Modeling community-wide molecular networks of multicellular systems

Kakajan Komurov*

Divisions of Experimental Hematology and Cancer Biology, Human Genetics and Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Multicellular systems, such as tissues, are composed of different cell types that form a heterogeneous community. Behavior of these systems is determined by complex regulatory networks within (intracellular networks) and between (intercellular networks) cells. Increasingly more studies are applying genome-wide experimental approaches to delineate the contributions of individual cell types (e.g. stromal, epithelial, vascular cells) to collective behavior of heterogeneous cell communities (e.g. tumors). Although many computational methods have been developed for analyses of intracellular networks based on genome-scale data, these efforts have not been extended toward analyzing genomic data from heterogeneous cell communities.

Results: Here, we propose a network-based approach for analyses of genome-scale data from multiple cell types to extract community-wide molecular networks comprised of intra- and intercellular interactions. Intercellular interactions in this model can be physical interactions between proteins or indirect interactions mediated by secreted metabolites of neighboring cells. Applying this method on data from a recent study on xenograft mouse models of human lung adenocarcinoma, we uncover an extensive network of intra- and intercellular interactions involved in the acquired resistance to angiogenesis inhibitors.

Contact: kakajan.komurov@cchmc.org

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Individual cells within a heterogeneous cell population interact with each other through secreted molecules and membrane proteins, sometimes referred to as cross-talk (Frankenstein *et al.*, 2006; Jahoda and Christiano, 2011; Kiel and Morrison, 2008). At the molecular level, this population can be viewed as a community-wide network of molecular interactions comprising intracellular interactions within each cell as well as intercellular interactions of molecules of different cells. Since population characteristics as a whole are highly dependent on the intra- as well as intercellular networks, the global architecture of the community-wide molecular network (CMN, made of intra- and intercellular

molecular interactions) can determine the collective behavior of heterogeneous cell communities.

Network-based analyses of genomic data have shed light on the global organization of intracellular networks contributing to normal and malignant behavior of cells (Basso *et al.*, 2005; Calvano *et al.*, 2005; Chuang *et al.*, 2007; Tomlins *et al.*, 2007). Mounting evidence now suggests that interplays of cells within a microenvironment can give rise to complex population behavior (Frankenstein *et al.*, 2006; Jahoda and Christiano, 2011). Such complex interactions of cells within a population have been observed in developmental processes (Kirouac *et al.*, 2010; Lai, 2004), in stem cell niches (Jahoda and Christiano, 2011; Kiel and Morrison, 2008) and in tumor microenvironments (Boersma *et al.*, 2008; Coussens and Werb, 2002). However, most of the studies on deciphering the complex pattern of molecular network interactions in such multicellular systems and their role in population-wide collective behavior have been focused on a limited number of molecules. Although some notable large-scale studies in some experimental systems have been undertaken (Frankenstein *et al.*, 2006; Kirouac *et al.*, 2010), the computational methodology of analysis primarily involved candidate-based approaches, limiting the scope of analysis. More powerful computational methods for analyses of genomic data from heterogeneous cell populations would therefore greatly enhance our ability to gain insight into the organization of CMNs and their role in the collective population behavior.

Here, in order to enable modeling of molecular networks of whole-cell populations, we developed a network model of community-wide molecular interactions by combining intracellular interactions from each cell type and their intercellular connections into a single global network (community molecular network) (Fig. 1). We use this global network in conjunction with the genome-wide gene expression data from different cell types to extract networks of interest showing community-wide molecular interactions most highlighted by the data. We integrate genomic data with the global network using NetWalk (Komurov *et al.*, 2010), a computational algorithm for seamless integration of genomic data with molecular networks. The advantage of NetWalk compared to other network analysis tools is that NetWalk takes into account the whole-data distribution without requiring statistical cutoffs or predetermined gene lists of interest. NetWalk output is a distribution of Edge Flux (EF) values containing numeric score of relevance assigned to each interaction in the network. EF values, just like in gene expression data, can be used for further statistical analyses, allowing for direct network-based statistical analyses. Using this platform, we present an analysis of recently published gene expression data from epithelial and
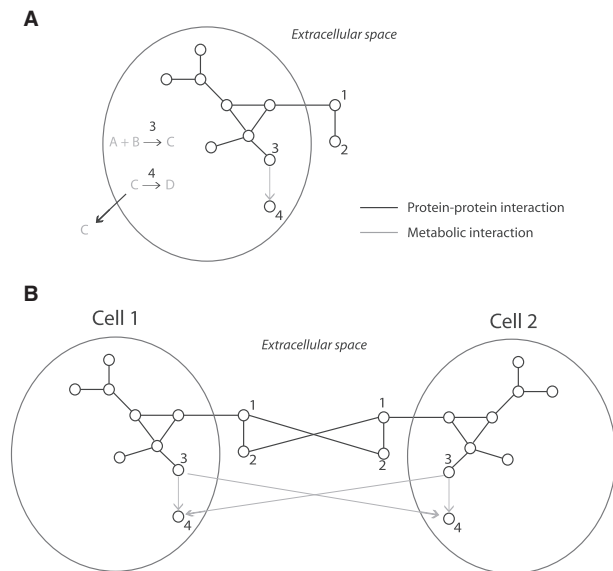
---

*To whom correspondence should be addressed.

