Journal of
Emerging Investigators

Pancreatic Adenocarcinoma: An Analysis of Drug Therapy Options through Interaction Maps and Graph Theory

Anvita Gupta, Sejal Aggarwal, and Sangeeta Agrawal

BASIS Scottsdale High School, Scottsdale, Arizona; Wright State University, Dayton, OH

Summary

Pancreatic cancer is the fourth leading cause of cancer-related deaths in humans (ASCO 2012). Pancreatic cancer cells exhibit a different gene expression profile from normal cells, with approximately 122 over-expressed proteins. A novel method was created to find the most important areas for future drug development based on influential disease-causing proteins in pancreatic cancer that currently lack drug treatments.

Protein-protein interaction maps were created, and proteins were ranked based on the number of connections each protein exhibited. A protein-drug interaction map was then constructed to analyze which influential proteins have no drugs developed for them or that have a very low drug association level. Afterward, the proteins were graphically and mathematically profiled to further determine which proteins are necessary for immediate research.

Through this method, KRAS, CDKN2A, and RBBP8 were found to be important proteins that lacked drug treatments. By comparing the chemical structure of KRAS to similar chemical structures of other GTPase enzymes and proteins with Walker A motifs, potential drugs were found that could inhibit KRAS and significantly slow the advancement of pancreatic tumors. This approach is applicable to several other types of cancers, such as renal cell carcinoma, melanoma, and prostate cancer.

Received: Sept 6, 2013; **Accepted:** Nov 27, 2013; **Published:** Feb 4, 2014

Introduction

Pancreatic cancer is the fourth leading cause of cancer-related deaths in humans, with an average five-year survival rate of 6% (1). According to the American Cancer Society, approximately 45,220 people will die from pancreatic cancer in 2013 alone (1). Pancreatic adenocarcinoma, like many other cancers, occurs because of the over-expression of oncogenes (*i.e.*, genes that cause cancers), the inactivation of tumor suppressor genes, and the deregulation of various signaling proteins (2). A number of abnormalities in protein pathways cause the changes in cells that lead to pancreatic cancer and tumor growth (3).

Research is often focused on inhibiting proteins that are over-expressed in pancreatic cancer cells (39, 40). However, this approach results in some problems. First,

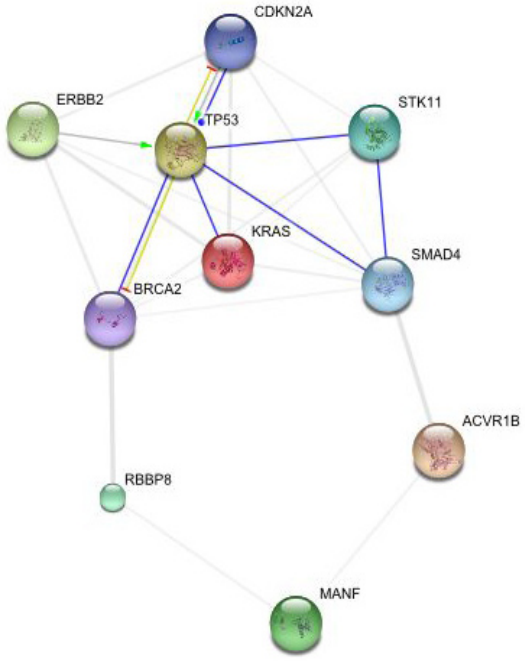


Figure 1. A protein-protein interaction map of pathway starters for pancreatic cancer. Each protein is represented by a labeled circle or node. Line connections between the proteins represent protein-protein interactions and are drawn if the two proteins bind together to carry out a common function. Made using the STRING database (14).

	P-S-PPI	P-G-PPI	S-S-PPI	AbnormalityPrev	NumMut	NumDiseases[32]
KRAS	12.0	9.4	8.1	10.5[26]	10.7[29]	5.1
CDKN2A/p16	10.0	5.0	10.0	9.3[26]	4.3[30]	6.0
TP53	12.0	4.4	8.8	7.3[26]	4.3[31]	10.3
ARMET	4.0	0.6	1.3	3.5[28]	0.7[32]	0.3
RBBP8	4.0	2.8	3.8		1.4[32]	0.9
SMAD4	14.0	4.4	8.8	5.8[26]	2.9[32]	1.4
BRCA2	10.0	4.4	11.3	0.8[26]	4.3[32]	1.1
ACVR1B	4.0	2.2	5.0		1.4[32]	0.3
STK11	10.0	3.9	10.0	0.5[27]	2.9[27]	0.6
ERBB2	10.0	10.6	10.0	2.6[26]		2.6

Figure 2: A data table for protein importance factors for the heat map shown in Figure 3. The data has been normalized by mean normalization (*i.e.*, subtracting the mean for each factor from the value for that factor and dividing by the range).

an over-expressed protein may be just a biomarker that may not cause the tumor to grow (4). A biomarker is "a biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or other condition or disease" (5). Also, there are around 122 over-expressed proteins in pancreatic cancer, and some are more important than others in causing tumor growth (6). Focusing on so many proteins can have negative effects on drug development and success. Current medicines for pancreatic cancer, such as Gemcitabine and Abraxane, extend the life expectancy of patients by about 8.5 months (7). But they cost around $6000 to $8000 per month and cause considerable nerve damage (7).

The authors propose a novel method for finding the most important proteins in pancreatic cancer for future drug research using protein-protein interactions (PPIs). PPIs occur when two or more proteins bind together to carry out their biological function and are important in signal transduction pathways inside the cell (8). A PPI map depicts these various interactions by creating a node on the graph for each protein and connecting the two nodes with an edge if the proteins interact physically. Some examples of PPI maps are shown in **Figures 1**, **6**, **7**, and **8**. By constructing PPI maps with the proteins involved in pancreatic cancer, one can determine the relative amount of influence a protein has in pancreatic cancer protein pathways and thus its importance as a drug target. In the same way that an entrepreneur tries to market to the person with the most friends since they have the greatest influence in their network of acquaintances (9), the proteins with the highest degree of connections should be targeted for future drug research because they interact with the greatest number of cancer-causing pathways.

Afterward, another interaction map can be constructed between these proteins and the drugs currently available on the market to inhibit them to find gaps in current drug development. Factors of importance pertaining to a protein, such as the number of PPIs and the percent of other diseases that display a mutation in that protein, can be used to create a heat map. A heat map is a visualization for a matrix that shows a correlation between a protein and a factor by color, with higher correlations depicted by using darker colors. Examples of heat maps are shown in **Figures 2** and **3**.

