UNIVERSITY OF CALIFORNIA
SANTA CRUZ

Inferring Genetic Co-dependencies To Identify New Vulnerabilities In Cancer

A dissertation submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOMOLECULAR ENGINEERING AND BIOINFORMATICS

by

Thomas J. Matthew

September 2019

The Dissertation of Thomas J. Matthew
is approved:

Professor Joshua Stuart, Ph.D., Chair

Professor Phillip Berman, Ph.D.

Professor Manuel Ares, Ph.D.

Quentin Williams
Vice Provost and Dean of Graduate Studies

Table of Contents

Figure List

Figure List, cont.

List of Tables

Abstract

Inferring Genetic Co-dependencies To Identify New Vulnerabilities In Cancer

Thomas J Matthew

Translation of cancer genomic data into cancer therapies and companion diagnostics remains a primary challenge in personalized medicine. Much of this challenge is due to the difficulty of identifying genetic co-dependencies that lead to clinically actionable drug targets. Targeting many of the known essential gene products are not always selectively efficacious because these targets may be common to both malignant and benign cells. However, essential genes that are associated with particular genomic alterations in cancer cells, like those from synthetic lethality, can potentially provide a source of tumor-specific drug targets.

To help aid novel drug discovery, I developed a computational approach called CLOvE, a multi-omic approach that identifies co-dependencies in pairs of genes. These co-dependencies are inferred from context-dependent changes in expression, where CLOvE assigns high scores to those genes with the greatest compensatory change in expression. These scores may suggest synthetic lethal interactions, which may uncover clinically actionable essential genes. These methods were developed in CCLE cell lines and validated with RNAi and CRISPR viability data. CLOvE identifies meaningful expression changes, assigns high scores

to known essentials, reveals known synthetic lethal connections, and implicates many possible new connections. This approach could provide a tool to accelerate both target discovery and biomarker discovery, to develop drugs suitable for a specific cancer, and identify and stratify patients who may benefit from these treatments.

To Grace.
May you live in a world without cancer.

Acknowledgements

This work would not be possible without the support and guidance I received from my PI Dr. Josh Stuart. He inspired many of the ideas here, and through many hours of discussion (with a great deal of patience), helped me to chart the course for the research presented here. I also want to thank my committee members Dr. Manuel Ares and Dr. Phillip Berman for committing their time and expertise to evaluate my work and encourage my progress throughout this program. Thank you to the members of the Stuart Lab who, through many lab meetings, provided expert advice and posed important questions that helped to shape this work.

CHAPTER 1

Motivation: Beginning in Thymoma

Summary

- Served as an analyst for the Thymoma Analysis Working Group as part of The Cancer Genome Atlas (TCGA) consortium, June 2015 to October 2017
- Used PARADIGM and TumorMap to investigate and define the molecular subtypes of Thymic Epithelial Tumors (TETs)
- Co-Authored the largest study of TETs to date (Radovich et al. 2018)

  "The Integrated Genomic Landscape of Thymic Epithelial Tumors"
- Identified molecular basis for aggressiveness in Thymic Carcinomas
- Used an integrative approach to find similarities between molecular tumor state and World Health Organization subtypes of TETs
- Informed my interest in network biology and differential analysis

Introduction

Thymoma is the most common form of thymic epithelial tumors, or TETs. TETs are the most rare form of malignancies, affecting almost exclusively adults (with an average of 70 years of age) TETs have an incidence of just 0.15 cases in 100,000 person-years in the United States (Engels 2010). The standard care for TETs is surgical resection of the thymus. Prognosis for most TETs is relatively good. In advanced cases of TETs, the median 5-year conditional survival is 69%, but patients have an 81% 5-year survival if they live to 10 years post-diagnosis (E. Kim and Thomas 2015; Scorsetti et al. 2016). Despite these encouraging survival rates, the International Thymic Malignancy Group Interest Group has maintained that all TETs have malignant potential, and should not be considered benign (Travis et al. 2015). Common malignancy routes are shown in Figure 1. More malignant cases were shown to cause a 3-fold increase in the risk of metastasis due to some pre-existing and unknown etiologic factor (Engels and Pfeiffer 2003; Souadjian, Silverstein, and Titus 1968). What is known, however, is the clear link between TET and autoimmune diseases.

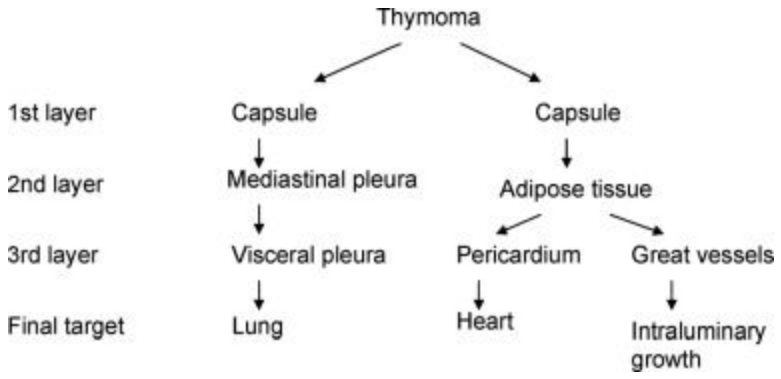

Figure 1. Process of Invasion of Thymoma. Reproduced from "Staging system of thymoma" (Masaoka 2010).

TETs are neoplasms of the anterior mediastinum, which is centered in the upper thorax, posterior to the sternum, anterior to the heart, between the thoracic inlet and the diaphragm (Wychulis et al. 1971). The anterior mediastinum contains predominantly fat, lymph nodes, and the thymus gland (Strollo, Rosado de Christenson, and Jett 1997). The thymus is a component of the immune system where its primary function is to host the maturation of T-cells, which function to determine and control the immune response (J. F. Miller 1961). Due to the central role of the thymus in the immune system, TETs may often be associated with autoimmune diseases, in particular thymoma-associated myasthenia gravis (TAMG) (Castleman 1966). TAMG is an autoimmune syndrome caused by inappropriate binding of antibodies to signaling proteins in the neuromuscular junctions, leading to failure of neuromuscular transmission (Conti-Fine, Milani, and Kaminski 2006). This leads to weakness and ultimately failure of control in the ocular, bulbar, and limb

muscles, controlling sight, speech, and movement, respectively. Risk of both TAMG and degree of malignancy is described in the classification of TETs.

TETs are classified into both stages of advancement and histological subgroups. The Masaoka system refined earlier attempts by Bergh and Wilkins and Castleman to stage thymomas, using more surgically-relevant criteria to accurately distinguish pleural and pericardial invasions (Bergh et al. 1978; Wilkins and Castleman 1979). The Masaoka-Koga system stages TETs from one (I) to four (IV), in increasing order of metastasis, summarized in TABLE (Detterbeck et al. 2011; Masaoka et al. 1981).

Stage I	Macroscopically completely encapsulated and microscopically no capsular invasion
Stage II	1. Macroscopic invasion into surrounding fatty tissue or mediastinal pleura, or 2. Microscopic invasion into capsule
Stage III	Macroscopic invasion into neighboring organs, i.e., pericardium, great vessels, or lung
Stage Iva	Pleural or pericardial dissemination
Stage IVb	Lymphogenous or hematogenous metastasis

Table 1. Criteria of Clinical Stages. Adapted from Masaoka et al. (Masaoka et al. 1981).

Besides staging, TETs can be separated into two main groups, thymic carcinoma (TC) and thymoma, which is the largest of the two (Engels 2010). Further classification of TETs into subgroups is defined by histology according to the World Health Organization (WHO) histological classification system developed for thymic epithelium (Okumura et al. 2002). Thymoma is separated into those with a spindle

or ovoid shape (type A), dendritic or “plump” shape resembling epithelium (type B), further separated into B1, B2, and B3 that reflect increasing epithelioid appearance and degree of lymphocyte infiltration, and finally thymic carcinoma (type C) (Okumura et al. 2002). Part of the TCGA investigation was to establish some genomic underpinnings to these different subclasses of thymoma across multiple molecular platforms

A	A tumor comprising a homogenous population of neoplastic epithelial cells with spindle/oval shape, lacking nuclear atypia, and accompanied by few or no non-neoplastic lymphocytes
AB	A tumor in which foci with the features of type A thymoma are admixed with foci rich in lymphocytes: the segregation of two patterns can be sharp or indistinct
B1	A tumor that resembles the normal functional thymus in that it combines large expanses with an appearance practically indistinguishable from that of normal thymic cortex with areas resembling thymic medulla
B2	A tumor in which the neoplastic epithelial component appears as scattered plump cells with vesicular nuclei and distinct nucleoli among a heavy population of lymphocytes; perivascular spaces are common
B3	A tumor comprising predominantly epithelial cells with a round or polygonal shape and exhibiting mild atypia admixed with a minor component of lymphocytes; foci of squamous metaplasia and perivascular spaces are common

Table 2. Definitions of WHO Classification of Thymic Epithelial Tumors. Adapted from Okumura et al. (Okumura et al. 2002).

Previous attempts to define molecular classes of TETs focus on separate platforms and do not take an integrative approach of multiple platforms. In 2002, using an analysis of microarrays, a positive correlation was identified between c-JUN, a component of the proto-oncogenic transcription factor AP-1, and Masaoka stage in thymoma (H. Sasaki et al. 2002). Another group extended this study to include a significance test of an additional 50 genes (Lee et al. 2007). Later in 2009, another effort investigated the molecular profiles of thymoma and thymic carcinoma, demonstrating TCs were subject to increased copy aberrations (both amplifications and deletions) and somatic mutations in KIT, a receptor tyrosine kinase and proto-oncogene (Girard et al. 2009). Associations between distinct clusters of RNA expression and histologic classification, as well as pathway enrichment analysis that revealed an expression signature associated with recurrence and metastasis (Badve et al. 2012). This pathway analysis led to the development DecisionDx-Thymoma, a nine-gene expression profile test used to predict metastasis risk in thymoma patients (Gökmen-Polar et al. 2013). Exome sequencing performed across 28 patients revealed some high-frequency somatic mutations in type A and AB, which include GTF2I, general transcription factor II (Petrini et al. 2014). Overexpression of microRNA (miRNA) cluster on chromosome 19 associated with PI3K/AKT activation was also identified in type A and type AB TETs (Radovich et al. 2016). Many other attempts to characterize thymoma led to the identification of genes and gene signatures pertaining to anti-apoptosis, histone modification, DNA methylation, and chromatin remodelling (Bellissimo et al. 2017; Huang et al. 2013; Petrini et al. 2014; Y. Wang et al. 2014).

Despite the body of work to analyze TETs, there have been few instances of more comprehensive analysis. This may be to blame for lack of clinical success in targeted treatments for TETs (Chen, Gharwan, and Thomas 2014). To this end, the goal of the TCGA-THYM was to present a deeply comprehensive, integrative, multi-platform approach to uncover the genomic landscape of these tumors.

Data

Clinical and pathological characteristics for 117 patient samples was collected and evaluated according to WHO criteria. A total of 105 patients had thymoma, 10, 48, 12, 25, and 10 having type A, AB, B1, B2, and B3 respectively. Of the remaining patients, 10 had TC and 2 had micronodular thymoma (MNT). 32 patients (27%) had TAMG, with another 7 (6%) having some other form of autoimmune disease, diagnosed at the median age of 52 and 62 years, respectively. Histological subtypes A and B were associated with TAMG ($p=0.00015$), but not with with TC ($p=0.0015$). Upon a media follow-up time of 38.3 month, 10 patients had recurrent malignancies, and 8 patients died. Recurring malignancies presented as pleural involvement in 80% of those cases. Higher rates of progression free survival was most strongly associated with early Masaoka stage ($p=0.000375$), but also with lower T stage and non-Hispanic ethnicity. Overall survival improved most with

types A and B1-2 (p=0.008), but also younger age of diagnosis and higher tumor lymphocyte component.

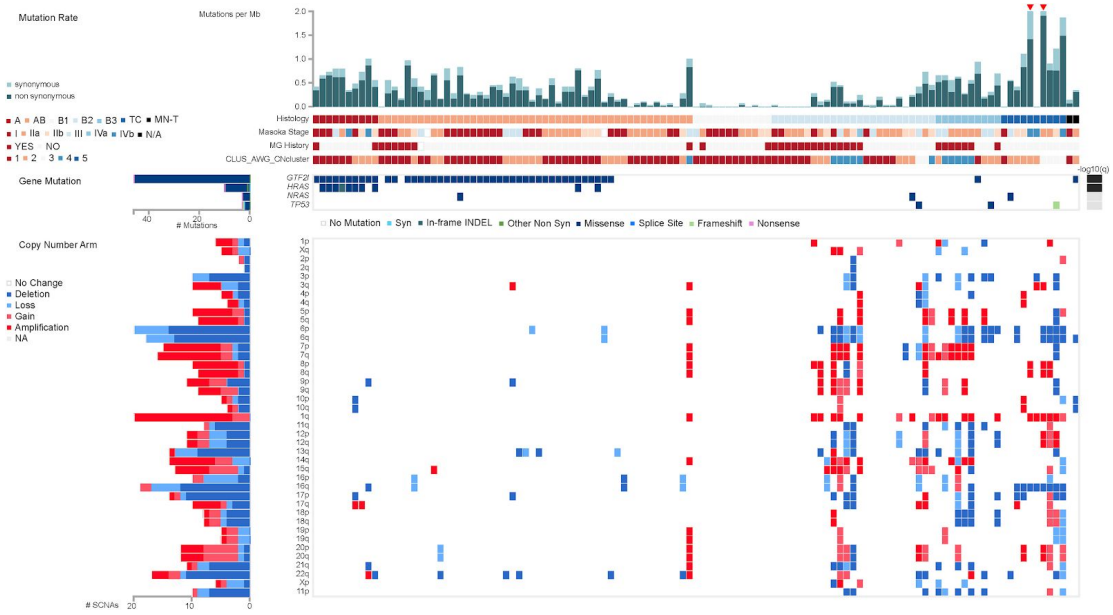

Figure 2. Landscape of DNA Mutation in TETs. Reproduced from "The Integrated Genomic Landscape of Thymic Epithelial Tumors" (Radovich et al. 2018).

First, whole exome sequencing was performed and subsequently analyzed with MutSig2CV to identify any recurrent somatic mutations (Lawrence et al. 2013). HRAS, TP53, NRAS, and GTF2I were the top four most significantly mutated genes, with GTF2I having the highest mutational frequency (39%) in type A and AB thymomas (Figure 2). This analysis also revealed that TETs have the lowest mutational burden of any adult cancer of the 21 other cancers profiled by TCGA, besides the childhood cancers of rhabdomyosarcoma (RHAB) and medulloblastoma (MED). Somatic copy number alterations (sCNAs) were observed only in the minority tumors with few recurrent mutations (in B2, B3, and TC), and were rarely

seen in all other majority of other samples. RNAseq was also performed, which revealed no fusion events, either viral or bacterial components.

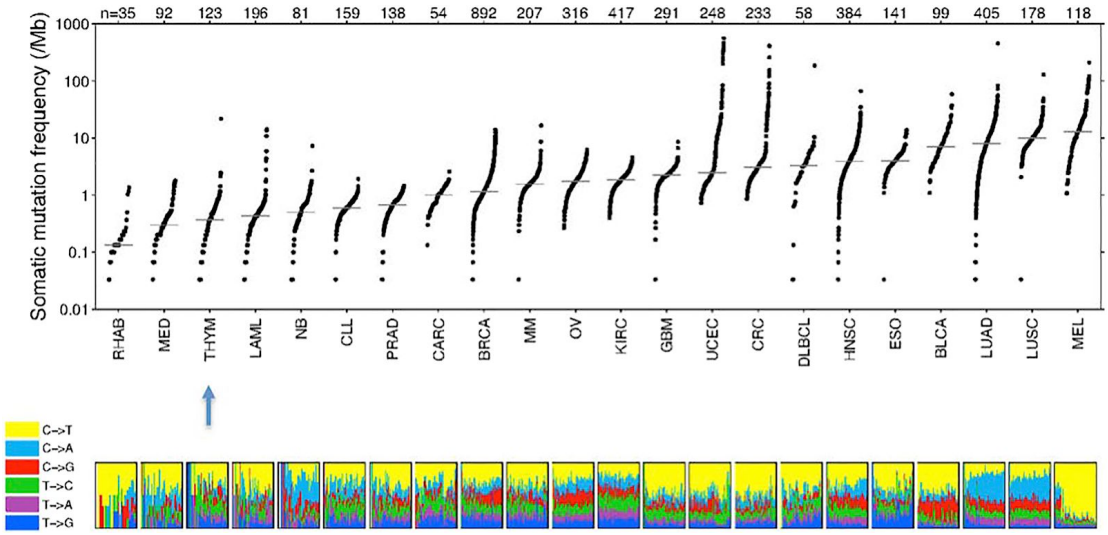

Figure 3. Tumor Mutational Burden of THYM compared to the rest of TCGA samples. Reproduced from "The Integrated Genomic Landscape of Thymic Epithelial Tumors" (Radovich et al. 2018).

Given the goal of investigating thymoma with a multi-platform approach, gathering data from all possible platforms for these 117 patients was the utmost importance. Data from platforms including sCNV, mRNA (messenger RNA), miRNA, DNA methylation, and reverse phase protein array (RPPA) were collected. Pavana Anur, Katherine Hoadley, and myself were responsible for an integrated analysis of these platforms. In particular, I was responsible for analytically and visually representing the key molecular differences among the TET subtypes using two tools developed

in the Stuart Lab of UC Santa Cruz: PARADIGM and TumorMap. The culmination of this effort would ultimately become Figure 3 of the paper.

Methods

Like other previous efforts to define subtypes, subtyping TETs was difficult. The complexity and subtlety of histological structure, as well as the subjective nature of such exercises, leads to inconsistencies in the field. Using an integrative approach was necessary to understand more objectively molecular differences between these subtypes.

To this end, we adopted a cluster-of-cluster assignments approach described previously by the TCGA (Cancer Genome Atlas Network 2012). Using the centroid of each cluster formed in each particular platform, we calculated a correlation matrix that related each sample to each centroid from each platform. Subsequently, the R package ConsensusClusterPlus_1.24.01 was used to calculate a hierarchical clustering with pearson correlation distance over 1000 iterations and a 90% sampling rate. consensus clustering of samples from this weighted matrix. Consensus clustering allowed integration of the disparate platforms and the ability to identify relatedness between and across subtypes

PARADIGM is a computational model that identifies significantly altered pathways from an integrated analysis of copy number and gene expression of a patient or sample (Vaske et al. 2010). This integrated analysis is performed in the context of pathway entities. These entities comprise biological molecules, small molecules, complexes, or abstract concepts that represent cellular processes such as apoptosis or endothelial cell migration (Schaefer et al. 2009). The PARADIGM graphical model represents such entities as nodes and generates an integrated pathway activity (IPA) for each entity of a patient. A gene IPA score, for example, refers to the final active protein inferred from copy number, expression, and signaling from other genes in the pathway Figure 4. Here we use PARADIGM to generate IPA scores for each of the 117 TET patients.

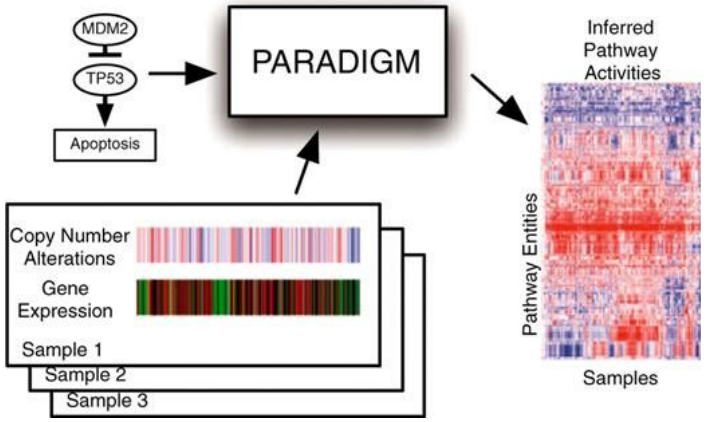

Figure 4. Overview of the Paradigm Method. Reproduced from Vaske et al. (Vaske et al. 2010).

We clustered PARADIGM IPAs using the UCSC consensus clustering RData script v.1.0.0 available though medbook.ucsc.edu. These PARADIGM inputs were merged real IPAs to be clustered by samples (patients). The clustering conditions used a k-means algorithm, average final linkage, and 500 repetitions with a k-max of 10. Clustered heatmaps of patient IPAs were graphed with attribute color assignments. As a quality check, we computed a silhouette score for each k to measure goodness of fit for each patient in a cluster. The silhouette score used Euclidian distance to compute both similarity of a patient with other patients in a cluster, and separation of patients in different clusters (Rousseeuw 1987). To perform significance testing on cluster attribute enrichment, we applied Benjamini-Hochberg p value correction (BH FDR) on Fisher Exact p values to compute False Discovery Rate (FDR). This produced a ranked list of clinical attributes based on p value per cluster, annotated with FDR.

We also employed a complementary approach known as TumorMap, which generates a map of samples for interactive exploration, statistical analysis, and data visualization using the Google Maps API (application program interface). Samples are arranged on a hexagonal grid based on similarity: samples with similar genomic profiles are placed near each other in the map, whereas dissimilar samples are placed further away. Clusters of samples that appear as “islands” in the map indicate groups of samples that share genomic features.

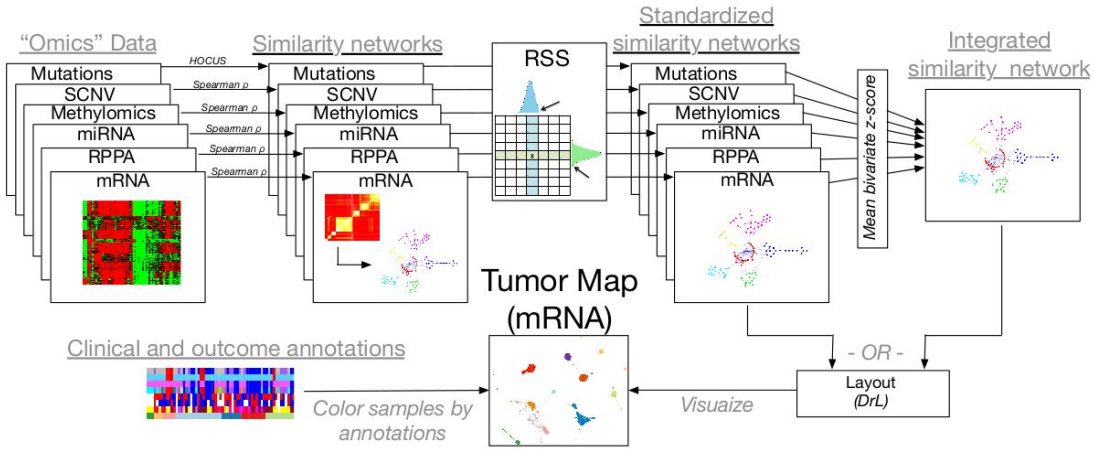

Figure 5. Tumor Map Construction. Reproduced from Yulia Newton (Newton 2016).

TumorMap is a tool that generates a map of cancer samples for interactive exploration, statistical analysis, and data overlay visualization. These visualizations, which employ the Google Maps API, arrange samples on a hexagonal 2-dimensional grid based on a sample-by-sample similarity matrix. Samples can be annotated according to different attributes to allow the user to explore new associations in clinical data. Maps can be from a single platform or multiple merged platforms. This later approach uses Bivariate Standardization similarity space Transformation (BST), adapted from Faith et al's CLR, to integrate multiple similarity matrices into a single matrix of sample-sample associations (Faith et al. 2007). Here we use an integrative approach to reveal the relationship between the molecular and clinical attributes of TET patients based on a multi-platform co-cluster analysis.

We first created a sparse sample-by-sample similarity matrix from each of the platform clusters. This similarity matrix comprises the top 10 highest ranked

Spearman correlations per sample as implemented by the sklearn.metrics.pairwise submodule (Pedregosa et al. 2011). We performed this rank on each of the platform clusters, and combined them with the BST pipeline. The BST pipeline averages Z scores of each sample-sample similarity between platform matrices, resulting in a single similarity matrix. This single matrix was inputted to the physics-based layout engine DrL, an open source version of VxOrd created by Sandia National Labs (Martin et al. 2011). DrL treats sample similarities as spring constants and searches for a spatial configuration among samples that minimizes system tension. This ultimate spatial configuration was mapped in 2D using TumorMap v0.5, using an interactive session on medbook.ucsc.edu, and colored samples by clinical attribute.

Results

PARADIGM pathway analysis, TumorMap clustering, and the complementary COCA approach yielded several key associations between histological subtype and molecular landscape. COCA revealed 4 cluster groups that associated with WHO subtypes ($p<0.0005$): group 1 with type B, group 2 with TC, group 3 with type AB, and group 4 with a mix of both A and AB, as shown in FIGURE-INTEGRATIVE. The results of the TumorMap analysis agreed with the results of COCA, revealing four distinct molecular clusters that were highly correlated with WHO histological subtype and COCA classifications.

