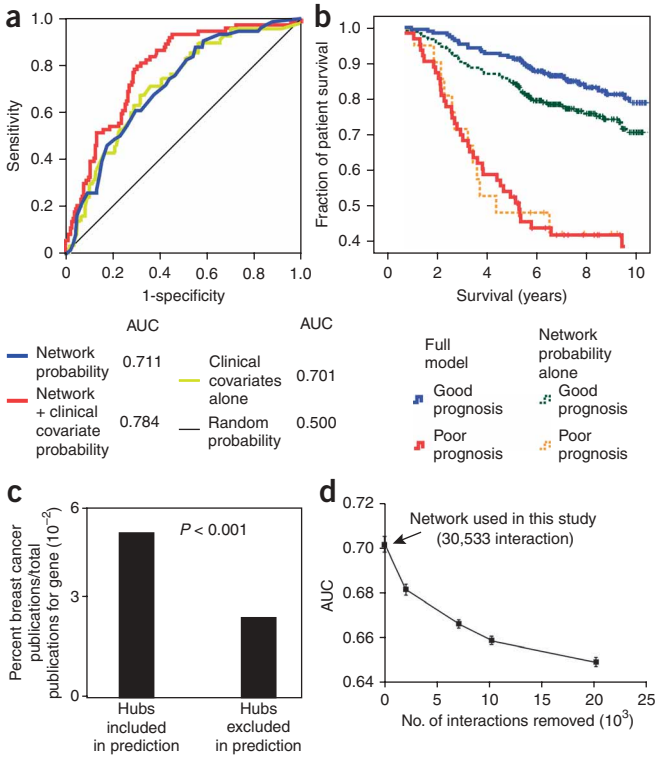

Figure 4 Dynamic network properties predict breast cancer outcome. (**a**) ROC curve of the probabilities for prognostic group membership from the affinity propagation clustering of patient dynamic network properties using fivefold cross-validation runs. Outcome prediction performances are shown for network probabilities alone (blue line), TNM tumor classifications alone (yellow line) and combining network properties of each tumor and TNM tumor classifications (red line). Random division of patients is shown with the black diagonal). (**b**) Kaplan-Meier disease-free survival curves. Patients were grouped into good and poor prognostic groups based on a fivefold cross-validation analysis of patient data. Patient survival is plotted for network probability alone (green and orange lines, as indicated) or network probability controlling for clinical covariates (red and blue lines). (**c**) Genes encoding hub proteins that are included in the prediction algorithm are cited significantly more frequently in the breast cancer literature than excluded hubs. (**d**) Algorithm performance declines as a function of decreasing interactome size. Interactions were randomly removed from the current interactome as indicated and performance of the dynamic network modularity algorithm assessed. Average AUC (+s.d.) at each of the reduced interactome sizes is plotted (black squares) and was calculated from 5-fold cross-validation runs performed in triplicate.

For example, no significant difference in the expression level of the oncogene product SRC was observed between groups (**Fig. 3b**, inset); however, the coordinated co-expression of SRC and its regulators or effectors (see inset **Fig. 3b**) was clearly affected. Unbiased analysis of the 256 hubs in this aberrant network demonstrated over-representation in literature (**Fig. 4**, Fisher's exact test, $P<0.001$) and microarray studies[20,24–26] of breast cancer (**Supplementary Fig. 8** online, Fisher's exact test, $P<0.02$) when compared to a similar network that did not change significantly between groups. These hubs include signaling proteins (MAP3K1, GRB2, SHC and SRC), an estrogen receptor (ESR1) and DNA damage response proteins (BRCA1, RAD51, MRE11). Single-nucleotide polymorphisms in MAP3K1 are associated with breast cancer susceptibility[27]. Thus, there are changes in dynamic network modularity that are associated with poor outcome in breast cancer, and these may provide a prognostic signature in breast cancer.

