# MOLECULAR AND CLINICAL NETWORK ANALYSIS OF COLORECTAL CANCER

By

Ertugrul Dalkic

## A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

## DOCTOR OF PHILOSOPHY

Cell and Molecular Biology

#### ABSTRACT

## MOLECULAR AND CLINICAL NETWORK ANALYSIS OF COLORECTAL CANCER By

#### Ertugrul Dalkic

Cancer is a large class of diseases and colorectal cancer is one of the leading types of cancer. Systems level analysis of complex diseases like cancer requires the analysis of relationships between different types of clinical data as well as molecular data. Common or specific network features of colorectal cancer together with the other cancer types could be identified by using different network approaches, such as the analysis of clinical data associations, molecular signaling pathways of cancers, and specific interaction networks of cancers. Firstly, a clinical network analysis has been performed on relationships between different types of cancer and the drugs. We generated two cancer networks, one of cancer types that share Food and Drug Administration (FDA) approved drugs, and another of cancer types that share clinical trials of FDA approved drugs. Breast cancer is the only cancer type with significant weighted degree values in both cancer networks. Lung cancer is significantly connected in the FDA approval based cancer network, whereas ovarian cancer and lymphoma are significantly connected in the clinical trial based cancer network. We defined global and local lethality values representing death rates relative to other cancers vs. within a cancer. Correlation and linear regression analyses suggests that global lethality impacts the drug approval and trial numbers, whereas, local lethality impacts the amount of drug sharing in trials and approvals. However, this effect may not apply to pancreatic, liver, and esophagus cancers as the sharing of drugs for these cancers is very low. We also showed a weak overlap between the mutation and drug target based cancer networks. Secondly, we analyzed the cancer pathways in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database, which provides a collective

of signaling pathway members involved in cancer progression. However, the KEGG cancer pathways, unlike signaling pathways, were analyzed extensively with gene expression and mutation data. We transformed the colorectal cancer pathway into subgroups based on their position and analyzed the relative expression levels of adenoma and carcinoma samples as well as the distribution of mutation targets. The gene expression values of the early stage pathway members are significantly higher than the rest of the pathway members in colorectal adenoma tissues. The colorectal cancer pathway shows some degree of coherence in only the carcinoma samples. The correlated gene pairs responsible for the coherence of the colorectal cancer pathway in the carcinoma samples are supported, in part, by the literature and may suggest novel regulatory associations. Thirdly, we compared colorectal cancer samples not only to a control sample set but against a wide variety of samples and conditions, in contrast to current integrative network approaches that identify specific genes by comparing pair-wise control (i.e. normal) to treated (i.e. disease) samples. We were able to identify a distinctly expressed set of genes which were significantly associated with colorectal cancer in the literature unlike the pair-wise approach. We integrated these specific genes with the PPI data to construct a colorectal cancerspecific network. We identified a potential regulatory relationship between glucocorticoid receptor (GR) and ring finger protein 43 (RNF43) which may play a role in colorectal cancer. In HCT116 colorectal cancer cell line, knocking-down GR levels with siRNA resulted in increased RNF43 levels and inducing the colorectal cancer cells with dexamethasone, which is an activating ligand for GR, resulted in decreased RNF43 levels. On the other hand, knocking-down RNF43 levels with siRNA resulted in decreased GR levels. Our study suggests GR might regulate RNF43 negatively, whereas there might not be such a negative regulation from RNF43 to GR.

#### ACKNOWLEDGMENTS

I would like to thank my family; my father Mesut Dalkic, my mother Emine Dalkic, and my brother Mete Dalkic for their support for not only my studies but for my life.

I would like to thank Prof. Christina Chan for supporting me during my PhD. I would like to thank her for being very open to any scientific discussion, organizing research meetings and courses for systems biology.

I would like to thank my committee members; Prof. David Arnosti, Prof. Phillip Duxbury, Assoc. Prof. Shin-Han Shiu, and Assoc. Prof. Michael Feig.

I would like to thank to the current and previous members of Cellular and Molecular Biology Lab at the Department of Chemical Engineering and Materials Science; Li, Hyun-Ju, Linxia, Amanda, Xueruei, Betul, Jason, Phil, Danielle, and all others.

I would like to thank to the current and previous Computational Biology/Systems Biology people of our group; Zheng, Xuewei, Ming, Aritro, Hussein, Daniel, Mohammad Kasim.

I would like to thank to the members of Cell and Molecular Biology Program at MSU.

I would like to thank Prof. Christina Chan for nominating me for Sigma Xi Graduate Student Award and Prof. David Arnosti and others in Sigma Xi for the Graduate Student Award.

## **TABLE OF CONTENTS**

LIST OF TABLES	vii
LIST OF FIGURES	ix
INTRODUCTION	1
BIBLIOGRAPHY	13
CHAPTER 1: CANCER-DRUG ASSOCIATIONS	15
Introduction	15
Results and Discussion	18
FDA cancer drug approvals and clinical cancer drug trials	18
Global and local lethality values for cancer types	19
Effect of lethality on FDA approvals and clinical trials	19
Weighted cancer networks	21
Effect of lethality on the cancer networks	25
Specific and originally approved drugs	26
Clinical and molecular target based cancer networks	27
Conclusion	30
Materials and Methods	32
Drug-cancer pairs	32
Cancer death and survival statistics	33
Network construction	34
Statistical analysis	35
Appendices	38
Appendix A	38
Appendix B	55
Bibliography	76
CHAPTER 2: INTEGRATIVE ANALYSIS OF CANCER PATHWAYS	78
Introduction	78
Materials and Methods	82
Pathway data	82
X/Y scale	83
Expression data	84
Drug and mutation targets	85
Statistical analysis	86
Results and Discussion	88
Colorectal cancer pathway subgroups	88
Coherence of the colorectal cancer pathway	91
Mutation target analysis	94
Conclusion	96

Appendix	
Bibliography	
CHAPTER 3: SPECIFIC INTERACTION NETWORK ANALYSIS FOR	
COLORECTAL CANCER	120
Introduction	
Results and Discussion	
Colorectal cancer specific genes	123
Colorectal cancer specific network	124
GR-RNF43 regulation in colorectal cancer	125
Conclusion	
Materials and Methods	
Transcriptome and Interactome data	127
Determining the colorectal cancer specific gene list	127
Cell culture	128
Quantitative real-time polymerase chain reaction	129
Western blot	129
Dexamethasone and siRNA treatment	130
Appendix	132
Bibliography	142
CONCLUSION	
BIBLIOGRAPHY	

## LIST OF TABLES

Table 1	Clinical trial and FDA approvals of drugs for different cancer types	38
Table 2	FDA approvals, clinical trial, weighted degree values and death statistics of cancers in this study	44
Table 3	Correlation values of weighted degree, approval number values and of FDA specific drug percentage with global and local lethality values	46
Table 4	Cancer pairs with a weight difference of at least 0.5 or lower than 0	47
Table 5	Weighted degree values of drug target and mutation target based networks	48
Table 6	Mutation target- and drug target-based weight values of cancer pairs which have a positive difference between the drug and mutation target-based values	49
Table 7	Cancers with at least one common mutation and drug target	50
Table 8	Drug and cancer-type association with a year tag, based on FDA labels	55
Table 9	Targets of FDA approved cancer drugs	59
Table 10	Mutation targets of different cancer types	62
Table 11	Global lethality ratio values for different cancers from 2001 to 2007	65
Table 12	Local lethality ratio values for different cancers from 2001 to 2007	66
Table 13	Correlation values of weighted degree, approval number values with global lethality values for 2001-2007	67
Table 14	Correlation values of weighted degree, approval number values with local lethality values for 2001-2007	68
Table 15	Weighted degree values for FDA cancer network from 2000 to 2007	69
Table 16	Weighted degree p-values for FDA cancer network from 2000 to 2007	70
Table 17	Clinical trial numbers along with distinct drug number and specific drug number for clinical trials	71
Table 18	ANOVA p values	98
Table 19	Pairwise t-test p values	98

Table 20	Apoptosis and colorectal cancer pathway coherence at correlation coefficient of 0.5	98
Table 21	X/Y group coherence at correlation coefficient of 0.5	99
Table 22	Correlated pairs in the colorectal cancer pathway for normal, adenoma and carcinoma samples	100
Table 23	Mutation frequency values of X/Y groups	102

## LIST OF FIGURES

Figure 1	Cancer drug approval and clinical trial percentages. FDA cancer drug approval and clinical drug trial percentages for 23 cancers	51
Figure 2	Weighted degree values breast and lung cancers in the previous years	52
Figure 3	FDA cancer network weighted degree vs. local lethality ratio	53
Figure 4	FDA specific drug percentages	54
Figure 5	Time dependent characteristics of the FDA approvals and FDA cancer network	72
Figure 6	Cluster dendogram of cancer types based on global and local lethality values	74
Figure 7	The ratio of carcinoma expression ratio to the adenoma expression ratio given for the maximum and minimum two genes and the average	103
Figure 8	X/Y-dependent analysis of colorectal cancer pathway gene expression levels	104
Figure 9	Pathway correlation and cumulative fraction distributions	110
Figure 10	Vend diagrams of correlated pairs of genes in colorectal normal, adenoma and carcinoma samples	114
Figure 11	Literature comparison of differentially expressed genes in colorectal cancer samples	132
Figure 12	Construction of the colorectal cancer specific network	134
Figure 13	Modulation of mRNA levels of GR and RNF43	136
Figure 14	Modulation of protein levels of GR and RNF43	138

#### **INTRODUCTION**

Cancer is a large class of different diseases which result from the deregulation of cell growth [1]. There are more than 200 cancer types, such as lung cancer, breast cancer, colorectal cancer, etc. [2]. Cancer is a major cause of death in the United States. The United States has the highest number of cancer deaths in the world [2]. Approximately 12 million Americans alive in 2008 had a history of cancer, some of whom are currently cancer free. Around one and a half million Americans are expected to be diagnosed with cancer and approximately half a million Americans are expected to die of cancer in 2012 [1]. Cancer is the second leading cause of death in the United States. These statistics support the significance of cancer is one of the leading types of cancers based on the number of new cases and number of expected deaths [1]. However, it does not have as many clinical drug trials and FDA drug approvals as lung cancer, leukemia, etc. This suggests that further clinical studies for colorectal cancer are needed, which should also be accompanied by basic molecular research.

Clinical treatment of cancer includes chemotherapy, radiotherapy, surgery, transplantation, etc. [3]. Chemotherapy is the use of drugs to slow down or prevent the growth of the cancer cells. The goal of chemotherapy might be to cure the cancer by destroying the tumor cells, control the growth of cancer by preventing the spread of the tumor cells, and slow the growth of the cancer, or ease the symptoms (i.e., pain) associated with the growth of the cancer. Classical cancer drugs include i) mitotic poisons or tubulin inhibitors such as Paclitaxel, which disrupt the formation of mitotic spindle and thereby progression of the cell cycle, ii) antimetabolites like Cytarabine, which are nucleic acid base analogues that are incorporated into the DNA to stop DNA synthesis, iii) topoisomerase inhibitors like Epipodophyliotoxins, which are

inhibitors of topoisomerase I or topoisomerase II, that prevent DNA replication, and iv) cell cycle phase-independent agents such as Cisplatin, Doxorubicin, etc., which alkylate the guanine bases to crosslink two strands of DNA, or intercalate between the neighbor bases, or locate in the external grooves of the DNA to disrupt the double helix topology [2]. There may be side-effects of chemotherapy such as nausea, diarrhea, hair loss, fatigue, secondary cancers, etc., because the non-cancer cells might also be damaged [4]. Managing the dosage and the delivery of the cancer drugs is also important aspect of the chemotherapy. There is only a small range between the toxicity and efficacy of most cytotoxic cancer drugs, which necessitates caution in determining the dosage of the chemotherapy [4].

Radiation therapy is the use of high energy radiation, such as X-rays, gamma rays, etc., to damage the DNA and thereby destroy the tumor cells [3]. There also might be side effects to radiotherapy as the normal cells are damaged with this therapy. Furthermore, different types of cancer therapies might be combined; for instance, surgery can be followed by radiotherapy and chemotherapy [3]. A challenge of cancer therapy is to avoid recurrence of the tumor growth [2]. For this reason, tumor markers are monitored to check the progression of the disease.

As mentioned previously, cancer drugs generally target the DNA structure, function, and synthesis to halt or stop cell division. Such drugs are not likely to selectively kill a specific tumor type or even differentiate between normal and tumor cells. Especially in recent years, there has been progress in targeted cancer therapies, which are drugs that target molecular markers of a specific cancer type [3]. For example, the SRC family protein tyrosine kinases are downstream factors of several cell membrane receptors and relay the signal to proliferation-promoting genes. Dasatinib is a direct inhibitor of the SRC family tyrosine kinases and is approved for the treatment of leukemia. Targeted therapies are more promising, as they are more specific to the

tumor cells, and thus are less likely to harm other (i.e. normal) cells. The success of the cancer treatment is dependent on the cause of the cancer, external causes such as smoking or internal causes such as mutation, and the type of the cancer. For example, cancers caused by smoking can be prevented [3]. Also, if a certain mutation is found in a patient, if there is a targeted chemotherapy for that molecule, it would be more likely to be successful. For example, the cancer drug Vemurafenib inhibits the activated mutated form of the BRAF serine/threonine kinase protein in melanoma patients [3]. In recent years, combination chemotherapy for cancer has gained significant attention [2]. When a single protein is targeted by a drug, the cell can use a redundant pathway to continue proliferation. In combination chemotherapy, different drugs can target different molecules, avoiding the redundant pathway activity; as well as reducing toxicity due to the lower doses of the drugs. Another targeted therapy for cancer is vaccine therapy, in which the immune system of the patient is exploited to recognize and destroy the tumor cells [3]. Cancer vaccines contain inactivated cancer cells, tumor antigen-expressing viruses, or overexpressed tumor antigens along with supporting factors for the immune system such as interleukin-2. Cancer vaccines enhance the immune activity of the patient towards the cancer by helping it to recognize the tumor cells.

Since most of the current cancer drugs target general mechanisms such as DNA synthesis, which do not provide specificity and also result in drug resistance, clinical trials for new targeted cancer drugs are of high priority [2]. For this purpose, both basic and translational research are important. While basic cancer research aims to find novel molecular mechanisms involved in the cancer, translational research aims to improve cancer therapy by capitalizing on the basic research findings [2]. Basic research investigates the modified genes in cell lines through functional studies, such as analyzing the upregulated genes with apoptosis assays. On

the other hand, translational research investigates the potential of clinical usage of this target gene, i.e., screening drugs for inhibitors of the upregulated protein product. This is followed by *in vivo* studies in transgenic mouse models and then clinical trials [2].

Clinical trials for novel drugs provide evidence for the efficacy, risks and optimal use of the drugs in patients [5]. Phase 1 clinical trials are performed on small groups of individuals to evaluate the pharmacodynamics, pharmacokinetics, and dose toxicity of the drug [2]. Phase 2 clinical trials are performed to assess the safety and efficacy of the drug. Phase 3 clinical trials are performed on a large group with similar purposes as well as to provide comparisons with previously approved drugs. The US Food and Drug Administration (FDA) regulates the approval and labeling of the drugs by considering the safety, efficacy, and security of the cancer drugs for patient use [6]. Since basic molecular studies are fundamental to the cancer drug development, it is necessary to monitor the relationship between the basic cancer research results with the clinical results. If a common molecular mechanism is found for two different cancer types, this raises the question of whether they also share clinical drug trials or FDA drug approvals. This provides an idea of whether the molecular research findings are reflected in the cancer therapy that is developed.

In recent years the biomedical and clinical sciences have embraced the era of systems level research which is quantitative, integrative, and predictive. Systems science is an interdisciplinary research that focuses on the principles of abstract organization rather than specific properties and produces models to describe the global characteristics [7]. Analyzing part of a system in isolation from the rest cannot provide a complete understanding of its properties and thus cannot readily yield predictions of its behavior. This is significant for biomedical research since the interactions between the genes, proteins, cells, tissues, and organs are crucial

for the proper function of the organism. In other words, biomedical systems cannot be completely understood by analyzing the properties of the molecules in isolation, but rather is better understood by analyzing the relationships between the molecules [7]. For example, the identification of feedback mechanisms is a systems level observation of a biomedical system. Systems medicine investigates the regulatory interactions in the clinical systems such as feedback and feed-forward loops between the patients, clinicians, literature knowledge and data from both basic and translational research [7]. The relationships between different disease types, drugs, and clinical trials are also important just like the relationships between different molecules. Analyzing the links between the diseases provides a system level observation of the relationships between the different diseases [8]. These studies are based on the assumption that two diseases that share a common molecular mechanism, like a mutation target, should have a common origin. This is supported by the finding that similar disease phenotypes, like different cancer types, cluster together when they are linked by mutation targets [9]. These network analyses of disease networks provided new insight into comorbidities, which are secondary diseases that occur concurrently with the primary disease [8]. A patient with a particular disease was shown to be more likely to develop a secondary disease that is a neighbor in the disease network [8]. Such analysis of disease networks provides a systems level approach to understanding disease relationships as compared with the traditional classification of human diseases that neglects this higher level of connectivity of diseases. For example, there might be a mutation in a gene which can cause a primary disease and the interacting neighbor of that specifically mutated gene can lead to secondary disease phenotypes, thereby linking these diseases to each other [8].

Previously, a network level analysis was performed for all the approved drugs and disease therapies [10]. The global topological analysis of the drug and disease associations provided some emergent properties of the FDA approved drug therapies, i.e., most diseases are connected to each other at most through 3 connections; revealing the high compactness of the network as well as the presence of drug hubs that connect many disease therapies to each other [10]. These results can be used in future clinical trials, i.e., the hub drugs may be given priority. My hypothesis is that systems level analysis can improve our understanding and treatment of cancer, which should have many subtypes that are interconnected with each other in a complex way. Therefore I propose an isolated in-depth systems analysis of cancer. The network of cancer types is a smaller-scale network than the disease and drug network [8, 9, 10] but it provides a more focused analysis of the individual members and their interactions in the network. The presence of a common drug therapy between two different cancer types hints to a relationship between these drugs, therefore I analyzed the collection of these relationships at the network level in Chapter 1. The list of cancer drugs in this study included both the drugs that target general cell division mechanisms and those that pursue specific targets, as mentioned before.

Systems medicine investigates the relationships between different types of information at the clinical or basic molecular level, such as the clinical decisions for different diseases vs. death statistics. Only by performing such systematic analyses is it possible to have a more comprehensive clinical therapy that is regulated by clinical and research results. Therefore, I analyzed the potential influences of death statistics on the FDA approved cancer drugs in Chapter 1 and showed the presence of only a few cancers with a significant connection in the FDA approval based cancer network such as breast and lung cancers. I also showed a significant correlation of the death statistics with the drug sharing of different cancers. Furthermore, for a

complete understanding of diseases such as cancer, it is vital to perform such analysis together with molecular biology studies. For example, different cancer types were shown to be closely related to each other as they share some common mutation targets like TP53 [9]. This raises the need for comparing the molecular target information for the various cancers with the drug target information in order to assess for any overlaps or differences. I addressed this issue in Chapter 1 and showed through network analysis that there is a low overlap between the cancer networks based on drug target based associations versus mutation target based associations, suggesting a low impact of molecular biological research findings on clinical decisions for cancer.

Even within the same cancer type, there are several sub-classes of the cancer. Thus it is also important to describe the differences and relationships between these sub-classes. One of the differentiating factors is the severity level of the cancer. For example, colorectal cancer can be classified into colorectal adenoma and colorectal carcinoma, where the latter is differentiated by the invasive properties of the tumor. The KEGG (Kyoto Encyclopedia of Genes and Genomes) database provides cancer pathways which gives an opportunity to analyze the signaling events in different sub-classes of a cancer [11]. The KEGG cancer pathways are different from other pathways such as signaling pathways, metabolic pathways, etc., as they are different signaling pathway members that are integrated into a single pathway. These cancer pathways contain information about the molecular mechanisms on the different stages of a specific cancer like colorectal cancer, pancreatic cancer, glioma, etc., and the different signaling pathways involved in the different stages of the specific cancer.

The molecular events which play a major role in the different stages of colorectal cancer are described in databases such as KEGG; which is based on current literature information. [12, 13, 14, 15]. For example, alteration of the Wnt pathway leads to activation of  $\beta$ -catenin, which in

turn activates its target genes and thereby leads to the progression of colorectal cancer [12, 13]. Because these events are described as earlier molecular mechanisms of colorectal cancer, they take place in the upper portions of the KEGG colorectal cancer pathway. On the other hand, TGF-beta pathway is another important pathway implicated in colorectal cancer. The members of the TGF-beta pathway are located in the lower portions of the pathway image since they play a role in the more advanced stages of colorectal cancer [14, 15]. The actual value of the KEGG cancer pathways is the combination of these different signaling pathways such that a cancer is described as an interconnection of different signaling modules. Module level analysis is an important systems level methodology that aims to find significantly altered modules in a phenotype of interest, shifting towards the analysis of crosstalk between modules recently [16]. Therefore, the analysis of the cancer pathways could provide a systems level identification of the relationships between the different pathway modules in colorectal cancer.

Previously, pathways and protein complexes were studied for the collective behavior of their members in terms of their gene expression levels. Although integrating signaling pathways with genome level expression data was studied widely, it had yet to be realized with cancer pathways. For example, the members of the protein complexes or pathways, such as ribosome, proteasome, and certain metabolic and signaling pathways, have been found to have a strong correlation to each of the members within the pathway with respect to their gene expression levels or cis-element profiles, which is an indicator of similar transcriptional regulation [17, 18]. In another study with a similar goal, the coherence of a pathway was found to have a significant fraction of its members correlated with each other based on their gene expression levels, as compared with a random set of genes of the same size within the pathway of interest [19]. In order to evaluate whether a cancer pathway, which itself is composed of several other pathways,

could be coherent, I performed a similar analysis with the KEGG colorectal cancer pathway in Chapter 2 and found that the colorectal cancer pathway as a whole is coherent.

Since these pathway analyses depend on the current literature information, it is necessary to improve the current information on the molecular events involved in cancer. Network level analysis of protein-protein interactions (PPI) is an important tool for generating specific networks and adding novel regulatory information to our understanding of cancer [20, 21, 22, 23]. Network biology and network medicine are closely related to each other [8, 24]. As network biology investigates the interdependencies between the genes, proteins, metabolites, etc., network medicine relates these dependencies with human diseases. Therefore, the relationships between a collection of genes is as important as the characteristics of a single gene for understanding the function of biological systems and diseases. PPIs are one of the common ways genes and their products regulate or interact with each other, in addition to other mechanisms such as transcriptional interactions. PPI datasets are large-scale collections of interactions between proteins and their topological analysis constituted some of the earliest network biology studies [24]. For example, yeast PPI network was found to have a specific degree distribution, namely scale-free, where the degree of a protein is defined as the number of connections to other proteins in the network. In a scale-free topology very few proteins are hubs, i.e. highly connected, whereas most proteins are lowly connected [24]. Nevertheless, these previous studies are limited to the global statistical observations of the characteristics of the network structure. These global analyses did not provide enough information on the role of the interactions or the network modules in the biological process, phenotype or disease.

PPI datasets are a collection of protein interactions obtained from various sources and conditions and are not condition-specific [25]. However, certain PPIs, excluding permanent

protein complexes, can be condition-specific [26]. Since such condition-specific interaction information is not available on a large-scale to perform network level research, therefore an alternative approach is required to constructed specific interaction networks from the large-scale datasets. Only then can these specific networks be associated with a phenotype such as cancer. Condition-specific datasets such as microarray datasets, which contain global mRNA expression levels of a particular condition, was integrated with the PPIs, genetic interactions, etc. in order to construct condition-specific interaction networks [20, 21]. In these studies, the PPI datasets provide the pairs of proteins which were shown to interact physically at some condition and the differentially expressed genes or differentially correlated gene pairs are calculated from condition-specific microarray datasets. These condition-specific pairs of genes are checked to determine if they interact. If so, they are considered to take place in the specific network. Such networks constructed for cancer metastasis was shown to be represent the hallmarks of cancer like apoptosis, cell growth, metabolism, etc. [20]. These specific networks can be used to generate hypotheses about the involvement of genes and their interactions in the different conditions, such as the involvement of a regulatory interaction between two genes in a particular cancer [21].