We performed single-platform analyses as well as multi-platform PARADIGM analysis, employing copy number and RNA expression. PARADIGM allowed is to identify unique pathways and genomic hallmarks of each cluster, and overlaid these onto the TumorMap to differentiate the clusters FIGURE-INTEGRATIVE-D. Single-platform analyses demonstrated that type A and AB tumors are characterized by GTF2I mutations and overexpression of a large miRNA cluster on chromosome 19q13.4, while type C tumors are characterized by loss of chromosome 16q. Examination of the PARADIGM findings revealed upregulation of tumor suppression (p53) and downregulation of oncogenes (MYC/Max, MYB, and FOXM1) in the A-like cluster (Figure 3D). The opposite is seen in the AB-, B-, and C-like clusters where tumor suppression is downregulated (p53, and TAp73a), and oncogenes are upregulated (MYC/Max, MYB, FOXM1, and E2F1) (Figure 3D). These results agree with the known increased clinical aggressiveness observed in type B and TC TETs and their overall poorer overall survival compared to other subtypes.

Further analysis characterized tumors with GTF2I mutations, which is summarized in Figure 6. These mutations were found to be exclusively mutated at the codon L424H in type A and 70% of type B tumors. GTF2I L424H mutations were not observed in more than 10,000 TCGA tumor samples. Analysis of both RPPA and expression, GTF2I mutants were found to be strongly associated with pathways pertaining to cell morphogenesis, tyrosine kinase signaling, and WNT signalling pathways. These agree with previous observations from a knockout experiment in mice of the family member GTF2IRD1 (Corley et al. 2016). Conversely, GTF2I

mutants saw less pathway activity in terms of apoptosis, cell cycle, DNA damage, androgen signalling, RAS and mTOR signalling. GTF2I mutants also did not have a strong clinical associating with MG.

Autoimmunity is strongly associated with TETs and no other TCGA cancer type, with MG occuring in 30% of TET cases (Zekeridou, McKeon, and Lennon 2016). Given the significant link between thymomas and autoimmunity, the group focused on the integrated molecular features of TAMG. Figures relevant to TAMG related analyses are shown in Figure 7. As expected from previous observations, reported cases of MG were higher with type B thymoma, and in cases of aneuploidy (Cufi et al. 2014; Zettl et al. 2000). The link between aneuploidy and TAMG has been thought to be just due to its association with type B. However, the association depends more on the aneuploidy than developing type B thymoma, as MG was detected before thymoma. TAMG samples also had increased transcript levels of previously described auto-antigens that react with AChR, titn, and RYR receptors (Dalmau et al. 1992). These include CHRNA1 (3.0 fold increase), and RYR1/RYR2/RYR3 (5.5 fold increase), and NEF, a midsize filament gene, which contains sequences that express epitopes for AChR and titin (Marx et al. 1992; Masuda et al. 2012; Mygland et al. 1994; Schultz et al. 1999).

Thymic carcinomas were the last single group that was analyzed. Though they represent less than 10% TET cases, they lead to a more difficult course of treatment and poorer prognoses on average. Therefore, it was necessary to go beyond the

traditional histological definition, which notes that TC appears more similar to epithelial tumors, and explore their molecular landscape. These findings are summarized on Figure 8. An analysis of sCNA in TC showed a chromosomal arm loss at chromosome 16q in 8 of the 10 samples, in line with previous sCNA investigations of TC (Zettl et al. 2000). Chromosome 16q is known to harbor tumor suppressors including, CYLD, CBFB, CDH1, CDH11, CTCF, and ZFHX3, which may explain its more aggressive rates of metastasis. In addition, TC also had the highest TMB of all tested TET subtypes, 19% enrichment of indels vs 5% for all other tested TETs, and microsatellite instability leading to loss of the DNA repair gene MLH1.

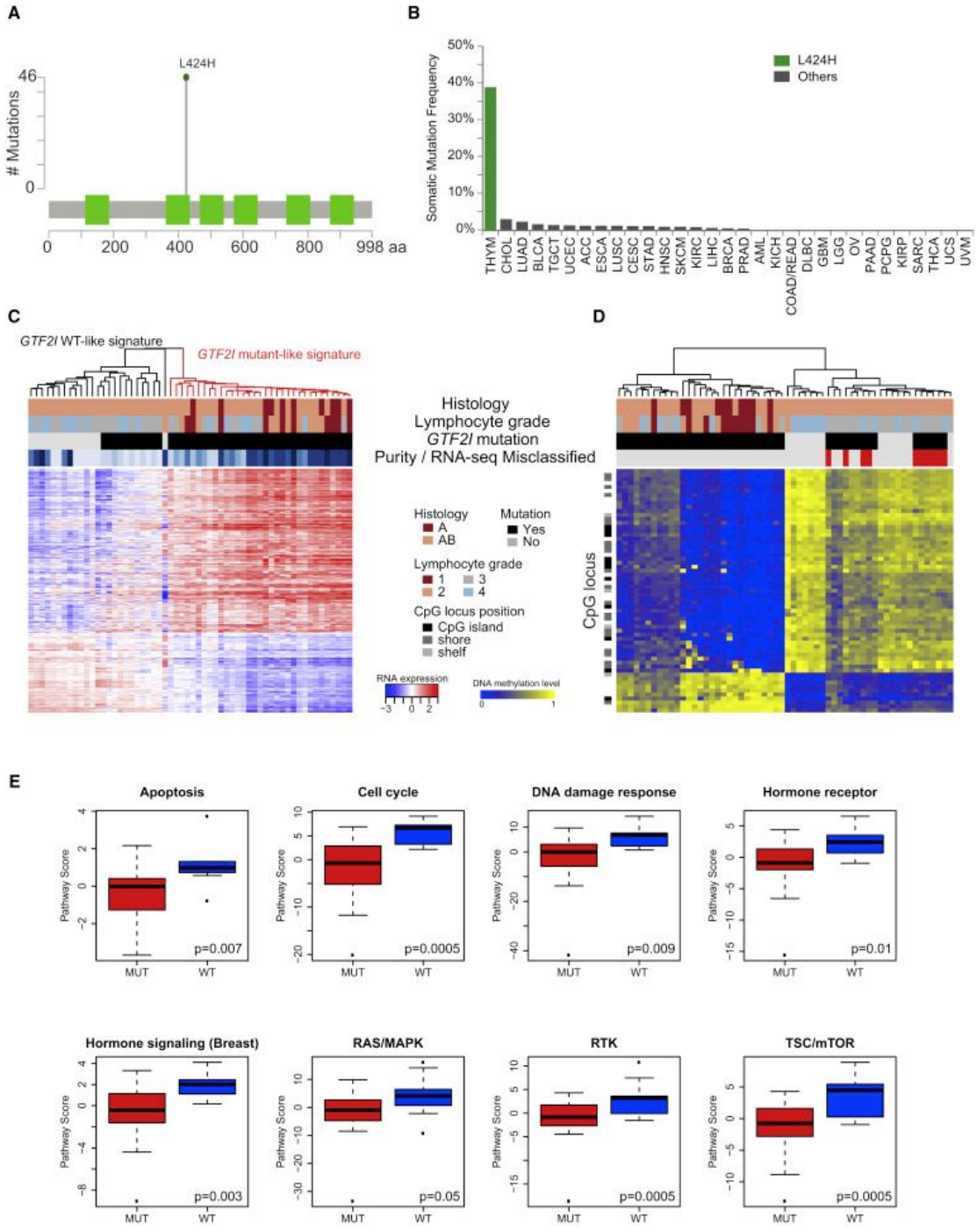

Figure 6. Integrative Unsupervised Clustering of Subtypes from Five Data Platforms. Reproduced from "The Integrated Genomic Landscape of Thymic Epithelial Tumors" (Radovich et al. 2018).

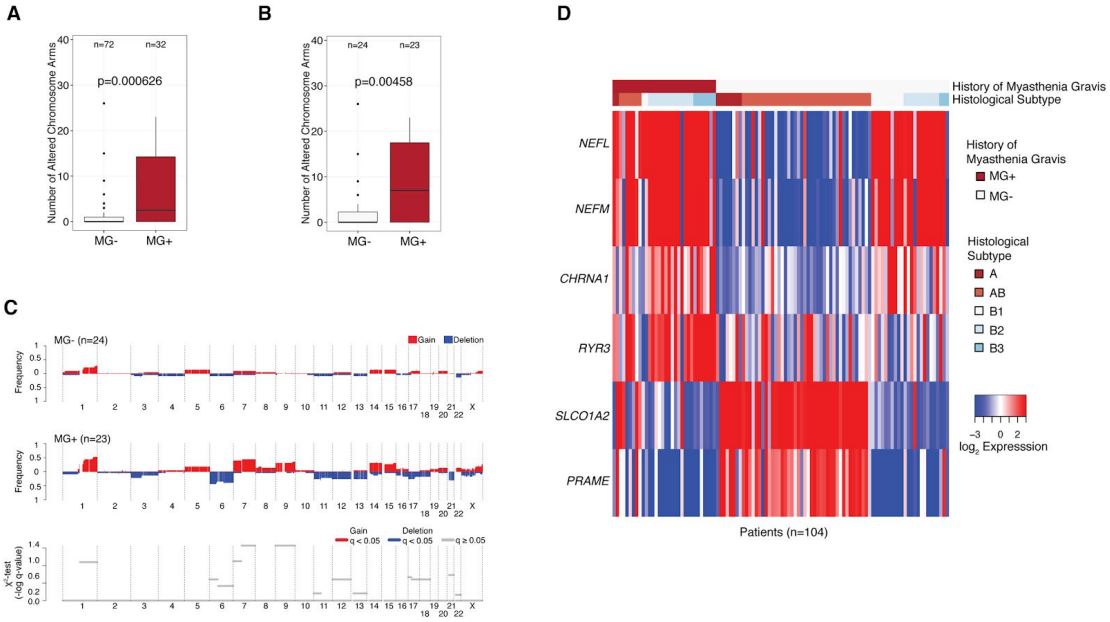

Figure 7. Patterns of sCNA and gene expression Associated with Autoimmunity. Reproduced from "The Integrated Genomic Landscape of Thymic Epithelial Tumors" (Radovich et al. 2018).

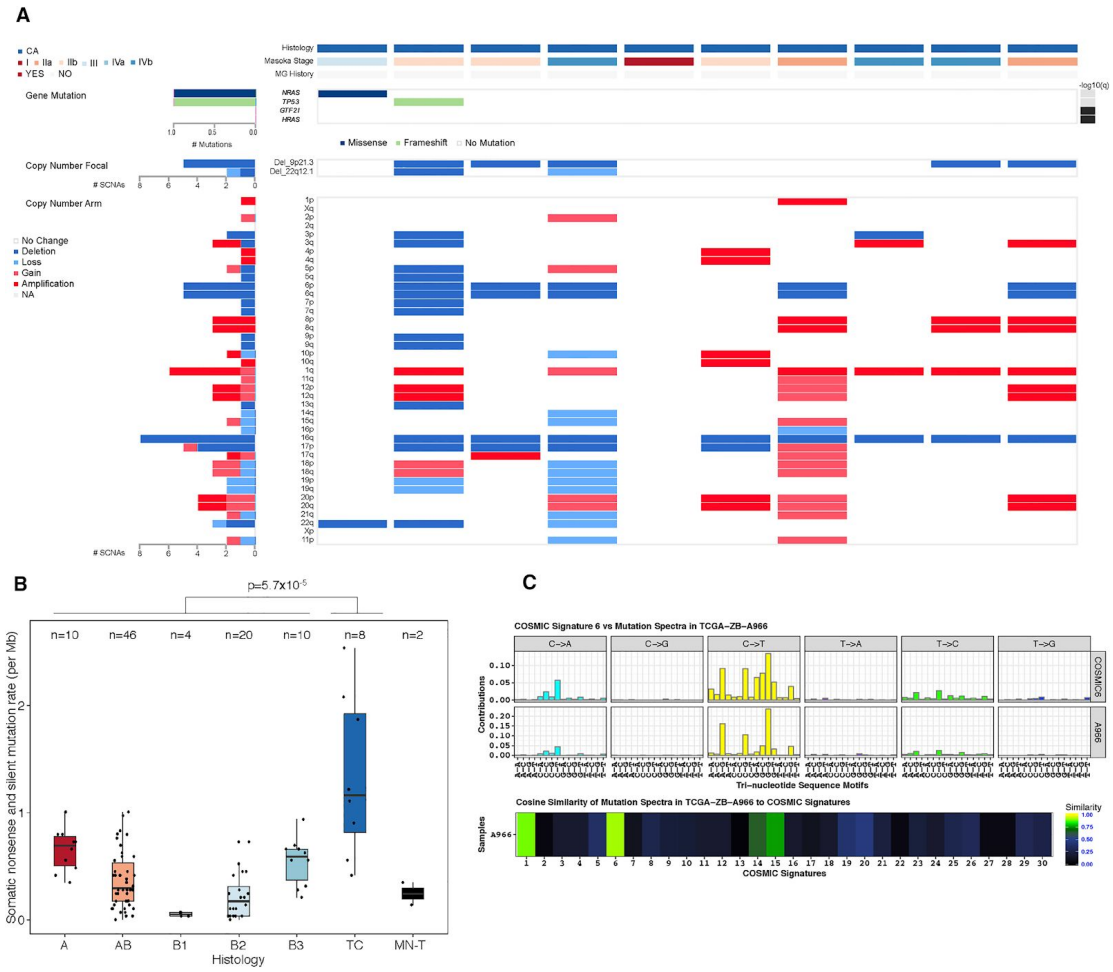

Figure 8. Genomic Analysis of Thymic Carcinomas. Reproduced from "The Integrated Genomic Landscape of Thymic Epithelial Tumors" (Radovich et al. 2018).

Conclusion

TETs are a rare and heterogeneous form of cancer. In this investigation, TET subtypes, previously defined by histology, have now been defined based on the integrated molecular landscape of thymoma. This effort represents the largest study to date on cancers of the thymus, comprising 5 data platforms on 117 patient samples. Despite inconsistencies in histological studies, we have uncovered part of the molecular basis that defines these observed subtypes. In the area of TAMG, we found expression of certain self-reacting antigens can now be added to the hallmarks of autoimmunity in TETs. We also further showed the very specific mutations of GTF2I that are unique to thymoma that may lead to future advancements in the diagnosis and treatment of TETs.

CHAPTER 2

Background of Synthetic Lethality: From Bench, to CPU, to Clinic

Summary

- Cancer represents an unmet medical need worldwide, necessitating new understanding and approaches to targeted therapy
- Cancer is inherently chaotic, and thus inherently vulnerable with the concept of Synthetic Lethality (SL)
- SL was first screened for at the bench, later inferred computationally, and now being introduced to the clinic
- Inferring more SL gene pairs is a worthwhile approach in identifying new therapeutic targets in cancer

Unmet Need

The world is becoming a safer place, and yet cancer is becoming more prevalent. In 2018, cancer was responsible for 1 out of every 6 deaths worldwide, claiming 9.6 millions lives (Bray et al. 2018). Socioeconomic factors at play contribute to this paradox. Whether it is attributed to rising income, increased access to education, favorable fiscal policy, improved human rights conditions, or dozens of other factors, the human population is getting older. By 2100, people over the age of 60 years will comprise 33% of the worldwide population, increased from just 10% in the year 2000 (Lutz, Sanderson, and Scherbov 2008; Desa 2015). As the average age increases, so does the prevalence of cancer: 18.1 million new cases in 2018 is projected to grow 24 million cases in 2035 (Bray et al. 2018; Ferlay et al. 2015; Pilleron et al. 2019). Cancer is both a public health concern, as well as a serious economic concern.

It is estimated that in 2014, cancer had a total economic impact of US$1.16 trillion (Stewart, Wild, and Others 2014). The most direct cost can be seen in the cost of medicinal therapies to treat cancer, with over $150 billion spent in 2018, according to the IQVIA Global Oncology Trends Report (IQVIA 2019). These medicines play a major role in the management of cancer, and as cancer becomes more prevalent worldwide, these medicines will fall into greater demand. To reach a “grand

convergence" in health, new treatments must be developed to meet this demand, but not with a one-size-fits-all approach (Boyle et al. 2015; Jamison et al. 2013).

Personalized medicine is the practice of assigning clinical intervention based on the specific needs of a patient and their disease. Though standards of care are becoming more patient-specific, current diagnosis and treatment campaigns adopt a more one-size-fits all approach along general stratum of cancer types defined by histology (Sudhakar 2009).

Cancer patients, or those who may develop cancer, stand to reap the greatest benefit from personalized medicine since generally-defined tumor strata do not reflect the molecular heterogeneity of cancer. This rough understanding of cancer classification was reflected in first-generation chemotherapeutics. These past oncology interventions hinged on potent but nonspecific therapies that were generally cytotoxic and known for their radical side-effects. Drugs like generally-acting DNA-damaging drugs, topoisomerase inhibitors, and mitotic inhibitors have both narrow therapeutic windows and high therapeutic indices, defined by a narrow dose range and a low lethal-to-effective dose ratio, respectively (Kaelin 2005). While effective in destroying tumor cells, non-specific pharmacology also leads to the destruction of healthy cells, promoting harmful or potentially fatal off-target effects in the patient taking these drugs.

Fortunately, a greater understanding of cancer genomics has informed a new age of targeted drug discovery. Support for the effectiveness of essential gene targeted therapies in oncology is exemplified in the clinical success of HER2 drugs, especially trastuzumab and lapatinib, and in the role of biomarkers on the development of drugs and companion diagnostics (Cheng, Koch, and Wu 2012). However, many other forms of cancer lack such specific treatment and therefore represent an unmet medical need. To address this need, it is crucial to understand the evolutionary basis for tumor heterogeneity and the essential genes associated with poor clinical prognoses.

Evolution & Vulnerability

Cancer is a disease characterized by dynamic perturbations of the genome (Jerby-Arnon et al. 2014). These perturbations can exist in pathways and processes common among both healthy and cancerous cells. However, unlike in cancerous cells, normal cells exhibit lower rates of spontaneous mutations because changes can be adequately detected and addressed by the mechanisms that protect genomic fidelity. Those mechanisms provide a selective pressure on the cancer cells, leading to an increased rate of mutation sufficient for tumorigenesis. It is this rate of mutation and the nature of the surviving mutations that cause the emergence of cancer subclones that come to underlie tumor heterogeneity and further separate

the cancer genome from host genome. To understand the potential drivers of tumorigenesis, it is important to recognize cancer as a genomic disease marked by the evolutionary cycles of replication, mutation, and selection.

More than 6 decades ago, pathological analysis of tumors showed stepwise progression through increasingly malignant states (Greene and Newton 1948). This demonstrated that cancer is a dynamic disease, with greater incidence of genetic instability in tumors attributed to sequential selection of more aggressive forms of cancer (Nowell 1976). The experimental process of transforming human cells to models of cancer demonstrate the multitude of genetic steps required for tumorigenesis: initiation of proliferation, down-regulation of apoptosis, and activation of angiogenesis (Bergers, Hanahan, and Coussens 1998). These iterative and stepwise changes at the genotypic level, conferring advantage at the phenotypic level, support the comparison between cancer evolution and Darwinian evolution (Foulds 1954; D. Hanahan and Weinberg 2000). Seeing common phenotypic trends in cancer evolution allowed researchers to standardize classes of biological pathways unique to cancer.

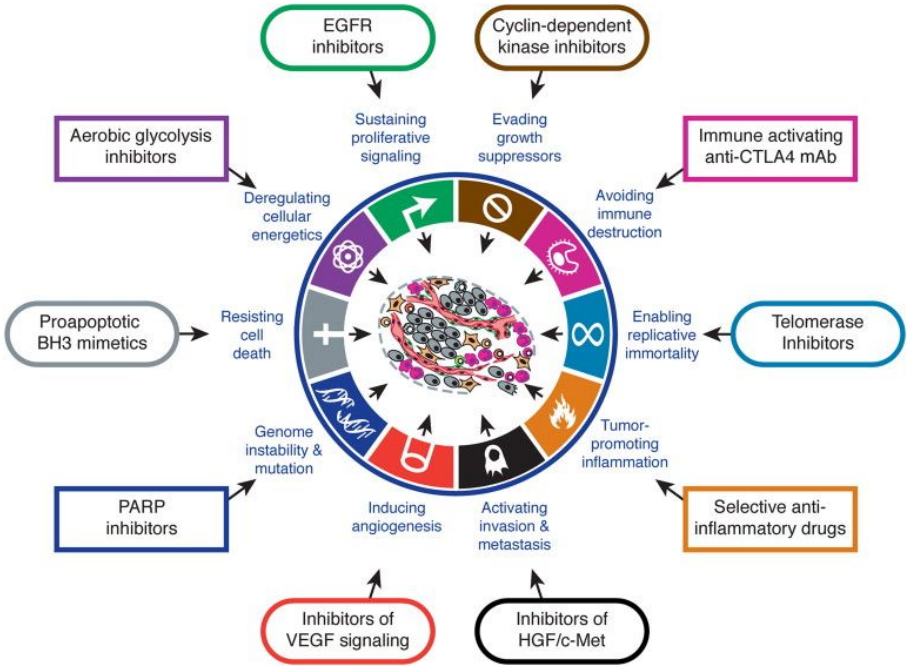

Figure 9. Therapeutic Targeting and the Hallmarks of Cancer. The new hallmarks of cancer are shown with examples illustrating their druggability. Genome instability and mutation (dark blue), shown with the treatment example of PARP inhibitors, is an enabling characteristic that both drives tumorigenesis and represents a therapeutic "Achilles' Heel". Synthetic lethality may be used to identify additional genes essential for tumorigenesis arising from genome instability, as they are with PARP inhibitors. Reproduced from Hanahan and Weinberg (Douglas Hanahan and Weinberg 2011).

Evolutionary pressure within the tumor microenvironment drives cells towards favorable mutations in pathways corresponding to proliferation and metastasis. In particular, these pathways relate to the 6 cancer hallmarks first proposed by Hanahan and Weinberg (D. Hanahan and Weinberg 2000). These initial hallmarks include: self-sufficiency in growth signals, insensitivity to anti-growth signals, tissue invasion and metastasis, limitless replicative potential, sustained angiogenesis, and evasion of apoptosis (D. Hanahan and Weinberg 2000). Though the genetic basis for cancer typically lies in one of these hallmarks, these hallmarks were expanded to

include dysregulation of cellular energetics and avoidance of immune detection (Figure 9). Along with these new additions to the hallmarks, the idea of genome instability is now believed to underlie the premise of the first 6 hallmarks and indeed the concept of the evolving cancer genome (Douglas Hanahan and Weinberg 2011). This concept of instability is both a mechanism of tumorigenesis and potential "Achilles' Heel" in revealing druggable targets and predicting treatment sensitivity. One such mutation type that guides our understanding of cancer is the loss-of-function mutation.

Cancer is a disease characterized by two complementary mutation types: dominant gain-of-function (GoF) and recessive loss-of-function (LoF) mutations (D. Hanahan and Weinberg 2000). These mutations produce phenotypes in cancer that may include any or all of the 6 hallmarks described by Hanahan and Weinberg. GoF are activating mutations that induce expression of some new function, acting as a dominant allele, and produce some new phenotype (Figure 10.e) (Griffiths et al. 2000). A differential phenotype in a GoF mutation, like over-expression of a protein in cancer relative to normal, allows for a target-driven paradigm of drug discovery. This paradigm just focuses on targets present in cancer that are not otherwise present. Genes involved in GoF mutations are also called oncogenes.

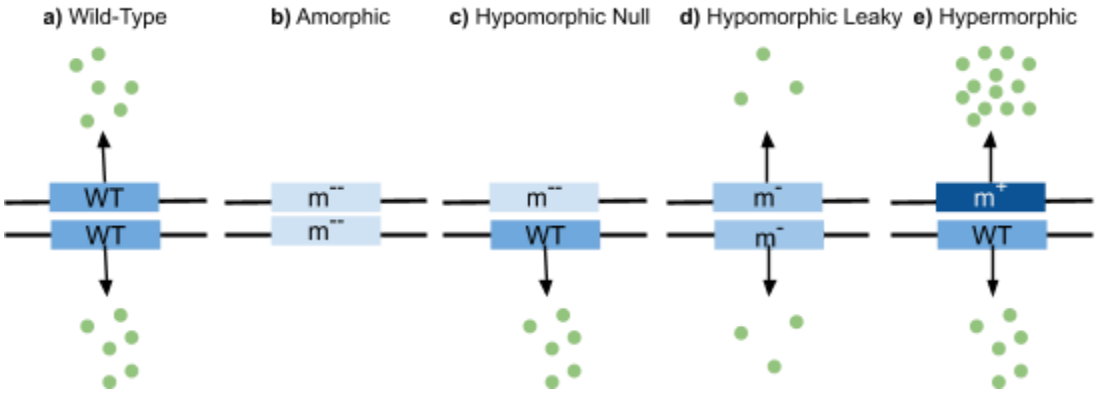

Figure 10. Schematic of various types of alterations. The wild type locus (a) contains no mutations and produces a sufficient amount of gene product. A null or complete loss-of-function (b) occurs in both chromosomal arms that causes total loss of the gene product. Two possibilities for hypomorphic loss-of-function mutations are represented with heterozygous null mutation (c) and a homozygous leaky mutation (d) that results in less total gene product than in the wild type scenario. Haploinsufficiency occurs when a critically low concentration of gene product is produced, as in (c) or (d), which results in an aberrant state. In a hypermorphic state (e) an excess of gene product is produced on one of the chromosome arms, but can also be homozygous. Adapted from Griffiths (Griffiths et al. 2000).