To develop a prognostic signature that could be used to classify gene expression profiles from individual patients, we computed the relative expression of hubs with each of their interacting partners, determined for which hubs the relative expression differed significantly between patients who survived versus those who died from disease, and then employed affinity propagation clustering[28]. Affinity propagation is a clustering algorithm that takes similarity measures between data points and iteratively refines them until there are high quality exemplars. Clustering of test patients using affinity propagation allowed us to assign a probability of poor prognosis for each patient (**Supplementary Methods** and **Supplementary Fig. 9** online). We used a fivefold cross-validation strategy in which the hub selection process was incorporated on the training set within the cross-validation loop to avoid overfitting and assessed performance using receiver operator characteristic (ROC) curves. This revealed a typical area under the curve (AUC) of 0.711 (**Fig. 4a**) and accuracy, sensitivity and specificity of 76%, 86% and 81%, respectively. This compared favorably with the retrospective[29] or prospective[30] performance of commercially available genomic breast cancer diagnostics (53%, 41% and 68% in ref. 30 and 70%, 71% and 67% in ref. 29 for predicting 10-year survival[31]).

We also assessed performance using interactomes in which hubs were randomly removed. We observed that the performance of the classifier was reduced as hubs were removed (**Fig. 4d**), indicating that our accuracy may be limited by the interactome density. As current interactomes are likely incomplete and contain biases[32], further interactome mapping by systematic approaches may lead to improved prognostic performance.

To test the ability of the classifier to predict survival, we grouped patients using the poor outcome probabilities. The threshold for probability of prognosis was set to 0.4 as this consistently yielded the highest accuracy of prediction. Analysis of these two groups revealed significantly different 5-year survival (Mantel-Cox Log Rank test, nominal $P < 0.001$). Only 48% of patients possessing the poor-prognosis modularity signature survived for >5 years (**Fig. 4b**). Conversely, 85% of those with a good prognostic signature survived for 5 years. The average overall error rate of prognosis using the test-set data at this prognostic cutoff was 29.1%.

We next asked whether prognostic accuracy could be improved by incorporating clinical data (patient age, tumor stage and tumor grade). A logistic regression model that incorporated these variables along with network probabilities resulted in better performance (AUC = 0.784) (**Fig. 4a**) and enhanced prognostic classification (error rate, 25%) (**Fig. 4b**). Clinical covariates alone showed similar performance as the network probability score (AUC = 0.701, **Fig. 4a**). We also repeated these analyses using expression data from the TransBIG[30] cohort of breast cancer patients and observed similar, if not better, performance (AUC = 0.718–0.827; **Supplementary Fig. 9a** online) and Kaplan-Meier survival curves. Thus, >80% of predicted good-prognosis patients survived >10 years compared with <35% of those in the poor-prognosis group (**Supplementary Fig. 9b**). These results demonstrate that the molecular changes of the tumor that are captured by measuring changes in the network modularity of tumor interactomes are significant and independent predictors of patient disease outcome and suggest that measuring these changes may improve the predictive value of prognostic indicators already used in the clinic.

Previous approaches have employed network information to improve classification performance of gene signatures by extracting co-expressed pathways (that is, functional modules) and then using these pathways to assess cancer outcome[1]. In contrast, we have

searched for changes in the global modularity of the human interactome that indicate altered organization and information flow. Other mechanisms that affect network connectivity, such as alterations in protein stability and post-translational modification, may also influence network modularity on a global scale during cancer progression. Our studies motivate the search for multi-modal therapies that target hubs in networks that display altered modularity in disease. Furthermore, the favorable performance of our classification algorithms suggests that changes in network modularity may be a defining feature of tumor phenotype that, in turn, determines patient prognosis.

METHODS

Data integration to determine PCC of co-expression in interaction networks. We used a method analogous to that previously described[7]. The complete interactome from OPHID[2] as well as subsets of interactions mapped from yeast to man[33] or literature-curated interactions[4] were downloaded as well as expression data from 79 human tissues[3]. Hubs were defined to be nodes with more than five interactions, as these proteins are in the top 15% of the degree distribution of the network. For each hub the average PCC of co-expression for each interaction and the hub was assessed using a similar algorithm as previously described[7]. Random reassignment of the expression values to nodes in the network was used to ascertain if the observed network was nonrandom. The network was visualized using Cytoscape 2.5.1 (ref. 34).