Fig. 1. CMN model. (**A**) An imaginary network of intracellular interactions. The circle represents plasma membrane. Four nodes have been numbered to demonstrate intercellular interaction types. Nodes 1 and 2 represent extracellular proteins, and 3 and 4 are enzymes performing consecutive metabolic reactions (product of the reaction catalyzed by 3, C, is a reactant of that catalyzed by 4). (**B**) CMN of two different cell types. Since proteins 1 and 2 are extracellular, protein 1 in one cell type can interact with protein 2 in the other, and vice versa. In addition, since the metabolite C is secreted into the extracellular space, the reaction performed by enzyme 3 in one cell can affect the reaction performed by enzyme 4 in the other.

stromal cells of a mouse xenograft model of acquired resistance to bevacizumab (angiogenesis inhibitor) (Cascone *et al.*, 2011), and identify previously unrecognized CMNs of stromal and epithelial cells involved in acquired resistance to this targeted therapy.

2 METHODS

2.1 Global network of molecular relationships

First, we compiled a comprehensive network of biomolecular interactions between human genes. In order to account for physical and indirect functional interactions, we compiled (i) protein–protein interactions; (ii) transcription factor–target pairs; (iii) interactions based on neighboring metabolic reactions; and (iv) neighboring functional interactions. Protein–protein interactions were obtained from HPRD [Human Protein Reference Database, Mishra *et al.* (2006)], BIND [Biomolecular Interaction Database, Bader *et al.* (2001)], MINT (Chatr-aryamontri *et al.*, 2007), BioGRID (Stark *et al.*, 2006) and IntAct (Kerrien *et al.*, 2007). Directed signaling interactions were obtained from Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) and National Cancer Institute (NCI) Pathway Interaction Database (http://pid.nci.nih.gov/). Interactions from MINT, BioGRID, IntAct and NCI were obtained from Pathway Commons (Cerami *et al.*, 2011). Transcription factor–target interactions were obtained from BIND (queried as protein–DNA interactions), Reactome (Joshi-Tope *et al.*, 2005) (obtained from Pathway Commons), TRANSFAC (Wingender *et al.*, 2000) and NCI Pathway Interaction Database (obtained from Pathway Commons). Neighboring metabolic reactions are assigned to a pair of genes if the product of the reaction catalyzed by one gene product is the reactant of the reaction catalyzed by the other. As such, metabolic interactions are directed (A → B, product of reaction catalyzed by A is a reactant for B).

For example, hexokinase II (HK2) catalyzes the reaction glucose + ATP → glucose-6-phosphate + ADP, and glucose phosphate isomerase (GPI) catalyzes the reaction glucose-6-phosphate → fructose-6-phosphate. Since glucose-6-phosphate is the product of the former and the reactant of the latter, these two genes are assigned an interaction HK2 → GPI in the network. See Supplementary Figure S1 for the network of glycolytic genes. Information on genes and their metabolic reactions was obtained from KEGG. Neighboring functional interactions were obtained from Reactome. Altogether, the global network consists of 13 256 unique genes connected by 170 735 interactions.

2.2 Community molecular network model

The global network of biomolecular relationships described above accounts for intracellular interactions between genes. This network is defined as a_{ij}, which is 1 if genes i and j interact, 0 otherwise. In case of multiple cell types within the same community, the intracellular networks for each cell are defined separately, a_{kij}, where k is the cell type k. We define the CMN (of two cell types, for simplicity) as

$$A = \begin{bmatrix} a_1 & r_{12} \\ r_{21} & a_2 \end{bmatrix} \tag{1}$$

where the entries in the diagonal (e.g. a_1) are intracellular adjacency matrices for each cell type, and r_{12} is the intercellular adjacency matrix between molecules of cell types 1 and 2. The matrix r_{12ij} is 1, only if the gene i in cell type 1 interacts with gene j in cell type 2, and 0 otherwise. Gene i in cell type 1 interacts with gene j in cell type 2 if one of the following is true:

1. The interaction ij is a protein–protein interaction or Reactome neighboring interaction and gene i encodes an extracellular secreted protein and gene j encodes an extracellular secreted protein or a membrane protein.

2. The interaction ij is a neighboring metabolic reaction and the common metabolites in the respective reactions catalyzed by gene i and j is a secreted metabolite.