An oncogene results in protein products that induce or maintain tumorigenesis through processes that relate, in one way or another, to the aforementioned cancer hallmarks. Perhaps the most well-characterized GoF targets are in the family of protein tyrosine kinases (PTKs), a broad family of 90 kinases involved in cell-signaling pathways involved in growth, differentiation, adhesion, and apoptosis (Levitzki and Gazit 1995; Robinson, Wu, and Lin 2000). G-coupled protein receptors (GPCRs), a PTK subfamily, are signal transducers often overexpressed on the surface of cancer cells. Their differential overexpression in cancer, and status as surface proteins, make GPCRs an obvious drug target typical of a GoF mutation. In fact, 42 GPCRs are identified to play a role of 14 cancer types (Dorsam and

Gutkind 2007), showing their role in a target-driven paradigm. LoF mutations, by contrast, represent a more intractable target, warranting a systematic effort to find druggable essential genes for patients with oncogenic LoF mutations.

LoF mutations are referred to as inactivating mutations because they attenuate the function of a gene or gene product. A gene with homozygous LoF that results in a complete loss of gene function is known as amorphic (Figure 10.b), while a heterozygous LoF marked by decreased gene activity is known as hypomorphic (Figure 10.c) or leaky (Figure 10.d) (Wilkie 1994) . Even in this hypomorphic state, diploid cells that retain a copy of the wild type (WT) gene can often overcome a LoF mutation and recover a WT phenotype from the remaining WT allele (Griffiths et al. 2000). This particular example heterozygosity characterizes LoF as a recessive mutation, since the mutant phenotype is not expressed over the dominant WT phenotype.

However, some LoF mutations show a differential phenotype, termed haploinsufficiency, which is both a feature of cancer genome instability, and a possible source of synthetic lethality. Haploinsufficiency occurs when a diploid organism has just one functional copy of a gene as a result of a mutation on the other copy (Deutschbauer et al. 2005). In this case, the single functional copy does not produce enough gene product sufficient for a WT phenotype, thus resulting in an aberrant phenotype. In cancer, heterozygous LoF mutations can act on tumor-suppressor genes, reducing their protein products to below functional

threshold, and promoting tumorigenesis. Haploinsufficiency can be a causal factor of cancer, (e.g., BLM, ATM) or cancer driver occurring in pathways pertaining to cell cycle regulation (e.g., Rb), proliferation (e.g., TGFb), DNA damage checkpoints (e.g., p53), apoptosis (e.g., PTEN, COPA), and DNA repair enzymes (e.g., SETD2) (Dittmer et al. 1993; Goss et al. 2002; Spring et al. 2002; Lodish 2008; Sudo et al. 2010; Song, Salmena, and Pandolfi 2012). Haploinsufficiency is an important concept in identifying and targeting vulnerabilities in cancer.

Developing drugs for these "lost" genes with a target-driven paradigm remains a difficult problem in cancer drug discovery. Reasons for this challenge are threefold: 1) absence of a tumor suppressor gene in cancer is the absence of a druggable target, 2) restoring the function of a mutated gene is intractable, and 3) determining the status of a tumor suppressor gene is often ambiguous (Y. Liu et al. 2015; H. Wang et al. 2006). One strategy to overcome the challenges of directly drugging a LoF mutation involve taking pathway-level approach to look beyond the primary mutation for a druggable target. A pathway level approach to drug discovery, often called the context-driven paradigm, requires bioinformatic tools to assess genetic interactions that give cancer cells new and unique druggable vulnerabilities. Development of these tools hinge on biological phenomena that guide the mechanism of modern small-molecule cancer treatments. One therapeutic mechanism is based on the concept of synthetic lethality.

Concept

Synthetic lethality (SL) occurs when simultaneous loss of two genes causes a dead or sick phenotype, while a single loss in either gene does not cause a decrease in viability (Figure 11). You can think of SL in terms of furniture. Consider a table to represent a cell, and each leg of the table to represent a group of four essential genes, either related or unrelated in function. If you remove just one table leg, the table, with some balancing, can manage to stand on its remaining three legs. Those remaining legs are now more important to keeping the table upright. The remaining legs are now essential. Remove one additional leg, or any two legs simultaneously, and the table will topple over and become non-functional.

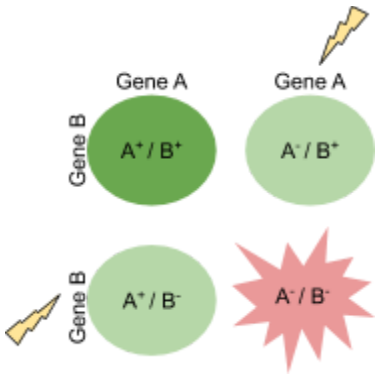

Figure 11. Definition of synthetic lethality. Basic mechanism of synthetic lethality (SL) in regards to a Loss-of-Function (LoF) event. Wild type (WT) cells, those without LoF event (lighting bolt) in either gene A or gene B (dark green), are normal cells. Cells with a heterozygous LoF in either gene A or gene B (light green) may have poorer or unchanged health relative to WT. A homozygous LoF in both gene A and gene B (red) may induce an SL outcome, including a sick or apoptotic phenotypic.

Sometimes a SL relationship is best visualized as a pathway essential to cell survival. In the figure below, SL is represented in a pathway diagram between a normal patient sample, S_N, and a malignant patient sample, S_M (Figure 12). In S_N, the hypothetical sub-pathway pathway p_x originates at source, node 1, and proceeds sequentially through the products of gene A and gene B to the sink, node 2. This preference for sub-pathway pathway p_x is reflected in a normal copy number of genes A and B, and a simultaneously low expression of genes C, D, and E, used in the alternative sub-pathway p_y.

However, in S_M, both gene B copies are lost among genomic perturbations occurring in that particular cancer type, perhaps by arm-level deletion. S_M thus loses the ability to carry out the preferred sub-pathway p_x. This malignant cell would be fated for destruction if it were not for an alternative pathway p_y occurring through genes C, D, and E. S_M can therefore use the redundant sub-pathway to maintain the overall function of this essential pathway.

Using the redundant pathway p_y in the context of a loss of p_x puts an increased importance on the genes C, D, and E. This establishes an acquired vulnerability in S_M of any of these three genes, or gene products, that could be targeted by a drug therapy. Such a drug therapy would cause a loss of redundancy in this essential pathway, and given the loss of the preferred sub-pathway, would trigger preferential death of the malignant cell S_M. However, a loss context could occur in any gene of either sub-pathway to create a potential SL link (in red), and therefore a collection of

essential genes. Though topical in the treatment cancer, researchers observed these SL links long before the establishment of cancer as a genomic disease.

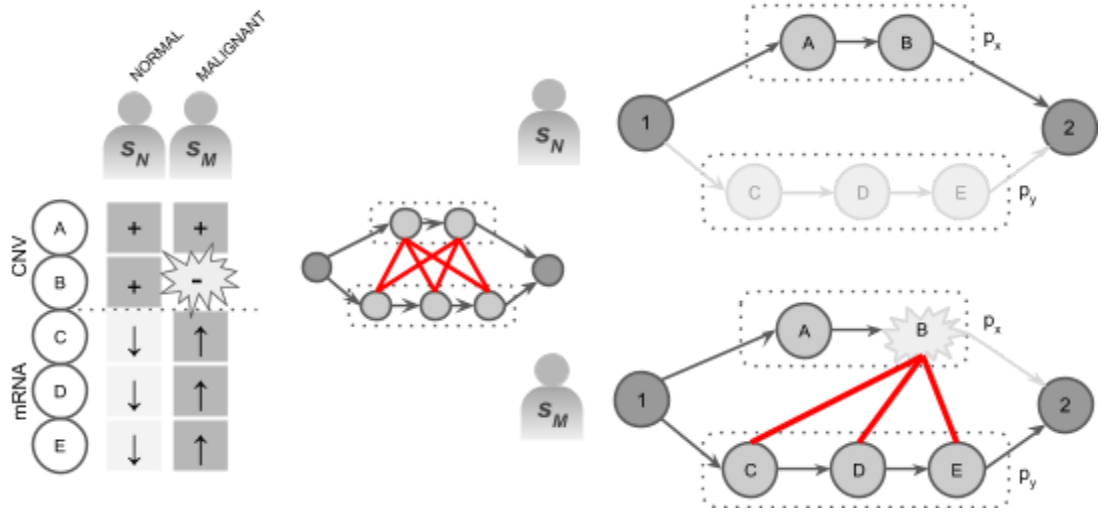

Figure 12. Pathway view of synthetic lethality (SL). The fictional pathway originating at node 1 and progressing to node 2 performs a function is essential to cell survival (right). A sample from a normal patient, S_N, proceeds through the preferred sub-pathway p_x, which relies on the products from genes A and B (top right). To remain energetically favorable, S_N reveals down regulation of genes in the alternative sub-pathway p_y (left). In a malignant patient sample, S_M, pathway p_x is lost due to a copy number deletion in gene B. However, S_M can still rely on the alternative sub-pathway p_y perform to crucial functions, as seen in the increased expression of genes C, D, and E (left). If any of the genes or gene products belonging to sub-pathway p_y were to be targeted in S_M, then the malignant cell would lose both sub-pathways, lose the essential function of the entire pathway, and die. Such potential targets are shown as SL links of genes between the loss context of B and the genes C, D, and E (right, red lines). However, these SL links could occur transitively, and in any combination of loss nodes and essential nodes (center, red lines).

Bridges et al. described the first SL pairing in 1922, when attempting to assemble the first large linkage map in *Drosophila melanogaster* (C. B. Bridges 1922; Calvin Blackman Bridges, Brehme, and Others 1944). Dobzhansky later coined the term synthetic lethality in 1946 to describe complementary lethal phenotypes in *Drosophila pseudoobscura* when crossing WT heterozygote flies with

complementary lethal genes (Dobzhansky 1946). Crossing over between two chromosomes, containing genes that are not individually lethal to homozygotes, created a heterozygous offspring with a recessive lethal phenotype(Dobzhansky 1946). These two studies laid the foundation for experiments meant to intervene and interrogate SL relationships.

Interventional SL relationships first involved drugging an essential “secondary” gene with a similar systemic function to a primary loss-of-function gene. These first cases of synthetic genetic interactions, as they were called, were observed in yeast: first with histone H2B subtypes, and then in the Ras family (Rykowski et al. 1981; Kataoka et al. 1984; Tatchell et al. 1984). Guarente formalized the concept as synthetic lethality in 1993 (Rykowski et al. 1981; Kataoka et al. 1984; Tatchell et al. 1984; Guarente 1993). Hartwell later performed an experiment in yeast showing single knockout of either a DNA proofreading or DNA mismatch repair caused high mutation rate and proliferation, while a double knockout induced apoptosis (Hartwell et al. 1997). This experiment shows that SL pairs can have similar functions, serving as the biological basis for the clinical success of PARP inhibitors mentioned later in this proposal. The work of Hartwell and others laid the groundwork for a number of systematic high-throughput screening approaches in model organisms.

Bench

Budding yeast was one of the first model organisms to be used in the systematic search of SL interactions. In 2001, Tong et al used a synthetic genetic array (SGA), which employed laboratory automation to cross a haploid query strain, or the “loss context”, with an array of haploid deletion mutants (Tong et al. 2001). The subsequent double-mutant meiotic progeny could be examined for “sick” phenotypes of reduced fitness, such as slowed growth rate. Double mutants showing these sick, or SL, phenotypes indicate a functional relationship between query and a mutant gene in the array. SGA was applied to 204 genes with roles predominantly in cytoskeletal organization, cell polarity, and DNA synthesis to generate a network of 291 genetic interactions. The genetic interactions showed a proliferation of SL pairs within gene modules, but also showed SL pairs can bridge orthogonal modules, such as DNA synthesis and DNA repair. The study lacked enough potential combinations, however, to produce a comprehensive network map in yeast.

SGA was later applied to all modules of the *S. Cerevisiae* genome and uncovered relationships between gene function and synthetic lethality. The study by Costanzo et al captured approximately 5.4 million gene pair candidates, comprising nearly 75% of the yeast genome, and found that 13 of the 17 modules were enriched for SL interactions between genes of similar function (Costanzo et al. 2010). This use

of SGA also uncovered SL pair enrichment that bridged disparate functions. Bridging commonly occurred from processes related to chromatin and transcription, whose genes were often the highest connected nodes in their network. However, many of the non-related processes had wide variation in number of SL bridge events. Costanzo et al demonstrated that a linear model of gene-specific properties, derived from ANOVA, could predict genetic interaction hubs, thus explaining variance in SL pairs (Costanzo et al. 2010). For example, a lack of SL bridge events in drug and ion transport modules was attributed to their higher rate of duplication and other copy number variabilities, two properties tested to be negatively correlated with SL pairs (Dunham et al. 2002).

Advancements in systematic genotypic-phenotypic screening of model organisms led to high-throughput screening approaches in cancer cell lines. This was because the task of identifying gene function by phenotypic changes in lower model organisms was similar to screens identifying essential genes in cancer. The biggest difference in cancer screens are the phenotypic endpoints they measure. In cancer cells, the desired phenotypic to measure pertains to viability. These viability phenotypes would later be explored in myriad screens using small molecules, RNAi, and other gene editing techniques that allowed a tumor-centric investigation of essential genes. Applying such essentiality screens to cell lines in the context of SL is a natural extension of target discovery in cancer.

Identifying druggable SL pairs in cancer cells can be performed in the context of combinatorial therapies using a chemogenomics screen. Whitehurst et al designed

a cell-based, high-throughput combination screen, which transfects cells with a library of siRNA while simultaneously treating the cells with compounds that inhibit tubulin (Whitehurst et al. 2007). The siRNA and the compound act like the partners in an SL pair, with the compound acting as the LoF context in its target and the siRNA blocking the translation of an essential gene. Since the cells were treated with subthreshold concentrations of antineoplastic compounds, a siRNA condition that resensitizes the cells to the compound must repress a tubulin-specific essential gene.

Paclitaxol, which stabilizes microtubule assembly, was one such antineoplastic drug Whitehurst et al used as an SL surrogate (Schiff, Fant, and Horwitz 1979). Many of the essential genes that surfaced from this chemosensitizer screen are implicated in microtubule assembly, proteasome subunits, and mitotic checkpoints (Whitehurst et al. 2007). The function of these essential genes, especially the microtubule assembly, are supported in the in previous literature to modulate sensitivity to paclitaxol (Jordan and Wilson 1998; Oyaizu et al. 2001). Due to the positive results, this SL chemosensitizer screening strategy can be considered a success. This screen and associated chemogenomics screens represent both a possible validation strategy for SL pairs found *in silico*, as well as a trove of data from which to infer essential genes from physical features of screened compounds (Rognan 2007).

Computation

The first attempts to identify SL interactions used data from high-throughput genetic and genetic-compound screens. Data produced in these screens, combined with the proliferation of omic data, allowed recent efforts to computationally infer SL interactions. However, the first inference of SL pairs in cancer did not come directly from SL experiments performed in cell lines. These inferences were originally drawn from the corpus of SL interactions experimentally found in yeast. While there is currently sufficient data in human cell lines and tissues to infer SL interactions, examining a previous approaches is worthwhile to highlight a possible route for filtering SL pair candidates of false positives, providing a training data for an inference scheme, or for validating predicted essentiality.

Conde-Pueyo et al used phylogenetic inference that mapped genetic orthologs from yeast SL networks to human genes to identify new essential genes (Conde-Pueyo et al. 2009). At the time, the SL screening data was predominantly available in the form of genetic profiling in yeast (Tong et al. 2001; Pan et al. 2004; Ooi et al. 2006; Boone, Bussey, and Andrews 2007). Mapping orthologs in this way is justified since genome integrity and cell-cycle related genes from yeast are conserved in higher organisms and have close relation with drivers of tumorigenesis (Yuen et al. 2007). Although not all SL pairs may be conserved between yeast and human, SL interactions at the pathway level have a higher probability of being evolutionarily

conserved, as seen in the high rate of SL pair conservation observed among eukaryotes (Yuen et al. 2007; Measday et al. 2005; Tarailo, Tarailo, and Rose 2007; Dixon et al. 2008). The authors stress that this approach not about supporting SL conservation between distantly-related species. Rather, the authors emphasise the role this methodology has in obviating large chemical SL screens, a utilitarian purpose that belies SL inference. What makes this strategy work is the reliance on known biology from a model organism.

The above approach shows promise in its use of genetic ontologies to incorporate biological knowledge into an SL network. Given the high rate of false positives from experimentally derived SL pairs, it is necessary to use a filtering method to remove these extraneous results (von Mering et al. 2002; J. P. Miller et al. 2005). Filtering occurred after Conde-Pueyo et al applied yeast networks from BioGRID genetic interaction repository to human orthologs using Ensembl (Stark et al. 2006; Hubbard et al. 2007). Three databases were used for filtering: 1) COSMIC, used to filter genes already associated with cancer, 2) GO, used to filter genes by function (ie. cell cycle, or DNA repair), and 3) DrugBank, used to map gene targets to existing drugs for filtering on tractability (Bamford et al. 2004; Ashburner et al. 2000; Wishart et al. 2008). The use of these genetic and pharmacological ontologies to filter false positives of little biological relevance or druggability is an important step to consider in the search for actionable synthetic lethal pairs.

DAta mIning SYnthetic lethality identification pipeline (DAISY) is a recent attempt to infer SL maps in cancer, whose methods can be applied to the CLOvE approach. This pipeline used neural network model predictions in three different inference procedures to each produce a set of SL and SDL pairs: 1) genomic survival of the fittest (SoF), 2) shRNA-based functional screens, and 3) pairwise gene expression (Figure 13) (Jerby-Arnon et al. 2014). SoF uses somatic copy number mutations (SCNA) with somatic mutation data to identify pairs of inactive genes without a corresponding partner gene deletion (SL) or overactive genes without a corresponding partner deletion (SDL). The group carried out the same inference strategy on functional shRNA screens, looking at SCNA data in conjunction with shRNA knockdowns. Finally, pairwise gene coexpression created a gene set by which to restrict all possible gene pairs from the first two inferential procedures, to create a final network comprising the intersection of all three gene sets.

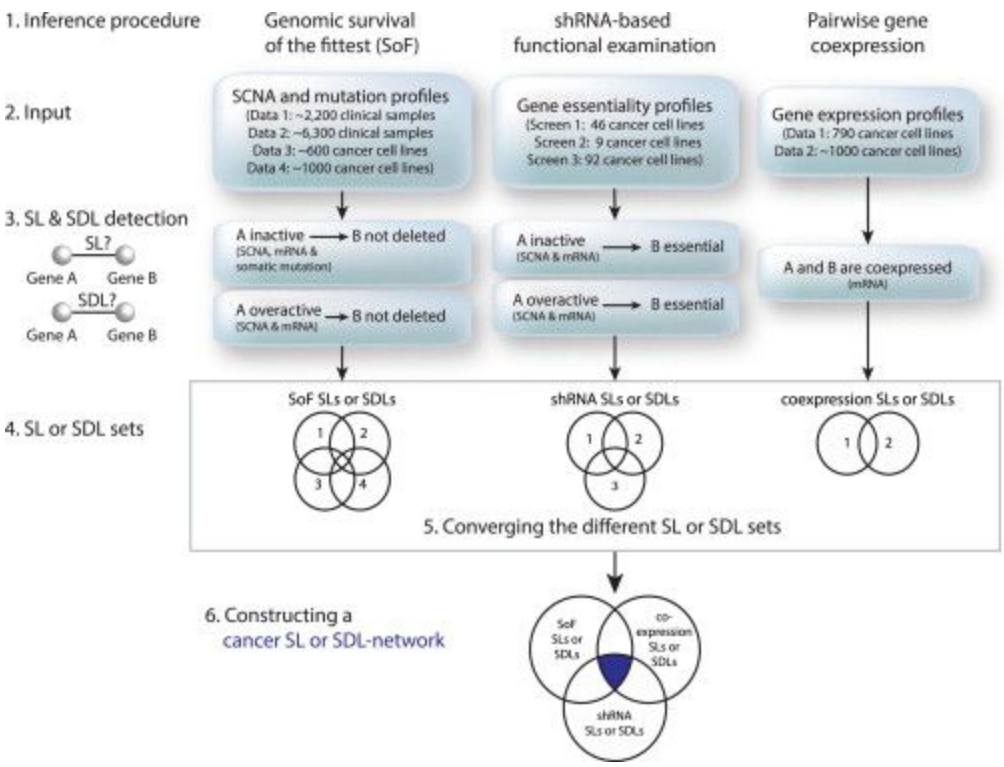

Figure 13. The DAISY Workflow. Three different inference procedures (1) are applied in parallel to identify SL and SDL gene pairs. The SL and SDL networks are then assembled from gene-pairs that are identified by all three procedures as SLs or SDLs, respectively (the intersection colored in blue). Reproduced from Jerby-Arnon ((Jerby-Arnon et al. 2014)

DAISY also employed two neural net models that could predict the relation between target gene and drug sensitivity. One model predicted gene-cell pair relation to infer gene essentiality by cell line. The other model predicted drug-cell pair relation to infer drug efficacy by cell line. Each type of model used a set of 53 genomic features extracted from gene neighborhood and other genomic features by each cell line. The group used two models from two gene essentiality datasets: one based on

a compendium of pooled shRNA screens, and the other from a genome-wide essentially assessment termed Project Achilles (Cheung et al. 2011; Marcotte et al. 2012; Cowley et al. 2014). Training data for drug-pair relations came from IC50 values from the Cancer Genome Project (CGP) and from the Cancer Therapeutics Response Portal (CTRP) (Garnett et al. 2012; Basu et al. 2013)[1]. Use of these biological and pharmacological features revealed known SL pairs BRCA1-PARP, BRCA2-PARP, and DHFR-MSH2, and related them to drug efficacy and ultimately patient survival. The final relation to drug selection and prognosis is of great interest in the area of personalized oncology, and one of the potential applications of the discussed approach.

DAISY does, however, have several limitations which can be addressed with an improved gene recommender. One limitation is the inherently noisy data from large scale genomic data and shRNA screening, which can be normalized across a library of non-targeting shRNA. DAISY also doesn't account for other mechanisms of gene regulation like epigenetics that could be incorporated into the SL sets with an appropriate weighting scheme. The ability of DAISY to predict gene essentiality to relevant biological models also remains unknown, but could call for a limited target validation campaign in cell lines and perhaps in animal models. Coexpression can also be affected by genomic linkage and thus chromosomal proximity. DAISY also

[1] IC50 is a pharmacological measurement of drug sensitivity defined as a half-maximal dose, or a drug concentration that destroys 50% of cancer cells via inhibition. EC50 is related to IC50 in terms of IC50, but is typically applied to agonist drugs. Area under activity curve (AUAC) is the integral of either IC50 or EC50, and is another measurement of drug sensitivity.

lacks a filter for SDL pairs based on proximity, which could be a feature of the CLOvE approach. Though DAISY successfully identified known SL pairs in relation to clinical outcomes, this methodology has clear limitations that can be addressed in an improved gene inference strategy.

There is an effort to aggregate and curate known SL pairs, including those from DAISY, which the discussed approach could use as both a predictive aid and a distribution platform. Guo and colleagues created SynLethDB as a repository for SL pairs, in addition to serving as a platform for statistical analysis and visualization of SL interactions (Guo, Liu, and Zheng 2016). The 34,089 SL pairs in SynLethDB are 58.5% human and 38.8% yeast, with the remainder split between the model species mouse, fruit fly, and worm (Guo, Liu, and Zheng 2016). What makes SynLethDB particularly useful is the system it uses to score the confidence of SL activity in a gene-gene interaction pair.

SynLethDB stores and scores SL pairs collected from biochemical assays (chemogenomics and RNAi), computational inference (DAISY), other SL databases (BioGRID, Syn-Lethality, and GenomeRNAi), and literature mining (Stark et al. 2006; Li et al. 2014; Schmidt et al. 2013). Pairs are given continuous confidence scores based on weights assigned to the type of experiment method used to validate that particular pair. Paris with more than one experimental type are combined using the probability disjunction formula. These scores can serve as a gene-gene interaction filter, a source of important feature extractions based on a large corpus of SL

research, or as a tool to validate predictions of gene essentiality. The latter use is recommended in this proposal as a final validation measure. It is also worth noting that SynLethDB could also incorporate both the results of the discussed approach, as well as a distribution of the recommender itself for user-supplied data.

