TECHNISCHE UNIVERSITÄT MÜNCHEN

Max-Planck-Institut für Biochemie

Systems and integrative approaches in oncogenomics

Gopinath Ganji

Vollständiger Abdruck der von der Fakultät für Chemie
der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigte Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Chr. S. W. Becker
Prüfer der Dissertation:	1. Priv.-Doz. Dr. N. Budisa
	2. Univ.-Prof. Dr. H. Kessler

Die Dissertation wurde am 10.12.2008 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 10.02.2009 angenommen.

"If one advances confidently in the direction of his dreams, and endeavors to live the life which he has imagined, he will meet a success unexpected in common hours. He will put some things behind, will pass an invisible boundary; new, universal, and more liberal laws will begin to establish themselves around and within him; or the old laws be expanded, and interpreted in his favor in a more liberal sense, and he will live wih the license of a higher order of beings. In proportion as he simplifies his life, the laws of the universe will appear less complex, and solitude will not be solitude, nor poverty poverty, nor weakness weakness. If you have built castles in the air, your work need not be lost; that is where they should be. Now put the foundations under them."

Henry David Thoreau – *Walden*.

Acknowledgements

A thesis advisor plays the most seminal role in the journey to a finished dissertation, but the true extent is rarely acknowledged. I am perpetually indebted to Axel and nothing I could say would ever measure up the gratitude, appreciation and respect I have for him today and always. He is the greatest mentor one could ask for. I've benefited tremendously not only from his keen intellect and refined wisdom, but also his boundless magnanimity.

Nediljko as my doctor father was the steering force behind this dissertation and I owe him tremendously for his support, encouragement and guidance. This thesis would not have been a reality without his involvement.

Very special thanks are in order for Lars who has been extremely generous in giving of his time and help to guide me through the process. He has been a buddy through thick and thin and it's been a slice knowing him. His expert Deutsch has saved me from getting in trouble with using *Google Translate* to frame my Zussamenfassung or translate the departmental forms!

Sushil and Kirti were the much needed catalysts that got me started and have done me the greatest favors I could ask for.

I am deeply indebted to all past and present Lilly colleagues for their technical assistance and camaraderie over the years. I've enjoyed technical and social interactions with Li Yue (statistics), Jaga (survival analysis), Yang (arrayCGH analysis) and Mahesh (annotation). Intellectual exchanges about cancer drug discovery with Kerry have been defining and inspiring. Ketan (who also presented me with the spectacular quote from *Walden*), Vinisha, Hai, Yang, Jude and Santosh have constantly prodded and cushioned me to make this happen.

My salutations to collaborators at TGen for a fruitful and enjoyable partnership. Quick Que and Holly Yin have grown to be great friends along the way. It's been an absolute privilege, but a heart wrenching loss to see Quick pass away this year due to terminal cancer. His memories will be cherished forever.

This list would be incomplete without mentioning Jason, Pooja, Subodh, Tariq and Mourad for just being there and making every rejection, mishap, challenge, trial and tribulation encountered during the course of my PhD journey seem trivial and momentary. Their reassurances have always driven me in the right direction.

Perhaps, the work of several researchers in my thesis and the availability of publicly available resources and tools deserve special mention since several people must've spent countless pain staking late hours to minimize my own blood, sweat and tears!

Above all, I owe everything to the unconditional love and undying support from my parents, grandmothers (who recently passed away due to terminal cancer and who I dedicate this work to), family and friends. They truly complete me.

Zusammenfassung

Auf Grund bedeutender technologischer Fortschritte konnte in der Vergangenheit auf molekularer Ebene ein systematisches Profiling von Krebs erstellt werden, wobei eine überwältigende Anzahl an Genomik-Daten (Oncogenomics) generiert wurde. Daraus ergibt sich ein Bedarf an innovativen und integrierten Ansätzen, die diese Reichhaltigkeit an Information in Wissen umwandeln. In der vorliegenden Dissertation wurden drei Fallstudien analysiert, die Hochdurchsatz-Datensätze wie z. B. RNAi-Screens, Mutation Profiling und Microarrays beinhalten. Durch das Kombinieren verschiedener Datensätze wurden Hypothesen erstellt und getestet, die zur Charakterisierung genetischer Determinanten in der Tumorbiologie und deren Relevanz für die Entwicklung neuer Medikamente dienen sollten. Die erzielten Ergebnisse identifizieren neue Gene, die in Zusammenhang mit Krebs stehen, geben Aufschluss über den Mechanismus der kürzlich entdeckten genetischen Fehlentwicklungen und führen zu rationellen therapeutischen Anwendungen, die nun in Labor und Klinik geprüft werden müssen. Die verwendeten globalen Ansätze sind vielversprechend und können erweitert werden, um unser Verständnis des „Onkogenoms“ zu verbessern. Außerdem bieten sie die Möglichkeit zur Entwicklung und Optimierung neuer bzw. bestehender Krebstherapien.

Table of Contents

1 Introduction

Advances in high throughput technologies such as large-scale sequencing and functional genomics have created a wealth of high resolution and high content information. The completion of several genome projects (including the Human Genome Project), uncovering protein-protein interaction networks, large scale knock-out/mutagenesis experiments, ever increasing molecular profiling and imaging experiments, construction of predictive models and generation of synthetic genomes are all a testament to a modern age of unprecedented information explosion that has shaped and continues to change the landscape of basic and applied biomedical research. Nowhere is this more apparent than in the field of oncology where large datasets have been generated and analyzed at various levels of molecular detail – genes, proteins, metabolites. Integration of such genome-wide datasets, aided by creative unconventional analysis, has begun to provide a systems level understanding of tumor biology. As a result, these powerful discoveries can be translated into clinical applications for better prevention, detection, diagnosis, prognosis and personalizing treatment for improved outcomes.

1.1 Cancer as a paradigm for systems analysis

Researchers at the Institute for Systems Biology (ISB, Seattle, WA) have nicely summarized the properties of biological systems that make them attractive for systems level exploration—emergent properties, robustness and modularity [1]. Emergence is a trait in which the whole is greater than the sum of the parts; robustness is characteristic of resilience to fluctuations in the immediate environment resulting from redundancy and control mechanisms; modularity is a phenomenon that explains the 'clustering' of parts into a functional or structural entity. Several aspects of cancer pathobiology make it particularly interesting for global investigations. A case in point for emergent properties is the accumulated genetic and epigenetic changes that collectively transform a normal cell into a cancer cell demonstrating the hallmarks of disease – self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of programmed cell death (apoptosis), limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis [2]. Robustness is a characteristic seen when tumors that are in initial remission after treatment frequently relapse and become resistant to anti-tumor therapy.

Modularity is manifested in how genetic aberrations drive disease progression such as an amplification of the EGFR genetic locus that triggers the ERK/MAPK cascade of downstream activation events leading to neoplasia. Taken together, these examples demonstrate that the cancer genome or 'oncogenome' provides a rich opportunity for large-scale systems and integrative analyses.

1.2 Systems level 'oncogenomic' profiling efforts

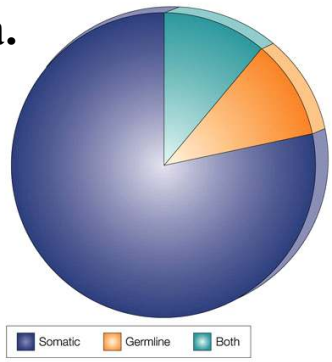
Great strides have been made scientifically and technologically in trying to dissect out the molecular 'parts list' of cancer genomes. A sampling of genomic surveys of resequencing, array profiling and RNAi screens will be presented below. Excellent reviews of systems approaches to understand epigenomic mechanisms in a global context can be found in [3-5]. These functional genomic approaches provide a top down view of the cancer system being investigated.

1.2.1 Genomic resequencing efforts

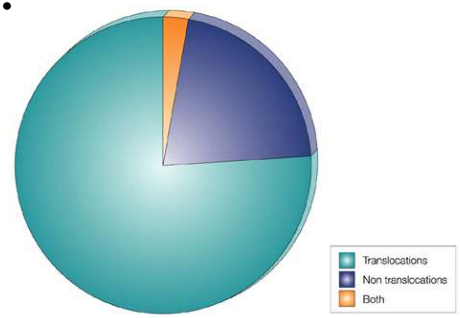
Cancer is a complex heterogeneous genetic disease that is acquired as a phenotypic consequence of the collective action of multiple genomic alterations. These can be broadly classified into growth promoting activation events (oncogenes) and growth inhibiting inactivating changes (tumor suppressors) (Figure 1). The dependence on such events drives the multi-step pathological process and is the basis of oncogene addiction. This has provided avenues for pharmacological inhibition as demonstrated by successful 'magic bullet' targeted drugs such as Herceptin® (also known as trastuzamab which targets Her2/Neu in breast cancer), Gleevec® (also known as imatinib which targets Bcr-Abl in CML, c-kit/PDGFRA in GIST) and EGFR inhibitors (gefitinib (Iressa®), erlotinib (Tarceva®), cetuximab (Erbitux®) or panitunimab (Vectibix®)). Furthermore, several papers [6-8] have prospectively analyzed retrospective data and shown that a subset of patients harboring EGFR mutations are responsive to EGFR specific tyrosine kinase inhibitors such as gefitinib and erlotinib. Therefore, undoubtedly, unbiased sequencing projects will yield valuable insights into the mechanisms of cancer and suggest novel means for disease treatment.

Automated sequencing has allowed systematic study of genetic alterations in cancer relevant gene families. Victor Velculescu's team carried the first ever analysis of the kinome [9] and phosphatome [10] in colorectal cancer. Among other genes (8 kinases, 6 phosphatases), their work pointed to PIK3CA as an oncogene that is frequently mutated in several human cancers [11] and that the PI3K pathway was the most frequently (~50%) mutated pathway in colorectal cancers [12] . From a more focused analysis by another group, BRAF mutations were discovered to be highly prevalent (66%) in malignant melanomas and relatively less common in other human cancers [13]. The Sanger group has carried out similar work to identify commonly altered kinases by sequencing 518 kinases in breast [14], lung [15] and testicular germ cell tumors [16]. Based on these studies and others, they generally found low frequencies of non-synonymous somatic mutations (e.g., 1 in 7 seminomas, ~40 in 26 primary lung neoplasms) with significant differences between individual cancers in the number (some having none) and pattern of mutations due to mutagen exposure, mutator phenotype or tissue of origin. They conclude that several mutations are likely to be 'passenger' or 'bystander' effects that do not contribute to tumorigenesis, but ~120 genes harbor 'driver' or 'causal' mutations [17].

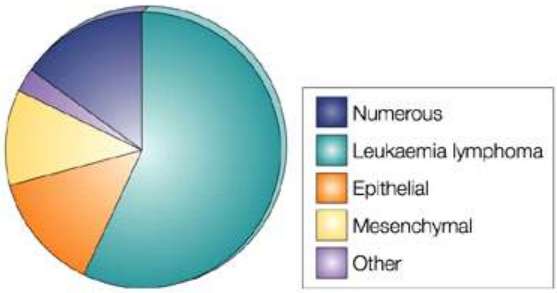
a.



b.



c.



d.

Gene (synonym)	Mutation type	Cancer type
ABL1 (ABL)	Translocation	Chronic myelogenous leukemia
AKT2	Amplification	Ovarian, breast
ALK	Translocation	Anaplastic large cell lymphoma
BAX	Inactivating codon change	Colon, stomach
BCL2	Translocation	Lymphomas
BCL6	Translocation	Lymphomas
BRAF	Activating codon change	Melanoma, colorectal, thyroid
CCND1 (cyclin D1)	Amplification, translocation	Leukemias, breast
CTNNB1 (-catenin)	Activating codon change	Colon, liver, medulloblastomas
EGFR	Amplification, activating codon change	Glioblastomas, non?small cell lung cancers
EPHB2	Inactivating codon change	Prostate
ERBB2	Amplification	Breast, ovarian
EVI1	Translocation	Leukemias
EWSR1	Translocation	Ewing's sarcomas, lymphomas, leukemias
FBXW7 (CDC4)	Inactivating codon change	Colon, uterine, ovarian, breast
FES	Activating codon change	Colon
FGFR1?3	Translocation	Lymphomas, gastric cancers, bladder cancers
FLT3, 4	Activating codon change	Leukemias, angiosarcomas
FOXO1A, 3A	Translocation	Rhabdomyosarcomas, leukemias
GLI	Amplification, translocation	Brain, sarcomas
HMGA2	Translocation	Lipomas
HOXA9, 11, 13; HOXC13, HOXD11, 13; HOX11, HOX11L2	Translocation	Leukemias
HPVE6	HPV infection	Cervical
HPVE7	HPV infection	Cervical
JAK2	Translocation	Leukemias
KRAS2, N-RAS	Activating codon change	Colorectal, pancreatic, non?small cell lung cancer
MAP2K4 (MKK4)	Inactivating codon change	Pancreas, breast, colon
MDM2	Amplification	Sarcomas
MLL	Translocation, activating codon change	Leukemias
MYC, MYCN, MYCL1	Amplification	Lymphomas, neuroblastomas, small cell lung cancers
NOTCH1	Translocation	Leukemias
NTRK1, 3	Translocation, activating codon change	Thyroid, secretory breast, colon
PDGFB	Translocation	Dermatofibrosarcomas and fibroblastomas
PDGFRB	Translocation	Leukemias
PI3KCA	Activating codon change	Colon, stomach, brain, breast
PTNP1, 11	Activating codon change	Leukemias, colon
RARA	Translocation	Promyelocytic leukemia
RUNX1	Translocation	Leukemias
SMAD2	Inactivating codon change	Colon, breast
SS18	Translocation	Synovial sarcomas
TAL1	Translocation	Leukemias
TFE3	Translocation	Kidney, sarcomas
TGFBR1, TGFBR2	Inactivating codon change	Colon, stomach, ovarian
TNFRSF6 (FAS)	Activating codon change	Lymphomas, testicular germ cell tumors

Figure 1. Summary of genetic alterations in human cancers. The Sanger group cataloged a census of 291 genes whose genetic alterations are widely studied in various human cancers [18]. ~90% harbor somatic (dark blue), ~20% germline (orange) and ~10% harbor both (light blue) types of mutations (a). (b)Majority of the reported somatic alterations involve translocations (light blue, e.g. ABL1, FLT3) as opposed to non-translocation events (dark blue, e.g. amplifications, missense mutations) or a combination (orange). (c) These cancer genes have been studied in a wide variety of indicated tumors, leukemias/lymphomas being the largest group. (d) A sampling of these genes which are somatically altered and not inherited (adapted from [19]). Oncogenes typically involve activating mutations, amplifications, translocations (except genes like RUNX1) while inactivating mutations or deletions occur in tumor suppressor genes. (a), (b), (c) were adapted from [18].

These results are broadly consistent in 2 subsequent consecutive publications [20, 21] by Vogelstein and colleagues. By sequencing 20, 857 cDNAs corresponding to 18, 191 genes in 11 breast and 11 colorectal cancers, they found that ~ 90 mutant genes make up an individual tumor genome, but only a handful of these form commonly mutated gene 'mountains' and several low frequency mutations form gene 'hills' (<5%). Out of these,

they suggested that ~80 mutations were non-consequential but <15 were 'driver' events that were responsible for initiation, progression or maintenance of disease. In total, they validated 183 genes in colorectal and 189 genes in breast cancers (on average, 11 per tumor) that were largely novel, affect diverse cellular functions and are frequently mutated. Although the numbers were similar for breast and colorectal, the actual genes and their patterns differed and no two tumors of the same type overlapped to a large extent which is likely due to tumor heterogeneity. Furthermore, they were able to cluster the large number of mutant genes into commonly altered pathways. Their findings suggest that the gene 'hill's and not 'mountains' dominate disease genetic landscape by providing collective incremental fitness advantage.

In a more recent study, the Ullrich lab carried out cDNA based sequencing of 254 established tumor lines, representing 19 different tissues, to identify 155 polymorphisms and 234 somatic mutations in 72MB of the tyrosine kinase gene family [22]. They found that the germ-line polymorphisms followed a Gaussian-like distribution with an average of 12.3 variations per cell line while somatic alterations were unevenly distributed. They did not find any somatic mutations in the tyrosine kinome of 119 cell lines which is in agreement with the low frequency of kinase mutations in breast [14], lung [15] and testicular germ cell [16] tumors seen by the Sanger group. On the other hand, 9-14 somatic mutations were detected in LNCaP, Jurkat (T-cell leukemia), MeWo (melanoma), MKN-1 (gastric), HCT-15 (colorectal) and DLD-1 (colorectal). While several polymorphisms were previously reported (e.g. NTRK1 R780Q), their relevance to cancer had not been established. Also, the authors compared frequencies of occurrence in normal versus cancer cells to suggest polymorphisms that maybe more oncogenic (e.g. TNK M598delinsEVRSHX) or tumor suppressive (e.g. EGFR R521K) in nature. It must be noted that an explosion of genome-wide studies [23-33] has bolstered the relevance of SNPs in cancer disease etiology. For somatic mutations, a total of 28 recurring events were identified. Furthermore, they were able to confirm several novel (e.g. FGFR4 Y367C and CSK Q26X) and previously known observations (e.g. EGFR G719S) in 165 tumor and 90 healthy blood DNA specimens. Interestingly, 70 kinases harbored at least one somatic mutation and only 9 of all the sporadic alterations were in common with previously published reports.

Collectively all the large scale resequencing studies underscore the relevance of fewer driver versus several passenger events in tumor genomes, the existence of biological relevant low mutation frequencies, the significance of kinase alterations, the pathway relevance of mutations and the overabundance of SNPs. Next generation sequencers such as Illumina, ABI and Helicos are likely to significantly expand this data by increasing coverage in larger number of individual genomes. Nonetheless, we already have an excellent repertoire of targets to begin therapeutic and diagnostic characterizations.

1.2.2 Genome-wide array profiling studies

The advent of gene chip or microarray technology has provided a genome-wide analytical tool to assess relative abundance of RNAs or DNA copies and identify SNPs or bindings sites of DNA binding proteins (e.g., transcription factors) in a high-throughput, parallelized format. This field has made leaps and bounds since the mid-90s overwhelming scientists with a bout of data as evidenced by thousands of freely accessible datasets collated in public data ware houses such as ArrayExpress[34] and GEO[35] Clearly, transcriptomic analysis by expression arrays, which are in principle analogous to large-scale RT-PCR, has dominated this area. Improvements in technology, access and cost have empowered widespread use of microarray studies in cancer research – lymphomas [36-42], lung cancer [43-46], breast cancer [47-60], ovarian cancer [61, 62], colon cancer [63-65], prostate cancer [66-70], brain cancer [71-77] and others [78, 79]. The underlying theme of these diverse studies has been definition of distinct molecular sub-classes of cancer based on gene expression profiles, development of prognostic signatures and demonstration of superiority over conventional pathological diagnoses.

One of the earliest breakthrough studies by Alizadeh et al [36] aimed to classify clinically heterogeneous diffuse large B-cell lymphomas (DLBCL) based on microarray derived gene expression profiles of 96 patient samples. They came up with 2 distinct previously unknown molecular subtypes indicative of different stages of B-cell differentiation from peripheral blood B cells – ‘Germinal centre B (GCB) like’ which had significantly improved prognosis and better outcome after CHOP therapy, and ‘activated B (ABC) like’ (Figure 2). Perou and colleagues [80] published similar work to capture the transcriptional blueprint of 65 primary breast cancer specimens into ~8K cDNA array

derived 'molecular portraits' which revealed new cancer subtypes that were associated with cell type origin – luminal A /ER+, luminal B/ER+, normal breast-like, ERBB2+, basal-like. They identified the underlying signatures which provided novel mechanistic insights and tested their stability and reproducibility to classify new patient populations into disease entities associated with clinical phenotype [56]. While such signatures have been unraveled in aforementioned studies for a whole host of tumor types, breast cancer has been most extensively characterized by genomics efforts. Of particular note is work [58, 81] by the Netherlands Cancer Institute (NKI) group in deriving a 70-gene classifier with the power to predict 10-year disease recurrence of node-negative early stage (N0, T1/T2) breast cancer patients under 53 years old who would otherwise unwarrantedly receive debilitating standard of care cytotoxic treatment. This signature outperformed classification by conventional histopathological risk factors and was far superior to St Gallen's and National Institute of Health guidelines in determining patient eligibility for not receiving adjuvant therapy. This clinically useful finding has been validated in large multi center studies to stratify patients for improved outcome with adjuvant systemic therapy [47, 82]. An often cited success story of the power of genomics, it was translated into a diagnostic tool, which received regulatory approval in 2007 and is currently marketed as MammaPrint® by Agendia.

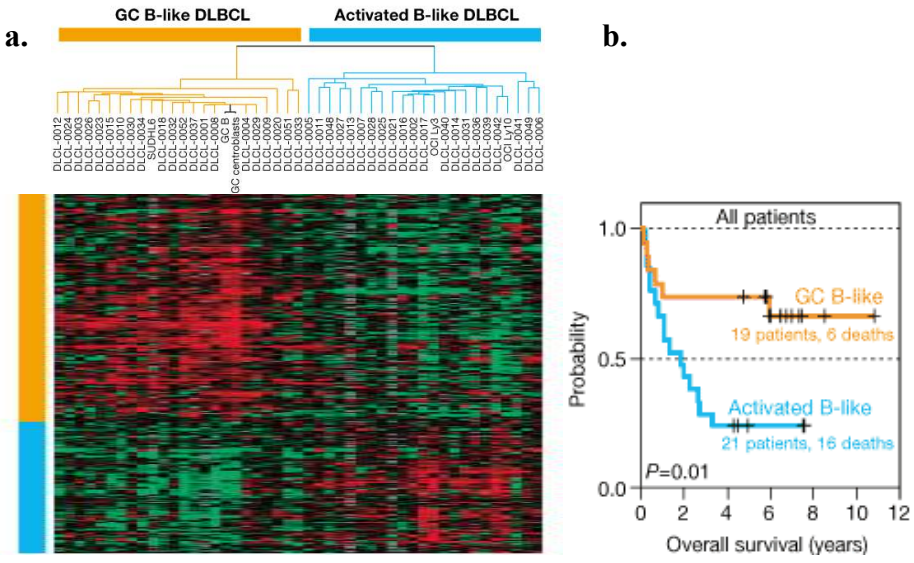


Figure 2. DLBCL subgroups with differential prognosis defined by gene expression profiling. (a) Hierarchical clustering of 128 cDNA microarrays corresponding to 96 samples of normal and malignant lymphocytes revealed 2 distinct subgroups, GC B-like DLBCL (orange) and activated B-like DLBCL

(blue), based on germinal center B cell (black) gene expression signature. Genes that are selectively expressed in each subtype are shown where each row represents a gene on the microarray and each column a tumor sample. Values depicted in the heatmap represent $\log_2$ based hybridization ratios for each sample (to a common reference) from red to green indicating high to low relative gene expression, respectively. (b) Kaplan-Meier plot of overall survival shows statistically significant clinical relevance, in terms of distinct prognosis, of these molecularly defined patient groups. Adapted from [36].

Many groups have similarly discovered and evaluated novel correlates of biological effect such as pathway activity and clinical end points such as tumor grade, disease progression, prognosis, survival and response to therapy. However, they suffer from lack of statistical power, sound validation and/or prospective analysis.

While we discussed genetic alterations that affect nucleotide sequence (e.g. point mutations, insertions, deletions, translocations), cancer genomes acquire changes in gene dosage by amplifications (such as oncogenes) or deletions (such as tumor suppressors) of a genetic locus that confer a growth selective advantage. DNA microarrays have also been effectively used in elucidating such chromosomal aberrations in a wide variety of tumor types. Aneuploidy is better detected by traditional cytogenetics methods (e.g. SKY, FISH), but array-based comparative hybridization (aCGH) can deliver genome-level assessments of high resolution gains and losses that are likely to be, like mutations, passenger or driver events. Genomic copy number alterations that manifest in transcriptional changes have a high likelihood of separating signal from noise. A case in point is the first such examination of primary breast tumor aCGH results where the investigators found that 60% of genomic amplifications corresponded to coordinately overexpressed genes. SNP arrays are also widely used in copy number analyses and while their application is limited in revealing large alterations, they are effective at elucidating copy number neutral loss-of-heterozygosity (LOH) events.

1.2.3 High throughput RNAi screens

Andrew Fire and Craig Mello were awarded a Nobel Prize in 2006 for their work on RNA interference (RNAi) which was first discovered in the worm Caenorhabditis elegans [83]. RNAi is an endogenous cellular process by which long double stranded RNAs are cleaved by the RNAse III-like ribonuclease enzyme Dicer into short 20-25 basepair fragments with 3' overhangs called small interfering RNA (siRNA), one strand of which (guide strand) binds to and activates the ribonucleoprotein complex called RISC

(RNA-induced silencing complex), containing endonucleases, to target complementary regions on messenger RNAs for degradation (Figure 3). In this manner, RNAi causes sequence-specific target silencing, regulates gene expression and causes a loss-of-function phenotype. The availability of RNAi reagent libraries, ranging from unbiased full genome coverage to customized subsets (e.g. gene family, druggable genome), has opened the doors to rapid, systematic, large scale, genome-wide loss of function screening in cells and whole organisms (e.g. C. elegans and Drosophila) by use of automated assays. The reagents often used for mammalian cells are synthetic silencing RNAs (siRNAs), short hairpin RNA (shRNAs) or endoribonuclease derived siRNAs (esiRNAs) [84]. 21-23 nucleotides long siRNAs are generally used due to activation of the interferon response by long double-stranded RNAs. In contrast to traditional gene 'knockout' experiments, RNAi is essentially a forward genetics screen using a reverse genetics technique that is empowered by flexibility and speed due to apriori knowledge of sequence information. However, except in invertebrates, significant caveats are phenotypic variability as a result of incomplete or inefficient knockdowns and off-target effects which can lead to artifacts in large scale screens. Nonetheless, this powerful technology has far reaching applications in target validation and drug discovery efforts [85]. Guidelines highlighting the importance of sound experimental design and analysis in carrying out robust screens and validating hits can be found in [86-88]. A sampling of published screens is presented below.

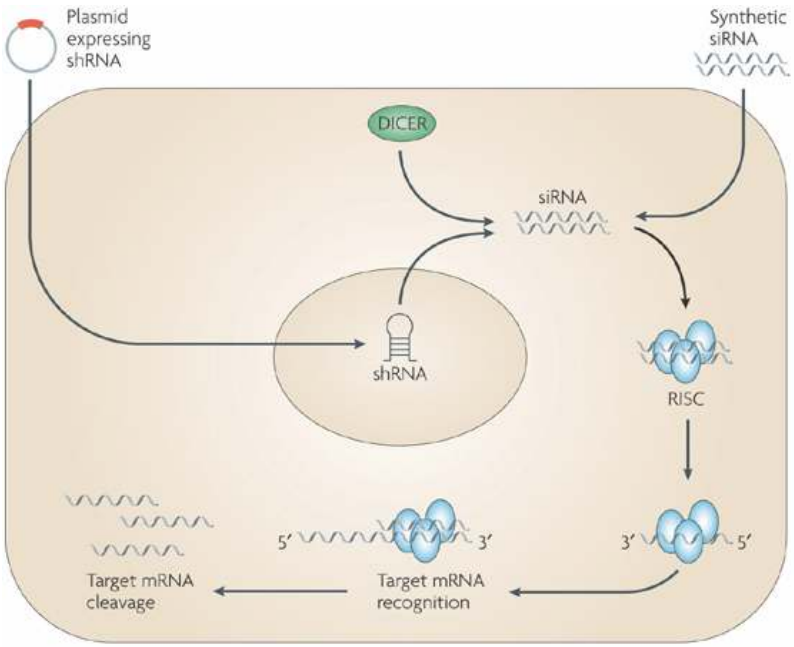


Figure 3. Mechanism of experimental RNAi. Chemically synthesized siRNAs or vector encoded shRNAs (short hairpin RNAs), processed by RNAse DICER, are loaded in ribonucleoprotein complex RISC (RNA induced silencing complex) to recognize and cleave complementary target mRNAs ('target expression silencing') in a sequence specific manner. Adapted from [89]. See text for details.

Among the first mammalian screens performed, Rene Bernard's group screened 50 human de-ubiquitinating enzymes to identify CYLD as a novel suppressor of NF-kappaB which functions by deubquitination of TRAF2 and consequently causes resistance to apoptosis [90]. Since inactivating mutations in CYLD are associated with familial cylindromatosis, a rare genetic disorder that predisposes individuals to skin tumors, the authors hypothesized aberrant NFkB pathway activation as the culprit. They went on to show interesting trial data where topical application of aspirin derivatives, salicylic acid (NFkB inhibitor), elicited a favorable clinical response in small group of patients. In subsequent work, the same group [91] used a barcode RNAi library containing 23,742 shRNAs targeting 7914 human genes to identify one known (TP53) and 5 novel proteins (RPS6KA6, HTATIP, HDAC4, KIAA0828, CCT2) that modulate p53-mediated cell cycle arrest. In 2005, MacKeigan and colleagues [92] reported the first ever RNAi screen to categorically knockdown kinases and phosphatases in HeLa cells by siRNAs to identify genes that offers a selective growth advantage to promote survival and escape apoptosis. They validated several known and novel survival kinases and presented intriguing data on several previously unknown tumor suppressive and oncogenic

phosphatases. Furthermore, they showed that RNAi-mediated silencing of specific survival kinases sensitized resistant cells to cytotoxic agents (e.g. PINK1 kinase suppression that enhanced Taxol mediated killing in HeLa and BT474 cells), suggesting new targets for therapeutic intervention. In a different study, the Genomics Institute of Novartis Research Foundation [93] applied an innovative 384-well wound healing assay coupled with microscopy to screen ~ 5K genes covered by a library of ~ 10K siRNAs for genes that were associated with migration in SKOV3, an ovarian cancer cell line. Out of 4 genes (CDK7, DYRK1B, MAP4K4, SCCA-1) whose inhibition blocked motility, MAP4K4's effect on invasiveness was found to be mediated through c-Jun N-terminal kinase, proposing the rational use of a MAPK pathway inhibitor to arrest tumor cell migration. RNAi-based genetic screens have generally been utilized to study function of oncogenes in cancer cells, but a group at Harvard [94] used a shRNA screen to identify a novel tumor suppressor candidate gene REST, previously implicated in neuronal gene expression, which inhibited transformation of mammary epithelial cells. The authors also uncovered well known tumor suppressors such as TGFBR2 and PTEN, but found REST to be a frequently deleted or mutated gene in colorectal cancer and was dependent on PI3K signaling for cellular transformation.

Due to the aforementioned incomplete or inefficient knockdown associated with siRNAs, reagent redundancy in producing consistent phenotype is widely accepted as the best parameter to confirm target specificity in producing a phenotype. To this end actives or hits are screened with multiple independent siRNAs individually or in pools, the latter being more likely to introduce artifacts due to off target effects. On the other hand, vector based shRNAs offer several advantages. They are generally cheaper and can be used to transfect or infect cells, especially untransfectable primary non-dividing cells, through packaged lentiviruses or retroviruses which can also be leveraged to produce stable knockdown clones. In barcoding screens [91, 95-97], pooled shRNA bearing viruses with a unique barcode that selectively integrate into cells to produce the desired phenotype are uniquely identified by sequencing or microarrays containing the barcode sequences. These screens have also been carried out in arrayed format by other groups [98]. shRNA screens, however, are limited by production of high titers of virus and selection of 'low hanging fruit' due to pooling. Improvements in library design and use as well as assay

formats are rapidly evolving and are highly likely to mitigate existing drawbacks. Compared to conventional functional genomics studies which are essentially global surveys that aim to provide 'associations or correlations', high throughput RNA interference screens are powerful methods to carry out targeted knockdowns that help assign 'causal' relationships.

A creative use of RNAi has been in synthetic lethality experiments in concert with drugs to identify enhancers or suppressors of drug efficacy *in vitro*. These are analogous to genetic modifier screens in model organisms. Whitehurst et al [99] recently published a genome-wide screen for sensitizers of paclitaxel in NCI-H1155, a human non-small-cell lung cancer line, and found 87 gene hits (false discovery rate < 5%) that compromised cell viability in the presence of sublethal concentrations of paclitaxel. Several of these genes were involved in microtubule biology and mitosis. Similarly, Berns and others [100] employed a shRNA barcode screen to uncover genetic determinants of Herceptin® resistance and showed that activation of PI3K pathway (PTEN loss and PIK3CA mutational activation) was a predictor of chemosensitivity *in vitro* and *in vivo*. This is in agreement with previous findings of PTEN-deficient tumors being poor responders to trastuzamab therapy and loss of PTEN conferring resistance *in vitro* [101]. Therefore, drug modifier screens hold great promise to reveal clinically useful insights into predictive biomarkers of response for patient selection, understanding mechanism of action and potentially discovering alternative indications as well as combination therapy opportunities.

