Zhao, et al. Supplementary Materials

Supplementary Materials and Methods

1. Cancer signaling bridges (CSBs)

1.1 Definition

CSBs are specific instances of network motifs in the protein-protein interaction (PPI) network; for example, triangles and squares, which play important roles in cancer study and drug discovery (1, 50). Each CSB can be denoted by a protein set with a PPI set between them, such as a square containing the protein set of P = (P_1, P_2, P_3, and P_4) and a PPI set for the four proteins, i.e. $P_1 \leftrightarrow P_2$, $P_1 \leftrightarrow P_3$, $P_4 \leftrightarrow P_2$, and $P_4 \leftrightarrow P_3$. To ensure that a CSB can be used to extend a known signaling pathway s (a signaling protein set) to a cancer protein set C, in which each protein's coding gene (or genes) has a close relationship with cancer genetic disorder, the protein set P of the CSB satisfies that $|\mathbf{P} \cap \mathbf{S}| > 0$, $|\mathbf{P} \cap \mathbf{C}| > 0$, and $|\mathbf{P}| > |\mathbf{P} \cap (\mathbf{S} \cap \mathbf{C})|$. The data for PPIs are collected from IntAct (51), DIP (52), MINT (53), MIPS (54), and BioGrid (55), those for signaling pathways are merged from Pathway Interaction Database (NCI-PID) (2, 3), BioCarta, and Kyoto Encyclopedia of Genes and Genomes (KEGG) (4), and those for cancer proteins are defined by the cancer genes in the Online Mendelian Inheritance in Man (OMIM) database (5, 6).

For more details, one can refer to (50).

1.2 High-quality CSBs

We merged all of the physical interactions in the databases of IntAct (51), DIP (52), MINT (53), MIPS (54), and BioGrid (55). The experimental methods for identification of the PPIs are merged for every PPI. If the number of experimental methods is higher than 2, we keep the PPI in the filtered PPI set. The high quality physical PPI set comprises 2,912 proteins and 4,034 PPIs. The high quality PPI data set facilitates to identify the high-quality CSBs. If all of the interactions of a CSB are derived from the filtered PPI set, the CSB will be filtered out. The high-quality CSB set includes 673 proteins and 1,759 PPIs.

2. Core signaling analysis

There are three parameters in the heuristic method for the mathematical model, that is, the path length l, and the parameters a and b. We set l as 5 in this study. The reason is that there is a small world property in the complex network (56), such as protein interaction network. We computed the average path length for the high-quality CSBs is between 5 and 6. For other two parameters, we extensively tests all of the combinations of a and b. The following supplementary table shows the scales of the core signaling networks while using different parameter combinations.

The parameterization process can help to narrow the decomposed paths from CSBs. We didn't use the stringent rules for parameters a and b. Instead, we chose those parameter combinations

to keep relative large numbers of paths and proteins for brain, lung, and bone metastases of breast cancer patients. The final parameters enable that there are about 300 paths and 100 proteins in the core signaling networks for the 3 metastases. Obviously, if one loosens the parameters, the identified protein paths will still appear in the output networks despite more paths are included in the output. So taking the present parameters is not loss of generality.

Parameters of a and b and corresponding network scales

Metastasis site	Parameter *a*	Parameter *b*	Network scale	
			# paths	#proteins
Brain	3	3	192	61
	3	2.5	192	61
	3	2	302	98 ✔
	2.5	3	278	99
	2.5	2.5	278	99
	2.5	2	544	179
	2	3	522	148
	2	2.5	522	171
	2	2	918	223
Lung	2	2	0	0
	1.5	2	20	23
	1.5	1.5	20	23
	1.5	1	96	74
	1	1.5	333	144✔
	1	1	988	222
Bone	1.5	2	0	0
	1.5	1.5	0	0
	1	1.5	21	24
	1	1	325	134✔

3. **Survival analysis**

The paths in this analysis component take gene probes in the cDNA microarrays as their nodes. Because most genes have multiple probes, the number of paths representing the combinations of probes is increased dramatically. The numbers of paths for brain, lung, and bone are 26,897, 11,956, and 16,542.