A shortcoming of these previous integrative studies that use PPI is that they were based on differential gene expression using a pairwise comparison approach, such as tumor versus normal samples; metastatic versus non-metastatic samples [20, 21]. However, it was recently shown that the pairwise comparison approach is insufficient for detecting the most relevant genes of a condition of interest [27]. This limitation was addressed with a multiple comparison approach, in which a sample set from a condition of interest was compared not only to a single reference set such as normal samples but rather to a collection of samples from many different

conditions; including different normal samples, different cancer types, different human diseases, etc. The diverse condition space more readily uncovers distinctive expression patterns for a particular condition. For example p53, a well-known mutated target in cancer was not differentially expressed based on a pairwise analysis of the gene expression data but through this multiple comparison approach p53 was identified to be an important gene in irradiation [27]. Based on this, I used the multiple comparison approach to identify the top ranked, significantly and differentially expressed genes in colorectal cancer in Chapter 3 and constructed a network of these genes by integrating the gene expression with the human PPI data. I showed that the colorectal cancer specific network included more genes that are significantly related to 'Colorectal cancer' in the literature, than the genes selected based on the pairwise differential expression approach. I also showed that there might be a regulatory mechanism between ring finger protein 43 (RNF43) and glucocorticoid receptor (GR, NR3C1) in this network which is confirmed experimentally. BIBLIOGRAPHY

#### BIBLIOGRAPHY

- American Cancer Society. Cancer Facts & Figures. Atlanta: American Cancer Society. 2012
- 2. Rachel, A. Cancer Chemotherapy: Basic Science to the Clinic. Wiley-Blackwell, Hoboken, NJ, USA, 2009
- 3. http://www.cancer.gov/cancertopics/treatment/types-of-treatment
- 4. Priestman, T. Cancer Chemotherapy in Clinical Practice. Springer, New York, USA, 2012
- 5. http://www.clinicaltrials.gov
- 6. http://www.fda.gov
- 7. Deisboeck TS, Kresh JY. Complex Systems Science in Biomedicine. Springer, New York, USA, 2006
- 8. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011, 12(1):56-68
- 9. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci U S A. 2007, 104(21):8685-90
- 10. Nacher JC, Schwartz JM. A global view of drug-therapy interactions. BMC Pharmacol. 2008, 8:5
- 11. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000, 28, 27-30
- 12. Behrens J. The role of the Wnt signalling pathway in colorectal tumorigenesis. Biochem Soc Trans. 2005, 33, 672-5
- 13. Burgess AW, Faux MC, Layton MJ, Ramsay RG. Wnt signaling and colon tumorigenesis--a view from the periphery. Exp Cell Res. 2011, 317(19):2748-58
- 14. Roman C, Saha D, Beauchamp R. TGF-beta and colorectal carcinogenesis. Microsc Res Tech. 2001, 52, 450-7
- 15. Lampropoulos P, Zizi-Sermpetzoglou A, Rizos S, Kostakis A, Nikiteas N, Papavassiliou AG. TGF-beta signalling in colon carcinogenesis. Cancer Lett. 2012, 314(1):1-7

- 16. Wang X, Dalkic E, Wu M, Chan C. Gene module level analysis: identification to networks and dynamics. Curr Opin Biotechnol. 2008, 19(5):482-91
- 17. Jansen R, Greenbaum D, Gerstein M, Relating whole-genome expression data with protein-protein interactions. Genome Res. 2002, 12, 37-46.
- 18. Hannenhalli S, Levy S. Transcriptional regulation of protein complexes and biological pathways. Mamm Genome. 2003, 14, 611-9
- 19. Yang HH, Hu Y, Buetow KH, Lee MP. A computational approach to measuring coherence of gene expression in pathways. Genomics 2004, 84, 211-7
- 20. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol. 2007, 3:140
- Ahn J, Yoon Y, Park C, Shin E, Park S. Integrative gene network construction for predicting a set of complementary prostate cancer genes. Bioinformatics. 2011, 27(13):1846-53
- 22. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics. 2002, 18 Suppl 1:S233-40
- 23. de Lichtenberg U, Jensen LJ, Brunak S, Bork P. Dynamic complex formation during the yeast cell cycle. Dynamic complex formation during the yeast cell cycle. Science. 2005, 307(5710):724-7
- 24. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004, 5(2):101-13
- 25. Brown KR, Jurisica I. Online predicted human interaction database. Bioinformatics. 2005, 21(9):2076-82
- 26. Levy ED, Pereira-Leal JB. Evolution and dynamics of protein interactions and networks. Curr Opin Struct Biol. 2008, 18(3):349-57
- 27. Wu M, Liu L, Chan C. Identification of novel targets for breast cancer by exploring gene switches on a genome scale. BMC Genomics. 2011, 12:547

#### **CHAPTER 1**

#### **CANCER-DRUG ASSOCIATIONS**

#### **INTRODUCTION**

Cancer is a complex disease, with many subtypes, affecting various tissues in diverse ways, thus giving rise to an abundance of chemotherapies. Taken together, cancers are the second leading cause of death in the United States [1]. The common features of cancer include uncontrolled cell growth, reduction in apoptosis, and loss of cell cycle regulation, while other features are more tissue specific and thus differentiate them and their chemotherapies.

In a global network level analysis of different diseases, where the vertices represented diseases and the edges represented connections between diseases that share common genetic background, most diseases were less connected, while a limited number of diseases, mostly cancers, were highly connected hubs [2]. Similarly, a network analysis of drugs, where the vertices represented drugs and the edges represented connections between drugs that share common protein targets, showed that drugs of similar types clustered together, and most proteins were targeted by a few drugs, whereas only a few proteins were targeted by many drugs [3, 4]. Cancers have fewer drugs that are used to treat them as compared with the other diseases, and the targets for the cancer drugs are at a shorter distance from the genes that are mutated in the cancers [3]. Quantitative analysis of the drug targets showed that proteins with at least 3 proteinprotein interactions are more likely to be targeted by drugs [5]. A recent network study characterized the global map of many diseases, including cancers, and their associations with drugs, where the vertices represented diseases and the edges represented connections between diseases that share common drugs [6]. This study was also concerned with the global description of the network, and found that only a few diseases are highly connected by drugs, while most

diseases are less connected; and most diseases, even those unrelated to each other, are connected by a few links [6]. These studies constitute the global topological analysis aspect of the emerging areas of network medicine [7] and network pharmacology [8]. However, these studies do not focus on the specific relationships between diseases and drugs, to address questions, such as, how might these relationships arise, or what factors may affect these relationships.

The field of medical sciences includes both basic molecular and clinical research, the latter involves clinical trials. Clinical trials apply biomedical protocols to humans that aim to intervene or observe a disease, e.g., testing drugs on cancers (http://clinicaltrials.gov). Clinical trials provide preliminary evidence of the efficacy, risks and optimum usage of the drugs. Phase 1 and 2 clinical trials are performed on small groups of individuals to evaluate their safety and efficiency. Phase 3 clinical trials are performed on a large group of individuals, to evaluate their efficiency, side effects and how they compare with approved drugs. Phase 4 clinical trials are performed after the drug has been approved for use, to obtain additional information. The United States Food and Drug Administration (FDA) regulates the approval and labeling of the drugs with regard to their safety, efficacy, and security to humans (http://www.fda.gov). In addition to the clinical drug trial and FDA approval data, death statistics, such as the estimated cases and estimated deaths over the years are available for the different cancer types [9]. Cancer is a large class of disease with various types, each with its own specific approvals, trials, death statistics, and molecular information, i.e., mutation targets. These diverse data provide opportunities to perform an integrative, systems level analysis of the cancers to reveal potential relationships between the various types of cancer and the drugs used to treat them and possible trends or factors that influence these relationships.

Global network analyses have been previously applied to describe the overall topology of disease and drug relationships, i. e, very few diseases and drugs are highly connected, while most members of these networks are less connected [2, 3, 4, 6]. Smaller network systems, such as in this study, are amendable to a more focused analysis of individual members of the network, whereas larger networks are not, and hence are more amendable to statistical topological analyses, such as degree distribution analysis [10]. We propose that a drug approved or used in clinical trials for treating several cancers may hint to a relationship between those cancers. Similarly, a mutation involved in or a drug target used in treating different cancers may suggest a relationship between these cancers. System level analysis of these relationships could reveal potential factors involved in the development of these complex relationships that are not readily apparent from the data itself. In contrast to the previous medical network analyses, the analysis of smaller networks of cancer-drug and cancer-target associations permits a more detailed evaluation of the specific relationships between individual cancers. Through correlation and linear regression analyses of the number of approvals and trials, and weighted degree values, with the cancer lethality values, we assessed whether the death statistics impact the formation of associations between the cancers and drugs. Our analyses suggest that global lethality has an effect on the number of FDA approved and clinical trial cancer drugs. Comparative analysis of the cancer networks based on the FDA approved drugs and clinical trial drugs showed that some cancers are significantly and highly connected in the clinical trial cancer network but not in the FDA cancer network, and vice versa. Correlation and linear regression analyses suggest that local and global lethality differentially impact the sharing of FDA approved cancer drugs and the sharing of clinical trial drugs. Further, a comparison of the mutation target-based with the FDA

drug target-based cancer networks suggests that the molecular information about a cancer does not strongly influence the cancer drug approvals.

#### **RESULTS AND DISCUSSION**

#### FDA cancer drug approvals and clinical cancer drug trials

We collected the drugs approved through 2009 by the FDA for 23 cancer types and the clinical trials completed by 2009 for these same drugs (see Appendix B). We compared these 81 drugs for the 23 cancer types, and checked which drugs had i) completed Phase 1 and 2 trials but were not listed under Phase 3 clinical trials and thus were not FDA approved, ii) completed Phase 3 clinical trial but were not FDA approved, iii) were FDA approved and in Phase 3 clinical trial (Table 1), and iv) were FDA approved and were not in clinical trials. There are several drugs for which Phase 3 clinical trial was completed but were not FDA approved (item ii). For example, cisplatin was approved for only testicular and bladder cancers, and has undergone and completed Phase 3 clinical trials for many types of cancer but has yet to be listed as approved by the FDA for those cancers (Table 1). The clinical trial data is incomplete (see Materials and Methods section for details). For example, there are some drugs which were FDA approved but not listed under any past clinical trials, completed or otherwise, which suggests that the analysis of the clinical trials will not be comprehensive. Leukemia, breast cancer, lung cancer, and lymphoma have the highest number of drug approvals and the highest number of clinical trials (Table 2). The percentage of clinical trials or FDA approvals for the different cancers were calculated as the number of clinical drug trials or FDA drug approvals for a specific cancer type, divided by the total number of clinical drug trials or FDA drug approvals for the 23 cancers analyzed in this study. The clinical trial and FDA approval percentages are similar for many of

the cancers in this study (Figure 1). There are a few notable exceptions, namely breast cancer and myeloma, which have much higher percentages of FDA approvals than of clinical trials.

#### Global and local lethality values for cancer types

Death and survival ratios have been predominantly used to describe the values of global and local significance of cancer deaths [9]. It is confusing to use these values since one uses death and the other uses survival numbers to describe global and local death statistics of a specific cancer. Therefore, we defined two different death-based statistics, a global and a local lethality rate by using the estimated death and new case numbers of each cancer (Table 2, see Appendix B). The percentage of global lethality is calculated as the ratio of estimated number of deaths for a cancer to the estimated number of deaths for all cancers. The percentage of local lethality is calculated as the ratio of estimated number of deaths to the estimated number of cases for a particular cancer. The global lethality provides a perspective of a particular cancer with respect to the other cancers, whereas, the local lethality is specific to each cancer type. A cancer with a high local lethality suggests that it has a high number of deaths within its own incidences, while its global lethality may or may not be high. For example, pancreatic cancer is a locally lethal but not globally lethal cancer; it has a local lethality value of 0.91 but a global lethality value of 0.06 (Table 2). This is because most of the pancreatic cancer patients have low survival rates, but comparatively there are fewer cases of pancreatic cancer.

#### Effect of lethality on FDA approvals and clinical trials

We hypothesize that there are factors, such as the lethality values of a cancer, that may influence the number of clinical trials and, in turn, FDA approvals. To quantitatively evaluate whether lethality values are related to the number of FDA drug approvals and clinical drug trials, Spearman correlation coefficients were calculated between the global/local lethality measures

and the trial/approval numbers. The correlation analyses suggest that global lethality is correlated, whereas local lethality is not correlated, to both the clinical trial and FDA approval numbers (Table 3). To further evaluate the impact of lethality values on the FDA drug approvals and clinical drug trials, we performed a linear regression analysis. Linear fit of the clinical trial numbers with global lethality suggests a slight but albeit significant relationship ( $r^2 = 0.25$ , p =0.03). This suggests the higher clinical drug trial numbers could be explained, in part, by the higher global lethality rates. Next, we considered whether the relationships found by correlation and linear regression analyses are affected by lung cancer, the most globally lethal cancer, and pancreatic, esophagus, and liver cancers, the most locally lethal cancers (see Table 2 and the Materials and Methods section). We re-calculated the correlations by removing the globally or locally lethal cancers. No significant change in the correlations resulted upon removing lung cancer. However, a linear fit of the FDA approval numbers with global lethality suggests a slight relationship which is significant, when lung cancer is excluded ( $r^2 = 0.20$ , p = 0.05). The significance of the correlation and the linear fit between local lethality with FDA approval and clinical trial numbers increased upon removing the most locally lethal cancers, pancreatic, liver and esophagus cancers (Table 3). Local lethality has a significant correlation with clinical trial drug numbers for the cancers other than the most locally lethal ones. This suggests the number of FDA approvals and clinical trials are much lower for pancreatic, liver and esophagus cancers as compared to other cancers despite their very high local lethality. Linear fits suggest that the values of the lethal cancers affect the linear relationship, therefore we estimated the values based on the equation with the other cancers. Based on the linear regression of the local lethality ratio with the FDA cancer network weighted degree values (y = 0.96x + 0.40), pancreatic, liver and

esophagus cancers should have been expected to have a weighted degree of around 1.2; however, they have a weighted degree of around 0.07-0.5.

Although, the linear fit p-values of local lethality with FDA approval numbers and clinical trial numbers decreased, when pancreatic, liver and esophagus cancers are excluded, they are not very significant. We also analyzed whether the FDA approval numbers from previous years correlated with the lethality values. The correlation of global lethality with the FDA approval numbers has mostly been present in previous years (see Appendix B). The correlation and linear regression analyses suggest that global lethality has an impact on the drug trial and approval numbers, for the cancers in this study.

#### Weighted cancer networks

The global relationships between drugs and diseases have been analyzed topologically in large-scale networks of drugs and diseases [2, 3, 4, 6]. Complex relationships between the types of cancer and drugs constitute a smaller network structure. Unlike the larger networks, a smaller network system, as in this study, are amendable to a more focused analysis of individual members of the network rather than statistical topology-based parameters [10]. We applied this more focused analysis, where individual members and interactions in the networks were studied rather than their global structure, to elucidate the drug therapy based relationships between various cancers and the factors that may influence these relationships.

The collection of cancer-drug pairs make up a bipartite network, which we transformed into a unipartite weighted network consisting of only cancers. To construct a weighted network of cancers, an edge between any two cancers was assigned, if there is at least one drug which was approved by FDA to treat both types of cancers. The weight of an edge was defined by the Jaccard index, which is the fraction of drugs which were approved for both cancers over all the

drugs which were approved for each of the two cancers, separately (see Materials and Methods). Weighted degree values were not significantly correlated with the number of FDA approvals (Pearson correlation coefficient of 0.34, p = 0.11), suggesting that the number of drugs approved for a cancer does not implicate the number of drugs shared with other cancers. We, further, assessed the significance of the weighted degree values by a permutation test, while keeping the number of drugs per cancer type constant, and found the degree of drug sharing is not significant for most of the cancers (Table 2), except for lung and breast cancer These two cancers have significant weighted degree values in the FDA cancer network. Lung cancer shares FDA drugs with many other cancers. Leukemia, the cancer type with the highest number of FDA approvals, does not have a significant weighted degree value in the FDA cancer network (Table 2). This is because leukemia does not share many of its FDA approved drugs with other cancers. Indeed, as discussed later, leukemia has many specific drugs (see section "Drugs specific to particular cancer types"). We also analyzed the FDA cancer network over time, by including the cancer drug approvals for the different years.

Using the date at which a drug was approved by the FDA, we analyzed the cancer networks at earlier time points. Based on the first cancer drug, mechlorethamine, approved in 1949, leukemia, lymphoma and lung cancer can be connected in a FDA cancer network. These 3 cancers for a long time were the only cancers for which a drug was approved by the FDA. This initial network began to grow as drugs for other cancers were approved by the FDA, joining the network from 1986 onwards. We defined the average weight as the average of the weights of all the edges of a network (0 weights are not included). The average weight of the FDA cancer networks was calculated to capture the changes in how the drugs are shared between the different cancers over the years (see Appendix B). The average weight of the FDA cancer network was at

the lowest in 1997, suggesting that the sharing of drugs between cancer types was not very high in 1997 as compared to the other years. Sharing of drugs between cancer types increased between 1998 and 2004. We also determined the component numbers of the FDA cancer networks. The network consisted of a single component except between 1991-1998 when the network contained 2 components (see Appendix B). There have been only two cases where a drug led to a jumping growth as new drugs were approved for cancers. Pamidronate was approved for breast cancer and myeloma in 1991, making a connection between the two cancers, separate from the other cancers. Paclitaxel was approved for skin cancer in 1997, making a connection to ovarian cancer, causing another jump in the growth of the network. Some FDA approvals could be terminated later, however, this analysis is only concerned with the drugs which were not terminated by 2009 (the time of completion of the clinical trials is not available for the majority of the trials, therefore a time dependent study of the clinical trial-based cancer networks was not performed). Based on the average weight values of the networks, there is no major change over the years. Weighted degree values for most of the cancers also are not significant in the previous years' networks. However the breast cancer weighted degree value has been significant since 2000 and the lung cancer degree value has become significant recently. Weighted degree values of lung and breast cancer have been increasing and significantly higher than the other cancers since 2006 (Figure 2). In recent years, FDA approved drugs for these cancers (the 1<sup>st</sup> and 3<sup>rd</sup> most globally lethal) have a high overlap with other cancers.

A weighted clinical trial-based cancer network was also constructed (herein denoted as clinical trial cancer network), where two cancers were connected if there is at least one FDA approved drug (approved for at least a cancer) in the clinical trial data for both cancers. The clinical trial cancer network is almost a complete network, because of the large number of drugs

that were used in clinical trials for the different cancers, thereby connecting many of the cancers, albeit not all, to each other. The significance of the weighted degree values was evaluated by a permutation test, with the number of drug trials kept constant. Breast cancer, ovarian cancer, and lymphoma have significant weighted degree values in the clinical trial cancer network (Table 2). Also, the weighted degree values of lung cancer and head and neck cancer are close to being significant. This indicates that these cancers shares clinical trial drugs, significantly, with other cancers. In addition, we calculated the difference in the edge weights between the FDA and clinical trial cancer networks for each cancer pair, and identified that most pairs are strongly connected in the clinical trial but not in the FDA cancer network. For example, stomach and esophagus cancers are strongly connected in the clinical trial cancer network (Table 4). There are many drugs used in clinical trials for both types of cancers, i.e., capecitabine, cisplatin, doxorubicin, erlotinib, fluorouracil, irinotecan, ixabepilone, leucovorin, oxaliplatin, paclitaxel, and vinorelbine, and thus strongly connecting these two cancers. However, they are not connected in the FDA cancer network, i.e. no drug is approved by the FDA for both stomach and esophagus cancers; porfimer was approved for esophagus cancer while docetaxel, fluorouracil, imatinib, and sunitinib were approved for stomach cancer. There are a few pairs of cancers which are more highly connected in the FDA cancer network than in the clinical trial cancer network (Table 4). For example, sarcoma and endometrial cancer pair has a weight of 0.5; they share methotrexate which is the only drug approved for endometrial cancer and one of the two drugs approved for sarcoma. On the other hand, there are many drugs in the clinical trial data for each of these two cancers which are not shared between them, such as altretamine, capecitabine, etoposide, etc. Weighted networks of cancers based on FDA approvals and clinical trials show different characteristics. Breast cancer is the only cancer with a significant degree value in both

the FDA and the clinical trial cancer networks. While lung cancer is more significantly connected only in the FDA cancer network, ovarian cancer and lymphoma are more significantly connected in the clinical trial cancer network (Table 2). This suggests that ovarian cancer and lymphoma have a high overlap of drugs in clinical trials but not in FDA approvals.

#### Effect of lethality on the cancer networks

Given that the lethality of a cancer impacts the number of drug trials and approvals, it raises the question of whether it could also influence the FDA and clinical trial cancer networks and if there could be differences in their influence on these two networks. We analyzed the correlation and the linear fit between the weighted degree values of the FDA/clinical trial cancer networks and the global/local lethality values. The weighted degree values for the clinical trial cancer network are correlated with local lethality (Table 3). Linear regression between the weighted degree values and the lethality values shows a partial but significant relationship between local lethality and clinical trial network weighted degree ( $r^2 = 0.26$ , p = 0.02). This suggests that sharing of drugs in clinical trials is impacted positively by local lethality values.

The weighted degree values of the FDA cancer network are not significantly correlated with the global and local lethality values (Table 3). Next, we analyzed the effect of the most globally lethal (lung cancer) and the most locally lethal cancers (pancreatic, esophagus and liver cancers) on these correlations and linear fits. Weighted degree values of the FDA cancer network are significantly correlated with local lethality after removing pancreatic, liver, and esophagus cancers (Table 3, Figure 3A-3B). Linear fit analysis suggests that the weighted degree of a cancer in the FDA cancer network tend to be high if its local lethality value is high. However, the most locally lethal cancers (pancreatic, esophagus and liver cancers) are excluded from this effect since they have lower than expected weighted degree values, as compared to the other

cancers. We also analyzed if the FDA cancer networks from previous years correlated with the lethality values. Global lethality and local lethality do not have a significant correlation in the older FDA cancer networks. However, more recently (2007) the cancer network has become correlated with local lethality, with the exclusion of pancreatic, liver, and esophagus cancers (see Appendix B).

Analysis of the weighted degree values of the cancer networks provides information on the level of drug sharing between cancers. We showed that local lethality has an effect on the clinical cancer drug trial sharing as well as FDA approved drug sharing, the latter appears to be a recent trend. However, the most locally lethal cancers, pancreatic, liver, and esophagus cancers, are biased towards having lower levels of sharing of FDA approved drug. For the most local lethality cancers, although sharing of drugs in clinical trials correlates positively with local lethality values, the sharing of the approved drugs does not correlate with local lethality values.

#### Specific and originally approved drugs to particular cancer types

Network analysis captured the overlap in cancer drug use, however, only 26 of the total 81 cancer drugs were approved for more than one cancer type. Therefore we analyzed the distribution of the remaining 55 drugs which were approved specifically for only one type of cancer. A drug which was approved by the FDA solely for a single cancer is denoted as a "specific" FDA drug. We calculated the specific drug percentage for a cancer as the ratio of the number of specific drugs to the total number of drugs approved by the FDA. Prostate cancer, leukemia, breast cancer, and lymphoma have the highest specific drug percentage approved by the FDA (Figure 4). The most locally lethal cancers, pancreatic, liver, and esophagus cancers, have no specific drugs (Table 2). Globally lethal cancer, i.e., lung cancer, has a low percentage of FDA specific drugs (Table 2, Figure 4). The number of specific drugs in clinical trials is very

low, therefore it was not analyzed further (see Appendix B). We also analyzed the possible effect of lethality on the percentage of FDA specific drug approvals and showed that there is no significant effect (Table 3).

There is also a notable difference among the non-specific (shared) drugs, such that some of the drugs were first approved for a cancer type and then approved for other cancer types, while other drugs might be approved for more than one type of cancer at the same time. We defined whether a drug was "originally approved" by the FDA for a specific cancer type and then approved for other cancers after at least a year. Colorectal cancer has the highest number of "originally approved" FDA drugs (Table 2). There is only one originally approved FDA drug, erlotinib for lung cancer (Table 2, see Appendix B). Many more drugs were approved for other cancers that were subsequently approved for lung cancer (11 drugs) than were "originally approved" for lung cancer (only one).

#### Comparison of clinical and molecular target based cancer networks

In addition to the death statistics, we asked whether molecular information impacted the cancer-drug associations. To compare the molecular target-based relationships to the clinical target-based relationships for the different cancer types, we constructed weighted molecular and clinical cancer networks based on mutation targets and FDA approved drug targets respectively. The edges between two cancers in the mutation target based network was assigned if there is at least one mutation target associated with both cancers and the edges between two cancers in the drug target based network was assigned if there is at least one drug target associated with both cancers. The weights of the edges were defined by the Jaccard index. To compare the mutation target-based and the drug target-based cancer networks, we included only the cancers that have both mutation and drug target data. We calculated the weighted degree values for the different
cancers and evaluated the significance of the weighted degrees with permutation test, keeping the distribution of target numbers for each cancer constant (Table 5). The weighted degree values of the mutation and drug-target based cancer networks are not strongly correlated (Pearson correlation coefficient of 0.37, p = 0.11). Lung and breast cancers have significant and high weighted degree values in the drug target-based network but not in the mutation target-based network (Table 5). On the other hand, colorectal, ovarian and brain cancers have significant weighted degree values in the mutation target-based network but not in the drug-target based network (Table 5). Leukemia is the only cancer which has significant weighted degree values in both networks.

The overlap between the two networks is very low and for the overlapping edges, we calculated the difference in mutation and drug target weight values (Table 6). Colorectal-ovarian, ovarian-endometrial, and endometrial-colorectal cancer pairs have higher mutation target-based weights than drug target-based weight values (Table 6). These cancers are connected to each other in the mutation target-based network through the following mutations: PMS1, PMS2, MLH1, MSH2, and MSH6, which are proteins responsible for DNA mismatch repair. On the other hand, all three cancers share no drug targets. Since they share many mutation targets, this suggests that they could have similar molecular mechanisms, and thus raises the question if they should share drug targets. On the other hand, kidney and liver cancers, which do not share any mutation targets, have a high overlap of drug targets. They share drug targets such as FLT4, PDGFRB, BRAF, etc. There could be mutation targets common to these cancers which may not have been identified or is absent in the current dataset. Alternatively, they could share molecular mechanisms without sharing mutation targets, i.e., similar pathways may be affected in both cancers despite different mutated genes.

We also evaluated the cancers that are associated with proteins that are both mutation and drug targets (Table 7). Only leukemia, lung, and breast cancers have mutation targets that are also drug targets. For example, ERBB2, a member of EGFR family, has long been known as a mutation target for breast cancer [11]. Lapatinib, letrozole, and trastuzumab are drugs that target ERBB2 in our data and all have been used in clinical drug trials for only breast cancer and approved by the FDA for only breast cancer. Furthermore, ERBB1, a member of the EGFR family, is known as a mutation target for lung cancer [12]. There are several drugs which target ERBB1, such as cetuximab, erlotinib, gefitinib, lapatinib, panitumumab, and trastuzumab, among which, only erlotinib and gefitinib are approved only for lung cancer (see Apendix B). The remaining drugs have not completed Phase 3 clinical trials. Cetuximab, and trastuzumab have completed Phase 1 and 2 trials, whereas clinical trials using lapatinib and panitumumab for lung cancer have not yet completed Phase 1 and 2 trials (Table 1). Overall, very few mutations have been approved as targets for cancer therapy.

Comparison of mutation and drug-target based cancer networks indicate that the overlap is very low. Various cancers have strong associations in one but not in the other network. For instance, lung and breast cancers have significant drug-target based associations but not mutation-target based associations. Similarly there are pairs of cancers, such as the pair of colorectal and endometrial cancers, with relatively high weights in the mutation-target based network but not in the drug-target based network. This analysis suggests that the influence of molecular information on the cancer-associations is not strong, and there are very few proteins which are both mutation-targets and drug-targets.

#### CONCLUSION

In this study, we present a systems level view of the cancer drugs. Comparing clinical trial and FDA approval based cancer networks, we showed that only breast cancer is significantly connected in both networks. Lung cancer is significantly connected in the FDA cancer network, whereas ovarian cancer and lymphoma are significantly connected in the clinical trial cancer network. This suggests that lung cancer has a high degree of sharing of FDA approved drugs with the other cancers. Indeed, it has the highest number of FDA approved drugs which are shared with other cancers. In contrast, ovarian cancer and lymphoma have a high degree of drug sharing in clinical trials but not in FDA approvals.