1.3 Examples of integrative analysis

The abundance of genes and proteins with therapeutic or diagnostic potential that are unraveled by functional genomics studies necessitates innovative and integrative analyses to reveal underlying mechanisms, establish cause-effect links and triage and prioritize this information for biopharmaceutical applications. A few examples are discussed below.

Some of the earliest attempt to combine diverse sets of functional genomics data included microarray-based gene expression profiles and chemosensitivity correlations.

The NCI-60 is a 60 cell line panel, representing various tumor types, that has been routinely in use to screen anti-tumor compounds for several years as part of the NCI Developmental Therapeutics Program [102]. In 2000, researchers at the NCI attempted to correlate cDNA gene expression profiling studies (~3700 genes) of the NCI-60 cell lines with their growth inhibition responses (GI50) to 1400 compounds [103]. Since these were untreated cells, their goal was to identify molecular patterns of drug activity analogous to selecting therapy based on basal characteristics of patient tumors and predicting response. Several gene-gene, gene-drug, drug-drug correlations were uncovered, showing known (5-FU and asparaginase) and novel relationships. In a follow up publication [104] using an Affymetrix platform with more genes, they showed successful predictions using gene expression based correlates of chemosensitivity for 88 out of 232 compounds. More recently, they reported a novel algorithm, "coexpression extrapolation" (COXEN), that can accurately predict drug sensitivity of bladder cancer cell lines and clinical responses of breast cancer [105]. McDermott et al. [106] have expanded on this to profile 500 diverse cancer cell lines for sensitivity to 14 kinase inhibitors and showed mutually exclusive toxicity in small subsets of cell lines. In their analysis, EGFR, HER2, MET, or BRAF inhibitors were selectively efficacious in cells with underlying activating mutations or amplifications for the respective target, suggesting that genetic context or genotype, regardless of tissue type, can predict response and guide early clinical development to kinase inhibitors. This is reinforced in promising results from Joseph Nevins' group who generated pathway activation signatures in cell line models and applied them to tumor gene expression data to predict sensitivity to agents that target members of the pathway, thus enabling guided use [45]. This continues to be an area of active research in pharmacogenomics applications as well as methodological improvements to identify genetic determinants of sensitivity/resistance in *in vitro* models as well as patient samples.

Three interesting studies recently showed the discovery of novel genetically altered and therapeutically relevant oncogenes by integrating diverse high throughput genome-wide profiling datasets. Garraway et al [107] integrated high-density single-nucleotide polymorphism (SNP) array-based genotypes with gene expression data for the NCI-60 cell panel. They elegantly identified a novel MITF amplification in melanoma cell lines.

They went to provide clinical support in patient samples with metastatic disease and decreased survival and reported co-occurrence with BRAF and p16 mutations. The authors leveraged the power of SNP arrays in unveiling LOH (loss of heterozygosity) events and copy number changes to identify 3p12-3p14 as a region of high gain in the melanoma cell line cluster and used transcript profiling to hone in on MITF as the only statistically significant and highly expressed gene in this region and confirmed the alteration in 10% primary and 20% metastatic melanoma tumors. MITF is a master regulator of melanocyte lineage commitment and survival and increase in copy number may well cause commonly found resistance to standard chemotherapeutic agents which could be addressed by combination therapy, as suggested in the study. In a different study, Boehm and colleagues [108] combined gain-of-function and loss-of-function screens with whole genome cell line SNP arrays and tumor arrayCGH to discover IKKepsilon (IKBKE) as an tumorigenic kinase that is frequently amplified in breast cancer. The team first identified IKBKE, among other hits, as a strong substitute for myristoylated-Akt in inducing transformation of immortalized non-tumorigenic HEK cells expressing activated MEK pathway members from a myristoylated kinase library screen and discovered that it was the only amplification seen in SNP array based copy number survey of 49 breast cancer cell lines and arrayCGH analysis of 30 primary breast cancer specimens. They proceeded to confirm over expression and show convincing data to implicate IKBKE in activation of the NFkB survival pathway. Interestingly, they found that 3/5 shRNAs targeting IKBKE compromised cell proliferation and viability of MCF7 cells in a separate shRNA screen. In our third case study, Kim et al [109] employed a comparative oncogenomics approach to identify NEDD9 as an orthologously conserved oncogene in human and mouse melanoma. They found that an acquired focal amplification (850kb on chromosome 13) in an inducible H-Ras mouse model of melanoma was shared with human metastatic melanoma samples, subjected to array-based comparative genomic hybridization, and that NEDD9 was the only over expressed gene among 8 genes in this region. Rigorous functional analyses supported an invasive role *in vitro* and a metastatic role in animal studies and human tumors. An intriguing thought to note is that mutations in MITF, IKBKE or NEDD9 have not been reported to date, demonstrating the power of

integrative approaches in discovering and validating novel alterations that are essential for cancer development and progression and hold translational potential.

Compendia analyses such as combining large sets of microarray data represent another attractive means to intelligently extract patterns. One such example is the Connectivity Map [110] which connects disease, genes and perturbagens (compounds) by matching transcript profiles of interest against a reference database comprising gene expression data from cells treated with various small molecules. Toward this end, the Broad Institute has compiled >400 gene-expression profiles derived from treating cultured human cells (MCF7, PC3, HL60, SKMEL5, HepG2, SHSY5Y) with a large number of perturbagens to populate a reference database. Pattern-matching algorithms using Kolmogrov-Smirnov statistics score each reference profile for the direction and strength of enrichment with a query signature. Perturbagens are ranked by this "connectivity score" where those at the top ("positive") and bottom ("negative") are suggested as being functionally connected with the query signature and thus provide data-driven 'leads' for experimental followup. While the method has several challenges and considerations, the group reported a variety of interesting applications such as finding molecules sharing similar mechanisms of action (e.g. HDAC inhibitors) for compound signatures; positively and negatively associated compounds (e.g. estrogens and anti-estrogens); mechanism of action from gene expression fingerprints of unknown compounds; compounds related to disease signatures (e.g. mTOR inhibitor sirolimus phenocopies dexamethasone sensitivity and reverts dexamethoasone resistance in ALL). They show that signatures can be agnostic to contextual (cell line, concentration) parameters and can produce real, confirmable *in silico* findings. A significant challenge with such large scale compendia analysis is collecting, parsing, standardizing, analyzing and making the data available to do a variety of global analyses. Oncomine [111, 112] (http://www.oncomine.org) is a large-scale initiative that has embarked on such a mission for the oncology community by collating > 18, 000 diverse cancer gene expression microarrays and enabling extremely valuable analysis of genes, pathways and networks that are affected in different cancers and their genetic or histological subtypes.

1.3.1 Challenges and considerations in integrative analyses

Genome-wide approaches inherently suffer from experimental and computational drawbacks. High-throughput technologies are sensitive to the way in which the samples are collected and handled, and a variety of factors such as RNA and protein degradation and presence of contaminating tissue can influence gene expression and proteome analysis. Efforts such as MAQC [113], a consortium of over 150 regulatory, industrial and academic scientists, have reassuringly shown comparability of cross-platform DNA microarray data from two commercially available RNA samples. Similar analyses are warranted for other technologies. The variability and lack of reproducibility across platforms and between laboratories is exacerbated by tumor heterogeneity for primary tumor tissue studies. Therefore, well annotated clinical samples from growing tissue banks and tumor repositories are essential for discovery and validation purposes. Furthermore, generating multiple molecular readouts (e.g. mutations, arrayCGH, expression) on the same sample and consistent measurements across samples would enable within-sample and between-sample cross-comparisons of diverse data types which can lead to powerful testable hypotheses of translational value. Such explorations, as exemplified by the Cancer Genome Atlas (http://cancergenome.nih.gov/), will allow interesting associations of genetic determinants with clinical covariates (histopathology, survival, disease stage, etc.). Since biological systems and in particular cancer models can be complex, multi-faceted, context-dependent and inherently dynamic, care should be taken to sample sufficient and informative time points

It must be noted, however, that due to the descriptive or observational nature of global profiling in a few samples, the designs are usually statistically underpowered and results are likely to yield false positives. In such cases, error propagation ought to be considered due to noisy correlations in fusing two disparately generated datasets in a ‘fishing expedition’. Several quantitative and statistical approaches, with varying degrees of sensitivity and specificity, have been developed to analyze, mine and model individual data types, but they can produce confounding and non-overlapping answers. These issues can be overcome by targeted hypothesis-driven analyses where specific questions are asked and different approaches are cleverly combined to find concordant results to minimize false positives. Particularly in cancer, these are likely to differentiate driver

events from bystander effects since, intriguingly, the mutation rate of sporadic cancers is apparently not higher than that of normal cells [114]. Also, rather than trying to understand single gene/protein changes in 'lists' of differential expression, pathway level analyses can provide powerful hypotheses for follow up analyses. While still in its infancy, there is a serious need for powerful knowledge bases and creative/unconventional integrative methods that yield high confidence hypotheses. Efforts such as Oncomine [112] are a step in this direction where large compendia of sub-optimal datasets are fused with apriori information to achieve an *in silico* genomic understanding of genes and pathways in a wide variety of tumor tissues. Data visualization becomes a formidable challenge with such integrative analyses. Novel insights and hypotheses generated by such approaches necessitate thorough and rapid validation as well. Advances in microfluidics, nanotechnology and non-invasive molecular imaging are beginning to enable this in a cost-effective and robust manner.

One of the greatest challenges of applying systems approaches is the curse of dimensionality and the complexity therein. Scientists are able to generate large volumes of data relatively easily and quickly and the rate-limiting step clearly is knowledge discovery for real world applications. This is in part due to a focus on generating 'parts lists' rather than understanding deeper biological meaning. Also, new data types (e.g. micro RNA profiles) with better understanding of regulation and technological innovation, new data sources, databases, tools and systems are constantly emerging. These together with pre-existing heterogeneity of data sources and lack of standardization in experiments, data types, tools and analysis pose significant challenges in data management, storage, processing, analysis , integration and interpretation. This impedes realization of the full power of systems level data for hypothesis generation. Efforts towards standardizing information contained in high throughput experiments as well as data exchange standards such as MIAME, PSI-MI, MIARE (reviewed in [115]) are definitely going to be helpful in this regard.

1.4 Specific aims of thesis

This dissertation was aimed at a systems level characterization of genetic determinants of tumor biology and their relevance, if any, to drug discovery applications. To this end, a

variety of integrative approaches were applied to three case studies, involving high-throughput genome-wide molecular profiling datasets of human cancer:

(1) Druggable genome-wide loss-of-function siRNA screens in four cancer cell lines were analyzed and integrated with orthogonal datasets (genetic alterations, transcriptomics, pathways, survival) to identify strong candidate target genes that are essential for cancer cell survival.

(2) A recent large resequencing effort [22] uncovered SYK, an unconventional tyrosine kinase tumor suppressor in breast cancer, as the most frequently mutated gene. By fusing heterogeneous data such as microarrays, pathways and compound sensitivity from SYK altered contexts, we generated hypotheses on the biological significance of the mutations identified as well as novel aspects of SYK biology.

(3) Mining gene expression profiling data from compound treated cancer cells can provide clues on acquired resistance mechanisms which are a clinical challenge. 5-FU treated colon cancer cell transcriptional profiles were combined with other sources to shed light on relevant druggable targets, survival pathways and testable compounds for combination/adjuvant therapy opportunities.

2 Materials and Methods

2.1 Computational methods

2.1.1 Datasets and tools

High throughput mutational screens published by the Singapore Oncogenome Group (SOG) [22], Sanger's COSMIC effort (http://www.sanger.ac.uk/genetics/CGP/) and the Johns Hopkins group (JHU) [20, 21] were the source of genome-wide mutation data. ArrayCGH data used in this dissertation came from genomic DNA from a panel of cell lines, including Calu6, HCT116, MCF7 and U87, were hybridized onto 44A/B Agilent CGH oligo arrays (Agilent, Santa Clara, CA, USA) and a novel method was employed to detect copy number changes (Xiang Y et al., unpublished). Publicly available microarray datasets were compiled from primary literature and compendia such as GNF (http://symatlas.gnf.org/), GEO (http://www.ncbi.nih.gov/geo/) or Oncomine (http://www.oncomine.org) and have been described elsewhere. Wherever applicable, differential gene expression analysis was performed using Significance Analysis of Microarrays (SAM) [116]. Clustering and heatmap analyses were carried out in TIGR's MeV4.0, a Java-based, open-source software (http://www.tm4.org/documentation/MeV_Manual_4_0.pdf). Overrepresentation analysis of Gene Ontology (GO) Biological Process (BP) terms was done using NCI's DAVID tool (http:///david.abcc.ncifcrf.gov/). Connectivity Map from the Broad Institute [110]was leveraged for compound derived gene expression profiles. Gene Set Enrichment Analysis (GSEA) [117] helped in determining enrichment of custom assembled signatures. Oncomine (http://www.oncomine.org) [112] and Ingenuity Pathway Analysis, IPA 6.0 (Ingenuity Systems, Redwood City, CA, USA) served as the platform for pathway and network analyses. We also queried Oncomine for published datasets where a statistically significant ($p<10\text{-}e6$) differential expression profile was noted for cancer versus normal tissues for SYK. SpotFire (TIBCO Spotfire, Somerville, USA) was used for visualization and R/BioConductor [118] (http://www.r-project.org) was employed for all other analyses.

2.1.2 Gene expression analysis

Affymetrix gene chip datasets were downloaded and processed, if not done already, by the Micro Array Suite 5.0 (MAS5) algorithm (http://www.affymetrix.com/support/technical/index.affx). The signal estimates were scaled by setting the target intensity to 500 to account for systematic differences in intensity between chips for cross comparability. Probesets with Absent calls across all arrays were dropped. Differential expressed genes were determined by SAM [116] for $\log_2$ transformed values. Significant results were filtered for a false discovery rate (FDR) <10% and fold change >2 or <-2. Functional grouping into gene families – ion channel, phosphate, kinase, transporter, receptor, enzyme, secreted, transcription regulator, other – were based on IPA's classification. Kaplan Meier statistics were implemented for survival analysis (see 2.1.4). 1-way or 2-way hierarchical clustering was applied in MeV 4.0 using Pearson correlation distance metric with average linkage on log transformed, normalized and median centered data.

2.1.3 SYK_interactions_network generation

Direct and indirect gene interactions of SYK were extracted from IPA using the Neighborhood Explorer feature (118 human genes). Protein-protein interactions involving SYK were also mined from Human Protein Reference Database, HPRD (http://www.hprd.org; 45 genes). A master list of 125 non-redundant genes was thus compiled that broadly represented a SYK molecular network. These genes were mapped to probesets demonstrating considerable variation across microarray profiles of a panel of 13 cell lines selected for varied SYK background (See relevant section for details). Probesets with a coefficient of variation (ratio of standard deviation to mean), CV > 0.4 were considered. A total of 109 genes mapped to 201 probesets comprised the SYK network that was used in clustering and gene set enrichment analyses.

2.1.4 Survival analysis

We compiled a list of publicly available Affymetrix microarray datasets from primary tumor patients (see below) with associated survival information. Further details of each study can be found in the primary citation. These datasets were pre-processed and signal

values were generated using MAS5 algorithm, as described in 2.1.2. Probesets with 100% Absent calls were filtered out. Samples were grouped based on median, quantiles (0-25%, 25-50%, 50-75%, 75-100%) or extreme quantiles (0-25%, 75-100%) of $\log_2$ transformed gene expression values for each probeset. Kaplan-Meier plots for each probeset were calculated in R/Bioconductor and statistical significance, unless otherwise specified, was established by a log-rank test $p < 0.05$. Results were summarized by the number of significant instances for each gene across the datasets grouped by tumor type.

Dataset	Tumor type	Microarray Platform	n	GEO Accession No.	Ref.
Bild_Lung	NSCLC; Adenocarcinoma & Squamous cell carcinoma (Lung)	U133Plus2	111	GSE3141	[45]
Beer_Lung	NSCLC; Squamous cell carcinoma (Lung)	U133A	130	GSE4573	[43]
Bhatt_Lung	NSCLC; Adenocarcinoma (Lung)	U95Av2	125	NA	[44]
Bild_Ovarian	Ovary	U133A	146	GSE3149	[45]
Phillips_Astrocytoma	Astrocytoma (Brain)	U133A&B	100	GSE4271	[75]
Freije_Glioma	Glioblastoma (Brain)	U133A&B	85	GSE4412	[71]
Nutt_Glioma	Glioblastoma (Brain)	U95Av2	50	NA	[76]
Hummel_Lymphoma	B-cell lymphomas (Lymphoma)	U133A	221	GSE4475	[40]
Miller_Breast_A	Breast	U133A&B	251	GSE3494	[54]
Bild_Breast	Breast	U95Av2	158	GSE3143	[45]
Sotiriou_Breast	Breast	U133A	178	GSE2990	[45]
Cromer_H&NSCC	Head and Neck squamous cell carcinoma	U95Av2	31	GSE2379	[78]
Shipp_Lymphoma	Lymphoma	Hu6800	77	NA	[41]
Pomeroy_Medulloblastoma	Medulloblastoma (Brain)	Hu6800	94	NA	[77]

2.1.5 Gene Set Enrichment Analysis (GSEA)

Custom gene sets were compiled for EGFR/MAPK/ERK pathway – 42 genes, NFkB pathway – 45 genes, PI3K/Akt pathway – 99 genes using IPA's canonical pathway content. SYK direct and indirect interactions (SYK_interactions_network) – 125 genes were derived as described in 2.1.3. GSEA was implemented by using the desktop GSEA

version 2.0 program as published previously [117]. Enrichment of each geneset was calculated in a ranked list of genes based on a signal-to-noise, SNR ($\mu_{class\ 0} - \mu_{class\ 1}$)/($\sigma_{class\ 0} + \sigma_{class\ 1}$) score that discriminated transcript profiles in any 2 group comparisons of NULL and/or MUT cell lines with respect to WT. Normalized enrichment scores, NES were calculated based on a weighted Kolmogorov–Smirnov statistic and statistical significance was assessed by 1000 permutations to produce FDR q-values. To explore other pathways and biological processes that are differentially modulated in NULL & MUT vs. WT cell lines, we filtered 1000 genes with SNR> 0.5 or <-0.5 that were differentially expressed in MUT & NULL groups relative to WT cell line.

2.1.6 Pathway and network analysis

2.1.6.1 GO analysis

Overrepresentation analysis of Gene Ontology (GO) Biological Process (BP) terms was done using NCI's DAVID tool (http:///david.abcc.ncifcrf.gov/). This was assessed by comparing the frequency of GO BP, level 5 categories represented in the nonredundant list of genes versus the global frequency of GO categories in the reference gene set which corresponded to all known genes in the human genome. Given the small numbers in each of the lists, statistical significance in terms of p-values was deemed to be less informative; therefore, we focused on examining relative enrichment or over representation of members to rank the categories. A fold enrichment >=1.5) in BP categories containing >=5 genes were considered significant.

2.1.6.2 IPA analysis

Canonical pathways and biological functions were queried for a given list of genes in IPA. 2 parameters were calculated for each pathway represented in the Ingenuity pathways knowledgebase. Ratio measures the fraction of genes in the list to the total number of genes making up that pathway. – log P values are calculated with the right-tailed Fisher's Exact test and can be used to support a non-random association. Relevant pathways containing >=2 members are shown with corresponding ratio and –logP values. Due to sparse lists of genes, canonical pathway analysis can be limited. In such cases, network analysis was performed in IPA. Functional networks, comprising <35 network eligible

molecules each were generated by the Network Generation Algorithm (https://analysis.ingenuity.com/pa/info/help/ingenuity_pathways_analysis_network_generation.htm). These networks were ranked by scores based on a hypergeometric distribution and calculated with a right-tailed Fisher's Exact Test as well. The topmost or top two overlapping networks were analyzed for tissue specific expression, top biological functions as well as inhibitors. Supplementary information on methods can be found at https://analysis.ingenuity.com/pa/info/help/ipa_help.htm

2.1.7 Connectivity Map analysis

The Connectivity Map [110] contains a compilation of >400 gene-expression profiles derived from treating cultured human cells (MCF7, PC3, HL60, SKMEL5, HepG2, SHSY5Y) with a large number of perturbagens to populate a reference database. 131 upregulated and 68 downregulated probesets, processed and analyzed in manner described earlier (2.1.2), represented our query signature of 5FU modulated gene changes in GC3 cells. Enrichment of these up and down 'tags' in each compound treatment instance in the Connectivity Map was calculated using a Kolmogorov-Smirnov statistic as reported [110] and combined into a 'connectivity score'. Each compound instance was ranked in this manner. Negative scores which were presumably negatively connected with the input signature were examined for multiple occurrences of compounds with similar chemistry or mechanism. These were suggested as putative combination therapy molecules. Supplementary information can be found at http://www.broad.mit.edu/cmap/

2.2 Experimental methods

2.2.1 Cell lines and reagents

A total of 7 cancer cell lines used in the primary RNAi screen and confirmation steps were obtained from American Tissue Culture Collection, ATCC (Rockville, MD, USA): HCT116 (Cat.# CCL-247), U87 (Cat.#HTB-14), MCF7 (Cat.#HTB-22), Calu6 (Cat.#HTB-56), A549 (Cat.#CCL-185), BxPC3 (Cat.#CRL-1687), SKOV3 (Cat.#HTB-77). All cells were cultured in the recommended growth medium supplemented with 2mM L-glutamine and 10 % fetal bovine serum (FBS), in a humidified 37°C incubator with 5 % CO2. Human Druggable Genome siRNA Set V2.0 (Qiagen, Valenica, CA, USA), covering 6992 genes X 4 siRNA duplexes per gene arrayed in 96-well format, served as the screening library. Reverse transfections were done using Lipofectamine™ 2000 (Catalog# 11668-500; Invitrogen, Carslabd, CA, USA). Cell Titer Glo® (Cat.# G7570; Promega) and ToxiLight cytotoxicity assay kit (Cat.# LT07-217; Cambrex Bio Science Inc., Rockland, MD, USA) were used to measure cell viability and cell death, respectively. Forward and reverse primers and probes for Taqman® QPCR were obtained from ABI Taqman Gene Assay Catalog (Applied Biosystems, Foster City, CA, USA): PIK3CA (Cat.#Hs00180679_m1), AKT1 (Cat.#Hs00178289_m1), AURKA (Cat.#Hs00269212_m1), ILK (Cat.#Hs00177914_m1).

2.2.2 siRNA high-throughput screen (HTS)

Prior to HTS, 300 384-well plates were pre-printed using 2 siRNAs per target in the screening library such that each well contained 13nM of an individual siRNA duplex. High throughput reverse transfections were performed by adding transfection agent and seeding ~1500 cells into each well as previously reported [119]. 72hr (for HCT116) or 96 hr (for Calu-6, U87MG and MCF7) later, cell viability was measured using chemiluminescence based CellTiter Glo assay readout, according to manufacturer's recommendations. This assay was selected due to greater dynamic range and low variability. Plates that passed QC in terms of high transfection efficiency, performance of controls, plate uniformity, low plate to plate variation, edge effects or systematic errors

were taken forward for analysis. Several screening parameters (cell growth, transfection, controls, etc.) were thoroughly optimized before commencing HTS. UBB siRNA was chosen as a positive cell killing control and a scrambled non-silencing siRNA or GFP siRNA as negative controls. Only half of the library was screened for HCT116.

2.2.3 Hit selection

Raw signal values from the screen were normalized to untreated control wells to compare across plates. Viability ratios are calculated for normalized signals of each target siRNA with respect to the negative control. Based on prior experience, an arbitrary cutoff of > 60% lethality or <40% viability was used to filter siRNAs with cytotoxic properties for the Cell Titer Glo readout. To short-list genes that are broadly essential to cell survival (general hits), we applied a cutoff of <40% viability in 2 or more cell lines where phenotypic concordance for duplicate siRNAs was seen in at least 2 cases. For cell specific lethality (cell specific hits), we picked out genes where the cognate siRNAs showed consistent cell kill phenotype for the cutoff used in a particular cell line. In cases, where concordance was not observed, single hits demonstrating <40% viability for a given cell line, but >80% viability for the remaining cell lines were picked.

2.2.4 Cell toxicity assays

Follow up analyses of screen actives by selecting all 4 siRNA from the library for each short-listed target was done using Cell Titer Glo (CTG) or ToxiLight (TXL) assay readouts according to the manufacturer's instructions. These experiments were performed using the same conditions as developed for the screen. Out of the short-listed targets that caused general cytotoxicity upon siRNA mediated inhibition, PLK1 and KIF11 were confirmed by both assays in 7 cell lines (A549, BxPC3, SKOV3, HCT116, Calu6, U87, MCF7). Out of the cell-specific targets, NOTCH4, AKT1, MCL1 were confirmed by both assays in the same 4 cell lines (HCT116, Calu6, U87, MCF7) used in the primary screen. Based on prior experience, we applied a threshold of >60% loss of viability for the CTG readout and >1.5 fold difference for the TXL assay relative to control siRNA.

2.2.5 RT-PCR

Cells were reverse transfected as described above and incubated with siRNAs for 72 hours at 37^0C and washed with 1X PBS using a plate washer (BIO-TEK ELx 405) before lysis. RNA was extracted using magnetic beads (Ambion, MagMax-96 Total RNA Isolation Kit, Cat # 1830) according to the manufacturer's protocol. Total RNA concentration of the samples was measured using a NanoDrop-1000 spectrophotometer. BioRad's iScript cDNA Synthesis Kit (Cat # 170-8891) is used for cDNA synthesis and reactions were run on MJ Research's DNA Engine Tetrad Peltier Thermal Cycler according to the manufacturer's recommendation. 5ng final concentration of cDNA was used per 10ul qPCR reaction volume. Gene expression was determined using TaqMan® probe chemistry (ABI, Foster City, CA, USA) and qPCR was run on an ABI 7900HT Fast Real-time PCR System. The reactions were carried out in triplicate per sample with endogenous (GAPDH), buffer, scrambled and non-template controls. Gene expression values were normalized to GAPDH and calculated by the relative quantification (RQ) method ($\Delta\Delta C_T$ method) using ABI's SDS RQ Manager 1.2 software. Knockdown of a gene of interest by a particular siRNA relative to endogenous expression is given by:

$$(\Delta C_T)_{test} = \text{Average Target Gene } C_T - \text{Average GAPDH } C_T$$

$$(\Delta C_T)_{control} = \text{Average Target Gene } C_T - \text{Average GAPDH } C_T$$

$$\Delta\Delta C_T = (\Delta C_T)_{test} - (\Delta C_T)_{Control}$$

$$RQ = 2^{-\Delta\Delta C_T}$$

$$\%KD = (RQ_{si} - RQ_{buffer})*100/RQ_{buffer}$$

where 'test' refers to siRNA treated (si) or buffer control (buffer); 'control' refers to scrambled siRNA control, a negative control. RQ_{si} and RQ_{buffer} are calculated as shown above to determine relative gene expression values for a target of interest with and without (endogenous levels) siRNA treatment, respectively. These RQ values are then used to measure % knockdown (KD). Standard error is estimated by calculating $2^{-(\Delta\Delta C_T + SD)}$ and $2^{-(\Delta\Delta C_T - SD)}$. Further details of using QPCR to measure target gene expression knockdown in RNAi experiments are given in [120].

2.2.6 High-content imaging

Live/dead assays involving each cell line were prepared under the same parameters established for each cell line in HTS. Live and dead cells were determined by two widely used fluorescent probes, calcein acetoxymethyl (calcein AM) and ethidium bromide homodimer (EthD). Live cells are determined by the enzymatic conversion of the non-fluorescent, membrane-permeable calcein AM to calcein, a polar lipid-insoluble green fluorescent product that is retained by viable cells. On the other hand, EthD enters cells with damaged membranes and undergoes a 40-fold enhancement of fluorescence upon binding to nucleic acids, thereby producing a red fluorescence in dead cells. Microscopy images were captured and analyzed using the IN Cell Analyzer 3000 (GE Healthcare, Piscataway, NJ, USA) and analyzed using the cell viability analysis module.

Briefly, the analysis module uses colors of the fluorescent dye assays and reports viability and/or toxicity events through changes in fluorescent intensity. The algorithm requires the use of a fluorescent marker dye such as Hoechst (blue) to identify each nucleus as an individual object or cell (object definition) in the image based on user defined thresholds. Once the thresholds are set, the algorithm identifies every object surrounded by a white mask and gives a total object count output. Next, the green channel (Calcein AM) is used to detect number of live cells and the red channel (EtDh-1) is use to detect dead cells. In this manner, high content images were captured for ILK and AURKA inhibition by 4 siRNAs in Calu6, HCT116, MCF7 and U87 cells.

3 Results and Discussion

3.1 Genome-wide RNAi profiling to determine contexts of vulnerability in cancer cells

A genome-wide loss-of-function siRNA screen of druggable targets (~7000 genes) in 4 cell lines – Calu6 (lung), HCT116 (colon), MCF7 (breast), U87 (glioblastoma) – representing different tumor types and genetic backgrounds was performed (see Figure 4). Following an extensive assay development phase, high throughput screens (HTS) were implemented using a cell viability assay readout using 2 siRNAs per target. Screen actives that caused significant lethality ('essential' or 'survival' genes) in a general or cell-specific context were analyzed by integrative and systems approaches with a variety of oncogenomics datasets (mutation, arrayCGH, microarray data) to identify targets and pathways that cancer cells depend on for proliferation, survival and evasion of cell death. Several genes were confirmed and validated by additional siRNAs, assays and cell lines. Our results provide a rich repertoire of rational targets and druggable pathways/networks to tailor existing or future cancer therapies.

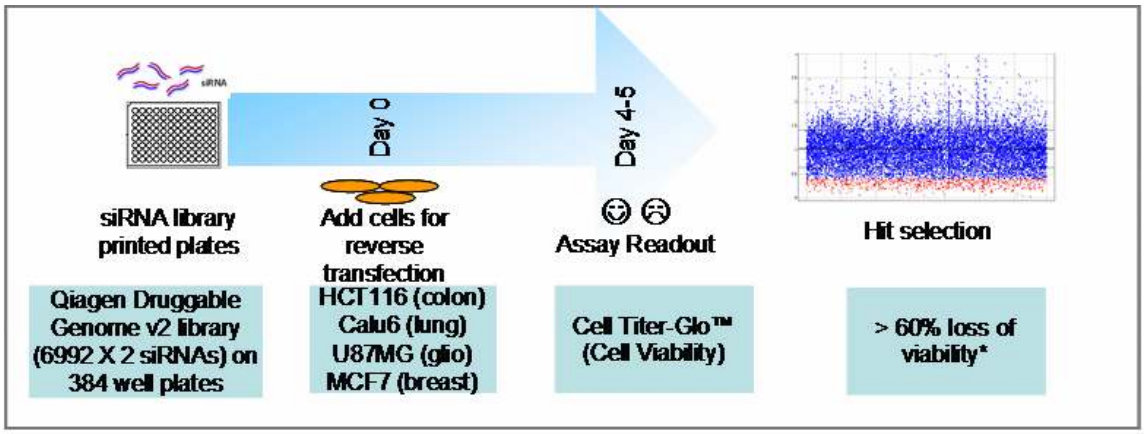


Figure 4. Experimental design of high throughput cell-based RNAi screen. Following an extensive phase of assay development and validation where a variety of assay parameters were optimized, HTS in 384-well plate format was carried out in duplicate. siRNAs from the Qiagen Druggable Genome v2 library (6992 targets X 2 siRNAs each) were printed on 384-well plates and reverse transfected into each of the 4 indicated cell lines on Day 0 and 4-5 days later, cell viability readouts were measured by Cell Titer-Glo assay, a cell viability assay. HCT116, Calu6, U87MG and MCF7 are colon, lung, glioblastoma, breast cancer cell lines, respectively. *Details of cut-offs for hit selection are explained elsewhere.

3.1.1 Distribution of hits and hit selection

From a histogram analysis of hits from HTS for each cell line (see Figure 5) a right skewed distribution is observed that suggests that the screen results are biased towards genes that are essential for cell viability. This may not be too surprising since the library used in the primary screen contained targets from the druggable genome. As expected, most of siRNAs have little effect on cell survival (~70-80% have minimal effect) as seen in most high throughput screens where outliers correspond to significant hits that drastically increase or decrease cell survival. In our case, we focused on identifying siRNA hits that correspond to genes essential for cell survival.