Collateral vulnerability is related to the concept of SL by virtue of genome instability in cancer, and is worth noting as a possible strategy in identifying essential genes in cancer. Whereas one of the SL partner genes is typically a LoF event in a known tumor suppressor gene (TSG), collateral vulnerability involves LoF events in bystander genes (Knudson 2001). These bystander genes may not be TSGs themselves, but they create a loss of genetic redundancy that can be exploited in a similar fashion to an SL-based inference scheme and subsequent treatment campaign.

Amplifications of oncogenes and deletions of tumor suppressor genes often happen at the level of the chromosome arm (Y. Liu et al. 2015). Though these events are positively selected as tumorigenic driver mutations, these mutational events lead to non-advantageous alterations, or passenger mutations, in neighboring genes on the affected chromosome (Nijhawan et al. 2012). Functional redundancy, in the form of homologous genes, attenuates viability effects in the case of heterozygous LoF in an essential housekeeping gene, and enables cell survival (Vavouri, Semple, and Lehner 2008). As demonstrated in yeast, reduced redundancy that leads to homologous LoF in an essential gene causes decrease of fitness and death (DeLuna

et al. 2008). Therefore, pharmacological inhibition of redundant gene products, or collateral vulnerabilities, in cancer could be used as an effective strategy for target discovery.

Muller et al tested the collateral vulnerability hypothesis in GBM cells and identified a druggable target in the enolase (ENO) family of genes. Investigators identified the homozygous loss of ENO1 in GBM using TCGA data to query homozygously deleted essential-redundant housekeeping genes (Muller et al. 2012). ENO1, a gene product responsible converting 2-phosphoglyceric acid into phosphoenolpyruvate during glycolysis, is mutated in chromosome 1 in GBM cells but is permitted due to the paralog ENO2 widely expressed in neurons (Wold 1971; Poyner and Reed 1992; Kobayakawa et al. 2007).

Muller et al found that shRNA knockdowns in ENO1-deficient mouse xenografts lead to tumor ablation (Muller et al. 2012). This was also tested in ENO-1 deficient cell lines using phosphonoacetohydroxamate (PhAH), a potent inhibitor of Recombinant *T. brucei* enolase (Navarro, Gomes Dias, and Mello 2007). Homozygous ENO-1 mutant cell lines experienced an 80% decrease at <5uM of ENO inhibitor PhAH versus WT controls, while heterozygous mutants experienced similar loss of viability at a 50uM dose (Muller et al. 2012). This experiment demonstrates that tissue-specific vulnerabilities can be exploited for therapeutic effect, but require a systematic approach to identification as recommended in this proposal.

Another important example of collateral vulnerability in cancer is in the mutual exclusivity of ARID1A and ARID1B. The ARID gene products are responsible for one of the five subunits of the SWItch/Sucrose Non-Fermentable (SWI/SNF) nucleosome remodeling complex and are mutated across several cancer types (Cairns et al. 1994; Wiegand et al. 2010; Guan, Wang, and Shih 2011; Jones et al. 2012). Though ARID1A is mutated in many human cancers, Helming et al identified ARID1B, a mutually exclusive homolog of ARID1A that stabilizes the SWI/SNF complex in cancer cells and allows proliferation (Helming et al. 2014). Mutual exclusivity refers to the presence of either ARID1A or ARID1B in the SWI/SWF complex (Wu and Roberts 2013).

Helming and colleagues identified this complex in a systematic screen of the essentiality dataset from Project Achilles. This dataset, v2.4, comprised essentiality data from pooled lentiviral shRNA screens in 216 cell lines with a proliferation/survival endpoint (Cheung et al. 2011). From these data, Helming et al used the Probability Analysis by Ranked Information Score (PARIS) to find genes essential to specific mutational contexts (Helming et al. 2014). PARIS is an information-based feature selection method that identifies a target profile *t*, like the LoF of a tumor suppressor, and matches essentiality profiles *x* at the level of either gene or shRNA. This match is based on a rescaled normalized information score (RNMI). The RNMI takes the normalized mutual information (NMI), a ratio of the mutual information *I* and joint entropy *H* (NMI) of both *t* and *x*, multiplied by the

correlation coefficient of the two profiles, adding directionality to the score (Cowley et al. 2014). This scoring metric enables a wide dynamic range and sensitivity to non-linear correlations, which could serve as an important feature to predict SL pair importance.

One important consideration in collateral vulnerability are those cases of hemizygous loss, which are far more prevalent in cancers than homozygous loss (Nijhawan et al. 2012). One important approach, CYCLOPs (copy number alterations yielding cancer liabilities owing to partial loss), investigated hemizygous essentiality (Nijhawan et al. 2012). This approach used shRNA screening data from Achilles with copy number data from CCLE and TCGA to identify genes where RNAi knockdown produced a viability phenotype in cell lines with a partial loss of that same gene (Nijhawan et al. 2012). One CYCLOPs gene, PSMC2, encodes an essential member of the 19S proteasome (Kaneko et al. 2009). While, normal cells express excess PSMC2, cancer cells with partial PSMC2 copy number loss cannot form the proteasome and enter apoptosis upon PSMC2 knockout. Such an example demonstrates partial collateral vulnerability in an otherwise innocuous passenger mutation.

The CYCLOPs approach classified cell lines as homozygous, hemizygous, or copy neutral using GISTIC log2 copy number ratios less than -1.28, less than -0.1, and between -0.05 and 0.05 respectively. Nijhawan and collaborators assessed mean gene dependency scores from the Achilles data using ATARiS, which scores genes

according to greatest effect on viability phenotype in an RNAi screen (Shao et al. 2013). These mean scores for samples with partial and total loss per gene were both compared to cell class permutations that maintained similar tissue distribution. Classifying cell lines by complete or partial loss and comparing some form essentiality scores within each group represents a reasonable method to identify SL pairs from collateral vulnerability.

Hypermutation, an outcome of the hallmark of genome instability, can provide further elucidation of driver mutations in cancer. Mutations exist throughout the cancer genome, but are characterized as either driver or passenger mutations, as described previously. Driver mutations aid in tumorigenesis and are therefore positively selected. Identifying such genes require showing a difference, or bias, in accumulating these mutations over the background rate of mutations. Creating a score based on this bias can then be used to rank or filter SL pairs, since genes with higher mutational bias scores may have greater reliance on their partner SL.

Early efforts to identify mutational bias examined recurrence of alterations. This approach ranked genes based on the probability of observing a number, or recurrence, of non-synonymous single nucleotide variants (nsSNV) in a set of samples (Sjöblom et al. 2006; Greenman et al. 2007). Assigning bias score based on recurrence may assign higher importance to driver genes that arose earlier in clonal expansion, and thus assign low importance to later mutations that may have low recurrence despite high functional impact (FI) (Shah et al. 2012). It is thus

important to consider FI and recurrence when creating identifying driver genes in an SL pair.

One solution is to use several bioinformatics methods to score the FI of somatic mutations, average these scores at the gene level, and assess their statistical bias compared to rate of background mutation. Lopez-Bigas and Gonzalez-Perez first recommended this statistical approach in their description of Oncodrive-fm using the homology-based SNV detectors SIFT, PPH2, and Mutation Assessor (Gonzalez-Perez and Lopez-Bigas 2012). SIFT examines protein sequence homology using PSI-BLAST and scores residue changes position-specific substitution matrices with Dirichlet priors(Altschul et al. 1997; Kumar, Henikoff, and Ng 2009). PPH2 (PolyPhen2) uses both sequence conservation and structural features structure from SWISS-PROT(Boeckmann et al. 2003; Adzhubei et al. 2010). Mutation Assessor scores patterns in amino acid residues that deviate from aligned homologs using a clustering algorithm called combinatorial entropy formalism (Reva, Antipin, and Sander 2011). Two possible extensions to these SNV detections are MAPP and Align-GVGD, which incorporate physicochemical features of amino acid substitutions (Stone and Sidow 2005; Mathe et al. 2006).

As suggested by Lopez-Bigas and colleagues, the above SNV detection methods can be averaged at the gene level to create a unified FI score. Other non-synonymous mutations, like stop SNVs (stSNVs) and frameshift indels (fsindels), can affect tumorigenesis. These mutations' impact on protein function is

not captured with the above mutation detectors, but can be inferred. High nsSNV scores can be attributed to stSNVs and fsindels since they cause the most significant amino residue changes. Lopez-Bigas and colleagues suggested these very assumptions, then averaged the FI for a gene across all tissues, and compared the difference to the null set to generate an empirical *P*-value (Gonzalez-Perez and Lopez-Bigas 2012). The *P*-value for a gene was calculated by selecting a set of possible mutations equal in length to the number of mutations observed in that gene. Selection occurred 1 million times with replacement, producing a fraction of selections (*P*-value) that had an FI that was greater than or equal to the observed FI of the gene (Gonzalez-Perez and Lopez-Bigas 2012). Such significance testing is a reasonable approach in assessing the importance of mutants in a SL pair, and is considered in the CLOvE approach.

There exists a range of potential vulnerabilities in a cancer cell. Hypermutation, copy aberrations, and subsequent expression changes all serve as clues on the path to identify these new cancer vulnerabilities. However, these clues must be considered in totality, examined for their relative importance, and used to build a case for a compensatory mechanism that gives rise to synthetic lethality.

Clinic

SL has seen much success in the clinic due to the increased selective targeting of malignant cells, and simultaneous indifference to normal somatic cells. One such class of SL cancer therapies are the inhibitors of poly ADP-ribose polymerases (PARPs). These include compounds Olaparib, Rucaparib, and Niraparib, which have been approved for ovarian cancer following the trials Olympiad (Robson et al. 2017), ARIEL2 (Swisher et al. 2017), and NOVA (Mirza et al. 2016), respectively. These medications are shown to be effective in ovarian cancers harboring a BRCA mutation. The relationship between BRCA and PARP is without question the best known example of SL in cancer.

SL interactions in cancers with BRCA1 and BRCA2 loss were described as a set of 'BRCAness' hallmarks and were later shown to be useful selectively targeting cancer (Y. Liu et al. 2015). The rationale for initially pursuing this target is revealed when one explores the role of BRCA in the cell. BRCA genes are responsible for homologous recombination (HR). HR is a type of genetic recombination during which cells repair double-stranded breaks (DSBs) in DNA. BRCA genes are therefore essential in protecting genomic fidelity during cell proliferation and are an important biomarker for predicting risk of breast or ovarian cancer (Antoniou et al. 2003). The family of poly ADP-ribose polymerases (PARPs) play an analogous role

to BRCA in DNA repair, but do so at the single-strand level using either nucleotide or base excision repair single stranded breaks (SSBs) (Lindahl et al. 1995).

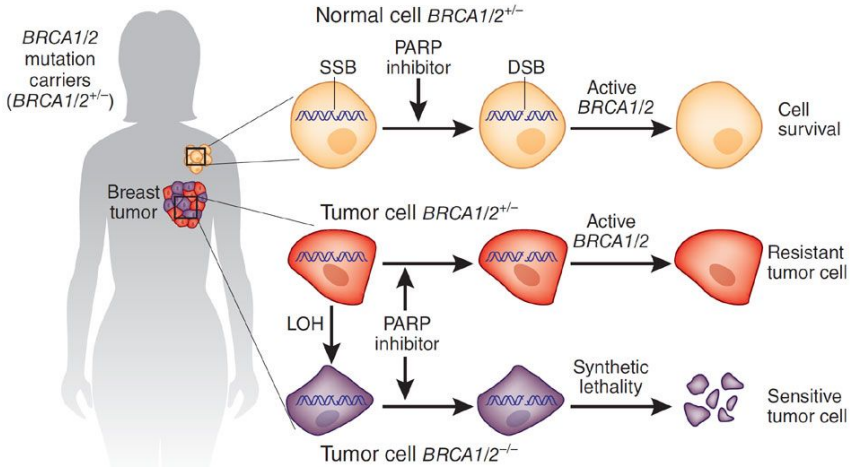

Figure 14. Synthetic lethality of PARP in BRCA mutation carriers. BRCA and PARP represent genes whose gene products are responsible for repairing double strand breaks (DSBs) and single strand breaks (SSBs), respectively. Each representation of a cell type comes from a patient with breast cancer who is treated with a PARP inhibitor. Normal somatic cells, in yellow, loose their ability to repair SSBs when treated with a PARP inhibitor. These cells accumulate DSBs during replication, but are then repaired due the presence of at least one copy of BRCA, allowing normal cells to survive. Unlike these normal cells, tumor cells that experience a loss of heterozygosity (LoH) event in BRA, in purple, cannot rely on BRCA to repair the accumulated DSBs, and thus initiate apoptosis. However, some tumor subclones that retain a wild-type copy of BRCA, in red, can fix their accumulated DSBs and lead to PARP inhibitor resistance. Resistance may also occur due to the reliance of another gene product which may initiate an alternate DNA repair mechanism. Such a compensatory mechanism can be inferred from a network view of synthetic lethality as recommended in this proposal. Reproduced from Polyak and Garber (Polyak and Garber 2011).

In cancers with a LoF mutation of BRCA and an inhibition of PARP through therapeutic intervention, single-strand breaks are allowed to accumulate throughout the cancer genome. When the cell is in S-phase, the single-strand breaks are

copied as double strand breaks at the DNA replication fork, and double-strand breaks accumulate in turn (McLornan, List, and Mufti 2014). Without BRCA to provide HR to repair this new DNA damage, the cancer genome becomes so aberrant as to induce apoptosis (Figure 14). Normal cells with active BRCA are able to overcome the formation of DSBs from replicated SSBs caused by PARP inhibition. This leads to a wider therapeutic window: survival of normal cells with a simultaneous sensitization and lowered viability of BRCA deficient cancer cells.

The SL approach first gained therapeutic support in-vitro with BRCA deficient cells dosed with PARP inhibitors, as performed by Bryant et al. The BRCA2-negative cell line V-68 was found to have a 90% decrease in viability compared to a similar BRCA-positive cell line V-79 dosed with 20uM of a PARP inhibitor NU1025 (Bryant et al. 2005). Mouse xenografts of these cell lines also experienced at least 50% tumor debulking within 10 days of NU1025, with 60% of mice experiencing either complete or partial remission (Bryant et al. 2005). These early experiments gave way to RNAi screens that uncovered DNA damage response regulator CDK5 and the role it plays as a sensitizer to PARP inhibition (Turner et al. 2008). PARP inhibitors eventually appeared in the clinic. In its clinical trial, olaparib, an FDA-approved PARP inhibitor, showed a 41% objective response rate in patients with BRCA-deficient ovarian cancers (Tutt et al. 2010). The premise of SL demonstrates practical therapeutic utility, and warrants a method to identify additional druggable targets.

An important result of the PARP-BRCA pair was the translation to other cancers with DSB mutations, namely endometrial cancers. Approximately 80% of endometrial cancers have mutations in phosphatase and tensin homolog (PTEN) (Mutter et al. 2000). In addition to regulating the PI3K-AKT-mTOR pathway, a LoF event in PTEN contributes to genetic instability due to defective HR repair of accumulated DSBs (Shen et al. 2007). Given the analogous role of BRCA and PTEN, there was clear motivation to test for PARP sensitivity in this major class of cancer. Dedes et al dosed endometrial cancers with PARP inhibitor KU0058948 and found a ~90% viability reduction in PTEN deficient cancers versus cancers without PTEN mutations (Dedes et al. 2010). This sensitivity ultimately resulted in the development of other PARP inhibitors, like ABT-888, with clinical investigations expanded to different cancers prone to missing DSB repair mechanisms, like lymphomas and colorectal cancer (Kummar et al. 2011; S. K. Liu et al. 2008/8; Pishvaian et al. 2011).

Conclusion

PARP inhibitors are a triumph of SL in the clinic, which validate continued drug discovery efforts based on SL, but they are only the beginning. Continuing these efforts requires the use of new SL pair-finding methodology in target identification. Historically, this methodology SL relied on the extending SGA-related techniques to experimentally identify and validate SL targets. Computational approaches have

given way to new ideas about how to interrogate these SL connections further. The new approach demonstrated here, CLOvE, seeks to address the limitations of these previous approaches and extend our understanding of SL interactions to further drug discovery in cancer.

CHAPTER 3

Description of CLOvE: The Method & The Data

Summary

- CLOvE is a computational approach to infer SL interactions arising from context-dependent changes in expression
- Developed in Cancer Cell Line Encyclopedia (CCLE) cell line samples
- Description of CLOvE pre-processing, filtering, computation, and post-processing steps

Introduction

Modern cancer interventions rely on targeted therapies to achieve maximal efficacy and minimal harm. Most of these targeted therapies arose from identifying and exploiting genetic vulnerabilities unique to cancer (J. Luo, Solimini, and Elledge 2009). One such vulnerability, oncogene addiction, describes a cancer cell's overreliance on a driver gene. Although inhibiting oncogenes like EGFR or MET have clinical merit, there are potentially more targets that exist as passenger drivers (Haber et al. 2005; Gherardi et al. 2012). These passenger drivers are often due to genetic dependencies that arise from genomic instability within the cancer cell (J. Luo et al. 2009). Synthetic Lethality is one approach to uncover the genetic interactions that define these non-oncogenic drivers, and perhaps identify new possibilities for cancer treatment.

Synthetic lethality (SL) refers to a relationship between two genes where a defect in a single gene is permissible, whereas defects in both genes cause cell death (Ashworth, Lord, and Reis-Filho 2011). A better understanding of these interactions led to the development of the most recent wave of chemotherapeutic agents, which include PARP inhibitors for BRCA-deficient breast cancer (Tutt et al. 2010; Bajrami et al. 2012). The ongoing search for these SL interactions might lead to more specific and efficacious therapeutic approaches.

Initially, SL interaction networks were derived exclusively in the laboratory via high-throughput screens in model organisms (Byrne et al. 2007; Costanzo et al. 2010; Typas et al. 2008; Ma, Tarone, and Li 2008). Similar techniques were applied to the study of cancer through screening cell lines with libraries of RNAi and small molecules (Bassik et al. 2013; Brough et al. 2011; Laufer et al. 2013; McDonald et al. 2017). However, even with advances in laboratory automation, it is not feasible to test every combination of possible SL gene pairs. Computational approaches, therefore, can provide a more systematic view of all possible SL interactions that may exist in a type of cancer.

Though there are several recent approaches to computationally infer SL networks, they are not without limitations. Many early computational efforts to predict SL were carried out in model organisms as a way to validate and build upon previously known SL pairs in those organisms, which do not accurately reflect the genomic landscape of cancer (Wong et al. 2004). However, those approaches carried out on cancer have their own particular limitations. One method uses mutual exclusion to infer SL across mutation and copy number, but is not adequately validated and is constrained by prior pathway knowledge (Ciriello et al. 2012). Another method uses genomic and transcriptomic data to capture SL pairs in a single cell line, but relies heavily on noisy shRNA screening data (Jerby-Arnon et al. 2014). Yet another proposed method also uses mutual exclusion in copy number and mutation separately across a subset of Pan-Cancer TCGA samples, but fails to make comparisons across these platforms and ignores the role of the transcriptome in

determining genetic dependencies (Mina et al. 2017). There is a clear need for a tool to infer synthetic lethality that takes into consideration of genomic and transcriptomic data across both cell types and patient samples.

Here we propose a method to identify context-dependent changes in expression. Like some computational approaches before it, this approach finds pairwise genetic dependencies. However, unlike previous approaches, this method looks takes a multi-omic approach to find meaningful increases in expression that may be associated with genetic dependency. The transcriptome is a reasonable proxy for proteome, especially for those gene products implicated in cancer (Young et al. 2001; Shankavaram et al. 2007; B. Kim et al. 2015; Vogel and Marcotte 2012). Therefore, identifying meaningful changes in expression may serve as an indicator for possible synthetic lethal interactions, and potentially new drug targets.

Population Description

CLOvE was first developed using cell line data. The first source of molecular data comes from the Broad Institute's Cancer Cell Line Encyclopedia (CCLE). CCLE was a collaborative effort between the Broad Institute, Novartis Institutes for Biomedical Research, and the Genomics Institute of the Novartis Research Foundation. It contains a collection of at least five platforms of molecular data, including copy

Number, mRNA expression, RPPA, RRBS, and mRNA expression (RNAseq), taken from at least 1046 Cell Lines spread among 24 tissue types (Fig 8) (Barretina et al. 2012). Mutation, copy number, and expression are the primary sources of data used in the development of CLOvE.

One reason for using CCLE data is that it served as a pilot to future SL networks inferred from TCGA samples. By comparing the predicted SL results from cells and tissues, it is possible to glean insights into the applicability of cell line models of tumor essentiality. Possible changes, if any, of final SL predictions between cells and tissues may affect the use and interpretation of experiments that use immortalized cells recapitulate tumors. Secondly, predicting essential genes in cell lines is a goal unto itself. While many can debate the contrived nature of cell models, cell-based assays and screens are an established mainstay in drug development and target identification (Balis 2002; Allen et al. 2005; Sharma, Haber, and Settleman 2010; Weinstein 2012; Wilding and Bodmer 2014). Predicting SL networks in cells alone supports the current trend of targeted cell-based screens in target discovery (Garnett and McDermott 2014).

Using CCLE data has three distinct advantages. First, a high diversity and respectable depth of many cell lines can capture tissue-specific SL interactions. Second, the Broad hosts varied and comprehensive molecular data from which to infer SL targets, including mRNA, copy-number, mutation calls, and pharmacological profiling. The wide use of CCLE cell lines as model systems also

opens the possibility of one-off screens. Third, cell lines are a cost effective and feasible system for in-vitro validation, as mentioned as an extension to CLOvE validation in chapter 5. Apart from these advantages, cell lines are a favorable system to model because they are an important stepping stone to patient samples.

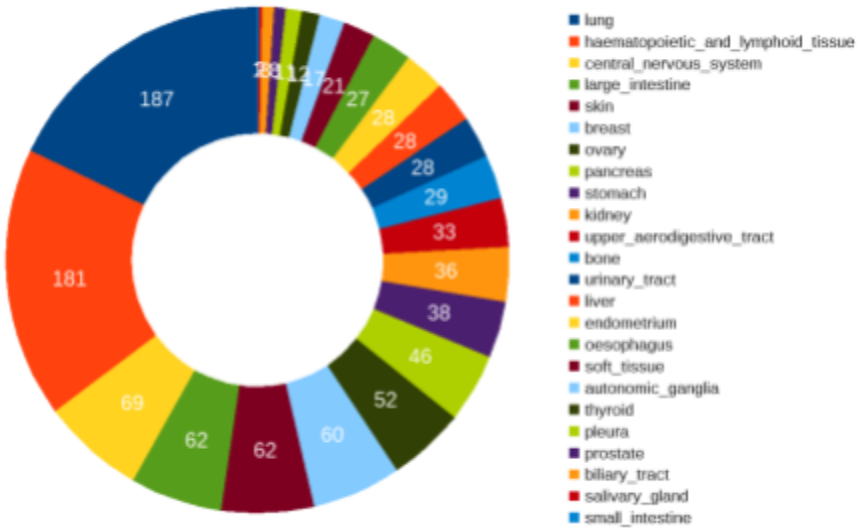

Figure 15. An overview of tissue site origins for CCLE cell lines. The CCLE data set includes 1046 Cell Lines spread among 24 tissue biopsy locations. The most represented cell lines come from lung cells and blood cells (comprising hematopoietic and lymphoid tissue), at 187 and 181 cell lines, respectively. All but 4 tissues (prostate, biliary tract, salivary gland, and small intestine) are represented by more than 20 cell line samples.

Data Preparation

Creating the initial network of SL candidates was inspired by the search for compensatory changes in expression following a copy-loss event. However, compensatory changes, as measured by expression, are often associated with the tumor types from which each of these cell lines are derived (Domcke et al. 2013). Tissue specific expression is shown in Figure 16. In order to remove potentially confounding expression signals between patients of different tumor types. Though other means to remove confounding factors exist in the CLOvE workflow (discussed in Post-Processing), it is recommended that CLOvE be run at least within tumor groups. This subsampling of cell lines, and their respective data from each platform, will allow a fairer intragroup comparison of possible compensatory changes. These compensatory changes were first examined in two matrices of genomic data from CCLE, mRNA expression (RNA) and copy number variation (CNV) and later combined with mutational data.