We also assessed whether death statistics and molecular information are related to the cancer-drug associations. We showed that the cancer-drug associations are differentially impacted by the type of lethality. Global lethality appears to have an effect on the number of FDA approved drugs and clinical drug trials, but not on the FDA approval and clinical trial-based drug sharing, as determined by the cancer network weighted degree values. On the other hand, local lethality has an effect on the FDA approval and clinical trial-based drug sharing, but not on the number of FDA approved drugs and clinical drug trials. The effect of local lethality on the sharing of FDA approved drugs is not present or captured by the most locally lethal cancers, pancreatic, liver and esophagus cancers. These cancers are biased towards having very low overlap of FDA approved drugs with other cancers. For example, there is only one drug approved for liver cancer, sorafenib, which is shared with lung cancer; however there are 13 more FDA approved drugs for lung cancer, which are not approved for liver cancer, leading to the lower weight for liver cancer. Although sharing of drugs in clinical trials correlates positively with local lethality values, however, it does not translate to increase sharing of the approved

drugs for the most locally lethal cancers. There could be a number of reasons for this; the drugs in clinical trial are not being approved for the most locally lethal cancers or they have not been approved yet. For example, liver cancer and lung cancer share 13 drugs out of total 32 drugs used in clinical trials for these cancers. 5 of these 15 common/overlapping clinical trial drugs are approved for lung cancer by FDA but they are still in clinical trials for liver cancer. Therefore they have a higher connection weight in the clinical trial cancer network than the FDA cancer network. These findings support network-based analysis and their ability to reveal relevant information distinct from the raw data. It is not surprising that clinical decisions may be impacted by death statistics. However, it is interesting that different types of death statistics (global lethality vs. local lethality) show different results. It should be kept in mind that this study does not capture all aspects of the clinical drug data. For example, this analysis does not account for the differential efficiencies of the various drugs used in treating a particular cancer, which could have an impact on why some cancers have few while others may have many more drugs that target it. The current analysis of the clinical trials is limited to those which have already been approved by the FDA for at least one cancer type and therefore do not include all cancer drugs currently in clinical trial.

Currently, most cancer drugs are designed to target the general mechanisms of cell division, which may not directly address the specific molecular mechanisms that drive the development of the type of cancer it aims to treat. We compared mutation and drug targets for various cancer types. We identified a number of differences and noted that some cancer types share mutation targets but not drug targets while others share drug targets but not mutation targets, thereby hinting at the possibility that new drug targets or mutation targets could be identified for these cancers. Nevertheless, there are many other factors to consider when

evaluating the data. Although two cancer types may not have the same mutation targets, they may have the same genes that are differentially expressed, which could suggest the involvement of similar molecular mechanisms. Given that cancer treatment includes surgery, radiotherapy in addition to chemotherapy (<u>http://www.cancer.gov/cancertopics/treatment/types-of-treatment</u>) thus, this study provides a systems level analysis of the trends of one aspect of clinical cancer research, namely from the perspective of the drugs that are FDA approved or undergoing clinical trials.

In closing, we demonstrated a systems level view of the drugs that have been approved and how they have been shared between cancer types. Thus we envision that this study could be informative to medical researchers from both the basic and clinical sciences alike. The trends revealed in this study could be monitored in the following years for any changes and these analyses could be used to guide more in-depth analysis of potential targets that could be involved in future clinical cancer drug trials and approvals. For example, one could followed whether the FDA approved drug sharing continues to be significant for breast and lung cancers which appears to be recent trends, beginning in the 2000s, and whether the overlap between the molecular target based and the drug target based cancer networks increases.

#### MATERIALS AND METHODS

#### **Drug-cancer pairs**

We obtained lists of cancer drugs from the National Cancer Institute Drug Information Summaries (<u>http://www.cancer.gov/cancertopics/druginfo/alphalist</u>), and the FDA Center for Drug Evaluation and Research (<u>http://www.fda.gov/Drugs/default.htm</u>). We used the indication information by 2009 from the drug labels from the Drugs@FDA database

(http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm?fuseaction=Search\_Search\_Dr ug Name) to generate a list of drug-cancer associations that included 23 types of cancer (see Appendix B). We renamed some cancers, for example, Kaposi's sarcoma is listed under skin cancer, glioma is listed under brain cancer, and different types of leukemia and lymphoma are listed more generally as leukemia and lymphoma, respectively. The time information tag of the FDA approved label files is also used. Drugs discontinued in the market were excluded. We obtained clinical trial information for all drug trials completed by 2009 from the Clinical Trials database (http://clinicaltrials.gov) and collected the clinical trials for the drugs and the cancer types that are in the list of FDA data. We differentiated between Phase <sup>1</sup>/<sub>2</sub> and Phase 3 trials since the Phase  $\frac{1}{2}$  are initial trials on small groups of patients, whereas Phase 3 trials are performed on large groups of patients. We excluded Phase 4 trials since they are post-approval. We did not include neoplasms in our analysis. Names of drugs and cancers have been organized according to the FDA data. In addition, we only collected the trials which were listed as drug trials. These limitations could lead to loss of information, such that we have FDA approval information for some drugs without completed clinical trial information (Table 1). We observed that these limitations could affect the clinical trial information prior to 2000s, namely, there could be cases in which there is an approval of a drug earlier than the trial dates. Therefore, we did not perform a time analysis of the clinical trials.

#### Cancer death and survival statistics

The cancer statistics for 2001-2008 of the estimated number of new cases and the estimated number of deaths for the different types of cancers were obtained from the American Cancer Society [9]. We defined two kinds of lethality values. Global lethality is defined as the ratio of deaths of a particular cancer over all cancers. Local lethality is defined as the ratio of

deaths of a particular cancer over the cases of that particular cancer. For breast, ovarian, cervical cancers only the female population values were considered. Likewise, for prostate and testicular cancers only the male population values were used. For the other cancer types, both the male and female population values were included. Lung cancer has the highest global lethality value, whereas pancreatic cancer has the highest local lethality value. To determine which other cancers are similar to lung and pancreatic cancers with respect to their global and local lethality values, we performed hierarchical clustering, based on Euclidean distance of lethality values with single linkage. Lung cancer clustered by itself, and pancreatic, liver and esophagus cancers clustered together (see Appendix B). Therefore only lung cancer is considered globally lethal cancer, whereas pancreatic, liver and esophagus cancers are considered locally lethal cancers.

#### **Network construction**

We constructed weighted clinical networks of cancer types, FDA cancer network and clinical trial cancer network, from the drug-cancer pairs. In the clinical cancer networks an edge was defined between two cancer types when there is at least one drug which was approved or used in clinical trials for both types of cancer. The weight of the edge was defined by the Jaccard index, which is the fraction of common drugs for both cancer types over all the drugs for each of the cancer types. For example, there is only one drug which was approved for both pancreatic and stomach cancers, fluorouracil, whereas there are 2 more drugs, erlotinib and gemcitabine, which were approved for pancreatic cancer but not for stomach cancer, and there are 3 more drugs, docetaxel, imatinib, and sunitinib, which were approved for stomach cancer but not for pancreatic cancer (see Appendix B). Therefore the weight of the edge between these two cancers is 1/(1+2+3)=0.17 (Table S3). The resulting FDA drug approval-based cancer network (herein denoted as FDA cancer network) contains 23 types of cancer (vertices or nodes) with 70

interactions (edges). We defined the weighted degree value for a cancer as the sum of the weights of the edges for that cancer. For example, pancreatic cancer shares drugs with stomach, lung, colorectal and breast cancers, therefore its weighted degree is the sum of the weights of the edges with these cancers, which is 0.17+0.13+0.1+0.1=0.5 (Table 1). This parameter provides an account of the allocation of drugs for a particular cancer and its neighbors in the network. If more drugs, which are approved for other cancers, are approved for pancreatic cancer (regardless of whether the drug is shared with stomach, lung, colorectal and breast cancers or other cancers) in the future, its weighted degree value will increase. Its weighted degree value will decrease if more drugs are approved for stomach, lung, colorectal and breast cancers but not for pancreatic cancer.

Similarly, we also constructed molecular target and clinical target-based cancer networks, using mutation target data from the Cancer Gene Census database (<u>http://www.sanger.ac.uk/genetics/CGP/Census/</u>) [13] and FDA approved drug target data from the DrugBank database (<u>http://www.drugbank.ca</u>) (Dataset S4), respectively. Mutation target data from the Cancer Gene Census database used was updated in January 2009 and includes mutation targets which have been implicated in the cancer. This database was chosen because it is based on literature curation and thus captures information on the molecular mechanisms that clinical researchers should have information on. Cytoscape version 2.4 was used to visualize the networks [14].

#### **Statistical analysis**

The significance of the weighted degree values in the cancer networks was analyzed by permutation tests. The distribution of the number of drugs or drug and mutation targets was kept constant while the cancer-drug or cancer-target associations were randomized, respectively. The

p-value for the weighted degree of a cancer type is calculated as the fraction of the randomly generated networks with a weighted degree value for a particular cancer which is equal to or greater than the actual weighted degree value of that particular cancer (Table 2 and 5). Conventional cutoff of 0.05 was used as a significance threshold. No multiple test correction has been applied to the p-values. Therefore, given the number of statistical tests performed, some of the associations reported, particularly borderline significant, could be spurious. In the FDA cancer network, Wilcoxon test is used to determine if the weighted degree values of breast and lung cancer are higher than the rest of the cancers in the network.

Shapiro-Wilk test suggests that some of the datasets used in this study are not normally distributed. If p-values obtained from this test are lower than 0.05, the null hypothesis that data are normally distributed is rejected. p-values for the 23 cancers and the cancers minus head and neck cancer, mesothelioma and sarcoma are, respectively, 4.775e-05 and 0.0003658 for FDA approval numbers, 0.002419 and 0.02321 for specific FDA drug percentage values, 0.001606 and 0.003714 for clinical trial numbers, 0.1582 and 0.3365 for FDA cancer network weighted degree values, 0.07316 and 0.0341 for clinical trial cancer network weighted degree values. p-value is 1.075e-06 for the global lethality values, and 0.07739 for the local lethality values. Therefore, we used Spearman correlation coefficient values for the analysis of the relationships between lethality values and the clinical trial and approval numbers, and the network weight values (Table 3). The significance of the correlations was determined by a permutation based algorithm [15].

APPENDICES

### APPENDIX A

**Table 1** Clinical trial and FDA approvals of drugs for different cancer types

Cancer type	Phase 3 trial only	Phase 1, 2 only	Phase 3 and FDA	FDA only
bladder cancer	doxorubicin, gemcitabine, paclitaxel	carboplatin, ifosfamide, bortezomib, trastuzumab, ixabepilone	cisplatin	-
brain cancer	procarbazine, cisplatin, ifosfamide, carboplatin, thalidomide, etoposide	cladribine, irinotecan, bortezomib, gefitinib, lenalidomide, busulfan, erlotinib, oxaliplatin, imatinib, temsirolimus, ixabepilone, topotecan, methotrexate, lapatinib, capecitabine, sorafenib	carmustine, temozolomide, cyclophosphamide	-
breast cancer	carboplatin, carmustine, cisplatin, doxorubicin, zoledronate, vinorelbine	decitabine, busulfan, etoposide, fludarabine, leucovorin, temozolomide, gefitinib, oxaliplatin, pemetrexed, dasatinib, irinotecan, bortezomib, erlotinib, imatinib, vorinostat, alemtuzumab	anastrozole, capecitabine, docetaxel, fulvestrant, gemcitabine, lapatinib, paclitaxel, trastuzumab, cyclophosphamide	epirubicin, pamidronate, toremifene, fluorouracil

cervical cancer	cisplatin	gemcitabine, capecitabine, paclitaxel, fluorouracil, oxaliplatin, arsenic trioxide, erlotinib, docetaxel, gefitinib	topotecan	-
colorectal cancer	cisplatin, doxorubicin	gefitinib, erlotinib, gemcitabine, trastuzumab, carmustine, ixabepilone, imatinib	bevacizumab, capecitabine, cetuximab, irinotecan, oxaliplatin, fluorouracil, leucovorin	panitumumab
endometrial cancer	cisplatin, doxorubicin, paclitaxel	capecitabine, topotecan, pemetrexed, oxaliplatin, raloxifene, thalidomide	-	methotrexate
esophagus cancer	cisplatin, fluorouracil	oxaliplatin, capecitabine, carboplatin, paclitaxel, irinotecan, topotecan, decitabine, vinorelbine, doxorubicin, docetaxel, erlotinib, arsenic trioxide, leucovorin, ixabepilone	-	porfimer
eye cancer	-	busulfan, carboplatin, topotecan	-	cyclophosphamide

head and neck cancer	cisplatin, fluorouracil, paclitaxel	irinotecan, oxaliplatin, bevacizumab, erlotinib, azacitidine, capecitabine, carboplatin, temozolomide, ixabepilone, doxorubicin, topotecan, carmustine, cyclophosphamide, etoposide, porfimer, thalidomide, gefitinib, bortezomib, sorafenib, gemcitabine	cetuximab, docetaxel	-
kidney cancer	carboplatin,cyclophosph amide,doxorubicin, etoposide	paclitaxel, irinotecan, oxaliplatin, temozolomide, busulfan, fludarabine, bevacizumab, cetuximab, erlotinib, fluorouracil, pentostatin, topotecan, arsenic trioxide, gefitinib, methotrexate, capecitabine, gemcitabine, lenalidomide, imatinib, denileukin diftitox, thalidomide	-	sunitinib
leukemia	cytarabine, etoposide, leucovorin, doxorubicin, ifosfamide	topotecan, bexarotene, carboplatin, bortezomib, temozolomide, rituximab, bevacizumab, sorafenib, thalidomide, denileukin diftitox, docetaxel, ixabepilone, temsirolimus	alemtuzumab, busulfan, daunorubicin, fludarabine, idarubicin, imatinib, mitoxantrone, methotrexate, cyclophosphamide	mechlorethamine, nilotinib, teniposide, bendamustine

liver cancer	cisplatin, fluorouracil, doxorubicin	irinotecan, oxaliplatin, temozolomide, erlotinib, gemcitabine, pemetrexed, capecitabine, carboplatin, topotecan, thalidomide, docetaxel, epirubicin	-	-
lung cancer	cisplatin,carboplatin	sunitinib, busulfan, cyclophosphamide, fludarabine, cetuximab, decitabine, imatinib, irinotecan, doxorubicin, bortezomib, ifosfamide, sorafenib, fluorouracil, azacitidine, trastuzumab, temozolomide, thalidomide, temsirolimus	docetaxel, erlotinib, etoposide, gefitinib, gemcitabine, paclitaxel, pemetrexed	mechlorethamine, nofetumomab, porfimer, methotrexate
lymphoma	etoposide, doxorubicin, ifosfamide, leucovorin, cisplatin, idarubicin, mitoxantrone, daunorubicin	topotecan, paclitaxel, irinotecan, oxaliplatin, busulfan, imatinib, temozolomide, fludarabine, cladribine, decitabine, alemtuzumab, arsenic trioxide, altretamine, gemcitabine, carboplatin, azacitidine, pentostatin, thalidomide, ixabepilone, temsirolimus, tretinoin, bevacizumab, sorafenib	carmustine, cytarabine, rituximab, methotrexate, cyclophosphamide	bexarotene, methoxsalen, procarbazine, vorinostat, bendamustine
mesothelioma	cisplatin	decitabine, doxorubicin, gemcitabine, epirubicin, gefitinib, bevacizumab	pemetrexed	-

myeloma	-	arsenic trioxide, fludarabine, etoposide, cisplatin, clofarabine	bortezomib, thalidomide, cyclophosphamide	carmustine, doxorubicin, lenalidomide, zoledronate
ovarian cancer	epirubicin, mitoxantrone	vinorelbine, temozolomide, docetaxel, ixabepilone, cisplatin, capecitabine, etoposide, ifosfamide, gemcitabine, bortezomib, lapatinib, erlotinib, imatinib, gefitinib, anastrozole, letrozole, pemetrexed, oxaliplatin, alemtuzumab, leucovorin, methotrexate, irinotecan, sorafenib, toremifene, bevacizumab, cetuximab, vorinostat	carboplatin, paclitaxel, cyclophosphamide	altretamine
pancreatic cancer	oxaliplatin	lapatinib, irinotecan, bevacizumab, cetuximab, cisplatin, pemetrexed, imatinib, trastuzumab, capecitabine, leucovorin, paclitaxel, docetaxel, ixabepilone, bortezomib, arsenic trioxide, temsirolimus, cyclophosphamide	erlotinib, gemcitabine, fluorouracil	-
prostate cancer	mitoxantrone, zoledronate, toremifene	doxorubicin, paclitaxel, carboplatin, epirubicin, temsirolimus, ixabepilone, pemetrexed, oxaliplatin, sunitinib, azacitidine, imatinib, trastuzumab, arsenic trioxide, bevacizumab, thalidomide, lapatinib	leuprolide, degarelix	-

sarcoma	cisplatin, doxorubicin, etoposide, ifosfamide, daunorubicin, cyclophosphamide, topotecan	irinotecan, oxaliplatin, temozolomide, busulfan, erlotinib, carboplatin, altretamine, leucovorin, paclitaxel, thalidomide, gemcitabine, trastuzumab, ixabepilone, bevacizumab, cytarabine	methotrexate	-
skin cancer	cisplatin	lenalidomide, decitabine, irinotecan, oxaliplatin, busulfan, cyclophosphamide, etoposide, fludarabine, docetaxel, leucovorin, sorafenib, temozolomide, thalidomide, denileukin diftitox, tretinoin, carmustine, temsirolimus, bortezomib, ixabepilone	-	daunorubicin, doxorubicin, imiquimod
stomach cancer	-	irinotecan, cisplatin, gemcitabine, vinorelbine, doxorubicin, paclitaxel, leucovorin, oxaliplatin, ixabepilone, erlotinib, capecitabine	-	imatinib, sunitinib
testicular cancer	carboplatin, cyclophosphamide, paclitaxel	busulfan, fludarabine, temozolomide, topotecan, ixabepilone, alemtuzumab, arsenic trioxide, imatinib	etoposide, ifosfamide, cisplatin	-

Cancer type	FDA drug approval number	FDA specific drug approval number	FDA specific drug percentage (%)	FDA original drug approval number	Clinical drug trial number	FDA w.d.	FDA w.d.p.	Clinical trial w.d.	Clinical trial w.d.p	Global lethality ratio	Local lethality ratio
lung cancer	14	3	21.4	1	121	1.32	0.029	7.58	0.066	0.286	0.753
colorectal cancer	8	4	50.0	3	61	0.46	0.840	6.15	0.899	0.088	0.336
breast cancer	19	10	52.6	2	97	1.17	0.003	6.81	0.037	0.072	0.222
pancreatic cancer	3	0	0.0	1	35	0.5	0.898	6.96	0.576	0.061	0.910
prostate cancer	4	3	75.0	0	48	0.41	0.937	4.32	0.998	0.051	0.154
leukemia	23	16	69.6	0	170	0.65	0.157	5.16	0.656	0.038	0.490
lymphoma	16	9	56.3	0	121	0.83	0.195	6.41	0.013	0.036	0.276
liver cancer	1	0	0.0	0	19	0.33	0.899	6.57	0.714	0.033	0.862
endometrial cancer	1	0	0.0	0	10	1.06	0.653	4.43	0.996	0.028	0.186
ovarian cancer	6	2	33.3	2	71	1.16	0.488	7.42	0.018	0.027	0.717
esophagus cancer	1	0	0.0	1	26	0.07	0.970	6.8	0.605	0.025	0.867
bladder cancer	2	1	50.0	0	10	0.25	0.967	4.34	0.997	0.025	0.205
brain cancer	3	1	33.3	1	62	0.89	0.748	6.57	0.606	0.023	0.599
kidney cancer	3	1	33.3	1	30	0.5	0.899	7.22	0.221	0.023	0.239
skin cancer	4	1	25.0	2	31	0.48	0.885	5.84	0.924	0.020	0.165
myeloma	8	3	37.5	1	13	0.86	0.563	2.81	1.000	0.019	0.537

**Table 2** FDA approval numbers, clinical trial numbers, weighted degree values (w.d.), weighted degree p-values (w.d.p.) and death statistics (global and local lethality ratio) of cancers in this study

stomach	4	0	0.0	1	19	1.13	0.609	6.32	0.759	0.019	0.506
cancer											
cervical	1	0	0.0	0	20	0.24	0.939	5.48	0.932	0.007	0.350
cancer											
testicular	3	1	33.3	2	11	0.31	0.966	5.31	0.984	0.001	0.047
cancer											
eye cancer	1	0	0.0	0	2	0.78	0.691	1.58	1.000	0.000	0.100
head and neck	3	0	0.0	0	45	1.35	0.558	8.03	0.056	-	-
cancer											
mesothelioma	1	0	0.0	0	40	0.07	0.984	3.23	1.000	-	-
sarcoma	2	0	0.0	1	7	1.21	0.636	7.39	0.259	-	-

**Table 3** Correlation values of weighted degree, approval number values and of FDA specific

 drug percentage with global and local lethality values

	Global lethality		Local lethality		
	All cancer types	All cancer types except globally lethal cancers (lung cancer)	All cancer types	All cancer types except locally lethal cancers (pancreatic, esophagus and liver cancers)	
FDA approval number	0.50	0.44	0.05	0.42	Spearman statistic
	0.03	0.06	0.85	0.09	Spearman p-value
Clinical trial	0.67	0.63	0.34	0.53	Spearman statistic
number	0.00	0.00	0.15	0.03	Spearman p-value
FDA cancer network	0.25	0.12	0.14	0.53	Spearman statistic
weighted degree	0.29	0.62	0.57	0.03	Spearman p-value
Clinical trial cancer network	0.42	0.33	0.61	0.55	Spearman statistic
weighted degree	0.06	0.17	0.00	0.03	Spearman p-value
FDA specific	0.35	0.44	-0.32	-0.05	Spearman statistic
drug percentage	0.13	0.06	0.17	0.85	Spearman p-value

Cancer type 1	Cancer type 2	Clinical	FDA	Difference
		trial	cancer	
		cancer	network	
		network	weight	
		weight		
stomach cancer	esophagus cancer	0.71	0.00	0.71
head and neck	kidney cancer	0.56	0.00	0.56
cancer				
kidney cancer	lung cancer	0.54	0.00	0.54
ovarian cancer	head and neck	0.54	0.00	0.54
	cancer			
leukemia	lymphoma	0.68	0.15	0.53
ovarian cancer	breast cancer	0.61	0.09	0.53
cervical cancer	esophagus cancer	0.50	0.00	0.50
head and neck	brain cancer	0.50	0.00	0.50
cancer				
head and neck	liver cancer	0.50	0.00	0.50
cancer				
stomach cancer	cervical cancer	0.50	0.00	0.50
brain cancer	myeloma	0.17	0.22	-0.05
ovarian cancer	myeloma	0.10	0.17	-0.06
head and neck	endometrial	0.25	0.33	-0.08
cancer	cancer			
ovarian cancer	eye cancer	0.06	0.17	-0.11
eye cancer	myeloma	0.00	0.13	-0.13
brain cancer	eye cancer	0.12	0.33	-0.21
sarcoma	endometrial	0.22	0.50	-0.28
	cancer			

Table 4 Cancer pairs with a weight difference of at least 0.5 or lower than 0

	Drug	Drug	Mutation	Mutation
	target	target	target	target
	based	based	based	based
	network	network	network	network
	weighted	weighted	weighted	weighted
	degree	degree	degree	degree p-
	value	p-value	value	value
leukemia	2.28	0.000	0.25	0.003
lung cancer	3.35	0.000	1.12	0.210
breast cancer	3.29	0.001	1.08	0.160
colorectal cancer	2.75	0.290	1.57	0.000
ovarian cancer	2.53	0.634	1.43	0.011
brain cancer	1.66	0.674	0.9	0.016
sarcoma	1.76	0.961	0.42	0.509
pancreatic cancer	0.62	1.000	0.82	0.706
endometrial	0.66	0.996	0.78	0.732
cancer				
eye cancer	1.79	0.309	0.17	0.880
stomach cancer	2.12	0.774	0.53	0.909
lymphoma	2.17	0.287	0.19	0.952
testicular cancer	1.83	0.817	0.33	0.980
skin cancer	2.22	0.808	0.35	0.981
bladder cancer	2.32	0.330	0.14	0.998
head and neck	2.75	0.275	0.06	1.000
cancer				
kidney cancer	1.38	0.996	0.13	1.000
liver cancer	1	0.999	0.26	1.000
myeloma	1.24	0.999	0.16	1.000
prostate cancer	0.88	1.000	0.19	1.000

**Table 5.** Weighted degree values of drug target and mutation target based networks

**Table 6** Mutation target- and drug target-based weight values of cancer pairs which have a positive difference between the drug and mutation target-based values

Cancer type 1	Cancer type 2	Difference of mutation
		target from
		drug target
		based
1		weight
colorectal cancer	endometrial	0.20
coloractal cancor		0.24
colorectal cancer	ovarian cancer	0.24
cancer	Ovarian Cancer	0.15
brain cancer	colorectal cancer	0.14
ovarian cancer	pancreatic cancer	0.14
ovarian cancer	liver concer	0.14
brain cancer		0.1
brain cancer	salcollia	0.09
brain cancer	endometrial	0.08
breast cancer	stomach cancer	0.08
ondometrial	stomach cancer	0.08
endometrial	stomach cancer	0.08
liver concor	noncreatic concer	0.08
	tasticular concer	0.08
brain concer	lung concer	0.08
		0.07
baad and naak	ballereatic cancer	0.06
neau anu neck	Kiulley calleel	0.00
liver concer	overion concor	0.06
hrein een een		0.00
oralli cancer	prostate cancer	0.05
endometrial	prostate cancer	0.05
	lung concer	0.05
	iung cancer	0.05
lung cancer	stomach cancer	0.05
brain cancer	breast cancer	0.03
brain cancer	liver cancer	0.03
brain cancer	pancreatic cancer	0.03
brain cancer	stomach cancer	0.03
breast cancer	eye cancer	0.03
breast cancer	kidney cancer	0.03
colorectal cancer	sarcoma	0.03
pancreatic cancer	skin cancer	0.03
bladder cancer	sarcoma	0.02
brain cancer	kidney cancer	0.02
breast cancer	sarcoma	0.02
eye cancer	sarcoma	0.02
prostate cancer	sarcoma	0.01

 Table 7 Cancers with at least one common mutation and drug target

Cancer	<b>Common mutation</b>
type	and drug target name
	and Entrez Gene ID
lung	ERBB1(1956)
cancer	
breast	ERBB2(2064)
cancer	
leukemia	FCGR2B(2213),
	ABL1(25),
	PDGFRB(5159),
	KIT(3815), ABL2(27),
	LCK(3932),
	BCL2(596)



Figure 1 Cancer drug approval and clinical trial percentages. FDA cancer drug approval and clinical drug trial percentages for 23 cancers

**Figure 2** Weighted degree values breast and lung cancers in the previous years. Weighted degree values of breast cancer, lung cancer, and the remaining cancers in the FDA cancer networks from 2000 to 2008. Average and the standard deviation of the weighted degree values are shown. Wilcoxon test was performed for greater values of lung and breast cancer than the other cancers. The networks with p-values lower than 0.05 are indicated by asterisk (\*)



**Figure 3** FDA cancer network weighted degree vs. local lethality ratio. FDA cancer network weighted degree values are plotted against local lethality ratio for (A) 23 cancers ( $r^2 = 0.01$ , p = 0.78), (B) the cancers except pancreatic, liver and esophagus cancers ( $r^2 = 0.35$ , p = 0.01). Lung cancer is shown as an open triangle and pancreatic, liver, esophagus cancers are shown as open circles.