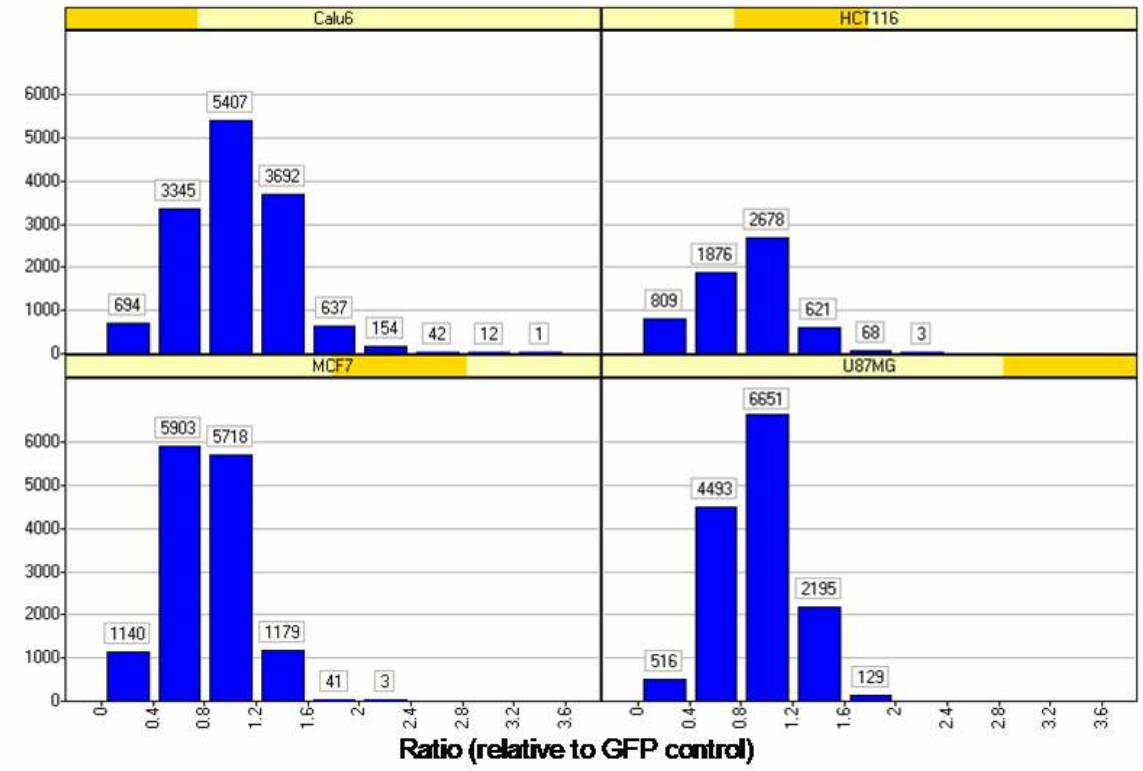


Figure 5. Distribution of hits from HTS for each cell line. Average control (i.e. GFP) normalized values are shown along the X-axis and the frequency counts are shown on the Y-axis. Only half the library was used for the HCT116 HTS run.

While many population-based statistical methods for determining a cutoff may exist, based on prior experience we applied an arbitrary threshold of >60% loss of viability to average normalized readouts for each cell line independently to select hits. These results

for Calu6, MCF7, HCT116 and U87MG are summarized in Figure 6. Our chosen cutoff provided an average hit rate of <15% which is consistent with similar published reports of large scale RNAi screens.

We then proceeded to perform unsupervised hierarchical clustering of these hits and clusters of siRNAs that that are broadly toxic to all 4 cell lines and selectively toxic to one or more cell lines were observed. These represent and will be referred to as 'general' and 'cell/context-specific' survival genes henceforth (Figure 6).

a.

	Calu6	MCF7	HCT116	U87MG
# of siRNAs/ genes in QDGv2 library	13984/ 6992	13984/ 6992	6055/ 3027	13984/ 6992
# of siRNAs/ genes causing > 60% loss of viability	694/ 654 (~4%)	1140/ 1040 (~8%)	809/ 740 (~13%)	516/ 489 (~4%)

*Only half the library was used for the HCT116 HTS run

b.

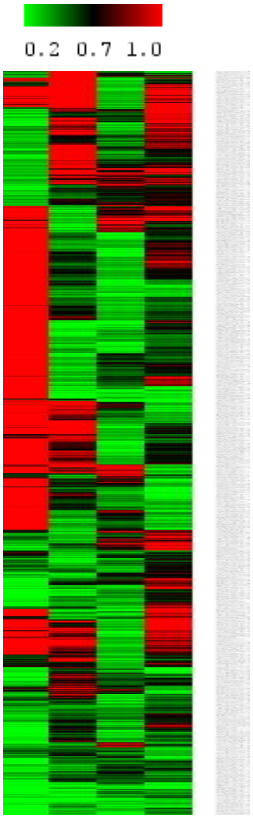


Figure 6. Hit selection and patterns of lethality. (a) An empirical cut-off value of 0.4 was applied to the control normalized averaged HTS readouts from each cell line. The breakdown of these filtered siRNAs or gene hits that cause >60% loss of viability in Calu6, MCF7, HCT116 and U87MG and the respective hit rates (in parentheses) are shown. (b) Hierarchical clustering of these hits reveals patterns of general and cell or context-selective lethality. Normalized (with respective to GFP) values from green (high killing) to red (low killing) are shown in the heat map where individual siRNAs are depicted along the rows and cell lines are along the columns.

3.1.2 General survival genes

By teasing out the pattern of siRNA hits that were broadly toxic (>60% lethality with respect to control) across all cell lines, a total of 147 genes were selected. We applied the following stringent cut-off criteria to reduce false positives and maximize high confidence hits: <40% viability in 2 or more cell lines where phenotypic concordance for duplicate siRNAs was seen in at least 2 cases (19genes, see Figure 7 and Table 1) or <40% viability in 3 or more cell lines where single hits occurred. Since siRNAs are known to have non-specific off-target effects, in the absence of target specific knockdown information, observing reagent redundancy in cellular phenotypic outcome is a well accepted criterion for target specific effects [86]. In other words, when 2 siRNAs show concordant phenotype (i.e. significant loss of viability, here) it increases the confidence in the gene hit. This was enabled by the use of 2 siRNAs per target in the primary screen. Therefore, we applied this parameter in generating lists for general and cell-specific survival genes (see 3.1.3). Arguably, when phenotypic siRNA concordance is seen in 2 or more cell lines for the same target, the emerging gene list is likely to be highly robust. 19/147 genes were, therefore, short-listed in this manner (see Figure 7 and Table 1).

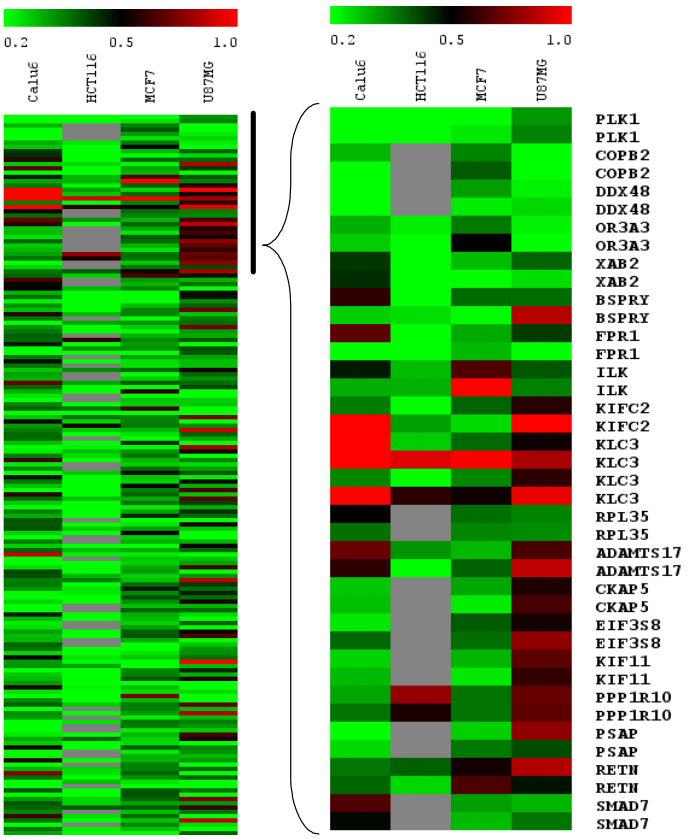


Figure 7. General survival genes. (a) A total of 147 genes were selected for <40% viability in 2 or more cell lines (b) 19 genes that are a subset of this list cause a significant lethal effect in 2 or more cell lines with siRNA phenotypic concordance in at least 2 cases. Green represents high degree cell kill, red represents insensitivity to cell kill and missing values are shown in grey.

Table 1: 19 short-listed essential genes. siRNAs targeting these genes show a broad range of toxicity (selected for <40% viability in 2 or more cell lines with phenotypic concordance in at least 2 cases).

Gene	Description
ADAMTS17	ADAM metallopeptidase with thrombospondin type 1 motif, 17
BSPRY	B-box and SPRY domain containing
CKAP5	cytoskeleton associated protein 5
COPB2	coatomer protein complex, subunit beta 2 (beta prime)
EIF3S8	eukaryotic translation initiation factor 3, subunit C
DDX48	eukaryotic translation initiation factor 4A, isoform 3
FPR1	formyl peptide receptor 1
ILK	integrin-linked kinase
KIF11	kinesin family member 11
KIFC2	kinesin family member C2
KLC3	kinesin light chain 3
OR3A3	olfactory receptor, family 3, subfamily A, member 3
PLK1	polo-like kinase 1 (Drosophila)
PPP1R10	protein phosphatase 1, regulatory (inhibitor) subunit 10
PSAP	prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy)
RETN	resistin
RPL35	ribosomal protein L35
SMAD7	SMAD family member 7
XAB2	XPA binding protein 2

3.1.3 Cell-specific survival genes

To identify contexts of vulnerability, we filtered siRNAs that showed <40% viability relative to control (GFP) in one cell line but were relatively less toxic in other cell lines. Using the rationale for siRNA concordance explained above, we picked out genes where the cognate siRNAs showed consistent cell kill phenotype for the cutoff used in a particular cell line. In cases, where concordance was not observed, single hits demonstrating <40% viability for a given cell line, but >80% viability for the remaining cell lines were picked. Within the limits of the chemiluminesence based Cell Titer Glo assay, this represents a minimum 2-fold window of selectivity over other cell lines. Together, these hits comprise the list of cell-specific or context-specific survival genes: 59hits (38genes) in Calu6, 134 hits (74 genes) in HCT116, 233 hits (162 genes) in MCF7 and 54 hits (39 genes) in U87 (Table 2). Out of these 21, 57, 145 and 22 genes showed siRNA concordance for Calu6, HCT116, MCF7 and U87, respectively (Table 2). It is interesting to note that although half the library was screened for HCT116, a relatively large set of cell selective survival genes were observed. This could be attributed to the

MMR deficient nature of HCT116 which causes genomic instability. Several groups [121, 122] have similarly reported the highly variable nature of MCF7 which may partially explain the large number of survival genes found in this cell line. Heat map visualizations illustrating these results along with the identities of these can be seen below (Figure 8 and Table 2).

Due to off-target effects of siRNAs, variable transfection and technical and/or biological variation, there are likely several false positive still, which can be eliminated by further testing with additional siRNAs. That being true, it is also possible to prioritize targets of interest by integrating this data (general or cell specific essential genes) with other genomics datasets.

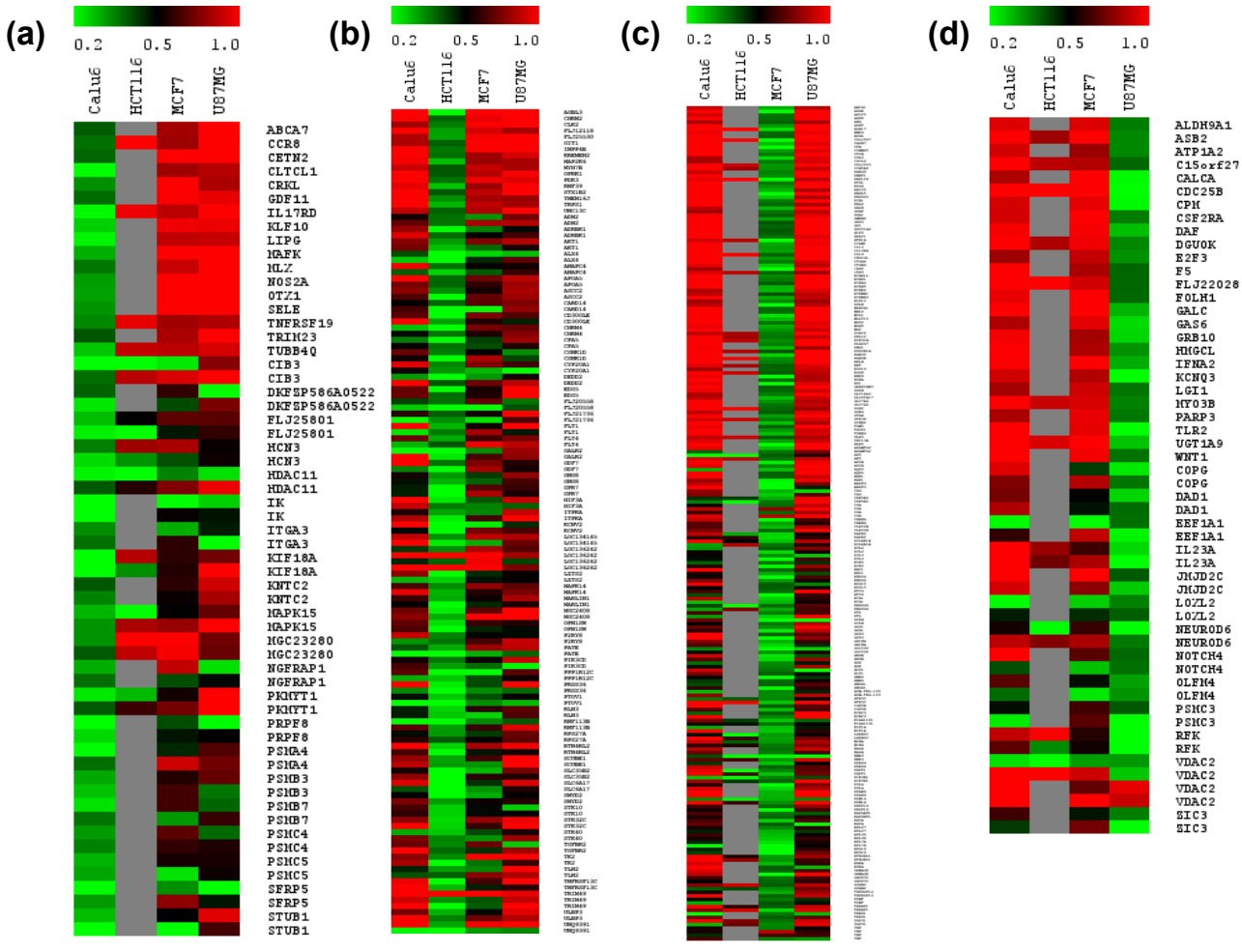


Figure 8. Cell-specific survival genes. siRNA hits that demonstrated >60% lethal effect (relative to control) and strong selectivity for each cell line were picked. Details on filtering can be found in the text

above. Heatmap visualizations for 59 hits in Calu6 (a), 134 hits in HCT116 (b), 233 hits in MCF7 (c) and 54 hits in U87 (d) runs are presented. Green represents high degree cell kill, red represents insensitivity to cell kill and missing values are shown in grey.

Table 2: List of cell specific survival genes. Genes, inhibited by siRNAs, that demonstrated >60% lethal effect (relative to control) and strong selectivity for each cell line were picked. Details on filtering can be found in the text above. The number of genes filtered for each cell line screened are indicated in brackets and those with duplicate siRNA hits are marked with a *.

Calu6 (38)	U87 (39)	HCT116 (74)		MCF7 (162)			
ABCA7	ALDH9A1	ADM2*	MARLIN1*	ABCG1	EPS8	JPH3	RET
CCR8	ASB2	ADRBK1*	MGC2408*	ADAMTS2*	ERCC5	KCNC3*	RGS14
CETN2	ATP1A2	AGBL3	MYH7B	ADAR	ETV4*	KCNE1L	RGS8
CIB3*	C15orf27	AKT1*	OPN1SW*	ADCY5	EYA1*	KCNF1	RHOA*
CLTCL1	CALCA	ALX4*	OPRK1	ADFP	FBXL5	KCNH4	RNF6
CRKL	CDC25B	ANAPC4*	P2RY8*	AK5*	FBXO22	KCNH5	RORA
DKFZP586A0522*	COPG*	APOA5*	PATE*	ANG	FBXO44*	KCNK2	RP1
FLJ25801*	CPM	ASCC2*	PDK3	APOH*	FCN1	KCNMB1	RPL27*
GDF11	CSF2RA	CARD14*	PIK3CD*	APRT	FRS2	KCNMB4	RPL36*
HCN3*	DAD1*	CD300LE*	PPP1R12C*	AQP2*	FTL*	KIAA1161*	RPL7A*
HDAC11*	DAF	CHRM2	PRSS36*	ASB17	GALE	KIF1A*	RPS19*
IK*	DGUOK	CHRM4*	PTOV1*	BAK1*	GCDH*	KIF1C	RTN4RL1*
IL17RD	E2F3	CLK2	RLN3*	BMP4	GDNF	LDLR	RXRA*
ITGA3*	EEF1A1*	CPA5*	RNF113B*	BPHL	GGA1	LRRN6C*	SEMA4D*
KIF18A*	F5	CSNK1D*	RNF39	BRPF3*	GMEB2	MAN2A1	SERPINB3
KLF10	FLJ22028	CYP20A1*	RPS27A*	C9orf97	GNG3	MBL2	SGSH
KNTC2*	FOLH1	DEDD2*	RTN4RL2*	CA4*	GPI	MCL1	SH2D3C*
LIPG	GALC	EDG5*	SCUBE1*	CAPN7	GPX1*	MINA*	SLC18A3
MAFK	GAS6	FLJ12118	SLC35B2*	CDK5R2*	GPX3*	MLLT10	SLC25A17
MAPK15*	GRB10	FLJ20558*	SLC6A17*	CFH	GRINA*	MPP2	SLC7A4
MGC23280*	HMGCL	FLJ21736*	SMYD2*	CGA*	GUCY1A2	MSH5	SLC7A6
MLX	IFNA2	FLJ25530	STK10*	CHRNG*	GUCY2F*	MX2	SSH2
NGFRAP1*	IL23A*	FLT1*	STK32C*	CLEC2B*	H1F0	NAGA*	SSR4
NOS2A	JMJD2C*	FLT4*	STK40*	COMMD5	HDAC1	NME3*	STAR
OTX1	KCNQ3	GALR2*	STX1B2	CTSW	HEXB*	P2RY6	STK38
PKMYT1*	LGI1	GDF7*	TGFBR2*	CUL5	HGD*	PDE6G*	SUHW2*
PRPF8*	LOXL2*	GIT1	TK2*	CXCL6	HIP1*	PELI2	SYNE2
PSMA4*	MYO3B	GNG8*	TLN2*	CXorf23	HMBS*	PHPT1*	TGM5
PSMB3*	NEUROD6*	GPR7*	TMEM16J	CYB5R2	HMGA1*	PIK3CA	THOP1
PSMB7*	NOTCH4*	HIF3A*	TNFRSF13C*	DAPK2*	HOM-TES-103*	PIK3R1*	TNFRSF14
PSMC4*	OLFM4*	INPP4B	TRIM49*	DCLRE1B*	HTR1A	PLA2G7	TPMT*
PSMC5*	PARP3	ITPKA*	TRPS1	DHRS9	HTR3C*	PMS1	TRERF1*
SELE	PSMC3*	KCNV2*	ULBP3*	DMBT1	ICAM1	POLA*	TREX1*
SFRP5*	RFK*	KREMEN2	UNC13C	DRD1IP	IGF2R*	PPARG*	TUBB4
STUB1*	TLR2	LOC134145*	UNQ9391*	DTX1	IL10	PPM1G*	ULK3
TNFRSF19	UGT1A9	LOC136242*		DVL2*	IL10RA	PPP2R1A	UNC13B
TRIM23	VDAC2*	LZTS2*		DVL3*	IL19	PRPF19*	USP31*
TUBB4Q	WNT1	MAP2K6		ECE2*	INSIG1	RAB30	VWF*
	ZIC3*	MAPK14*		ENC1*	ITGAX	RAB9B	WDR5
				ENDOG*	ITGB6	RAPGEF1*	
				EPS15*	JAG2	RELN	

3.1.4 Integration with array-based comparative hybridization data

We proceeded to overlay general and cell specific survival genes with genome-wide copy number alterations derived from array-based comparative hybridization (aCGH) profiling. Genomic DNA from a panel of cell lines, including 4 used in our study, were hybridized onto 44A/B Agilent CGH oligo arrays and a novel method was employed to detect copy number changes (Xiang Y et al., unpublished). By combining copy number information

with lethality caused by siRNA targeted knockdowns from the primary screen, we can identify target dependence that is governed by genomic aberrations such as amplification and partially deleted loci, conferring a survival advantage. Figure 9(a) depicts hits from the primary screen causing <40% viability that have a corresponding high copy (>2 copies) gain or loss (1 copy losses representing putative loss-of-heterozygosity events). This is detailed for genes in our cell specific lethal gene list in Table 5. It is highly likely that several targets are missed by the thresholds applied and the library or aCGH platform used, but their dependence could be attributed to other genetic or epigenetic mechanisms such as mutation or expression. Nonetheless, interesting examples were noted among genes that were broadly lethal upon inhibition. TPX2 is phosphorylated by Aurora kinase A (STK6/AURKA) and is involved in G2/M checkpoint by regulating mitotic spindle assembly and centrosome duplication. TPX2 is found in high copy number for the first time, to our knowledge, in Calu6 aCGH data and has been reported to be commonly amplified in lung cancers [123]. Although TPX2 (20q11) was not a Calu6 cell-specific hit in our analysis, knocking down the gene leads to potent cell killing (see Figure 9). 20q amplifications are widely reported in a variety of neoplasias, including breast, colon, bladder, ovarian and pancreatic (references in [124]). It is therefore interesting to note that, AURKA (20q13) was identified as an essential gene whose inhibition resulted in significant killing of Calu6, MCF7, HCT116 cells and relatively less for U87 (Confirmation in Figure 16). MCF7 cells in particular, carry high gains of AURKA (see Figure 9). Furthermore, several studies have utilized AURKA siRNAs as positive cell kill controls. Efforts are underway to design inhibitors for Aurora kinase inhibitors which would presumably have broad activity in several tumors.

Since several gains or losses are likely to be ‘passenger’ events and several hits from a screen are likely to be false positives, by integrating these two orthogonal data types, we can determine ‘driver’ events that may be causally responsible for ‘true’ screen positives. Such genes with frequent aberrations in human tumors can be prioritized for therapeutic intervention.

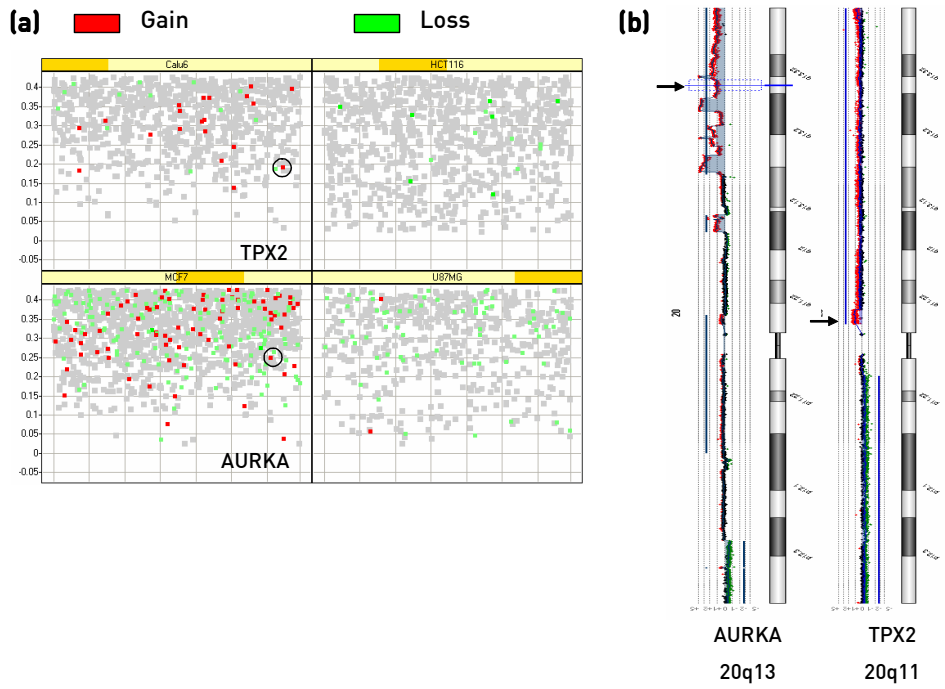


Figure 9. Integration of screening data with array-based comparative hybridization. (a) Scatter plots of siRNA hits that cause loss of viability (values < 0.4, as described above, are shown along the Y-axis) are overlayed with high gain (red) and loss calls (green) of gene lists for each corresponding cell line from array-based CGH data generated from Agilent 244K oligo chips. (b) 3 examples are shown with corresponding genomic copy number alteration (assessed by normalized $\log_2$ ratios) views. AURKA and TPX2 are amplified and cause low survival upon siRNA inhibition in MCF7 and Calu6 cell lines, respectively.

3.1.5 Integration with mutation data

Several high throughput mutational screens have been recently performed as described in Introduction. We compiled data published by the Singapore Oncogenome Group (SOG) [22], Sanger's COSMIC effort (http://www.sanger.ac.uk/genetics/CGP/) and the Johns Hopkins group (JHU) [20, 21]. From our list of general survival genes, 34 unique genes (see Table 3) mapped to mutations occurring in a broad range of tumor types from COSMIC or Hopkins collections. KDR, also known as VEGFR/VEGFR2, is one such example. It is the receptor tyrosine kinase that binds VEGF and result in epithelial to mesenchymal transition in the tumor microenvironment by endothelial cell activation and altering tumor vasculature. As an angiogenesis target, it is integral to a variety of drug discovery projects in various companies. Mutations in kidney and lung cancers were identified by Sanger while JHU identified several mutations in colorectal primary tumor

samples. In the SOG dataset, 2 somatic mutations were found in skin cancer cell lines – K107K in BOW-G and P1280S in MM-Du. In our screen, KDR siRNA caused >60% lethality in Calu6, U87 and HCT116 cells and ~ 45% killing in MCF7 cells. Interestingly, Avastin which is a clinically approved anti-VEGF monoclonal antibody is effective against a variety of cancers as well. These facts suggest that KDR inhibitors are likely to have broad spectrum activity.

When overlapping our lists of cell-specific genes, we were interested in commonly mutated genes of the same tissue type (see Table 5). No mutations were found in cognate tumors for Calu6 and U87 specific lethal genes. ADAR (JHU), DVL3 (JHU), GUCY2F (Sanger and JHU), PIK3CA (Sanger and JHU), PIK3R1 (JHU), SYNE2 (JHU), ULK3 (Sanger), VWF (JHU) came up as survival genes in the MCF7 screen and showed corresponding mutations in breast tumors from the indicated sources. PIK3CA is mutated in both MCF7 and HCT116 cells, yet appears to cause relatively higher killing in MCF7 cells. This can be explained by the KRAS mutation in HCT116 cells which may render them less sensitive to PI3K inhibition alone. ANAPC4 (JHU), LZTS2 (JHU), MARLIN (JHU), STK32C (JHU) and TGFBR2 (Sanger and JHU) were similarly identified from HCT116 RNAi screen and mutations in colorectal samples. ANAPC4 is a subunit of the anaphase promoting complex and is critical for cell cycle. Interestingly, ANAPC4 was seen in a shRNA dropout screen using pools of half hairpin barcodes in cell lines including HCT116 and DLD1 [96]. Intriguingly, LZTS2 is a leonine zipper tumor suppressor gene that regulates cell cycle and TGFBR2 is another tumor suppressor that is commonly truncated in colorectal tumors with microsatellite instability (MSI). Collectively, these targets are probably highly relevant to breast and colorectal subpopulations.

Table 3: Mutations identified in general survival genes from 3 published data sources. SOG refers to somatic mutations identified by the Singapore Oncogenome Group [22]; Sanger refers to cell line and tumor mutations published by COSMIC (http://www.sanger.ac.uk/genetics/CGP/); JHU refers to mutations in colorectal and breast transcriptomes published John Hopkins University [20, 21].

Gene Symbol	SOG	Sanger	JHU
AKAP13		X (kidney)	
ALS2CR2			
AMPD2			X (breast)
ARHGEF4		X (skin; breast)	X (breast)
AURKA		X (large_intestine; lung; skin)	
BSPRY			X (breast)
CACNA1H			X (breast)
CLCN1			X (breast)
EPHA5	X (bladder; colorectal; breast; lung; skin; ovary; stomach)	X (lung)	
FASTK		X (lung; stomach)	
FLJ23356		X (lung)	
FZR1		X (lung)	
GRK4		X (lung)	
JMJD1C			X (breast)
KCNK17		X (skin)	
KDR	X (skin)	X (kidney; lung)	X (colorectal)
KHSRP			
KIAA0664			X (breast)
KIAA1632			X (breast)
MAML2			
MAP3K1		X (skin; ovary)	
PKN1			X (breast)
PPFIA4		X (skin)	
PPP1R10		X (lung; kidney)	
PTK9L		X (lung; ovary)	
RACGAP1		X (kidney)	
SLAMF1			X (breast)
SMG1		X (breast; lung; stomach)	
TAOK3		X (lung)	
TESK1		X (breast)	X (breast)
TGM2			X (colorectal)
TRIB1		X (lung)	
XAB2		X (brain; skin)	X (breast)
ZNRF4			X (colorectal)

3.1.6 Integration with clinical outcome

To understand the clinical relevance of the various targets identified in this study, we cross-referenced the essential and cell specific survival gene lists against publicly available compendia of primary tumor microarray gene expression datasets with associated clinical covariates to see if they stratify patient subgroups informatively.

Datasets corresponding to lung, breast, glioma, medulloblastoma, head and neck tumors (see Materials and Methods) were downloaded and analyzed by Kaplan Meier curves. Significant ($p<0.05$) associations of survival were determined by grouping patient expression profiles on median, quantile or extreme quantile expression values of the target of interest. 50 general essential genes correlated with poor outcome in lung, ovarian, breast, brain or lymphoma cancers. 7 Calu6 specific survival genes showed poor survival in primary lung cancer datasets. 11 U87 specific survival genes showed poor survival in brain tumor datasets. 34 MCF7 specific survival genes showed poor outcome in primary breast cancer compendia. Since gene expression and clinical information coupled datasets were not accessible for colorectal tumors, we did not query for HCT116 specific gene list. MCL1 is anti-apoptotic protein whose overexpression in a variety of breast cancers has been linked to poor survival [125] and in this regard, it is interesting to note MCL1 in the list of MCF7 specific survival genes. Figure 10 shows a set of representative targets in general and cell-specific lists where overexpression from microarray profiling data corresponds to poor prognosis. This implies that inhibition of such targets is likely to enhance clinical outcome. Therefore, these clinically relevant genes make ideal targets for drug discovery.

Table 4: List of clinically relevant general survival genes. Summary of Kaplan Meier survival analysis of 50 genes, a subset of those that are broadly toxic upon siRNA knockdown. The number of probesets corresponding to a given gene in each of the tumor types (see Materials and Methods) analyzed are shown.

Gene	Total probesets	Total tumor type	Lung	Ovarian	Lymphoma	Breast	Brain
CD47	9	5	1	2	3	1	2
FASTK	4	4	0	1	1	1	1
PHKA2	5	3	3	0	1	0	1
FXYD5	5	3	0	1	2	0	2
HSDL2	5	3	3	0	0	1	1
KIF11	4	3	1	0	0	2	1
SMG1	4	3	2	0	0	1	1
COPB2	4	3	1	0	2	0	1
OGFR	4	3	1	2	0	0	1
PFKL	4	3	1	0	0	2	1
CKAP5	3	3	1	0	1	1	0
FZR1	3	3	1	1	1	0	0
RPL35	3	3	1	0	0	1	1
TAOK3	3	3	1	1	1	0	0
TGM2	3	3	0	1	1	1	0
WDHD1	6	2	0	0	0	3	3
PLK1	5	2	0	0	0	3	2
MARVELD3	5	2	3	0	0	2	0
OPRS1	5	2	0	0	2	3	0
KIF9	5	2	2	0	0	0	3
AKAP13	4	2	3	0	0	0	1
SNRP70	4	2	3	0	0	0	1
TPCN2	4	2	3	0	0	1	0
TSHR	4	2	1	3	0	0	0
ADORA3	3	2	1	0	0	2	0
ARHGEF4	3	2	2	0	0	0	1
MRPS17	3	2	0	0	0	2	1
PSPH	3	2	0	0	0	1	2
RACGAP1	3	2	0	0	0	2	1
SFRS2IP	3	2	0	1	0	0	2
XAB2	3	2	2	1	0	0	0
DDX54	3	2	1	0	0	0	2
ECE1	3	2	0	2	0	0	1
MGST1	3	2	0	0	0	2	1
ALDH16A1	2	2	0	0	0	1	1
BSPRY	2	2	1	0	1	0	0
EPHA5	2	2	0	0	0	1	1
FBXL18	2	2	1	1	0	0	0
FZD5	2	2	0	0	1	0	1
GEMIN6	2	2	0	0	0	1	1
HK1	2	2	1	0	0	1	0
LMBR1L	2	2	1	1	0	0	0
MUC17	2	2	1	0	0	0	1
NACA	2	2	0	0	0	1	1
OR3A3	2	2	0	1	0	0	1
PKN1	2	2	0	1	0	1	0
PSAP	2	2	0	1	0	1	0
SPATA20	2	2	0	0	1	1	0
TNFAIP2	2	2	1	0	0	0	1
TOP1MT	2	2	1	0	0	0	1

Table 5: Cell-specific survival genes with associated information. Subsets of the indicated cell-specific survival genes with corresponding mutations (as described in Table 3), copy number alterations (based on internal aCGH data) in the same cell line context or association with survival in the same tumor type (see Materials and Methods) are shown below. * refers to findings in [126].