The reason why we employed the heuristic cut tree algorithm to narrow the core signaling network is that the general classification methods, especially those based on feature selection, such as Support Vector Machine (SVM), can induce the overfitting problem when tackling a large number of features. Instead, the cut tree algorithm based on hierarchical clustering can justify the robustness of classification and help to remove abundant paths from the clustered tree. The description on the heuristic cut tree algorithm is as following framework:

4. Heuristic cut tree algorithm:

Step 1. Clustering on paths and get a tree **T**

Step 2. Cut tree into branches [$\mathbf{T}_1$, $\mathbf{T}_2$, …, $\mathbf{T}_n$]

Step 3. For each branch $\mathbf{T}_i$, remove the subtree

Clustering on the patient samples and get another tree **T'**

Cut tree into branches [$\mathbf{T'}_1$, $\mathbf{T'}_2$, …, $\mathbf{T'}_m$]

for each branch $\mathbf{T'_j}$, do a classification (Take the samples in the T'j as one class with indicator 1 and other samples as another with indicator 0)

do Kaplan-Meier survival analysis

output the p-value, P_j^{K-M}, for K-M survival analysis

Select a best *j* for the *i* with highest $R_j = \frac{\left|\mathbf{Patients}_j\right|}{\left|\mathbf{T}_j'\right|}$ or $P_j = -\log(P_j^{K-M})$

Step 4. Remove the best *i* with highest R_i or P_i.

Step 5. If #paths < p, stop. Else, go back to Step 1.

In the algorithm, there are three parameters, that is, *m, n* and *p*. The parameter *p* is taken as a relative small value, such as 50 or 100. In this study, we set *p* as 50.The parameter *m* is chosen from the integers between 2 and 20 and the parameter *n* is selected from the integers between 2 and 10. To not lose the generality, we run the algorithm by using all of the combinations of the parameters. The best parameters, *m* and *n*, for brain metastasis of breast cancer are 10 and 7 (EMC192), 8 and 10 (EMC82), 7 and 10 (EMC286); those for lung metastasis of breast cancer are 9 and 6 (the combined cohort of EMC 82 and EMC286); and those for bone metastasis of breast cancer are 8 and 8 (the combined cohort of EMC82 and EMC286).

To determine the best subtree from the outputs of the algorithm, we put all of the subtrees as points in a 3-dimesional cube with the evaluations on the path number, classification ratio (e.g. R_i), and K-M survival analysis *P* value (e.g. P_i). Generally, the points far away from (0, 0, 0) are chosen as the final output core signaling networks (Supplementary Fig 12).

5. Antibodies for Western Blotting

Anti-phospho-RET (Tyr905) (dilution 1:500, rabbit polyclonal, cat. #3221), anti-AKT (dilution 1:500, rabbit monoclonal, cat. #4685), anti-phospho-AKT (Ser473) (dilution 1:1000, rabbit monoclonal, cat. #4060) antibodies were purchased from Cell Signaling; anti-RET (dilution 1:1000, mouse monoclonal, cat. #sc-365943), anti-phospho-Fyn (Tyr416) (dilution 1:1000, goat

polyclonal, sc-16848) were obtained from Sata Cruze Biotechnology; anti-FYN (dilution 1:1500, BD Bioscience, mouse monoclonal, cat. #610164), anti-phospho-KDR (Y1214) (dilution 1:1000, rabbit polyclonal, R&D Systems, cat #AF1766), anti-KDR (dilution 1:500, R&D Systems, mouse monoclonal, cat. #MAB3573), Anti-phospho-p70 S6 Kinase (Thr389) (pS6K1, dilution1:5000; rabbit monoclonal, cat. #04-392), and anti-p70 S6 Kinase (S6K, dilution1:200; rabbit monoclonal, cat. #04-391) antibodies were obtained from Upstate Biotechnology; Anti-mitogen activated protein kinase kinase (MEK, dilution1:10000; rabbit whole antiserum, cat. #5795), anti-phospho-MEK1 (pThr292) (pMEK, dilution1:1000; rabbit affinity isolated antibody, cat. #M2943), anti-ERK1/2 (dilution1:40000; rabbit whole antiserum, cat. #M5670); anti-activated ERK-1&2 (diphosphorylated) (pERK1/2, dilution1:2000; mouse monoclonal, cat. #M9692); anti-GFP (1:200; mouse monoclonal, cat. #G6539); and anti-β actin antibody (dilution1:5000; mouse monoclonal, cat. #WH0002597M1) were obtained from Sigma-Aldrich. All secondary horseradish peroxidase-conjugated antibodies were obtained from GE Healthcare (Anti-mouse IgG, cat. #NXA931; Anti-rabbit IgG, cat. #NA934).