**Figure 4** FDA specific drug percentages. FDA and clinical trial specific drug percentages for the cancers except cervical, endometrial, esophagus, liver, pancreatic, eye, sarcoma, mesothelioma, and stomach cancers, which do not have specific drugs



### APPENDIX B

Drug name	Other names of the drug	Cancer type	Label date
Alemtuzumab	Campath	leukemia	2001
Altretamine	Hexalen	ovarian cancer	1990
Anastrozole	Arimidex	breast cancer	1995
Arsenic trioxide	Trisenox	leukemia	2000
Azacitidine	Vidaza	leukemia	2004
Bendamustine	Treanda, Bendamustine	leukemia	2008
	hydrochloride		
Bendamustine	Treanda, Bendamustine	lymphoma	2008
	hydrochloride		
Bevacizumab	Avastin	colorectal cancer	2004
Bevacizumab	Avastin	lung cancer	2006
Bevacizumab	Avastin	breast cancer	2008
Bexarotene	Targretin	lymphoma	1999
Bortezomib	Velcade	myeloma	2003
Bortezomib	Velcade	lymphoma	2006
Busulfan	Myleran, Busulfex	leukemia	1954
Capecitabine	Xeloda	breast cancer	1998
Capecitabine	Xeloda	colorectal cancer	2001
Carboplatin	Paraplatin	ovarian cancer	2003
Carmustine	Gliadel, Bicnu	brain cancer	1996
Carmustine	Gliadel, Bicnu	lymphoma	1997
Carmustine	Gliadel, Bicnu	myeloma	1997
Cetuximab	Erbitux	colorectal cancer	2004
Cetuximab	Erbitux	head and neck	2006
		cancer	
Cisplatin	Platinol, Platinol-AQ, Cisplatin	testicular cancer	1978
Cisplatin	Platinol, Platinol-AQ, Cisplatin	bladder cancer	1993
Cladribine	Leustatin	leukemia	1993
Clofarabine	Clolar	leukemia	2004
Cyclophosphamide	Lyophilized cytoxan	brain cancer	2000
Cyclophosphamide	Lyophilized cytoxan	breast cancer	2000
Cyclophosphamide	Lyophilized cytoxan	eye cancer	2000
Cyclophosphamide	Lyophilized cytoxan	leukemia	2000
Cyclophosphamide	Lyophilized cytoxan	lymphoma	2000
Cyclophosphamide	Lyophilized cytoxan	myeloma	2000
Cyclophosphamide	Lyophilized cytoxan	ovarian cancer	2000
Cytarabine	DepoCyt, Cytosar-U	lymphoma	1999
Dasatinib	Sprycel	leukemia	2006
Daunorubicin	Daunorubicin citrate,	skin cancer	1996
	DaunoXome, Daunorubicin		
	hydrochloride, Cerubidine		

**Table 8** Drug and cancer-type association with a year tag, based on FDA labels

Daunorubicin	Daunorubicin citrate,	leukemia	1998
	DaunoXome, Daunorubicin		
	hydrochloride, Cerubidine		
Decitabine	Dacogen	leukemia	2006
Degarelix		prostate cancer	2008
Denileukin diftitox	Ontak	lymphoma	1999
Docetaxel	Taxotere	breast cancer	1996
Docetaxel	Taxotere	lung cancer	1999
Docetaxel	Taxotere	prostate cancer	2004
Docetaxel	Taxotere	head and neck	2006
		cancer	
Docetaxel	Taxotere	stomach cancer	2006
Doxorubicin	Doxil, Doxorubicin	skin cancer	1995
	hydrochloride		
Doxorubicin	Doxil, Doxorubicin	ovarian cancer	1999
	hydrochloride		
Doxorubicin	Doxil, Doxorubicin	myeloma	2007
	hydrochloride		
Epirubicin	Ellence, Epirubicin	breast cancer	1999
	hydrochloride		
Erlotinib	Tarceva, Erlotinib	lung cancer	2004
	hydrochloride		
Erlotinib	Tarceva, Erlotinib	pancreatic cancer	2005
	hydrochloride		
Estramustine	Emcyt, Estramustine phosphate	prostate cancer	1981
	sodium		
Etoposide	Etopophos, Etoposide	testicular cancer	1983
	phosphate, Vepesid		
Etoposide	Etopophos, Etoposide	lung cancer	1986
	phosphate, Vepesid		
Exemestane	Aromasin	breast cancer	1999
Fludarabine	Fludara, Fludarabine phosphate	leukemia	1991
Fluorouracil	Fluoroplex	colorectal cancer	1962
Fluorouracil	Fluoroplex	breast cancer	1998
Fluorouracil	Fluoroplex	pancreatic cancer	1998
Fluorouracil	Fluoroplex	stomach cancer	1998
Fulvestrant	Faslodex	breast cancer	2002
Gefitinib	Iressa	lung cancer	2003
Gemcitabine	Gemzar, Gemcitabine	pancreatic cancer	1996
	hydrochloride		
Gemcitabine	Gemzar, Gemcitabine	lung cancer	1998
	hydrochloride		
Gemcitabine	Gemzar, Gemcitabine	breast cancer	2004
	hydrochloride		
Gemtuzumab	Mylotarg, Gemtuzumab	leukemia	2000
	ozogamicin		

Ibritumomab	Zevalin	lymphoma	2002
tiuxetan			
Idarubicin	Idamycin PFS, Idarubicin	leukemia	1997
	hydrochloride, Idarubicin		
	hydrochloride PFS,		
Ifosfamide	Ifex, Ifex/Mesnex kit,	testicular cancer	1988
	Ifex/Mesna kit, Mesna		
Imatinib	Gleevec, Imatinib mesylate	leukemia	2003
Imatinib	Gleevec, Imatinib mesylate	stomach cancer	2003
Imatinib	Gleevec, Imatinib mesylate	sarcoma	2006
Imiquimod	Aldara	skin cancer	2004
Irinotecan	Camptosar, Irinotecan	colorectal cancer	1996
	hydrochloride		
Ixabepilone	Ixempra, Ixempra kit	breast cancer	2007
Lapatinib	Tykerb, Lapatinib ditosylate	breast cancer	2007
Lenalidomide	Revlimid	myeloma	2006
Letrozole	Femara	breast cancer	1997
Leucovorin	Leucovorin Calcium	colorectal cancer	1952
Leuprolide	Eligard, Leuprolide acetate,	prostate cancer	1998
1	Lupron, Lupron depot, Lupron	1	
	depot-3, Lupron depot-4,		
	Lupron depot-ped, Viadur		
Mechlorethamine	Mustargen, Mechlorethamine	leukemia	1949
	hydrochloride		
Mechlorethamine	Mustargen, Mechlorethamine	lung cancer	1949
	hydrochloride		
Mechlorethamine	Mustargen, Mechlorethamine	lymphoma	1949
	hydrochloride		
Methotrexate	Methotrexate LPF,	sarcoma	1988
	Methotrexate sodium, Trexall		
Methotrexate	Methotrexate LPF,	breast cancer	2003
	Methotrexate sodium, Trexall		
Methotrexate	Methotrexate LPF,	endometrial cancer	2003
	Methotrexate sodium, Trexall		
Methotrexate	Methotrexate LPF,	head and neck	2003
	Methotrexate sodium, Trexall	cancer	
Methotrexate	Methotrexate LPF,	leukemia	2003
	Methotrexate sodium, Trexall		
Methotrexate	Methotrexate LPF,	lung cancer	2003
	Methotrexate sodium, Trexall		
Methotrexate	Methotrexate LPF,	lymphoma	2003
	Methotrexate sodium, Trexall		
Methoxsalen	Uvadex	lymphoma	1999
Mitoxantrone	Novantrone, Mitoxantrone	leukemia	1987
	hydrochloride		
Nelarabine	Arranon	leukemia	2005
Nelarabine	Arranon	lymphoma	2005

Nilotinib	Tasigna, Nilotinib	leukemia	2007
	hydrochloride monohydrate		
Nofetumomab	Verluma	lung cancer	1996
Oxaliplatin	Eloxatin	colorectal cancer	2005
Paclitaxel	Abraxane, Taxol	ovarian cancer	1992
Paclitaxel	Abraxane, Taxol	skin cancer	1997
Paclitaxel	Abraxane, Taxol	breast cancer	1998
Paclitaxel	Abraxane, Taxol	lung cancer	1998
Pamidronate	Aredia, Pamidronate disodium	breast cancer	1991
Pamidronate	Aredia, Pamidronate disodium	myeloma	1991
Panitumumab	Vectibix	colorectal cancer	2006
Pemetrexed	Alimta, Pemetrexed disodium	lung cancer	2004
Pemetrexed	Alimta, Pemetrexed disodium	mesothelioma	2004
Pentostatin	Nipent	leukemia	1991
Porfimer	Photofrin, Porfimer sodium	esophagus cancer	1995
Porfimer	Photofrin, Porfimer sodium	lung cancer	1998
Procarbazine	Matulane, Procarbazine	lymphoma	1969
	hydrochloride		
Raloxifene	Evista, Raloxifene	breast cancer	2007
	hydrochloride		
Rituximab	Rituxan	lymphoma	1997
Sorafenib	Nexavar, Sorafenib tosylate	kidney cancer	2005
Sorafenib	Nexavar, Sorafenib tosylate	liver cancer	2007
Sunitinib	Nexavar, Sorafenib tosylate	stomach cancer	2006
Sunitinib	Nexavar, Sorafenib tosylate	kidney cancer	2007
Temozolomide	Sutent, Sunitinib malate	brain cancer	1999
Temsirolimus	Torisel	kidney cancer	2007
Teniposide	Temodar	leukemia	1992
Thalidomide	Thalomid	myeloma	2006
Topotecan	Hycamtin, Topotecan	ovarian cancer	1996
	hydrochloride		
Topotecan	Hycamtin, Topotecan	lung cancer	1998
	hydrochloride		
Topotecan	Hycamtin, Topotecan	cervical cancer	2006
	hydrochloride		
Toremifene	Fareston, Toremifene citrate	breast cancer	1997
Tositumomab	Bexxar, Tositumomab and	lymphoma	2003
	iodine I 131 tositumomab		
Trastuzumab	Herceptin	breast cancer	1998
Tretinoin	Vesanoid	leukemia	2004
Valrubicin	Valstar	bladder cancer	1998
Vinorelbine	Navelbine, Vinorelbine tartrate	lung cancer	1994
Vorinostat	Zolinza	lymphoma	2006
Zoledronate	Zometa, Zoledronic acid	myeloma	2002

Table 9 Targets of FDA approved cancer drugs

Drug	Target
Alemtuzumab	CD52(1043), FCGR1A(2209), FCGR3B(2215), C1R(715),
	C1QA(712), C1QB(713), C1QC(714), FCGR3A(2214),
	FCGR2A(2212), FCGR2B(2213), FCGR2C(9103)
Altretamine	DNA
Anastrozole	CYP19A1(1588)
Arsenic trioxide	ATP2C1(27032), AKT1(207), IL6(3569), MAPK1(5594),
	ERGIC2(51290), CCND1(595), ABCB1(5243), JUN(3725),
	MAPK3(5595), IKBKB(3551)
Azacitidine	DNA, DNMT1(1786)
Bevacizumab	VEGF(7422), FCGR1A(2209), FCGR3B(2215), C1R(715),
	C1QA(712), C1QB(713), C1QC(714), FCGR3A(2214),
	FCGR2A(2212), FCGR2B(2213), FCGR2C(9103)
Bexarotene	RXRB(6257)
Bortezomib	PSMD1(5707), PSMD2(5708), PSMB1(5689), PSMB5(5693),
	PSMB2(5690)
Busulfan	DNA
Capecitabine	TYMS(7298), DPYD(1806)
Carboplatin	DNA, ALB(213)
Carmustine	DNA, GSR(2936)
Cetuximab	FCGR1A(2209), EGFR(1956), FCGR3B(2215), C1S(716),
	C1R(715), C1QA(712), C1QB(713), C1QC(714),
	FCGR3A(2214), FCGR2A(2212), FCGR2B(2213),
	FCGR2C(9103)
Cisplatin	DNA
Cladribine	NP(4860), DNA, ADA(100)
Clofarabine	RRM1(6240), POLA1(5422)
Cyclophosphamide	DNA
Cytarabine	POLB(5423), ALB(213)
Dasatinib	ABL1(25), PDGFRB(5159), KIT(3815), ABL2(27), SRC(6714),
	FYN(2534), YES1(7525), EPHA2(1969), LCK(3932),
	STAT5B(6777)
Daunorubicin	ABCC1(4363), DNA, ABCB1(5243)
Decitabine	DNA, DNMT1(1786)
Degarelix	-
Denileukin diftitox	IL2RB(3560), IL2RA(3559), IL2RG(3561)
Docetaxel	TUBB1(81027), BCL2(596)
Doxorubicin	TOP2A(7153), DNA
Epirubicin	CHD1(1105), TOP2A(7153)
Erlotinib	EGFR(1956)
Estramustine	ESR1(2099), ESR2(2100), ABCB1(5243), MAP2(4133),
	MAP1A(4130)
Etoposide	TOP2A(7153), MAP2K7(5609)
Exemestane	CYP19A1(1588)
Fludarabine	BCL2(596), RRM1(6240), POLA1(5422), STAT1(6772).
	ADA(100)

Fluorouracil	TYMS(7298), DPYD(1806)
Fulvestrant	ESR1(2099)
Gefitinib	EGFR(1956)
Gemcitabine	TYMS(7298), RRM1(6240), CMPK1(51727)
Gemtuzumab	CD33(945), FCGR1A(2209), FCGR3B(2215), C1S(716),
	C1R(715), C1QA(712), C1QB(713), C1QC(714),
	FCGR3A(2214), FCGR2A(2212), FCGR2B(2213),
	FCGR2C(9103)
Ibritumomab	MS4A1(931), FCGR1A(2209), FCGR3B(2215), C1S(716),
tiuxetan	C1R(715), C1QA(712), C1QB(713), C1QC(714),
	FCGR3A(2214), FCGR2A(2212), FCGR2B(2213),
	FCGR2C(9103)
Idarubicin	TOP2A(7153)
Ifosfamide	DNMT1(1786)
Imatinib	ABL1(25), PDGFRB(5159), KIT(3815), PDGFRA(5156),
	CSF1R(1436), ABCB1(5243), NTRK1(4914), ABCG2(9429),
	RET(5979), DDR1(780)
Imiquimod	TLR7(51284)
Irinotecan	TOP1MT(116447), TOP1(7150)
Ixabepilone	-
Lapatinib	ERBB2(2064), EGFR(1956)
Lenalidomide	PTGS2(5743)
Letrozole	ESR1(2099), ERBB2(2064), CYP19A1(1588)
Leucovorin	TYMS(7298)
Leuprolide	GNRHR(2798)
Mechlorethamine	DNA
Methotrexate	DHFR(1719), ALB(213)
Methoxsalen	DNA, CYP2A6(1548)
Mitoxantrone	TOP2A(7153), ABCG2(9429)
Nelarabine	DNA
Nilotinib	-
Nofetumomab	-
Oxaliplatin	DNA
Paclitaxel	TUBB1(81027), BCL2(596)
Pamidronate	FDPS(2224)
Panitumumab	EGFR(1956)
Pemetrexed	TYMS(7298), DHFR(1719), GART(2618)
Pentostatin	ADA(100)
Porfimer	LDLR(3949), FCGR1A(2209)
Procarbazine	DNA
Raloxifene	ESR1(2099), ESR2(2100)
Rituximab	MS4A1(931), FCGR1A(2209), FCGR3B(2215), C1S(716),
	C1R(715), C1QA(712), C1QB(713), C1QC(714),
	FCGR3A(2214), FCGR2A(2212), FCGR2B(2213),
	FCGR2C(9103)

Sorafenib	FLT4(2324), RAF1(5894), FLT3(2322), PDGFRB(5159),
	KDR(3791), KIT(3815), BRAF(673)
Sunitinib	FLT4(2324), FLT1(2321), FLT3(2322), PDGFRB(5159),
	KDR(3791), KIT(3815), PDGFRA(5156), CSF1R(1436),
	RET(5979),
Temozolomide	DNA
Temsirolimus	FRAP1(2475)
Teniposide	TOP2A(7153)
Thalidomide	PTGS2(5743), TNF(7124), NFKB1(4790)
Topotecan	TOP1MT(116447), TOP1(7150), ABCG2(9429)
Toremifene	ESR1(2099)
Tositumomab	MS4A1(931), FCGR1A(2209), FCGR3B(2215), C1R(715),
	C1QA(712), C1QB(713), C1QC(714), FCGR3A(2214),
	FCGR2A(2212), FCGR2B(2213), FCGR2C(9103)
Trastuzumab	ERBB2(2064), EGFR(1956), FCGR1A(2209), FCGR3B(2215),
	C1S(716), C1R(715), C1QA(712), C1QB(713), C1QC(714),
	FCGR3A(2214), FCGR2A(2212), FCGR2B(2213),
	FCGR2C(9103)
Tretinoin	NR0B1(190), RXRB(6257), RARG(5916), ALDH1A2(8854),
	ALDH1A1(216), RARRES1(5918), GPRC5A(9052),
	RXRG(6258)
Valrubicin	TOP2A(7153)
Vinorelbine	TUBB(203068)
Vorinostat	HDAC8(55869), HDAC1(3065), HDAC2(3066), HDAC3(8841),
	HDAC6(10013)
Zoledronate	FDPS(2224)

Cancer	Mutation targets
type	
Bladder	FGFR3(2261) HRAS(3265)
cancer	
Brain	ALK(238) APC(324) ATM(472) BRAF(673) COPEB(1316)
cancer	EGER(1956) GOPC(57120) HRAS(3265) IDH1(3417)
cuncer	KIAA1549(57670) MDM2(4193) MLH1(4292) MN1(4330)
	MYCN(4613) NBS1(4683) NBS1(4683) NF1(4763) NF2(4771)
	PAL $B^{2}(79728)$ PHOX2B(8929) PIK3CA(5290) PIK3B1(5295)
	PMS2(5395) PTCH(5727) PTEN(5728) ROS1(6098) SDHB(6390)
	SDHC(6391) SDHD(6392) SUFU(51684) TP53(7157) WRN(7486)
Breast	AKT1(207) BRCA1(672) BRCA2(675) BRIP1(83990) CCND1(595)
cancer	CDH1(999) CHEK2(11200) EP300(2033) ERBB2(2064) ETV6(2120)
cancer	MAP2KA(6A16) NTRK3(A016) PAL B2(70728) PIK3CA(5200)
	PR1(5025) TD53(7157)
coloractal	AET(3323), HIS(7137) AET(207), AEC(224), BEAE(672), CTNNE1(1400), EE200(2022)
conor	AKT1(207), AFC(324), DKAF(073), CTININD1(1499), EF300(2033), EDVW7(55204), VDAS(2845), MADH4(4080), MAD2V4(6416)
cancer	MDM2(4102) ML $H1(4202)$ MSH2(4426) MSH6(2056)
	MUTVH(4505), $MUT1(4292)$ , $MUT12(4450)$ , $MUTVH(4505)$ , $MU2CA(5200)$ , $MUTVH(4505)$ , $MU2CA(5200)$ , $MU2D1(5205)$ , $DMS1(5278)$
	MOTTH(4393), FIRSCA(3290), FIRSRT(3293), FIRST(3378), DMS2(5205), TD52(7157)
andomatrial	EDVW7(55204) = ECED2(2262) = IA7E1(221805) = MI H1(4202)
endomethal	FDAW / (35294), FUFR2(2205), JAZF1(221095), WILT1(4292), MSU2(4426), MSU6(2056), DMS1(5278), DMS2(5205), DTEN(5728)
cancer	MSD2(4450), MSD0(2950), FMS1(5576), FMS2(5595), F1EN(5726), SU212(22512)
	SUZ12(25512)
eye cancer	KB1(5925)
nead and	ME1(4233)
neck cancer	$\mathbf{DUD}(2011(2), \mathbf{EU}(2071), \mathbf{CDC}(20710), \mathbf{MET}(4022), \mathbf{NONO}(4041)$
Kidney	BHD(201163), FH(22/1), GPC3(2/19), ME1(4233), NONO(4841), DAL D2(70729), DDCC(55.4C), DDC1072(20005), SEDC((421))
cancer	PALB2(79728), PRCC(5546), PRO1073(29005), SFPQ(6421),
	1FE3(7030), 1FEB(7942), 1SC1(7248), 1SC2(7249), VHL(7428), 1SC2(7249), VHL(7428), 1SC2(7249), 1SC2(726), 1SC2(726), 1SC2(726), 1SC2(726), 1SC2(726), 1SC2(726), 1SC2(726), 1
	W11(7490), W1X(139285)
leukemia	ABL1(25), ABL2(27), AF15Q14(57082), AF1Q(10962), AF3p21(51517),
	AF5q31(2/125), ARHGEF12(23365), ARN1(405), ATM(4/2),
	BCL11A(53335), BCL11B(64919), BCL2(596), BCL3(602), BCL5(603),
	BCL6(604), BCL9(607), BCR(613), BLM(641), BRCA2(675),
	BRIP1(83990), BTG1(694), CBFA2T1(862), CBFA2T3(863),
	CBFB(865), CBL (867), CCND1(595), CCND1(595), CCND2(894),
	CDK6(1021), CDX2(1045), CEBPA(1050), CHIC2(26511),
	CREBBP(1387), D10S170(8030), DDX10(1662), DEK(7913),
	ELF4(2000), ELL(8178), ELN(2006), EP300(2033), EPS15(2060), ERG
	(2078), ETV6(2120), EVI1(2122), EWSR1(2130), FACL6(23305),
	FANCA(2175), FANCC(2176), FANCD2(2177), FANCE(2178),
	FANCF(2188), FANCG(2189), FBXW7(55294), FCGR2B(2213),
	FLT3(2322), FNBP1(23048), FOXO3A(2309), FOXP1(27086),
	FSTL3(10272), FUS(2521), GAS7(8522), GATA1(2623), GATA2(2624),
	GMPS(8833), GPHN(10243), GRAF(23092), HEAB(10978), HLF(3131),
	HLXB9(3110), HOXA11(3207), HOXA13(3209), HOXA9(3205),
	HOXC11(3227), HOXC13(3229), HOXD11(3237), HOXD13(3239)

 Table 10 Mutation targets of different cancer types

leukemia	IGH@(3492), IKZF1(10320), JAK2(3717), JAK3(3718), KIT(3815),
	KRAS(3845), LAF4(3899), LASP1(3927), LCK(3932), LCX(80312),
	LMO1(4004), LMO2(4005), LPP(4026), LYL1(4066), MDS1(4197),
	MKL1(57591), MLF1(4291), MLL(4297), MLLT1(4298),
	MLLT10(8028), MLLT2(4299), MLLT3(4300), MLLT4(4301),
	MLLT6(4302), MLLT7(4303), MN1(4330), MSF(10801),
	MSI2(124540), MTCP1(4515), MYC(4609), MYH11(4629),
	MYST4(23522), NCOA2(10499), NOTCH1(4851), NPM1(4869),
	NRAS(4893), NSD1(64324), NUMA1(4926), NUP214(8021),
	NUP98(4928), OLIG2(10215), PALB2(79728), PAX5(5079),
	PBX1(5087), PCM1(5108), PDGFRB(5159), PER1(5187),
	PICALM(8301), PML(5371), PMX1(5396), PNUTL1(5413),
	PRDM16(63976), PSIP2(11168), PTPN11(5781), RANBP17(64901),
	RAP1GDS1(5910), RARA(5914), RBM15(64783), RPL22(6146),
	RPN1(6184), RUNX1(861), RUNXBP2(7994), SBDS(51119),
	SEPT6(23157), SET(6418), SH3GL1(6455), SIL(6491),
	SSH3BP1(10006), STL(7955), TAF15(8148), TAL1(6886), TAL2(6887),
	TCF3(6929), TCL1A(8115), TCL6(27004), TFPT(29844), TIF1(8805),
	TLX1(3195), TLX3(30012), TOP1(7150), TRA@(6955), TRB@(6957).
	TRD@(6964), TRIP11(9321), TTL(150465), WHSC1L1(54904),
	ZNF145(7704), ZNF384(171017), ZNF521(25925), ZNFN1A1(10320)
liver cancer	APC(324), CTNNB1(1499), IL6ST(3572), TCF1(6927)
lung cancer	AKT1(207), ALK(238), BRAF(673), EGFR(1956), EML4(27436),
8	ERBB2(2064), FGFR2(2263), KRAS(3845), MYCL1(4610), RB1(5925).
	STK11(6794), TP53(7157)
lymphoma	ALK(238), ALO17(57714), ARHH(399), ATIC(471), ATM(472),
<b>J F - - -</b>	BCL10(8915), BCL2(596), BCL6(604), BCL7A(605), BIRC3(330),
	BLM(641), CARD11(84433), CARS(833), CCND2(894), CEP1(11064),
	CLTC(1213), CLTCL1(8218), DDX6(1656), EIF4A2(1974),
	ETV6(2120), FGFR1(2260), FGFR1OP(11116), FGFR3(2261),
	FVT1(2531), HIST1H4I(8294), HSPCA(3320), HSPCB(3326),
	IGH@(3492), IGK@(50802), IGL@(3535), IL2(3558), IL21R(50615),
	IRTA1(83417), ITK(3702), LCP1(3936), MALT1(10892).
	MHC2TA(4261), MSN(4478), MUC1(4582), MYC(4609), MYH9(4627),
	NACA(4666), NBS1(4683), NFKB2(4791), NPM1(4869),
	PAFAH1B2(5049), PAX5(5079), PCSK7(9159), PIM1(5292),
	POU2AF1(5450), REL(5966), SFRS3(6428), SOCS1(8651),
	SYK(6850), TFG(10342), TFRC(7037), TNFRSF17(608).
	TNFRSF6(355), TPM3(7170), TPM4(7171), WAS(7454),
	ZNF198(7750), ZNFN1A1(10320)
myeloma	CCND3(896), FGFR3(2261), HCMOGT1(92521), HIP1(3092),
5	IGH@(3492), IRF4(3662), MAF(4094), MAFB(9935), NRAS(4893),
	PDGFRB(5159), PER1(5187), PTPN11(5781), RAB5EP(9135),
	WHSC1(7468)
ovarian	AKT1(207), AKT2(208), BRAF(673), BRCA1(672), BRCA2(675),
cancer	CTNNB1(1499), ERBB2(2064), MLH1(4292), MSH2(4436).
•••••••	MSH6(2956), PIK3R1(5295), PMS1(5378), PMS2(5395), STK11(6794)
## Table 10 (cont'd)

pancreatic	AKT2(208), APC(324), BRCA2(675), CDKN2A(1029), EP300(2033),
cancer	KRAS(3845), MADH4(4089), MAP2K4(6416), MEN1(4221),
	STK11(6794)
prostate	C15orf21(283651), COPEB(1316), ERG (2078), ETV1(2115),
cancer	ETV4(2118), ETV5(2119), HNRNPA2B1(3181), PTEN(5728),
	SLC45A3(85414), TMPRSS2(7113)
sarcoma	ASPSCR1(79058), BUB1B(701), CHN1(1123), CIC(23152),
	COL1A1(1277), CREB1(1385), CREB3L2(64764), DDIT3(1649),
	DUX4(22947), ERG (2078), ETV1(2115), ETV4(2118), ETV6(2120),
	EWSR1(2130), EXT1(2131), EXT2(2132), FEV(54738), FLI1(2313),
	FOXO1A(2308), FUS(2521), FUS(2521), HRAS(3265), MDM2(4193),
	NBS1(4683), NCOA1(8648), NR4A3(8013), NTRK3(4916),
	PAX3(5077), PAX7(5081), PDGFB(5155), POU5F1(5460), RB1(5925),
	RECQL4(9401), SS18(6760), SS18L1(26039), SSX1 (6756),
	SSX2(6757), SSX4(6759), TAF15(8148), TCF12(6938), TFE3(7030),
	TP53(7157), WRN(7486), ZNF278(23598)
skin cancer	ATF1(466), BLM(641), BRAF(673), CDK4(1019), CDKN2A(1029),
	DDB2(1643), ERCC2(2068), ERCC3(2071), ERCC4(2072),
	ERCC5(2073), GNAQ(2776), KIT(3815), MITF(4286), NRAS(4893),
	PTCH(5727), RECQL4(9401), SMO(6608), TNFRSF6(355), XPA(7507),
	XPC(7508)
stomach	CDH1(999), ERBB2(2064), FGFR2(2263), PIK3CA(5290)
cancer	
testicular	KIT(3815), STK11(6794), TNFRSF6(355)
cancer	
thyroid	AKAP9(10142), BRAF(673), D10S170(8030), ELKS(23085),
cancer	GOLGA5(9950), HMGA1(3159), HRPT2(3279), KRAS(3845),
	KTN1(3895), MEN1(4221), NCOA4(8031), NRAS(4893),
	NTRK1(4914), PAX8(7849), PCM1(5108), PPARG(5468),
	PRKAR1A(5573), RET(5979), TFG(10342), TPM3(7170), TPR(7175),
	TRIM33(51592), TSHR(7253), ZNF331(55422)