We propose that the novel approach of interaction maps can be used to integrate disparate databases to find important drug targets in pancreatic cancer based on disease-causing proteins that currently lack effective drugs. The authors found that KRAS, CDKN2A, RBBP8, and ACVR1B are proteins in the pancreatic cancer network that currently lack FDA-approved drugs. KRAS, CDKN2A, and TP53 were found to be the three most important mutations in pancreatic cancer, which is also true according to the Sanger COSMIC database; this correlation supports the presented approach and data (10).

The authors then applied this methodology to melanoma, renal cell carcinoma, and prostate cancer, all of which have been successfully treated by drug therapy. In each case, the inhibition of the protein with the most PPIs in its cancer network significantly slowed tumor growth. Either this protein was targeted directly by an end drug as with prostate cancer, or the protein targeted had a high confidence interaction (confidence > 0.95) to the most important protein in the cancerous network. This step both validated the authors' hypothesis and illustrated the potential application of this methodology to other cancers to develop successful cancer medication in a more time-effective manner.

In the third stage, KRAS was determined to be the most important protein among the four proteins lacking drugs in pancreatic cancer since, according to The Kyoto Encyclopedia of Genes and Genomes (KEGG), a KRAS mutation starts the pancreatic cancer pathway that inhibits apoptosis in tumor cells. This mutation is present in 90% of tumors (11) and has been linked with increased resistance to pancreatic cancer medication. KRAS is a member of the RAS family, one of the most frequently mutated classes of oncogenes in pancreatic cancer (12). The presence of RAS, a type of GTPase protein, in human cancers was discovered in 1982 (12). This discovery spurred research into the chemical structure of RAS, which lead to the discovery that small GTPases regulate almost all cellular processes (12). Eighty-five percent of all RAS mutations are in KRAS, and these occur most often at codons 12, 13, and 61(12). Since binding directly to the GTPase protein has not been successful, researchers have focused on inhibiting the downstream effectors of KRAS, such as RAF, MEK, ERK, and PI3K. However, since RAS does not rely solely on the MAP signaling pathway, which contains the effectors that are currently targeted, blocking it by

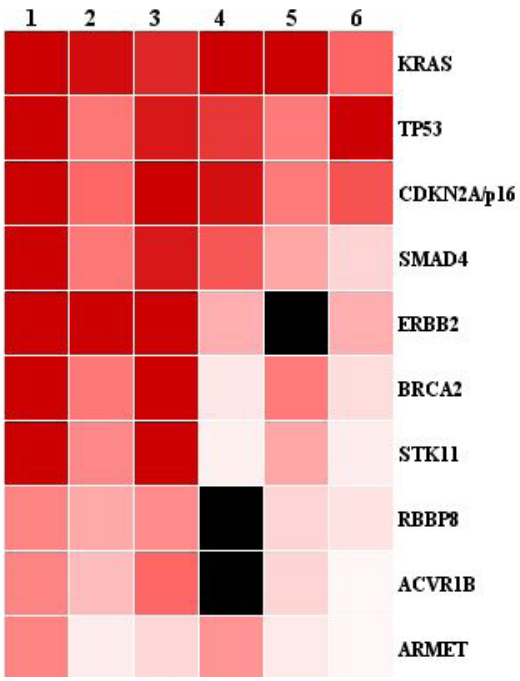


Figure 3: A heat map of factors in protein importance, with protein pathway starters in pancreatic cancer. Data for protein-protein interactions was retrieved from STRING (14). (**1**) **Primary PPI-Pathway Starters**: how many primary PPIs the protein had with the ten other proteins on this list and in **Figure 1**. (**2**) **Primary PPI- General Pancreatic Cancer Proteins**: how many interactions a protein had with all other proteins involved in pancreatic cancer rather than just the protein pathway starters shown in **Figure 1**. (**3**) **Secondary PPI- Pathway Starters**: how many PPIs those proteins that the initial protein connected with had with other proteins. (**4**) **% Tumors Mutated In**: the percentage of malignant pancreatic tumors that display a mutation in that protein. (**5**) **# Mutations in Tumors**: the number of different mutations for this protein that can occur in tumors. (**6**) **# Other Diseases Protein Mutated In**: the number of other diseases this protein is mutated in. Darker shades of red indicate higher values for that factor. Higher values for these factors indicate an increased chance that a protein will be important in the pancreatic cancer network, and thus there is a greater need for focused drug research. So that # Other Diseases Protein Mutated In, which ranges from 1-6, would be on the same scale as % Tumors Mutated In, which ranges from 1-100, and the other factors, the mean was calculated for each factor and subtracted from each value, and the resulting number was divided by the range of values for that factor. The resulting table is shown in **Figure 10** and was used to create this heat map.

inhibiting its downstream effectors has not proved successful (12). Thus, concurrent blocking of two or more pathways downstream of KRAS will be needed. The authors studied the chemical structure of KRAS and drugs that inhibit proteins with a similar structure to KRAS. Through this method, Tipifarnib, Lonafarnib, Clonodine, and three other drugs with a similar chemical structure to Clonodine were determined to be likely to inhibit KRAS and thus merit future testing.

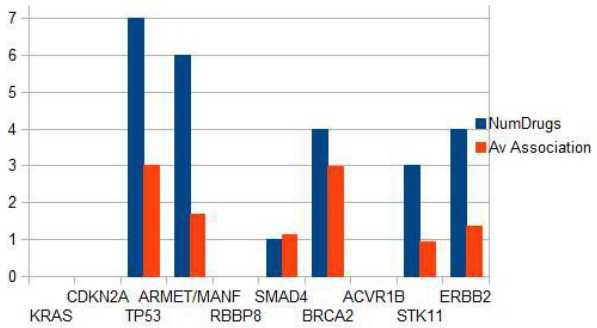


Figure 4: A bar graph showing ten protein pathway starters and proteins produced by disease genes in pancreatic cancer, with the number of drugs and average association on the y-axis and the protein on the x-axis. Each entry in the Connectivity Maps database that states a particular drug that targets cells with mutations in a certain protein is considered a "protein-drug interaction." The blue bars represent the number of protein-drug interactions, and the orange bars represent the average association of the drug with the protein. Drug Association refers to how effective the drug is in targeting cells with a mutation in that protein and thus inhibiting the effects of the protein. Thus, the average drug association is the sum of the associations for all protein-drug interactions divided by the number of protein-drug interactions for a particular protein.

Results

A PPI map was created for pancreatic cancer proteins whose mutations lead to changes in protein pathways causing cancer (KEGG). Additional mutated proteins were found from the database Diseasome, which depicts gene-cancer interactions by mapping diseases to the genes whose differential expression are associated with the disease (13). The interaction map of 10 proteins from KEGG and Diseasome is shown in **Figure 1**.