Gene Symbol	Gene Description	Cell line	Mutation?	Copy number alteration?	Association with survival?
GDF11	growth differentiation factor 11	Calu6			Yes
IL17RD	interleukin 17 receptor D	Calu6			Yes
KLF10	Kruppel-like factor 10	Calu6			Yes
KNTC2	Kinetochore associated 2	Calu6			Yes
PRPF8	PRP8 pre-mRNA processing factor 8 homolog (S. cerevisiae)	Calu6			Yes
PSMA4	proteasome (prosome, macropain) subunit, alpha type, 4	Calu6			Yes
TNFRSF19	tumor necrosis factor receptor superfamily, member 19	Calu6			Yes
ADRBK1	adrenergic, beta, receptor kinase 1	HCT116	Sanger		
AKT1	v-akt murine thymoma viral oncogene homolog 1	HCT116	*		
ANAPC4	anaphase promoting complex subunit 4	HCT116	JHU		
LZTS2	leucine zipper, putative tumor suppressor 2	HCT116	JHU		
MARLIN1	janus kinase and microtubule interacting protein 1	HCT116	JHU		
PDK3	pyruvate dehydrogenase kinase, isozyme 3	HCT116	JHU		
STK32C	serine/threonine kinase 32C	HCT116	JHU		
TGFBR2	transforming growth factor, beta receptor II (70/80kDa)	HCT116	JHU; Sanger		
ABCG1	ATP-binding cassette, sub-family G (WHITE), member 1	MCF7		Gain	
ADAR	adenosine deaminase, RNA-specific	MCF7	JHU		
ADFP	Adipose differentiation-related protein	MCF7			Yes
APOH	Apolipoprotein H (beta-2-glycoprotein I)	MCF7		Gain	
APRT	adenine phosphoribosyltransferase	MCF7			Yes
BRPF3	Bromodomain and PHD finger containing, 3	MCF7			Yes
CA4	carbonic anhydrase IV	MCF7		Gain	
CLEC2B	C-type lectin domain family 2, member B	MCF7			Yes
COMMD5	COMM domain containing 5	MCF7			Yes
CUL5	cullin 5	MCF7		Loss	
DCLRE1B	DNA cross-link repair 1B (PSO2 homolog, S. cerevisiae)	MCF7		Gain	
DVL3	dishevelled, dsh homolog 3 (Drosophila)	MCF7	JHU		
ENDOG	endonuclease G	MCF7			Yes
ERCC5	excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	MCF7		Loss	
FBXO22	F-box protein 22	MCF7			Yes
GALE	UDP-galactose-4-epimerase	MCF7			Yes
GGA1	golgi associated, gamma adaptin ear containing, ARF binding protein 1	MCF7	JHU		
GPI	Glucose phosphate isomerase	MCF7			Yes
GPX1	glutathione peroxidase 1	MCF7			Yes
GRINA	glutamate receptor, ionotropic, N-methyl D-aspartate-associated protein 1 (glutamate binding)	MCF7			Yes
GUCY1A2	guanylate cyclase 1, soluble, alpha 2	MCF7		Loss	Yes
GUCY2F	guanylate cyclase 2F, retinal	MCF7	JHU; Sanger		
HDAC1	histone deacetylase 1	MCF7			Yes
HMBS	hydroxymethylbilane synthase	MCF7		Gain	
HMGA1	high mobility group AT-hook 1	MCF7			Yes
ICAM1	intercellular adhesion molecule 1 (CD54), human rhinovirus receptor	MCF7			Yes
IGF2R	insulin-like growth factor 2 receptor	MCF7			Yes
IL10	interleukin 10	MCF7		Gain	
IL10RA	interleukin 10 receptor, alpha	MCF7		Gain	
IL19	interleukin 19	MCF7		Gain	
INSIG1	insulin induced gene 1	MCF7			Yes
ITGB6	integrin, beta 6	MCF7			Yes
KCNH4	potassium voltage-gated channel, subfamily H (eag-related), member 4	MCF7			Yes
KCNK2	potassium channel, subfamily K, member 2	MCF7		Gain	Yes
KIAA1161	KIAA1161	MCF7	JHU		
MCL1	myeloid cell leukemia sequence 1 (BCL2-related)	MCF7		Gain	Yes
MLLT10	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 10	MCF7			Yes
MX2	myxovirus (influenza virus) resistance 2 (mouse)	MCF7		Gain	
PIK3CA	Phosphoinositide-3-kinase, catalytic, alpha polypeptide	MCF7	JHU; Sanger		
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	MCF7	JHU		
PLA2G7	phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)	MCF7			Yes
PPP2R1A	protein phosphatase 2 (formerly 2A), regulatory subunit A , alpha isoform	MCF7			Yes
RET	ret proto-oncogene	MCF7			Yes
RORA	RAR-related orphan receptor A	MCF7			Yes
RPL27	ribosomal protein L27	MCF7			Yes
RPS19	ribosomal protein S19	MCF7			Yes
RXRA	retinoid X receptor, alpha	MCF7			Yes
SERPINB3	serpin peptidase inhibitor, clade B (ovalbumin), member 3	MCF7		Loss	
SGSH	N-sulfoglucosamine sulfohydrolase (sulfamidase)	MCF7			Yes
SLC25A17	solute carrier family 25 (mitochondrial carrier; peroxisomal membrane protein, 34kDa), member 17	MCF7		Loss	Yes
STK38	Serine/threonine kinase 38	MCF7			Yes
SUHW2	suppressor of hairy wing homolog 2; zinc finger protein 280B	MCF7			Yes
SYNE2	spectrin repeat containing, nuclear envelope 2	MCF7	JHU		
VWF	von Willebrand factor	MCF7	JHU		
WDR5	WD repeat domain 5	MCF7			Yes
CPM	carboxypeptidase M	U87			Yes
CSF2RA	colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)	U87			Yes
E2F3	E2F transcription factor 3	U87			Yes
EEF1A1	eukaryotic translation elongation factor 1 alpha 1	U87			Yes
FLJ22028	pyridine nucleotide-disulphide oxidoreductase domain 1	U87			Yes
GALC	galactosylceramidase	U87			Yes
GRB10	growth factor receptor-bound protein 10	U87			Yes
IFNA2	interferon, alpha 2	U87		Loss	
LOXL2	Lysyl oxidase-like 2	U87			Yes
OLFM4	olfactomedin 4	U87			Yes
PARP3	poly (ADP-ribose) polymerase family, member 3	U87			Yes
RFK	riboflavin kinase	U87			Yes

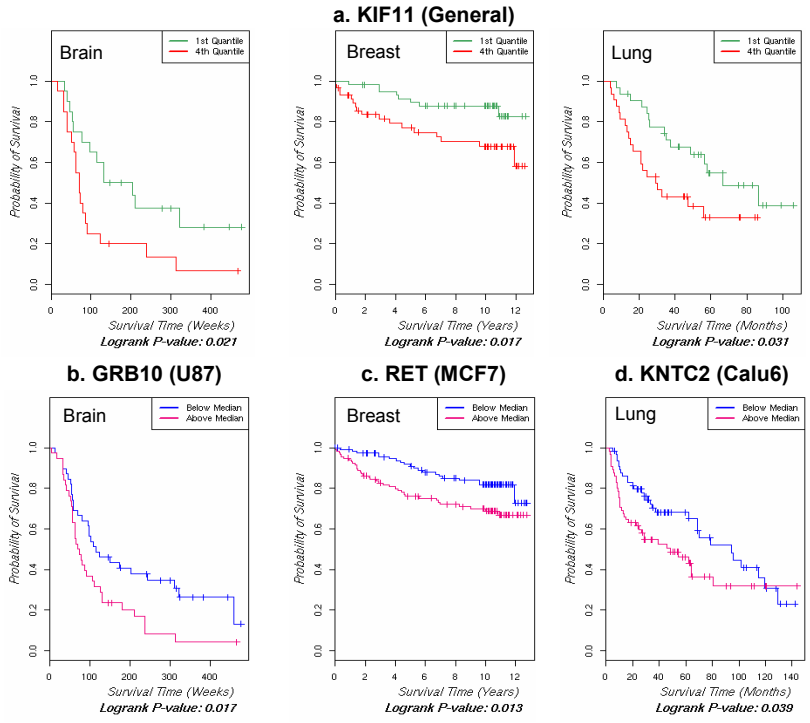


Figure 10. Kaplan Meier survival plots for a few representative screen actives. Shown above, are statistically significant (log-rank $p<0.05$) patient stratifications where over expression (from publicly available microarray gene expression data, see text and Materials and Methods for details) is associated with poor prognosis for a few select hits from general (a) and indicated cell-specific survival genes (b,c,d). (a) KIF11 (general survival gene) quantile expression patterns show significant and distinct survival in primary brain [75], breast [54] and lung [44] cancer datasets. Similar results were achieved upon grouping tumors based on median expression of (b) GRB10 (U87-specific), (c) RET (MCF7-specific), and (d) KNTC2 (Calu6-specific) in the respective tissue-specific cancer datasets, as indicated. Complete lists and gene descriptions can be found in Table 4 and

Table 5. For details on the method and datasets employed, see Materials and Methods.

3.1.7 Integration with pathways and networks

3.1.7.1 Pathway mapping results

By fusing pathway information with the general and cell specific survival gene lists, a mechanistic understanding of these 'contexts of vulnerability' can be achieved. Given the small numbers in each of the lists, statistical significance in terms of p-values is less informative; therefore, we focused on examining relative enrichment or over representation of members to rank canonical pathways (Ingenuity's IPA) or biological processes (NCI's DAVID).

Table 6: Pathway mapping of genes essential for cancer cell survival. Genes from Figure 7(a) were analyzed by GO (Biological Process, level 5) and relevant top scoring (Fold enrichment >=1.5) processes (containing >=5 genes) are shown below.

GO BP(5) Term	Count	Genes	Fold enrichment
GO:0016310~phosphorylation	21	NEK3, GRK4, ILK, PLK1, HIPK4, NRBP1, ATP6V1B1, EPHA5, TRIB1, MAP3K1, TESK1, ALS2CR2, SMG1, FLJ23356, PKN1, BRSK2, KDR, AURKA, TAOK3, LYK5, FASTK,	1.55
GO:0006396~RNA processing	10	ELAC2, TXNL4B, DDX54, SNRPD3, KHSRP, GEMIN6, SFRS2IP, SNRP70, XAB2, DDX48,	4.23
GO:0019932~second-messenger-mediated signaling	10	ADORA3, HTR6, TGM2, P2RY4, MC2R, TSHR, ADORA2A, FPR1, AURKA, TRAT1,	2.08
GO:0016071~mRNA metabolic process	9	TXNL4B, SMG1, SNRPD3, KHSRP, GEMIN6, SFRS2IP, SNRP70, XAB2, DDX48,	5.76
GO:0043549~regulation of kinase activity	7	PKIB, PKN1, TRIB1, ALS2CR2, FPR1, LYK5, CDK5R1,	1.69
GO:0009967~positive regulation of signal transduction	6	TGM2, CARD9, UBE2V1, TAOK3, TRAT1, TICAM2,	2.49
GO:0007017~microtubule-based process	6	KIFC2, KIF11, KIF9, KIF17, CKAP5, AURKA,	2.15
GO:0000087~M phase of mitotic cell cycle	6	NEK3, FZR1, KIF11, TXNL4B, PLK1, AURKA,	1.80
GO:0007067~mitosis	6	NEK3, FZR1, KIF11, TXNL4B, PLK1, AURKA,	1.80
GO:0006412~translation	6	MRPS17, RPL35, NACA, DDX48, KIAA0664, EIF3S8,	1.59
GO:0008284~positive regulation of cell proliferation	6	CD47, SLAMF1, ILK, TGM2, EFNB1, TSHR,	1.53
GO:0000279~M phase	6	NEK3, FZR1, KIF11, TXNL4B, PLK1, AURKA,	1.48

Gene Ontology offers a widely accepted and published curated systematic hierarchical classification of molecular function, cellular component, biological process for genes in the order of increasing specificity or complexity [127]. We examined level 5 terms of the GO Biological Process tree to identify the top common events that are at play in the general survival gene list (see Table 6). While there are several redundant and broadly classified biological processes, basic survival and metabolic functions are affected by genes in the general survival list. RNA processing and metabolism, cell cycle (mitosis) and several members of various growth factor signaling and their regulators feature

prominently. Cancer (39 genes) and cell death (37 genes) are the top significant functions and diseases associated with this list (Data not shown).

Table 7: Pathway mapping of genes essential for Calu6 cell survival. Genes from Figure 8(a) were analyzed by Ingenuity Pathway Analysis tools and relevant top scoring (-logP >1.3; p<0.05) canonical pathways (containing >5 genes) are shown below.

Pathway	Count	Genes	-Log(P-value)
Protein Ubiquitination Pathway	6	PSMB3 PSMB7 STUB1 PSMC4 PSMA4 PSMC5	5.04

Table 8: Pathway mapping of genes essential for MCF7 cell survival. Genes from Figure 8(c) were analyzed by Ingenuity Pathway Analysis tools and relevant top scoring (-logP >1.3; p<0.05) canonical pathways (containing >5 genes) are shown below.

Pathway	Count	Genes	-Log(P-value)
		NME3 PDE6G AK5 ADCY5 POLA1 GUCY2F GUCY1A2 ADAR	
Purine Metabolism	10	MPP2 APRT	2.59
Integrin Signaling	7	PIK3CA RHOA PIK3R1 CAPN7 ITGB6 ITGAX RAPGEF1	2.05
Acute Phase Response Signaling	6	PIK3CA FTL MBL2 APOH PIK3R1 VWF	1.83
G-Protein Coupled Receptor Signaling	6	PIK3CA ADCY5 P2RY6 PIK3R1 RGS14 HTR1A	1.57
Notch signaling pathway	5	DVL3 JAG2 DTX1 DVL2 HDAC1	2.18

Table 9: Pathway mapping of genes essential for HCT116 cell survival. Genes from Figure 8(b) were analyzed by Ingenuity Pathway Analysis tools and relevant top scoring (-logP >1.3; p<0.05) canonical pathways (containing >5 genes) are shown below.

Pathway		Molecules	-Log(P-value)
G-Protein Coupled Receptor Signaling	6	AKT1 ADRBK1 OPRK1 CHRM4 PIK3CD CHRM2	3.92
Inositol Phosphate Metabolism	5	MAP2K6 INPP4B CSNK1D PIK3CD ITPKA	3.69

We probed pathway information (in Ingenuity Pathway Analysis, see Materials and Methods) to cluster cell-specific lists cluster into unique pathways that may suggest predominant signaling pathways at play in different cancers. Sparse representation was found, perhaps due to the size of the query lists and the coverage of well annotated canonical pathways. That being said, several members (PSMC4, PSMB7, STUB1, PSMA4, PSMB3, PSMC5) of the proteolytic machinery in the protein ubiquitination pathway can be noted in the Calu6 specific survival list. U87 specific hits could not be collectively grouped into relevant pathways for the criteria employed. On the other hand, HCT116 and MCF7 specific survival lists contained relatively higher numbers of the genes, allowing us to see multiple hits that can be significantly binned into 2 and 5 pathways, respectively. Generic GPCR signaling pathways are represented in both lists. Interestingly, several members of the inositol phosphate metabolic pathway are essential

for the survival of HCT116 cells. On the other hand, as expected, depletion of purine metabolism in MCF7 cells confers loss of viability. Similarly, targeting Notch, integrin and acute response pathways can induce lethality. Taken together, these common pathways provide insight into biological mechanisms for genetic disruption and hence, can strengthen target selection and prioritization hypotheses for subsequent experimental follow up.

3.1.7.2 Functional interaction network analysis results

While canonical pathways provide valuable information, they can be sparse and restrictive in revealing the underlying web of functional interactions (gene-gene, gene-protein, protein-protein). To understand if the genes identified in our screen were truly connected we carried out a network analysis (see Materials and Methods). 116 genes from the general survival list were involved in 16 large or small networks. The top scoring network, containing 25 genes, depicting direct and indirect interactions is illustrated in Figure 11. Cancer, cell death, cellular growth and proliferation, cellular assembly and organization and cell morphology are the top significantly associated biological functions with this network. Several kinases that are essential for cell growth can be seen along with several members being targets for drugs that are marketed or in development. Two such examples AURKA and PLK1 are check point kinases involved in G2/M phase of the cell cycle and have been the subject of pharmacological investigations. Similar network analyses were performed for the cell-specific survival genes and results are shown in the figures below.

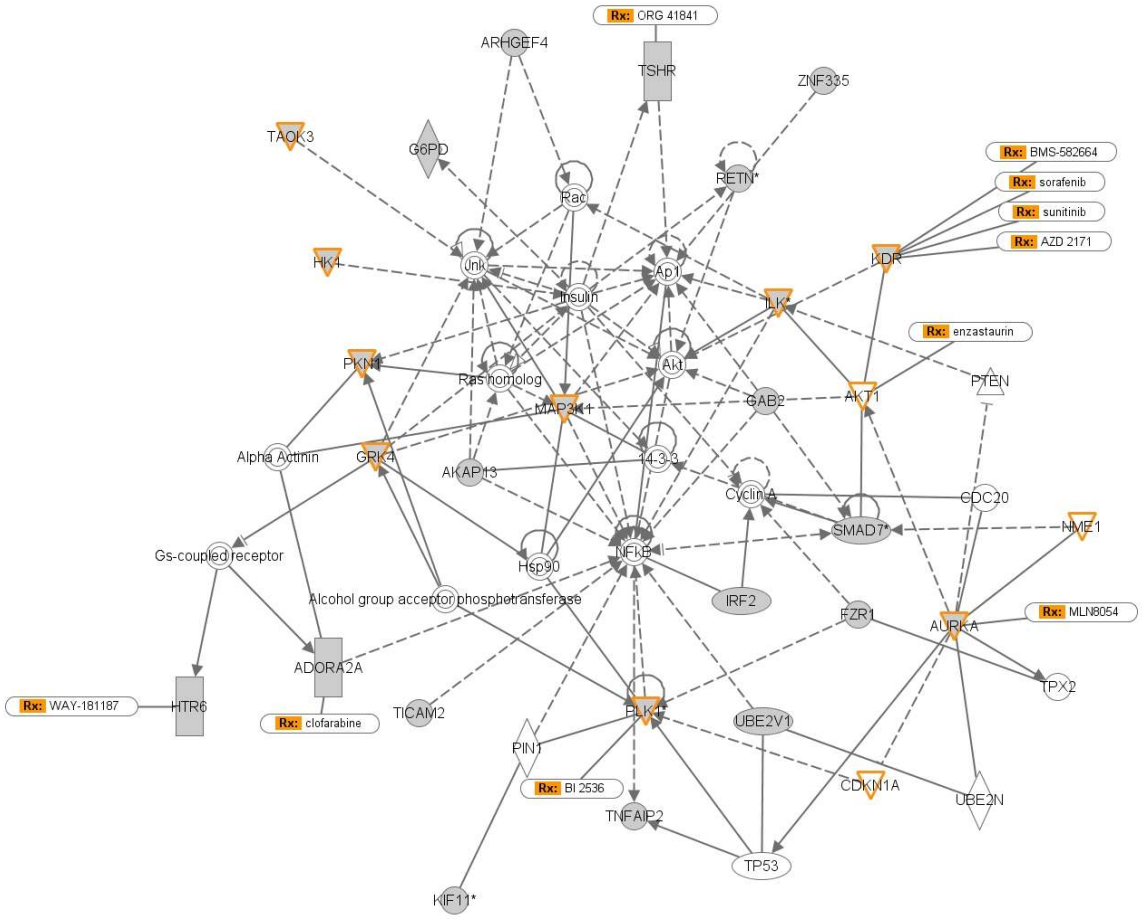


Figure 11. Network analysis of general survival genes. Top ranking network representing 25 genes (colored in grey) that are essential for cancer cell survival. Kinases are highlighted and targets for drugs approved or in development are shown. Genes in the list which caused concordant loss of viability with 2 siRNAs are marked with an *.

30 genes from the Calu6 specific survival list were involved in 8 large or small networks. The top scoring network, containing 11 genes, composed of direct and indirect interactions are illustrated in Figure 12. Cell to cell signaling, cellular movement and cancer are the top significant biological functions associated with this network. Interestingly several genes are linked to NFkB and PKCs which may explain their role in cell survival. Furthermore, several genes in this sub-network are expressed in lung cancer cell lines (highlighted) suggesting their relevance and applicability to lung cancer biology (e.g. NOS2A and ITGA3).

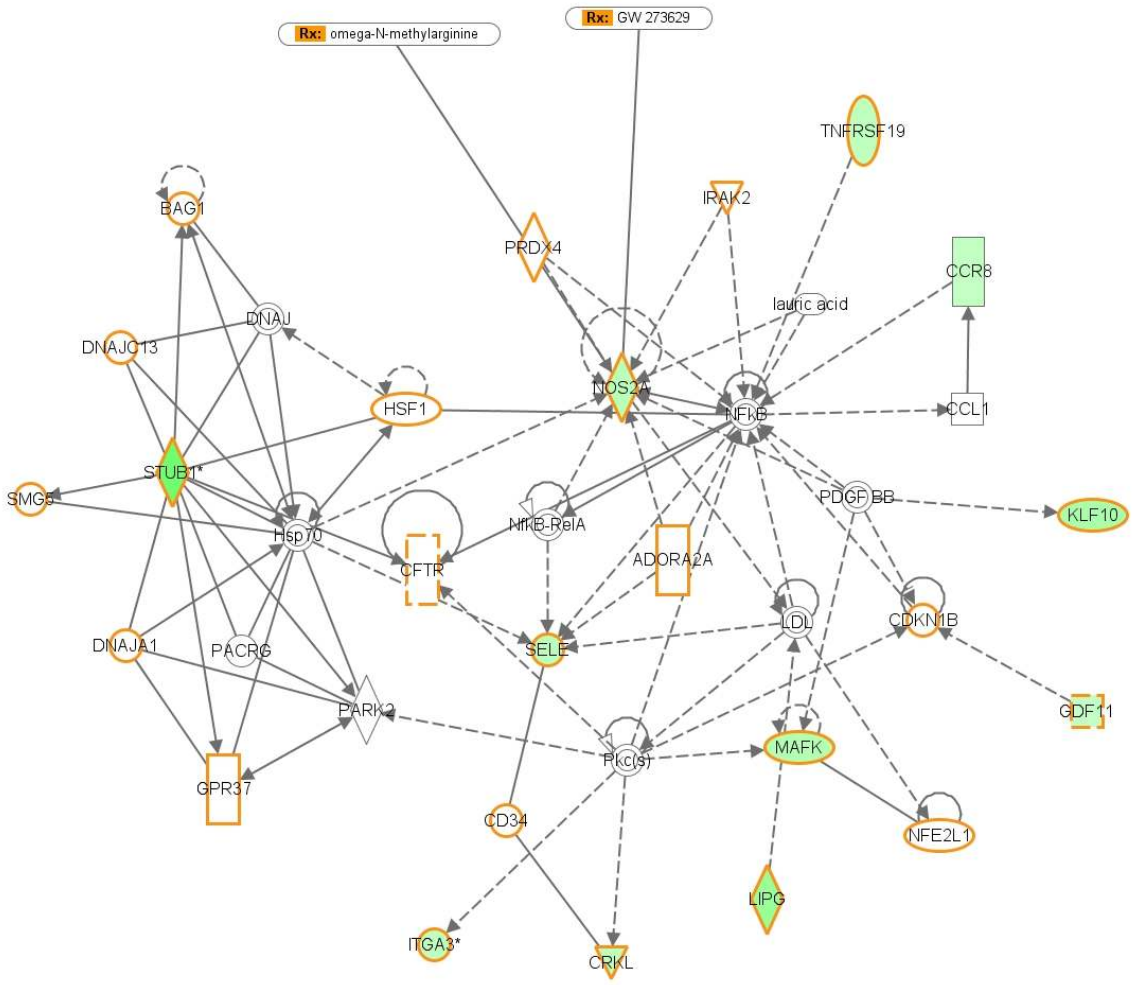


Figure 12. Network analysis of Calu6 specific survival genes. Top ranking network representing 11 genes (colored in green) that are essential for Calu6 cell survival. Genes that are expressed in other lung cancer cell lines are highlighted and targets for drugs approved or in development are shown. Targets in the list which caused concordant loss of viability upon knockdown with 2 siRNAs are marked with an *.

34 genes from the U87 specific survival list were involved in 5 large or small networks. The two overlapping top scoring networks, containing 25 genes in total, are shown in Figure 13. Cancer, cell death, cell growth and proliferation are the top significant biological functions associated with this combined network. Several drug targets can be noted as well such as FOLH1, IL23A and CSF2RA. It is also interesting to that multiple members of this merged network are also expressed in other CNS cancer cells.

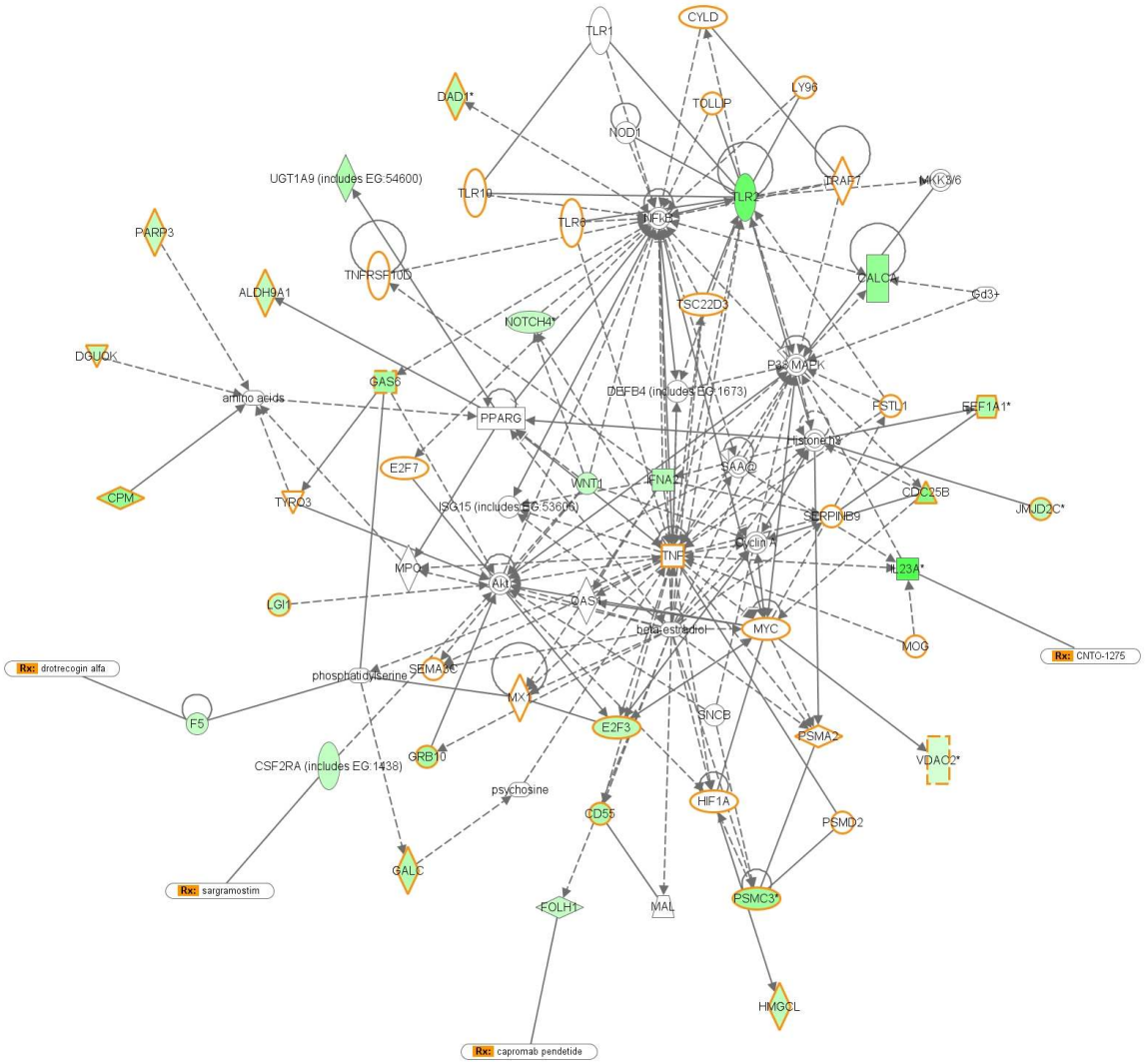


Figure 13. Network analysis of U87 specific survival genes. Top ranking networks representing 25 genes (colored in green) that are essential for U87 cell survival. Genes that are expressed in other CNS cancer cell lines are highlighted and targets for drugs approved or in development are shown. Targets in the list which caused concordant loss of viability upon knockdown with 2 siRNAs are marked with an *.

136 genes from the MCF7 specific survival list were involved in 16 large or small networks. The two overlapping top scoring networks, containing 42 genes in total, are shown in Figure 14. Cancer, cell death, cell growth and proliferation are the top significant biological functions associated with this combined network as well. 4 well known drug targets exist in this merged network – POLA, HDAC1, RXRA and PPARG. Genes that interact with caspase (e.g. MCL1, BAK) are likely to cause programmed cell

death. Many of these genes are likely to be relevant to breast cancer as they are expressed in other breast cancer cell lines as well.

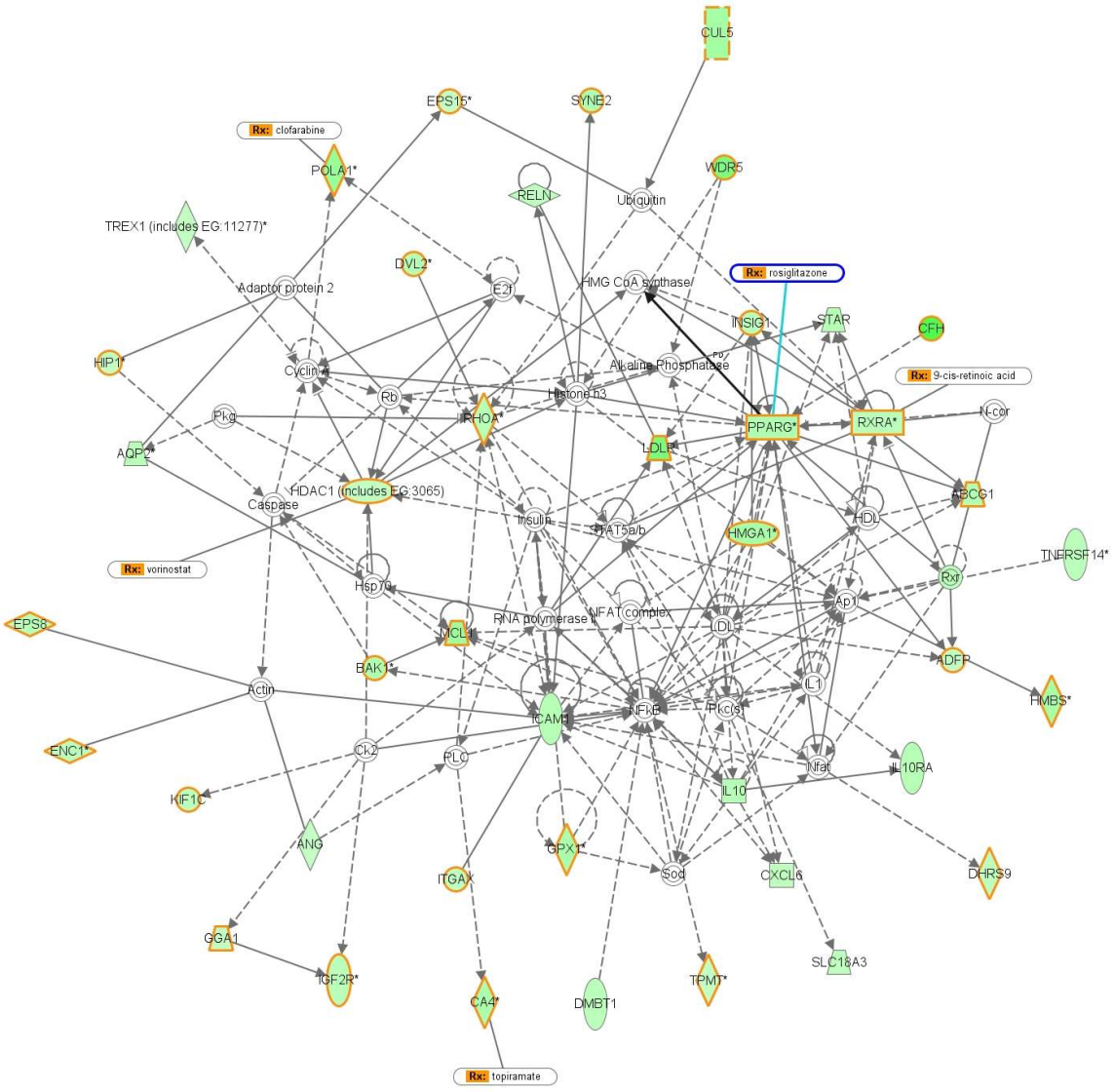


Figure 14. Network analysis of MCF7 specific survival genes. Top ranking networks representing 42 genes (colored in green) that are essential for MCF7 cell survival. Genes that are expressed in other breast cancer cell lines are highlighted and targets for drugs approved or in development are shown. Targets in the list which caused concordant loss of viability upon knockdown with 2 siRNAs are marked with an *.