These two rules ensure that only those gene pairs capable of functionally interacting at the intercellular level are assigned an interaction. The first rule is for physical interactions that are possible only if the two proteins can be in a physical proximity of each other (type 1 intercellular interactions). The second rule is for indirect interactions between metabolic enzymes, where the reaction performed by one can affect the reaction performed by the other, which can only happen if the common metabolite of the two reactions is secreted [i.e. subcellular location of the metabolite is defined in the Human Metabolome Database (Wishart *et al.*, 2009) as 'extracellular'] (type 2). Although analyses of cell–cell interactions within a cell population have primarily focused on physical interactions defined by type 1 interactions above, metabolic interactions between enzymes of different cells can also be of high importance in defining population behavior. One such mechanism exists in normal brain (Dienel and Cruz, 2003) and in some tumors (Pavlides *et al.*, 2009) for shuttling of lactate from lactate-producing cells to lactate-oxidizing cells, thereby forming a metabolic symbiosis within the multicellular system. A recent study of large-scale modeling of metabolic interactions between cell types in the human brain is also an excellent example underlining the importance of this type of interactions (Lewis *et al.*, 2010).

Overall, the network defined by A is a single molecular network of the whole-cell community (CMN) that contains information on intra- as well as intercellular network interactions (Fig. 1). As expected, most of the CMN is composed of intracellular interactions, with intercellular interactions making up only slightly >6% of all interactions in the CMN. Approximately 20% of intercellular interactions in our model are composed of type 2 interactions defined above.

2.3 Data integration and NetWalk algorithm

Gene expression data used in the analyses were obtained from Cascone *et al.* (2011) (NCBI GEO: GSE26644). Briefly, *t*-tests were performed

on each gene to score significance of expression change between tumor or stromal cells before and after bevacizumab resistance. This generated a *t*-value of significance for each gene in tumor and stromal cells, respectively. Before analyses, mouse genes in stromal dataset were converted to corresponding human orthologous genes by using the mapping provided at the Mouse Genome Informatics Database (http://www.informatics.jax.org/orthology.shtml).

We utilize a data integration approach that takes into account the whole distribution of the data without relying on manual statistical cutoffs. Such seamless integration of the data with the network is done using the data-biased random walk algorithm, NetWalk. Briefly, NetWalk integrates genomic data with networks of interactions between genes to score the relevance of each interaction based on both the data values of the genes as well as their local network connectivity. This results in a distribution of EF values that can be used for dynamic reconstruction of user defined networks. EF values can be further subjected to statistical analyses such as clustering, allowing for direct numerical comparisons of context-specific networks between different conditions.

First, the experimental data are integrated with the network to form a transition probability matrix for random walk

$$p_{ij} = a_{ij} \frac{w_j}{\sum_{k \in N_i} w_k} \qquad (2)$$

where p_{ij} is the transition probability from node i to node j, w_j is the experimental value for node j, and N_i is the set of immediate downstream neighbors (undirected edges are considered bidirectional) of node i. The final relevance score (g) for each node is calculated by

$$g = g\left((1-q)P\right) + \frac{1}{\sum_{k \in n} w_k} q \times w^T \qquad (3)$$

where q is the restart probability (we use $q = 0.01$). EF values are calculated by

$$e_{ij} = g_i P_{ij} \qquad (4)$$

where e_{ij} is the flux through edge ij and represents the score of importance of the interaction based on the data. In order to control for topological bias in the network, we also calculate

$$e_{r\_ij} = g_{r\_i} P_{r\_ij} \qquad (5)$$

where e_r is the edge score distribution vector calculated by letting $w_i = 1$ for all i. Finally, the NetWalk output is the distribution of normalized EF values

$$EF_{ij} = \log\left(\frac{e_{ij}}{e_{r\_ij}}\right) \qquad (6)$$

This gives the final normalized score distribution of edges, which reflects EFs of nodes relative to what would be expected by topology alone in the given network. The EF distribution contains relevance scores assigned to interactions within each cell type as well as to interactions between cells. The EF distribution can be queried for networks most highlighted by the data within the whole-cell community (intracellular and intercellular interactions combined), within individual cell types or only for intercellular interactions.

NetWalk is available in a software suite NetWalker (Komurov *et al.*, submitted for publication, http://research.cchmc.org/netwalker). Implementation and inclusion of the CMN model into NetWalker are currently in progress. However, an R file-containing interaction list objects for our comprehensive network and a pre-constructed CMN for two cell types can be found in the Supplementary Material.

3 RESULTS

As a proof of concept analysis using our CMN model, we chose to use the data from a recent study utilizing mouse xenograft models to study acquired resistance of human lung adenocarcinomas to

angiogenesis inhibitors (Cascone *et al.*, 2011). The authors in the aforementioned study injected H1975 cells (human non-small cell lung cancer cell line that is sensitive to bevacizumab) into mice and after a while started treating them with a vehicle control or bevacizumab. Despite initial response, majority of tumors developed resistance to the drug. Gene expression microarray analyses were then performed on control and bevacizumab-resistant tumors ($n = 3$ for each) using Illumina mouse-specific (WG-6 v2) and human-specific (WG-6 v3) arrays. Mouse-specific arrays were used to measure stromal gene expression and the human array was used to measure the tumor gene expression.