Additional data on how likely a protein is to be important in pancreatic tumor growth was collected through database and literature mining and condensed into the table shown in **Figure 2** (11, 14-20). A heat map was created for the pancreatic cancer proteins based on various parameters outlined in the methods section and summarized in **Figures 2**, **3** , and **4** . The heat map was used to predict the likelihood that a protein was important in the pancreatic cancer pathway; proteins TP53, KRAS, and CDKN2A were determined to be proteins with a high degree of importance in their protein pathways.

A corresponding heat map of existing FDA-approved drugs for these proteins and the effectiveness of the existing drugs was created from data collected from the Connectivity Maps (Cmaps) database (21). This map investigated the effectiveness of existing drugs for a protein. If a protein has a high degree of drug development, then further drug development is not needed as urgently for that protein. Each entry in the Cmaps database that states that a particular drug targets cells with a mutation in a certain protein has been designated a "protein-drug interaction." Each such interaction is supported by a number of articles, and the number of protein-drug interactions per protein is mapped in **Figure 4**. The

	# Drugs[25]	Drug Effectiveness (Av. Assoc)[25]	# Evidence Articles linking drug (av) [25]
KRAS	0.0	0.0	0.0
CDKN2A	0.0	0.0	0.0
TP53	1.0	1.0	0.6
ARMET	0.9	0.6	0.4
RBBP8	0.0	0.0	0.0
SMAD4	0.1	0.4	0.3
BRCA2	0.6	1.0	1.0
ACVR1B	0.0	0.0	0.0
STK11	0.4	0.3	0.4
ERBB2	0.6	0.5	0.9

Figure 5: An unscaled data table for the heat map of drug factors, which is shown in Figure 3. The data has been normalized by mean normalization (*i.e.*, subtracting the mean for each factor from the value for that factor and dividing by the range).

average number of evidential articles was also recorded for all drugs for a particular protein. How effectively or to what degree a drug targeted cells with a particular protein mutation is referred to in Cmaps and in this paper as the drug's "association" with the protein. The completed, normalized table of these factors is shown in **Figure 5**, and the heat map is shown in **Figure 6**.

The two heat maps were then compared to find proteins for which future drug therapy is needed (**Figure 3**) and to determine which current drugs are most effective (**Figure 6**). Out of the ten proteins from KEGG and Diseasome that initiated protein pathways in pancreatic cancer, four had no known drug and three had drugs with an average association lower than 1.96. An average association level of 1.96 was observed for all drug-protein interactions for the studied proteins in pancreatic cancer; thus, an association above 1.96 meant that the drug-protein association and drug effectiveness was above average. Successful drugs for other diseases usually have an association above 1.96 (21). By analyzing this data, the authors found that KRAS, CDNK2A, RBBP8, and ACVR1B all lack effective drug treatments. Although a mutation in ARMET/MANF is characteristic of some pancreatic cancer patients, ARMET/MANF is shown to be the protein least likely to be important in pancreatic tumor growth of those proteins examined based on PPIs; yet, it has six drugs associated with inhibiting it. On the other hand, KRAS is a protein with more PPIs than ARMET/MANF, is found more frequently in tumors, exhibits a number of mutations in cancers, and begins the PI3K-AKT pathway in pancreatic cancer; however, it has no drug associated with inhibiting it. The PI3K-AKT signaling pathway regulates transcription, translation, growth, and proliferation of the cell and is activated by the mutation of the gene K-RAS (22). Out of the 15 proteins from the KEGG, Diseasome, and drug databases that were shown to be most important in pancreatic cancer pathways, five had no known drug, and four had drugs with extremely low association scores. STK11, ARMET, SMAD4, and ERBB2 had drugs with an average association of less than two. ERBB2, whose over-expression is strongly linked with cancer recurrence (23), is identified as one of the two primary oncogenes in pancreatic cancer, along

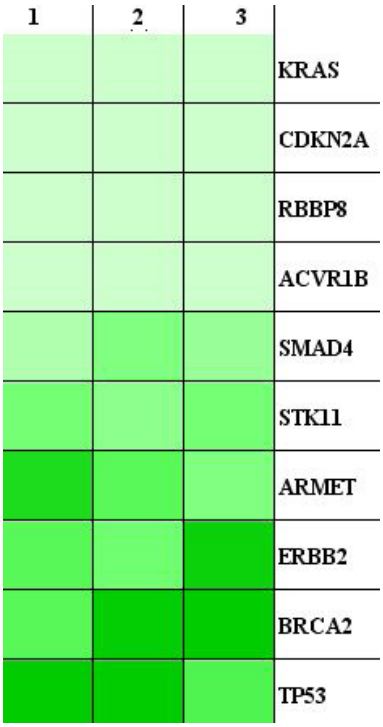
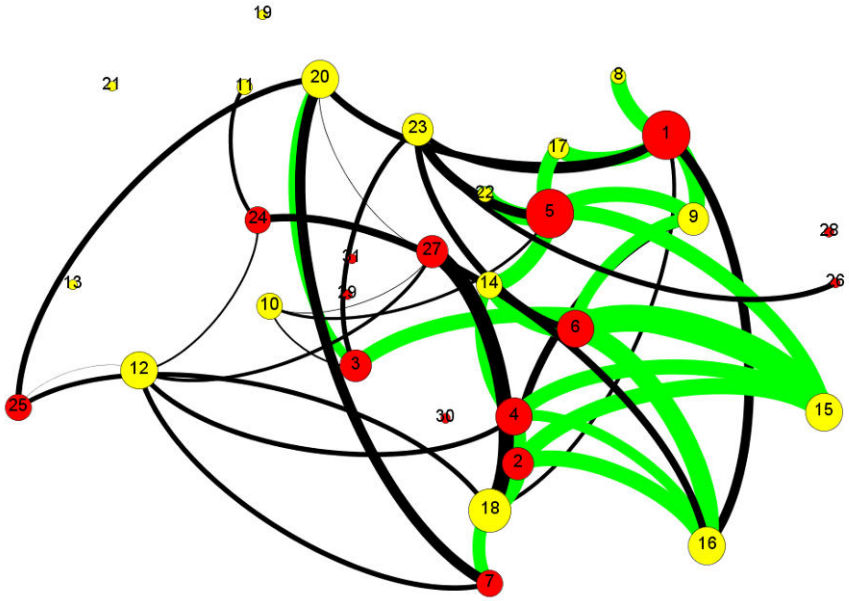