46 genes from the HCT116 specific survival list were involved in 14 large or small networks. The two overlapping top scoring networks, containing 25 genes in total, are shown in Figure 15. Cancer, cell signaling, cell growth and proliferation are the top significant biological functions associated with this combined network as well. Networks

surrounding Akt and NFkB pathways along with angiogenic factors such as FLT1, FLT4 and p38 survival genes form the biological basis for target dependency in these cells. Furthermore, it is worth noting that several genes in this merged network are expressed in other colorectal lines.

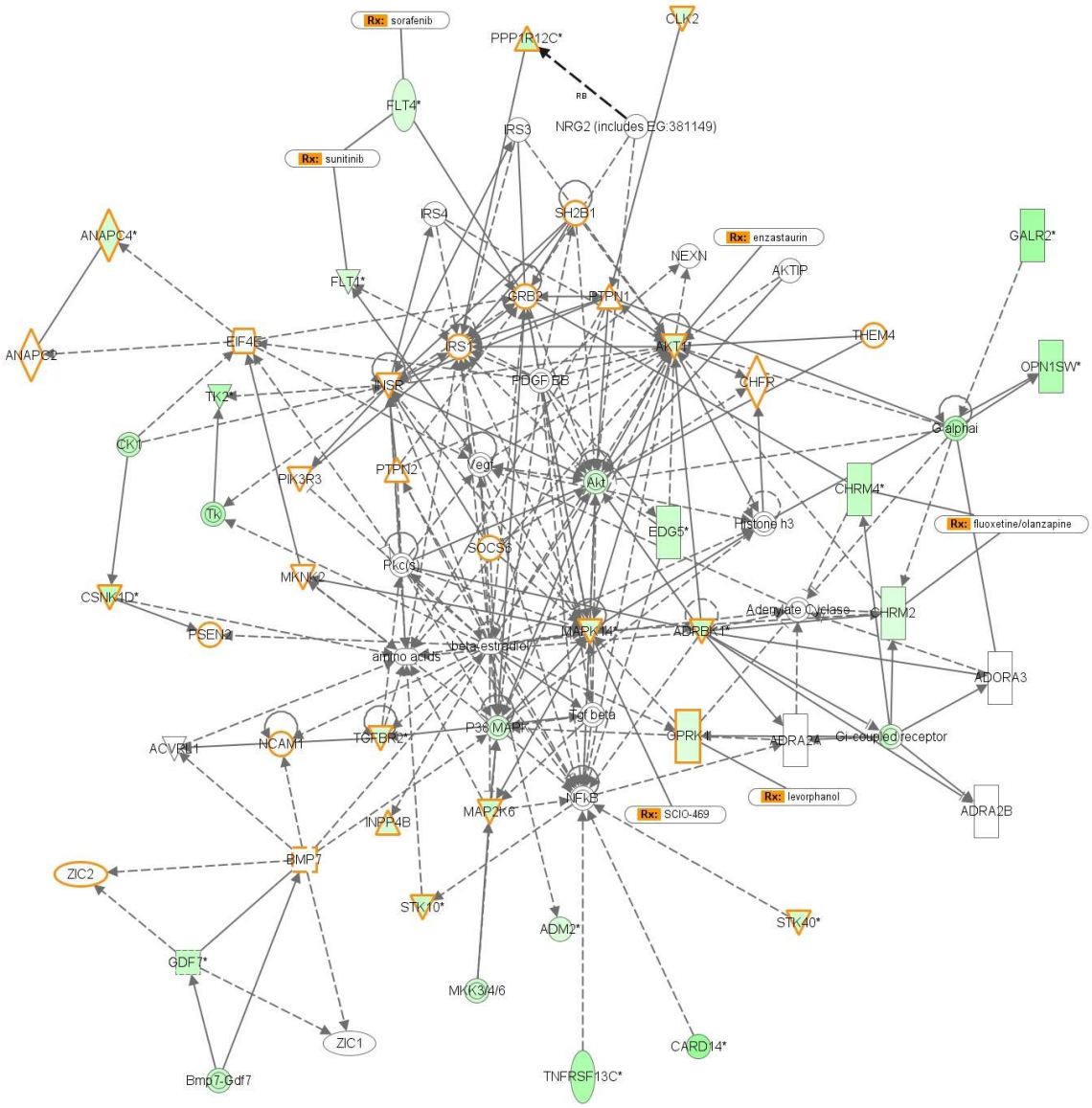


Figure 15. Network analysis of HCT116 specific survival genes. Top ranking networks representing 25 genes (colored in green) that are essential for HCT116 cell survival. Genes that are expressed in other colorectal cancer cell lines are highlighted and targets for drugs approved or in development are shown. Targets in the list which caused concordant loss of viability upon knockdown with 2 siRNAs are marked with a *.

3.1.8 Experimental confirmation

Based on all our analyses, there were several leading hypotheses that need to be validated experimentally. A major consideration is the off-target effects mediated by siRNAs which can complicate screen outputs. Target knockdown experiments by RT-PCR are essential. We confirmed knockdown for a select set of targets (see Figure 16(a)) using previously tested siRNA reagents. Furthermore, it is widely accepted that multiple siRNAs causing similar phenotypes demonstrate a target-specific effect. Therefore, we proceeded to take some of the top leads from above and tested them with multiple siRNAs for concordant phenotype in cell viability, cell death and live/dead high-content imaging assays.

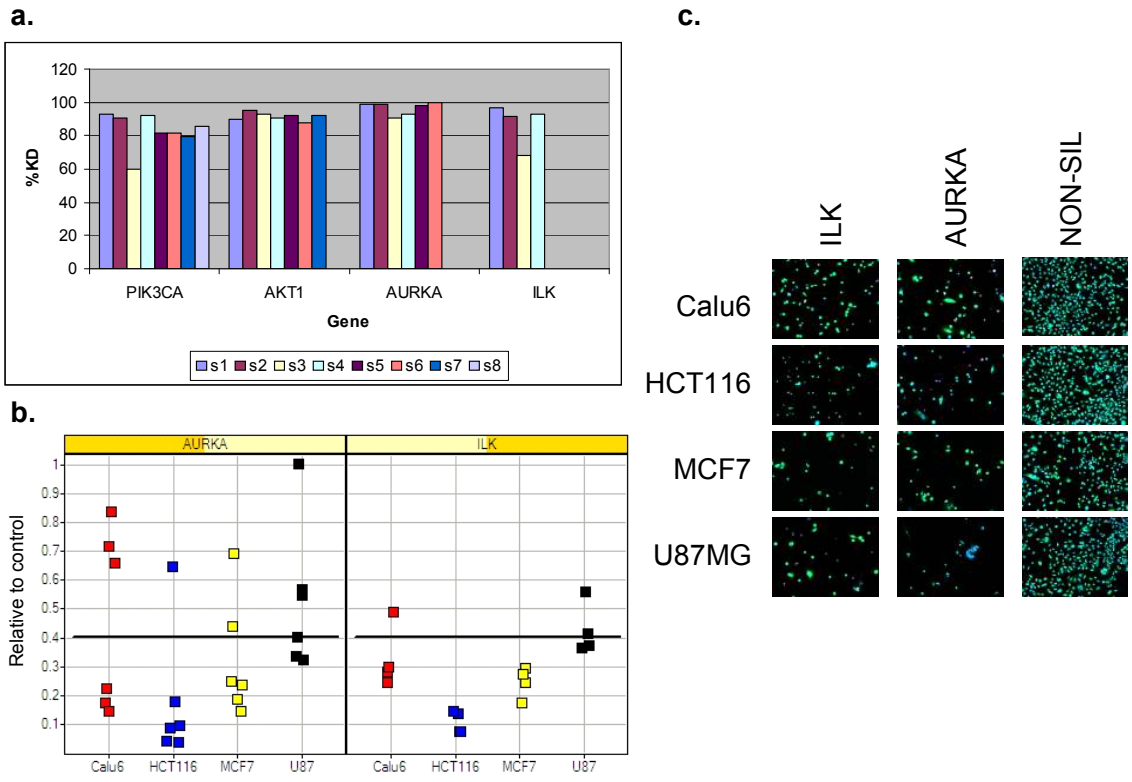


Figure 16. Validation of selected gene targets by knockdown and high content assays. (a) Real-time QPCR was performed to test 4 or more siRNAs (shown as s1, s2, s3, etc.) targeting the indicated genes for depletion of gene expression. %KD values on the Y-axis are GAPDH normalized values relative to those of a non-silencing negative control siRNA. (b) General toxicity caused by siRNA mediated knockdown of AURKA and ILK was tested by these siRNAs in the 4 indicated cell lines using a cell viability readout. GFP normalized values are shown along the Y-axis with a cutoff line drawn around 0.4 (see text for details). (c) High content images showing inhibition of ILK, AURKA or Non-sil (non-silencing negative control) by a representative siRNA in the indicated cell lines. The images represent phenotypic results for 3 or more siRNAs for a given gene in a given cell line. Reduction in cell count can be seen by a decrease in live cells (green) and/or increase in dead (red) cells.

We chose AURKA, KIF11, PLK1 and ILK genes that were essential for cell survival across several cell lines. AURKA, PLK1 and KIF11 are involved in G2/M arrest and are commonly upregulated in several cancers and their expression is associated with poorer prognosis. The latter was reflected in some of our own analysis using publicly available gene expression datasets (Table 4). Furthermore, several drug discovery efforts have produced small molecules, targeting these genes, with favorable toxicity profiles that constitute the next generation of cell cycle inhibitors. ILK, on the other hand, is an integrin-linked kinase, that is known to trigger apoptosis and is widely deregulated in multiple cancers [128]. As we expected, multiple siRNAs that knockdown AURKA and ILK demonstrate significant cell death in all 4 cell lines, as measured by cell viability readouts (Figure 16(b)) or % live cells parameter in live/dead high content assays (Figure 16(c)). Almost all ILK siRNAs reduce number of live cells by < 50% in all 4 cell lines. A point to note is that although a few AURKA siRNAs appear to be less effective, majority of the siRNAs induce lethality in all 4 cell lines. The outliers can be attributed to variable transfection, half-life or potency issues, as we have noted from earlier experiments. To follow up on PLK1 and KIF11, we tested 4 siRNA in 7 different cell lines (A549, BxPC3, SKOV3, HCT116, Calu6, U87, MCF7), representing various tumor types, by cell viability (Cell Titer Glo) and cell death (ToxiLight) assays. In Figure 17(a) and (b), it is evident that significant lethality is caused in all cell lines for both assays with multiple siRNAs. While U87 cells are relatively less susceptible to target inhibition by the cell viability assay, they certainly undergo extensive cell killing with the ToxiLight assay. Also, it must be pointed out the phenotypic effect in MCF7 is significant, albeit less pronounced. The ToxiLight assay measures cell death, a different readout from our initial screen, and has lower dynamic range. As a result, it is highly likely that not all screen actives are likely to produce similar results when tested with different assays. For instance, target effects that are anti-proliferative in nature may or may not be detected by an apoptosis assay readout. Conversely, targets that are scored positive by multiple siRNAs and more than one assay are more likely to be high confidence, robust hits. Taken together, our data supports our initial hypothesis that these genes are generally

essential for cancer cell survival, regardless of tumor type. Anti-tumor agents targeting these genes are likely to have broad spectrum activity.

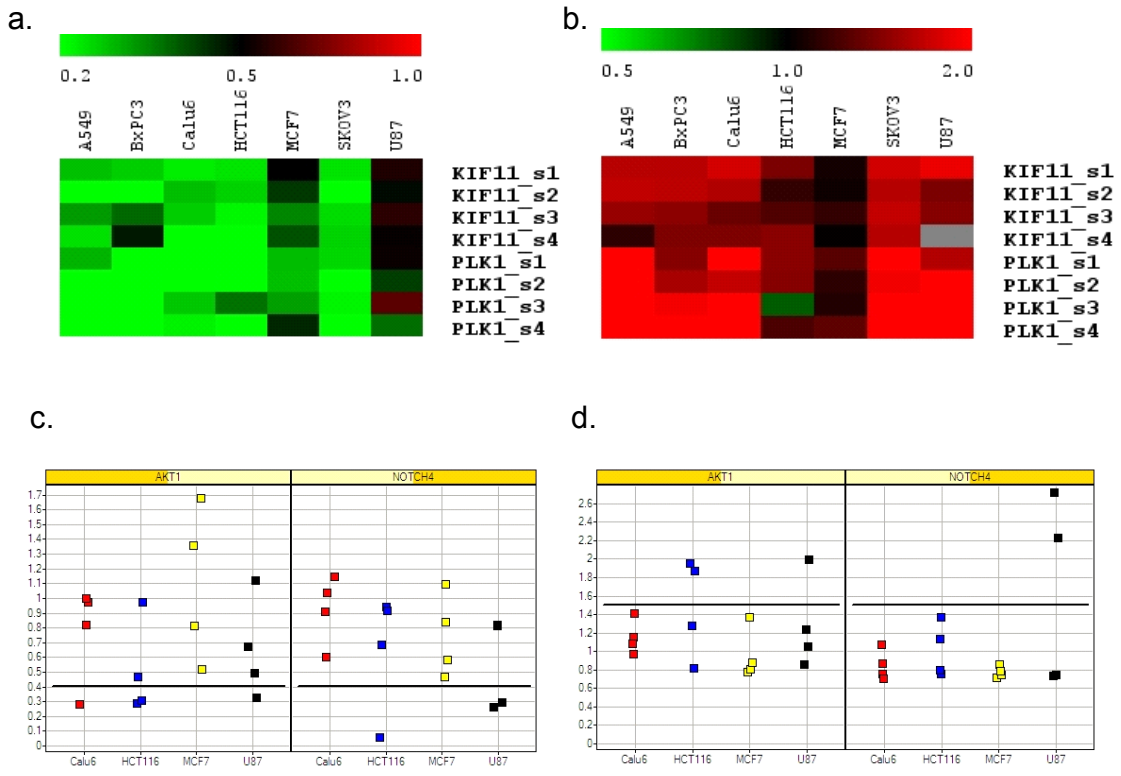


Figure 17. Validation of selected general and cell specific survival genes by proliferation and cell death assays. KIF11 and PLK1 were tested for their ability to cause toxicity in a broad range of 7 cell lines upon knockdown by 4 validated siRNAs by cell viability (a), measured by Cell Titer Glo, and cell death (b), measured by ToxiLight, assays. Green in (a) represents lower viability and red in (b) portrays higher cell death. AKT1 and NOTCH4 were similarly tested by cell viability (c) and cell death (d) assays for cell-specific lethal effects in HCT116 and U87, respectively. Normalized values are shown along the Y-axis with a cutoff line drawn around 0.4 (see text for details).

Based on the above integrative analyses and literature support, we hand-picked NOTCH4, AKT1 and MCL1 for follow up as cell-specific survival genes in U87, HCT116 and MCF7, respectively.

2/4 NOTCH4 siRNAs inhibited U87 proliferation by >70% with respective to negative control and gave a ~2-fold selectivity over the other 3 cell lines (Figure 17(c) and (d)). The same siRNAs enhanced cell death as measured by a ToxiLight assay by 2-fold which

is a significant readout for this assay. There was minimal to no effect seen with the other cell lines. The considerable loss of viability and cell death seen in HCT116 cells with 1 siRNA is possibly an off-target effect. We believe that the remaining 2 siRNAs did not produce a robust effect in U87 cells due to lower target knockdown (Data not shown) and follow up with additional siRNAs would support our findings. NOTCH4 is a less studied member of the NOTCH pathway and it has been mainly studied as a vascular endothelium specific signaling receptor. NOTCH4 is a target gene of F-box nuclear protein Fbxw7 which is tumor suppressor and is mutated in several cancers. Interestingly, Hagedorn and colleagues [129] have recently shown that grade IV gliomas (glioblastoma multiforme) are deficient in Fbxw7 and show an upregulation of AuroraA and NOTCH4. They also showed that knocking down Fbxw7 leads to mitotic defects in U87 cells. These findings along with our results make NOTCH4 - a highly druggable, oncogenic signaling receptor - a particularly intriguing target to pursue in brain tumors overexpressing Notch4 or lacking Fbxw7/4q13.3 locus.

AKT1 was identified in HTS as giving HCT116 cells a ~2X higher survival advantage over other cell lines. Integrative analysis has also shown that multiple members of the inositol phosphate metabolism pathway could be at play (Table 9). Confirmation experiments have shown that 3 out of 4 siRNAs cause >50% loss of viability and 2 out of 4 siRNAs cause 2X increase in cell death with a significant degree of selectivity over Calu6 and MCF7 (Figure 17(c) and (d)). AKT1 is a survival kinase with several downstream functions that affect cell proliferation, survival, metastasis and apoptosis. Several growth factors and cellular insults are known to stimulate this pathway by signaling through PI3K and Ras. While Calu6 cells have a KRAS mutation and MCF7 cells have a PIK3CA mutation, HCT116 cells harbor both mutations which in all likelihood leads to hyperactivated Akt pathway thereby causing a dependence on AKT1. PIK3CA depletion by siRNAs resulted in significant lethality for MCF7 cells, but, contrary to published findings, did not result in a non-viable phenotype for HCT116 (Data not shown). This context for target dependence and cellular vulnerability can be confirmed by following up in a broader panel of cell lines with genetic aberrations in the Ras/PI3K pathway. Another relevant point to note is that the cell viability assay data (Figure 17 (c)) show AKT1 siRNA knockdown could be similarly detrimental to U87.

This can be explained by the fact that U87 is PTEN-null which leads to increased levels of phosphorylated Akt resulting in target dependence. Therefore, AKT inhibitors could have a higher degree of therapeutic success in KRAS/PIK3CA double mutant or PTEN null genetic backgrounds.

MCL1 was selected as an MCF7-specific hit from HTS and represents an example of the power of integrative analysis. MCL1 siRNA caused significant lethality in MCF7 cells, offering a > 2-fold selectivity over other cell lines (Figure 18(a)). Array CGH data have shown that it is amplified in MCF7 cells (Figure 18 (b)), thereby providing a genetic driver for target dependence. As revealed by network analysis, it is also connected by direct and indirect interactions with several other MCF7-specific hits identified in the screen (Figure 18 (c)). Upon knocking down MCL1 with 4 siRNAs in cell viability assays, relatively higher loss of survival (>50-60%) was observed in MCF7 cells (Figure 18 (d)). To understand implications more broadly for breast cancer, we analyzed arrayCGH for primary breast tumors for a gain in copy number and discovered that the frequency was ~30%. Interestingly this frequency of gain was noted in other tumors (e.g. colon, ovarian) as well. To our knowledge, the presence of these genetic lesions is novel and unreported. Furthermore, we evaluated a published breast tumor transcription profiling dataset [130] and found that that MCL1 was generally over expressed relative to normal tissue (Figure 18(e)). IHC results on Protein Atlas showed 9/12 breast cancer tissues that had strong to moderate staining when probed with MCL1 antibody (Figure 18(f)). A statistically significant association of MCL1 overexpressing tumors with poor outcome and higher grade was recently reported as well [125] (Figure 18(g)). In our own Kaplan-Meier survival analysis as well, we found a significant association of RNA expression with survival in breast cancer (Table 5) as well as a glioma dataset (Data not shown). The latter is intriguing in the light of a comparable lethal effect mediated by MCL1 siRNAs in U87 cells (see Figure 18 (d)).

MCL1, is an anti-apoptotic BCL2 family protein, that has been predominantly studied in hematological and lymphoid malignancies. BCL2 family proteins, consisting of BCL2L, BCLXL, MCL1 share BCL2-homology regions and are required for cell survival by playing a key role in mitochondria-mediated intrinsic pathway of apoptosis. Perhaps due to assay sensitivity or mechanism, we were not able to see a noticeable effect with MCL1

siRNAs using the ToxiLight assay. It has been shown that BCL2 family proteins are overexpressed in multiple cancers and may contribute to chemoresistance. Lin and colleagues at Abbott laboratories reported an siRNA screen to identify mechanisms of resistance to their BCL2 inhibitor in development, ABT-737, and after an intriguing investigation, found that MCL1 knockdown drastically increased ABT-737 mediated lethality in a SCLC cell line [131]. These results could have a major impact on clinical use of standard chemotherapy or targeted agents, including ABT-737 for counteracting resistance. Moreover, the genetic lesions of MCL1 could also have major ramifications for patient selection when treating cancer patients, including breast and other types, with an MCL1 inhibitor as a single agent, as seen in our MCF7 example.

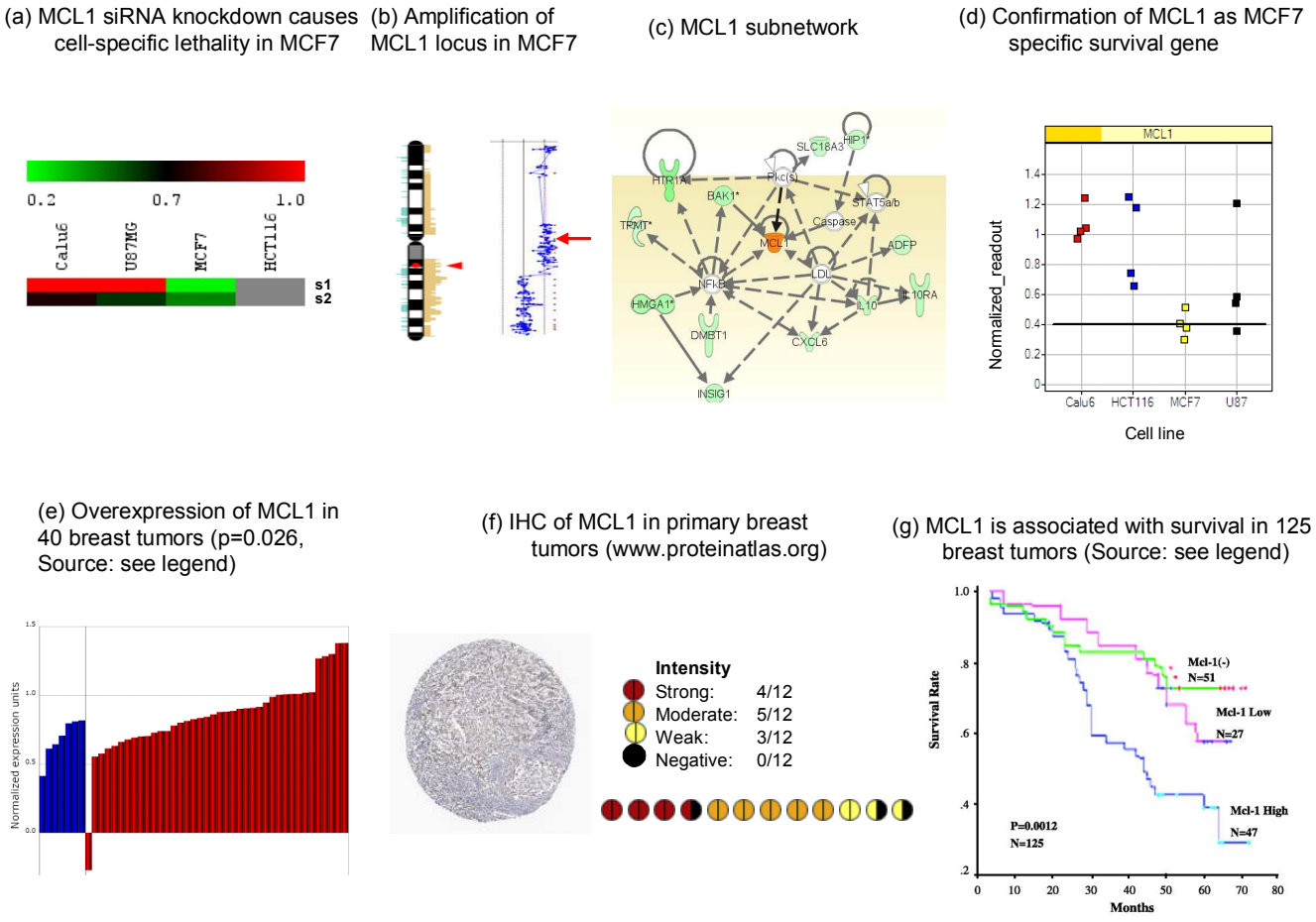


Figure 18. Integrative analysis of MCL1 as a rational target in breast cancer. MCL1 siRNAs had a profound lethal effect on MCF7 cells relative to others in our primary screen (a) which was reproduced upon confirmation with 4 siRNAs using the cell viability readout (d). (b) Genomic amplification of the MCL1 locus in MCF7 as well as primary breast tumors (not shown) was seen in internal aCGH data. (c) Functionally, MCL1 was also connected to various other MCF7 selective hits as shown by a subset of

network interactions seen in Figure 14. Mining public data revealed that MCL1 was overexpressed in a large subpopulation of breast tumors by microarray analysis (e) of 40 primary breast tumors versus 7 normal controls [130] as well as by immunohistochemistry (f) of 12 tumor samples (ProteinAtlas, http://www.proteinatlas.org). (g) This was associated with statistically significant poor survival in a separate population of 125 breast tumors [125]. Collectively, this data presents MCL1 as a promising target for therapeutic intervention in breast cancer.

3.1.9 Discussion

To our knowledge, this is the first ever high-throughput RNAi screen done in 4 different cell lines representing different tumor types followed by the most comprehensive integrative analysis of screen actives. It must be noted that several limitations exist with such large-scale screens. The library used in our screen was focused on the druggable genome, so there certainly are several other genes that are critical for tumor cell growth, proliferation and survival. Several hits may have been missed due to the choice of assay, cell line/context and time point. Targets with an exclusive pro-apoptotic or anti-angiogenic function are less likely to be picked up by a cell viability screen. Also, genes that are essential for specific cell types, genetic backgrounds or tumor subtypes will likely be missed. Other major concerns with the use of RNAi reagents are specificity, selectivity and potency. Complications arising from off-target effects of siRNAs can provide false positives. Furthermore, incomplete knockdown of targeted proteins can lead to false negatives, not to mention the possibility of hypomorphic phenotypes. Nonetheless, this approach provides rich information on key targets for appropriate follow up.

We have provided a rich list of prioritized targets and generated robust hypotheses by integrating the screening data with orthogonal datasets (expression, aCGH, mutation, pathways). Specifically, we have shown that (1) overlaying clinical outcome data can provide a repertoire of novel cancer targets of clinical relevance and (2) incorporation of array CGH/mutation/expression data and pathways can provide genetic and mechanistic rationale, respectively, for prioritization and downstream followup of screen actives.

While limited follow up has been reported, we believe we now have excellent starting points for a variety of future experiments: (1) Several targets have corresponding small molecule chemical inhibitors that are available and it would be intriguing to test them to

see if they would phenocopy the siRNA results. For instance, do the 'general' survival provide us with broad-spectrum agents that are agnostic to tumor type while the 'cell-specific' survival genes are likely to be context specific? (2) Furthermore, siRNAs or compounds that inhibit our list of general survival genes may work well in combination with commonly used cytotoxics. John Blenis' team [92] has shown similar results with hits from a kinome RNAi screen done in HeLa cells. PLK1 is one such example from our results that has been published as a potentiator of chemotherapeutics [132]. (3) While we have systematically tested single gene knockdowns in our screens and follow up experiments, it would be interesting to see the effects of additivity/synergy/antagonism in terms of lethality by knocking down 2 or more targets which might provide rational leads into effective multi-targeted agents or polypharmacology. (4) The lists of general or cell specific survival genes are excellent candidates for resequencing, aCGH, expression, epigenomic studies as they are causally implicated in cell survival. (5) To truly explore the 'contexts of vulnerability' we will need to follow up on hits with or without combination studies. If a gene exists in inactivated form (mutation, expression, deletion) in certain tumor types, does it enhance killing by an anti-tumor agent? If the gene in question is commonly amplified or overexpressed, does it confer resistance and would a combination therapy approach be more useful? We believe these questions can be answered by interrogating genetic background rather than cancer type. By testing a panel of cell lines of same tumor types but varied genetic backgrounds (e.g. KRAS mutant vs. non mutant colorectal lines), such synthetic lethal oncogene addictions can be unraveled. This is likely to shed light on pharmacogenomics development strategies for targeted agents for different cancers.

A central problem with genome-wide profiling approaches such as microarrays, proteomics, metabonomics is that they are associative in nature. Differentiating driver versus passenger effects is not trivial. Several reports from sequencing, aCGH and expression profiling have implicated hundreds of genes in various types and stages of cancer, but not all are likely to be relevant to disease or therapeutic intervention. High-throughput RNAi, on the other hand, offers a powerful approach to perform targeted knockdowns and hence provide causal information. By understanding target dependence, novel 'oncogene addictions' can be uncovered. This approach helps complement

conventional functional genomics approaches in prioritizing genomic alterations that are likely to be of significance amongst those are likely to be bystander effects of tumorigenesis.

A variation of this technique is where screens can be performed with or without a drug. Several recent reports (see Introduction) have shown the power of such chemosensitization or response modifier or synthetic lethality screens to identify modulators (enhancers or suppressors) of drug response and putative combination therapy targets.

In conclusion, we believe this is a powerful systems technique which when integrated informatively with other high dimensional datasets can yield valuable insights for target validation and predictive pharmacogenomic applications in cancer.

3.2 Integrative analysis of mutation profiling of human cancer

The Singapore Oncogenome Group (SOG) published an impressive genome-wide survey of mutations in the tyrosine kinase trascriptome of 254 cancer cell lines followed by an independent assessment of a select few frequent mutations in a variety of primary tumors [22]. SYK was the most frequently mutated PTK – 33.3 sporadic alterations per 1MB expressed coding sequence – all of which were somatic in nature (Figure 19). Spleen tyrosine kinase (SYK) regulates transcription, signaling, cell proliferation, neutrophil phagocytosis, leukocyte chemotaxis and lamellipodium biogenesis. It belongs to the SYK family of PTKs and contains 2 adjacent SH2 domains and a kinase domain. SYK biology has been extensively studied in hematopoietic cells – B and T lymphocytes, NK cells , mast cells, etc. where SYK protein plays a scaffolding role in downstream signaling of ITAM-containing immunoreceptors [133]. The biological role of SYK in solid tumors is just being uncovered. Recent reports of SYK being an unconventional tyrosine kinase tumor suppressor in breast (Figure 19), gastric and melanoma suggest a broader role for this gene in other tumor types. In the light of these observations, the biological significance of the mutations identified in by Ruhe et al. [22] is intriguing, yet currently unknown. We attempted to discern clues by mining a variety of datasets and fusing various analyses to generate hypotheses for guided experimentation.