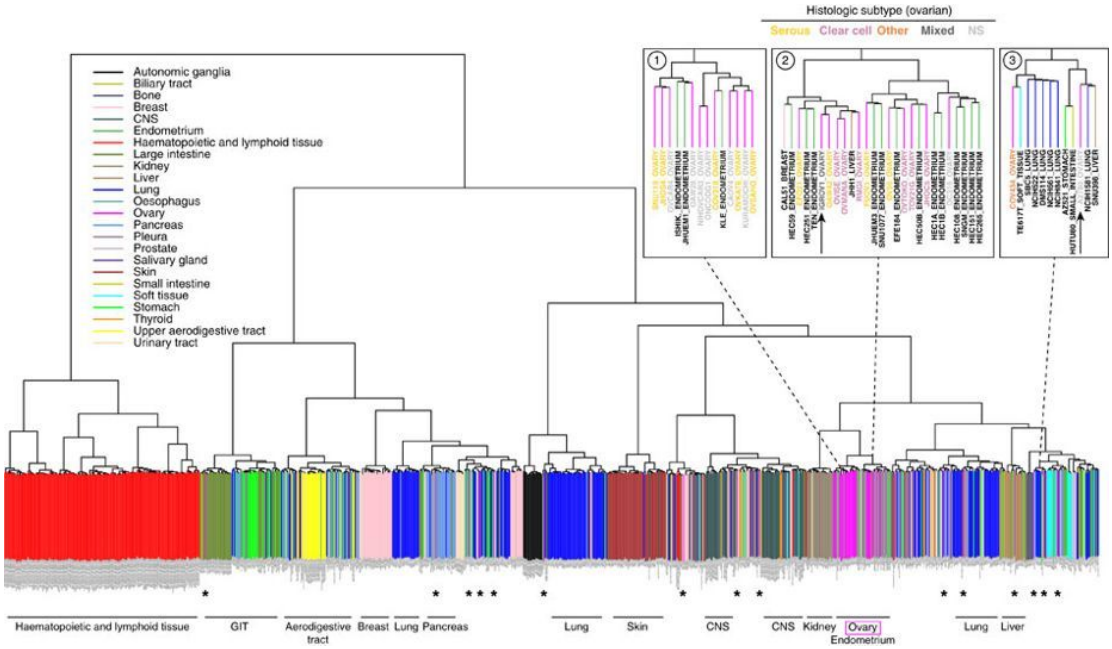

Figure 16. Expression Tends to Cluster by Tissue-Specific Origin. Cell lines were clustered based on mRNA expression data that was limited to the top 5000 genes with the greatest variation across 949 cell lines. Colored bars represent different tumor origins. Though some cell types, like ovarian (*) are dispersed across many clusters, most cell lines tend to separate from one another. This is illustrates that inter-tissue comparisons in CLOvE may show differential expression between tissue types, rather than true compensatory changes. However, in heterogeneous tissues like ovarian, different subtypes may present the same confounding factors of expression (upper right panel). These differences warrant correction and need to be addressed. Reproduced from Domcke (Domcke et al. 2013).

CNV data is taken from Affymetrix SNP6.0, a high-density single nucleotide polymorphism (SNP) array. These arrays were converted to single probe values used to infer copy number based on linear calibration curves specific to that probe set, normalized to normal HapMap samples as $\log_2(CN/2)$, and segmented with the circular binary segmentation (CBS) (International HapMap Consortium 2003; Olshen et al. 2004; Barretina et al. 2012). CBS segments these array copy number experiments into contiguous groups to not only model contiguous losses or

amplifications that reflect biology, but to minimize the noise from individual markers. Next, segments were median centered to zero, quality checked, and run through the Genomic Identification of Significant Targets in Cancer (GISTIC) 2.0 algorithm (Mermel et al. 2011). GISTIC accounts for underlying alterations, whether at specific foci or for an entire arm, to form an improved background rate of alternation. These CNA coordinates were mapped to the hg19 (NCBI build 37) coordinates to provide gene-level copy number using the UCSC Genome Browser (Kent et al. 2002; Wolfsberg 2011). A boolean mask of *n x m* matrix of *n* genes and *m* samples was then created from this data, where 1 reflects a loss and 0 reflects no loss of a particular gene in a particular sample. These calls were used to define the context across the cohort of samples in the CLOvE analysis.

Mutation, in addition to copy number, can be responsible for a LoF context. Mutation data was downloaded from the COSMIC database v56, and cell lines were annotated according to the tumor types, and with a set of 1651 genes from a hybrid capture set (Barretina et al. 2012). Further characterization of mutations at specific loci were generated using mass spectroscopy genotyping based on the OncoMap platform (Thomas et al. 2007). This consisted of 456 genotyping assays comprising 392 mutations and validated with non-multiplexes assays. Later, REVEALER, a computational method to identify genomic alterations correlated with LoF phenotypes, generated binary mutation calls for each gene in each of the 1651 cell line samples (J. W. Kim et al. 2016). These REVEALER mutation calls were used to populate a *n x m* matrix of *n* genes and *m* samples. Later, the union of mutation

and copy data were combined into a single context with inclusive disjunction, often referred to the logical **or** operator. The rationale for combining mutation and copy number in this way is based in functional loss "by any means necessary." In this case, CLOvE is cause-agnostic, focusing only on loss despite the reason for that loss.

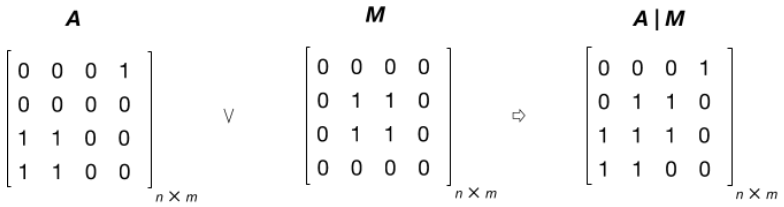

Figure 17. Copy number alterations combined with mutations. Both copy number, *A*, and mutation, *M*, data are binarized into *n X m* matrices of *n* genes and *m* samples. Each binary variable represents the presence (1) or absence (0) of a context gene. When combined disjointly as a union A | M, a context may be called if a context is detected in either *A* or *M*. If either matrix used has a different dimension, an inner merge is taken to preserve samples or genes with data in both platforms.

Expression data is taken from Affymetrix U133+2 arrays and converted to a single probe using Robust Multi-array Average (RMA). RMA is a three step normalization procedure for Affymetrix data comprising background correction, quantile normalization and summarization (Irizarry et al. 2003; Barretina et al. 2012). This normalization removes batch effects and improves the quality of comparison between samples. Expression data was further wrangled into a *n x m* matrix of *n* genes and *m* samples containing only "present" calls. These expression measurements are measured as a response to changes in context.

Pre-Processing

CLOvE determines the significance in expression change in a population of samples with a loss context versus samples without a loss context. The general pre-processing work-flow is represented as a schematic Figure 18. CLOvE requires at least two matrices of data: a binary matrix of context, and a matrix of expression values. The data is read, filtered for confounding copy number effects in expression, filtered again according to user-defined cutoffs, then prepared for CLOvE analysis. CLOvE then runs iteratively over all possible pairs with a randomized null model running in parallel. Upon completion, CLOvE prepares a results for each pair, comprising the CLOvE score, p-value, and other summary statistics, sorted in descending order of significance. These results can be pivoted to form a matrix or a bipartite graph.

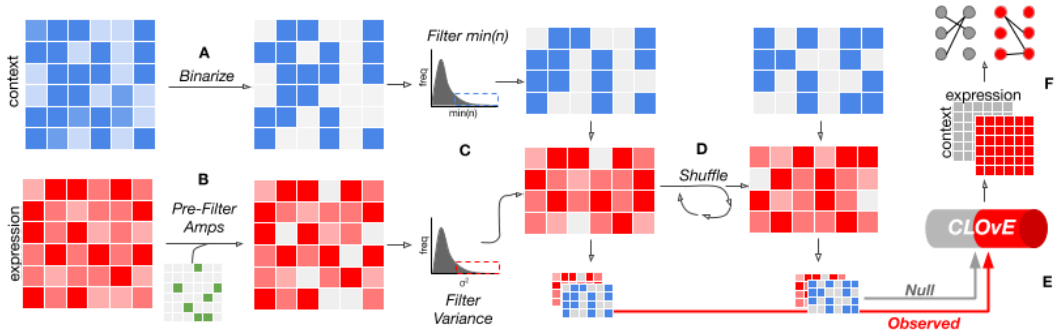

Figure 18. Overview of the pre-CLOvE workflow. Prepared context and expression files are read from CSV or TSV format into a matrices of context x sample and expression x sample, respectively. Each matrix is standardized to contain the intersection of samples and genes (not shown). (**A**) The context matrix is binarized, and (**B**) the expression matrix is pre-filtered for any increase in expression due to an amplification. (**C**) Pre-filtered matrices are then filtered again according to user parameters. This includes a minimum threshold, applied across the cohort of samples, for number of samples in a gene context condition, *n*, and minimum variance for a gene's expression, σ^2. (**D**) A copy of the filtered expression matrix is randomly permuted (shuffled) to randomize associations between context and expression, and paired with the filtered expression matrix to form the null set (grey arrow). Similarly, the un-permuted and filtered expression matrix is paired with the filtered context matrix to form the observed set (red arrow). (**E**) These respective sets are fed into CLOvE, which computes the observed and null statistical tests in parallel. The results of which can be transformed into context x sample matrices, or visualized as a bipartite graph.

A context matrix of counts (e.g. SCNA, mutation calls) and a matrix of expression measurements (e.g. microarray, RNAseq counts) is read in from a comma or tab delimited file into a dataframe. Both matrices are coerced into the same *m* dimensions of samples by subsampling intersecting samples are present in both data sets. This is done to allow masking context samples completely onto expression samples. Additionally, the expression gene vector is subsampled to those genes also intersecting with the context gene vector. This is to ensure that no gene with expression measurements lacks context measurements. An example of

the latter is an important factor in prefiltering, or exclusion, of expression genes that are associated with an amplification signal. Such genes introduce a confounding effect, potentially leading to misattribution in the rise in expression, i.e. believing the increase expression is due to a compensatory mechanism, rather than an underlying amplification that is actually driving higher expression. Note that genes unique to the context dataset are not excluded, as having a larger m dimension of these genes does not affect this filtering step and includes more potential SL pairs.

After pre-filtering, CLOvE prompts the user for filtering parameters. These parameters filter out genes with very low variance in expression, σ^2, and genes with very low numbers, n, of either context positive or context negative samples. Filtering these genes is important for the statistical test the CLOvE later performs, as very low variance and small sample sizes may cause an increase in type I errors, with offending pairs scoring artificially high. The user selects cutoff thresholds based on graphs rendered when context and expression data are loaded. These graphs show the distribution of each parameter, and is marked with the first, second, and third standard deviation, in increasing order of stringency, as a guide. To ease the cutoff threshold decision, they may refer to a plot showing drop off in number of comparisons based on parameter selection (Figure 19). However, in a test of all combinations of parameters for both drop off and mean separation from null, default values of n=7 and v=0.2 offered the most optimal trade off (t=7.64, p=2.53e-14).

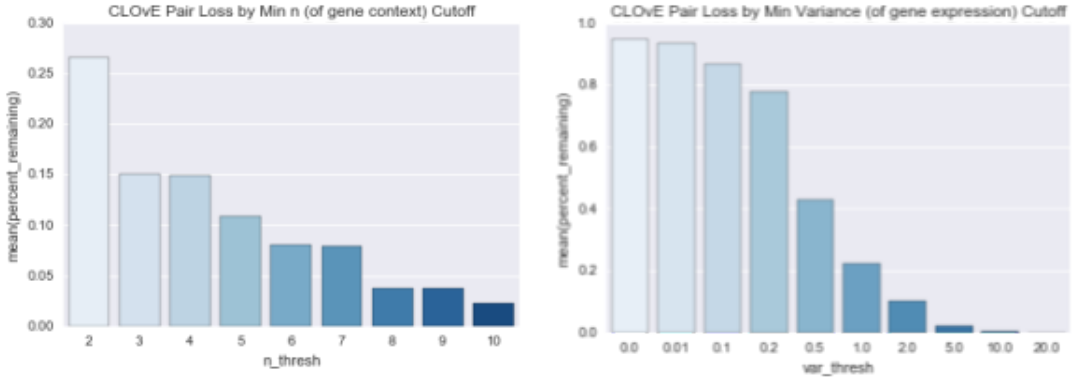

Figure 19. Pair drop-off across range of CLOvE parameters. Minimum variance threshold (left) and minimum context samples (right) are plotted against percent of pairs remaining. Choosing no filters would mean 100 percent remain, but the bars begin to show a decrease in the number of CLOvE pairs as the parameters get more stringent. These data were generated by performing CLOvE computations in breast cancer cell lines with each parameter, while holding the other parameter constant. Recommended default values are n=7 and v=0.2.

After selecting the appropriate filters for variance and population size, a copy of the expression matrix is permuted in order to generate a null distribution to perform significance testing on CLOvE. The null distribution was performed as above for the non-CLOvE and CLOvE comparison, but performed with a permuted RNA labels set from same tissue. Permutation of the labels is performed with the numpy permute module. This module is an implementation of the Mersenne Twister, a widely used pseudorandom number generator (Matsumoto and Nishimura 1998). The null distribution maintains structural variants like CNV while only permuting the a copy of the transcriptome in the same tissue. This effectively destroys any pattern of meaningful CLOvE pairing within tissue and represents the null hypothesis that expression does not change with copy-loss context.

Calculation

CLOve then begins calculation of both sets of context and expression matrices. This process is depicted in Figure 20. CLOvE seeks to determine if two expression vectors taken from different contexts are significantly different than each other using a variation of the Student's t-test called Welch's t-test (Welch 1947). Welch's t-test was chosen because it offers a more reliability over a Student's t-test when the two samples to be compared have unequal variances or unequal sample sizes, or both (Derrick, Toher, and White 2016; Ruxton 2006). However, like the Student's T-test, Welch's test also assumes that both samples are normally distributed, and this is true given the normalization procedure for expression data discussed previously and represented in Figure 20.A.

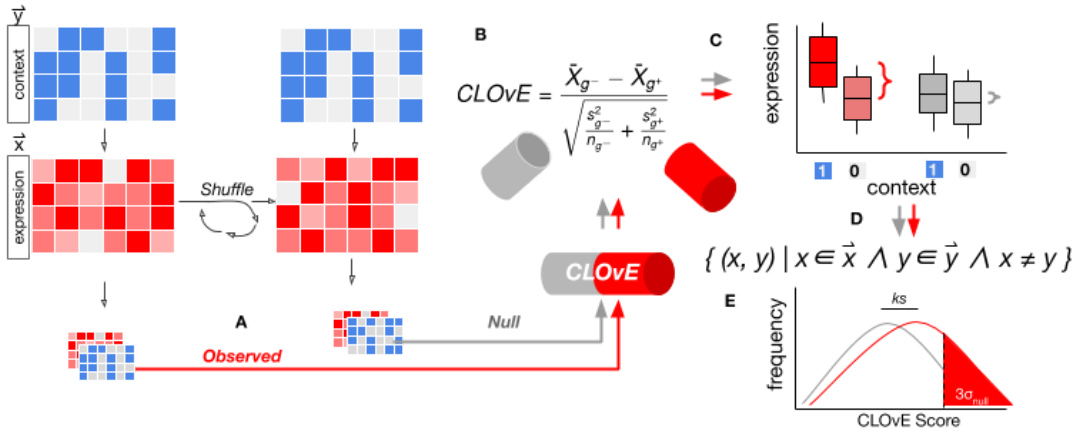

Figure 20. CLOvE calculation of context-dependent expression. (**A**) After shuffling, a filtered context and expression matrices are paired up in observed (red) and permuted (grey) sets. Each of these pairs contains a vector of *y* genes and *x* genes, corresponding to context and expression, respectively. (**B**) These are fed into CLOvE scoring expression in parallel. (**C**) CLOvE scoring is illustrated here at the gene level using a box plot. Here, expression difference of a single gene are compared between another gene's conditions having a loss context (context=1) and not having a loss context (context=2). (**D**) CLOvE is an iterative process, so this analysis is repeated for the cartesian product of gene vectors x and y to get all possible pairs. (**E**) These pairs are aggregated into distributions of CLOvE scores, converted into cumulative density functions (not shown), and a KS statistic is computed between them as a quality control step.

CLOvE therefore employs a Welch's t-test to compute a score for a pair of context and expression genes, where $\bar{x}$, S, and n are the sample mean, variance, and size, respectively (Figure 20.B). This t-test is computed using the stats package as part of SciPy (Oliphant 2007). Graphically, a single CLOvE analysis can be represented as a t-test performed across two groups of a box plot (Figure 20.C). However, CLOvE scores all possible combinations of genes that could exist in a potential compensatory pair. The set of compensatory pairs is represented by a Cartesian product of the context vector of genes with the expression vector genes, except in the case of self-links (Figure 20.D). Self-links are cis-acting links that cannot

represent a typical compensatory pair, since the deletion of a LoF cannot compensate for its own expression[2]. CLOvE pairs are then iteratively calculated over this set of possible pairs in both the null and permuted set to create two distributions of scores (Figure 20.E). To test the significance of the scores over all, the two distributions are transformed into cumulative distributions, and the non-parametric Kolmogorov-Smirnov (ks) test is performed between them to generate a D-value of difference and a p-value of significance.

Post-Processing

Along with the t-statistic for each pair, a p-value is also computed as the probability of observing a t-statistic at least as extreme as observed for that pair. This gives a large distribution of p-values, many of which will fall victim to what is commonly termed the multiple comparison problem. This problem is characterized by a prevalence of type 1 error, or a large number of false positives due to the relatively large number of tests that could receive exceptionally low p-values purely by chance. Therefore, it is important to adjust the significance of each pair using a modified "Bonferronitype" procedure described by Benjamini and Hochberg (Benjamini and Hochberg 1995). These adjusted p-values are included along with

[2] However, these cis-acting self-links could be an interesting source of essential genes that increase transcription depeding copy loss, and are investigated later.

the original p-values, in addition to the summary statistics from each pair for potential downstream analyses.

Some pairs may also be removed to limit the amount of confounding factors present in computed CLOvE scores. Some of these confounding factors are accounted for in the selection of tissue-specific samples at the outset, as mentioned previously. This limits the number of comparisons showing differential expression between tissue types, and increases the likelihood of top scoring CLOvE pairs to reflect actual compensatory changes. However, many cell types comprise molecularly and histologically distinct subtypes that need to also be accounted for. Part of CLOvE post-processing includes running a differential expression analysis between user-specified conditions, and eliminating compensatory pairs harboring any of the top k most significantly different genes.

Given the purpose of CLOvE to identify new cancer vulnerabilities, it becomes necessary to collapse pairs into single genes. In this scheme, a single score corresponds with a single gene in a specific sample, rather than the scored interaction between two genes in a pair calculated across all samples. This transformation of CLOvE scores is termed here as vulnerability vectorization. The resulting sample-specific vector is product between the CLOvE matrix, *C*, and the context vector, *y*, of a particular sample for the vulnerability vector, *v*, shown in Figure 21. This post-processing step is needed to measure the connectedness, or degree, that a particular gene has within a CLOvE interaction network. Higher

counts indicate more connections to a particular differentially expressed gene, and may indicate a compensatory mechanism to many simultaneous contexts in the cell. From a practical standpoint, vulnerability vectorization is also needed to transform CLOvE scores to a non-pairwise format that can be used to compare this approach against genes essentiality scores from other external sources for validation. These validation efforts are discussed later.

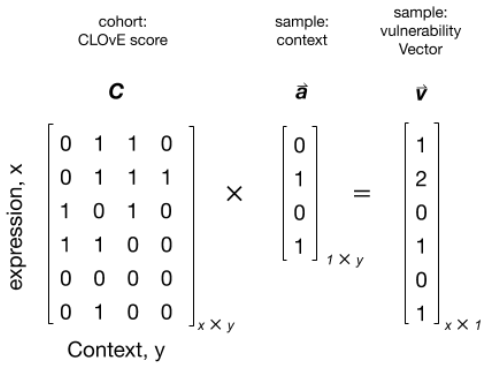

Figure 21. Vulnerability Vector represents degree of connections in a CLOvE network. CLOvE scores genes based on the differences in expression level of each gene in vector *x*, between the presence and absence of each context gene in vector *y* across a cohort of samples. This is represented as the CLOvE matrix *C*. A context vector *a* is referenced from a particular sample from that cohort. The product of matrix C with sample context vector, *a,* generates a vulnerability vector for that particular sample, *v*. Therefore, *v* represents the sum of all CLOvE relationships that tie to a specific gene's expression across all genes in *x*. Each score represents the connectedness, or degree, of a gene in a CLOvE network. Vulnerability vectors are important for understanding the relative importance of a gene to a particular sample.

Ultimately, CLOvE produces a bipartite graph comprising interactions between expression and context. In addition to a single CLOvE *t*-test for each pair, a p-value, and BH-adjusted p-value, as well as components that define Welch's t-test are also included for potential downstream analysis. These same results are

produced for the permuted set of CLOvE pairs that make up the null distribution. The comparison between the observed and null is an important first step in assessing the informational quality of the CLOvE approach.

Dependencies

CLOvE is run entirely in python 3.7 using common packages available through the Anaconda package management system (Analytics 2016). These are summarized in Table 3. CLOvE was developed in a Jupyter-IPython notebook, an interactive python environment. Pandas, a library of optimized python data structures, was used heavily in conjunction with computational libraries numpy and scipy. Graphing was handled with Matplotlib, with a Seaborn interface. All of the packages used are free and open source.

Package	Function	Citation	Website
Pandas	Data structures	(McKinney and Others 2010)	pandas.pydata.org
Numpy	General Computation	(Van Der Walt, Colbert, and Varoquaux 2011)	numpy.org
Scipy	Statistics	(Oliphant 2007)	scipy.org
Matplotlib	Graphing	(Hunter 2007)	matplotlib.org
IPython	Environment	(Pérez and Granger 2007)	ipython.org

Table 3. Python Packages used by CLOvE.

CHAPTER 4

Exploring CLOvE: Known Entities and Potential Suspects

Summary

- CLOvE shows a significant difference from random across different cell types
- Known synthetic lethal pairs have higher CLOvE scores compared to random
- CLOvE pairs occurring in-cis may also reveal dependencies
- Gene pairs involved in apoptosis pathways have high CLOvE scores
- Essentiality scores in RNAi and CRISPRi viability screens show agreement with CLOvE scores

QC and Validation

To assess the significance of the CLOvE analysis, it was necessary to compare the scores of observed pairs with the permuted distribution of CLOvE pairs. This comparison was done to test whether CLOvE pairs are informative, or if they merely reflect random noise. The first validation was performed in context selection, first in just breast cell lines. Breast cell lines were initially chosen because of their reasonably large population size on CCLE, well defined subtypes, and known SL pairs for comparison to CLOvE results.

The specific question was whether CLOvE could show us pairs that were significantly different than pairs computed on a permuted expression profile, the null distribution, over three context conditions. These context conditions included copy number, mutation, and copy number and mutation. The combinatory condition represents a logical *OR* operation between mutation and copy number, as previously described.

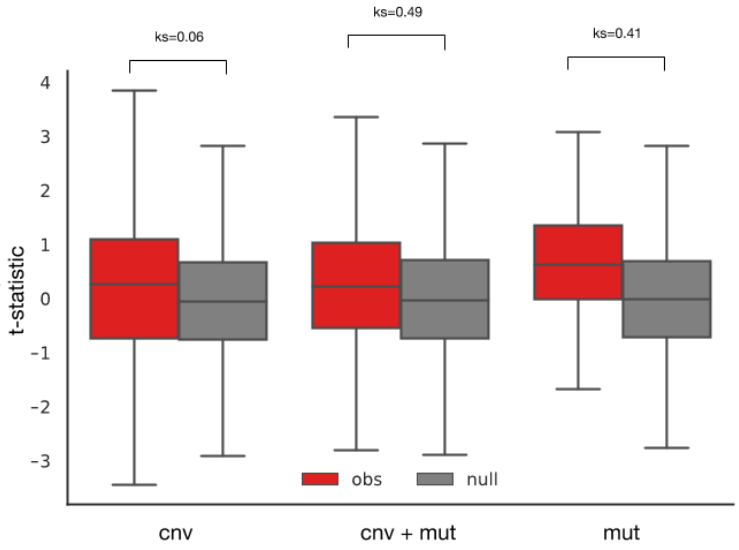

Figure 22. Comparison of possible CLOvE contexts. CLOvE was run in CCLE breast using binary context values of either only copy number (left), only mutation (right), and a combination of the two (center). In each context, a distribution of observed (red) and null (gray) *t*-statistics from the CLOvE output was generated, and compared within each group. In order of greatest separation from mean was copy number, mutation, and the combined context, with KS values of 0.06, 0.41, and 0.49, respectively ($p<0.001$). All observed values had positive medians, meaning higher expression on average when in the presence of a loss context. This shift in context-associated expression detected by CLOvE may provide some insight into possible compensatory mechanisms.

Comparing these three context conditions showed a meaningful separation from the null distribution calculated in each case. Such separation is represented graphically as a grouped boxplot, where all observed distributions show positive median *t*-statistics compared to the grouped null distributions from each condition (Figure 22). A positive score indicates a higher level of expression when another gene is lost than when it is not lost. Since each observed distribution has higher median *t*-statistics compared to random expression profiles, each of these three potential

context choices is telling us something informative about possible compensatory changes to the transcriptome, some of which may reflect SL interactions.