6. RNA interference mediated knockdown

A pool of three different shRNA lentiviral particles was used to knockdown FYN (sc-35425-V), and RET (sc-36404-V).

Hairpin sequences for RET shRNA are:
sc-36404-VA:
ATCCGGATCCTGTTTGTGAATGATTCAAGAGATCATTCACAAACAGGATCCTTTTT
sc-36404-VB:
GATCCCACCACGCAAAGTGATGTATTCAAGAGATACATCACTTTGCGTGGTGTTTTT
sc-36404-VC:
GATCCCCTTCCACATGGATTGAAATTCAAGAGATTTCAATCCATGTGGAAGGTTTTT
Hairpin sequence for FYN shRNA is:
GATCCCATCGAGCGCATGAATTATTTCAAGAGAATAATTCATGCGCTCGATGTTTTT
Lentiviral Particles contain shRNA construct encoding a scrambled sequence is used as control shRNA (sc-108080, Santa Cruz Biotechnology). Cells were transfected with the siRNA Reagent System (sc-45064, Santa Cruz, CA. U.S.A.).

7. Determination of drug concentrations in mouse plasma and brain using UPLC-MS/MS

Linear ion trap quadrupole LC/MS/MS 3200Q trap mass spectrometer (Applied Biosystem/MDS SCIEX, Foster City, CA, USA) equipped with UPLC system was used to determine Sunitinib and Dasatinib in aqueous and biological matrices using Testosterone as an internal standard (IS). The positive ion mode for MS/MS analyses was selected. The quantification was performed using MRM method with the transitions of m/z 400→m/z 284 for Sunitinib, m/z 488→m/z 401 for Dasatinib and m/z 289→m/z 109 for Testosterone (IS). The separation was performed by injecting 10μL of the sample on Acquity UPLC BEH C18 column (50×2.1 mm I.D., 1.7μm, Waters, Milford, MA, USA). Gradient elution with a mobile phase consisted of, 0.1% formic acid (A) and 100% acetonitrile (B) was adapted in the following sequence 0–5% B at 0–0.5 min, 5–90% B at 0.5–2.4 min, 90–5% B at 2.4–3.1 min to separate Sunitinib, Dasatinib and Testosterone (IS) from the matrices. The flow rate was 0.45 ml/min and

the column temperature was kept at 45◦C. The retention time of Sunitinib, Dasatinib and IS were 1.62, 1.58 and 1.85 min, respectively. A good linear relationship with coefficients of determination ≥0.99 was achieved over the Sunitinib's concentration ranges of 119–13300 ng/mL for plasma and 89–9961 ng/g for the brain tumor, Dasatinib's concentration ranges of 35–45000 ng/mL for plasma and 18–21940 ng/g for the brain tumor.

Reference

50. Jin G, Zhao H, Zhou X, Wong ST. An enhanced Petri-net model to predict synergistic effects of pairwise drug combinations from gene microarray data. Bioinformatics. 2011;27:i310-6.

51. Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, et al. IntAct--open source resource for molecular interaction data. Nucleic Acids Res. 2007;35:D561-5.

52. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Res. 2002;30:303-5.

53. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, et al. MINT: the Molecular INTeraction database. Nucleic Acids Res. 2007;35:D572-4.

54. Mewes HW, Frishman D, Guldener U, Mannhaupt G, Mayer K, Mokrejs M, et al. MIPS: a database for genomes and protein sequences. Nucleic Acids Res. 2002;30:31-4.

55. Breitkreutz BJ, Stark C, Reguly T, Boucher L, Breitkreutz A, Livstone M, et al. The BioGRID Interaction Database: 2008 update. Nucleic Acids Res. 2008;36:D637-40.

56. Barabasi AL, Albert R. Emergence of scaling in random networks. Science. 1999;286:509-12.