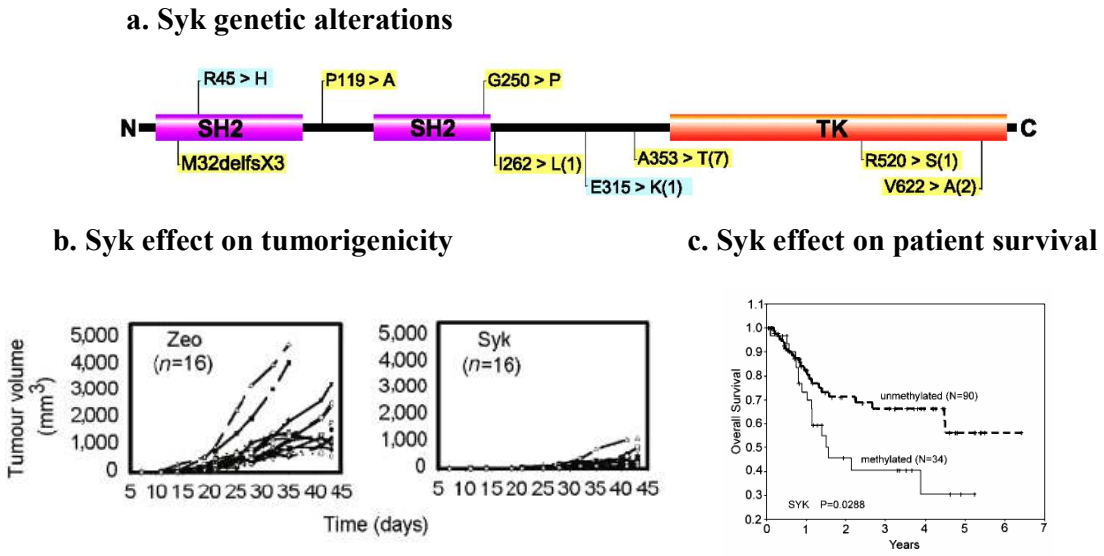


Figure 19. Molecular and biological aspects of SYK biology. (a) Domain organization of SYK protein displaying known polymorphisms and novel somatic alterations (identified in [22]). SH2: Src homology 2 domain; TK: tyrosine kinase domain. Adapted from [22]. (b) Inhibition of *in vivo* tumor growth of SYK-transfected MDA-MB-435BAG(SYK negative) breast cancer cell lines in athymic nude mice relative to control (Zeo). Adapted from [134]. (c) Poor overall survival of 124 hepatocellular carcinoma patients as a result of SYK methylation status as shown by Kaplan-Meier analysis. Adapted from [135]

3.2.1 Molecular consequences of SYK mutations

In the SOG dataset, all 15 SYK mutations identified were somatic in nature (Table 10). 3 nonsense frameshift and 12 missensense mutations were uncovered in SH2 or interdomain/tyrosine kinase domains, respectively. Out of these, Jurkat, MeWo and BM-1604 harbored homozygous mutations. Notably, M34fx3 (SH2_1 domain), A353T (Interdomain) and V622A (Kinase domain) were recurring mutations. Precedence of a genetic role in gastric carcinoma, melanoma, prostate cancer, lymphoma and myeloma provide strong support for the relevance of these mutations (reviewed in [133]). Given that much is known, intriguingly, no genetic alterations were found in breast cancer cells.

Table 10. SYK mutations in the SOG dataset. LS-174T and LS-180; DLD-1 and HCT-15 share the same genotype and are presumably derived from the same individual. Homozygous mutations are highlighted. * refers to previously reported alteration in Jurkat cells [136, 137]. Homo, homozygous; Het, heterozygous; TK, tyrosine kinase; SH2, Src homology domain

Cell line	Tissue	Mutation	Homo/Het	Domain
LS-174T	Colon	M34fsX3	Het	SH2_1
Jurkat	Hematopoietic and lymphoid	M34fsX3*	Homo	SH2_1
LS-180	Colon	M34fsX3	Het	SH2_1
KG-1	Hematopoietic and lymphoid	I262L	Het	Interdomain
MeWo	Skin	E315K	Homo	Interdomain
SK-N-SH	Brain	A353T	Het	Interdomain
MES-SA	Cervix and vulva	A353T	Het	Interdomain
DLD-1	Colon	A353T	Het	Interdomain
MM-Arn	Skin	A353T	Het	Interdomain
MKN-1	Stomach	A353T	Het	Interdomain
HCT-15	Colon	A353T	Het	Interdomain
MM-Leh	Skin	A353T	Het	Interdomain
A-498	Kidney	R520S	Het	TK
BM-1604	Prostate	V622A	Homo	TK
DU-145	Prostate	V622A	Het	TK

Cell line information was sought from literature to cover as many different SYK genetic backgrounds possible to discern molecular consequences of the mutations. To this end, a total of 77 cell lines (15 from Table 1 and 62 from Table 2) from different tumor types with varying levels of SYK were compiled : 32 (Null), 29 (wild type, WT), 1 (Copy number gain), 15 (mutations, in SOG). By overlapping with the GNF Affymetrix transcription profiling resource (http://www.symatlas.gnf.org), a panel of 13 cell lines with associated basal microarray data was assembled – 5 wildtype (WT); 4 mutant from SOG datasets (MUT); 4 deficient in expression (NULL). This panel is listed in Table 12 and was utilized for all subsequent analyses. A point to note is that Jurkat, a NULL cell line, has been reported to expresses RNA but not protein [136, 137]. RL, a WT cell line, likely overexpresses SYK, as confirmed by [138] and SK-MeL 28, labeled WT in our panel, has been shown to express very low levels of SYK (Table 11).

Table 11: 62 cell lines with varied SYK genetic backgrounds extracted from literature. RT- RT-PCR; WB-Western blot; MSP – Methylation specific PCR' Methylated – generally rescued by 5-aza cytidine. Normal can be considered as wild type (WT).

Cell line	Tissue	Defect_type	Defect_detail	PubMed ID	Hypothetical effect
BT20	Breast	Normal		11454707	Normal
BT474	Breast	Normal		10963601	Normal
BT483	Breast	Normal		11454707	Normal
BT549	Breast	No expression	No RNA or protein expression (RT, WB)	10963601	Null
DU4475	Breast	Normal		11454707	Normal
HCC1428	Breast	Normal		11454707	Normal
HCC1954	Breast	Normal		11454707	Normal
Hs578T	Breast	No expression	No RNA? or protein expression (RT, WB)	10963601	Null
Hs854T	Breast	No expression	No RNA (RT); methylated	11454707	Null
MCF7	Breast	Normal		10963601	Normal
MCF7-ADR	Breast	No expression	No protein expression (WB)	10963601	Null
MDA-MB-130	Breast	Normal		11454707	Normal
MDA-MB-134-VI	Breast	No expression	No RNA (RT); methylated	11454707	Null
MDA-MB-231	Breast	No expression	No RNA or protein expression (RT, WB)	10963601	Null
MDA-MB-361	Breast	Normal		11454707	Normal
MDA-MB-415	Breast	Normal		11454707	Normal
MDA-MB-435	Breast	No expression	No RNA or protein expression (RT, WB)	10963601	Null
MDA-MB-436	Breast	No expression	No RNA? or protein expression (RT, WB)	10963601	Null
MDA-MB-453	Breast	No expression	No RNA (RT); methylated	11454707	Null
MDA-MB-468	Breast	Normal		10963601	Normal
SKBr3	Breast	Normal		10963601	Normal
T47D	Breast	Normal		11454707	Normal
ZR75.1	Breast	Normal		11454707	Normal
MCF10A	Breast_non tumorigenic	Normal		10963601	Normal
0013"	Head&Neck	Normal		17699797	Normal
005B	Head&Neck	Normal		17699797	Normal
006/1	Head&Neck	No expression	No RNA or protein expression (RT, WB)	17699797	Null
011A	Head&Neck	No expression	No RNA or protein expression (RT, WB)	17699797	Null
CAL27	Head&Neck	Normal		17699797	Normal
D562	Head&Neck	Normal		17699797	Normal
HM6	Head&Neck	Normal		17699797	Normal
HN3	Head&Neck	No expression	No RNA or protein expression (RT, WB)	17699797	Null
HN4	Head&Neck	No expression	No RNA or protein expression (RT, WB)	17699797	Null
HN5	Head&Neck	Normal		17699797	Normal
Daudi	Hematopoietic and lymphoid	Normal		12717427	Normal
Granta-519	Hematopoietic and lymphoid	No expression		16409295	Normal
H9	Hematopoietic and lymphoid	No expression	No RNA or protein expression (MSP, WB)	12717427	Null
Hut78	Hematopoietic and lymphoid	No expression	No RNA or protein expression (MSP, WB).	12717427	Null
JeKo-1	Hematopoietic and lymphoid	Increased expression	arrayCGH, gene expression and FISH	16409295	Activated
K562	Hematopoietic and lymphoid	Normal		12717427	Normal
Karpas-422	Hematopoietic and lymphoid	Normal		16912221	Normal
Molt3	Hematopoietic and lymphoid	No expression	No RNA or protein expression (MSP, WB)	12717427	Null
Nalm6	Hematopoietic and lymphoid	Normal		12717427	Normal
NCEB-1	Hematopoietic and lymphoid	No expression		16409295	Normal
REC	Hematopoietic and lymphoid	No expression		16409295	Normal
RL	Hematopoietic and lymphoid	Normal		16912221	Normal
Hep3B	Liver	Normal		17121887	Normal
HepG2	Liver	No expression	No RNA (RT); methylated	17121887	Null
518A2	Skin; melanoma	No expression	No protein expression (WB)	15955106	Null
607B	Skin; melanoma	No expression	No protein expression (WB)	15955106	Null
A375	Skin; melanoma	No expression	No protein expression (WB)	15955106	Null
C8161	Skin; melanoma	No expression	No RNA or protein expression (RT, WB)	17145863	Null
Carney	Skin; melanoma	No expression	No RNA or protein expression (RT, WB)	17145863	Null
MelJuSO	Skin; melanoma	No expression	No RNA or protein expression (RT, WB)	17145863	Null
MH	Skin; melanoma	No expression	No protein expression (WB)	15955106	Null
Neo6/C8161	Skin; melanoma	No expression	No RNA or protein expression (RT, WB)	17145863	Null
Roth	Skin; melanoma	No expression	No RNA or protein expression (RT, WB)	17145863	Null
SKMel-28	Skin; melanoma	Very low expression	Low protein expression (WB)	15955106	Very low
UACC903	Skin; melanoma	No expression	No RNA or protein expression (RT, WB)	17145863	Null
WM1205Lu	Skin; melanoma	No expression	No RNA or protein expression (RT, WB)	17145863	Null
WM35	Skin; melanoma	Normal, reduced expression	No RNA, some protein expression (RT, WB)	17145863	Low/Normal
WM455	Skin; melanoma	No expression	No RNA or protein expression (RT, WB)	17145863	Null

Table 12: 13 cell line basal expression panel. By overlapping with the GNF Affymetrix transcription profiling resource (http://www.symatlas.gnf.org), a panel of 13 cell lines with associated basal microarray data was assembled. For details on the cell lines and associated SYK genetic defects (Label), refer to Table 10 and Table 11.

Cell line	Tissue	Label
A-498	Kidney	MUT
DLD-1	Colon	MUT
HCT-15	Colon	MUT
Jurkat	Hematopoietic and lymphoid	MUT
HepG2	Liver	NULL
Hs578T	Breast	NULL
MDA-MB-231	Breast	NULL
MDA-MB-435	Breast	NULL
K562	Hematopoietic and lymphoid	WT
MCF7	Breast	WT
RL	Hematopoietic and lymphoid	WT
ZR75.1	Breast	WT
SKMel-28	Skin; melanoma	WT

We used the microarray data to examine if molecular profiles cluster cell lines based on their SYK genetic defects. The signal values across the 13 cell lines for SYK expression were far too low (data not shown), but in general, NULL cell lines showed far lower values than WT or MUT. Unsupervised hierarchical clustering of basal genome-wide transcript profiles was performed to see if the microarray profiles would cluster cell lines based on their SYK background. To specifically attribute relevance of expression profiles to SYK biology, a custom gene set of direct and indirect SYK interactions, called SYK_interactions_network (see Materials and Methods) comprising 109 genes was custom assembled, which mapped to 201 probesets with an appreciable degree of variation (coefficient of variation, CV>0.4) across all cell lines. A general pattern of all MUT samples co-occurring with NULLs while WT samples clustering separately was observed (Figure 20). This suggests that the mutations A498_R420S, HCT15_A353T, DLD1_A353T maybe inactivating mutations and may show a SYK-NULL like signature. Jurkat was a notable exception and clustered with RL7, possibly a SYK-overexpressing cell line. This may be due to 2 reasons: (1) the expression data was generated from a Jurkat clone without the biallelic truncating mutation (2) compensatory changes that are known to occur in T-cell leukemias [139]. Although SKMel-28 is labeled as WT, this cell

line is known to express very low levels of protein [140] which may explain why it groups with the NULL and MUT cell lines. It was interesting, however, to note that HepG2 groups with WT cell lines despite literature evidence of no RNA expression (Table 11). Although 13 cell lines may be too small a sample set to give a robust clustering result, it is noteworthy that a clear separation exists, as expected, between WT and NULL samples with respect to their underlying SYK network gene expression.

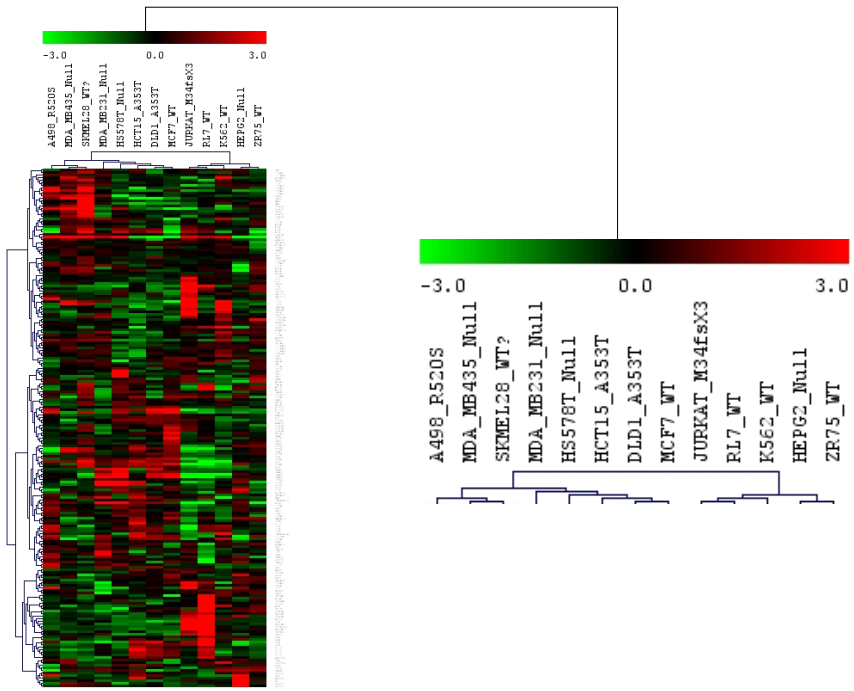


Figure 20. Unsupervised clustering of SYK network transcript profiles across 13 cell line panel with varied SYK genetic backgrounds. Log_2 transformed signal values were analyzed by unsupervised hierarchical clustering using a Pearson correlation coefficient metric and average linkage. Red and green represent relatively overexpressed and underexpressed genes, respectively.

3.2.2 Pathway analysis of transcriptional profiling data from varied SYK genetic backgrounds

To identify patterns of pathways that are upregulated or downregulated genomically as a consequence of SYK status, the 13 cell line expression profiles were interrogated for differential enrichment of pathways relevant to SYK biology. Therefore, 4 genesets were compiled: EGFR/MAPK/ERK pathway – 42 genes, NFkB pathway – 45 genes, PI3K/Akt pathway – 99 genes, SYK direct and indirect interactions (SYK_interactions_network) – 125 genes. Gene Set Enrichment Analysis (GSEA) is a powerful method that calculates enrichment of user defined groups of genes in a 2 class comparison by examining whole array expression data and taking into account subtle but consistent profiles [117]. GSEA was used for MUT vs. WT, NULL vs. WT, MUT+NULL (REST) vs. WT comparisons to examine enrichment of the above mentioned pathways. EGF/ERK/MAPK pathway was seen to be consistently and significantly (FDR <0.15) upregulated in MUT and NULL lines relative to WT (see Table). This suggests that SYK deficiency upregulates EGFR signaling. This has been previously reported in MCF10A cells by Ruschel and Ullrich [141]. Due to limitations in number of samples (very small n), statistical significance is not seen for the other pathways. However, if trends are noted, PI3K and NFkB pathways are also enriched in NULL and MUT cell lines and SYK_interactions_network is enriched in WT cell lines, as expected. SYK mediated repression of PI3K and NFkB pathways has been previously reported [133].

Table 13: GSEA results for SYK relevant gene sets. EGFR/ERK/MAPK, PI3K, NFKB and SYK network interactions were analyzed for enrichment by GSEA in the indicated comparisons of WT, NULL and MUT cell line expression profiles. REST refers to a combination of NULL and MUT cell lines. Size refers to the number of probesets that the genes in each gene set mapped to. NES, normalized enrichment score; FDR q-val, false discovery rate after correcting for multiple testing. Interesting results are highlighted and described in the text above.

Comparison	Gene set	Size	NES	FDR q-val
REST vs WT	EGF_ERK_MAPK	41	1.41604	0.127193
NULL vs WT	EGF_ERK_MAPK	41	1.399081	0.125
MUT vs WT	EGF_ERK_MAPK	41	1.315885	0.075758
NULL vs WT	PI3K	94	0.934409	0.851563
NULL vs WT	NFKB_IPA	42	0.928829	0.586458
MUT vs WT	NFKB_IPA	42	0.897361	0.686869
REST vs WT	NFKB_IPA	42	0.770963	0.869518
WT vs MUT	SYK_INTERACTIONS	116	-0.899716	0.633333
WT vs REST	SYK_INTERACTIONS	116	-0.944864	0.541246
WT vs NULL	SYK_INTERACTIONS	116	-0.956538	0.519558

Having observed a greater prevalence of EGFR/MAPK activation in the transcript profiles of cell lines with NULL or MUT backgrounds, we investigated an inverse relationship, if any, between EGFR and SYK in a panel of 11 breast cancer cell lines that were reported in [142]. SYK status in these cell lines was obtained from Table 11. As seen in Table 14, 5/8 lines with low/absent EGFR expression harbor normal levels of SYK. Interestingly, 3 out these 5 are sensitive to erlotinib as well. It would be intriguing to test endogenous SYK transcript/protein in MDA-MB-453. In this panel, there wasn't enough data to confirm an inverse trend, but 2 cases (MDA-MB-231 and MDA-MB-435) showed that absence of SYK was not inversely related to EGFR. These observations support the hypothesis that SYK expression may regulate EGFR expression/signaling. Not surprisingly, erlotinib IC50 values are agnostic to EGFR levels since it is widely known that EGFR expression is not associated with response to EGFR inhibitors. It is, therefore, tempting to ask if SYK status singly or in combination with EGFR would serve as a predictive biomarker to stratify patients for anti-EGFR therapy.

Table 14: Relation between SYK, EGFR and erlotinib sensitivity in breast cancer cells. ? refers to unknown status of SYK. Erlotinib sensitivity (IC50) and EGFR protein expression (Western blots) data were extracted from [142].

Cell line	Syk	EGFR	Erlotinib_IC50
MDA-MB-453	?	-	>20uM
A431	?	++	1.53uM
MCF7	Normal	-	>20uM
MDA-MB-361	Normal	-	>20uM
SKBr3	Normal	-	3.98uM
BT474	Normal	-	5.01uM
T47D	Normal	-	9.8uM
MDA-MB-468	Normal	++	>20uM
BT20	Normal	++	>20uM
MDA-MB-231	Null	-	>20uM
MDA-MB-435	Null	-	>20uM

To explore other pathways and biological processes that are differentially modulated in NULL & MUT vs. WT cell lines, we filtered 1000 genes with a signal-to-noise ratio (SNR) > 0.5 or <-0.5 that were differentially expressed in MUT & NULL groups relative to WT cell line microarray profiles. Pathway analysis was done in Ingenuity Pathway Analysis. From Figure 21, it is evident that several cancer relevant growth factor signaling and survival pathways (e.g. PDGFR, Ephrin receptor signaling) are constitutively upregulated in MUT and NULL groups relative to WT.

Taken together, this data supports the previous observation that the mutations maybe inactivating in nature and functionally similar to NULLs in terms of pathways and genomic profiles.

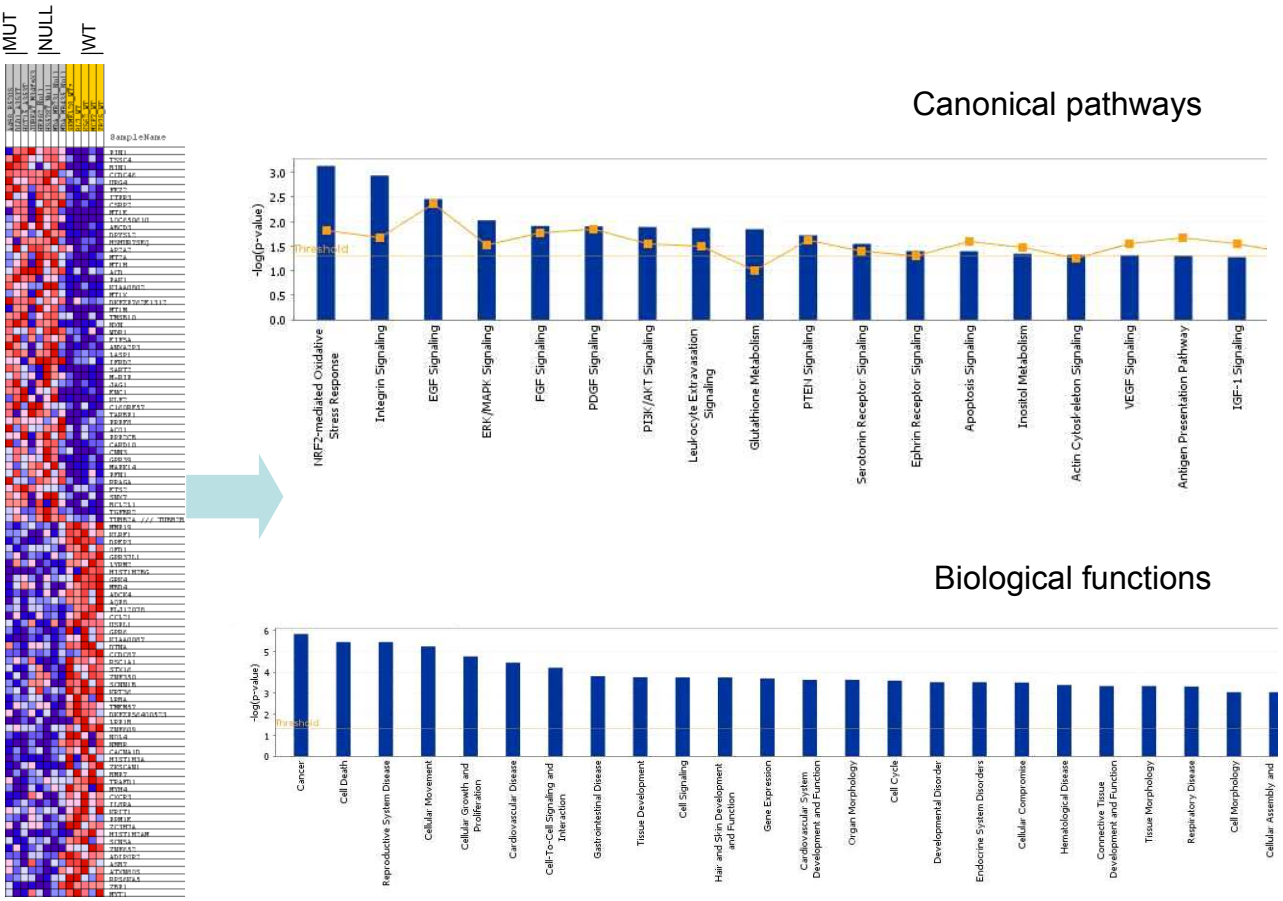


Figure 21. Pathway analysis of differentially expressed genes in SYK MUT and NULL cell lines relative to WT. Gene expression profiles were filtered for SNR > 0.5 or <-0.5 in NULL and MUT cases relative to WT (left) for pathway analysis in Ingenuity Pathway Analysis. Canonical pathways and biological processes (right) that are above the indicated threshold values are shown. Enrichment of growth factor signaling (e.g. PDGF and Ephrin receptor signaling) and survival pathways in the NULL and MUT cell lines can be seen.

3.2.3 Relationship between differential SYK expression and clinical outcome in various tumor types

A few publicly available gene expression datasets with corresponding clinical covariates (e.g. survival) were queried (see Materials and Methods) to check if SYK transcript levels were associated with clinical outcome in various primary tumors by univariate Kaplan-Meier (KM) survival analysis. Figure 22 demonstrates that significant stratification ($p<0.06$) of patients is achieved in non-small cell lung, breast, ovarian, lymphoma and glioblastoma datasets. In the NSCLC dataset, KM survival curves for both

adenocarcinoma (AD) and squamous (SQ) subtypes demonstrate that lower expression of SYK corresponds to worse outcome, thereby suggesting a poor prognosis. Interestingly, we noted a relatively lower expression in lung cancer samples relative to normal controls (Figure 23). This supports a currently unknown tumor suppressor role for SYK in lung tumors. Examining gene expression profiles of NSCLC cell lines and follow up experiments may provide additional insights. Also, survival analysis of primary breast cancer gene expression data supports SYK playing a tumor suppressor role. This is in agreement with the well studied functional role of SYK in breast cancer where reduced expression is associated with invasion and overexpression is associated with better outcome [133, 134]. Similar trends are also seen in an ovarian cancer dataset, where an association between high expression (above median) and poor clinical outcome is observed. A tumor suppressive role for SYK in ovarian cancer is unknown.

In contrast, similar analysis of DLBCL gene chip data is in agreement with previously reported findings of SYK playing a proliferative role in B-cell malignancies. Similarly in glioblastoma multiforme (GBM) , a form of brain cancer, while the trend is conserved in general, relatively significant ($p=0.055$) separation by quantiles of expression values is seen where the first and third quantiles show maximum separation demonstrating that higher expression is associated with poor prognosis. This observation suggests that SYK may have a potential role in cell proliferation/survival/migration/invasiveness in GBMs. An oncogenic role for SYK in brain tumors is unknown. Interestingly, a somatic mutation was identified in SYK in SK-N-SH (neuroblastoma) cells by the SOG group. Taken together, based on the microarray based gene expression levels and association with clinical outcome in terms of survival, it appears that SYK may play context-dependent oncogene or tumor suppressor roles. It appears to be a tumor suppressor in lung, breast and ovarian cancers while playing a more oncogenic role in lymphomas and glioblastomas.

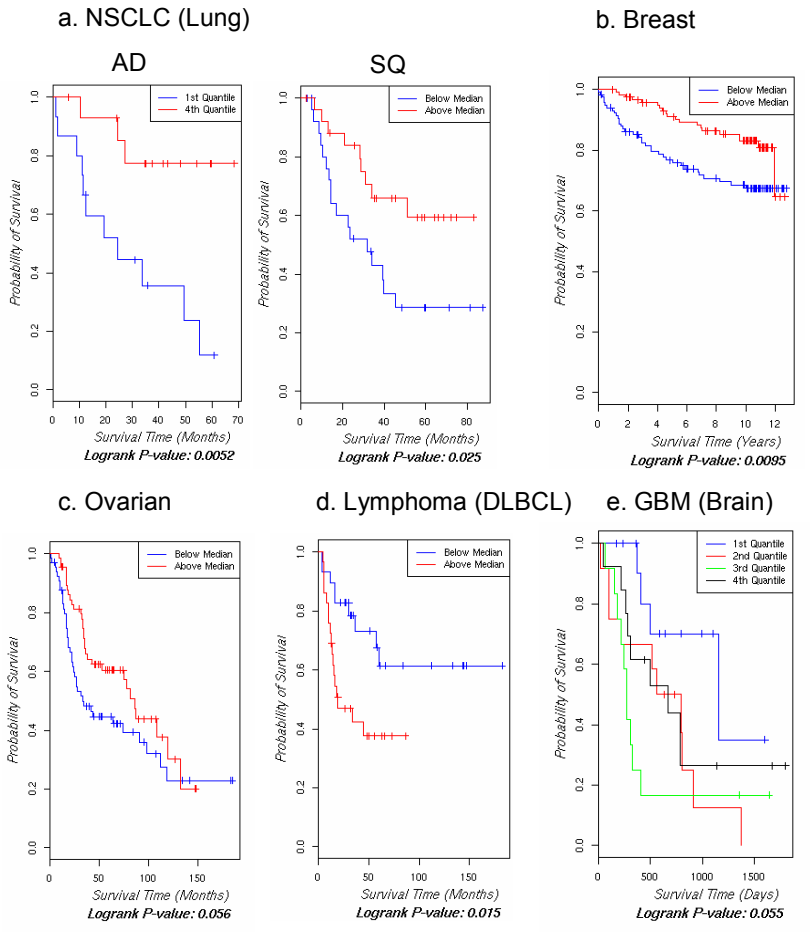


Figure 22. Kaplan Meier survival plots of SYK transcript profiling in multiple tumors. Prognostic patient stratifications where median or quantile expression pattern is associated with distinct survival outcome. Several publicly available microarray gene expression datasets were employed that corresponded to (a) NSCLC, non-small cell lung cancer (AD, adenocarcinoma and SQ, squamous) [45] (b) breast cancer [54] (c) ovarian tumors [45] (d) diffuse large B-cell lymphoma (DLBCL) [41] and (e) glioblastoma multiforme (GBM) brain tumors[76]. Statistical significance was determined by logrank p-values. For details on the method and datasets employed, see Materials and Methods.

To understand its function role, we examined profiles of normal versus cancer in other publicly available epithelial tumor datasets with large sample numbers (n>=20) in Oncomine (http://www.oncomine.org). It is noteworthy that lower expression of SYK is seen in primary prostate carcinoma relative to normal cases (Figure 23). Several independent datasets were found where this trend was conserved (data not shown). This is consistent with previous findings [143, 144] where the promoter is likely methylated. Methylation has also been observed in multiple myeloma [145] which is consistent with a lower expression in myeloma patients relative to normal plasma controls (Figure 23).

This epigenetic inactivation of SYK suggests that it has a significant role as a tumor suppressor in disease progression as seen in breast cancer where progressive loss of SYK was observed in normal mammary to ductal carcinoma in situ to invasive carcinoma[134]. In a similar way, lower expression in renal cancer may be relevant in the light of the somatic mutation identified in A498 (see Table 10). A potential tumor suppressor role in renal cancer is unknown to date.

Therefore, it appears that SYK has context-dependent roles of an oncogene or a tumor suppressor. This is consistent with the current knowledge of SYK function in various tumors.

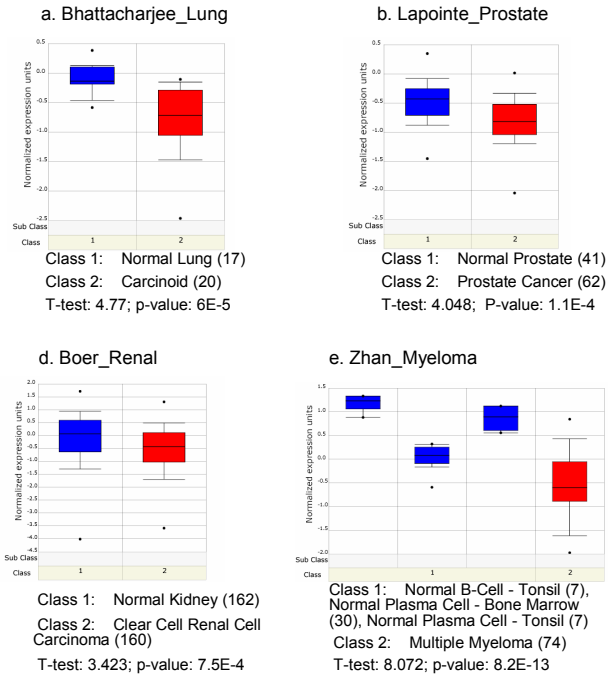


Figure 23. Differential SYK expression in Normal vs. Cancer tissues. Box plots showing the range of normalized expression values that show a statistically significant pattern in normal vs. cancer microarray studies of (a) lung cancer [44]; (b) prostate cancer [146]; (c) renal cancer[147]; (d) multiple myeloma [148]. Class1, normal tissue counter parts; Class2, tumor samples. The number of samples in each class is shown in brackets along with T-test scores and p-values. These analyses were carried out in Oncomine (http://www.oncomine.org)

3.2.4 Insights into compound sensitivity

From a therapeutic standpoint, understanding SYK levels and how they relate to the type or stage of disease is critical. Based on our analysis and published evidence, this would

translate to either rationally inhibiting or activating SYK. For instance, it may not be effective to inhibit SYK activity in breast, lung or prostate cancers. In these instances, pending experimental follow up, it may be altogether desirable to effect cytoprotection by enhancing SYK. In fact, after mining the Connectivity Map database in Oncomine (http://www.oncomine.org), we found that SYK expression is induced in MCF7, SKMEL5 and HL60 cells (top 5% of upregulated genes) upon treatment with several compounds, including different concentrations of HDAC inhibitors (TSA, valproic acid, vorinostat), geldanamycin, LY294002 (PI3K inhibitor). On the other hand, SYK upregulation in DLBCL and GBM (our analysis) may warrant a therapeutic intervention strategy. To explore what potent chemical entities might already available and understand their differential efficacy, if any, we scanned the literature and analyzed publicly available datasets.