Figure 6: A heat map of drug-related factors for protein-pathway starters in pancreatic cancer. Each entry in the Connectivity Maps database that states a particular drug targets cells with a mutation in a certain protein is considered a "protein-drug interaction." (**1**) **# Drugs Available**: how many protein-drug interactions a particular protein has in the Cmaps database. (**2**) Drug Association refers to how effective the drug is in targeting cells with a mutation in that protein and thus inhibiting the effects of the protein. Thus, the **Average Drug Association** is the sum of the associations for all protein-drug interactions divided by the number of protein-drug interactions for a particular protein. (**3**) Each protein-drug interaction is supported by a number of research papers from PubMed; thus, **# Evidential Articles (Av)** is the average number of evidential articles supporting each protein-drug interaction. Darker shades of green indicate higher values. Lower values indicate a greater need for further research and drug development since mutations of the proteins at the bottom of the heat map are targeted by numerous effective drugs, whereas mutations of those at the top are not. So that Average Drug Association, which ranges from 0-1, would be on the same scale as # Drugs Available, which ranges from 1-6, and # Evidential Articles, the mean was calculated for each factor and subtracted from each value for that factor, and the resulting value was divided by the range of values for that factor. The resulting table is shown in **Figure 9** and was used to create this heat map.

with KRAS. **Figure 4** shows that ERBB2 activates TP53, the most influential of the cancer genes, and yet has only four available drugs to inhibit it and the third lowest average association score. Drug development is further needed in this area.

According to the protein-drug interaction map in **Figure 7**, the most effective drug is Methyl Methanesulfonate (MMS) when taking into account association and the



Id	Label	Type of Molecule
1	P53	Protein
2	BRCA2	Protein
3	ARMET	Protein
4	BRCA1	Protein
5	ATM	Protein
6	RAD51	Protein
7	SMCA4	Protein
8	4-Nitroquinoline-1-Oxide	Drug
9	Camptothecin	Drug
10	Dexamethasone	Drug
11	Dextrans	Drug
12	Estradiol	Drug
13	Hydrocortisone	Drug
14	Hydroxyurea	Drug
15	Methyl Methanesulfonate	Drug
16	Mitomycin	Drug
17	Nocodazole	Drug
18	Tamoxifen	Drug
19	Teniposide	Drug
20	Tretinoin	Drug
21	Triamcinolone Acetonide	Drug
22	Topotecan	Drug
23	Bleomycin	Drug
24	STK11	Protein
25	PTEN	Protein
26	SMAD4	Protein
27	ERBB2	Protein
28	KRAS	Protein
29	CDKN2A	Protein
30	RBBP8	Protein
31	ACVR18	Protein

Figure 7: A Drug Effectiveness Map for the set of ten proteins whose mutations lead to cancerous protein pathways, both with and without already developed drugs. Connections between a drug and protein are created if the drug and protein have a "protein-drug interaction" recorded in the Cmaps database. Edge thickness between the drugs and proteins indicate the association of the protein-drug interaction. Drug Association refers to how effective the drug is in targeting cells with a mutation in that protein and thus inhibiting the effects of the protein. The proteins are indicated by red nodes (circles) and the drugs by yellow in order to distinguish them. Node size corresponds to the number of connections each protein/drug has. Green Edges/Connections correspond to an association > 1.98, the mean association level.

number of proteins inhibited. Association between a drug and protein refers to how effectively a drug targets cells with mutations in that protein. MMS targeted cells with mutations in BRCA2, BRCA1, ARMET, RAD51, and ATM. All of these proteins, with the exception of ARMET, had many PPIs. MMS proved very effective in eliminating cells with mutations of these closely related proteins, but not effective for other proteins less closely related. It causes hyperacetylation of cytoplasmic and nuclear proteins, a process directly leading to the cell death pathway (24). Cancer cells show increased susceptibility to MMS. However, MMS also can act as a carcinogen by increasing the risk of secondary cancers (25). Mitomycin is the second most effective drug from the drugs mapped and is used in a variety of cancers after chemotherapy as a further attack against cancer cells. However, prolonged use can cause significant nerve damage and has exhibited carcinogenic activity in rats (24).

When studying the chemical structure of KRAS, the authors found that KRAS is a member of the P-loop NTPase domain superfamily, a member of the RAS subfamily, and associated with GTPase because it contains the Walker A motif (26). Through literature mining, the authors found that Lonafarnib has the potential to inhibit KRAS in clinical trials (27). Therefore, drugs closely related to Lonafarnib also have a higher chance of inhibiting KRAS. Lonafarnib is a member of the Farnesyltransferase inhibitor family of drugs, which, by targeting Farnesyltransferase, prevents the addition of a farnesyl group to KRAS and thus prevents KRAS from embedding in the cellular membrane and carrying out its function (28). The drug Tipifarnib belongs to the same class of drugs, and shares a maximum common substructure of 14 with Lonafarnib (29). Maximum common substructure (MCS) refers to the length of the largest substructure present in both molecules, and chemical structures with a larger MCS have similar functions (30). KRAS has a small-molecule binding pocket between its Helix Alpha 2 and core beta sheet, which can fit a chloro and benzine group. Clonodine/DCAI can bind to this pocket and inhibit KRAS (31). Using PubChem, Clonodine's molecular structure was studied, and other chemical compounds were clustered based on their structural similarity to DCAI. The three drugs with chemical structure most similar to DCAI were CN5710355, CID12699319, and CID12699317 (32). These three chemicals also have a higher likelihood of binding to the KRAS binding pocket and preventing KRAS from carrying out its function in cancer.

To validate the hypothesis that PPI maps can be used to find important proteins that lack drug development within a cancer network, the authors then studied the protein pathways of prostate cancer, melanoma, and renal cell carcinoma. All of these cancers have been treated with drug therapy with some measure of success; Enzalutamide has reduced serum levels of the biomarker prostate specific antigen by greater than 50% by targeting the androgen-specific receptor protein AR (33). The FDA has approved the BRAF inhibiting drug Vemurafenib, and CDK4 inhibitors are in clinical trials to treat patients with advanced metastatic melanoma (34). Everolimus is used to inhibit the mTOR signaling pathway in renal cell carcinoma (RCC). As mTOR activates