In order to identify community-wide molecular networks involved in the acquired resistance to bevacizumab, we first performed *t*-tests on each gene in the stromal and tumor datasets to score the extent of their change in resistant versus control cases. The resultant *t*-values (the value of *t*-statistic from *t*-tests) for tumor and stromal gene expression data had different distributions; stromal cells had a wider *t*-value distribution (not shown), consistent with the original study's report that stromal gene expression changes were more pronounced than that in tumor data. Therefore, we transformed the two *t*-value distributions using quantile normalization to obtain identical distributions for tumor and stromal *t*-values. These *t*-values (two sets, one for stromal and one for tumor cells) were used as node weights for our NetWalk-based network analyses (Fig. 2).
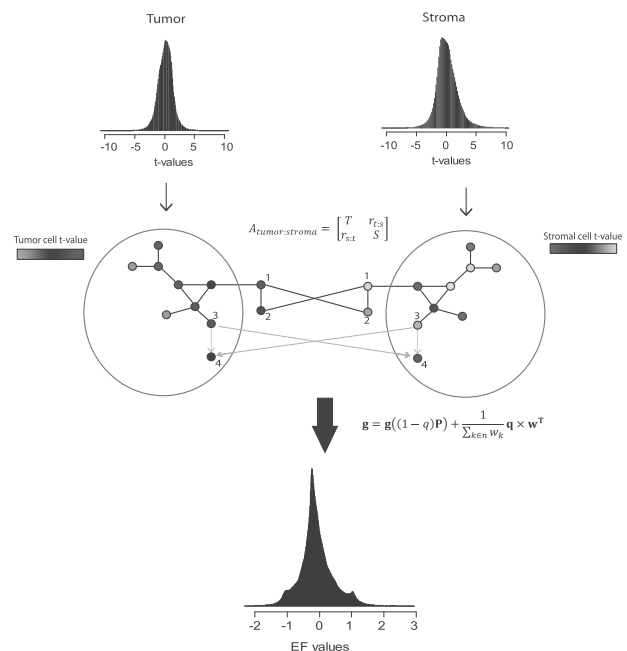

Fig. 2. Workflow of the method in this study. First, *t*-tests were performed for each gene to score the difference of their expression between control and bevacizumab-resistant samples in tumor and stroma datasets, respectively. CMN was created for two cell types, and the resultant *t*-values were overlaid onto this CMN. NetWalk analysis was performed to obtain an EF distribution, scoring each interaction in this CMN based on the tumor and stromal *t*-values. Note that the EF distribution contains scores for intracellular interactions in each cell type as well as intercellular interactions.
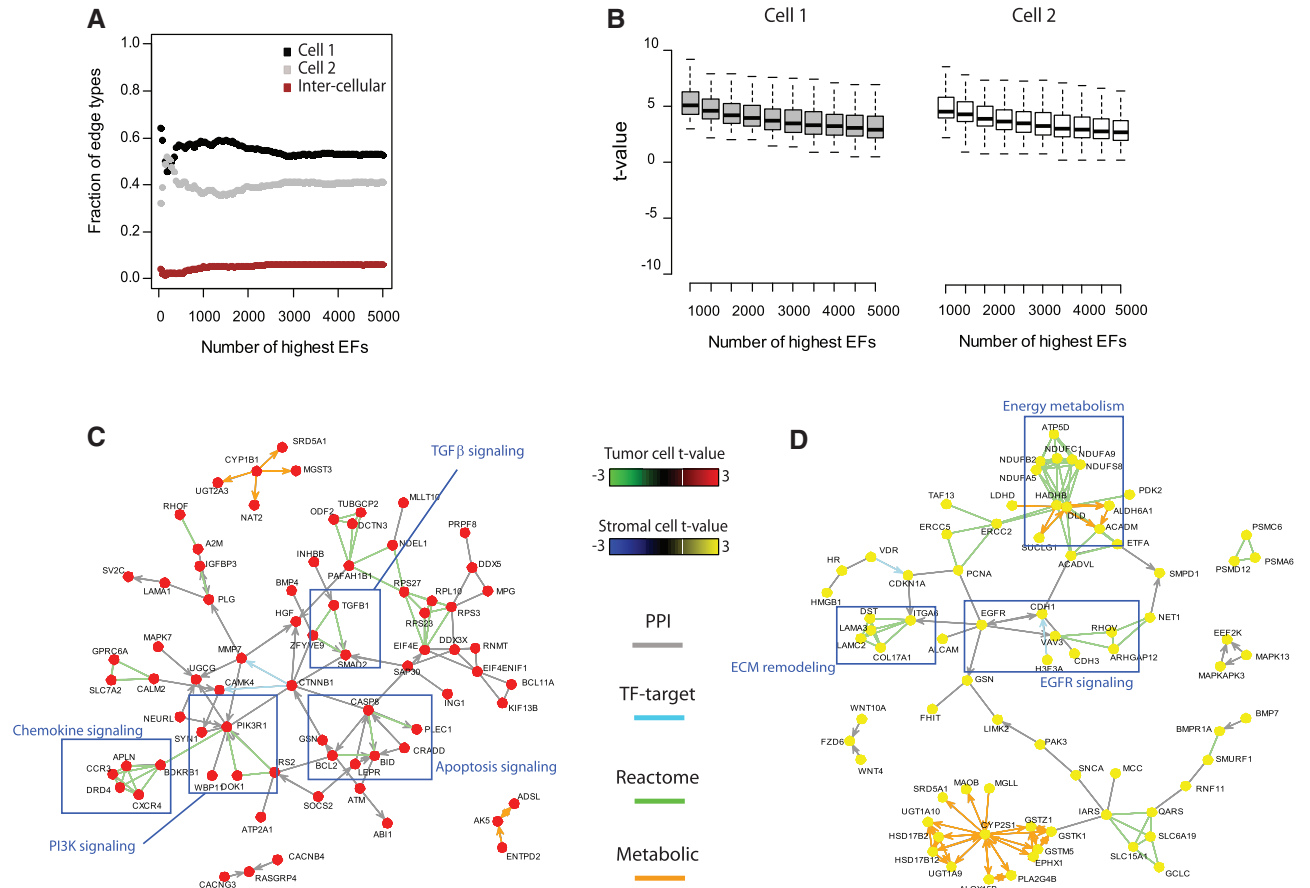
Fig. 3. NetWalk analysis of CMN. (**A**) Fraction of intracellular interactions in cells 1 and 2 (i.e. tumor and stroma) and intercellular interactions within the indicated number of highest scoring EF values from the CMN of two cell types. (**B**) Boxplot of *t*-value distributions of genes for cells 1 and 2 within the networks generated from indicated number of highest scoring EF values. Network of highest-scoring EF values for the (**C**) tumor and (**D**) stromal *t*-values. Nodes are colored by their *t*-values according to the color key. Interactions are colored by their type as indicated. Some notable subnetworks have been highlighted by boxes.