Analysis of growth inhibition data of piceatannol, which is purportedly a SYK inhibitor, did not show differential activity in the NCI60 panel of cell lines (obtained from http://dtp.nci.nih.gov/). This could potentially be due to non-specific activity of the compound and/or assay artifacts associated with the way kill curves are generated for the NCI-60 panel. Other groups have used piceatannol to inhibit SYK protein [138, 149, 150]. Interestingly, EKB-569 (an irreversible pan-EGFR inhibitor – ERBB1,2,4) showed activity against SYK ~1.2uM in [151] and BMS-BMS-354825 (dual Src/Abl inhibitor) showed activity against SYK ~3uM in [152] in the panel of kinases examined. EKB-569 was a potent inhibitor of proliferation in NHEK, A431, and MDA-468 cells (IC(50) = 61, 125, and 260 nM, respectively) but not MCF-7 cells (IC(50) = 3600 nM).

3.2.5 Discussion

We attempted to provide an *in silico* explanation of the molecular consequences of SYK mutations identified in a systems level survey for PTK mutations. We have generated several hypotheses by integrative analyses of SYK biology at large that warrant experimental validation.

SYK mutant cell lines (HCT15, DLD1 and A498) showed a downregulated transcriptional profile with respect to a SYK signature, similar to NULL cell lines, implying that the mutations may play an inactivating role. Based on GSEA, we found that

SYK-deficient lines (MUT and NULL) were associated with enhanced EGFR signaling, among other proliferation and survival pathways, relative to WT. Interestingly, SYK expression also appears to inversely correlate with EGFR status in a panel of well characterized breast cancer cell lines. This is in agreement with previous reports [141] and begs to questions if SYK status (mutation or expression) could be a predictor of EGFR pathway activity. It remains to be seen if SYK is upregulated in EGFR deficient lines or upon treatment with anti-EGFR compounds such as gefinitib or erlotinib. Interestingly, SYK is generally underexpressed (data not shown) in several lung cancer cell lines where overexpression or mutant EGFR is commonly seen. These observations upon testing are likely to have considerable ramifications in tailoring EGFR inhibitor therapy. Since several mutations were found in a wide variety of tumors, we analyzed clinical relevance of SYK by KM survival plots in various publicly available primary tumor microaray datasets. We found that depletion of SYK was associated with good prognosis in ovarian, lung, glioma. These are novel observations that suggest that SYK may play a tumor suppressor role in several cancers. Inactivation by methylation and loss of expression are commonly reported, but the mutations uncovered by the SOG group [22] suggest another molecular mechanism to turn off the potentially tumor suppressing activity by SYK. That said, SYK biology is far more complex and it has been reported play an oncogenic role in lymphomas [145] and head and neck squamous carcinomas [153]. In our own analysis, higher levels of SYK correspond to poorer outcome in GBM. Interestingly, we found several compounds that appeared to trigger SYK expression such as HDAC inhibitors (TSA, valproic acid, vorinostat), geldanamycin, LY294002 (PI3K inhibitor) which may have tumor stimulatory or inhibitory effects. Taken together, these results support a complex, context dependent role for SYK in solid tumors.

There are a few caveats associated with our analysis. Our observations have been derived from heterogeneous sources of data – different cell lines, conditions, doses, times, platforms, etc. We have mostly used RNA expression profiling datasets where expression may not always be a good surrogate for pathway activation or inactivation. Also, not all cell lines had readily accessible gene chip data which limited us to analyzing genomic transcript profiles for 4/15 cell lines. Given the dual roles of SYK, it is likely that some of the mutations may play an activating role (e.g. E15K). A point to note is that most of the

mutations identified were heterozygous. Nonetheless, unlike classical tumor suppressor genes like TP53, they may still play a role by secondary allele inactivation or haploinsufficiency or allelic insufficiency as seen is 5q deletion syndrome [154] or PAX5 mutations in acute lymphoblastic leukemia [155]. Additionally, basal expression is likely to be different from stimulated expression (by growth factors, for example) which may impact our observations of other pathways associated with the genetic defects. Also, integrative analyses such as GSEA or pathway mapping have been done on a limited number of samples which can limit the power of the results.

Nonetheless, these integrative analyses offer powerful hypotheses to begin rational experimentation in the lab and clinic. Increasing availability of high throughput functional genomics data (e.g. arrayCGH, methylation, proteomics) in the future, profiling larger test panels of cell lines, would improve "bottoms up" integration and increase the power of such hypotheses by minimizing false positives. Furthermore, once proven, this approach can be extended to analyze the functional consequences other novel and unexplained mutations to enhance our systems understanding of the 'oncogenome'.

3.3 Mining compound-treated cancer gene expression data for combination opportunities

Transcriptomics data provide a molecular finger print of cellular state.
When this is applied to drug-treated cancer cells, where arguably tell tale signs of resistance arise early and cells expressing such survival factors would dominate over time, combination or adjuvant therapy opportunities can be extracted. For instance, 5-fluorouracil (5-FU) is the standard-of-care chemotherapeutic agent used for treatment of colorectal carcinoma. By analyzing colon cancer cells treated with 5-FU, we filtered compound induced gene changes for druggable targets and pathways which were associated with poor outcome. Furthermore, these molecular readouts were used to suggest suitable compounds for follow up. Figure 24 provides an overview of this analysis workflow.

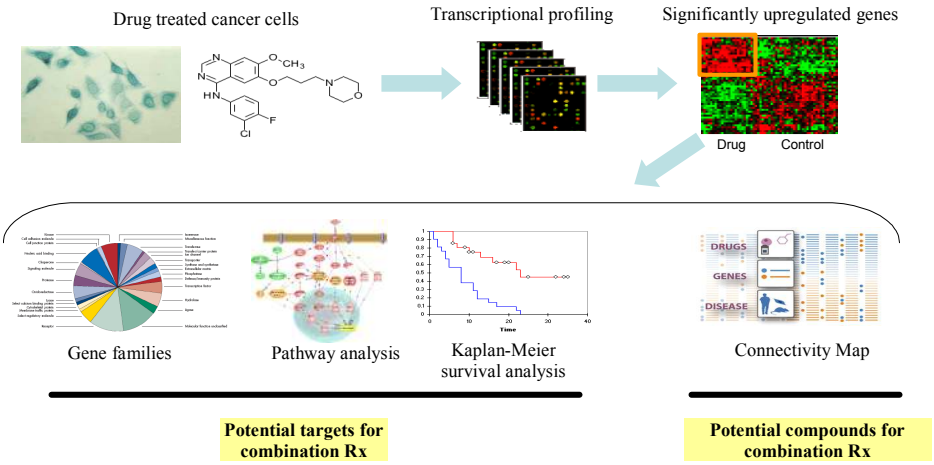


Figure 24. Overview of analysis to mine gene expression profiles for combination therapy targets and compounds. Statistically significant (FDR <0.1) upregulated (>2 fold) gene changes were extracted from cancer cells treated with standard chemotherapeutic drugs and hybridized onto high resolution gene chips. These genes were further analyzed for druggable classes and pathways. Clinical relevance was determined by querying publicly available tumor gene expression data coupled with survival information through Kaplan Meier survival analysis. Furthermore, the transcript profiles were matched against hundreds of compounds in the Connectivity Map [110] to identify compounds for follow up combination studies.

3.3.1 Microarray dataset analysis

We identified an internal microarray dataset where GC3 colon carcinoma cells were treated with 5-FU over time (12 and 24h) and transcript profiling was done on the

Affymetrix human HU-133A Plus 2 chip, representing 54 675 probesets. After analyzing for statistically significant gene expression changes (see Materials and Methods), 364 probesets corresponding to 285 genes were obtained where at least a 2 fold change in expression was seen at 12 or 24 hour time point. While a pleiotropic response is noted, several genes that are involved in cell death and apoptosis are modulated as expected since 5-FU is an anti-metabolite with strong cytotoxic properties.

3.3.2 Targets that are upregulated by compound treatment

We classified genes modulated by 5-FU treatment into gene families and restricted the list to targets that belonged to potential druggable classes. This resulted in 184 probesets (148 genes) that mapped to 4 kinases, 12 cell surface receptors and 16 secreted (growth factor, cytokine, soluble peptidases and receptors) proteins. A detailed breakdown is shown in Table 15 and complete details corresponding to genes in a partial list are given in Table 17.

Table 15: List of gene families for genes modulated by 5-FU treated GC3 colon carcinoma cells. Gene level counts are shown for probesets filtered for FDR <0.1 and foldchange >2.

Gene family	Count
ion channel	1
phosphatase	2
kinase	4
transporter	6
receptor	12
enzyme	13
secreted	16
transcription regulator	27
other	48

It is interesting to note that PLAU and PLAUR are both secreted and consistently upregulated at both time points. PLAU is urokinase plasminogen activator that binds to PLAUR and converts plasminogen to plasmin by cleaving an Arg-Val bond and hence, causes degradation of extracellular matrix and possibly contributes to cell migration, angiogenesis and metastasis. Arguably, upregulation of these genes would lead to an invasive phenotype leading to chemoresistance, as reported by several groups [156]. Intriguingly, Alfano et al have shown that PLAUR expression can promote resistance to apoptosis by increasing BCL-XL levels through MEK/ERK and PI3K/AKT dependent pathways [157]. In a similar way, upregulation of MMP1, an extracellular protease, can

contribute to enhanced migration and proliferation. In addition, increased expression of growth factors such as amphiregulin (AREG) and epiregulin (EREG), EGF-like ligands, are noted. These trigger salvage pathways that increase downstream signaling to promote growth. Amphiregulin over expression, in particular, has been observed in 5-FU resistant HCT116 derivative lines [158]. IL8 is a pro-inflammatory cytokine that promotes NFkB mediated anti-apoptotic and pro survival pathways and is significantly upregulated at both times. This is further supported by BCL10 expression which plays a role in NFkB activation by forming a complex with MALT1 and CARD-family proteins. On the other hand, FLT1 and PLK are two popular kinase targets that are modulated by 5-FU treatment. FLT1, also known as VEGFR1, is a receptor tyrosine kinase that binds VEGF and other ligands to promote angiogenesis. Polo-like kinases are cell cycle regulators whose aberrant expression leads to uncontrolled mitosis and prevention of growth arrest. Surprisingly, PLK1 is downregulated (data not shown), but PLK2 is upregulated. PLK2 is a relatively less studied member of the Plk family, but shares a high degree of homology with PLK1 and may have a redundant compensatory function.

While we have noted several interesting genes that are over expressed upon 5-FU treatment, determining the ramifications at a pathway level would be warranted. We therefore carried out pathway analysis to evaluate enrichment, if any, of relevant canonical pathways (see Figure 25). Consistent with our initial observation, several growth (p53, cell cycle, IGF1, EGF), survival (p38, NFkB) and signaling pathways (PPAR, PI3K/Akt) can be seen.

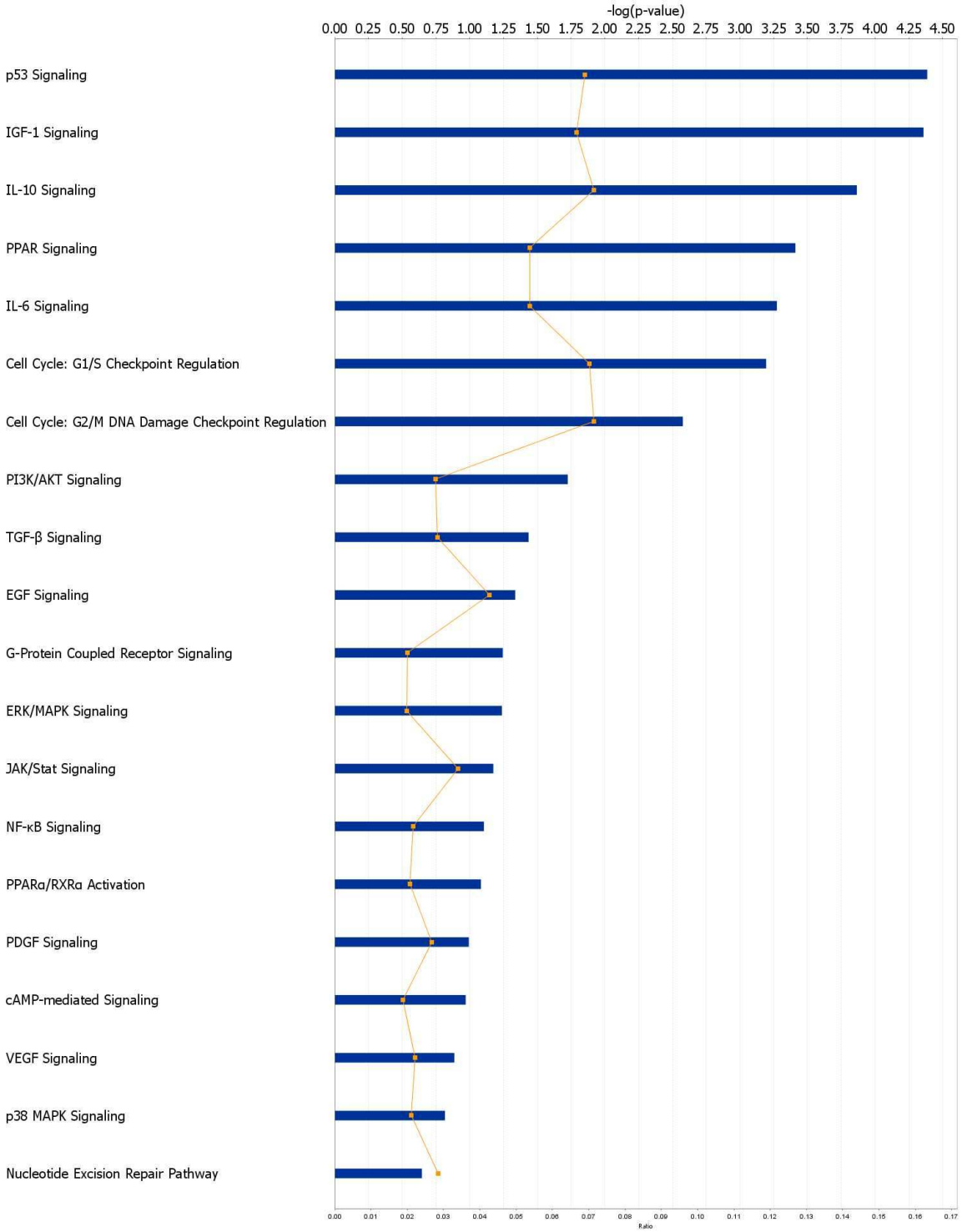


Figure 25. Pathways upregulated by 5-FU treated GC3 colon carcinoma cells. A snapshot of canonical pathways for upregulated genes in GC3 cells treated with 5-FU at 48h. Analysis was done using Ingenuity Pathway Analysis (IPA). Several cell growth, survival and signaling pathways can be seen. Ratio of observed and expected genes as well as –logP for each pathway are shown.

3.3.3 Survival data

To further prioritize clinically relevant targets, we cross-referenced the above gene list with publicly available primary tumor microarray data with associated clinical information. Univariate Kaplan Meier survival analyses were performed (see Materials and Methods), as mentioned before, to produce a list of genes whose over expression significantly ($p<0.05$) stratified patients with poor prognosis in a variety of tumors. A total of 60 probesets that corresponded to 51 genes were obtained (Table 16). These genes provide a diverse set of clinically relevant and druggable targets that are upregulated by 5-FU treatment and represent a rich repertoire of targets for inhibitors that can be administered concurrently or sequentially with 5-FU to decrease resistance and enhance therapeutic outcome. In our analysis, taking all the primary lung tumor gene expression datasets collectively, PTN, PLAU and PLAUR are most frequently observed as several instances of probesets (>3) seem to be associated with poor outcome. Pleiotrophin (PTN) is an angiogenic factor that causes endothelial vascularization and metastasis of tumor cells. Jager et al. [159] have measured plasma and serum levels of PTN in lung cancer patients and found that it positively correlates with stage of disease and negatively with response to treatment. Our results are also is in agreement with a recent study that found that increased levels of PLAU and PLAUR correlate with poor prognosis of small cell lung cancer patients [156]. Furthermore, inhibition of PLAUR in a mouse model of NSCLC enhanced tumor regression [160]. Similarly, most instances of probesets (>2) for ERCC1 and CYR61 were upregulated in brain tumors with lower survival. ERCC1 has been extensively studied and overexpression can lead to chemoresistance. ERCC1 confers protection of DNA damage of S-phase selective agents like 5-FU through activation of DNA repair. CYR61 is a cystein rich angiogenic inducer that has been shown to be prognostic for tumor progression and survival of glioma patients [161]. In our breast tumor data, EMP1, DUSP6 and JUN are overexpressed and have the highest number of associations (>3) with survival. EMP1 is a cell surface marker that appears to be prognostic in lung and breast tumors as well (Table 16). DUSP6 is a phosphatase that is upregulated in a variety of breast tumors and may predict

resistance to tamoxifen [162]. JUN is a transcription factor that regulates several target genes in breast tumors and increased expression is likely to have pleiotropic effects.

Table 16: Summary of Kaplan Meier survival analysis of 51 genes upregulated by 5-FU treated GC3 colon carcinoma cells. The number of probesets corresponding to a given gene in each of the tumor types (see Materials and Methods) analyzed are shown.

Genes	Lung	Ovarian	HNSCC	Brain	Lymphoma	Breast
TRIM29	2	2	1	1		2
PTN	4			1		3
EMP1	3					5
ERCC1	2		1	3		1
DUSP6	1	1		1		4
CYR61	1			3		3
PLAU	4	1		1		
JUN	1			1		4
ISG2	2			1	1	2
PLAUR	4		1			
FLT1	2		1	2		
F2RL1	1	1		1	1	1
EGR1				1	1	3
CCNE2	1	1				3
ZNF273	1		1	1		1
TAX1BP3				1		3
S1A11				1	1	2
PLK2	1		1	1		1
KLF6	1			1		2
ITGA2	2			1		1
FOS	1			1		2
BTG1	3	1				
BHLHB2	2			1		1
SMOX	1					2
NRP1	2					1
GJB5				1		2
DCAMKL1	2	1				
CREB5	2					1
BCL1		1		1		1
UPP1	1			1		
TFE3	1					1
MMP1					1	1
MAST4	1					1
LAMA3	1					1
IL8	1			1		
IL18	1	1				
GDF15				1		1
FOSL1	1					1
EGR4						2
CDKN2B			1	1		
CCNE1				1		1
AQP3	1					1
SEMA4B	1					
LRP1	1					
LOC28313	1					
GJB3						1
EREG	1					
EGR3						1
CDKN1A	1					
AREG	1					
ANXA3						1

Table 17: Summary of genes modulated by 5-FU treated GC3 cells. Filtered probesets (FDR<0.1; foldchange >2 at either 12h or 24h time point) were classified into relevant gene families (Table 15) and analyzed by Kaplan Meir survival analysis to determine prognosis in primary tumor patients (See text and Table 16 for details). The last two columns represent fold change values relative to control seen at the indicated time point.

Probeset	Gene Symbol	Gene Description	Family	Poor prognosis?	5FU_12h	5FU_24h
209369_at	ANXA3	annexin A3	enzyme	Yes	1.27	2.17
203720_s_at	ERCC1	excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense sequence)	enzyme	Yes	1.43	3.22
33304_at	ISG20	interferon stimulated exonuclease gene 20kDa	enzyme	Yes	1.17	2.52
1555680_a_at	SMOX	spermine oxidase	enzyme	Yes	2.07	3.13
210357_s_at	SMOX	spermine oxidase	enzyme	Yes	1.62	2.65
203234_at	UPP1	uridine phosphorylase 1	enzyme	Yes	1.46	2.54
202284_s_at	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	kinase	Yes	2.18	3.01
229800_at	DCLK1	doublecortin-like kinase 1	kinase	Yes	1.18	2.35
226498_at	FLT1	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)	kinase	Yes	2.21	3.47
40016_g_at	MAST4	microtubule associated serine/threonine kinase family member 4	kinase	Yes	1.38	2.54
201939_at	PLK2	polo-like kinase 2 (Drosophila)	kinase	Yes	1.79	2.09
208891_at	DUSP6	dual specificity phosphatase 6	phosphatase	Yes	1.49	2.29
208893_s_at	DUSP6	dual specificity phosphatase 6	phosphatase	Yes	1.77	2.17
1564796_at	EMP1	epithelial membrane protein 1	receptor	Yes	1.41	2.67

201324_at	EMP1	epithelial membrane protein 1	receptor	Yes	1.77	4.68
201325_s_at	EMP1	epithelial membrane protein 1	receptor	Yes	1.42	2.87
213895_at	EMP1	epithelial membrane protein 1	receptor	Yes	1.56	2.86
213506_at	F2RL1	coagulation factor II (thrombin) receptor-like 1	receptor	Yes	1.39	2.25
227314_at	ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	receptor	Yes	1.39	3.14
201412_at	LRP10	low density lipoprotein receptor-related protein 10	receptor	Yes	1.39	2.63
227252_at	LRP10	low density lipoprotein receptor-related protein 10	receptor	Yes	1.93	2.71
212298_at	NRP1	neuropilin 1	receptor	Yes	1.18	2.18
201289_at	CYR61	cysteine-rich, angiogenic inducer, 61	secreted	Yes	1.49	2.14
203726_s_at	LAMA3	laminin, alpha 3	secreted	Yes	2.44	11.05
200660_at	S100A11	S100 calcium binding protein A11	secreted	Yes	1.40	2.51
234725_s_at	SEMA4B	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4B	secreted	Yes	1.33	3.03
206295_at	IL18	interleukin 18 (interferon-gamma-inducing factor)	secreted	Yes	1.48	3.17
202859_x_at	IL8	interleukin 8	secreted	Yes	3.97	11.22
211506_s_at	IL8	interleukin 8	secreted	Yes	2.85	7.60
205239_at	AREG	amphiregulin (schwannoma-derived growth factor)	secreted	Yes	2.01	3.72
205767_at	EREG	epiregulin	secreted	Yes	2.47	5.52
221577_x_at	GDF15	growth differentiation factor 15	secreted	Yes	2.53	3.63

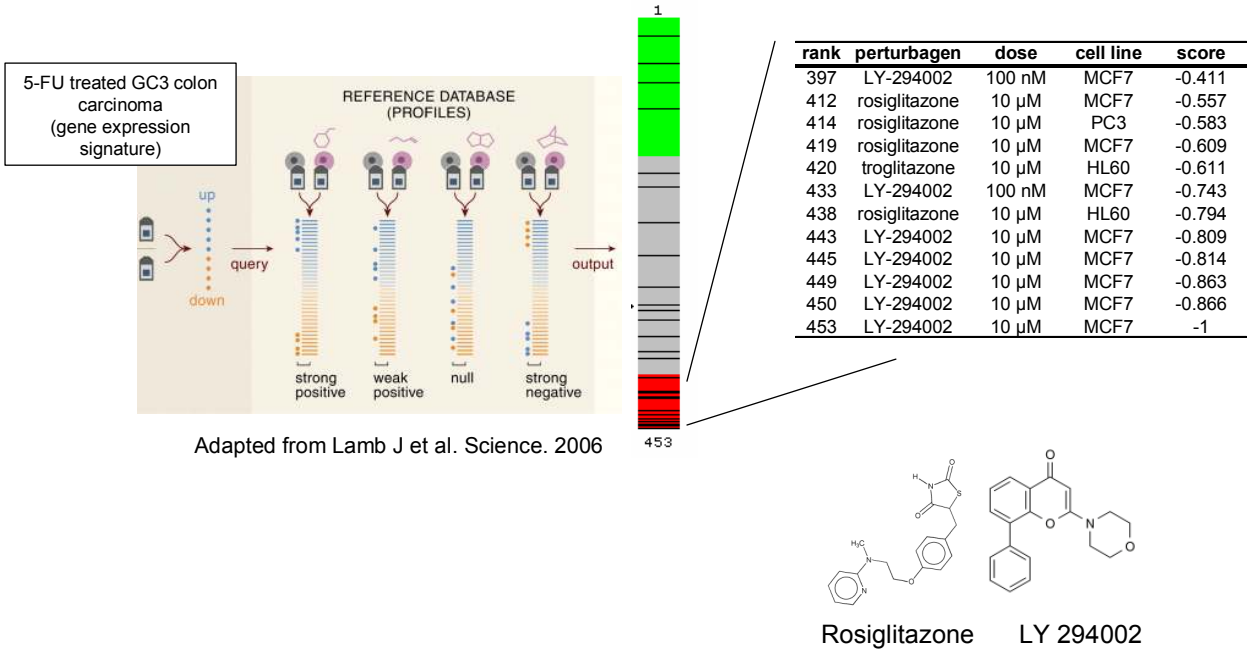
209465_x_at	PTN	pleiotrophin (heparin binding growth factor 8, neurite growth-promoting factor 1)	secreted	Yes	1.95	2.01
204475_at	MMP1	matrix metallopeptidase 1 (interstitial collagenase)	secreted	Yes	1.13	3.05
205479_s_at	PLAU	plasminogen activator, urokinase	secreted	Yes	2.79	5.76
211668_s_at	PLAU	plasminogen activator, urokinase	secreted	Yes	2.72	4.40
210845_s_at	PLAUR	plasminogen activator, urokinase receptor	secreted	Yes	1.52	3.12
214866_at	PLAUR	plasminogen activator, urokinase receptor	secreted	Yes	1.31	2.05
1557257_at	BCL10	B-cell CLL/lymphoma 10	transcription regulator	Yes	1.61	2.18
205263_at	BCL10	B-cell CLL/lymphoma 10	transcription regulator	Yes	1.43	2.27
201169_s_at	BHLHB2	basic helix-loop-helix domain containing, class B, 2	transcription regulator	Yes	2.03	4.29
200920_s_at	BTG1	B-cell translocation gene 1, anti-proliferative	transcription regulator	Yes	1.11	2.11
213523_at	CCNE1	cyclin E1	transcription regulator	Yes	2.18	2.16
205034_at	CCNE2	cyclin E2	transcription regulator	Yes	2.78	2.20
211814_s_at	CCNE2	cyclin E2	transcription regulator	Yes	2.59	1.78
236313_at	CDKN2B	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	transcription regulator	Yes	1.40	2.64
229228_at	CREB5	cAMP responsive element binding protein 5	transcription regulator	Yes	1.37	2.72
201693_s_at	EGR1	early growth response 1	transcription regulator	Yes	1.78	4.05
201694_s_at	EGR1	early growth response 1	transcription regulator	Yes	1.93	3.99
206115_at	EGR3	early growth response 3	transcription regulator	Yes	2.59	4.84
207768_at	EGR4	early growth response 4	transcription regulator	Yes	5.52	10.03

209189_at	FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog	transcription regulator	Yes	3.48	4.69
204420_at	FOSL1	FOS-like antigen 1	transcription regulator	Yes	1.68	3.21
201464_x_at	JUN	jun oncogene	transcription regulator	Yes	1.66	2.56
201466_s_at	JUN	jun oncogene	transcription regulator	Yes	1.92	2.35
1555832_s_at	KLF6	Kruppel-like factor 6	transcription regulator	Yes	1.12	2.23
208960_s_at	KLF6	Kruppel-like factor 6	transcription regulator	Yes	1.02	2.45
208961_s_at	KLF6	Kruppel-like factor 6	transcription regulator	Yes	1.02	2.51
224606_at	KLF6	Kruppel-like factor 6	transcription regulator	Yes	1.15	2.25
209154_at	TAX1BP3	Tax1 (human T-cell leukemia virus type I) binding protein 3	transcription regulator	Yes	1.19	2.29
215464_s_at	TAX1BP3	Tax1 (human T-cell leukemia virus type I) binding protein 3	transcription regulator	Yes	1.20	2.26
212457_at	TFE3	transcription factor binding to IGHM enhancer 3	transcription regulator	Yes	1.30	2.46
202504_at	TRIM29	tripartite motif-containing 29	transcription regulator	Yes	2.37	2.76
243661_at	ZNF273	zinc finger protein 273	transcription regulator	Yes	1.80	2.33
39248_at	AQP3	aquaporin 3 (Gill blood group)	transporter	Yes	2.36	3.82
205490_x_at	GJB3	gap junction protein, beta 3, 31kDa	transporter	Yes	1.79	3.14
215243_s_at	GJB3	gap junction protein, beta 3, 31kDa	transporter	Yes	1.46	2.21
206156_at	GJB5	gap junction protein, beta 5, 31.1kDa	transporter	Yes	1.96	3.13
236436_at	SLC25A45	solute carrier family 25, member 45	transporter	Yes	1.63	2.41

3.3.4 Connectivity Map analysis

The Connectivity Map [110] contains a compilation of >400 gene-expression profiles derived from treating cultured human cells (MCF7, PC3, HL60, SKMEL5, HepG2, SHSY5Y) with a large number of perturbagens to populate a reference database. Pattern-matching algorithms were employed to score each reference profile for the direction and strength of enrichment with our query signature of 5FU modulated gene changes. A connectivity score was used to rank all perturbagens from most positive (top) to negative (bottom) (see Figure 26). We carried out this analysis to identify compounds whose signatures were negatively connected to our 5-FU treated gene expression signature, suggesting possible options for combination or sequential therapy. While the actual score has little meaning, the relative ranking of perturbagens in the list and multiple occurrences of compound different doses or compounds of the same class increase the likelihood of a true positive. Taking these criteria into consideration, we found that multiple instances of LY294002 and TZDs (namely, rosiglitazone and troglitazone) are strongly associated by a negative connection. LY 294002 is a PI3 kinase inhibitor and TZDs are a well-studied class of PPAR agonists. In the light of these observations, it is interesting to note that PI3K/Akt signaling and PPAR signaling were seen in our pathway analysis results (see Figure 25). It is surprising though that PPAR agonists are likely candidates to combine or add onto 5-FU therapy when PPAR upregulation was seen with 5-FU treatment. We hypothesize that this may be due to downstream effects of PPAR activation that are not entirely intuitive, but effectively 'complement' 5-FU activity. As such, PPAR agonists have been frequently employed and are seen to provide favorable endpoints in colorectal carcinoma. To verify the effectiveness of the suggested agents in a combination drug setting in colon cancer as our signature was derived from GC3 cells, we found 2 supporting publications. In one study [163], HT29 colon cancer cells treated with rosiglitazone and 5-FU underwent higher cell death than with either drug alone (Figure 27a). In another study [164], LY 294002 and wortmannin (both PI3K inhibitors) resulted in significant apoptosis, as measured by DNA fragmentation, in combination with 5-FU relative to either agent alone in KM20 colon cancer cells (Figure 27b). Taken together, PI3K inhibition and TZDs are likely to be effective in enhancing the effect of 5-

FU in through simultaneous or sequential use in colon cancer. To our knowledge, this is the first application of the Connectivity Map to identify combination therapy agents.



rank	perturbagen	dose	cell line	score
397	LY-294002	100 nM	MCF7	-0.411
412	rosiglitazone	10 μM	MCF7	-0.557
414	rosiglitazone	10 μM	PC3	-0.583
419	rosiglitazone	10 μM	MCF7	-0.609
420	troglitazone	10 μM	HL60	-0.611
433	LY-294002	100 nM	MCF7	-0.743
438	rosiglitazone	10 μM	HL60	-0.794
443	LY-294002	10 μM	MCF7	-0.809
445	LY-294002	10 μM	MCF7	-0.814
449	LY-294002	10 μM	MCF7	-0.863
450	LY-294002	10 μM	MCF7	-0.866
453	LY-294002	10 μM	MCF7	-1

Figure 26. Connectivity Map analysis using 5-FU treated GC3 colon carcinoma gene expression signature. The Broad Institute has compiled >400 gene-expression profiles derived from treating cells with a large number of perturbagens (at various concentrations) to populate a reference database. Compounds are ranked by a "connectivity score"; those at the top ("positive") and bottom ("negative") are functionally connected with the query signature. Above, LY294002 and TZDs are strongly associated by a negative score to the 5-FU gene expression signature, suggesting complementarity in drug action.