Cancer	2001	2002	2003	2004	2005	2006	2007
type	Global						
	lethality						
	ratio						
lung cancer	0.284	0.280	0.283	0.285	0.287	0.288	0.287
colorectal	0.103	0.102	0.103	0.101	0.099	0.098	0.093
cancer							
breast	0.150	0.148	0.147	0.147	0.147	0.150	0.150
cancer							
pancreatic	0.052	0.054	0.054	0.056	0.056	0.057	0.060
cancer							
prostate	0.110	0.106	0.101	0.103	0.103	0.094	0.093
cancer							
leukemia	0.039	0.039	0.039	0.041	0.040	0.039	0.039
lymphoma	0.050	0.047	0.044	0.037	0.036	0.036	0.035
liver cancer	0.026	0.026	0.026	0.025	0.027	0.029	0.030
endometrial	0.025	0.025	0.025	0.026	0.027	0.027	0.027
cancer							
ovarian	0.052	0.052	0.053	0.059	0.059	0.056	0.057
cancer							
esophagus	0.023	0.023	0.023	0.024	0.024	0.024	0.025
cancer							
bladder	0.022	0.023	0.023	0.023	0.023	0.023	0.025
cancer							
brain	0.024	0.024	0.024	0.023	0.022	0.023	0.023
cancer							
kidney	0.022	0.021	0.021	0.022	0.022	0.023	0.023
cancer							
skin cancer	0.018	0.017	0.018	0.018	0.019	0.019	0.019
stomach	0.023	0.022	0.022	0.021	0.020	0.020	0.020
cancer							
myeloma	0.020	0.020	0.020	0.020	0.020	0.020	0.019
cervical	0.017	0.015	0.015	0.014	0.014	0.014	0.014
cancer							
testicular	0.001	0.001	0.001	0.001	0.001	0.001	0.001
cancer							
eye cancer	0.000	0.000	0.000	0.000	0.000	0.000	0.000

**Table 11** Global lethality ratio values for different cancers from 2001 to 2007

Cancer type	2001	2002	2003	2004	2005	2006	2007
	Local						
	lethality						
	ratio						
lung cancer	0.929	0.914	0.914	0.923	0.947	0.931	0.752
colorectal	0.360	0.328	0.680	0.267	0.268	0.259	0.232
cancer							
breast cancer	0.209	0.195	0.188	0.186	0.191	0.192	0.227
pancreatic	0.990	0.980	0.977	0.981	0.988	0.958	0.898
cancer							
prostate	0.159	0.160	0.131	0.130	0.131	0.117	0.124
cancer							
leukemia	0.682	0.704	0.716	0.697	0.648	0.635	0.492
lymphoma	0.434	0.424	0.405	0.333	0.323	0.305	0.276
liver cancer	0.870	0.849	0.832	0.754	0.879	0.875	0.876
endometrial	0.172	0.168	0.170	0.176	0.179	0.178	0.189
cancer							
ovarian	0.594	0.597	0.563	0.629	0.729	0.759	0.681
cancer							
esophagus	0.947	0.962	0.935	0.933	0.935	0.946	0.896
cancer							
bladder	0.228	0.223	0.218	0.211	0.208	0.213	0.205
cancer							
brain cancer	0.762	0.771	0.716	0.690	0.690	0.681	0.621
kidney cancer	0.393	0.365	0.373	0.349	0.350	0.330	0.252
skin cancer	0.174	0.165	0.167	0.173	0.160	0.156	0.167
stomach	0.590	0.574	0.540	0.519	0.528	0.513	0.527
cancer							
myeloma	0.778	0.740	0.747	0.725	0.707	0.683	0.542
cervical	0.341	0.315	0.336	0.371	0.358	0.381	0.329
cancer							
testicular	0.056	0.053	0.053	0.040	0.049	0.045	0.048
cancer							
eye cancer	0.095	0.091	0.091	0.086	0.108	0.097	0.094

 Table 12 Local lethality ratio values for different cancers from 2001 to 2007

**Table 13** Correlation values of weighted degree, approval number values with globallethality values for 2001-2007

	FDA appro number	FDA approval number		er network legree	
	All cancers	All cancers except lung	All cancers	All cancers except lung	
		cancer		cancer	
2001	0.507883	0.440047	-0.01325	-0.01431	Spearman statistic
	0.044594	0.100702	0.961142	0.959628	Spearman p-value
2002	0.493703	0.434419	0.010302	-0.00179	Spearman statistic
	0.05195	0.105646	0.969796	0.994956	Spearman p-value
2003	0.453218	0.384073	0.295524	0.207506	Spearman statistic
	0.067694	0.141915	0.249477	0.440621	Spearman p-value
2004	0.529063	0.454699	0.256285	0.133922	Spearman statistic
	0.028981	0.076807	0.320752	0.620973	Spearman p-value
2005	0.549142	0.489158	0.273761	0.159509	Spearman statistic
	0.018258	0.046291	0.271658	0.540859	Spearman p-value
2006	0.525673	0.4698	0.414946	0.313503	Spearman statistic
	0.020802	0.049159	0.077301	0.205218	Spearman p-value
2007	0.479393	0.415906	0.301958	0.185413	Spearman statistic
	0.032453	0.076549	0.195699	0.447281	Spearman p-value

**Table 14** Correlation values of weighted degree, approval number values with local lethality values for 2001-2007

	FDA appro number	FDA approval number		er network legree	
	All	All	All	All	
	cancers	cancers	cancers	cancers	
		except		except	
		pancreatic,		pancreatic,	
		liver and		liver and	
		esophagus		esophagus	
		cancers		cancers	
2001	0.062646	0.40361	0.267647	0.481319	Spearman statistic
	0.817717	0.152398	0.312408	0.082451	Spearman p-value
2002	0.078756	0.425737	0.270588	0.485714	Spearman statistic
	0.771877	0.129076	0.306974	0.079402	Spearman p-value
2003	0.203359	0.553168	0.144697	0.37891	Spearman statistic
	0.433722	0.03244	0.57952	0.163683	Spearman p-value
2004	0.164304	0.522921	0.137255	0.407143	Spearman statistic
	0.528597	0.045488	0.595315	0.131493	Spearman p-value
2005	0.197299	0.39853	0.168215	0.358824	Spearman statistic
	0.432609	0.126272	0.500542	0.170818	Spearman p-value
2006	0.174395	0.389871	0.124561	0.330882	Spearman statistic
	0.475188	0.121863	0.607889	0.192679	Spearman p-value
2007	0.072584	0.449279	0.171429	0.573529	Spearman statistic
	0.761048	0.070419	0.466237	0.017639	Spearman p-value

	2000	2001	2002	2003	2004	2005	2006	2007
breast	0.944	1.013	0.936	1.192	1.269	1.250	1.327	1.139
cancer								
lung cancer	0.774	0.771	0.763	1.049	1.092	1.158	1.315	1.296
leukemia	0.559	0.525	0.514	0.832	0.706	0.711	0.681	0.649
lymphoma	0.802	0.797	0.716	0.983	0.957	0.934	0.858	0.839
ovarian	1.258	1.255	1.223	1.031	0.967	0.962	1.097	1.164
cancer								
head and	NA	NA	NA	2.312	2.282	2.273	1.411	1.389
neck cancer								
myeloma	1.401	1.397	1.138	0.937	0.923	0.913	0.771	0.873
stomach	0.917	0.833	0.827	0.663	0.591	0.488	1.009	1.156
cancer								
sarcoma	0.000	0.000	0.000	2.312	2.282	2.273	1.237	1.223
endometrial	NA	NA	NA	2.312	2.282	2.273	1.086	1.072
cancer								
eye cancer	1.144	1.138	1.037	0.921	0.906	0.896	0.824	0.793
brain cancer	1.319	1.314	1.191	1.067	1.054	1.041	0.936	0.897
colorectal	0.655	0.593	0.583	0.525	0.391	0.336	0.434	0.422
cancer								
skin cancer	0.567	0.562	0.558	0.484	0.416	0.414	0.406	0.487
pancreatic	0.927	0.877	0.871	0.683	0.681	0.629	0.525	0.505
cancer								
kidney	NA	NA	NA	NA	NA	0.000	0.000	0.500
cancer								
liver cancer	NA	0.333						
prostate	0.000	0.000	0.000	0.000	0.125	0.125	0.488	0.479
cancer								
cervical	NA	NA	NA	NA	NA	NA	0.238	0.238
cancer								
testicular	0.341	0.341	0.341	0.327	0.317	0.317	0.313	0.313
cancer								
bladder	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250
cancer								
esophagus	0.111	0.111	0.111	0.091	0.077	0.077	0.071	0.071
cancer					0.077	0.0	0.071	0.071
mesotheliom	NA	NA	NA	NA	0.077	0.077	0.071	0.071
a								

 Table 15 Weighted degree values for FDA cancer network from 2000 to 2007

	2000	2001	2002	2003	2004	2005	2006	2007
breast cancer	0.063	0.035	0.027	0.028	0.007	0.012	0.007	0.009
lung cancer	0.327	0.289	0.343	0.164	0.088	0.073	0.022	0.024
leukemia	0.383	0.391	0.401	0.169	0.181	0.151	0.166	0.155
lymphoma	0.299	0.298	0.292	0.188	0.176	0.161	0.203	0.212
ovarian	0.315	0.291	0.286	0.515	0.535	0.528	0.509	0.466
cancer								
head and	NA	NA	NA	0.369	0.388	0.401	0.544	0.522
neck cancer								
myeloma	0.379	0.415	0.426	0.627	0.615	0.641	0.664	0.563
stomach	0.618	0.596	0.617	0.817	0.832	0.843	0.694	0.582
cancer								
sarcoma	1.000	1.000	1.000	0.367	0.386	0.374	0.612	0.623
endometrial	NA	NA	NA	0.380	0.387	0.379	0.632	0.617
cancer								
eye cancer	0.520	0.506	0.541	0.723	0.723	0.690	0.714	0.724
brain cancer	0.434	0.431	0.450	0.672	0.650	0.653	0.747	0.741
colorectal	0.706	0.732	0.711	0.838	0.897	0.878	0.856	0.868
cancer								
skin cancer	0.741	0.787	0.775	0.858	0.890	0.885	0.925	0.894
pancreatic	0.596	0.627	0.619	0.797	0.786	0.814	0.885	0.881
cancer								
kidney cancer	NA	NA	NA	NA	NA	1.000	1.000	0.891
liver cancer	NA	0.907						
prostate	1.000	1.000	1.000	1.000	0.966	0.983	0.880	0.898
cancer								
cervical	NA	NA	NA	NA	NA	NA	0.931	0.931
cancer								
testicular	0.862	0.862	0.858	0.919	0.929	0.931	0.942	0.945
cancer								
bladder	0.915	0.913	0.896	0.939	0.946	0.942	0.949	0.962
cancer								
esophagus	0.940	0.944	0.928	0.956	0.987	0.980	0.979	0.985
cancer								
mesothelioma	NA	NA	NA	NA	0.966	0.973	0.979	0.984

 Table 16 Weighted degree p-values for FDA cancer network from 2000 to 2007

**Table 17** Clinical trial numbers along with distinct drug number and specific drug number for clinical trials

Cancer type	Clinical	Clinical	Clinical
	drug	trials	trials
	trial	distinct	specific
	number	drug	drug
		number	number
leukemia	170	37	2
lymphoma	121	42	3
lung cancer	121	30	0
breast cancer	97	37	2
ovarian cancer	71	34	0
brain cancer	62	25	1
colorectal cancer	61	16	0
prostate cancer	48	23	3
head and neck cancer	45	26	1
mesothelioma	40	8	0
pancreatic cancer	35	21	0
skin cancer	31	21	0
kidney cancer	30	27	0
esophagus cancer	26	16	0
cervical cancer	20	11	0
liver cancer	19	16	0
stomach cancer	19	13	0
myeloma	13	9	1
testicular cancer	11	14	0
bladder cancer	10	10	1
endometrial cancer	10	9	0
sarcoma	7	24	0
eye cancer	2	3	0

**Figure 5** Time dependent characteristics of the FDA approvals and FDA cancer network (A) Number of cancers in the network from 1980-2008, (B) Number of FDA approvals from 1980-2008, (C) Average weight of the network from 1980-2008, (D) Number of components of the network from 1980-2008.









Figure 5 (cont'd)





D





Figure 6 Cluster dendogram of cancer types based on global and local lethality values Height

BIBLIOGRAPHY

### BIBLIOGRAPHY

- 1. Miniño AM, Heron MP, Murphy SL, Kochanek KD; Centers for Disease Control and Prevention National Center for Health Statistics National Vital Statistics System. Deaths: final data for 2004. Natl Vital Stat Rep. 2007, 55:1-119
- 2. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, et al. The human disease network. Proc Natl Acad Sci U S A. 2007, 104: 8685-8690
- 3. Yildirim MA, Goh KI., Cusick ME, Barabasi AL, Vidal M. Drug-target network. Nat Biotechnol. 2007, 25: 1119-1126
- 4. Ma'ayan A, Jenkins SL, Goldfarb J, Iyengar R. Network analysis of FDA approved drugs and their targets. Mt Sinai J Med. 2007, 74: 27-32
- 5. Sakharkar MK, Li P, Zhong Z, Sakharkar KR. Quantitative analysis on the characteristics of targets with FDA approved drugs. Int J Biol Sci. 2008, 4:15-22
- 6. Nacher JC, Schwartz JM. A global view of drug-therapy interactions. BMC Pharmacol. 2008, 8:5
- 7. Barabasi AL. Network Medicine From Obesity to the "Diseasome". N Engl J Med. 2007, 357: 404-407
- 8. Hopkins, AL. Network pharmacology. Nat Biotechnol. 2007, 25: 1110-1111
- 9. American Cancer Society. Cancer Facts & Figures. Atlanta: American Cancer Society 2001-2008
- Newman MEJ. The structure and function of complex networks. SIAM Review. 2003, 45: 167-256
- 11. Yokota J, Yamamoto T, Toyoshima K, Terada M, Sugimura T, et al. Amplification of cerbB-2 oncogene in human adenocarcinomas in vivo. Lancet. 1986, 1:765-767
- 12. Testa JR, Siegfried JM. Chromosome abnormalities in human non-small cell lung cancer. Cancer Res. 1992, 52:2702-2706
- 13. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, et al. A census of human cancer genes. Nat Rev Cancer. 2004, 4:177-183

- 14. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003, 13: 2498-2504
- 15. Best DJ, Roberts DE. Algorithm AS 89: The Upper Tail Probabilities of Spearman's rho. Applied Statistics 1975, 24: 377–379

#### **CHAPTER 2**

# INTEGRATIVE ANALYSIS OF CANCER PATHWAYS INTRODUCTION

Cancer is a complex disease, with many subtypes, affecting various tissues, and according to the severity of the abnormality, giving rise to different stages and classifications, such as carcinoma, sarcoma, primary, etc. Physiological and genetic studies identified different stages in the progression of the various types of cancers. For instance, colorectal tumorigenesis begins with normal epithelium which proceeds through stages of hyperproliferative epithelium, early adenoma, intermediate adenoma, late adenoma, carcinoma, and metastasis [1]. Genetic alterations, such as mutations or deletions of genes, accumulate as the cancer progresses to the next, more severe and proliferative stage. The accumulation of changes (e.g., mutations, deletions, etc.) is a key factor in tumor progression. For example, a more severe stage will have a higher probability than a less severe stage of having more mutations [1]. On the other hand, the specific number and the identity of the genetic alterations, such as point mutations, amplifications, or deletions in a stage, are not determining factors, but rather are general features that characterize the severity of a stage, and varies from one cancer to another [1].

Omics technologies, such as cDNA and oligonucleotide arrays, and comparative genome hybridization, have dominated tumor characterization [2]. Genome level analysis have been applied successfully to differentiate between the different stages of cancer, and revealed differentially expressed genes and genomic alterations that play important roles in the development of cancer [2]. More recently, gene set and pathway centric analysis of cancer progression have gained popularity. These analyses confirmed results currently known about the role of cell cycle in cancer development, as well as produced novel findings, i.e. the involvement

of the ERBB4 gene in primary prostate cancer [3]. Approaches that integrate gene expression and pathway analysis of colorectal cancer have proved useful in finding potential prognostic and diagnostic markers, as well as therapeutic targets [2].

Biological pathways consist of molecular interactions or other biochemical events among a group of proteins, genes, and other chemicals. They are used to characterize metabolism, signal transduction, specific cellular processes, diseases, or drug mechanisms. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database contains pathway information on the progression of different types of cancer, as combinations of signaling pathways

(http://www.genome.ad.jp/kegg/pathway.html#disease) [4, 5, 6]. Signaling pathways capture the molecular interactions and reactions that take signals from the outside to the nucleus of the cell, where transcriptional regulation occurs. Signaling pathways, e.g. the MAPK, Wnt, TGF-beta signaling, are well studied in the context of cell proliferation [7]. Cancer is the only human disease for which pathway information of the disease progression is available, i.e., at different stages of the cancer, from normal tissue to the advanced tumor phase in KEGG [4, 5, 6]. KEGG cancer pathways are different in that they contain members of different signaling pathway members within a single pathway. This is because they contain information on various stages of the cancer and the various stages involve different signaling pathways. The pathway information contained in KEGG, i.e., the genes in each stage of the disease progression that are genetically altered in the colorectal cancer pathway, is derived from the literature [8, 9]. In the KEGG colorectal cancer pathway, the Wnt pathway members are at the upper positions of the image, corresponding to normal or initial stages of the cancer progression, which is supported by the literature [8, 9, 10]. Constant activation of Wnt signaling targets occurs in adenomatosis polyposis coli (APC) or beta-catenin mutations, as well as upon phosphorylation by glycogen

synthase kinase-3-beta. These events lead to early dysplastic lesions, which is an early event in colorectal cancer [10]. Similar support exists in the literature to suggest that genes, i.e., DCC and KRAS, should be below the Wnt pathway members since their alterations occur later in the disease progression. Furthermore, DCC is located below KRAS in the KEGG pathway because DCC alterations are believed to occur later than KRAS alterations in the development of colorectal cancer [11]. Finally, TGF-beta pathway members are located in the lower portions of the pathway image because the literature suggests that their alterations occur during more advanced stages of the cancer [12]. The significance of the KEGG cancer pathways is the integration of the cancer stages with signaling pathways. Although the integration of signaling pathways with genome level expression data has been widely performed, it has yet to be realized with cancer pathways.

Using pathway information to understand genome level expression data has been extensively applied [13, 14, 15]. The approach integrates *a priori* knowledge of a gene's functional role with expression data to detect for concerted expression changes in a set of genes responsible for producing a phenotype [13]. Pathway-centric analysis of tumor microarray data has been successfully applied to identify signaling pathway members [16]. IL-1 and ER-induced pathways were found to be significantly coexpressed in breast cancer data. In addition to analyzing the entire dataset, analysis of individual samples or a subset of the data identified significant pathway activities that were relevant to the biological context of the tissue or organ. For example, pathway analysis uncovered the expected association of estrogen-induced pathways within a group of clinical breast cancer data [16], and signaling and metabolic pathways involved in the development of type 2 diabetes [17]. On the other hand, gene expression level analysis *within* pathways is an area that mostly has been ignored. Some studies

have analyzed the relationship among the members of protein complexes or pathways in terms of their gene expression levels [18, 19, 20, 21]. Protein complexes, such as ribosome and proteasome, show significant correlation in their gene expression levels [18]. In addition, the ciselement profiles are highly similar for members within a signaling pathway, such as the KEGG apoptosis pathway, and functionally related interacting proteins (i.e., protein complexes). This suggests that a strong relationship in the gene expression levels between members of these pathways should exist [19]. Coherence is a measure of the level of correlation among a group of genes. A coherent group of genes may share similar regulation of their gene expression levels. Indeed, genes in the same pathway with similar functions have been shown to be coherent as compared to a random group of genes from the genome [20].

Previous studies that integrate gene expression data with pathway information have not incorporated the dimensionality of the pathways. Most studies have focused on the members of the pathways. Thus far, there has been one study of pathways that incorporated the position in the pathways in the analysis of the genome level data. They developed a statistical impact analysis that used the pathway position to calculate the significance of the pathway [21]. With this approach they identified the Focal Adhesion Pathway as a significant pathway for lung cancer, which was not found using classical approaches, such as gene set enrichment and gene ontology analyses, thereby enhancing the information content extracted from analyzing genome level data. Impact analysis considers expression alteration in receptors, such as integrin, receptor tyrosine kinase (RTK), and the receptor ligand vascular endothelial growth factor (VEGF), as important parameters in the analysis, since they affect the downstream molecules in the pathway. The Focal Adhesion Pathway, on the other hand, was not found to be significant using classical approaches because the other genes in the pathway were not significantly altered. Classical

approaches analyze pathways as a whole without special emphasis on receptor molecules or other position(s) in a pathway.

In this study, we capitalized upon the progressive nature of the cancer disease captured in the biological knowledge represented in the KEGG cancer pathways. We analyzed the gene expression levels, coherence, and mutation target data of the pathway members to determine if there is a significant relationship or correlation within any group of pathway members from the rest of the pathways. Analyzing the KEGG colorectal cancer pathway with microarray data identified that different parts of the pathways were up-regulated or coherent at the mRNA level, at the different stages of cancer progression, i.e., adenoma vs. carcinoma. Since the KEGG cancer pathways integrate different signaling pathways of the various cancer stages, unlike classical signaling pathways, we analyzed the coherence of the colorectal cancer pathway, and found the carcinoma expression data was more coherent than the normal or adenoma data. In addition, mutation targets were found to be localized primarily in the nucleus of the cell and concentrated at the later stages of the cancer.

#### **MATERIALS AND METHODS**

#### Pathway data

We used KEGG as our source of pathway information [4, 5, 6]. We focused on the cancer, apoptosis, oxidative phosphorylation and proteasome pathways. We collected a list of human genes for these pathways. For each protein/gene in the pathways we used all the different homologues provided by KEGG database.