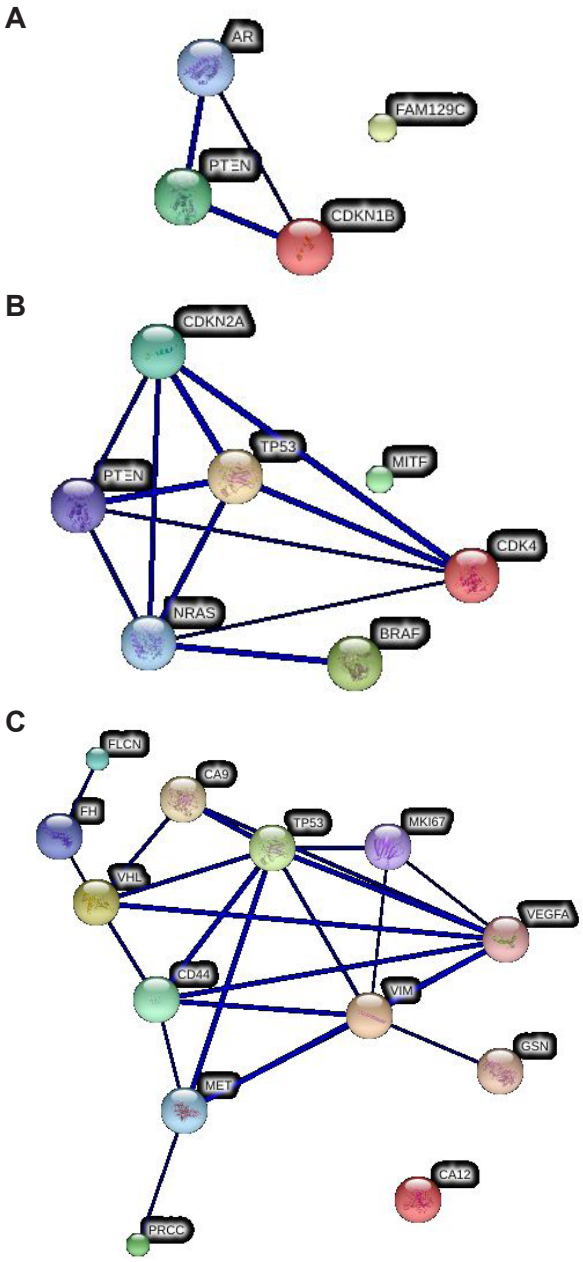


Figure 8: PPI networks in cancer. A. A PPI network for protein pathway starters in prostate cancer. The inhibition of the protein AR has reduced tumor growth in prostate cancer. **B. A PPI network for protein-pathway starters in melanoma.** The inhibition of the protein BRAF has reduced tumor growth in melanoma. BRAF interacts with a high confidence level to NRAS, which has one of the greatest numbers of PPIs in the melanoma network.**C. A PPI network for protein pathway starters in renal cell carcinoma.** The inhibition of the protein mTOR has reduced tumor growth in renal cell carcinoma. mTOR is not of primary importance in the renal cell carcinoma network, but interacts with high confidence to VEGF, which has one of the greatest number of PPIs in its network. These network was made using STRING. The thickness in the connection line corresponds to the confidence level for the interaction based on co-expression, literature mining, high throughput experiments, etc.

VEGF, which is critical in the RCC network, Everolimus is an effective treatment for RCC (35). The proteins that begin protein-pathways in these cancers were mapped in STRING (a database of known and predicted protein protein interactions) and their interactions examined. The PPI maps for prostate cancer, melanoma, and renal cell carcinoma appear in **Figure 8A-C**. Applying PPI networks can explain why these drug therapies have been a success, and PPI networks can be used in the future to develop effective medicines for targeted therapies of cancer.

Discussion

The use of PPIs as a method of finding the most important proteins in cancer was validated when the authors applied the approach to cancers such as melanoma, prostate cancer, and renal cell carcinoma, all of which have been successfully treated by targeted drug therapy. In the experiment, a "high-confidence" interaction is taken to be an interaction with a confidence score above 0.95, which is based on experimental data, how close the two genes are that produce the proteins, and other factors used to predict confidence by STRING. The authors found that successful targeted therapy seems to have two main trends; either the protein targeted has a high number of primary PPIs with pathway starters, as in the case of CDK4 or AR, or the protein targeted has a high-confidence interaction with the protein most influential in its cancer network. For example, VEGF has the greatest number of primary PPIs with pathway starters in renal cell carcinoma, but when the PPI network is re-centered on VEGF, then mTOR is shown to bind to and inhibit VEGF with a confidence level of 0.947 (14). The high degree of VEGF in the renal cell carcinoma network is illustrated in **Figure 8C**, and the interaction map of VEGF is shown in **Figure 9**, with the mTOR-VEGF interaction highlighted. Similarly, BRAF, the target of anti-melanoma drug Venurafenib, interacts with NRAS with a confidence level of 0.975; NRAS has four primary PPIs with pathway starters in its network, the highest level of any pathway starter for prostate cancer. Thus, centrality in cancerous networks, determined by PPIs, can be effectively used to find targeted therapy options for a variety of cancers in a more time-effective manner. As shown in **Figure 10**, the targeted proteins for melanoma and prostate cancer share a very high number of PPIs, a high percentage of tumors they occur in, and have drugs with a high average association. Since KRAS has a high number of protein-protein interactions and percentage of tumors it occurs in, its need for immediate drug development is increased.

PPI maps, heat maps, and the approaches outlined above can help bring attention to new drug targets, as well as rank these drug targets for how much influence they are likely to have in their cancerous pathways. However, researchers should also consider how easy the protein is to target with chemical compounds, or its "druggability." Some factors in druggability include the protein's structure, how likely it is to mutate, clinical toxicity, and resistance mechanisms. KRAS, in