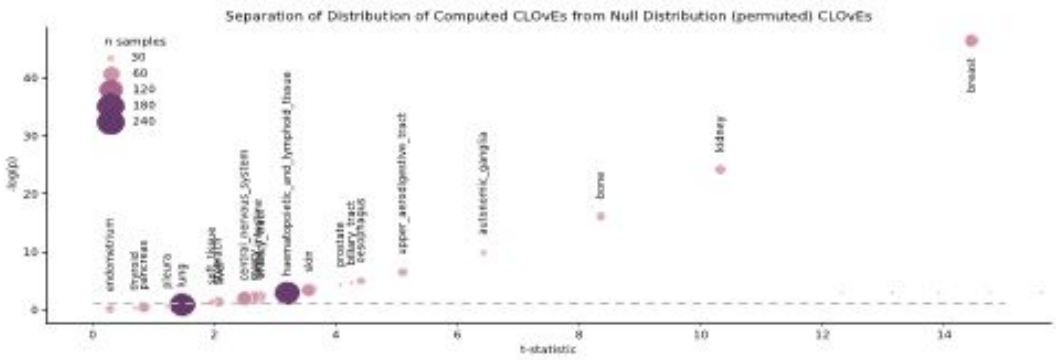

Figure 23. Observed versus null separation for CLOvE across 24 cell lines. For each cell's primary source, a KS test measured the separation of observed CLOvE scores from null CLOvE scores. The p-values of that separation are plotted here in negative log scale. A low p-value, or high -log(p), indicates a non-random relationship between context and expression, highlighting possible compensatory activity in each cell, as measured by CLOvE. Breast, kidney, and bone comprised the highest significance of all tissue comparison. A p-value threshold equivalent to p=0.01 is shown, where any sample above that line has a smaller, more significant p-value (dotted line). To show that this is not solely due to tissue population size, dots are scaled to the number of samples in that tissue.

Next, CLOvE was run in all available cell lines in CCLE by tissue of origin, to understand whether the approach can be generalized to other cell types besides breast. CLOvE was run in each of the 24 cell lines with enough samples to meet the requirements of the pre-filtering process. From this computation, CLOvE pairs and their differential scores from both the observed and null sets were placed into their respective distributions, and checked for similarity with a KS test. The p-value from each observed vs null KS test in each tissue was $-log_{10}$ transformed, with the highest transformed value representing the most significant deviation of the observed from the null distribution. The distribution of observed CLOvE scores

shows a significant difference from null across 20 of the 24 different cell types (Figure 23). Visually representing the differences in median observed CLOvE *t*-statistics compared to median permuted *t*-statistics, shows all observed medians are greater than their respective null medians and all greater than zero (Figure 24). The cell type where CLOvE has potentially the most utility is breast, as it has the most significant difference from the mean in terms of both distance from median to median and p-value from a KS test.

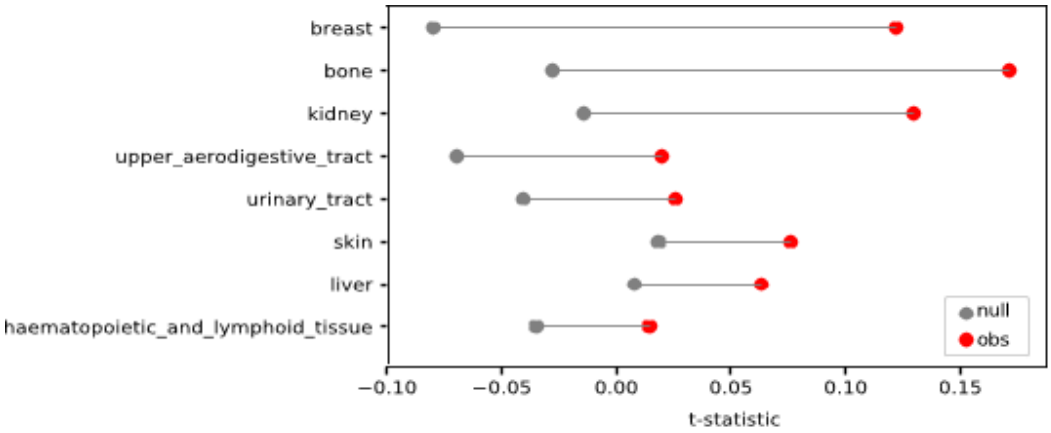

Figure 24. Comparison of tissue sources most significantly different from random. The top 8 tissues are displayed as a ball and stick plot of median *t*-test reported for the observed distribution (red) or null distribution (gray), ordered in decreasing order of p-value.

CLOvE pairs were further validated against known synthetic lethals. In this case, known synthetic lethals were pulled from the SynLethDB database of curated synthetic lethal pairs (Guo, Liu, and Zheng 2016). These results are summarized in Figure 25. An intersection of these values with the computed CLOvE values produced a distribution of known synthetic lethals. A sample of CLOvE scores that

did not appear in that database was subsampled to the same number of pairs, and are termed “potential” pairs. Similarly, a subsample of cis-acting pairs was also computed. Pairs in cis are defined here as being pairs where the compensatory gene is the same as the context gene. Of course, if a compensatory gene is completely lost, it cannot compensate for itself. Like in all previous analyses, the null distribution was also computed and randomly subsampled to the same number of pairs.

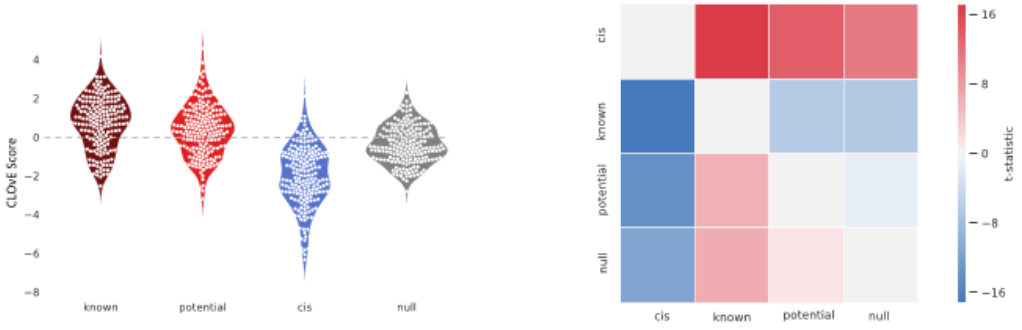

Figure 25. Comparison of CLOvE pairs to known synthetic lethals. (Left) CLOvE pairs that mapped to known SL pairs taken from SynLethDB, a repository for synthetic lethal interactions. Potential CLOvE pairs represent a subsample of all CLOvE pairs that are not known to be of a SL interaction. Cis pairs are those where the context gene is the same as the compensatory gene. The known distribution of CLOvE had a slightly higher median score from null (t=7.15, p<0.001) than did the distribution of potential CLOvE scores compared to null (t=6.72, p<0.001), since the known distribution of CLOvE scores is completely enriched for synthetic lethal pairs. This is encouraging, as known SL pairs would be expected to higher CLOvE scores. It is also expected that the cis-acting pairs would predominately less than zero, since copy or mutational loss would be expected to have a deleterious effect for the expression of that gene. However, some positive CLOvE scores from that set merit attention.

Known pairs tend to have higher median CLOvE score compared to SL pairs from the null distribution (t=7.15, p<0.001). This is encouraging, since CLOvE scores

would have to be higher in cases of known SL pairs if CLOvE is to be considered as effective in providing evidence of SL interactions. Potential CLOvE pairs also had a higher median score compared to null (t=6.72, $p<0.001$), suggesting that there may be some pairs that score highly in CLOvE that are unknown in the literature as being a compensatory or SL.

Anti-Apoptotic Subnetwork

Some of the highest scoring CLOvE pairs in breast tissue that are not recognized as known SL pairs, are found to participate in the apoptosis pathway. The fact that they exist in the same pathway supports the functional rationale as to how SL occurs. Looking at the most connected nodes in this sub network, CRIP1 (n=12) and ERBB3 (n=13), reveals an interesting relation to apoptosis.

ERBB3/HER3 encodes a gene for an epidermal growth factor (EGF) that is known to be highly expressed in some cancers, including the major subtypes of breast cancer. When it dimerizes to other EGFs, ERBB3 activates the PI3K/AKT/mTOR signaling pathways known to increase cell proliferation and differentiation (Carraway and Cantley 1994). One node connected to ERBB3 through a CLOvE in a pair is BCL2, a gene that codes for a pro-survival proteins. These BCL2 proteins prevent the release of caspase-9 and cytochrome c, effectively stalling apoptosis (Cory and

Adams 2002). From the CLOvE network, when BCL2 is lost, ERBB3 increases in expression. Since cell life is effectively shortened in BCL2 deficient cells, it is plausible that cells must then rely on ERBB3 and the anti-apoptotic pathway it promotes. This compensatory mechanism, identified by CLOvE, is the mechanism of action behind patritumab, lumretuzumab, istiratumab, and 12 other current investigational compounds to target ERBB3/HER3 (Fitzgerald et al. 2014; Kawakami et al. 2014; Mishra et al. 2018; Schneeweiss et al. 2018).

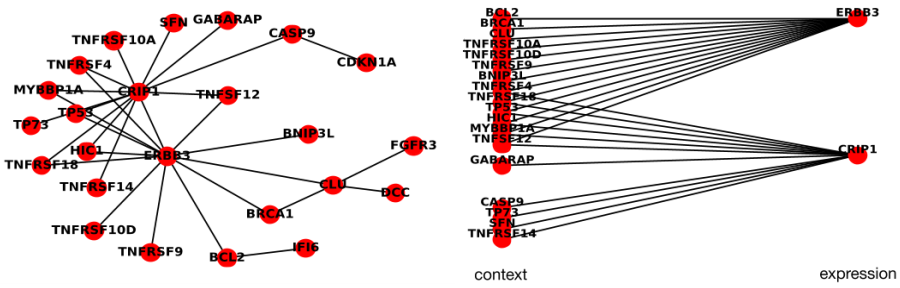

Figure 26. Significant CLOvE pairs in genes functionally related by apoptosis. (Left) The nodes are genes and the edges, or lines, that connect them depict CLOvE interactions of significance ($p < 0.01$) after correcting p-values multiple tests. These connections are occurring in genes not known to be synthetically lethal with each other. (Right) Interactions are transformed into a bipartite graph showing each node as belonging to context or expression. All genes are still involved in the apoptotic pathway.

CLOvE in-cis

CLOvE pairs occuring in-cis do not necessarily meet the canonical definition of synthetic lethal, which states that the loss states of two different genes must be

mutually exclusive. In the case of cis-CLOvE, this is where the loss of a gene is associated with an increase of its very own expression. The first experiment to identify cis-CLOvE pairs was modified to only include hemizygous loss events from the copy number data. This was done to reflect a hemizygous or leaky context loss.

As expected, a majority (n=2982) of the total usable cis pairs (n=3268) had a negative CLOvE score, reflecting a loss of expression during a context loss event in that same gene (Figure 27). The remaining 9% of usable samples (n=286) were found to have positive CLOvE scores, reflecting an increase in expression with partial loss of that gene. Upon correcting p-values, the most significantly positive gene was BACH2 (t=5.18, p=0.000139), which plays a role in cancer.

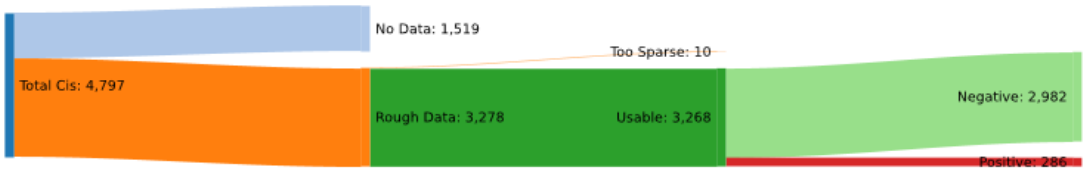

Figure 27. Description of cis-CLOvE pairs. Restricting the copy number to just cases of hemizygous loss, allows for potential cases of a leaky essential phenotype. Running CLOvE against this data produced 4797 pairs that were ultimately reduces to 3268 pairs due to reasons of missing data and sparsity. As expected, of those usable cis-CLOvE pairs, a 91% majority were negative (n=2982). A 9% minority of positive cis-CLOvE pairs (n=286) indicates a case where partial loss of a gene is associated with its expression gain.

BTB Domain And CNC Homolog 2, or BACH2, is a leucine zipper transcription factor (S. Sasaki et al. 2000). BACH2 coordinates transcription activation or repression along with another transcription factor, MAFK (Oyake et al. 1996).

Although overexpression of BACH2 has been found to increase sensitivity to some chemotherapies response, some studies have shown that tumor growth is impaired in mice deficient in BACH2 (Kamio et al. 2003; Roychoudhuri et al. 2016). In these BACH2-deficient mice, it was found that BACH2 inhibited CD8+ T-cells and the cytokine IFN-γ in the tumor microenvironment, effectively operating as a tumor promoter and immunosuppressive factor.

There are many explanations for why BACH2 is revealing itself as a high scoring cis-CLOvE. These could include sample heterogeneity, selective pressure from chemotherapies and other tumor factors. However, a leaky phenotype is a possible explanation of its status as a cis-CLOvE. This phenotype could be the result of a hemizygous loss of BACH2, causing a decrease in tumorsupressive ability, and favoring cells that can respond by increasing the level of these transcripts. Although exploring the possible mechanism is beyond the scope of CLOvE, it illustrates some of the insights to be gained from different aspects of this approach.

Congruency

Synthetic lethal pairs of genes which are related by function or pathway tend toward higher rates of essentiality. For example, PARP1 and BRCA1 are both involved in recombinatorial repair, MTOR and PTEN both regulate protein kinase B signalling,

and ATR and MYC are both involved in DNA damage response, to name a few. Though not all SL pairs have such an easily defined functional relationship, it was worthwhile to test if observed CLOvE scores, as a whole, had higher scores between genes in have higher CLOvE scores than unrelated genes. Of those related genes, it would be expected that two related genes would have similar CLOvE partners. Comparing similarities between the SL pairs to which two genes belong is the idea of congruence.

Thinking of the ancient Sanskrit proverb “the enemy of my enemy is my friend,” consider a SL network as a social network of enemies: if two entities share common enemies, they can be considered “friends” (Rangarajan 1992). This idea of congruence was made popular when using SL networks to infer compensatory pathways and gene function from congruent genes in yeast genomics (Ma, Tarone, and Li 2008; Ye et al. 2005). Here, congruent networks add a layer of corroborative evidence to SL candidates that serve a compensatory function across a well defined gene module or pathway.

BRCA and PARP can be corroborated by investigating the congruent network between the BRCA genes which cause PARP compensation. BRCA1 and BRCA2, both damage repair genes, are outliers with a pearson coefficient of 0.35 compared to their mean out-of-module correlation of 0.073 . This expected result shows CLOvE can identify genes with similar function, as BRCA1 and BRCA2 are both

"friends" that serve to repair double-strand break repair, with a common "enemy" of PARP.

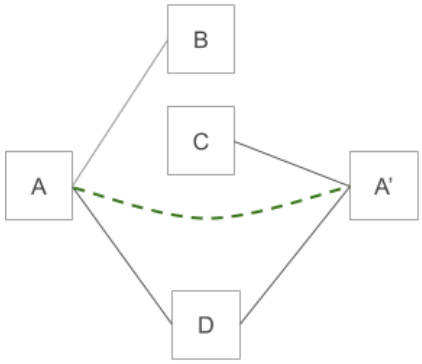

Figure 28. Congruent networks show degree of similarity between genes. A putative SL network shows SL pairs A-B, A-D, A'-C and A'-D. The congruence calculated as the pearson correlation between the CLOvE scores in common between A and A' (green dotted box).

Applying a congruence filter of above 0.9 revealed the top hit of HRAS (a proto-oncogene) and SDCCAG8 (Serologically Defined Colon Cancer Antigen 8). HRAS-SDCCAG8 was identified as a hit in both SynLethDB and in an RNAi screen that investigated SL relationships in RAS mutants (J. Luo et al. 2009). This shows the utility of combining multiple networks with other corroborating evidence to identify SL interactions.

Viability

CLOvE shows differential, context-dependent changes in expression, but are these expression changes at all indicative of essentiality? To help answer this question, experimental data from viability experiments was correlated with CLOvE scores to measure the strength of association between CLOvE score and essentiality. The first data we used to compare these breast cancer cell lines was DEMETER.

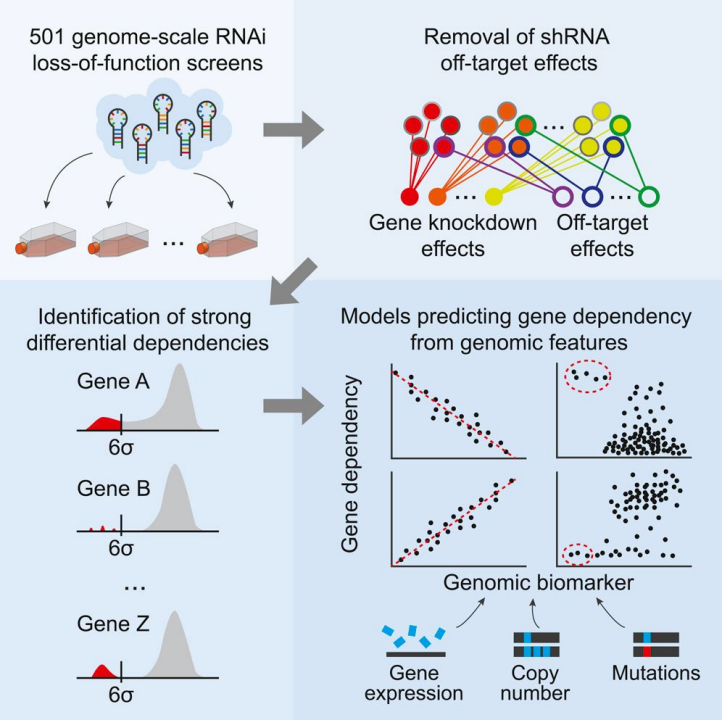

Figure 29. The DEMETER approach used to validate CLOvE. This approach aggregates RNAi data, segregates on-target effects from off-target effects, and uses a predictive model to assign dependency scores for each gene. RNAi experiment data from 501 cell lines was ultimately used to identify 769 dependencies. Reproduced from "Defining a Cancer Dependency Map" (Tsherniak et al. 2017)

DEMETER is an essentiality score based on viability data from cell-based screens in RNAi (Tsherniak et al. 2017). These screens were systematic attempts to identify genes in cancer essential for survival and proliferation, but were often confounded by off-target effects (Birmingham et al. 2006; Cowley et al. 2014; B. Luo et al. 2008; Marcotte et al. 2012, 2016). Rather than use these data directly to validate CLOvE, computational approaches like DEMETER offers a computational approach to aggregate these data, quantify on-target vs off-target effects, and predict essentiality (Tsherniak et al. 2017). A similar algorithm and its viability scores, ATARiS, was initially considered as a possible validation set (Shao et al. 2013). However, ATARiS was ultimately eschewed in favor for DEMETER, since DEMETER models miRNA seed sequences that are often the source of shRNA off-target effects, and does so without the need for prior knowledge.

To relate DEMETER scores with CLOvE, CLOvE scores were transformed to fit the same data structure as DEMETER. Pairwise scores from CLOvE were collapsed from a network to a single vulnerability vector, as explained previously, to produce a single score for each gene in each sample. This score is a count of the number of connections, or degree, which a particular gene has in the CLOvE network. Higher counts indicate more connections to a particular differentially expressed gene, and may indicate a compensatory mechanism to many simultaneous contexts in the cell.

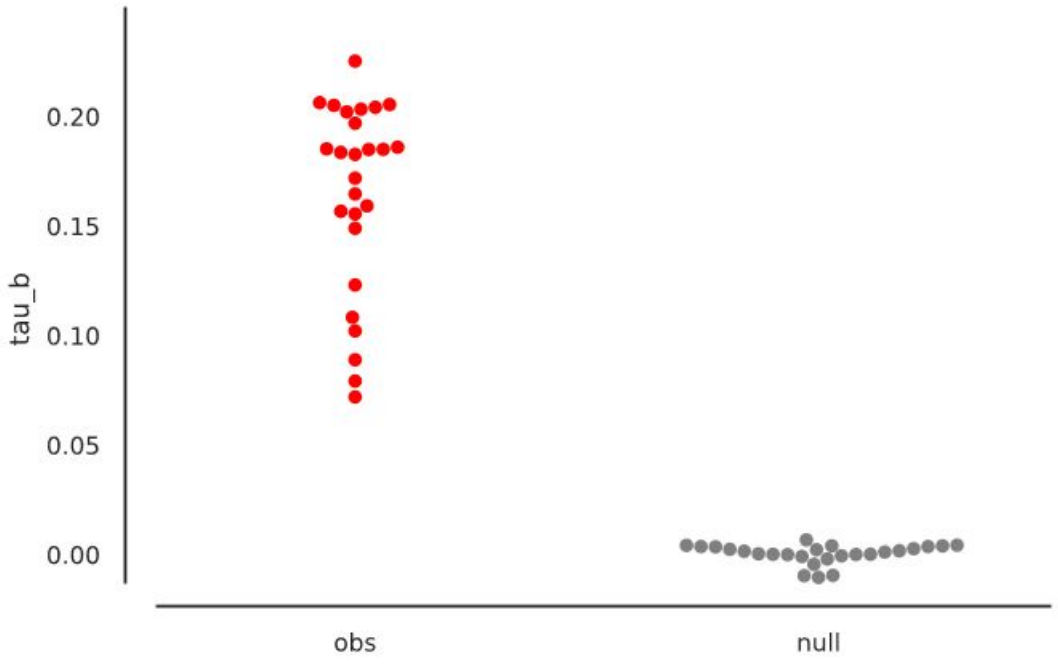

Figure 30. Correlation of CLOvE scores with DEMETER essentiality. CLOvE pairs transformed to vulnerability vectors were correlated with RNAi-based essentiality scores from the DEMETER computational model. These comparisons were made in each cell line, correlating the vulnerability vector with the DEMETER score using Kendall's tau-b, a rank-based concodence measurement. Cell-to-cell comparisons in observed CLOvE pairs (red), and permuted pairs (gray) are graphed above, where each dot represents a single cell line. The concordance for the observed CLOvE vulnerabilities is very high with DEMETER essentiality, and nearly zero in the null model (*t*=19.1, p=1.15e-24).

A subsample of intersecting breast cell lines were selected from both DEMETER and CLOvE, of which each was subsampled again for an intersecting set of genes. In an iterative approach, Kendall's tau-b, a rank-based correlation measurement, was applied to test the association between gene vulnerability in CLOvE and in DEMETER. This correlation was also repeated in the null dataset formed from the permuted CLOvE set (Figure 30). This demonstrated a positive association (τ_B>0)

between CLOvE, in strong contrast to the null distribution(t=19.1, p=1.15e-24). This indicates a non-random association between the degree of CLOvE vulnerability and DEMETER essentiality scores.

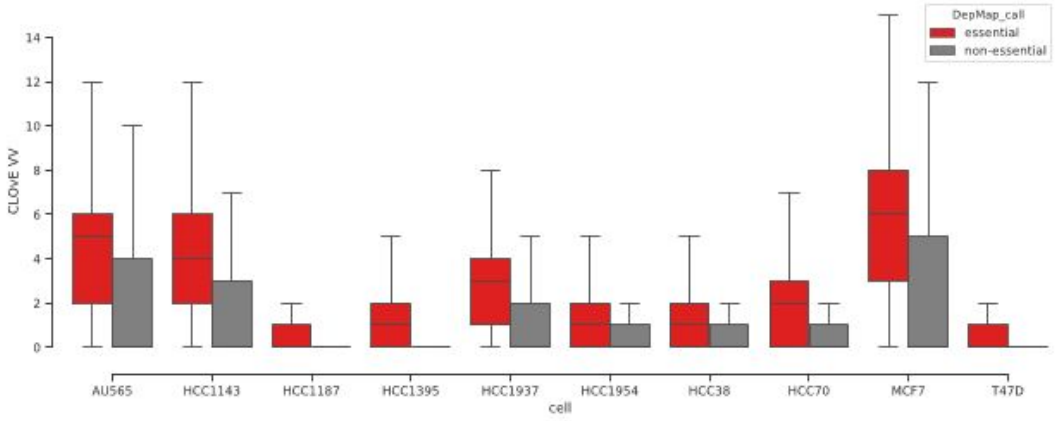

Figure 31. Comparison of CLOvE Vulnerability Vector to DepMap essential-status. CLOvE vulnerability vectors were computed for all breast cancer cell lines that intersect with those found in DepMap, a compendium of cancer dependency experiments using CRISPR and small molecule. Single gene vulnerability vectors were compared between groups of genes considered essential by DepMap (red), vs those considered to be non-essential (gray). All of the 10 intersecting cell lines had markedly higher VV counts for genes considered essential vs those considered non-essential (*t*=19.1, p=1.25e-80). This supports the tendency of high CLOvE VV scores to be associated with essentiality.