#### X/Y scale

The X scale represented the direction from receptor to nucleus. The Y scale (analyzed only for the colorectal cancer pathway) represented the direction from normal tissue to advanced tumor or metastasis. We ignored the sub-pathway distinction (such as Chromosome Unstable Pathway and Microsatellite Unstable Pathway in colorectal cancer pathway). We generated 3 Xdependent and 4 Y-dependent groups in the KEGG colorectal cancer pathway, and 2 Xdependent groups in the KEGG apoptosis pathway. For X, receptor ligands and receptors are designated by a value of 1. If there is a nuclear distinction in the pathway; molecules in the cytosol are designated by a value of 2, and molecules in the nucleus are designated by a value of 3, otherwise molecules in the cytoplasm are designated by a value of 2. For Y (analyzed only for the KEGG colorectal cancer pathway), the first stage of the pathway, which signified the first initial molecular events (in the colorectal cancer pathway, the first stage is represented by the transition from normal epithelium to early adenoma) is designated by a value of 1 and the next stages in the cancer progression are designated by values of 2, 3, 4 (the last stage). The last stage is defined by the latest events in the progression of the disease in the KEGG pathway. For example, in the colorectal cancer pathway, the last stage is represented by the transition from late adenoma to carcinoma. The biomolecules involved in the transition to the different stages provided by KEGG are supported by the literature, as discussed in the Introduction above. The KEGG information was used as is, unless the stage is an insignificant one. For example, dysplastic aberrant crypt foci stage was ignored in our analysis of the colorectal cancer pathway, because this stage covered a very small part of the pathway and it was unclear which molecules were associated with this stage. In the colorectal cancer pathway, normal epithelium to early adenoma is designated by a Y value of 1 and includes proteins such as, Frizzled, glycogen

synthase kinase-3-beta (GSK-3β), adenomatosis polyposis coli (APC), T cell factor/lymphoid enhancer factor (TCFLEF), Survivin, etc. Early adenoma to intermediate adenoma is designated by a Y value of 2 and includes proteins such as, RTK, K-Ras, protein kinase B (PKB), extracellular signal regulated protein kinase (ERK), C-Fos, etc. Intermediate adenoma to late adenoma is designated by a Y value of 3 and includes proteins such as, deleted in colon cancer (DCC), caspase 3 (CASP3), human mutL homolog 1 (hMLH1), etc. Late adenoma to carcinoma is designated by a Y value of 4 and includes proteins such as cytochrome c (Cytc), p53, etc. Since our analysis focused on the potential of the genes in contributing to the progression of cancer, we used the larger value of Y for a gene if there were more than one value associated with the gene. For consistency, the same approach was taken for the X value, i.e., the larger of the two values was used. For example, if a protein is present in more than one location, the larger value is assigned, for instance, CyclinD1 is present in two Y locations, 1 and 2, but was assigned a Y value of 2, and transforming growth factor beta receptor 2 (TGFBR2) is located in two X locations, 1 and 3, but was assigned a X value of 3.

#### **Expression data**

Normalized microarray data (GSE4183 and GSE8671) were downloaded from the NCBI GEO database (http://www.ncbi.nlm.nih.gov/geo/) GSE4183 includes normal colon tissue, colon adenoma and colon carcinoma gene expression datasets for several biopsy samples obtained from individuals. The GSE8671 dataset includes the whole genome mRNA expression level for 32 colorectal adenomas paired with the normal mucosa from the same individuals. In these large-scale datasets, we focused on the expression values of the pathway members, as oppose to analyzing all the genes. Both datasets contained expression values for all the gene members in the colorectal cancer pathway. If there were more than one value for a gene in a sample (same

column), the mean values were used. We combined the normal and the adenoma samples from both datasets, to obtain a set of expression values for normal, adenoma and carcinoma samples. We calculated expression ratio for adenoma (or carcinoma) by dividing the average value of adenoma (or carcinoma) samples to the average value of normal samples. For adenoma, the ratio was calculated by dividing the average of the combined adenoma values to the average of the combined normal values. Since carcinoma data is present only in GSE4183, the average of the carcinoma values in GSE4183 were divided by the average of the normal values in GSE4183. For the calculation of carcinoma/adenoma ratio, the carcinoma ratio was divided by the adenoma ratio for each gene.

#### **Drug and mutation targets**

We collected drug information from the National Cancer Institute web page (http://www.cancer.gov/cancertopics/druginfo/alphalist). We limited the drug list to ones which targeted a cellular signaling protein. Drugs that targeted molecules with a general role in the cell were excluded; for example, Bortezomib, a drug approved for leukemia, was excluded because it targeted the proteasome. We collected mutation target information from two sources. The first source was the Cancer Gene Census of the Cancer Genome project [22]

(http://www.sanger.ac.uk/genetics/CGP/Census/). We used only the cancer types provided by this dataset. For some mutation targets, such as p53, "others" was mentioned in addition to the cancer type listed, therefore we were not able to include the cancer types referred to as "others" in the analysis. The second source was a list of genetically altered genes provided by KEGG for each pathway. We combined these two sources of data to obtain a list of mutation targets for each cancer pathway. The list from the two sources did not overlap, for example, BRAF was identified as a mutation target only in the Cancer Gene Census dataset, while TGFBR2 was

identified as a mutation target only in the KEGG dataset. For mutation targets we calculated the frequency as the number of mutation targets in a group divided by the total number of genes in the group.

#### **Statistical Analysis**

Analysis of groups of genes or members of a pathway was performed for their coherence in the normal, adenoma, and carcinoma tissues. We calculated the Pearson's correlation coefficient for the expression values of every pair of genes in both datasets. We selected the same number of random genes from the entire microarray dataset, whether the genes belonged in the pathway or not, and performed the same calculation for this random group. For the range between 0-1, at increments of 0.01, we calculated the fraction of pairs with a correlation coefficient for both the real groups (pathways, subgroups inside pathways) and the random groups. For each gene, we performed the randomizations 1000 times for the pathways and 100 times for the subgroups within the pathways, and calculated the fraction of random pairs with a correlation coefficient threshold of 0.5. In addition to the colorectal cancer pathway, we analyzed the apoptosis pathway because it was previously shown to be coherent in colorectal cancer [20]. We also analyzed the oxidative phosphorylation and proteasome pathways which were shown to be coherent both in normal and cancer tissues [20]. The oxidative phosphorylation pathway has more genes than the colorectal cancer and apoptosis pathways, while the proteasome has fewer genes. Therefore any size effect of the pathway should be accounted for. For each X/Y dependent subgroup, randomizations of the same size were performed and whether the subgroups differed significantly from random was determined at a correlation coefficient threshold of 0.5. For the analysis of expression ratios, we performed t-test on groups of 2 (i.e., comparing gene expression profiles of pathway members in group Y=1 to members in the rest of

the pathway, namely Y=2, 3, 4). For more than two groups (i.e., comparing gene expression profiles of pathway members in groups X=1, X=2, and X=3), we used ANOVA.

#### **RESULTS AND DISCUSSION**

#### **Colorectal cancer pathway subgroups**

We analyzed the tumor expression levels of the members of the KEGG colorectal cancer pathway. We investigated whether the pathway members are differentially expressed with respect to their X (*cellular location*), and Y (*stage of the tumor*) values in the KEGG pathway, see Materials and Methods for details. The colorectal cancer pathway in KEGG provides some of the molecular events that underlie the progression of cancer, from normal to carcinoma. We assessed whether a relationship existed between the progression of colorectal cancer and the gene expression levels of the members of the colorectal cancer pathway. The normal and the adenoma samples from both datasets (GSE4183, GSE8671) were combined and the carcinoma dataset came from GSE4183.

Analyzing the genes that were highly differentiated among the pathway members in the carcinoma and adenoma datasets, we found AXIN2 and FZD3 genes were the most down-regulated in the carcinoma and the most upregulated in the adenoma samples. These genes had the lowest carcinoma to adenoma expression ratio, whereas PDGFRB and FZD2 were the most upregulated genes in the carcinoma samples and thus had the highest carcinoma to adenoma expression ratio (Figure 7). The average expression level of all the genes in the pathway did not change significantly for the adenoma and carcinoma samples (or stages), which centered around a ratio of 1. Analyzing the members at particular locations of the pathway suggested that different genes are differentially expressed and thus possibly differentially regulated, depending on the stage of the cancer tissues from which the expression data were obtained (Figure 7). This demonstrated one of the findings that could be obtained with this integrative approach and would have been lost by analyzing all the pathway members. Analysis of all the pathway members as a

whole, suggested no difference between the stages (i.e., ratio = 1) and did not identify a potential for any genes to be differentially regulated. FZD2 and FZD3 are different homologues of Frizzled, which is the receptor for Wnt signaling molecules and is known to be important in early development of colorectal cancer [8]. AXIN2 (Axin) is also a member of Wnt signaling and has been shown to be a mutation target in the development of colorectal adenoma [8]. Our analysis suggests that Wnt signaling members, such as different homologues of Frizzled receptor (FZD2, FZD3), may play a role in early (adenoma) and late (carcinoma) events of colorectal cancer progression. Wnt signaling is known to be activated during the earlier stages of colorectal cancer progression and is suggested to be involved in also the later stages of the progression [10]. Platelet-derived growth factor receptor beta (PDGFRB) is known to play a role in advanced and metastatic stages of colorectal cancer development [23]. It is noteworthy that most genes that showed differential expression in adenoma and carcinoma samples were receptors, i.e. Frizzled and PDGFRB. In support of this integrative pathway analysis, a previous study [21] also found that pathway information, i.e. whether it is a receptor, was important in identifying physiologically relevant functional groups. Currently, the drugs, Erbitux and Vectibix, used to treat colorectal cancer, target a receptor, EGFR (RTK). The location of the gene in the colorectal cancer pathway corresponds to the early cancer stage and the receptor region. The analysis appears to suggest that receptors could be important regulatory regions in colorectal cancer development, and as such, other receptors, i.e. FZD2, FZD3 and PDGFRB, could be possible candidates for drug development. However, more data is needed to confirm this relationship.

Next, we analyzed the possibility of dimensional (X and Y) distinction of the overall pathway, in other words, whether a group of pathway members defined by a stage (Y) or a cellular location (X) showed a significant difference in terms of expression levels, and the

presence of mutation and drug targets, as compared to the other members of the pathway. We set the X values to vary from 1 to 3, corresponding to the following locations: receptor/ligand (denoted as 1), cytosol (denoted as 2), and nucleus (denoted as 3). We examined the significance of grouping the expression profiles of the pathway members according to their X values. There was a significant grouping of both the adenoma and carcinoma expression profiles of the pathway members across the X groups (Table 18). In the adenoma and the carcinoma datasets, cytosolic members of the colorectal cancer pathway (X=2) had significantly lower gene expression values (Figure 8A, 8B, Table 19), on the other hand, nuclear members of the colorectal cancer pathway (X=3) had significantly higher gene expression values relative to other X groups (Figure 8C, 8D, Table 19).

Similar analysis was performed for the Y values, where the Y values ranged from 1 to 4, to correspond to the different stages, from normal tissue to early adenoma (denoted as 1), early to intermediate adenoma (denoted as 2), intermediate to advanced adenoma (denoted as 3), and advanced adenoma to carcinoma (denoted as 4). There was a significant difference across the Y groups for the adenoma but not the carcinoma datasets (Table 18), which may be attributed to a significant difference in the expression values of the pathway members with Y values of 1 as compared to the other pathway members (Y values of 2-4) in adenoma (Figure 8E, Table 19). In the adenoma tissue samples, the gene expression values of the pathway members, which play a role in the normal epithelium to early adenoma stage (Y=1) were expressed significantly higher than the other pathway members. The key genes that contribute to these results were BIRC5 (Survivin), FZD3, and AXIN2 (Axin), which are highly expressed and are located at Y=1, and PIK3CG, and PDGFRA, which are lowly expressed in the colorectal adenoma samples and located at Y=2. This result is in line with our previous observation that a group of genes in a

particular pathway could be differentially expressed from the other members in the pathway depending on the stage of the cancer. In addition to the adenoma tissue samples, we analyzed the expression values of the colorectal pathway members in the carcinoma tissue samples for possible distinctive patterns but found no significant grouping of the carcinoma expression with respect to Y (Figure 8F, Table 18, 19). If we consider that the colorectal cancer pathway includes the stages from normal epithelium to carcinoma, with several adenoma stages but only a single stage of carcinoma development, and also only one set of gene expression data from carcinoma tissues, there is likely insufficient information to distinguish the molecular events in the carcinoma pathway.

#### **Coherence of the colorectal cancer pathway**

The KEGG cancer pathways represent a collective behavior of a group of proteins that underlies the disease, and as such, are built from proteins that belong to multiple signaling pathways. A coherence indicator has been defined as the ratio of the number of correlated gene pairs to the total number of gene pairs in a pathway, which is deemed significant based upon a statistical measure [20]. Using this indicator, the gene expression levels of the signaling and metabolic pathway members were shown to be coherent, suggesting that coherence may be an important measure of functionally related genes [20]. Cancer pathways, due to their very nature of involving multiple signaling groups, may not be expected to show the same level of coherence. Therefore, we examined whether the gene expression levels of the KEGG colorectal cancer pathway members were coherent, as compared to apoptosis, oxidative phosphorylation and proteasome pathway members, the latter were previously shown to be coherent [20]. The degree of coherence is determined by plotting the correlation in gene expression of the pathway members. We analyzed normal colorectal, adenoma and carcinoma samples. In normal colorectal

tissue expression data, the oxidative phosphorylation and proteasome pathways show a distinct positive correlation among their pathway members, the apoptosis pathway show a slightly negative correlation, whereas the correlation distribution of the colorectal cancer members are closer to the random distributions and hence uncorrelated (Figure 9A). In addition to the correlation distribution of the expression levels, we analyzed the cumulative distributions of the absolute values of the correlation coefficients. Similar results for the cumulative and noncumulative correlation distributions are observed for normal colorectal tissue, i.e. the pathway is uncorrelated (Figure 9A, 9B). Oxidative phosphorylation and proteasome pathways are coherent in all 3 groups (normal, adenoma, and carcinoma tissues), while the colorectal cancer pathway is coherent only in the carcinoma samples (Figure 9B-9D). We compared whether the colorectal cancer and apoptosis pathways differed significantly from random at a correlation coefficient of 0.5 (Table 20). The colorectal cancer pathway appears to be coherent for colorectal carcinoma but not for the normal and adenoma datasets. This suggests that a cancer pathway may be coordinately regulated to achieve a biological function or phenotype, which, in this case, is the progression of the colorectal tissue to the tumor stage. The apoptosis pathway appears to be coherent in normal colorectal sample but not in the colorectal adenoma or carcinoma data (Table 20). This is in contrast to a previous report which used different gene expression data and found the data was coherent in the colorectal tumor but not in the normal samples [20]. In addition to the entire pathway, we studied the possibility of coherence of dimensional groupings in colorectal cancer and apoptosis pathways. Our analysis suggests that members within the pathways could be coherent, for example X=2 in the colorectal cancer pathway could be coherent in normal and adenoma samples even though the entire pathway may not be coherent (Table 20, 21). However, more data is required to confirm these analyses.

In order to determine which proteins contribute to the coherence of the colorectal cancer pathway in the carcinoma stage, we obtained pairs of genes with an absolute correlation coefficient of at least 0.5. There were 683 correlated gene pairs in the carcinoma samples, most of which were specific to carcinoma (Figure 10A). On the other hand, the normal and adenoma samples had fewer correlated gene pairs and shared more than half of them with each other. The pairs of correlated genes specific to normal and adenoma tissues included mostly genes in the early stage members of the pathway, such as AKT homologues (Gene ID of 207 and 208), DVL homologues (Gene ID of 1856 and 1857), and FZD homologues (Gene ID of 8321, 8322, 8323, 8324). On the other hand, pairs of correlated genes specific to carcinoma included mostly pairs of genes from different parts of the pathway, such as TGFBR2 homologue (Gene ID of 91) from late stage of the pathway, APC (Gene ID of 324) from early stage of the pathway, KRAS (Gene ID of 3485) and MET (Gene ID of 4233) from mid stage of the pathway. Next, we analyzed an absolute correlation coefficient of at least 0.8, suggesting these gene pairs are highly correlated, which are provided in Table 22 (0 indicates absolute correlation below 0.8 and 1 indicates absolute correlation of at least 0.8). We identified 10 gene pairs which were highly correlated in both the normal and adenoma but not in carcinoma samples (Figure 10B, Table 22). These pairs, unlike the carcinoma specific gene pairs, suggest coordination within the members of the Wnt signaling pathway, such as FZD, DVL and AXIN. In addition, these members are highly correlated with RAC, which is downstream of RAS oncogene (in the adenoma stage of the colorectal cancer pathway). The coordination between the members of the Wnt signaling pathway and RAC is supported by protein level interactions [10]. On the other hand, there were 38 gene pairs in the carcinoma samples, none of which were correlated in the normal or adenoma samples (Figure 10B, Table 22). These strongly correlated gene pairs may suggest direct or

indirect protein-protein or transcriptional interactions specific to the carcinoma stage of the colorectal cancer pathway. For example, the list included the gene pair TGFBR1 and TGFBR2, whose protein products are known to interact directly and play a role in advanced stages of colorectal cancer [12]. The presence of a strong correlation between these two genes suggests that the interaction of these two receptors may be relevant only in the carcinoma stage, for the samples we analyzed in this study. Another significantly correlated pair was MYC and MSH2 (Table 22). MYC is a transcriptional regulator of MSH2 [24]. Therefore, MYC driven regulation of MSH2 may be important in carcinoma but not in the normal or adenoma samples. In addition, BIRC5 was identified to be highly correlated with various other genes, such as ACVR1B and EGFR, suggesting that BIRC5 may also affect the regulation of ACVR1B and EGFR. While some of the strongly correlated gene pairs of the colorectal cancer pathway are supported by the literature, our analysis suggests there are others that may be correlated but currently are not supported by the literature. These results could suggest potentially novel direct or indirect interactions, or common regulations by upstream molecules. Since the analysis identified the colorectal cancer pathway to be significantly coherent only in the carcinoma samples, further experimental investigation of these correlations is needed to confirm these novel regulatory mechanisms for colorectal carcinoma.

#### **Mutation target analysis**

In addition to gene expression, the distribution of mutation targets of the colorectal cancer pathway members was compared with respect to their X and Y values. The pathway members with high X values had higher frequencies of mutation targets. The nuclear members (X=3) had the highest number of mutation targets, followed by the cytosolic members (X=2) and the receptor/ligand members (X=1) (Table 23). Most of the mutation targets were concentrated at

the later two stages of the pathway (Y=3, Y=4); covering the progression from intermediate adenoma to carcinoma, with the highest number of mutation targets found in the last stage, from late adenoma to carcinoma. In other words, colorectal cancer mutation targets were represented more in the late tumor stages of the KEGG pathway. This result is supported by the current view that mutations accumulate and increase as the tumor grows and cancer progresses [8]. Lastly, we analyzed whether the mutation targets were differentially expressed in the adenoma or carcinoma tissues as compared to the other pathway members, and found there was no statistically significant difference between the mutation targets and the other pathway members. This may be because that the expression and mutation data are sparse. Note that our list of mutations targets is a collection of several genes which were reported to be mutated in at least a single sample. Therefore only some of these mutation targets may actually be mutated in the samples from which the expression data were collected. Furthermore, we do not know whether these mutations are the cause or the effect of the gene expression level changes. Currently, not enough mutation and gene expression data are available to perform an extensive mutation analysis with respect to the X and Y groups.

Mutations alter important residues or domains of proteins that could lead to alteration in the binding properties of proteins to other proteins or regulatory DNA sequences, thus the levels of other genes and proteins may change. Previous reports have suggested an association between mutation events and changes in the gene expression levels [25, 26]. There is no evidence to suggest a correlation between the mutation of a gene and its own mRNA expression levels, however, a mutation in the TP53 gene has been shown to be correlated with its protein expression level in cancer [27]. This current study also did not find a strong relationship between a mutation target and its gene expression level for the colorectal cancer pathway members.

#### CONCLUSION

In this study, we analyzed expression and mutation data of the KEGG colorectal cancer pathway members. Previous studies focused predominantly on analyzing gene expression data for an entire signaling pathway and assessing whether the entire pathway is differentially expressed in a microarray dataset. Here we demonstrate an integrative analysis that investigates the distribution of expression values of the pathway members. Our analysis incorporated the location of the pathway members and found the expression values and the number of mutation targets varied depending on the cellular location (X) and stage of the cancer (Y) (Figure 9, Tables 19, 20). Previous studies found signaling and metabolic pathways to be coherent, and we show that the colorectal cancer pathway is coherent as well, depending on the stage at which the expression data was obtained. The members of the KEGG colorectal cancer pathway showed some degree of correlation in the expression profile of the carcinoma data (GSE4183).

This analysis has the potential to help uncover the roles of the different genes in the pathways in the progression of colorectal cancer. We were able to analyze only the colorectal cancer pathway in this study. However, we anticipate that as more information on cancer pathways and expression data for the various cancer stages becomes available, this approach to pathway analysis could be more widely applicable and may help contribute to our understanding of the similarities and differences in the progression of cancer in the different tissues. Similarly, this integrative analysis could be applied to analyze the progression of other diseases or biological processes.

APPENDIX

## APPENDIX

Table 18 ANOVA p values

Group	Sample set	P value
X=1, 2, 3	Adenoma	0.0005
X=1, 2, 3	Carcinoma	0.0058
Y=1, 2, 3, 4	Adenoma	0.0265
Y=1, 2, 3, 4	Carcinoma	0.6136

**Table 19** Pairwise t-test p values

Group 1	Group 2	Adenoma	Carcinoma
		p value	p value
X=1	X=2, 3	0.9470	0.2051
X=2	X=1,3	0.0038	0.0023
X=3	X=1, 2	0.0002	0.0149
Y=1	Y=2, 3, 4	0.0049	0.8826
Y=2	Y=1, 3, 4	0.1654	0.3116
Y=3	Y=1, 2, 4	0.0882	0.3050
Y=4	Y=1, 2, 3	0.7675	0.5209

 Table 20 Apoptosis and colorectal cancer pathway coherence at correlation coefficient of 0.5

Sample set	Apoptosis pathway	Colorectal cancer
	r ·····	pathway
Normal	0.021	0.255
Adenoma	0.182	0.244
Carcinoma	0.119	0.002

Pathway	Groups	Normal	Adenoma	Carcinoma
Colorectal	X=1	0.79	0.42	0.35
pathway	X=2	0.02	0.02	0.95
	X=3	0.91	0.72	0.21
	Y=1	0.84	0.94	0.19
	Y=2	0.21	0.07	0.92
	Y=3	0.34	0.07	0.18
	Y=4	1.00	1.00	1.00
Apoptosis	X=1	0.91	0.61	0.17
paniway	X=2	0.00	0.07	0.84

 Table 21 X/Y group coherence at correlation coefficient of 0.5
Gene ID 1	Gene Name 1	Gene	Gene Name 2	Normal	Adenoma	Carcinoma
91	ACVR1B	1956	FGFR	0	0	1
208	AKT2	4087	SMAD2	0	0	1
324	APC	5602	MAPK10	0	0	1
332	BIRC5	91	ACVR1B	0	0	1
332	BIRC5	324	APC	0	0	1
332	BIRC5	1956	EGFR	0	0	1
332	BIRC5	2956	MSH6	0	0	1
332	BIRC5	4436	MSH2	0	0	1
332	BIRC5	5602	MAPK10	0	0	1
332	BIRC5	6932	TCF7	0	0	1
332	BIRC5	8313	AXIN2	0	0	1
842	CASP9	6934	TCF7L2	0	0	1
2353	FOS	7046	TGFBR1	0	0	1
2956	MSH6	1630	DCC	0	0	1
2956	MSH6	8313	AXIN2	0	0	1
3845	KRAS	1630	DCC	0	0	1
3845	KRAS	2956	MSH6	0	0	1
3845	KRAS	6934	TCF7L2	0	0	1
4233	MET	5604	MAP2K1	0	0	1
4609	MYC	332	BIRC5	0	0	1
4609	MYC	4436	MSH2	0	0	1
4609	MYC	6932	TCF7	0	0	1
5880	RAC2	5293	PIK3CD	0	0	1
6654	SOS1	4436	MSH2	0	0	1
6654	SOS1	8322	FZD4	0	0	1
6655	SOS2	324	APC	0	0	1
7040	TGFB1	5159	PDGFRB	0	0	1
7043	TGFB3	23533	PIK3R5	0	0	1
7976	FZD3	5881	RAC3	0	0	1
8313	AXIN2	6932	TCF7	0	0	1
8322	FZD4	4436	MSH2	0	0	1
8326	FZD9	4087	SMAD2	0	0	1
83439	TCF7L1	332	BIRC5	0	0	1
83439	TCF7L1	4436	MSH2	0	0	1
83439	TCF7L1	5602	MAPK10	0	0	1
130399	ACVR1C	91	ACVR1B	0	0	1
130399	ACVR1C	1956	EGFR	0	0	1
130399	ACVR1C	5295	PIK3R1	0	0	1
1857	DVL3	5159	PDGFRB	0	1	0
1857	DVL3	10000	AKT3	0	1	0

**Table 22** Correlated pairs (with absolute correlation coefficient level of 0.8 as the threshold) in the colorectal cancer pathway for normal, adenoma and carcinoma samples

### Table 22 (cont'd)

5296	PIK3R2	23533	PIK3R5	0	1	0
5879	RAC1	1857	DVL3	0	1	0
5879	RAC1	5881	RAC3	0	1	0
5879	RAC1	8321	FZD1	0	1	0
5879	RAC1	8324	FZD7	0	1	0
5881	RAC3	23533	PIK3R5	0	1	0
7040	TGFB1	1630	DCC	0	1	0
8313	AXIN2	5159	PDGFRB	0	1	0
8323	FZD6	10000	AKT3	0	1	0
207	AKT1	1857	DVL3	1	0	0
207	AKT1	5881	RAC3	1	0	0
208	AKT2	207	AKT1	1	0	0
208	AKT2	1857	DVL4	1	0	0
208	AKT2	5291	PIK3CB	1	0	0
208	AKT2	5881	RAC3	1	0	0
1857	DVL3	5291	PIK3CB	1	0	0
1857	DVL3	8321	FZD1	1	0	0
7157	TP53	8323	FZD6	1	0	0
7157	TP53	8324	FZD7	1	0	0
8313	AXIN2	207	AKT1	1	0	0
8313	AXIN2	208	AKT2	1	0	0
8313	AXIN2	7157	TP53	1	0	0
8313	AXIN2	8321	FZD1	1	0	0
8313	AXIN2	8323	FZD6	1	0	0
8313	AXIN2	8324	FZD7	1	0	0
1857	DVL3	5881	RAC3	1	1	0
1857	DVL3	8323	FZD6	1	1	0
7157	TP53	5881	RAC3	1	1	0
8313	AXIN2	1857	DVL5	1	1	0
8313	AXIN2	5881	RAC3	1	1	0
8321	FZD1	5881	RAC3	1	1	0
8321	FZD1	8324	FZD7	1	1	0
8323	FZD6	5881	RAC3	1	1	0
8323	FZD6	8324	FZD7	1	1	0
8324	FZD7	5881	RAC3	1	1	0

Group	Average
	mutation
	frequency
X=1	0.0455
X=2	0.1702
X=3	0.3333
Y=1	0.0833
Y=2	0.0789
Y=3	0.3333
Y=4	0.5714

 Table 23 Mutation frequency values of X/Y groups

Figure 7 The ratio of carcinoma expression ratio to the adenoma expression ratio given for the maximum and minimum two genes and the average



**Figure 8** X/Y-dependent analysis of colorectal cancer pathway gene expression levels (A) Comparison of expression ratio values of X = 2 to other X groups in adenoma. (B) Comparison of expression ratio values of X = 2 to other X groups in carcinoma. (C) Comparison of expression ratio values of X = 3 to other X groups in adenoma. (D) Comparison of expression ratio values of X = 3 to other X groups in carcinoma. (E) Comparison of expression ratio values of Y = 1 to other Y groups in adenoma. (F) Comparison of expression ratio values of Y = 1 to other Y groups in carcinoma.