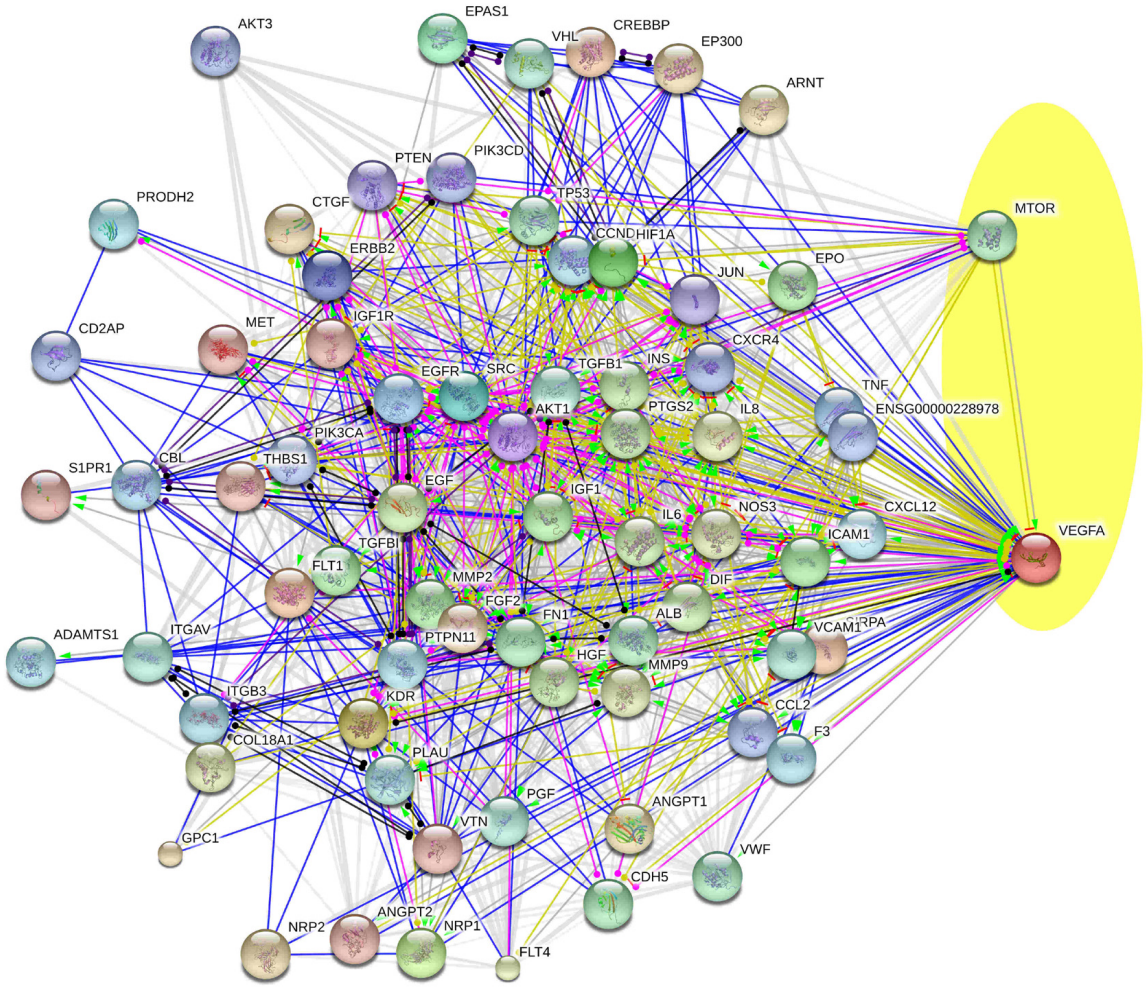


Figure 9: The interaction between mTOR and VEGF, with the PPI network centered on VEGF.

particular, has proved very difficult to target despite the amount of research that has gone into it. About 40% of KRAS-mutant colorectal cancers are genetically heterogeneous, so generic drug development is difficult. KRAS functions by GTP binding, which researchers have found extremely difficult to inhibit (17).

Future research should focus on identifying amino acid sequences that are homologous among multiple GTPase enzymes and associated with multiple types of cancer. In addition, there is a need to identify proteins with structures and amino acid sequences similar to KRAS. By finding proteins with amino acid sequences homologous to those of KRAS, researchers should be able to identify classes of drugs that inhibited proteins similar to KRAS and that could be used to inhibit KRAS itself. In the future, research should also focus on testing drugs that are effective on other members of the P-loop NTPase domain superfamily, of the RAS subfamily, associated with GTPase, and containing the Walker A motif. The drugs isolated in this study, Lonafarnib, Tipifarnib, CN5710355, CID12699319, and CID12699317, should also be tested because they have a greater chance of inhibiting KRAS effectively. Also, KRAS can be targeted like VEGF was in renal cell carcinoma: by designing drugs for another protein that has a high confidence interaction to KRAS and a higher "druggability" than KRAS.

In future work, the authors would also create a separate factor for the strength of a protein-protein interaction and confidence level with which that PPI is stated when ranking proteins by potential effect in their cancerous networks. After all, a protein with many weak interactions that are not necessarily well-supported may not be as likely to be important in a cancerous network as a protein with a few strong, well-supported interactions. Research should also be conducted on whether a protein that has many primary interactions with other proteins but few secondary interactions is more likely to be important in a cancerous network than proteins with few primary interactions but many secondary reactions, such as KRAS. For example, if protein X interacts with protein Y but protein Y interacts with five other proteins downstream, and protein Z interacts with three proteins but those proteins only interact with one other protein downstream, which protein is likely to be more important in its cancerous network? Further research is needed to address these questions on PPI maps. Interaction maps can elucidate new drug targets in cancer, allow

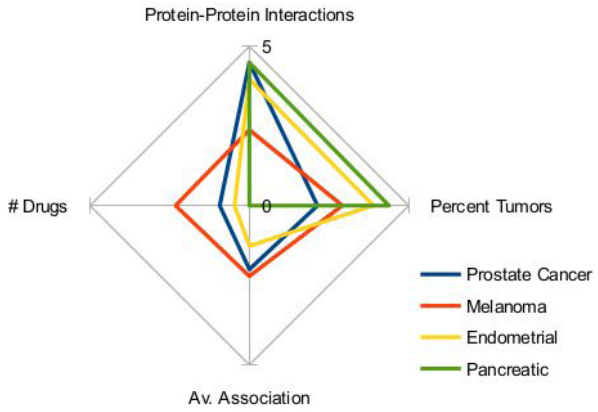


Figure 10: A Radar Chart depicting mTOR, BRAF, and AR. Inhibition of these proteins in renal cell carcinoma, melanoma, and prostate cancer, respectively, significantly reduced tumor growth (33, 34, 35), the percentage of tumors this protein was mutated in, the number of protein-protein interactions it had within its network, the number of drugs that inhibited it, and the average association of those drugs. KRAS is shown in green for pancreatic cancer, and its high value for PPIs and Percentage of Tumors, as well as its low values (0, 0) for the number of drugs and average association, exhibit how future research is needed for this protein. Feature normalization has been performed for each factor.

researchers to prioritize proteins for future research, and should be incorporated into the drug development process.

Methods

Creating a set of proteins whose protein-protein interactions should be examined

Using the KEGG database, the authors added the seventy proteins listed on the database as involved in pancreatic cancer pathways to the set of proteins whose PPIs were to be studied. Additional candidates for important disease-causing proteins produced by mutated genes were found at http://www.diseasome.ed/. Diseasome depicts gene-cancer interactions by mapping diseases to the genes whose mutations or differential expressions are associated with causing the disease (13). The proteins from KEGG and Diseasome were mapped using the STRING Database. In STRING, PPIs are predicted based on the following: 1) how close the two genes that produce that protein are on the chromosome, 2) their conservation in other species, 3) if the two proteins are expressed together, 4) high throughput experiments, 5) database mining, and 6) text mining. In this stage of the experiment, text mining was disabled. When text mining is counted as evidence, two proteins are considered to interact if they are mentioned together in PubMed abstracts. Conservation in other species is taken into account because proteins that are "functional partners" and thus interact often are expressed in the same species (14). The PPI map was examined to gain a rough idea of which proteins are most influential. Proteins with the greatest number of connections to other proteins, or greatest degree, will have a greater degree of influence in cancerous protein pathways. The "degree" of a protein refers to how many other proteins it interacts with, or how many outgoing links it has. With a PPI map of 80 proteins, however, the degrees of the individual proteins are much more difficult to determine. Thus, the 6 proteins that began protein-pathways in pancreatic cancer, according to KEGG, were isolated from the previous set of 70 proteins, and mapped separately using STRING along with the disease-causing proteins found from Diseasome, and each protein's degree was recorded.