Topological network analysis. Betweenness and CPL of networks were calculated using algorithms implemented by the tYNA web interface[13]. When assessing network robustness to hub removal, an equivalent number of intermodular and intramodular hubs were removed from the network in order of descending clustering coefficient.

To validate that the two hub classes are distinct, we investigated length, phosphorylation, linear motifs, globularity, domain architecture (**Supplementary Methods**). These were either computed directly from the hub sequence or by mapping to the appropriate database[35]. Significance levels were computed by sampling (**Supplementary Methods**).

Distribution of hub types by human disease phenotypes. For each hub, gene entries in OMIM[18] were extracted and manually curated for hubs (i) associated with cancer, malignancy or metastasis and (ii) found to be involved in oncogenic translocation fusions.

Network analysis between breast tumor samples. To assess differences in network organization between patients who were alive after extended follow-up versus those that died from disease, we used a nonparametric algorithm, within a cross-validation loop, to determine the difference in correlation of co-expression of hubs with their interactors. First, we calculated the PCC of hubs and their interactors for each patient group. We then calculated the absolute value of the difference of these PCCs. The magnitude is the difference in PCC of a hub between patient groups. To identify hubs that are significantly different between patient groups, we randomly assigned patients to one of two groups and repeated the analysis. This was done 1,000 times to calculate the random distribution. Real PCC differences for hubs between patient groups were compared to the random distribution to generate *P*-values. This defines a network signature of hubs whose modularity is different as a function of disease outcome. *P*-value cutoff and degree cutoff for hubs were optimized as a function of accuracy during cross-validation runs.

To measure prognostic accuracy of this network, we trained an affinity propagation algorithm[28] using the network signature to predict the patient outcome using fivefold cross-validation. Specifically, we partitioned the patient cohort into five approximately equally-sized portions, defined a network signature and trained our algorithm using four of these portions as described in detail in **Supplementary Methods**. To test the algorithm, we provided it with only the gene expression data for patients in this latter hold-out training set and compared its predictions of clinical outcome with the actual outcomes for these patients. We repeated this procedure for each hold-out set, amassing outcome predictions for every patient. To measure the variability in our predictions, we repeated the fivefold cross-validation procedure three times with different random partitions of the data.

Breast cancer patient prognostic predictive value is related to the total size of the protein interaction network. Interactions were randomly removed to obtain interactomes of reduced size, as indicated. The accuracy of prediction of outcome using dynamic network modularity at each indicated interactome size was then assessed by ROC curve analysis and is plotted as the average AUC (±s.d.) of three runs of fivefold cross-validation.

Kaplan-Meier survival curves were drawn for groups defined by the algorithm using patient survival data and drawn using SPSS for Mac, Rel.14.0.1.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We acknowledge Lars Juhl Jensen, EMBL-Heidelberg, for critical assessment of this manuscript. We would like to thank L. Attisano for critical review and helpful discussions. Work in J.L.W.'s lab was supported by funds from Genome Canada through the Ontario Genomics Institute and the Ontario chapter of the Canadian Breast Cancer Fund. D.W.-F. was supported by an Natural Sciences and Engineering Research Council of Canada operating grant assigned to Q.M. J.L.W. is a CRC chair and an International Scholar of the Howard Hughes Medical Institute. I.W.T. is supported by a fellowship from the Canadian Breast Cancer Foundation.