A



### Adenoma

B



Carcinoma

С



## Adenoma

D



# Carcinoma

E



## Adenoma

F



# Carcinoma

**Figure 9** Pathway correlation and cumulative fraction distributions (A) Pearson's correlation coefficient distributions among members of apoptosis, colorectal cancer, oxidative phosphorylation, and proteasome pathways shown together with 10 random selections of size 84 for normal colorectal tissue. (B–D) Cumulative absolute value fractions of correlation coefficient distributions among members of apoptosis, colorectal cancer, oxidative phosphorylation, and proteasome pathways shown together with 10 random selections of size 84 for (B) normal colorectal tissue, (C) colorectal adenoma, and (D) colorectal carcinoma.

A



### Normal Colorectal Tissue

B



## Normal Colorectal Tissue

r



## **Colorectal Adenoma**

r



## **Colorectal Carcinoma**

r

**Figure 10** Vend diagrams of correlated pairs of genes in colorectal normal, adenoma and carcinoma samples (A) Correlated pairs with at least 0.5 absolute correlation coefficient. (B) Correlated pairs with at least 0.8 absolute correlation coefficient.

#### A



### B



BIBLIOGRAPHY

#### BIBLIOGRAPHY

- 1. Fearon, E.R., and Vogelstein, B., A genetic model for colorectal tumorigenesis. Cell 1990, 61, 759-767
- 2. Cardoso, J., Boer, J., Morreau, H., Fodde, R., Expression and genomic profiling of colorectal cancer. Biochim Biophys Acta. 2007, 1775, 103-37
- 3. Edelman, E.J., Guinney, J., Chi, J.T., Febbo, P.G., Mukherjee, S. Modeling Cancer Progression via Pathway Dependencies. PLoS Comput Biol. 2008, 4, e28
- 4. Kanehisa, M., Goto, S., KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000, 28, 27-30
- Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F. et al., From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 2006, 34, D354-357
- 6. Kanehisa, M., Araki, M., Goto, S., Hattori, M. et al., KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008, 36, D480-D484
- 7. Dreesen, O., Brivanlou, A.H., Signaling Pathways in Cancer and Embryonic Stem Cells, Stem Cell Rev. 2007, 3, 7-17
- 8. Grady, W.M., Genomic instability and colon cancer. Cancer Metastasis Rev. 2004, 23, 11-27
- 9. Söreide, K., Janssen, E.A., Söiland, H., Körner, H., Baak, J.P., Microsatellite instability in colorectal cancer. Br J Surg. 2006, 93, 395-406
- 10. Behrens, J., The role of the Wnt signalling pathway in colorectal tumorigenesis. Biochem Soc Trans. 2005, 33, 672-5
- 11. Mehlen, P., Fearon, E.R., Role of the dependence receptor DCC in colorectal cancer pathogenesis. J Clin Oncol. 2004, 22, 3420-8
- 12. Roman, C., Saha, D., Beauchamp, R., TGF-beta and colorectal carcinogenesis. Microsc Res Tech. 2001, 52, 450-7
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S. et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005, 102, 15545-50

- Li, Z., Srivastava, S., Findlan, R., Chan, C., Using dynamic gene module map analysis to identify targets that modulate free fatty acid induced cytotoxicity. Biotechnol Prog. 2008, 24, 29-37
- 15. Li, Z., Srivastava, S., Yang, X., Mittal, S. et al., A hierarchical approach employing metabolic and gene expression profiles to identify the pathways that confer cytotoxicity in HepG2 cells. BMC Syst Biol. 2007, 11, 1-21
- 16. Breslin, T., Krogh, M., Peterson, C., Troein, C. Signal transduction pathway profiling of individual tumor samples. BMC Bioinformatics 2005, 6, 163
- 17. Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A. et al., PGC-1alpharesponsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003, 34, 267-273
- 18. Jansen, R., Greenbaum, D., Gerstein, M., Relating whole-genome expression data with protein-protein interactions. Genome Res. 2002, 12, 37-46
- 19. Hannenhalli, S., Levy, S., Transcriptional regulation of protein complexes and biological pathways. Mamm Genome. 2003, 14, 611-9
- 20. Yang, H.H., Hu, Y., Buetow, K.H., Lee, M.P., A computational approach to measuring coherence of gene expression in pathways. Genomics 2004, 84, 211-7
- 21. Draghici, S., Khatri, P., Tarca, A.L., Amin, K. et al., A systems biology approach for pathway level analysis. Genome Res. 2007, 17, 1537-45
- 22. Futreal, P.A., Coin, L., Marshall, M., Down, T. et al., A census of human cancer genes. Nat Rev Cancer 2004, 4, 177-83
- 23. Wehler, T.C., Frerichs, K., Graf, C., Drescher, D. et al., PDGFRalpha/beta expression correlates with the metastatic behavior of human colorectal cancer: a possible rationale for a molecular targeting strategy. Oncol Rep. 2008, 19, 697-704
- 24. Menssen, A., Hermeking, H., Characterization of the c-MYC-regulated transcriptome by SAGE: identification and analysis of c-MYC target genes. Proc Natl Acad Sci USA 2002, 99, 6274-9
- 25. Whyte, D.B., Holbeck, S.L., Correlation of PIK3Ca mutations with gene expression and drug sensitivity in NCI-60 cell lines. Biochem Biophys Res Commun. 2006, 340, 469-75
- 26. Austinat, M., Dunsch, R., Wittekind, C., Tannapfel, A. et al., Correlation between betacatenin mutations and expression of Wnt-signaling target genes in hepatocellular carcinoma. Mol Cancer. 2008, 7, 21

27. Martinez-Delgado, B., Robledo, M., Arranz, E., Infantes, F. et al., Correlation between mutations in p53 gene and protein expression in human lymphomas. Am J Hematol. 1997, 55, 1-8

#### **CHAPTER 3**

# SPECIFIC INTERACTION NETWORK ANALYSIS FOR COLORECTAL CANCER INTRODUCTION

Previous studies used gene expression levels to derive condition specific networks from large scale protein-protein interaction (PPI) datasets [1-4]. The limitations of having static PPI datasets were somewhat mitigated with the integration of sample-specific gene expression level changes so that condition specific/dynamic, and possibly novel regulatory interactions can be obtained. Differential gene expression analysis was used for groups of genes in a close neighborhood in the human PPI network to derive modules which were shown to increase the classification accuracy of cancer [1, 3]. Biologically relevant and significant condition specific molecular mechanisms have been revealed in such studies. Integrative analyses of PPI data have mostly been based on modular analysis in which the human PPI network is assumed to have a modular structure and the condition specific modules were retrieved by estimating the activity of the modules from the gene expression alterations of the interacting pairs. With this approach, modules representing deregulated events, such as cell cycle, apoptosis, etc. were revealed [1, 3, 4].

Integrative analyses of PPI data use microarray datasets to provide the condition specific network information. PPI datasets provide the pairs of proteins that are likely to interact physically. Differentially expressed genes obtained from microarray datasets in an experimental or disease condition can provide additional support for the interaction between the genes at the protein level. These pairs of genes are specific to the conditions in the microarray dataset unlike the PPI dataset pairs. Therefore the specific gene pairs could play a role under one condition

specifically. Indeed, with the integration of condition specific information, the PPI networks become more suitable for predicting novel regulatory interactions and providing novel hypotheses for the experimental condition under investigation. Previous integrative studies using PPI data were limited to comparisons of gene expression levels between two conditions such as comparing tumor to normal samples, long survival to poor survival samples, or metastatic to non-metastatic samples [1, 3, 4]. However, the analysis of a multi-conditional large-scale gene expression dataset also provides useful information, such as identifying genes with switch-like behavior that were not easily uncovered with the analysis of a pair-wise dataset [5]. Analyzing the expression levels of a gene in a diverse condition space provides a better understanding of the specificity of the expression level for a particular condition. For example p53, a well-known gene mutated commonly in many cancer types, does not appear to have a very significant differential gene expression in a small dataset that compares only two conditions, but was identified in the large dataset comprising of multiple conditions [5].

Colorectal cancer is one of the leading cancer types based on the number of new cases and number of expected deaths [6]. Colorectal cancer shares many mutation targets with other cancers such as endometrial cancer and ovarian cancer, but very few drug targets with other cancer types [7]. The insufficient number of drug therapies suggests that further studies are necessary for colorectal cancer to find novel targets. Although certain molecular mechanisms such as the involvement of TGF-beta pathway in colorectal cancer is known [8]; there is a need for more and specific molecular mechanisms to be elucidated for colorectal cancer.

RNF43 is an E3 ubiquitin ligase that resides in the ER and nuclear membranes, however no ubiquitination target for RNF43 has been identified and thus far only auto-ubiquitination activity has been shown [9, 10]. RNF43 was found to be upregulated in colorectal cancer and is

also a tumor associated antigen marker for colorectal cancer [11]. There are peptide vaccine clinical trials of RNF43 for colorectal cancer [12, 13]. GR is a nuclear receptor for glucocorticoid hormones, primarily involved in maintaining homeostasis in response to stress. However, it has diverse roles in various cell types and under different conditions. For example GR can induce the killing of lymphocytes, and promote proliferation of endometrial and liver cells, etc. [14]. Upon binding of glucocorticoids, GR is phosphorylated (activated) and transported into the nucleus where it can 1) directly bind to the promoters of its targets genes, or 2) modulate the activity of other transcription factors [14]. GR expression levels and function vary among different colorectal cancer patient samples and different cell lines. GR was found to be epigenetically downregulated by promoter hypermethylation or absent at the protein level in some patient tissue derived colorectal cancer samples as well as in certain cell lines [15-19]. Some studies have suggested an apoptotic role for glucocorticoids, and have associated it, in part, to GR activity. In those cases, GR enhanced the activity of drugs like genistein. GR expression has been shown to correlate with pRB and p16 tumor suppressor protein expression levels [18, 20].

In our study, we used a multi-condition dataset to define the top ranked significantly differentially expressed genes in colorectal cancer, which were found significantly associated with colorectal cancer in the literature. We constructed a network of these genes using the human protein-protein interaction data. Our analysis was not restricted to a modular structure, thus any gene could be linked to any other gene; thereby increasing the number of potential gene pairs. The ring finger protein 43 (RNF43), which is upregulated in colorectal cancer, is linked to the glucocorticoid receptor (GR, NR3C1), which is downregulated. In the HCT116 colorectal cancer

cell line, modulation of GR and RNF43 levels by siRNA and glucocorticoid treatment supported the link between these two proteins.

#### **RESULTS AND DISCUSSION**

#### **Colorectal cancer specific genes**

Unlike previous studies that defined specific genes based on the gene expression level changes by comparing tumor to normal samples, long survival to poor survival samples, or metastatic to non-metastatic samples, etc. [1, 3, 4], we defined the specific genes by differential expression of genes in colorectal cancer samples compared to a large set of samples from diverse conditions, including normal samples from various tissues, different cancer samples, cell lines, as well as other disease samples, etc. In other words, in prior approaches the comparison is made between condition A and an "other" condition (i.e. condition B). In contrast, we compare a condition against many conditions (i.e. condition A against conditions B,C,D, where B,C,D all fall under the "other condition") to identify more specific targets.

We calculated the separation of the expression values by D value, which is a normalized absolute difference of the mean values between two populations [5]. We compared the lists of differentially expressed gene based on a comparison of the colon cancer samples with the normal samples vs. all samples minus the colorectal cancer samples, for their enrichment of colorectal cancer related publications in the literature. We searched NCBI Pubmed database for the Official Full Name together with 'Colorectal cancer' or 'Colon cancer' and counted the number of genes with at least one publication. We also performed the same calculation for a random set of genes with the same number of gene expression profiles. Comparison to all the samples always yielded more genes and more significant genes that are related to colorectal cancer than the comparison to only normal samples (Figure 11). The greater fluctuation in the fraction of relevant genes for the multiple comparison approach is because of the smaller size of the list.

As D has to be at least 2 for two distinct populations, we used it as a threshold to obtain a list of colorectal cancer specific genes which has distinct expression profiles in colorectal cancer samples compared to a large spectrum of samples and which were significantly associated with colorectal cancer in the literature. In this way, we collected a list of specific genes for colorectal cancer which have significantly distinct expression profile from the other samples of various conditions.

#### **Colorectal cancer specific network**

We aimed to find the most significant regulatory interactions or associations between different genes in colorectal cancer such that these genes are distinctly expressed in colorectal cancer samples as compared to not only non-cancer colorectal samples but also all other cancer and disease samples of various types and conditions. Therefore our objective was to obtain novel molecular mechanisms specific to colorectal cancer, which is not readily revealed with previous two-condition or pair-wise comparisons. We used colorectal cancer specific genes to construct a colorectal cancer specific network so that we could generate hypotheses on potentially novel molecular mechanisms of colorectal cancer. We used the direct and indirect PPI neighborhood to construct a colorectal cancer specific network (Figure 12). We defined the resulting network as the colorectal cancer specific network (Figure 12). The colorectal cancer specific network includes genes that are well associated with colorectal cancer in the literature. In this network, only NR3C1 has a D value greater than 2, when the pairwise approach is used. This shows that

this network would not be obtained by a pairwise differential expression approach. The network also provides potential regulatory mechanisms for colorectal cancer. For example, a previously identified interaction between the HOXB7 protein and MAPK pathway [21], was identified in the network.

Moreover, we aimed to confirm the distinct expression levels of the members of the colorectal cancer specific network in an independent expression dataset. We used a patient derived paired normal and adjacent tissue sample set to calculate the D values. In the colorectal cancer specific network, only NR3C1 (GR) and RNF43 have D values greater than 2 in this independent dataset.

#### **GR-RNF43** regulation in colorectal cancer

NR3C1 (glucocorticoid receptor, GR) is the only gene in the colorectal cancer specific network that also had a high D value when the colorectal cancer samples were compared pairwise to normal samples. While NR3C1 was downregulated, one of its neighbors in the network, RNF43 was upregulated.

To validate a possible regulatory mechanism between these two proteins, we knocked down GR and RNF43 levels by specific siRNAs in HCT116 colorectal cancer cell lines. While knock-down of GR increased RNF43 mRNA levels, knock-down of RNF43 levels did not change the GR mRNA levels (Figure 13). This result suggests a possible transcriptional regulation of RNF43 by GR, therefore we induced GR with dexamethasone and measured the RNF43 mRNA levels. The induction of GR by dexamethasone led to decreased RNF43 mRNA levels that is abolished when GR is simultaneously silenced by siRNA treatment (Figure 13), suggesting that the effect of dexamethasone is specific to GR. This result suggests that GR is

transcriptional repressor of the RNF43 gene. Furthermore, GR has a binding site at 138 kb upstream of RNF43 transcription start site in lung carcinoma cells (FDR=5.9%) [22]. The position of this binding site is consistent with the results of other genes that are negatively regulated by GR, where GR binding is at distant sites (with a median of 146kb in contrast to the median of 11kb for positively regulated genes) from the transcription start site on the promoter of these genes. When GR was silenced, RNF43 protein levels were also upregulated (Figure 14). RNF43 protein levels were also downregulated in response to dexamethasone treatment in HCT116 colorectal cancer cells (Figure 14). These results confirm the negative regulation of RNF43 by GR. On the other hand, GR protein levels, unlike the mRNA levels, were downregulated in response to RNF43 silencing (Figure 14). Therefore there could be a positive regulation of GR by RNF43, which might be mediated by the ubiquitin ligase activity of RNF43 towards a negative regulator of GR.

#### CONCLUSION

In contrast to current approaches that identify targets by comparing pair-wise control (i.e. normal) to treated (i.e. disease) or across samples, we compared colorectal cancer samples not only to a control sample set but against a wide variety of samples and conditions. We were able to identify more specific genes for colorectal cancer which are significantly associated with colorectal cancer in the literature. We constructed a colorectal cancer-specific network based on the expression levels of the neighboring genes obtained from the human protein-protein interaction network. We identified a potential negative relationship between glucocorticoid receptor (GR) and ring finger protein 43 (RNF43) which may play a role in colorectal cancer. In HCT116 colorectal cancer cell line, knocking-down GR $\alpha$  levels with siRNA resulted in increased RNF43 levels and inducing the colorectal cancer cells with dexamethasone, which is

an activating ligand for GR, resulted in decreased RNF43 levels. On the other hand, knockingdown RNF43 levels with siRNA resulted in decreased GR $\alpha$  levels. Our study suggests GR negatively regulates RNF43 whereas there is no such a negative regulation from RNF43 to GR, indeed there could be a positive regulation.

#### MATERIALS AND METHODS

#### **Transcriptome and Interactome data**

We collected the expression data from an integrated multi-condition human microarray dataset (E-TABM-185) [23] from the ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) and integrated GSE24514 [24] from NCBI GEO database (http://www.ncbi.nlm.nih.gov/geo/) as there were not many colorectal tissue samples. After combining the raw data, we re-normalized the dataset by GCRMA [25]. We also used another expression data of a set of colorectal cancer patient-derived paired normal and tumor samples (GSE18105; GSM452629- GSM452662) [26] from NCBI GEO database. We used the human protein-protein interaction data from I2D database (Version 1.9) (http://ophid.utoronto.ca/ophidv2.201/), which includes experimentally confirmed as well as computationally predicted interactions [27, 28].

#### Determining the colorectal cancer specific gene list

We used the separation value (D value) to quantify the distinction of the expression values between sample sets similar to [5]. We calculated the D value between the colorectal cancer samples and the non-cancer colorectal samples as well as all the samples except the colorectal cancer samples. We tested the significance of the D values by permutation test. We also calculated Mann-Whitney test p-values. We used each affymetrix ID individually when calculating the D values and Mann-Whitney test p-values, which corresponded to different gene expression profiles. We used the Entrez Gene IDs matching to the affymetrix IDs to access the Official Full Name of the gene. In order to obtain the colorectal cancer associated genes, we used this full name to search Pubmed (http://www.ncbi.nlm.nih.gov/pubmed/) for this gene together with 'Colorectal cancer' or 'Colon cancer' using Bipython [29]. We collected the fraction of the number of genes (each distinct Gene ID) that was found to have at least one publication when searched together with 'Colorectal cancer' or 'Colon cancer' or 'Colon cancer' in Pubmed and divided by the number of genes (the number of distinct Gene ID s). We also obtained a random set of genes with the same number of distinct gene expression profiles (affymetrix IDs) and performed the same calculation for the random gene lists to obtain a random distribution for the colorectal cancer associated genes.

We used indirect interaction neighbors to construct the network, since it has been shown that they have significant functional similarity [30]. We integrated the colorectal cancer specific genes, which have a D value of at least two when colorectal cancer samples were compared to a multiple condition set of samples, with the first and second degree neighborhood in the parent network which is the human protein-protein interaction network from I2D database.

#### Cell culture

HCT116 colorectal carcinoma cell line (ATCC, Manassas, VA, USA) was cultured in McCoy's media (ATCC, Manassas, VA, USA) with 10% fetal bovine serum and 1% penicillin/streptomycin and maintained at 37°C and 5% CO<sub>2</sub>.

#### Quantitative real-time polymerase chain reaction

mRNA from HCT116 cells were isolated with the RNA isolation kit (Qiagen, Valencia, GA, USA). cDNA synthesis from mRNA was done with the cDNA synthesis kit (Bio-Rad, Hercules, CA, USA). For specific mRNA quantification by PCR, based on SYBR Green detection (Bio-Rad, Hercules, CA, USA) the following primers (Operon, Huntsville, AL, USA) were used; GR-alpha forward (5'-TCAGTTCCTAAGGACGGTCTG-3'), GR-alpha reverse (5'-CCACTTCATGCATAGAATCCAA-3'), RNF43 forward (5'-

CAAATTCACAGCCAGTGTGG-3'), RNF43 reverse (5'- GCTCCTCGAGTTCCTCCTCT-3'). The PCR cycle threshold values were based on the MyIQ software.

#### Western Blot

HCT116 cells were washed twice with cold PBS and lysed in RIPA lysis buffer with protease inhibitor. The cell lysate was centrifuged at 8000 rcf for 10 min, and the supernatant was collected. Total protein levels were quantified by BCA assay kit from Pierce Inc. (Rockford, IL, USA). A total of 20-40 g of total protein was resolved by SDS-PAGE and transferred to nitrocellulose membranes, and probed with primary antibodies with RNF43 antibody (ab84125 from Abcam), GRα (ab3580 from Abcam), and anti-beta-actin (Sigma-Aldrich). Biotinylated protein ladders (Cell Signaling, Beverly, MA, USA) were used together with the samples antibiotin antibody was used to detect the protein ladders on the western blots. Primary antibody incubation was done at 4 °C for overnight, secondary anti-rabbit and anti-mouse antibody (Pierce Biotechnology Inc) incubation was done for 1 hour at room temperature. Antibody detection was performed using the enhanced chemiluminescence kit from Pierce Biotechnology and imaged on the Molecular Imager ChemiDoc XRS System from Bio-Rad.

#### Dexamethasone and siRNA treatment

Dexamethasone (Sigma-Aldrich, St.Louis, MO, USA) was dissolved in ethanol and used as 250nM in HCT116 cell culture. Validated siRNA s for GR-alpha and RNF43 (Applied Biosystems, Carlsbas, CA, USA) and a scrambled negative control siRNA was transfected by reverse-transfection using Lipofactemine RNAiMAX (Invitrogen, Grand Island, NY, USA) for 48-72 hours. APPENDIX

#### APPENDIX

**Figure 11** Literature comparison of differentially expressed genes in colorectal cancer samples with respect to only non-cancer colorectal samples (pairwise) vs. all other samples (multiple) (A) Fraction of genes that are relevant to colorectal cancer in Pubmed database when different D value cut-offs around 2 are chosen. (B) Significance of the genes that are relevant to colorectal cancer in Pubmed database when different D value cut-offs around 2 are chosen. (B) Significance of the genes that are relevant to colorectal cancer in Pubmed database when different D value cut-offs around 2 are chosen, based on permutation test. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

A



Figure 11 (cont'd)





**Figure 12** Construction of the colorectal cancer specific network (A) Network construction method based on both direct and indirect protein-protein interactions, which are integrated with differentially expressed genes with D>2. (B) Colorectal cancer specific network



Figure 12 (cont'd)

### B


**Figure 13** Modulation of mRNA levels of RNF43 and GR $\alpha$  (A) All values were first normalized to the beta-actin levels. Specific siRNA treatment of RNF43 and GR was done in HCT116 colorectal cancer cell lines and normalized to the scrambled control siRNA treatment (B) HCT116 cells treated with 250nM dexamethasone, dissolved in EtOH, were compared with control cells treated with EtOH (less than 0.1%) treated cells. All values were first normalized to the beta-actin levels. Dexamethasone and GR siRNA treated HCT116 cells were normalized with EtOH and negative control siRNA treated cells. Real-time PCR results of primers specific for RNF43 and GR $\alpha$  are normalized with the results from  $\beta$ -actin primers. Fold change values are based on 9 values representing the comparison of all of the 3 treatment replicates to 3 control replicates. \* indicates p  $\leq 0.06$  based on the comparison of expression values of the treatment replicates.



A





**Figure 14** Modulation of protein levels of RNF43 and GRα (A) 250nM dexamethasone, dissolved in EtOH, treated HCT116 cells were compared with EtOH (less than 0.1%) treated cells. Dexamethasone and GR siRNA treated HCT116 cells were compared with EtOH and negative control siRNA treated cells. Western blot with RNF43 antibody (ab84125 from Abcam) (B) Specific siRNA treatment of RNF43 and GR was done in HCT116 colorectal cancer cell lines and compared to the negative control siRNA treatment. (C) Quantitated value for the RNF43 and GR western blot images in siRNA experiments by ImageJ (D) Quantitated value for the RNF43 and GR western blot images in Dexamethasone experiments by ImageJ. siRNA treatments are normalized with negative control siRNA treatment and dexamethasone treatments are normalized with EtOH treatments.