Assessing the association between CLOvE vulnerability vectors and gene essentiality continued at the level of the cell. Degrees of vulnerability derived from CLOvE were compared in genes determined be to essential or nonessential, according to binarized essentiality calls from the Achilles CRISPR interference dataset (Aguirre et al. 2016; Han et al. 2017). Comparing distributions of CLOvE vulnerabilities in essential versus nonessential genesets show a clear separation ((t=19.1, p=1.25e-80), with essential genes having a higher median CLOvE vulnerability than non-essential genes (Figure 31). Since this comparison is limited

due to availability of breast cell lines available in this dataset, it is necessary to repeat this analysis across all cell lines. Despite this limitation, it is clear that CLOvE vulnerability vectors associate with essentiality. However, more validation is needed.

CHAPTER 5

Continuing CLOvE: Can it recapitulate *in-vitro*?

Summary

- CLOvE requires biologically relevant validation for high-scoring, yet uncharacterized gene pairs
- A possible *in-vitro* RNAi-based phenotypic screen is proposed
- A possible *in-vitro* CRISPR-based phenotypic screen is proposed

Motivation

CLOvE was developed with the purpose to identify new essential genes in cancer based on synthetic lethality. To satisfy this purpose, future directions for CLOvE must include validating the synthetic lethal candidates found in this approach. These can first take the form of continued validation in-silico, but ultimately must extend to validation in biological systems in-vitro, or possibly in-vivo. This chapter proposes these next steps for future development of CLOvE.

Pre-Screen

CLOvE pairs have already been compared to compendiums of viability data begot from RNAi and CRISPR knockdown viability experiments. Many other datasets exist that use the same population of samples, CCLE cell lines, in small molecule screens that also test for viability. The Cancer Therapeutics Response Portal (CTRP) is one such compendium of small molecule viability data. The CTRP comprises annotated viability data for 481 compounds tested in 860 cell lines, in addition to statistical tools to perform pathway enrichment analysis, cell and sample clustering, and sensitivity correlations (Basu et al. 2013; Rees et al. 2016; Seashore-Ludlow et al. 2015). Using this data could be beneficial in relating CLOvE scores of particular high-scoring genes to compound viability in those same cell lines.

The first step would be to identify cell lines with copy loss contexts that are significantly associated with increased expression of another compensatory gene. Within this group of cell lines, examine the viability data for compounds targeting each gene product involved in all CLOvE pairs. Computing the correlation for all CLOvE scores against the viability measurements like AUC could reveal domains of the CLOvE distribution that yield the most predictive power in inferring drug sensitivity.

RNAi

Due to limitations in size of the CTRP, many significant genes implicated by CLOvE will not have viability data reported in the cell lines of interest. In this case, performing a cell-based phenotypic viability screen would be beneficial. This screen could be characterized as a chemogenomic, or a screen using selected panels of small molecule inhibitors to create a viability loss in a knockout cell. This screen would use either pooled shRNA or siRNA to silence genes of a particular context in cell lines that constitutively express that gene. Both groups of cells would then be treated with a library of compounds in serial dilution to produce dose response curves using a viability assay. Particular aspects of this experiment are discussed below and are shown in Figure 32.

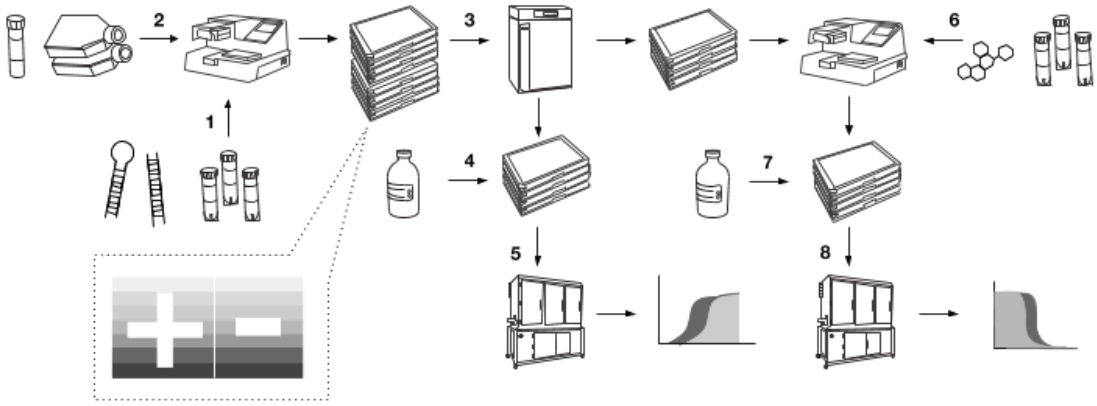

Figure 32. Schematic of possible RNAi validation screen. 1) shRNA or siRNA is loaded with transfection reagent onto multiwell plates. 2) Cell lines, previously frozen, are taken out of culture once stabilized and plates into the same multiwell plates to transfect. 3) Plates are made in duplicate and placed into an incubator for up to 3 days. 4) One set of plates are removed to check growth curves and get a neutral control for luminescence with added reagent, and then 5) read and analyze those growth plates. 6) To the second set of transfected plates, add a library of compounds in serial dilution. 7) Let the compound incubate briefly, then add a viability reagent, and 8) visualize the differences in the dose response curve created. (Inset) a model plate map is included, with half the columns dedicated to knockdown transfection vs no-knockdown transfection, and the rows in order of increasing compound concentration from top to bottom of plate.

Screens will be performed in either 96- or 384-well assay plates, and treatment conditions will be transferred from similarly-formatted library plates pre-stamped with commercially available compounds or RNAi. In either screening scenario, the treatment will be added to plates pre-seeded with cells 24-36 hours prior. Lipofectamine®, a lipid-based transfection reagent, will be chosen for an RNAi screen due to its reproducibility and ease of use. Treatment incubation will commence for an additional 24-36 hours before the endpoint assay. The assay endpoint, cell viability, will be assessed with luminescence using Cell-Titer Glo®, or fluorescence using LIVE/DEAD® Cell Vitality stain.

Though many experimental protocols exist in the literature, some assay development is required to optimize the experimental conditions of the screen. Such variables include RNAi/compound concentration, length of treatment incubation, and selection of controls. Compound screens could include staurosporine as a positive control, and vehicle (DMSO) as a negative control. RNAi screens could include common housekeeping genes (POLR2A, AKT, or PLK1) as positive controls, and commercially available non-targeting siRNA as negative controls. Non-targeting siRNA should be validated as having off-target proliferation loss of no less than 90% compared to transfection vehicle alone.

CRISPR

Using CRISPR interference may be a better alternative to the RNAi approach detailed above, mostly for the reduction in off target effects compared to RNAi. Such an experiment would include selecting cell lines that are copy-deficient for a particular gene. There are many cell types known for specific lesions, such as deficiency of VHL (von Hippel-Lindau, a well known SL context partner), which is often found in clear cell renal carcinomas. However, finding these cell lines for a lesser known SL gene would involve examining the context data for mutational or SCNA calls in that gene.

Whole genome libraries exist for sgRNA that have been well validated, including AVANA, which could be used for screening the possible compensatory gene (Doench et al. 2016). However, a focused approach would be favored over a library wide-approach, especially in the case of a combined small molecule screen. In such an experiment, CRISPR interference creates the loss-context and the small molecule inhibits the compensatory gene.

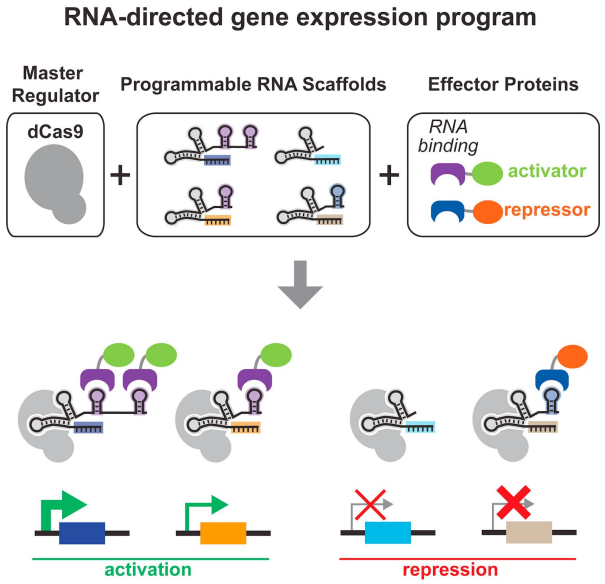

Figure 33. Transcriptional programming with engineered scRNA. Engineered CRISPR scRNA can be used for simultaneous activation and repression. Reproduced from "Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds" (Zalatan et al. 2015).

An extension of a basic CRISPR screen would test the viability effect of altering levels of context loss or expression gain. Specially-engineered CRISPR scaffold RNA (scRNA) have been described to offer variable levels of both activation and

repression (Zalatan et al. 2015). With the addition of different activators or repressors to the scRNA, variable levels of activation or repression can be achieved. The possibility for fine tuning these transcriptional programs increases with simultaneous activation and repression. For example, a scRNA experiment could repress context gene BCL2 (Com-KRAB repressor). This repression would occur simultaneously with repressing ERBB3 (KRAB repressor), partially activating ERBB3 (MCP or PCP activator), and fully activating ERBB3 (multivalent MCP+PCP activators). Measuring the effect of viability changes on these conditions for a variety of contexts and compensatory genes could reveal the molecular basis for SL sensitivity.

Conclusion

CLOvE identifies non-random, context dependent expression changes. The expression changes found with CLOvE have included genes that are known synthetic lethals. High scoring CLOvE pairs are also enriched in genes that participate in the same functional pathways. Finally, CLOvE scores and their resulting vulnerability vectors are seen to associate with essentiality, as validated in phenotypic screening data performed in with both RNAi and CRISPR. Some potential synthetic lethals have been discussed, especially those participating in the same functional pathways. However, more validation is needed, not just for the CLOvE approach itself, but for the uncharacterized pairs at the far right edge. Here

in this far flung tail may lie a druggable target, one that can be recapitulated *in-vitro*, *in-vivo*, and perhaps, one day, in a patient.

Bibliography

Adzhubei, Ivan A., Steffen Schmidt, Leonid Peshkin, Vasily E. Ramensky, Anna Gerasimova, Peer Bork, Alexey S. Kondrashov, and Shamil R. Sunyaev. 2010. “A Method and Server for Predicting Damaging Missense Mutations.” *Nature Methods* 7 (4): 248–49.

Aguirre, Andrew J., Robin M. Meyers, Barbara A. Weir, Francisca Vazquez, Cheng-Zhong Zhang, Uri Ben-David, April Cook, et al. 2016. “Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting.” *Cancer Discovery* 6 (8): 914–29.

Allen, David D., Raúl Caviedes, Ana María Cárdenas, Takeshi Shimahara, Juan Segura-Aguilar, and Pablo A. Caviedes. 2005. “Cell Lines as in Vitro Models for Drug Screening and Toxicity Studies.” *Drug Development and Industrial Pharmacy* 31 (8): 757–68.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. “Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs.” *Nucleic Acids Research* 25 (17): 3389–3402.

Analytics, Continuum. 2016. “Anaconda Software Distribution.” *Computer Software Vers*, 2–2.

Antoniou, A., P. D. P. Pharoah, S. Narod, H. A. Risch, J. E. Eyfjord, J. L. Hopper, N. Loman, et al. 2003. “Average Risks of Breast and Ovarian Cancer Associated

with BRCA1 or BRCA2 Mutations Detected in Case Series Unselected for Family History: A Combined Analysis of 22 Studies." *American Journal of Human Genetics* 72 (5): 1117–30.

Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25 (1): 25–29.

Ashworth, Alan, Christopher J. Lord, and Jorge S. Reis-Filho. 2011. "Genetic Interactions in Cancer Progression and Treatment." *Cell* 145 (1): 30–38.

Badve, Sunil, Chirayu Goswami, Yesim Gökmen-Polar, Robert P. Nelson Jr, John Henley, Nick Miller, Narjis A. Zaheer, et al. 2012. "Molecular Analysis of Thymoma." *PloS One* 7 (8): e42669.

Bajrami, Ilirjana, Asha Kigozi, Antoinette Van Weverwijk, Rachel Brough, Jessica Frankum, Christopher J. Lord, and Alan Ashworth. 2012. "Synthetic Lethality of PARP and NAMPT Inhibition in Triple-Negative Breast Cancer Cells." *EMBO Molecular Medicine* 4 (10): 1087–96.

Balis, Frank M. 2002. "Evolution of Anticancer Drug Discovery and the Role of Cell-Based Screening." *Journal of the National Cancer Institute* 94 (2): 78–79.

Bamford, S., E. Dawson, S. Forbes, J. Clements, R. Pettett, A. Dogan, A. Flanagan, et al. 2004. "The COSMIC (Catalogue of Somatic Mutations in Cancer) Database and Website." *British Journal of Cancer* 91 (2): 355–58.

Barretina, Jordi, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A. Margolin, Sungjoon Kim, Christopher J. Wilson, et al. 2012. "The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity."

Nature 483 (7391): 603–7.

Bassik, Michael C., Martin Kampmann, Robert Jan Lebbink, Shuyi Wang, Marco Y. Hein, Ina Poser, Jimena Weibezahn, et al. 2013. “A Systematic Mammalian Genetic Interaction Map Reveals Pathways Underlying Ricin Susceptibility.” *Cell* 152 (4): 909–22.

Basu, Amrita, Nicole E. Bodycombe, Jaime H. Cheah, Edmund V. Price, Ke Liu, Giannina I. Schaefer, Richard Y. Ebright, et al. 2013. “An Interactive Resource to Identify Cancer Genetic and Lineage Dependencies Targeted by Small Molecules.” *Cell* 154 (5): 1151–61.

Bellissimo, Teresa, Federica Ganci, Enzo Gallo, Andrea Sacconi, Claudia Tito, Luciana De Angelis, Claudio Pulito, et al. 2017. “Thymic Epithelial Tumors Phenotype Relies on miR-145-5p Epigenetic Regulation.” *Molecular Cancer* 16 (1): 88.

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 57 (1): 289–300.

Bergers, G., D. Hanahan, and L. M. Coussens. 1998. “Angiogenesis and Apoptosis Are Cellular Parameters of Neoplastic Progression in Transgenic Mouse Models of Tumorigenesis.” *The International Journal of Developmental Biology* 42 (7): 995–1002.

Bergh, N. P., P. Gatzinsky, S. Larsson, P. Lundin, and B. Ridell. 1978. “Tumors of the Thymus and Thymic Region: I. Clinicopathological Studies on Thymomas.” *The Annals of Thoracic Surgery* 25 (2): 91–98.

Birmingham, Amanda, Emily M. Anderson, Angela Reynolds, Diane Ilsley-Tyree, Devin Leake, Yuriy Fedorov, Scott Baskerville, et al. 2006. "3′ UTR Seed Matches, but Not Overall Identity, Are Associated with RNAi off-Targets." *Nature Methods* 3 (3): 199–204.

Boeckmann, Brigitte, Amos Bairoch, Rolf Apweiler, Marie-Claude Blatter, Anne Estreicher, Elisabeth Gasteiger, Maria J. Martin, et al. 2003. "The SWISS-PROT Protein Knowledgebase and Its Supplement TrEMBL in 2003." *Nucleic Acids Research* 31 (1): 365–70.

Boone, Charles, Howard Bussey, and Brenda J. Andrews. 2007. "Exploring Genetic Interactions and Networks with Yeast." *Nature Reviews. Genetics* 8 (6): 437–49.

Boyle, Colin F., Carol Levin, Arian Hatefi, Solange Madriz, and Nicole Santos. 2015. "Achieving a 'Grand Convergence' in Global Health: Modeling the Technical Inputs, Costs, and Impacts from 2016 to 2030." *PloS One* 10 (10): e0140092.

Bray, Freddie, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. 2018. "Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." *CA: A Cancer Journal for Clinicians* 68 (6): 394–424.

Bridges, Calvin Blackman, Katherine Suydam Brehme, and Others. 1944. "The Mutants of Drosophila Melanogaster." *The Mutants of Drosophila Melanogaster.*, no. 552. http://www.cabdirect.org/abstracts/19451601593.html.

Bridges, C. B. 1922. "The Origin of Variations in Sexual and Sex-Limited Characters." *The American Naturalist*. https://www-journals-uchicago-edu.oca.ucsc.edu/doi/pdfplus/10.1086/279847.

Brough, Rachel, Jessica R. Frankum, Sara Costa-Cabral, Christopher J. Lord, and Alan Ashworth. 2011. “Searching for Synthetic Lethality in Cancer.” *Current Opinion in Genetics & Development* 21 (1): 34–41.

Bryant, Helen E., Niklas Schultz, Huw D. Thomas, Kayan M. Parker, Dan Flower, Elena Lopez, Suzanne Kyle, Mark Meuth, Nicola J. Curtin, and Thomas Helleday. 2005. “Specific Killing of BRCA2-Deficient Tumours with Inhibitors of poly(ADP-Ribose) Polymerase.” *Nature* 434 (7035): 913–17.

Byrne, Alexandra B., Matthew T. Weirauch, Victoria Wong, Martina Koeva, Scott J. Dixon, Joshua M. Stuart, and Peter J. Roy. 2007. “A Global Analysis of Genetic Interactions in Caenorhabditis Elegans.” *Journal of Biology* 6 (3): 8.

Cairns, B. R., Y. J. Kim, M. H. Sayre, B. C. Laurent, and R. D. Kornberg. 1994. “A Multisubunit Complex Containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 Gene Products Isolated from Yeast.” *Proceedings of the National Academy of Sciences of the United States of America* 91 (5): 1950–54.

Cancer Genome Atlas Network. 2012. “Comprehensive Molecular Portraits of Human Breast Tumours.” *Nature* 490 (7418): 61–70.

Carraway, K. L., 3rd, and L. C. Cantley. 1994. “A Neu Acquaintance for erbB3 and erbB4: A Role for Receptor Heterodimerization in Growth Signaling.” *Cell* 78 (1): 5–8.

Castleman, B. 1966. “The Pathology of the Thymus Gland in Myasthenia Gravis.” *Annals of the New York Academy of Sciences* 135 (1): 496–505.

Cheng, Suzanne, Walter H. Koch, and Lin Wu. 2012. “Co-Development of a Companion Diagnostic for Targeted Cancer Therapy.” *New Biotechnology* 29

(6): 682–88.

Chen, Yuanbin, Helen Gharwan, and Anish Thomas. 2014. “Novel Biologic Therapies for Thymic Epithelial Tumors.” *Frontiers in Oncology* 4 (May): 103.

Cheung, Hiu Wing, Glenn S. Cowley, Barbara A. Weir, Jesse S. Boehm, Scott Rusin, Justine A. Scott, Alexandra East, et al. 2011. “Systematic Investigation of Genetic Vulnerabilities across Cancer Cell Lines Reveals Lineage-Specific Dependencies in Ovarian Cancer.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (30): 12372–77.

Ciriello, Giovanni, Ethan Cerami, Chris Sander, and Nikolaus Schultz. 2012. “Mutual Exclusivity Analysis Identifies Oncogenic Network Modules.” *Genome Research* 22 (2): 398–406.

Conde-Pueyo, Nuria, Andreea Munteanu, Ricard V. Solé, and Carlos Rodríguez-Caso. 2009. “Human Synthetic Lethal Inference as Potential Anti-Cancer Target Gene Detection.” *BMC Systems Biology* 3 (December): 116.

Conti-Fine, Bianca M., Monica Milani, and Henry J. Kaminski. 2006. “Myasthenia Gravis: Past, Present, and Future.” *The Journal of Clinical Investigation* 116 (11): 2843–54.

Corley, Susan M., Cesar P. Canales, Paulina Carmona-Mora, Veronica Mendoza-Reinosa, Annemiek Beverdam, Edna C. Hardeman, Marc R. Wilkins, and Stephen J. Palmer. 2016. “RNA-Seq Analysis of Gtf2ird1 Knockout Epidermal Tissue Provides Potential Insights into Molecular Mechanisms Underpinning Williams-Beuren Syndrome.” *BMC Genomics* 17 (June): 450.

Cory, Suzanne, and Jerry M. Adams. 2002. “The Bcl2 Family: Regulators of the

Cellular Life-or-Death Switch." *Nature Reviews. Cancer* 2 (9): 647–56.

Costanzo, Michael, Anastasia Baryshnikova, Jeremy Bellay, Yungil Kim, Eric D. Spear, Carolyn S. Sevier, Huiming Ding, et al. 2010. "The Genetic Landscape of a Cell." *Science* 327 (5964): 425–31.

Cowley, Glenn S., Barbara A. Weir, Francisca Vazquez, Pablo Tamayo, Justine A. Scott, Scott Rusin, Alexandra East-Seletsky, et al. 2014. "Parallel Genome-Scale Loss of Function Screens in 216 Cancer Cell Lines for the Identification of Context-Specific Genetic Dependencies." *Scientific Data* 1 (September): 140035.

Cufi, Perrine, Patrick Soussan, Frédérique Truffault, Rachid Fetouchi, Marieke Robinet, Elie Fadel, Sonia Berrih-Aknin, and Rozen Le Panse. 2014. "Thymoma-Associated Myasthenia Gravis: On the Search for a Pathogen Signature." *Journal of Autoimmunity* 52 (August): 29–35.

Dalmau, J., H. M. Furneaux, C. Cordon-Cardo, and J. B. Posner. 1992. "The Expression of the Hu (paraneoplastic Encephalomyelitis/sensory Neuronopathy) Antigen in Human Normal and Tumor Tissues." *The American Journal of Pathology* 141 (4): 881–86.

Dedes, Konstantin J., Daniel Wetterskog, Ana M. Mendes-Pereira, Rachael Natrajan, Maryou B. Lambros, Felipe C. Geyer, Radost Vatcheva, et al. 2010. "PTEN Deficiency in Endometrioid Endometrial Adenocarcinomas Predicts Sensitivity to PARP Inhibitors." *Science Translational Medicine* 2 (53): 53ra75.

DeLuna, Alexander, Kalin Vetsigian, Noam Shoresh, Matthew Hegreness, Maritrini Colón-González, Sharon Chao, and Roy Kishony. 2008. "Exposing the Fitness

Contribution of Duplicated Genes." *Nature Genetics* 40 (5): 676–81.

Derrick, Ben, Deirdre Toher, and Paul White. 2016. "Why Welch's Test Is Type I Error Robust." *Tutorials in Quantitative Methods for Psychology* 12 (1): 30–38.

Desa, U. N. 2015. "United Nations Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2015 Revision, Key Findings and Advance Tables." In *Technical Report*. Working Paper No. ESA/P/WP. 241.

Detterbeck, Frank C., Andrew G. Nicholson, Kazuya Kondo, Paul Van Schil, and Cesar Moran. 2011. "The Masaoka-Koga Stage Classification for Thymic Malignancies: Clarification and Definition of Terms." *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 6 (7 Suppl 3): S1710–16.

Deutschbauer, Adam M., Daniel F. Jaramillo, Michael Proctor, Jochen Kumm, Maureen E. Hillenmeyer, Ronald W. Davis, Corey Nislow, and Guri Giaever. 2005. "Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast." *Genetics* 169 (4): 1915–25.

Dittmer, D., S. Pati, G. Zambetti, S. Chu, A. K. Teresky, M. Moore, C. Finlay, and A. J. Levine. 1993. "Gain of Function Mutations in p53." *Nature Genetics* 4 (1): 42–46.

Dixon, Scott J., Yaroslav Fedyshyn, Judice L. Y. Koh, T. S. Keshava Prasad, Charly Chahwan, Gordon Chua, Kiana Toufighi, et al. 2008. "Significant Conservation of Synthetic Lethal Genetic Interaction Networks between Distantly Related Eukaryotes." *Proceedings of the National Academy of Sciences of the United*

States of America 105 (43): 16653–58.

Dobzhansky, T. 1946. “Genetics of Natural Populations. Xiii. Recombination and Variability in Populations of Drosophila Pseudoobscura.” *Genetics* 31 (3): 269–90.

Doench, John G., Nicolo Fusi, Meagan Sullender, Mudra Hegde, Emma W. Vaimberg, Katherine F. Donovan, Ian Smith, et al. 2016. “Optimized sgRNA Design to Maximize Activity and Minimize off-Target Effects of CRISPR-Cas9.” *Nature Biotechnology* 34 (2): 184–91.

Domcke, Silvia, Rileen Sinha, Douglas A. Levine, Chris Sander, and Nikolaus Schultz. 2013. “Evaluating Cell Lines as Tumour Models by Comparison of Genomic Profiles.” *Nature Communications* 4: 2126.