3.1 NetWalk analysis of CMN

First, we constructed a CMN model of two cell types: tumor (T) and stroma (S). This CMN is defined as

$$A_{\text{tumor:stroma}} = \begin{bmatrix} T & r_{t:s} \\ r_{s:t} & S \end{bmatrix} \qquad (7)$$

where $A_{tumor:stroma}$ is the adjacency matrix of this CMN, T is the adjacency matrix for tumor and S is the adjacency matrix for stromal intracellular networks (note that T and S are identical, we denoted them separately for clarity), $r_{t:s}$ is the adjacency matrix for intercellular interactions between tumor and stromal cells defined by rules in 'Methods' section. This model is composed of 26 512 genes (13 256 genes/cell type) connected by 375 570 interactions (all intra- and intercellular interactions). By overlaying both *t*-value distributions onto the respective nodes on this CMN, we performed NetWalk analysis to obtain numeric scores of relevance for each interaction in this network (i.e. EF values). First, in order to understand the performance of our approach with respect to identifying intra- and intercellular networks, we tested the composition of increasing numbers of EF values for intra-

and intercellular interactions. Figure 3A shows the compositions of draws of 50–5000 highest EF values (Fig. 3A). Relative compositions of the individual intracellular interactions are similar, with intercellular interactions comprising <10% of all interactions, as expected. Therefore, NetWalk does not bias the scoring of interactions to one or the other cell type.

We have shown that networks extracted from EF values are more faithful to the underlying data than other network extraction algorithms (i.e. high scoring networks contain genes with high data values and vice versa) (Komurov *et al.*, 2010). We wanted to test if the EF values from NetWalk analysis of the CMN are also faithful to the underlying data distribution from the two cell types. Figure 3B shows the *t*-value distributions of genes from each cell type within the networks of 50–5000 highest EF values. The boxplots show that even in the network of 5000 highest scoring interactions from this NetWalk analysis, the distribution of *t*-values of genes in both cell types is strictly at the higher end of the individual *t*-value distributions. We have also conducted additional randomization simulations to verify the strict coherence of NetWalk output within the CMN context to the input data (see Supplementary Fig. S2 and legend). This indicates that the CMN subnetworks

extracted from NetWalk analysis are highly related to the input data, so that high scoring networks are composed of genes with high *t*-values, rather than unrelated genes. This feature of NetWalk analysis increases confidence in the relevance of the extracted networks to the underlying data.