AUTHOR CONTRIBUTIONS

I.W.T. contributed to the development of the project, the execution of all experiments and the writing of the manuscript. R.L. contributed to experiments in **Figure 2** and writing of the manuscript. D.W.-F. and Q.M. contributed to the development and implementation of the prognosis classification algorithm and the cross-validation strategies, as well as to the writing the manuscript. Y.L. contributed programming support for data in **Figures 1–4**. C.P. & D.F. contributed to the semantic similarity experiment in Supplementary **Figure 2b**. S.B. contributed to the development of the statistical frameworks throughout the manuscript and writing the manuscript. T.P. contributed to writing of the manuscript. J.L.W. contributed to the development of the project and the writing of the manuscript.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D. & Ideker, T. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* **3**, 140 (2007).
2. Brown, K.R. & Jurisica, I. Online predicted human interaction database. *Bioinformatics* **21**, 2076–2082 (2005).
3. Su, A.I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **101**, 6062–6067 (2004).
4. Chatr-aryamontri, A. *et al.* MINT: the Molecular INTeraction database. *Nucleic Acids Res.* **35**, D572–D574 (2007).
5. von Mering, C. *et al.* STRING 7–recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* **35**, D358–D362 (2007).
6. Fraser, H.B. Modularity and evolutionary constraint on proteins. *Nat. Genet.* **37**, 351–352 (2005).
7. Han, J.D. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004).
8. Barabasi, A.L. & Oltvai, Z.N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
9. de Lichtenberg, U., Jensen, L.J., Brunak, S. & Bork, P. Dynamic complex formation during the yeast cell cycle. *Science* **307**, 724–727 (2005).
10. Tengowski, M.W., Feng, D., Sutovsky, M. & Sutovsky, P. Differential expression of genes encoding constitutive and inducible 20S proteasomal core subunits in the testis and epididymis of theophylline- or 1,3-dinitrobenzene-exposed rats. *Biol. Reprod.* **76**, 149–163 (2007).
11. Thomas, M.K., Yao, K.M., Tenser, M.S., Wong, G.G. & Habener, J.F. Bridge-1, a novel PDZ-domain coactivator of E2A-mediated regulation of insulin gene transcription. *Mol. Cell. Biol.* **19**, 8492–8504 (1999).
12. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
13. Yip, K.Y., Yu, H., Kim, P.M., Schultz, M. & Gerstein, M. The tYNA platform for comparative interactomics: a web tool for managing, comparing and mining multiple networks. *Bioinformatics* **22**, 2968–2970 (2006).
14. Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X. & Gerstein, M. Genomic analysis of essentiality within protein networks. *Trends Genet.* **20**, 227–231 (2004).
15. Puntervoll, P. *et al.* ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* **31**, 3625–3630 (2003).

16. Letunic, I. *et al.* SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res.* **34**, D257–D260 (2006).
17. Karnoub, A.E. & Weinberg, R.A. Ras oncogenes: split personalities. *Nat. Rev. Mol. Cell Biol.* **9**, 517–531 (2008).
18. McKusick, V.A. Mendelian inheritance in man and its online version, OMIM. *Am. J. Hum. Genet.* **80**, 588–604 (2007).
19. Futreal, P.A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
20. van de Vijver, M.J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
21. Roukos, D.H. Prognosis of breast cancer in carriers of BRCA1 and BRCA2 mutations. *N. Engl. J. Med.* **357**, 1555–1556, author reply 1556.
22. Soderlund, K. *et al.* Intact Mre11/Rad50/Nbs1 complex predicts good response to radiotherapy in early breast cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **68**, 50–58 (2007).
23. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98**, 5116–5121 (2001).
24. Chang, H.Y. *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2**, E7 (2004).
25. Liu, R. *et al.* The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N. Engl. J. Med.* **356**, 217–226 (2007).
26. Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. USA* **100**, 8418–8423 (2003).
27. Easton, D.F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
28. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007).
29. Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
30. Buyse, M. *et al.* Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J. Natl. Cancer Inst.* **98**, 1183–1192 (2006).
31. Haibe-Kains, B. *et al.* Comparison of prognostic gene expression signatures for breast cancer. *BMC Genomics* **9**, 394 (2008).
32. Bertin, N. *et al.* Confirmation of organized modularity in the yeast interactome. *PLoS Biol.* **e153** (2007).
33. von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399–403 (2002).
34. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
35. Linding, R. *et al.* Systematic discovery of *in vivo* phosphorylation networks. *Cell* **129**, 1415–1426 (2007).