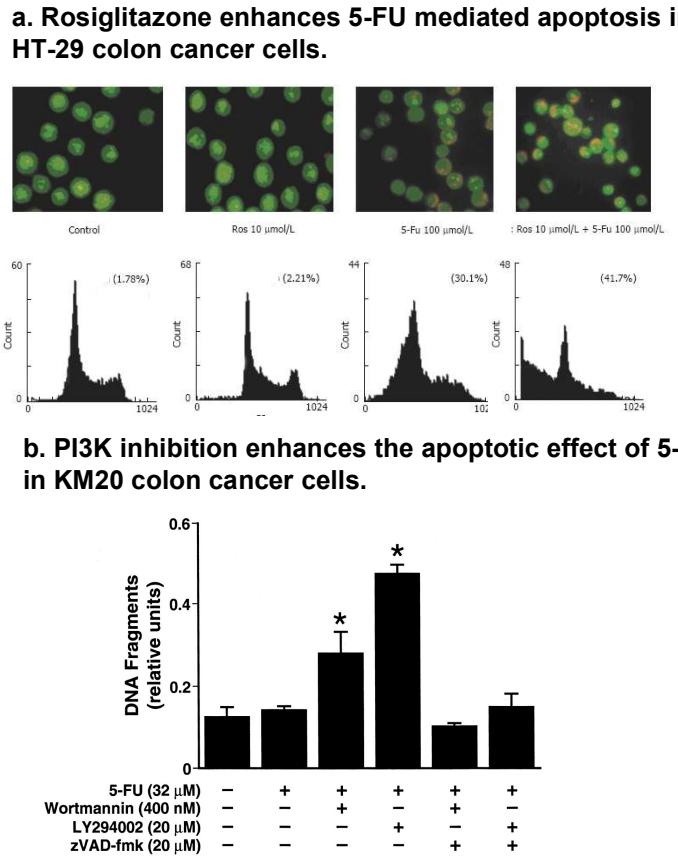


Figure 27. Published evidence of effective combinations of 5-FU with rosiglitazone and PI3K inhibitors. (a) Apoptosis of HT29 cells as measured by acridine orange/ethidium bromide staining (top panel) or flow cytometry (bottom panel) for Ros (rosiglitazone), control, 5-FU and the combination. Work done by Zhang et al.[163] (b) Apoptosis of KM20 cells as measured by DNA fragmentation assay for each treatment. *, $p<0.05$; zVAD-fmk, pan-caspase inhibitor. Work done by Wang et al. [164]. Figures were adapted from the original publications which contain details of materials and methods.

3.3.5 Discussion

Cancer is a complex disease with multiple genetic aberrations. Treatment by a single agent is likely to trigger resistance through innate or acquired mechanisms (see Introduction). A case in point is 5-FU which is highly effective and widely used to treat cancer patients, progressive chemoresistance is commonly encountered and poses a significant clinical challenge. To gain a better understanding of the molecular events at play, most current approaches have relied on extracting gene sets from pre- or untreated tumor samples based on chemosensitivity correlations. Essentially these are signatures of response to predict innate or basal sensitivity or resistance to a given compound treatment.

For instance, several *in vitro* and *in vivo* studies have demonstrated that increased levels of TYMS and DPD correlate and cause resistance to 5-FU.

What we have attempted here is to use compound treated gene expression profiles to identify complementary targets and therapies that would increase the probability of clinical success in combination or as adjuvant therapy and delay or prevent acquired chemoresistance. While this is not easily detected in an acute treatment experiment as the one employed in our study (i.e. GC3 cells treated with 5-FU at 12 and 24 hours), we hypothesized that due to systemic changes associated with drug action, pro-survival gene changes can be detected temporally and can be distilled into attractive drug targets based on biological and clinical relevance. Also, intelligent use of gene expression based signatures in Connectivity Map can provide compound leads with potential for combination or add on to primary chemotherapeutic regimen.

This was by no means a comprehensive survey of all druggable targets that are modulated by 5-FU treatment. Our analysis suffers from several drawbacks. We analyzed a sub-optimal dataset with a single dose of 5-FU administered at only two time points. Also, all of our results are based on a single cell line which is sensitive to the compound under study. Sensitive tumor samples as well as resistant cell lines and/or tumors are likely to reveal other genes. For future follow experiments, we recommend experimental confirmation of upregulated genes by RT-PCR and Western blots, wherever possible, in a panel of lines with varying conditions of compound dose, time effects followed by *in vitro* and *in vivo* combination studies. Furthermore, with the availability of large public databases such as GEO (http://www.ncbi.nih.gov/geo), this analysis can be extended to other relevant compound treated cell lines/patient profiling datasets. Incorporation of such standard-of-care/targeted agents treated gene expression datasets to identify common targets across compounds for a given tumor/cell line and common targets across all tumors/cell lines for a given compound would be informative in aiding drug development efforts. A major limitation with the current analysis is incomplete knowledge around gene changes that are specific to 5-FU versus any cytotoxic agent. Although we focused on overexpressed genes for practical reasons, analysis of underexpressed genes maybe warranted as well. There are likely other genes that were missed by our approach either due to the univariate Kaplan Meier analysis or the datasets

examined. Therefore, we recommend expanding the repertoire of datasets coupled with clinical covariates and employing multivariate tests such as Cox proportional hazards.

It can be argued that several of these targets could also serve as pharmacogenomic markers of response in pre-treated samples to detect innate resistance. Although we found limited overlap after re-analyzing their study (data not shown), Boyer and colleagues demonstrated an overlap with HCT116 parental lines treated with 5-FU or oxaliplatin and derived resistant lines [165].

Lastly, systematic and guided analysis of compendia of compound treated gene expression datasets in model systems (like Connectivity Map) can provide compound leads for rational combination therapy. However, care must be taken in using the most informative query signatures, generating meaningful reference signatures and interpretation of results.

In conclusion, we have demonstrated that despite sub-optimal datasets and analyses, analyzing gene expression data of cancer cells treated with a standard oncoloytic agents can reveal clinically relevant druggable targets. By using a novel approach, we have even provided compounds that may work well when administered simultaneously or sequentially. These valuable hypotheses hold great potential and with further followup, their utility in combination therapy or second/third line adjuvant opportunities for better clinical outcome in patients can be realized.

4 Summary

Cancer is a complex acquired genetic disease which is the result of molecular aberrations in multiple targets in various tissues, pathways and stages of disease progression. In recent times, a wealth of unprecedented genomics data has been generated and made accessible. We have demonstrated that innovative and integrated approaches can be applied to translate this rich information into knowledge. While RNAi screens have aforementioned limitations, it is noteworthy that several targets have been prioritized based on biological and/or clinical support of 'oncogene addiction' for cell survival in our first case study. From several intriguing results, we confirmed AURKA, KIF11, PLK1 and ILK as genes that had a broad spectrum lethal phenotype upon siRNA mediated inhibition. On the other hand, NOTCH4, AKT1 and MCL1 were followed up as cell-specific survival genes. MCL1, in particular, is amplified, overexpressed and a poor prognosticator in breast cancer and demonstrates the power of fusing different datasets. To our knowledge, this is the first comprehensive analysis of high-throughput RNAi screens combined with other high dimensional datasets. In our second case study, we collectively analyzed datasets from SYK wild type and defective contexts to generate valuable insights into SYK biology that warrant experimental validation. The recently uncovered mutations appear to be inactivating as the mutant cell lines share profiles and pathways in common with SYK-deficient cell lines and this is likely to have implications for EGFR inhibitor therapy. Survival analysis revealed novel information about the context-specific prognostic role of SYK expression in a variety of tumors. These investigations represents a 'bottoms up' systems approach and can be applied to a rapidly growing collection of novel genetic mutations, an active area of research empowered by advanced and cheaper sequencing technologies. To address acquired drug resistance in our last case study, 5-FU treated colon cancer cell line expression profiles were overlayed with druggable targets and pathways whose overexpression resulted in poor prognosis in primary tumor samples. Furthermore, based on Connectivity Map [110] analysis and literature support, TZDs and PI3K inhibitors may be effectively combined with 5-FU administration. We believe these analyses can be applied to other agents and tumor types to suggest rational combination or second/third line adjuvant therapy for testing in the lab and clinic. Collectively, these evolving systems and integrative approaches in

oncogenomics hold great promise and are necessary to enhance our understanding of tumor biology and to create opportunities for improvement of existing or future cancer treatment

5 References

1. Aderem, A., *Systems biology: its practice and challenges.* Cell, 2005. 121(4): p. 511-3.

2. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. 100(1): p. 57-70.

3. Fazzari, M.J. and J.M. Greally, *Epigenomics: beyond CpG islands.* Nat Rev Genet, 2004. 5(6): p. 446-55.

4. Callinan, P.A. and A.P. Feinberg, *The emerging science of epigenomics.* Hum Mol Genet, 2006. 15 Spec No 1: p. R95-101.

5. Jones, P.A. and S.B. Baylin, *The epigenomics of cancer.* Cell, 2007. 128(4): p. 683-92.

6. Lynch, T.J., et al., *Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib.* N Engl J Med, 2004. 350(21): p. 2129-39.

7. Paez, J.G., et al., *EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy.* Science, 2004. 304(5676): p. 1497-500.

8. Pao, W., et al., *EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib.* Proc Natl Acad Sci U S A, 2004. 101(36): p. 13306-11.

9. Bardelli, A., et al., *Mutational analysis of the tyrosine kinome in colorectal cancers.* Science, 2003. 300(5621): p. 949.

10. Wang, Z., et al., *Mutational analysis of the tyrosine phosphatome in colorectal cancers.* Science, 2004. 304(5674): p. 1164-6.

11. Samuels, Y., et al., *High frequency of mutations of the PIK3CA gene in human cancers.* Science, 2004. 304(5670): p. 554.

12. Parsons, D.W., et al., *Colorectal cancer: mutations in a signalling pathway.* Nature, 2005. 436(7052): p. 792.

13. Davies, H., et al., *Mutations of the BRAF gene in human cancer.* Nature, 2002. 417(6892): p. 949-54.

14. Stephens, P., et al., *A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer.* Nat Genet, 2005. 37(6): p. 590-2.

15. Davies, H., et al., *Somatic mutations of the protein kinase gene family in human lung cancer.* Cancer Res, 2005. 65(17): p. 7591-5.

16. Bignell, G., et al., *Sequence analysis of the protein kinase gene family in human testicular germ-cell tumors of adolescents and adults.* Genes Chromosomes Cancer, 2006. 45(1): p. 42-6.

17. Greenman, C., et al., *Patterns of somatic mutation in human cancer genomes.* Nature, 2007. 446(7132): p. 153-8.

18. Futreal, P.A., et al., *A census of human cancer genes.* Nat Rev Cancer, 2004. 4(3): p. 177-83.

19. Vogelstein, B. and K.W. Kinzler, *Cancer genes and the pathways they control.* Nat Med, 2004. 10(8): p. 789-99.

20. Sjoblom, T., et al., *The consensus coding sequences of human breast and colorectal cancers.* Science, 2006. 314(5797): p. 268-74.

21. Wood, L.D., et al., *The genomic landscapes of human breast and colorectal cancers.* Science, 2007. 318(5853): p. 1108-13.

22. Ruhe, J.E., et al., *Genetic alterations in the tyrosine kinase transcriptome of human cancer cell lines.* Cancer Res, 2007. 67(23): p. 11368-76.

23. Amos, C.I., et al., *Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1.* Nat Genet, 2008. 40(5): p. 616-22.

24. Cox, A., et al., *A common coding variant in CASP8 is associated with breast cancer risk.* Nat Genet, 2007. 39(3): p. 352-8.

25. Easton, D.F., et al., *Genome-wide association study identifies novel breast cancer susceptibility loci.* Nature, 2007. 447(7148): p. 1087-93.

26. Gold, B., et al., *Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33.* Proc Natl Acad Sci U S A, 2008. 105(11): p. 4340-5.

27. Gudmundsson, J., et al., *Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24.* Nat Genet, 2007. 39(5): p. 631-7.

28. Hunter, D.J., et al., *A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer.* Nat Genet, 2007. 39(7): p. 870-4.

29. Tenesa, A., et al., *Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21.* Nat Genet, 2008. 40(5): p. 631-7.

30. Thomas, G., et al., *Multiple loci identified in a genome-wide association study of prostate cancer.* Nat Genet, 2008. 40(3): p. 310-5.

31. Tomlinson, I.P., et al., *A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3.* Nat Genet, 2008. 40(5): p. 623-30.

32. Yeager, M., et al., *Genome-wide association study of prostate cancer identifies a second risk locus at 8q24.* Nat Genet, 2007. 39(5): p. 645-9.

33. Zanke, B.W., et al., *Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24.* Nat Genet, 2007. 39(8): p. 989-94.

34. Parkinson, H., et al., *ArrayExpress--a public repository for microarray gene expression data at the EBI.* Nucleic Acids Res, 2005. 33(Database issue): p. D553-5.

35. Barrett, T., et al., *NCBI GEO: mining millions of expression profiles--database and tools.* Nucleic Acids Res, 2005. 33(Database issue): p. D562-6.

36. Alizadeh, A.A., et al., *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.* Nature, 2000. 403(6769): p. 503-11.

37. Cuadros, M., et al., *Identification of a proliferation signature related to survival in nodal peripheral T-cell lymphomas.* J Clin Oncol, 2007. 25(22): p. 3321-9.

38. Dave, S.S., et al., *Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells.* N Engl J Med, 2004. 351(21): p. 2159-69.

39. Huang, J.Z., et al., *The t(14;18) defines a unique subset of diffuse large B-cell lymphoma with a germinal center B-cell gene expression profile.* Blood, 2002. 99(7): p. 2285-90.

40. Hummel, M., et al., *A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling.* N Engl J Med, 2006. 354(23): p. 2419-30.

41. Shipp, M.A., et al., *Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning.* Nat Med, 2002. 8(1): p. 68-74.

42. Monti, S., et al., *Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response.* Blood, 2005. 105(5): p. 1851-61.

43. Beer, D.G., et al., *Gene-expression profiles predict survival of patients with lung adenocarcinoma.* Nat Med, 2002. 8(8): p. 816-24.

44. Bhattacharjee, A., et al., *Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses.* Proc Natl Acad Sci U S A, 2001. 98(24): p. 13790-5.

45. Bild, A.H., et al., *Oncogenic pathway signatures in human cancers as a guide to targeted therapies.* Nature, 2006. 439(7074): p. 353-7.

46. Raponi, M., et al., *Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung.* Cancer Res, 2006. 66(15): p. 7466-72.

47. Buyse, M., et al., *Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer.* J Natl Cancer Inst, 2006. 98(17): p. 1183-92.

48. Sotiriou, C., et al., *Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis.* J Natl Cancer Inst, 2006. 98(4): p. 262-72.

49. Dai, H., et al., *A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients.* Cancer Res, 2005. 65(10): p. 4059-66.

50. Glinsky, G.V., T. Higashiyama, and A.B. Glinskii, *Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm.* Clin Cancer Res, 2004. 10(7): p. 2272-83.

51. Ivshina, A.V., et al., *Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer.* Cancer Res, 2006. 66(21): p. 10292-301.

52. Liu, R., et al., *The prognostic role of a gene signature from tumorigenic breast-cancer cells.* N Engl J Med, 2007. 356(3): p. 217-26.

53. Ma, X.J., et al., *A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen.* Cancer Cell, 2004. 5(6): p. 607-16.

54. Miller, L.D., et al., *An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival.* Proc Natl Acad Sci U S A, 2005. 102(38): p. 13550-5.

55. Pawitan, Y., et al., *Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts.* Breast Cancer Res, 2005. 7(6): p. R953-64.

56. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. 100(14): p. 8418-23.

57. Thomassen, M., et al., *Prediction of metastasis from low-malignant breast cancer by gene expression profiling.* Int J Cancer, 2007. 120(5): p. 1070-5.

58. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med, 2002. 347(25): p. 1999-2009.

59. Wang, Y., et al., *Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.* Lancet, 2005. 365(9460): p. 671-9.

60. Weigelt, B., et al., *Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer.* Cancer Res, 2005. 65(20): p. 9155-8.

61. De Cecco, L., et al., *Gene expression profiling of advanced ovarian cancer: characterization of a molecular signature involving fibroblast growth factor 2.* Oncogene, 2004. 23(49): p. 8171-83.

62. Spentzos, D., et al., *Gene expression signature with independent prognostic significance in epithelial ovarian cancer.* J Clin Oncol, 2004. 22(23): p. 4700-10.

63. Barrier, A., et al., *Stage II colon cancer prognosis prediction by tumor gene expression profiling.* J Clin Oncol, 2006. 24(29): p. 4685-91.

64. Barrier, A., et al., *Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling.* Oncogene, 2007. 26(18): p. 2642-8.

65. Giacomini, C.P., et al., *A gene expression signature of genetic instability in colon cancer.* Cancer Res, 2005. 65(20): p. 9200-5.

66. Dhanasekaran, S.M., et al., *Delineation of prognostic biomarkers in prostate cancer.* Nature, 2001. 412(6849): p. 822-6.

67. Glinsky, G.V., et al., *Gene expression profiling predicts clinical outcome of prostate cancer.* J Clin Invest, 2004. 113(6): p. 913-23.

68. Halvorsen, O.J., et al., *Gene expression profiles in prostate cancer: association with patient subgroups and tumour differentiation.* Int J Oncol, 2005. 26(2): p. 329-36.

69. Lin, B., et al., *Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease.* Cancer Res, 2005. 65(8): p. 3081-91.

70. Singh, D., et al., *Gene expression correlates of clinical prostate cancer behavior.* Cancer Cell, 2002. 1(2): p. 203-9.

71. Freije, W.A., et al., *Gene expression profiling of gliomas strongly predicts survival.* Cancer Res, 2004. 64(18): p. 6503-10.

72. Fuller, G.N., et al., *Molecular voting for glioma classification reflecting heterogeneity in the continuum of cancer progression.* Oncol Rep, 2005. 14(3): p. 651-6.

73. Kim, S., et al., *Identification of combination gene sets for glioma classification.* Mol Cancer Ther, 2002. 1(13): p. 1229-36.

74. Ligon, K.L., et al., *The oligodendroglial lineage marker OLIG2 is universally expressed in diffuse gliomas.* J Neuropathol Exp Neurol, 2004. 63(5): p. 499-509.

75. Phillips, H.S., et al., *Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis.* Cancer Cell, 2006. 9(3): p. 157-73.

76. Nutt, C.L., et al., *Gene expression-based classification of malignant gliomas correlates better with survival than histological classification.* Cancer Res, 2003. 63(7): p. 1602-7.

77. Pomeroy, S.L., et al., *Prediction of central nervous system embryonal tumour outcome based on gene expression.* Nature, 2002. 415(6870): p. 436-42.

78. Cromer, A., et al., *Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis.* Oncogene, 2004. 23(14): p. 2484-98.

79. Dyrskjot, L., et al., *Identifying distinct classes of bladder carcinoma using microarrays.* Nat Genet, 2003. 33(1): p. 90-6.

80. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. 406(6797): p. 747-52.

81. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, 2002. 415(6871): p. 530-6.

82. Foekens, J.A., et al., *Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer.* J Clin Oncol, 2006. 24(11): p. 1665-71.

83. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. 391(6669): p. 806-11.

84. Kittler, R., et al., *Genome-scale RNAi profiling of cell division in human tissue culture cells.* Nat Cell Biol, 2007. 9(12): p. 1401-12.

85. Bartz, S. and A.L. Jackson, *How will RNAi facilitate drug development?* Sci STKE, 2005. 2005(295): p. pe39.

86. Echeverri, C.J., et al., *Minimizing the risk of reporting false positives in large-scale RNAi screens.* Nat Methods, 2006. 3(10): p. 777-9.

87. Echeverri, C.J. and N. Perrimon, *High-throughput RNAi screening in cultured cells: a user's guide.* Nat Rev Genet, 2006. 7(5): p. 373-84.

88. Boutros, M., et al., *Genome-wide RNAi analysis of growth and viability in Drosophila cells.* Science, 2004. 303(5659): p. 832-5.

89. Iorns, E., et al., *Utilizing RNA interference to enhance cancer drug discovery.* Nat Rev Drug Discov, 2007. 6(7): p. 556-68.

90. Brummelkamp, T.R., et al., *Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-kappaB.* Nature, 2003. 424(6950): p. 797-801.

91. Berns, K., et al., *A large-scale RNAi screen in human cells identifies new components of the p53 pathway.* Nature, 2004. 428(6981): p. 431-7.

92. MacKeigan, J.P., L.O. Murphy, and J. Blenis, *Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance.* Nat Cell Biol, 2005. 7(6): p. 591-600.

93. Collins, C.S., et al., *A small interfering RNA screen for modulators of tumor cell motility identifies MAP4K4 as a promigratory kinase.* Proc Natl Acad Sci U S A, 2006. 103(10): p. 3775-80.

94. Westbrook, T.F., et al., *A genetic screen for candidate tumor suppressors identifies REST.* Cell, 2005. 121(6): p. 837-48.

95. Paddison, P.J., et al., *A resource for large-scale RNA-interference-based screens in mammals.* Nature, 2004. 428(6981): p. 427-31.

96. Schlabach, M.R., et al., *Cancer proliferation gene discovery through functional genomics.* Science, 2008. 319(5863): p. 620-4.

97. Silva, J.M., et al., *Profiling essential genes in human mammary cells by multiplex RNAi screening.* Science, 2008. 319(5863): p. 617-20.

98. Moffat, J., et al., *A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen.* Cell, 2006. 124(6): p. 1283-98.

99. Whitehurst, A.W., et al., *Synthetic lethal screen identification of chemosensitizer loci in cancer cells.* Nature, 2007. 446(7137): p. 815-9.

100. Berns, K., et al., *A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer.* Cancer Cell, 2007. 12(4): p. 395-402.

101. Nagata, Y., et al., *PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients.* Cancer Cell, 2004. 6(2): p. 117-27.

102. Alley, M.C., et al., *Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay.* Cancer Res, 1988. 48(3): p. 589-601.

103. Scherf, U., et al., *A gene expression database for the molecular pharmacology of cancer.* Nat Genet, 2000. 24(3): p. 236-44.

104. Staunton, J.E., et al., *Chemosensitivity prediction by transcriptional profiling.* Proc Natl Acad Sci U S A, 2001. 98(19): p. 10787-92.

105. Lee, J.K., et al., *A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery.* Proc Natl Acad Sci U S A, 2007. 104(32): p. 13086-91.

106. McDermott, U., et al., *Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling.* Proc Natl Acad Sci U S A, 2007. 104(50): p. 19936-41.

107. Garraway, L.A., et al., *Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma.* Nature, 2005. 436(7047): p. 117-22.

108. Boehm, J.S., et al., *Integrative genomic approaches identify IKBKE as a breast cancer oncogene.* Cell, 2007. 129(6): p. 1065-79.

109. Kim, M., et al., *Comparative oncogenomics identifies NEDD9 as a melanoma metastasis gene.* Cell, 2006. 125(7): p. 1269-81.

110. Lamb, J., et al., *The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease.* Science, 2006. 313(5795): p. 1929-35.

111. Rhodes, D.R., et al., *ONCOMINE: a cancer microarray database and integrated data-mining platform.* Neoplasia, 2004. 6(1): p. 1-6.

112. Rhodes, D.R., et al., *Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles.* Neoplasia, 2007. 9(2): p. 166-80.

113. Shi, L., et al., *The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.* Nat Biotechnol, 2006. 24(9): p. 1151-61.

114. Wang, T.L., et al., *Prevalence of somatic alterations in the colorectal cancer cell genome.* Proc Natl Acad Sci U S A, 2002. 99(5): p. 3076-80.

115. Taylor, C.F., *Progress in standards for reporting omics data.* Curr Opin Drug Discov Devel, 2007. 10(3): p. 254-63.

116. Tusher, V.G., R. Tibshirani, and G. Chu, *Significance analysis of microarrays applied to the ionizing radiation response.* Proc Natl Acad Sci U S A, 2001. 98(9): p. 5116-21.

117. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. 102(43): p. 15545-50.

118. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol, 2004. 5(10): p. R80.

119. Aza-Blanc, P., et al., *Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening.* Mol Cell, 2003. 12(3): p. 627-37.

120. Tuzmen, S., J. Kiefer, and S. Mousses, *Validation of short interfering RNA knockdowns by quantitative real-time PCR.* Methods Mol Biol, 2007. 353: p. 177-203.

121. Yoon, D.S., et al., *Variable levels of chromosomal instability and mitotic spindle checkpoint defects in breast cancer.* Am J Pathol, 2002. 161(2): p. 391-7.

122. Nugoli, M., et al., *Genetic variability in MCF-7 sublines: evidence of rapid genomic and RNA expression profile modifications.* BMC Cancer, 2003. 3: p. 13.

123. Tonon, G., et al., *High-resolution genomic profiles of human lung cancer.* Proc Natl Acad Sci U S A, 2005. 102(27): p. 9625-30.

124. Hidaka, S., et al., *Differences in 20q13.2 copy number between colorectal cancers with and without liver metastasis.* Clin Cancer Res, 2000. 6(7): p. 2712-7.

125. Ding, Q., et al., *Myeloid cell leukemia-1 inversely correlates with glycogen synthase kinase-3beta activity and associates with poor prognosis in human breast cancer.* Cancer Res, 2007. 67(10): p. 4564-71.

126. Carpten, J.D., et al., *A transforming mutation in the pleckstrin homology domain of AKT1 in cancer.* Nature, 2007. 448(7152): p. 439-44.

127. Harris, M.A., et al., *The Gene Ontology (GO) database and informatics resource.* Nucleic Acids Res, 2004. 32(Database issue): p. D258-61.

128. Hannigan, G., A.A. Troussard, and S. Dedhar, *Integrin-linked kinase: a cancer therapeutic target unique among its ILK.* Nat Rev Cancer, 2005. 5(1): p. 51-63.

129. Hagedorn, M., et al., *FBXW7/hCDC4 controls glioma cell proliferation in vitro and is a prognostic marker for survival in glioblastoma patients.* Cell Div, 2007. 2: p. 9.

130. Richardson, A.L., et al., *X chromosomal abnormalities in basal-like human breast cancer.* Cancer Cell, 2006. 9(2): p. 121-32.

131. Lin, X., et al., *'Seed' analysis of off-target siRNAs reveals an essential role of Mcl-1 in resistance to the small-molecule Bcl-2/Bcl-XL inhibitor ABT-737.* Oncogene, 2007. 26(27): p. 3972-9.

132. Spankuch, B., et al., *Rational combinations of siRNAs targeting Plk1 with breast cancer drugs.* Oncogene, 2007. 26(39): p. 5793-807.

133. Coopman, P.J. and S.C. Mueller, *The Syk tyrosine kinase: a new negative regulator in tumor growth and progression.* Cancer Lett, 2006. 241(2): p. 159-73.

134. Coopman, P.J., et al., *The Syk tyrosine kinase suppresses malignant growth of human breast cancer cells.* Nature, 2000. 406(6797): p. 742-7.

135. Yuan, Y., et al., *Frequent epigenetic inactivation of spleen tyrosine kinase gene in human hepatocellular carcinoma.* Clin Cancer Res, 2006. 12(22): p. 6687-95.

136. Fargnoli, J., et al., *Syk mutation in Jurkat E6-derived clones results in lack of p72syk expression.* J Biol Chem, 1995. 270(44): p. 26533-7.

137. Goodman, P.A., et al., *Hypermethylation of the spleen tyrosine kinase promoter in T-lineage acute lymphoblastic leukemia.* Oncogene, 2003. 22(16): p. 2504-14.

138. Leseux, L., et al., *Syk-dependent mTOR activation in follicular lymphoma cells.* Blood, 2006. 108(13): p. 4156-62.

139. Zipfel, P.A., et al., *Requirement for Abl kinases in T cell receptor signaling.* Curr Biol, 2004. 14(14): p. 1222-31.

140. Hoeller, C., et al., *The non-receptor-associated tyrosine kinase Syk is a regulator of metastatic behavior in human melanoma cells.* J Invest Dermatol, 2005. 124(6): p. 1293-9.

141. Ruschel, A. and A. Ullrich, *Protein tyrosine kinase Syk modulates EGFR signalling in human mammary epithelial cells.* Cell Signal, 2004. 16(11): p. 1249-61.

142. Yamasaki, F., et al., *Sensitivity of breast cancer cells to erlotinib depends on cyclin-dependent kinase 2 activity.* Mol Cancer Ther, 2007. 6(8): p. 2168-77.

143. Kniazev Iu, P., et al., *[Gene expression profiles of protein kinases and phosphatases obtained by hybridization with cDNA arrays: molecular portrait of human prostate carcinoma]*. Mol Biol (Mosk), 2003. 37(1): p. 97-111.

144. Wang, Y., et al., *Survey of differentially methylated promoters in prostate cancer cell lines.* Neoplasia, 2005. 7(8): p. 748-60.

145. Reddy, J., et al., *Differential methylation of genes that regulate cytokine signaling in lymphoid and hematopoietic tumors.* Oncogene, 2005. 24(4): p. 732-6.

146. Lapointe, J., et al., *Gene expression profiling identifies clinically relevant subtypes of prostate cancer.* Proc Natl Acad Sci U S A, 2004. 101(3): p. 811-6.

147. Boer, J.M., et al., *Identification and classification of differentially expressed genes in renal cell carcinoma by expression profiling on a global human 31,500-element cDNA array.* Genome Res, 2001. 11(11): p. 1861-70.

148. Zhan, F., et al., *Global gene expression profiling of multiple myeloma, monoclonal gammopathy of undetermined significance, and normal bone marrow plasma cells.* Blood, 2002. 99(5): p. 1745-57.

149. Wolter, F., et al., *Piceatannol, a natural analog of resveratrol, inhibits progression through the S phase of the cell cycle in colorectal cancer cell lines.* J Nutr, 2002. 132(2): p. 298-302.

150. Yuan, Y., et al., *Reactivation of SYK expression by inhibition of DNA methylation suppresses breast cancer cell invasiveness.* Int J Cancer, 2005. 113(4): p. 654-9.

151. Fabian, M.A., et al., *A small molecule-kinase interaction map for clinical kinase inhibitors.* Nat Biotechnol, 2005. 23(3): p. 329-36.

152. Carter, T.A., et al., *Inhibition of drug-resistant mutants of ABL, KIT, and EGF receptor kinases.* Proc Natl Acad Sci U S A, 2005. 102(31): p. 11011-6.

153. Luangdilok, S., et al., *Syk tyrosine kinase is linked to cell motility and progression in squamous cell carcinomas of the head and neck.* Cancer Res, 2007. 67(16): p. 7907-16.

154. Ebert, B.L., et al., *Identification of RPS14 as a 5q- syndrome gene by RNA interference screen.* Nature, 2008. 451(7176): p. 335-9.

155. Mullighan, C.G., et al., *Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia.* Nature, 2007. 446(7137): p. 758-64.

156. Gutova, M., et al., *Identification of uPAR-positive chemoresistant cells in small cell lung cancer.* PLoS ONE, 2007. 2(2): p. e243.

157. Alfano, D., I. Iaccarino, and M.P. Stoppelli, *Urokinase signaling through its receptor protects against anoikis by increasing BCL-xL expression levels.* J Biol Chem, 2006. 281(26): p. 17758-67.

158. de Angelis, P.M., et al., *Molecular characterizations of derivatives of HCT116 colorectal cancer cells that are resistant to the chemotherapeutic agent 5-fluorouracil.* Int J Oncol, 2004. 24(5): p. 1279-88.

159. Jager, R., et al., *Serum levels of the angiogenic factor pleiotrophin in relation to disease stage in lung cancer patients.* Br J Cancer, 2002. 86(6): p. 858-63.

160. Rao, J.S., et al., *Inhibition of invasion, angiogenesis, tumor growth, and metastasis by adenovirus-mediated transfer of antisense uPAR and MMP-9 in non-small cell lung cancer cells.* Mol Cancer Ther, 2005. 4(9): p. 1399-408.

161. Xie, D., et al., *Levels of expression of CYR61 and CTGF are prognostic for tumor progression and survival of individuals with gliomas.* Clin Cancer Res, 2004. 10(6): p. 2072-81.

162. Cui, Y., et al., *Elevated expression of mitogen-activated protein kinase phosphatase 3 in breast tumors: a mechanism of tamoxifen resistance.* Cancer Res, 2006. 66(11): p. 5950-9.

163. Zhang, Y.Q., et al., *Rosiglitazone enhances fluorouracil-induced apoptosis of HT-29 cells by activating peroxisome proliferator-activated receptor gamma.* World J Gastroenterol, 2007. 13(10): p. 1534-40.

164. Wang, Q., et al., *Augmentation of sodium butyrate-induced apoptosis by phosphatidylinositol 3'-kinase inhibition in the KM20 human colon cancer cell line.* Clin Cancer Res, 2002. 8(6): p. 1940-7.

165. Boyer, J., et al., *Pharmacogenomic identification of novel determinants of response to chemotherapy in colon cancer.* Cancer Res, 2006. 66(5): p. 2765-77.