3.2 Analysis of individual cell networks

First, we analyzed the individual intracellular networks by visualizing networks corresponding to highest EF values from each cell type. An individual analysis of the network from highest EF values in tumor cells shows an upregulation of networks involved in survival signaling centered around phosphoinositide 3 kinase (PI3K), chemokine signaling, TGF-*β* and apoptosis signaling, among others (Fig. 3C). PI3K in this network seems to be activated by multiple mechanisms, including Toll-like receptor (TLR2), insulin receptor substrate (IRS2) and G-protein signaling. Stromal cells, however, are characterized by upregulation of different networks, primarily in the extracellular matrix (ECM) remodeling, epidermal growth factor receptor (EGFR) signaling, oxidative energy metabolism and others (Fig. 3D). Upregulation of EGFR in stromal cells was also reported in the original study, suggesting that our approach, in addition to revealing novel associations, can reproduce previous findings. Although revealing potentially important information on the respective intracellular network interactions involved in acquired drug resistance, individual network analyses do not allow for analyses of interactions between these two networks.

3.3 Analysis of the community networks of tumor and stromal cells

In order to identify community-wide molecular networks involved in drug resistance in this experimental model, we visualized the network corresponding to highest scoring 400 EF values (Fig. 4A). As would be expected from the total percentage of intercellular interactions in the CMN (∼6%), most of the network is composed of intracellular interactions of the respective cell types. There are however a few potentially important interactions occurring between stromal and tumor cell gene products (highlighted in Fig. 4A). This subnetwork contains both physical and metabolic intercellular interactions.

3.3.1 Physical interactions Some notable examples for physical interactions include HMGB1 secreted by stromal cells acting on the TLR2 and PLG (plasminogen) in tumor cells to activate them (Fig. 4B), NID1–LAMC2 (nidogen and laminin C2) involved in the ECM remodeling, fibroblast growth factor signaling (tumor FGF8–stromal FGFR2), tumor cell bone morphogenetic protein (BMP) proteins acting on the BMP receptor on stromal cells, chemokine signaling between stromal and tumor cells, and neuropeptide signaling (neuropeptide (NPY) from stromal cells activating NPY5R, the receptor, on tumor cells). Upregulation of FGFR2 in stromal cells was reported in the original study of Cascone *et al.*, again showing that our analysis can capture findings from traditional approaches. However in addition, this analysis shows multiple potentially important intercellular physical interactions associated with drug resistance. Importantly, most of these intercellular physical interactions have been described in the literature to play roles in different aspects of cancer progression.

3.3.2 Metabolic interactions In addition to physical interactions, the subnetwork in Figure 4A contains several metabolic intercellular interactions. A particularly intriguing set of interactions is shown separately in Figure 4C, which contains several enzymes in tumor and stromal cells involved in energy metabolism. A sketch of the reactions performed by these enzymes in the stromal and tumor cells is shown in Figure 4D. Stromal cells have a significant upregulation (based on *t*-value) of the enzymes lactate dehydrogenase B (LDHB), dihydrolipoamide dehydrogenase (DLD), glutamic pyruvate transaminase (GPT2), pyruvate carboxylase (PC) and asparagine synthetase (ASNS), while tumor cells have upregulated lactate dehydrogenase C (LDHC) and asparaginase like 1 (ASRGL1). While other lactate dehydrogenase enzymes primarily catalyze conversion of pyruvate to lactate as the terminal reaction of glycolysis, LDHB mainly converts lactate back to pyruvate. DLD is a component of several enzyme complexes, including the pyruvate dehydrogenase complex, which catalyzes conversion of pyruvate to acetyl CoA, which enters the tricarboxylic acid cycle (TCA). Similarly, PC catalyzes conversion of pyruvate to oxaloacetate, which also enters the TCA cycle. Enzyme GPT2 in stromal cells catalyzes the reversible reaction pyruvate + glutamate ↔ alanine + 2-oxoglutarate, which is one of the important reactions in gluconeogenesis. ASNS and ASRGL1 catalyze, respectively, asparagine synthesis producing glutamate from glutamine, and breakdown of asparagine to aspartate producing NH_3.

It is obvious from this picture, that stromal cells have a significant upregulation of the enzymes involved in oxidative energy generation (LDHB, DLD, PC) and gluconeogenesis (GPT2), which is also evident from the significant upregulation of the mitochondrial electron transport chain components (Fig. 4A). Tumor cells, however, seem to mainly perform glycolysis, as evidenced from preferential expression of LDHC and glycolytic enzymes hexokinase I (HK1), GPI and the glucose transporter SLC2A10 (see the highlighted subnetwork 9 in Fig. 4A). Three metabolites, lactate, asparagine and aspartate mediate the metabolic interactions between the tumor and stromal cells in this scenario. While the functional significance of the interaction of ASNS in stromal cells and ASRGL1 in tumor cells potentially involving an aspartate/asparagine shuttle may not be immediately clear, the interaction of LDHC with LDHB and other stromal enzymes involved in oxidative energy metabolism is particularly intriguing. Based on the above observations, a hypothesis emerges where tumor cells primarily consume glucose by glycolysis and secreting lactate, which is then uptaken by stromal cells and oxidized through mitochondrial TCA cycle and/or used for gluconeogenesis to produce glucose or other building blocks (e.g. by GPT2) that may be used again by tumor cells. This hypothesis is not far-fetched, as a similar metabolic symbiosis involving lactate shuttling between stromal and tumor cells has been reported and has been implicated with increased tumorigenicity in several cancer models (Koukourakis *et al.*, 2006; Pavlides *et al.*, 2009; Sonveaux *et al.*, 2008). Therefore, the metabolic interactions identified in this study may underlie *bona fide* cross-talk mechanisms between tumor and stromal cells that play roles in meeting the metabolic demand of the tumorigenic and drug resistance phenotypes.