Studying the Protein-Protein Interaction Maps

PPIs were classified as either primary with pathway starters, primary with all other proteins involved in pancreatic cancer or secondary with pathway starters. Let an arbitrary protein be taken as protein X. If protein X interacts and binds with protein Y, then the interaction between X and Y was classified as primary. If Y was one of the ten proteins that started cancerous pathways or was a disease gene from Diseasome, then the interaction was further classified as primary with other pathway starters. If Y was simply an arbitrary protein involved in the pancreatic cancer network, the interaction between X and Y was classified as primary with general proteins. However, if protein X binds to protein Y, and protein Y further binds to protein Z, then the interaction between X and Z was classified as secondary. Secondary interactions were only mapped with pathway starters from KEGG and Diseasome.

Finding proteins for which drug development is needed through Drug-Protein Interaction Maps

The set of protein-pathway starters and proteins produced by disease genes was then compared against the list of proteins that have existing drug treatment options to see where future development is needed. This was done by visiting the database Cmaps, which links diseases to the proteins that cause them and the proteins to drugs that target cells with a mutation of that particular protein (21). Connectivity maps only displays proteins with drugs that target cells with that particular protein mutation. The set of ten proteins found from Diseasome and KEGG was searched and downloaded, and the number of drugs for each protein was recorded.

A drug-protein interaction map was then created using Gephi (36) to determine areas where drug development is lacking, and the effectiveness of current drug therapy options. A graph was created with a red node for each protein from the data set from Kegg and Diseasome, as well as the ten most important proteins with existing drugs from the Cmaps database. A yellow node for each drug found in the file downloaded from Cmaps was created. A new edge was created for each drug-protein interaction from Cmaps, where at least three evidential articles supported the effect of the drug on the protein. An edge weight was assigned for each edge based on the association level between the drug and protein shown in Cmaps. Association refers to how well the drug is able to inhibit the protein. The edges with an association > 1.98 (average association) were

assigned the color green. This map consolidates the data collected from Cmaps and KEGG to show which proteins lack drugs. One can easily examine which proteins from this set are the most necessary for drug development. Drug need can also be examined by the creation of a bar chart with the protein on the x-axis and the number of drugs on the y-axis, as shown in **Figure 4**.

Creating Heat Maps to rank proteins based on importance as potential drug targets

Heat maps can be used as a tool to sort proteins based on a number of factors in a qualitative manner, where each column represents a factor and each row represents a protein. A high value for that factor is represented by a darker shade (37). The first factor examined was the number of PPIs with pathway starters. The second factor examined is the number of primary PPIs with all proteins involved in pancreatic cancer. Let protein X again be an arbitrary protein that begins a pathway in pancreatic cancer. Protein X's interactions with one hundred other proteins was mapped through STRING, and the data filtered to include only protein X's interactions with other proteins that were involved in pancreatic cancer, out of those one hundred proteins. The third factor examined was the number of secondary PPIs with pathway starters for each protein. The authors performed a literature review of PubMed and PNAS to find articles stating the percentage of pancreatic tumors each protein was mutated in and the number of mutations of that protein that occur in cancerous tumors (11, 15-20). These two factors were also examined in the heat map. The final factor was the number of other diseases this protein is involved in, as recorded in the KEGG database (20). Feature normalization was performed by scaling the columns in each data table so that a high value on the mutation score is in the same range as a high value on the PPIs. Each value for a particular feature was divided by the range of that feature. The completed table is illustrated in **Figure 2**.

Another excel spreadsheet was created with the number of drugs per protein and average association for all drugs with that protein. We called each entry in the Cmaps database that states that a particular drug inhibits a certain protein a "protein-drug interaction." Each protein-drug interaction is supported by a number of evidential articles. The number of evidential articles was summed for a protein's protein-drug interactions and divided by the number of protein-drug interactions to create the average number of evidential articles. A high amount of scientific support for a protein's protein-drug interactions paired with a high association level for those protein-drug interactions indicates that a protein has less need for immediate drug development. Feature normalization was again performed on these three factors, and the completed, normalized table is shown in **Figure 5** . With the use of GiTools, two heat maps were created; one examined factors contributing to protein importance within pancreatic cancer, and the other examined how much research has already taken place for that protein to find gaps in drug therapy.

Filling Existing Gaps in Drug Therapy by focusing on KRAS

The authors then aimed to fill existing gaps in drug therapy by determining which amino acid sequences were important to KRAS, the highest ranked protein that lacked drugs. In the NCBI Database (38), RAS was searched for under "Conserved Domains." Using the first result, possible proteins were predicted that might share homologous sequences with KRAS. Lonafarnib, a member of the Farnesyltransferase inhibitor drug family, has been shown to inhibit KRAS in clinical trials (28). The structure of Tipifarnib, another Farnesyltransferase inhibitor, was compared to Lonafarnib using the ChemMine toolbox. Tipifarnib and Lonafarnib share a maximum common substructure of length 14 (29). Greater maximum common substructure implies greater similarity between the drugs in both structure and function (30). Thus, Tipifarnib should also be tested on KRAS. The binding pockets found in KRAS were then searched to find small molecule ligands that may bind there to inhibit KRAS.

References

1. Cancer.Net Editorial Board. "Pancreatic Cancer-Statistics." American Society of Clinical Oncology (ASCO) - Cancer.net, 25 June 2012. Web.
2. Ghaneh, Paula, Eithne Costello, and John P. Neoptolemos. "Biology and Management of Pancreatic Cancer." *Postgraduate Medical Journal*. 84.995 (2008): 478-97. Print.
3. Sarkar, F. H., S. Banerjee, and Y. Li. "Pancreatic Cancer: Pathogenesis, Prevention, and Treatment." *Toxicology and Applied Pharmacology* 3rd ser. 1.224 (2007): 326-36. Print.
4. Sawyers, CL. "The Cancer Biomarker Problem." *Nature* 7187th ser. 452 (2008): 548-52. Print.
5. "Biomarker." NCI Dictionary of Cancer Terms -. National Cancer Institute, n.d. Web.
6. Alldinger, I., D. Dittert, M. Peiper, A. Fusco, G. Chiappetta, E. Staub, M. Lohr, R. Jesnowski, G. Baretton, D. Ockert, HD Saeger, R. Grutzmann, and C. Pilarsky. "Gene Expression Analysis of Pancreatic Cell Lines Reveals Genes Overexpressed in Pancreatic Cancer." *Pancreatology* 5.4-5 (n.d.): 370-79. Print.
7. Pollack, Andrew. "Drug Is Shown to Help Pancreatic Cancer Cases." *New York Times*. 22 January, 2013. Web.
8. De Las Rivas J, Fontanillo C (2010) Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks. *PLoS Computational Biology* 6(6): e1000807. doi:10.1371/journal.pcbi.1000807
9. Domingos, P. "Mining Social Networks for Viral Marketing." *IEEE Intelligent Systems* 1st ser. 20 (2005): 80-82. Print.
10. Forbes S. et al. "COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer." *Nucleic Acids Research* (2011) 39 (suppl 1): D945-D950. doi: 10.1093/nar/gkq929. Print.
11. Konner Jason, O'Reilly Eileen. "Pancreatic cancer:

epidemiology, genetics, and approaches to screening." *Oncology* (Williston Park) 2002 Dec;16(12):1615–1638. Print.
12. Cox A., Der C. "Ras history. The saga continues." *Small GTPases* (2010) 1:1, 2-27. Print.
13. Reichardt, J. Alamino,R., Saad, D. "The Interplay between Microscopic and Mesoscopic Structures in Complex Networks." *PLoS ONE* 6.8. Print.
14. Jensen et al. "STRING 8--a global view on proteins and their functional interactions in 630 organisms." *Nucleic Acids Research*. 2009, 37(Database Issue):D412-416.Print.
15. Su GH, Hruban RH, Bansal RK, Bova GS, Tang DJ, et al. "Germline and somatic mutations of the STK11/LKB1 Peutz-Jeghers gene in pancreatic and biliary cancers." *American Journal of Pathology*. 1999;154:1835–1840. Print.
16. Shridhar, V., Rivard, S., Wang, X., Shridhar, R., Paisley, C., Mullins, C., Beirnat, L., Dugan, M., Sarkar, F., Miller, O.J., Vaitkevicius, V.K., Smith, D.I."Mutations in the arginine-rich protein gene (ARP) in pancreatic cancer."*Oncogene*(1997); 14(18):2213-6. Print.
17. Helwick,C. "Targeting KRAS in GI Cancers: The Hunt for the Holy Grail in Cancer Research." *ASCO Post* (2012). 3.6. Print.
18. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D. "GeneCards: integrating information about genes, proteins and diseases." *Trends in Genetics* 13: 163 (1997). Print.
19. Rivlin N, Brosh R, Oren M, Rotter V. Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. *Genes & Cancer*. 2011;14(4):466–474. Print.
20. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M. "KEGG for integration and interpretation of large-scale molecular datasets." *Nucleic Acids Research*. Oxford Journals (2012). Print.
21. Hui Huang, Xiaogang Wu, Ragini Pandey, Jiao Li, Guoling Zhao, Sara Ibrahim, Jake Y. Chen (2012) C2Maps: A network pharmacology database with comprehensive disease-gene-drug connectivity relationships BMC Genomics. Vol. 13, Supplement 6, S17, 2012. Print.
22. Morgensztern D, McLeod HL. "PI3K/Akt/mTOR pathway as a target for cancer therapy." *Anticancer Drugs*. 16.8 (2005): 797-803. Print.
23. Zhou, X.,et al. "Activation of the Akt/Mammalian Target of Rapamycin/4E-BP1 Pathway by ErbB2 Overexpression Predicts Tumor Progression in Breast Cancers." *Clinical Cancer Research* (2004) 10; 6779. Print.
24. International Agency for Research on Cancer. "Methane Methanesulfonate." IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. 10 (1974): 291. Print.
25. International Agency for Research on Cancer. "Mitomycin." IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. 10 (1976): 291. Print.
26. Carpeño, J.,Belda-Iniesta,C. "KRAS mutant NSCLC, a new opportunity for the synthetic lethality therapeutic approach." *Translational Lung Cancer Research*. 2013;2(2):142-151. Print.
27. Baines, A., Xu,D. , Der, CJ. "Inhibition of Ras for cancer treatment: the search continues." *Future Medicinal Chemistry*. 2011 October; 3(14): 1787–1808. Print.
28. Reuter C.W., Morgan M.A., Bergmann L. "Targeting the Ras signaling pathway: a rational, mechanism-based treatment for hematologic malignancies?"*Blood*. 2000;96(5): 655-1669. Print.
29. Backman TW, Cao Y, Girke T. "ChemMine tools: an online service for analyzing and clustering small molecules." *Nucleic Acids Research*. 2011 Jul; 39(Web Server issue): W486-91. Web.
30. Cao Y., Jiang T., Girke T. "A maximum common substructure-based algorithm for searching and predicting drug-like compounds." *Bioinformatics*. (2008) 24 (13): i366-i374. Print.
31. Maurer T., et al. "Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS mediated nucleotide exchange activity." *Proc Natl Acad Sci USA* (2012);109:5299–5304.Print.
32. Bolton E, Wang Y, Thiessen PA, Bryant SH. PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annual Reports in Computational Chemistry*, Volume 4, American Chemical Society, Washington, DC, 2008 Apr. Web.
33. Tran, et al. "Development of a Second-Generation Antiandrogen for Treatment of Advanced Prostate Cancer." *Science*. 2009 May 8; 324(5928): 787–790. Print.
34. Genetech. "FDA Approves Zelboraf (Vemurafenib) and Companion Diagnostic for BRAF Mutation-Positive Metastatic Melanoma, a Deadly Form of Skin Cancer." *Genetech*; 2011. Web.
35. Motzer, R. J., et al. "Phase 3 trial of evorolimus for metastatic renal cell carcinoma." *Cancer*, 116:4256-4265. Print.
36. Bastian M., Heymann S., Jacomy M. "Gephi: an open source software for exploring and manipulating networks." *International AAAI Conference on Weblogs and Social Media* (2009). Print
37. Wilkinson,L.,Friendly, M. "The History of the Cluster Heat Map." *The American Statistician* (2009); 63.2:179-184. Print.
38. The NCBI handbook [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2002 Oct. Chapter 18, The Reference Sequence (RefSeq) Project. Print.
39. Han H., et al. "Identification of differentially expressed genes in pancreatic cancer cells using cDNA microarray." *Cancer Research*. 2002. 62(15):4523. Print.
40. Lopez-Casas P., Lopez-Fernandez L. "Gene-expression profiling in pancreatic cancer." *Expert Review of Molecular Diagnostics*. 2010. 10.5:591-601. Print.