A

Figure 14 (cont'd)

С



GR siRNA

## RNF43 siRNA

Figure 14 (cont'd)

D



BIBLIOGRAPHY

## BIBLIOGRAPHY

- 1. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics. 2002, 18 Suppl 1:S233-40
- 2. de Lichtenberg U, Jensen LJ, Brunak S, Bork P. Dynamic complex formation during the yeast cell cycle. Dynamic complex formation during the yeast cell cycle. Science. 2005, 307(5710):724-7
- 3. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol. 2007, 3:140
- 4. Ahn J, Yoon Y, Park C, Shin E, Park S. Integrative gene network construction for predicting a set of complementary prostate cancer genes. Bioinformatics. 2011, 27(13):1846-53
- 5. Wu M, Liu L, Chan C. Identification of novel targets for breast cancer by exploring gene switches on a genome scale. BMC Genomics. 2011, 12:547
- 6. American Cancer Society. (2012) Cancer Facts & Figures. Atlanta: American Cancer Society.
- 7. Dalkic E, Wang X, Wright N, Chan C. Cancer-drug associations: a complex system. PLoS One. 2010, 5(4):e10031
- 8. Lampropoulos P, Zizi-Sermpetzoglou A, Rizos S, Kostakis A, Nikiteas N, Papavassiliou AG. TGF-beta signalling in colon carcinogenesis. Cancer Lett. 2012, 314(1):1-7
- 9. Yagyu R, Furukawa Y, Lin YM, Shimokawa T, Yamamura T, Nakamura Y. A novel oncoprotein RNF43 functions in an autocrine manner in colorectal cancer. Int J Oncol. 2004, 25(5):1343-8
- Sugiura T, Yamaguchi A, Miyamoto K. A cancer-associated RING finger protein, RNF43, is a ubiquitin ligase that interacts with a nuclear protein, HAP95. Exp Cell Res. 2008, 314(7):1519-28
- 11. Uchida N, Tsunoda T, Wada S, Furukawa Y, Nakamura Y, Tahara H. Ring finger protein 43 as a new target for cancer immunotherapy. Clin Cancer Res. 2004, 10(24):8577-86
- 12. Yoshimatsu K, Yokomizo H, Osawa G, Fujimoto T, Otani T, Tsunoda T, Nakamura Y, Ogawa K. Phase I study of combination therapy with peptide vaccine and anti-cancer drug for colorectal cancer]. Gan To Kagaku Ryoho. 2008, 35(12):2268-70

- 13. Yasuda S, Tsuchiya I, Okada K, Tanaka A, Suzuki T, Sadahiro S, Takeda K, Yamamoto S, Nakui M. Significant clinical response of advanced colon cancer to Peptide vaccine therapy: a case report. Tokai J Exp Clin Med. 2012, 37(2):57-61
- 14. Oakley RH, Cidlowski JA. Cellular processing of the glucocorticoid receptor gene and protein: new mechanisms for generating tissue-specific actions of glucocorticoids. J Biol Chem. 2011, 286(5):3177-84
- 15. Lind GE, Kleivi K, Meling GI, Teixeira MR, Thiis-Evensen E, Rognum TO, Lothe RA. ADAMTS1, CRABP1, and NR3C1 identified as epigenetically deregulated genes in colorectal tumorigenesis. Cell Oncol. 2006, 28(5-6):259-72
- 16. Ahlquist T, Lind GE, Costa VL, Meling GI, Vatn M, Hoff GS, Rognum TO, Skotheim RI, Thiis-Evensen E, Lothe RA. Gene methylation profiles of normal mucosa, and benign and malignant colorectal tumors identify early onset markers. Mol Cancer. 2008, 7:94
- Lien HC, Lu YS, Shun CT, Yao YT, Chang WC, Cheng AL. Differential expression of glucocorticoid receptor in carcinomas of the human digestive system. Histopathology. 2008, 52(3):314-24
- 18. Theocharis S, Kouraklis G, Margeli A, Agapitos E, Ninos S, Karatzas G, Koutselinis A. Glucocorticoid receptor (GR) immunohistochemical expression is correlated with cell cycle-related molecules in human colon cancer. Dig Dis Sci. 2003, 48(9):1745-50
- 19. Alford TC, Do HM, Geelhoed GW, Tsangaris NT, Lippman ME. Steroid hormone receptors in human colon cancers. Cancer. 1979, 43(3):980-4
- Park JH, Oh EJ, Choi YH, Kang CD, Kang HS, Kim DK, Kang KI, Yoo MA. Synergistic effects of dexamethasone and genistein on the expression of Cdk inhibitor p21WAF1/CIP1 in human hepatocellular and colorectal carcinoma cells. Int J Oncol. 2001, 18(5):997-1002
- Liao WT, Jiang D, Yuan J, Cui YM, Shi XW, Chen CM, Bian XW, Deng YJ, Ding YQ. HOXB7 as a prognostic factor and mediator of colorectal cancer progression. Clin Cancer Res. 2011, 17(11):3569-78
- 22. Reddy TE, Pauli F, Sprouse RO, Neff NF, Newberry KM, Garabedian MJ, Myers RM. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. Genome Res. 2009, 19(12):2163-71
- 23. Parkinson H, Kapushesky M, Kolesnikov N, Rustici G, Shojatalab M, Abeygunawardena N, Berube H, Dylag M, Emam I, Farne A, Holloway E, Lukk M, Malone J, Mani R, Pilicheva E, Rayner TF, Rezwan F, Sharma A, Williams E, Bradley XZ, Adamusiak T, Brandizi M, Burdett T, Coulson R, Krestyaninova M, Kurnosov P, Maguire E, Neogi SG, Rocca-Serra P, Sansone SA, Sklyar N, Zhao M, Sarkans U, Brazma A. ArrayExpress

update--from an archive of functional genomics experiments to the atlas of gene expression. Nucleic Acids Res. 2009, 37(Database issue):D868-72

- 24. Alhopuro P, Sammalkorpi H, Niittymäki I, Biström M, Raitila A, Saharinen J, Nousiainen K, Lehtonen HJ, Heliövaara E, Puhakka J, Tuupanen S, Sousa S, Seruca R, Ferreira AM, Hofstra RM, Mecklin JP, Järvinen H, Ristimäki A, Orntoft TF, Hautaniemi S, Arango D, Karhu A, Aaltonen LA. Candidate driver genes in microsatellite-unstable colorectal cancer. Int J Cancer. 2012, 130(7):1558-66
- 25. Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. Journal of the American Statistical Association 2004, 99(468): 909-917
- 26. Matsuyama T, Ishikawa T, Mogushi K, Yoshida T, Iida S, Uetake H, Mizushima H, Tanaka H, Sugihara K. MUC12 mRNA expression is an independent marker of prognosis in stage II and stage III colorectal cancer. Int J Cancer. 2010, 127(10):2292-9
- 27. Brown KR, Jurisica I. Online predicted human interaction database. Bioinformatics. 2005, 21(9):2076-82
- 28. Brown KR, Jurisica I. Unequal evolutionary conservation of human protein interactions in interologous networks. Genome Biol. 2007, 8(5):R95
- 29. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009, 25(11):1422-3
- 30. Chua HN, Sung WK, Wong L. ing indirect neighbours and topological weight to predict protein function from protein-protein interactions. Bioinformatics. 2006, 22(13):1623-30

## **CONCLUSION**

Systems level analysis of biomedical relationships, systems medicine or systems biology, not only interactions between genes, proteins, metabolites but also diseases, drugs, drug targets, and mutation targets, is necessary for understanding the global characteristics of diseases such as cancer [1, 2, 3]. For example, the structure of the PPI networks is a specific structure which contains only a few hub proteins that connect a lot of proteins to each other; therefore, a random dysfunction of a member in the network is very unlikely to hit one of these very few hub proteins, thus making the network robust against random problems. This finding can be used in a perturbation analysis; for instance, proteins with different number of interactions could be targeted separately to examine the differences, i.e., hub proteins could be more important for the survival of the organism, or cancer mutation targets are more connected from the rest of the proteins in a human PPI network [4, 5]. Therefore, systems medicine and systems biology approaches are useful for obtaining important characteristics of the organization of the biomedical system which could be exploited for advancing our understanding of cancer and its therapy. As cancer is a major cause of death in USA and colorectal cancer is one of the leading types of cancer, we aimed to identify common or distinct network features of colorectal cancer together with the other cancer types such as the analysis of clinical data associations, molecular signaling pathways of cancers, and specific interaction networks.

Previously, global network analysis was done on disease-gene, drug-target, disease-drug associations, however, driving factors of the networks or topological properties of the networks was not done extensively. For example, some disease phenotypes were shown to be highly connected in a disease network study, but, whether the high connectivity is related to the high number of cases or deaths for those diseases was not investigated [6].

Firstly, we collected the cancer-drug associations to generate cancer networks and analyzed the correlation of the network properties with cancer death statistics as a likely factor affecting cancer drugs. The cancers with the highest number of FDA approvals and clinical trials are leukemia, lung cancer, lymphoma and breast cancer; however, only breast and lung cancers have high and significant weighted degree values in the FDA cancer network, which is based on the cancer-drug associations. Leukemia and lymphoma have high number of drug approvals and trials but they don't share most of these drugs with other cancers unlike breast and lung cancers. This implies that the drug therapy for leukemia and lymphoma are more specialized for their own cases. Furthermore, cancer drug approvals and clinical drug trials are correlated to each other; a cancer type with a high number of drug approvals is more likely to have a high number of clinical drug trials. However, FDA cancer network, based on the sharing of drugs in FDA approvals, is very different from the clinical trial cancer network, based on the allocation of drugs in clinical trials. Moreover, lung cancer, which is significantly connected in the FDA cancer network, is not significantly connected in the clinical trial cancer network, whereas, lymphoma and ovarian cancer, which are not significantly connected in the FDA cancer network, are significantly connected in the clinical trial cancer network. As a result, there is a significant difference between the FDA approval based drug sharing and the clinical trial based drug sharing. This raises questions regarding the cancer drugs that need to be addressed, such as, if there is a bias in favor of lung cancer in FDA drug approvals since it shares many drugs in FDA approvals but not in clinical trials, also what the reason could be for less drug sharing for ovarian and lymphoma in FDA approvals in contrast to clinical trials.

In order to understand some of these differences, we compared the edge weights for the FDA and clinical trial cancer networks for each cancer pair and found out that most pairs are

strongly connected in the clinical trial but not in the FDA cancer network. As a specific example, stomach and esophagus cancers share many clinical drug trials such as capecitabine, cisplatin, doxorubicin, etc. but not any FDA drug approvals. There could be various reasons for this; i.e., the clinical trials for these cancers may not be successful. Our data includes any kind of clinical drug trial which is completed. It should be investigated further if the FDA approved drugs of these two cancer types could be used for each other. This analysis provides guidance for future FDA approvals and clinical decision making. When a decision is going to be made for a certain drug to be approved for a certain cancer type, the neighborhood of that cancer type in the clinical trial cancer network could be searched for the FDA approved drugs of the neighbor cancer types. For example, the approved drugs of stomach cancer; docetaxel, fluorouracil, imatinib, and sunitinib, could be given priority for future drug approvals for esophagus cancer. Unquestionably, there will be other factors such as the suitability of the deliverance of these drugs for esophagus cancer, but this information could be used as an additional factor in clinical decision making.

We also constructed FDA cancer networks for different years and observed that while breast cancer is significant since earlier years, lung cancer is significantly connected only in later years. In addition, these two cancers are significantly more connected than the rest of the network only recently. This shows that clinical decisions for cancer drugs in different years can change the structure of this network. There are more cancer drug approvals for lung cancer in later years. This observation shows a bias in favor of lung cancer that needs to be monitored in the following years.

We tested the correlation of cancer death statistics, global and local lethality values, with the network degree values and compared them with the FDA approval or clinical trial numbers.

This showed us that while there is a significant correlation between the FDA cancer network degree values and the local lethality values, there is not a significant correlation between the FDA approval numbers and the local lethality values. On the other hand, while the FDA approval numbers are significantly correlated with the global lethality values, the FDA cancer network degree values are not correlated. This implies that global lethality of cancer could affect the number of FDA approved drugs, but not the FDA approved drug sharing. On the other hand, local lethality might influence the FDA approval drug sharing, but not the number of FDA approved drugs. There could be similar effect for the clinical trial drugs also.

Interestingly, we also observed exceptions for certain groups of cancer types; the reasons for which should be investigated further. For example, the potential effect of local lethality on the sharing of FDA approved drugs is not present for the most locally lethal cancers, pancreatic, liver and esophagus cancers. These cancers have very low overlap of FDA approved drugs with other cancers. While liver and lung cancers have only one common FDA approved drug, they have 13 common drugs out of total 32 drugs used in clinical trials for both. 5 of these clinical trial drugs are approved for lung cancer by FDA, but they are still in clinical trials for liver cancer. In the future, these drugs might be expected to be given priority for liver cancer.

We also showed the unlikeness of the drug target based relationships of cancer types with the mutation target based relationships. While lung and breast cancers are significantly connected with respect to drug targets, they are not significantly connected based on mutation targets. This might indicate that their relatedness to the other cancer is not very high based on molecular mechanisms but they are highly related to other cancers in drug approvals. On the other hand, colorectal, ovarian and brain cancers could be highly related to the other cancers in drug approvals but not in molecular mechanisms based on their drug target based and mutation

target based degree values. These results might guide the future drug trials and approvals of colorectal, ovarian, and brain cancers to look for overlaps between these 3 cancer types and other cancer types. This is based on the assumption that the mutation target information is complete and represents the molecular events underlying cancer. It should be noted that there could be other mutation events, as well as other molecular alterations such as the transcriptional regulation of genes, protein modifications, etc. On the other side, we only analyzed the clinical drugs for cancer types, which is also not complete as there are many other drugs in clinical trials which are not in our list. In addition, there can be other drugs used in patients than the list of FDA approval drugs. In this study, we only analyzed the drug therapy aspect of cancer therapy excluding radiotherapy, surgery, etc. There may be some cancer types, for which chemotherapy might not be as important as other cancer types which can impact some of our conclusions.

In our next study, we showed that cancer pathways which consist of various signaling pathways can be analyzed in combination with gene expression levels to investigate the coherence of the pathway. Coherence of a pathway is defined as having close expression levels among the members of a pathway [7]. While previous research has focused on mostly metabolic, cellular, signaling pathways for coherence, the analysis of cancer pathways was ignored. We showed that the KEGG colorectal cancer pathway is coherent in only the carcinoma stage but not the normal or adenoma stages. This implies the collective regulation of the colorectal cancer pathway genes in carcinoma. These genes belong to several different signaling pathways such as Wnt signaling and TGF-beta pathways, which are well-known to be involved in colorectal cancer [8, 9]. Therefore our study implies enhanced communication between these pathways in later stages. Networks of pathways or modules can be analyzed to have global descriptions underlying biological processes and diseases like cancer [10, 11, 12]. For example, analysis of a pathway

crosstalk network for gene expression changes in colorectal cancer metastasis revealed a module of cell cycle related pathways and a modules cell migration related pathways [11]. Therefore, the potential interaction between Wnt signaling, MAPK, TGF-beta pathways should be investigated further. They can either have common regulators or they could co-regulate each other. We found more support for the latter case. A positive crosstalk between Wnt signaling and MAPK pathways was shown to be involved in colorectal cancer recently [13]. There is also evidence for crosstalk between Wnt signaling and TGF-beta pathways, mediated by Smad or Dishevelled proteins [14, 15]. In the future, the crosstalk between these pathways specifically only in colorectal carcinoma but in colorectal adenoma should be confirmed. For example, the association between Smad and Dishevelled was shown to be induced by Wnt signaling and this association prevented ubiquitination and deregulation of Smad [15]. First of all, it is necessary to show this association in colorectal carcinoma cells and test for its absence in colorectal adenoma samples. Then the reasons for the absence in colorectal adenoma samples can be analyzed further. Elucidation of such mechanisms is critical for understanding the progression of the invasive carcinoma phenotype from the adenoma phenotype.

We observed that Wnt signaling pathway members are upregulated compared to the other members of the KEGG colorectal cancer pathway. This may be one explanation for the loss of coherence in adenoma. Wnt pathway was known to be involved in the early progression of colorectal cancer [8]. Thus, our study suggests that the regulation of pathways in cancer can be captured at the gene expression level.

We showed that there are a lot more pairs of genes in the colorectal cancer pathway that have correlated gene expression levels in carcinoma samples compared to normal or adenoma samples. As a result of this, we observed a significant coherence of the pathway in only

carcinoma stage. In normal and adenoma samples, most pairs of correlated genes are from Wnt signaling pathway; such as pairs of Dishevelled and Frizzled genes. In contrast, carcinoma samples have correlated pairs of genes from different pathways corresponding to the different sections of the colorectal cancer pathway, such as Tgfbr2 from TGF-beta pathway with Apc from Wnt signaling pathway, Kras with Met, etc. These pairs of genes might have direct or indirect protein-protein interactions or they might regulate each other at the transcriptional level directly or indirectly. They can also be regulated by some common factors. For example, Myc and Msh2 genes are correlated only in the carcinoma samples. This might be because Myc is a positive transcriptional regulator of Msh2 or they have an indirect protein-protein interaction between themselves mediated by Max [16, 17]. Therefore, it is necessary to confirm the transcriptional regulation of Msh2 by Myc or the Max-mediated interaction between them in colorectal carcinoma samples. Then, these can also be checked in colorectal adenoma samples so that any differential regulation between these two proteins can be tested. This kind of further studies might support our observation for more integrative behavior of the colorectal cancer pathway in carcinoma stage. Then, the correlated pairs, specific for carcinoma stage, might be important candidates for perturbation analysis to check for their effect on the cancer progression.

Lastly, we analyzed specific networks for colorectal cancer in order to identify novel mechanisms involved in colorectal cancer. We used a novel approach to define the differentially expressed genes in colorectal cancer. We compared the expression levels in colorectal cancer samples to a large set of samples from various sources, conditions, etc. (multiple comparison approach) rather than the classical way to compare the cancer samples to normal samples (pairwise approach). The top ranked genes with a highly separated expression patterns obtained by using the multiple comparison approach are more significantly associated with colorectal

cancer in the literature. This observation is based on counting the genes which gives at least one paper in the NCBI Pubmed database when their official names are searched together with 'Colorectal cancer' or 'Colon cancer'. This analysis might miss some genes which have articles associated with colorectal cancer. Also the number of papers for a gene is omitted. A gene with many articles associated with colorectal cancer might be more related to colorectal cancer than a gene with only one article. Overall, our analysis suggests that multiple comparison approach to obtain the differentially expressed genes gives more relevant lists of genes to colorectal cancer. This should also be tested for other cancers, diseases, etc.

Interestingly, there is only one gene common to the specific genes obtained by pairwise and multiple comparison approaches. This shows that the multiple comparison approach gives us also a very different set of genes. Our study does not imply that the pairwise comparison approach should be completely abandoned and the multiple comparison approach should be used alone. Instead, the two approaches should be combined as they capture genes with different expression patterns.

The colorectal cancer specific network is constructed by mapping the specific genes obtained by the multiple comparison approach to the direct and indirect neighborhoods in the human protein-protein interaction network. It can be used for generating hypotheses for novel molecular mechanisms for colorectal cancer. One such example, that is between HOXB7 and MAPK13, is confirmed in the literature. Protein level expression of HOXB7 was upregulated in advanced, metastatic, highly proliferative stages of colorectal cancer patient samples and the induction of HOXB7 expression in colorectal cancer cells induce MAPK pathway [18]. The other interactions in this network can be tested in the future experimentally to check for their involvement in colorectal cancer. We focused on glucocorticoid receptor (GR, NR3C1) and ring

finger 43 (RNF43) proteins in this network as they shows a high and significant separation of expression levels for colorectal cancer in an independent paired dataset. GR is downregulated while RNF43 is upregulated, which is also confirmed in most colorectal cancer cell lines including HCT116.

We knocked down GR levels in HCT116 cell line by specific siRNA and showed that both mRNA and protein levels of RNF43 increase in response to GR downregulation. We also treated HCT116 cell line with Dexamethasone, which is a glucocorticoid ligand for GR, in order to activate GR function. We observed a decline in RNF43 levels in response to Dexamethasone, but it is less significant than the change of RNF43 levels in response to GR knock-down. However, when we induced the HCT116 cells with both Dexamethasone and GR siRNA the decrease in RNF43 is abolished. Therefore the negative effect of Dexamethasone on RNF43 might be significant. However, the most significant observation for the decline of RNF43 is when the GR is silenced without Dexamethasone treatment. GR has some isoforms which do not need glucocorticoid ligands to get activated and are always active in the nucleus [19]. These isoforms might be responsible for the effect we observed without Dexamethasone treatment. As a result, GR might be a negative regulatory factor for RNF43 transcription. This is partly supported by a study in lung cancer, showing that GR regulates genes from distant sites [20]. GR has binding sites further away from the transcription start site (TSS), for the genes that it regulates negatively. Genes activated by GR are around 11 kb of TSS, whereas genes repressed by GR are around 146 kb of TSS. They showed that GR has a binding site at 138 kb upstream of RNF43 transcription start site in lung carcinoma cells [20]. This indicates that GR could be a negative regulator of RNF43 in lung cancer. It should be tested in the future, whether GR could

bind to the distant promoter site of RNF43 and negatively regulate the transcription of RNF43 in colorectal cancer.

Recently,  $\beta$ -catenin was found to be a positive regulator of RNF43 transcription, by direct binding to its promoter together with TCF4 in HCT116 colorectal cancer cell line [21]. RNF43 is induced by Wnt signaling pathway in colorectal cancer [22]. Wnt pathway is well-known to induce the progression of colorectal cancer [8]. RNF43 is also upregulated in colorectal cancer and knock-down of RNF43 suppresses the growth of colorectal cancer [23]. Therefore, it is possible that Wnt signaling pathway, as an early event in colorectal cancer, induces RNF43, which may in turn downregulate proteins like p53, or other unknown targets [24]. On the other hand, GR downregulates Wnt signaling pathway by direct inactivation of  $\beta$ -catenin [25]. By downregulating Wnt signaling pathway GR represses the targets of the pathway such as Cyclin D1, therefore it is likely that GR could also downregulate RNF43, another target of Wnt pathway, through inactivation of  $\beta$ -catenin. It is necessary to show the inhibition of  $\beta$ -catenin by GR in colorectal cancer. Then the downregulation of RNF43 by GR could be explained by the inactivation of  $\beta$ -catenin in colorectal cancer.

When we knocked-down RNF43 by siRNA, GR mRNA levels didn't change whereas there was a decline in protein levels, yet not statistically significant. This might show a positive feedback from RNF43 to GR, yet it needs to be supported by further experiments. For example, there might be a negative regulator of GR, which in turn might be a ubiquitination target of RNF43. However, the analysis of RNF43 function in HCT116 cell line is questionable, as a recent study showed that RNF43 is mutated in this cell line [26]. Interestingly RNF43 was shown to inhibit Wnt signaling pathway and this was shown to be mediated by the ubiquitination of Frizzled receptor. This observation is abolished in HCT116 colorectal cancer cell line as RNF43

is mutated, thus it cannot ubiquitinate Frizzled receptor. Therefore, it is necessary to analyze the possible effect of RNF43 on GR in a different colorectal cancer cell line.

Overall, we were able to show a potential mechanism between GR and RNF43 in colorectal cancer as the downregulation of GR could be involved in the upregulation of RNF43. This observation has important clinical implications, as there are successful vaccine therapy studies using RNF43 derived antigens in colorectal cancer patients [27]. The potential use of glucocorticoids in colorectal cancer patients together with the RNF43 antigens needs to be investigated. BIBLIOGRAPHY

## BIBLIOGRAPHY

- 1. Deisboeck TS, Kresh JY. Complex Systems Science in Biomedicine. Springer, New York, USA, 2006
- 2. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011, 12(1):56-68
- 3. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004, 5(2):101-13
- 4. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001, 411(6833):41-2
- 5. Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. Bioinformatics. 2006, 22(18):2291-7
- 6. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci U S A. 2007, 104(21):8685-90
- 7. Yang HH, Hu Y, Buetow KH, Lee MP. A computational approach to measuring coherence of gene expression in pathways. Genomics 2004, 84, 211-7
- 8. Burgess AW, Faux MC, Layton MJ, Ramsay RG. Wnt signaling and colon tumorigenesis--a view from the periphery. Exp Cell Res. 2011, 317(19):2748-58
- 9. Lampropoulos P, Zizi-Sermpetzoglou A, Rizos S, Kostakis A, Nikiteas N, Papavassiliou AG. TGF-beta signalling in colon carcinogenesis. Cancer Lett. 2012, 314(1):1-7
- 10. Wang X, Dalkic E, Wu M, Chan C. Gene module level analysis: identification to networks and dynamics. Curr Opin Biotechnol. 2008, 19(5):482-91
- 11. Li Y, Agarwal P, Rajagopalan D. A global pathway crosstalk network. Bioinformatics. 2008, 24(12):1442-7
- 12. Liu KQ, Liu ZP, Hao JK, Chen L, Zhao XM. Identifying dysregulated pathways in cancers from pathway interaction networks. BMC Bioinformatics. 2012, 13(1):126
- Jeong WJ, Yoon J, Park JC, Lee SH, Lee SH, Kaduwal S, Kim H, Yoon JB, Choi KY. Ras stabilization through aberrant activation of Wnt/β-catenin signaling promotes intestinal tumorigenesis. Sci Signal. 2012, 5(219):ra30
- 14. Letamendia A, Labbé E, Attisano L. Transcriptional regulation by Smads: crosstalk

between the TGF-beta and Wnt pathways. J Bone Joint Surg Am. 2001, 83-A Suppl 1(Pt 1):S31-9

- Mamidi A, Inui M, Manfrin A, Soligo S, Enzo E, Aragona M, Cordenonsi M, Wessely O, Dupont S, Piccolo S. Signaling crosstalk between TGFβ and Dishevelled/Par1b. Cell Death Differ. 2012, doi: 10.1038/cdd.2012.50
- Menssen A, Hermeking H, Characterization of the c-MYC-regulated transcriptome by SAGE: identification and analysis of c-MYC target genes. Proc Natl Acad Sci USA 2002, 99, 6274-9
- 17. Mac Partlin M, Homer E, Robinson H, McCormick CJ, Crouch DH, Durant ST, Matheson EC, Hall AG, Gillespie DA, Brown R. Interactions of the DNA mismatch repair proteins MLH1 and MSH2 with c-MYC and MAX. Oncogene. 2003, 22(6):819-25
- Liao WT, Jiang D, Yuan J, Cui YM, Shi XW, Chen CM, Bian XW, Deng YJ, Ding YQ. HOXB7 as a prognostic factor and mediator of colorectal cancer progression. Clin Cancer Res. 2011, 17(11):3569-78
- 19. Oakley RH, Cidlowski JA. Cellular processing of the glucocorticoid receptor gene and protein: new mechanisms for generating tissue-specific actions of glucocorticoids. J Biol Chem. 2011, 286(5):3177-84
- 20. Reddy TE, Pauli F, Sprouse RO, Neff NF, Newberry KM, Garabedian MJ, Myers RM. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. Genome Res. 2009, 19(12):2163-71
- Bottomly D, Kyler SL, McWeeney SK, Yochum GS. Identification of {beta}-catenin binding regions in colon cancer cells using ChIP-Seq. Nucleic Acids Res. 2010, 38(17):5735-45
- 22. Hao HX, Xie Y, Zhang Y, Charlat O, Oster E, Avello M, Lei H, Mickanin C, Liu D, Ruffner H, Mao X, Ma Q, Zamponi R, Bouwmeester T, Finan PM, Kirschner MW, Porter JA, Serluca FC, Cong F. ZNRF3 promotes Wnt receptor turnover in an R-spondinsensitive manner. Nature. 2012, 485(7397):195-200
- 23. Yagyu R, Furukawa Y, Lin YM, Shimokawa T, Yamamura T, Nakamura Y. A novel oncoprotein RNF43 functions in an autocrine manner in colorectal cancer. Int J Oncol. 2004, 25(5):1343-8
- 24. Shinada K, Tsukiyama T, Sho T, Okumura F, Asaka M, Hatakeyama S. RNF43 interacts with NEDL1 and regulates p53-mediated transcription. Biochem Biophys Res Commun. 2011, 404(1):143-7
- 25. Takayama S, Rogatsky I, Schwarcz LE, Darimont BD. The glucocorticoid receptor represses cyclin D1 by targeting the Tcf-beta-catenin complex. J Biol Chem. 2006,

281(26):17856-63

- 26. Koo BK, Spit M, Jordens I, Low TY, Stange DE, van de Wetering M, van Es JH, Mohammed S, Heck AJ, Maurice MM, Clevers H. Tumour suppressor RNF43 is a stemcell E3 ligase that induces endocytosis of Wnt receptors. Nature. 2012, 488(7413):665-9
- 27. Yasuda S, Tsuchiya I, Okada K, Tanaka A, Suzuki T, Sadahiro S, Takeda K, Yamamoto S, Nakui M. Significant clinical response of advanced colon cancer to Peptide vaccine therapy: a case report. Tokai J Exp Clin Med. 2012, 37(2):57-61