4 DISCUSSION

In this study, we proposed a modeling approach of multicellular systems as community-wide molecular networks to increase our
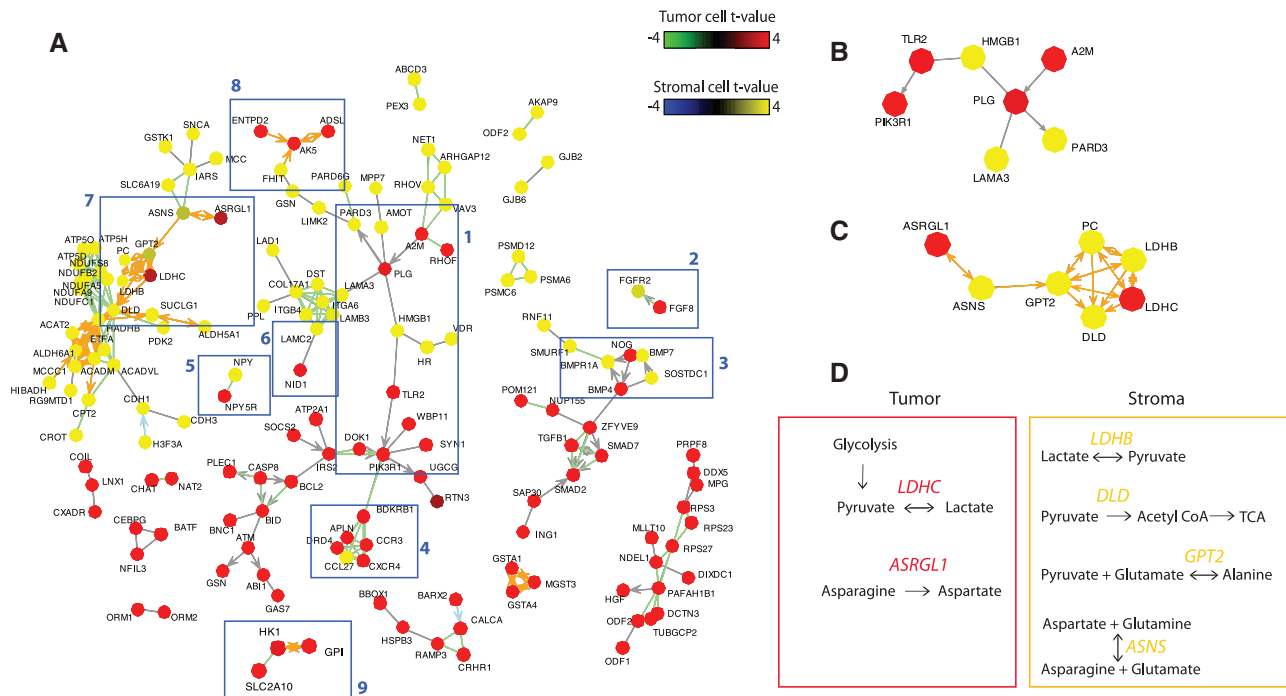
Fig. 4. CMN analysis of tumor and stromal *t*-values. (**A**) Highest scoring interactions from the CMN analysis of tumor and stromal *t*-values. Nodes are colored by the respective *t*-values of tumor and stromal cells as shown in the color key. Interactions are colored as in Figure 3. Some notable subnetworks of intercellular interactions are highlighted in boxes and numbered. (**B**) A blow-up of the network 1 in (A). (**C**) A blow-up of the network 7 in (A). (**D**) Metabolic reactions performed by enzymes in C in the tumor and stromal cells.

understanding of the complex interplay between intracellular networks of different cell types within the community. Using our previously developed data-biased random walk approach, NetWalk, together with CMN, we were able to obtain a view of the complex interplay between intracellular networks of tumor and stromal cells in acquired drug resistance. In addition to identifying the previously reported findings from the original study of Cascone *et al.*, our approach uncovered several novel intercellular interactions involving physical as well as indirect metabolic cross-talk with potential roles in the drug resistance phenotype. Although experimental validation of these findings is beyond the scope of the current study, which is to demonstrate the use of CMN in conjunction with NetWalk, the analyses presented here show that the results faithfully reflect the data distribution that was used as input for NetWalk (Fig. 3B and Supplementary Fig. S2).

Although the analyses demonstrated here used NetWalk for network integration, other methods, such as those based on data cutoffs could be used. Although insightful community networks can be constructed by using gene lists from such data cutoffs, most of the genes within these lists will not be used for network construction and therefore will not be accounted for during analyses (Supplementary Fig. S3). However, the interactions with highest EF values generated by NetWalk contain most of the genes with highest data values, and therefore information loss is minimized in NetWalk-based network analyses (Supplementary Fig. S3 and legend).

The community network model that we present here takes into account both physical and indirect interactions mediated by secreted metabolites between different cells within the community.

Although physical interactions through ECM components have been well-characterized in several systems, intercellular interactions at the metabolite level have not been studied to a similar extent (Lewis *et al.*, 2010), leaving much room for exploration. Our model enables analyses of large-scale functional genomics data from multicellular systems for the identification of complex scenarios involving functional interplay at the physical and metabolic levels between different cells of the multicellular system. Overall, our community-wide molecular network model is an invaluable tool in genome-wide studies of multicellular systems.

ACKNOWLEDGEMENTS

REFERENCES

Bader,G.D. *et al.* (2001) BIND–The Biomolecular Interaction Network Database. *Nucleic Acids Res.*, **29**, 242–245.

Basso,K. *et al.* (2005) Reverse engineering of regulatory networks in human B cells. *Nat. Genet.*, **37**, 382–390.

Boersma,B.J. *et al.* (2008) A stromal gene signature associated with inflammatory breast cancer. *Int. J. Cancer*, **122**, 1324–1332.

Calvano,S.E. *et al.* (2005) A network-based analysis of systemic inflammation in humans. *Nature*, **437**, 1032–1037.

Cascone,T. *et al.* (2011) Upregulated stromal EGFR and vascular remodeling in mouse xenograft models of angiogenesis inhibitor-resistant human lung adenocarcinoma. *J. Clin. Invest.*, **121**, 1313–1328.

Cerami,E.G. *et al.* (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.

Chatr-aryamontri,A. *et al.* (2007) MINT: the Molecular INTeraction database. *Nucleic Acids Res.*, **35**, D572–D574.

Chuang,H.Y. *et al.* (2007) Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.*, **3**, 140.

Coussens,L.M. and Werb,Z. (2002) Inflammation and cancer. *Nature*, **420**, 860–867.

Dienel,G.A. and Cruz,N.F. (2003) Neighborly interactions of metabolically-activated astrocytes in vivo. *Neurochem. Int.*, **43**, 339–354.

Frankenstein,Z. *et al.* (2006) The immune-body cytokine network defines a social architecture of cell interactions. *Biol. Direct.*, **1**, 32.

Jahoda,C.A. and Christiano,A.M. (2011) Niche crosstalk: intercellular signals at the hair follicle. *Cell*, **146**, 678–681.

Joshi-Tope,G. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.

Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.

Kerrien,S. *et al.* (2007) IntAct–open source resource for molecular interaction data. *Nucleic Acids Res.*, **35**, D561–D565.

Kiel,M.J. and Morrison,S.J. (2008) Uncertainty in the niches that maintain haematopoietic stem cells. *Nat. Rev. Immunol.*, **8**, 290–301.

Kirouac,D.C. *et al.* (2010) Dynamic interaction networks in a hierarchically organized tissue. *Mol. Syst. Biol.*, **6**, 417.

Komurov,K. *et al.* (2010) use of data-biased random walks on graphs for the retrieval of context-specific networks from genomic data. *PLoS Comput. Biol.*, **6**, e1000889.

Koukourakis,M.I. *et al.* (2006) Comparison of metabolic pathways between cancer cells and stromal cells in colorectal carcinomas: a metabolic survival role for tumor-associated stroma. *Cancer Res.*, **66**, 632–637.

Lai,E.C. (2004) Notch signaling: control of cell communication and cell fate. *Development*, **131**, 965–973.

Lewis,N.E. *et al.* (2010) Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat. Biotechnol.*, **28**, 1279–1285.

Mishra,G.R. *et al.* (2006) Human protein reference database–2006 update. *Nucleic Acids Res.*, **34**, D411–D414.

Pavlides,S. *et al.* (2009) The reverse Warburg effect: aerobic glycolysis in cancer associated fibroblasts and the tumor stroma. *Cell Cycle*, **8**, 3984–4001.

Sonveaux,P. *et al.* (2008) Targeting lactate-fueled respiration selectively kills hypoxic tumor cells in mice. *J. Clin. Invest.*, **118**, 3930–3942.

Stark,C. *et al.* (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.*, **34**, D535–D539.

Tomlins,S.A. *et al.* (2007) Integrative molecular concept modeling of prostate cancer progression *Nat. Genet.*, **39**, 41–51.

Wingender,E. *et al.* (2000) TRANSFAC: an integrated system for gene expression regulation *Nucleic Acids Res.*, **28**, 316–319.

Wishart,D.S. *et al.* (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.*, **37**, D603–D610.