Dorsam, Robert T., and J. Silvio Gutkind. 2007. “G-Protein-Coupled Receptors and Cancer.” *Nature Reviews. Cancer* 7 (2): 79–94.

Dunham, Maitreya J., Hassan Badrane, Tracy Ferea, Julian Adams, Patrick O. Brown, Frank Rosenzweig, and David Botstein. 2002. “Characteristic Genome Rearrangements in Experimental Evolution of Saccharomyces Cerevisiae.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (25): 16144–49.

Engels, Eric A. 2010. “Epidemiology of Thymoma and Associated Malignancies.” *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 5 (10 Suppl 4): S260–65.

Engels, Eric A., and Ruth M. Pfeiffer. 2003. “Malignant Thymoma in the United States: Demographic Patterns in Incidence and Associations with Subsequent

Malignancies." *International Journal of Cancer. Journal International Du Cancer* 105 (4): 546–51.

Faith, Jeremiah J., Boris Hayete, Joshua T. Thaden, Ilaria Mogno, Jamey Wierzbowski, Guillaume Cottarel, Simon Kasif, James J. Collins, and Timothy S. Gardner. 2007. "Large-Scale Mapping and Validation of Escherichia Coli Transcriptional Regulation from a Compendium of Expression Profiles." *PLoS Biology* 5 (1): e8.

Ferlay, Jacques, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. 2015. "Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012." *International Journal of Cancer. Journal International Du Cancer* 136 (5): E359–86.

Fitzgerald, Jonathan B., Bryan W. Johnson, Jason Baum, Sharlene Adams, Sergio Iadevaia, Jian Tang, Victoria Rimkunas, et al. 2014. "MM-141, an IGF-IR--and ErbB3-Directed Bispecific Antibody, Overcomes Network Adaptations That Limit Activity of IGF-IR Inhibitors." *Molecular Cancer Therapeutics* 13 (2): 410–25.

Foulds, L. 1954. "The Experimental Study of Tumor Progression: A Review." *Cancer Research* 14 (5): 327–39.

Garnett, Mathew J., Elena J. Edelman, Sonja J. Heidorn, Chris D. Greenman, Anahita Dastur, King Wai Lau, Patricia Greninger, et al. 2012. "Systematic Identification of Genomic Markers of Drug Sensitivity in Cancer Cells." *Nature* 483 (7391): 570–75.

Garnett, Mathew J., and Ultan McDermott. 2014. “The Evolving Role of Cancer Cell Line-Based Screens to Define the Impact of Cancer Genomes on Drug Response.” *Current Opinion in Genetics & Development* 24 (February): 114–19.

Gherardi, Ermanno, Walter Birchmeier, Carmen Birchmeier, and George Vande Woude. 2012. “Targeting MET in Cancer: Rationale and Progress.” *Nature Reviews. Cancer* 12 (2): 89–103.

Girard, Nicolas, Ronglai Shen, Tianhua Guo, Maureen F. Zakowski, Adriana Heguy, Gregory J. Riely, James Huang, et al. 2009. “Comprehensive Genomic Analysis Reveals Clinically Relevant Molecular Distinctions between Thymic Carcinomas and Thymomas.” *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 15 (22): 6790–99.

Gökmen-Polar, Yesim, Robert W. Cook, Chirayu Pankaj Goswami, Jeff Wilkinson, Derek Maetzold, John F. Stone, Kristen M. Oelschlager, et al. 2013. “A Gene Signature to Determine Metastatic Behavior in Thymomas.” *PloS One* 8 (7): e66047.

Gonzalez-Perez, Abel, and Nuria Lopez-Bigas. 2012. “Functional Impact Bias Reveals Cancer Drivers.” *Nucleic Acids Research* 40 (21): e169.

Goss, Kathleen Heppner, Mary A. Risinger, Jennifer J. Kordich, Maureen M. Sanz, Joel E. Straughen, Lisa E. Slovek, Anthony J. Capobianco, James German, Gregory P. Boivin, and Joanna Groden. 2002. “Enhanced Tumor Formation in Mice Heterozygous for Blm Mutation.” *Science* 297 (5589): 2051–53.

Greene, H. S. N., and B. L. Newton. 1948. “Evolution of Cancer of the Uterine Fundus in the Rabbit.” *Cancer* 1 (1): 82–99.

Greenman, Christopher, Philip Stephens, Raffaella Smith, Gillian L. Dalgliesh, Christopher Hunter, Graham Bignell, Helen Davies, et al. 2007. “Patterns of Somatic Mutation in Human Cancer Genomes.” *Nature* 446 (7132): 153–58.

Griffiths, Anthony J. F., Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, and William M. Gelbart. 2000. *Mutant Types*. W. H. Freeman.

Guan, Bin, Tian-Li Wang, and Ie-Ming Shih. 2011. “ARID1A, a Factor That Promotes Formation of SWI/SNF-Mediated Chromatin Remodeling, Is a Tumor Suppressor in Gynecologic Cancers.” *Cancer Research* 71 (21): 6718–27.

Guarente, L. 1993. “Synthetic Enhancement in Gene Interaction: A Genetic Tool Come of Age.” *Trends in Genetics: TIG* 9 (10): 362–66.

Guo, Jing, Hui Liu, and Jie Zheng. 2016. “SynLethDB: Synthetic Lethality Database toward Discovery of Selective and Sensitive Anticancer Drug Targets.” *Nucleic Acids Research* 44 (D1): D1011–17.

Haber, D. A., D. W. Bell, R. Sordella, E. L. Kwak, N. Godin-Heymann, S. V. Sharma, T. J. Lynch, and J. Settleman. 2005. “Molecular Targeted Therapy of Lung Cancer: EGFR Mutations and Response to EGFR Inhibitors.” *Cold Spring Harbor Symposia on Quantitative Biology* 70: 419–26.

Hanahan, Douglas, and Robert A. Weinberg. 2011. “Hallmarks of Cancer: The next Generation.” *Cell* 144 (5): 646–74.

Hanahan, D., and R. A. Weinberg. 2000. “The Hallmarks of Cancer.” *Cell* 100 (1): 57–70.

Han, Kyuho, Edwin E. Jeng, Gaelen T. Hess, David W. Morgens, Amy Li, and Michael C. Bassik. 2017. “Synergistic Drug Combinations for Cancer Identified

in a CRISPR Screen for Pairwise Genetic Interactions." *Nature Biotechnology* 35 (5): 463–74.

Hartwell, L. H., P. Szankasi, C. J. Roberts, A. W. Murray, and S. H. Friend. 1997. "Integrating Genetic Approaches into the Discovery of Anticancer Drugs." *Science* 278 (5340): 1064–68.

Helming, Katherine C., Xiaofeng Wang, Boris G. Wilson, Francisca Vazquez, Jeffrey R. Haswell, Haley E. Manchester, Youngha Kim, et al. 2014. "ARID1B Is a Specific Vulnerability in ARID1A-Mutant Cancers." *Nature Medicine* 20 (3): 251–54.

Huang, Bei, Djeda Belharazem, Li Li, Susanne Kneitz, Philipp Schnabel, Ralf Rieker, Daniel Körner, et al. 2013. "Anti-Apoptotic Signature in Thymic Squamous Cell Carcinomas--Functional Relevance of Anti-Apoptotic BIRC3 Expression in the Thymic Carcinoma Cell Line 1889c." *Frontiers in Oncology* 3: 316.

Hubbard, T. J. P., B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, et al. 2007. "Ensembl 2007." *Nucleic Acids Research* 35 (Database issue): D610–17.

Hunter, John D. 2007. "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering* 9 (3): 90–95.

International HapMap Consortium. 2003. "The International HapMap Project." *Nature* 426 (6968): 789–96.

IQVIA. 2019. "Global Oncology Trends 2019 - Therapeutics, Clinical Development and Health System Implications." IQVIA Institute. https://www.iqvia.com/institute/reports/global-oncology-trends-2019.

Irizarry, Rafael A., Bridget Hobbs, Francois Collin, Yasmin D. Beazer-Barclay, Kristen J. Antonellis, Uwe Scherf, and Terence P. Speed. 2003. “Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data.” *Biostatistics* 4 (2): 249–64.

Jamison, Dean T., Lawrence H. Summers, George Alleyne, Kenneth J. Arrow, Seth Berkley, Agnes Binagwaho, Flavia Bustreo, et al. 2013. “Global Health 2035: A World Converging within a Generation.” *The Lancet* 382 (9908): 1898–1955.

Jerby-Arnon, Livnat, Nadja Pfetzer, Yedael Y. Waldman, Lynn McGarry, Daniel James, Emma Shanks, Brinton Seashore-Ludlow, et al. 2014. “Predicting Cancer-Specific Vulnerability via Data-Driven Detection of Synthetic Lethality.” *Cell* 158 (5): 1199–1209.

Jones, Siân, Meng Li, D. Williams Parsons, Xiaosong Zhang, Jelle Wesseling, Petra Kristel, Marjanka K. Schmidt, et al. 2012. “Somatic Mutations in the Chromatin Remodeling Gene ARID1A Occur in Several Tumor Types.” *Human Mutation* 33 (1): 100–103.

Jordan, M. A., and L. Wilson. 1998. “The Use and Action of Drugs in Analyzing Mitosis.” *Methods in Cell Biology*. http://www.sciencedirect.com/science/article/pii/S0091679X0861986X.

Kaelin, William G., Jr., 2005. “The Concept of Synthetic Lethality in the Context of Anticancer Therapy.” *Nature Reviews. Cancer* 5 (9): 689–98.

Kamio, Takuya, Tsutomu Toki, Rika Kanezaki, Shinya Sasaki, Satoru Tandai, Kiminori Terui, Dai Ikebe, Kazuhiko Igarashi, and Etsuro Ito. 2003. “B-Cell–specific Transcription Factor BACH2 Modifies the Cytotoxic Effects of

Anticancer Drugs." *Blood* 102 (9): 3317–22.

Kaneko, Takeumi, Jun Hamazaki, Shun-Ichiro Iemura, Katsuhiro Sasaki, Kaori Furuyama, Tohru Natsume, Keiji Tanaka, and Shigeo Murata. 2009. "Assembly Pathway of the Mammalian Proteasome Base Subcomplex Is Mediated by Multiple Specific Chaperones." *Cell* 137 (5): 914–25.

Kataoka, T., S. Powers, C. McGill, O. Fasano, J. Strathern, J. Broach, and M. Wigler. 1984. "Genetic Analysis of Yeast RAS1 and RAS2 Genes." *Cell* 37 (2): 437–45.

Kawakami, Hisato, Isamu Okamoto, Kimio Yonesaka, Kunio Okamoto, Kiyoko Shibata, Yume Shinkai, Haruka Sakamoto, et al. 2014. "The Anti-HER3 Antibody Patritumab Abrogates Cetuximab Resistance Mediated by Heregulin in Colorectal Cancer Cells." *Oncotarget* 5 (23): 11847–56.

Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and David Haussler. 2002. "The Human Genome Browser at UCSC." *Genome Research* 12 (6): 996–1006.

Kim, B., S. Wang, J. M. Lee, Y. Jeong, T. Ahn, D-S Son, H. W. Park, et al. 2015. "Synthetic Lethal Screening Reveals FGFR as One of the Combinatorial Targets to Overcome Resistance to Met-Targeted Therapy." *Oncogene* 34 (9): 1083–93.

Kim, Ellen, and Charles R. Thomas Jr. 2015. "Conditional Survival of Malignant Thymoma Using National Population-Based Surveillance, Epidemiology, and End Results (SEER) Registry (1973-2011)." *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 10 (4): 701–7.

Kim, Jong Wook, Olga B. Botvinnik, Omar Abudayyeh, Chet Birger, Joseph Rosenbluh, Yashaswi Shrestha, Mohamed E. Abazeed, et al. 2016. "Characterizing Genomic Alterations in Cancer by Complementary Functional Associations." *Nature Biotechnology* 34 (5): 539–46.

Knudson, A. G. 2001. "Two Genetic Hits (more or Less) to Cancer." *Nature Reviews. Cancer* 1 (2): 157–62.

Kobayakawa, Ko, Reiko Kobayakawa, Hideyuki Matsumoto, Yuichiro Oka, Takeshi Imai, Masahito Ikawa, Masaru Okabe, et al. 2007. "Innate versus Learned Odour Processing in the Mouse Olfactory Bulb." *Nature* 450 (7169): 503–8.

Kumar, Prateek, Steven Henikoff, and Pauline C. Ng. 2009. "Predicting the Effects of Coding Non-Synonymous Variants on Protein Function Using the SIFT Algorithm." *Nature Protocols* 4 (7): 1073–81.

Kummar, Shivaani, Alice Chen, Jiuping Ji, Yiping Zhang, Joel M. Reid, Matthew Ames, Lee Jia, et al. 2011. "Phase I Study of PARP Inhibitor ABT-888 in Combination with Topotecan in Adults with Refractory Solid Tumors and Lymphomas." *Cancer Research* 71 (17): 5626–34.

Laufer, Christina, Bernd Fischer, Maximilian Billmann, Wolfgang Huber, and Michael Boutros. 2013. "Mapping Genetic Interactions in Human Cancer Cells with RNAi and Multiparametric Phenotyping." *Nature Methods* 10 (5): 427–31.

Lawrence, Michael S., Petar Stojanov, Paz Polak, Gregory V. Kryukov, Kristian Cibulskis, Andrey Sivachenko, Scott L. Carter, et al. 2013. "Mutational Heterogeneity in Cancer and the Search for New Cancer-Associated Genes." *Nature* 499 (7457): 214–18.

Lee, Gui Youn, Woo Ick Yang, Hei Cheul Jeung, Sang Chul Kim, Min Young Seo, Chan Hee Park, Hyun Cheol Chung, and Sun Young Rha. 2007. “Genome-Wide Genetic Aberrations of Thymoma Using cDNA Microarray Based Comparative Genomic Hybridization.” *BMC Genomics* 8 (September): 305.

Levitzki, A., and A. Gazit. 1995. “Tyrosine Kinase Inhibition: An Approach to Drug Development.” *Science* 267 (5205): 1782–88.

Lindahl, Tomas, Masahiko S. Satoh, Guy G. Poirier, and Arne Klungland. 1995. “Post-Translational Modification of poly(ADP-Ribose) Polymerase Induced by DNA Strand Breaks.” *Trends in Biochemical Sciences* 20 (10): 405–11.

Liu, Stanley K., Carla Coackley, Mechthild Krause, Farid Jalali, Norman Chan, and Robert G. Bristow. 2008/8. “A Novel poly(ADP-Ribose) Polymerase Inhibitor, ABT-888, Radiosensitizes Malignant Human Cell Lines under Hypoxia.” *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 88 (2): 258–68.

Liu, Yunhua, Xiaoxiao Hu, Cecil Han, Liana Wang, Xinna Zhang, Xiaoming He, and Xiongbin Lu. 2015. “Targeting Tumor Suppressor Genes for Cancer Therapy.” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 37 (12): 1277–86.

Li, Xue-Juan, Shital K. Mishra, Min Wu, Fan Zhang, and Jie Zheng. 2014. “Syn-Lethality: An Integrative Knowledge Base of Synthetic Lethality towards Discovery of Selective Anticancer Therapies.” *BioMed Research International* 2014 (April): 196034.

Lodish, H. 2008. *Molecular Cell Biology*. W. H. Freeman.

Luo, Biao, Hiu Wing Cheung, Aravind Subramanian, Tanaz Sharifnia, Michael Okamoto, Xiaoping Yang, Greg Hinkle, et al. 2008. “Highly Parallel Identification of Essential Genes in Cancer Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (51): 20380–85.

Luo, Ji, Michael J. Emanuele, Danan Li, Chad J. Creighton, Michael R. Schlabach, Thomas F. Westbrook, Kwok-Kin Wong, and Stephen J. Elledge. 2009. “A Genome-Wide RNAi Screen Identifies Multiple Synthetic Lethal Interactions with the Ras Oncogene.” *Cell* 137 (5): 835–48.

Luo, Ji, Nicole L. Solimini, and Stephen J. Elledge. 2009. “Principles of Cancer Therapy: Oncogene and Non-Oncogene Addiction.” *Cell* 136 (5): 823–37.

Lutz, Wolfgang, Warren Sanderson, and Sergei Scherbov. 2008. “The Coming Acceleration of Global Population Ageing.” *Nature* 451 (7179): 716–19.

Marcotte, Richard, Kevin R. Brown, Fernando Suarez, Azin Sayad, Konstantina Karamboulas, Paul M. Krzyzanowski, Fabrice Sircoulomb, et al. 2012. “Essential Gene Profiles in Breast, Pancreatic, and Ovarian Cancer Cells.” *Cancer Discovery* 2 (2): 172–89.

Marcotte, Richard, Azin Sayad, Kevin R. Brown, Felix Sanchez-Garcia, Jüri Reimand, Maliha Haider, Carl Virtanen, et al. 2016. “Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance.” *Cell* 164 (1-2): 293–309.

Martin, Shawn, W. Michael Brown, Richard Klavans, and Kevin W. Boyack. 2011. “OpenOrd: An Open-Source Toolbox for Large Graph Layout.” In *IS&T/SPIE Electronic Imaging*, 786806–786806 – 11. International Society for Optics and

Photonics.

Marx, A., M. Osborn, S. Tzartos, K. I. Geuder, B. Schalke, W. Nix, T. Kirchner, and H. K. Müller-Hermelink. 1992. “A Striational Muscle Antigen and Myasthenia Gravis-Associated Thymomas Share an Acetylcholine-Receptor Epitope.” *Developmental Immunology* 2 (2): 77–84.

Masaoka, Akira. 2010. “Staging System of Thymoma.” *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 5 (10 Suppl 4): S304–12.

Masaoka, Akira, Yasumasa Monden, Kazuya Nakahara, and Tsuneo Tanioka. 1981. “Follow-up Study of Thymomas with Special Reference to Their Clinical Stages.” *Cancer* 48 (11): 2485–92.

Masuda, Tomoko, Masakatsu Motomura, Kimiaki Utsugisawa, Yuriko Nagane, Ruka Nakata, Masahiro Tokuda, Taku Fukuda, Toshiro Yoshimura, Mitsuhiro Tsujihata, and Atsushi Kawakami. 2012. “Antibodies against the Main Immunogenic Region of the Acetylcholine Receptor Correlate with Disease Severity in Myasthenia Gravis.” *Journal of Neurology, Neurosurgery, and Psychiatry* 83 (9): 935–40.

Mathe, Ewy, Magali Olivier, Shunsuke Kato, Chikashi Ishioka, Pierre Hainaut, and Sean V. Tavtigian. 2006. “Computational Approaches for Predicting the Biological Effect of p53 Missense Mutations: A Comparison of Three Sequence Analysis Based Methods.” *Nucleic Acids Research* 34 (5): 1317–25.

Matsumoto, Makoto, and Takuji Nishimura. 1998. “Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number

Generator." *ACM Transactions on Modeling and Computer Simulation* 8 (1): 3–30.

Ma, Xiaotu, Aaron M. Tarone, and Wenyuan Li. 2008. "Mapping Genetically Compensatory Pathways from Synthetic Lethal Interactions in Yeast." *PloS One* 3 (4): e1922.

McDonald, E. Robert, 3rd, Antoine de Weck, Michael R. Schlabach, Eric Billy, Konstantinos J. Mavrakis, Gregory R. Hoffman, Dhiren Belur, et al. 2017. "Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening." *Cell* 170 (3): 577–92.e10.

McKinney, Wes, and Others. 2010. "Data Structures for Statistical Computing in Python." In *Proceedings of the 9th Python in Science Conference*, 445:51–56. Austin, TX.

McLornan, Donal P., Alan List, and Ghulam J. Mufti. 2014. "Applying Synthetic Lethality for the Selective Targeting of Cancer." *The New England Journal of Medicine* 371 (18): 1725–35.

Measday, Vivien, Kristin Baetz, Julie Guzzo, Karen Yuen, Teresa Kwok, Bilal Sheikh, Huiming Ding, et al. 2005. "Systematic Yeast Synthetic Lethal and Synthetic Dosage Lethal Screens Identify Genes Required for Chromosome Segregation." *Proceedings of the National Academy of Sciences of the United States of America* 102 (39): 13956–61.

Mering, Christian von, Roland Krause, Berend Snel, Michael Cornell, Stephen G. Oliver, Stanley Fields, and Peer Bork. 2002. "Comparative Assessment of

Large-Scale Data Sets of Protein–protein Interactions." *Nature* 417 (6887): 399–403.

Mermel, Craig H., Steven E. Schumacher, Barbara Hill, Matthew L. Meyerson, Rameen Beroukhim, and Gad Getz. 2011. "GISTIC2.0 Facilitates Sensitive and Confident Localization of the Targets of Focal Somatic Copy-Number Alteration in Human Cancers." *Genome Biology* 12 (4): R41.

Miller, J. F. 1961. "Immunological Function of the Thymus." *The Lancet* 2 (7205): 748–49.

Miller, John P., Russell S. Lo, Asa Ben-Hur, Cynthia Desmarais, Igor Stagljar, William Stafford Noble, and Stanley Fields. 2005. "Large-Scale Identification of Yeast Integral Membrane Protein Interactions." *Proceedings of the National Academy of Sciences of the United States of America* 102 (34): 12123–28.

Mina, Marco, Franck Raynaud, Daniele Tavernari, Nikolaus Schultz, Elisa Oricchio, Giovanni Ciriello, Elena Battistello, et al. 2017. "Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies." *Cancer Cell* 32 (2): 155–68.

Mirza, Mansoor R., Bradley J. Monk, Jørn Herrstedt, Amit M. Oza, Sven Mahner, Andrés Redondo, Michel Fabbro, et al. 2016. "Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer." *The New England Journal of Medicine* 375 (22): 2154–64.

Mishra, Rosalin, Hima Patel, Samar Alanazi, Long Yuan, and Joan T. Garrett. 2018. "HER3 Signaling and Targeted Therapy in Cancer." *Oncology Reviews* 12 (1): 355.

Muller, Florian L., Simona Colla, Elisa Aquilanti, Veronica E. Manzo, Giannicola Genovese, Jaclyn Lee, Daniel Eisenson, et al. 2012. “Passenger Deletions Generate Therapeutic Vulnerabilities in Cancer.” *Nature* 488 (7411): 337–42.

Mutter, G. L., M. C. Lin, J. T. Fitzgerald, J. B. Kum, J. P. Baak, J. A. Lees, L. P. Weng, and C. Eng. 2000. “Altered PTEN Expression as a Diagnostic Marker for the Earliest Endometrial Precancers.” *Journal of the National Cancer Institute* 92 (11): 924–30.

Mygland, A., O. B. Tysnes, R. Matre, J. A. Aarli, and N. E. Gilhus. 1994. “Anti-Cardiac Ryanodine Receptor Antibodies in Thymoma-Associated Myasthenia Gravis.” *Autoimmunity* 17 (4): 327–31.

Navarro, M. V., S. M. Gomes Dias, and L. V. Mello. 2007. “Structural Flexibility in Trypanosoma Brucei Enolase Revealed by X-ray Crystallography and Molecular Dynamics.” *FEBS*. http://onlinelibrary.wiley.com/doi/10.1111/j.1742-4658.2007.06027.x/full.

Newton, Yulia. 2016. “Discovering Molecular Patterns with Therapeutic Implications in Large-Cohort Heterogeneous Cross-Cancer Data.” UC Santa Cruz. https://escholarship.org/uc/item/9sd1t5jb.

Nijhawan, Deepak, Travis I. Zack, Yin Ren, Matthew R. Strickland, Rebecca Lamothe, Steven E. Schumacher, Aviad Tsherniak, et al. 2012. “Cancer Vulnerabilities Unveiled by Genomic Loss.” *Cell* 150 (4): 842–54.

Nowell, P. C. 1976. “The Clonal Evolution of Tumor Cell Populations.” *Science* 194 (4260): 23–28.

Okumura, Meinoshin, Mitsunori Ohta, Hisashi Tateyama, Katsuhiro Nakagawa,

Akihide Matsumura, Hajime Maeda, Hiroto Tada, Tadaaki Eimoto, Hikaru Matsuda, and Akira Masaoka. 2002. “The World Health Organization Histologic Classification System Reflects the Oncologic Behavior of Thymoma: A Clinical Study of 273 Patients.” *Cancer* 94 (3): 624–32.

Oliphant, Travis E. 2007. “Python for Scientific Computing.” *Computing in Science & Engineering* 9 (3): 10–20.

Olshen, Adam B., E. S. Venkatraman, Robert Lucito, and Michael Wigler. 2004. “Circular Binary Segmentation for the Analysis of Array-based DNA Copy Number Data.” *Biostatistics* 5 (4): 557–72.

Ooi, Siew Loon, Xuewen Pan, Brian D. Peyser, Ping Ye, Pamela B. Meluh, Daniel S. Yuan, Rafael A. Irizarry, Joel S. Bader, Forrest A. Spencer, and Jef D. Boeke. 2006. “Global Synthetic-Lethality Analysis and Yeast Functional Profiling.” *Trends in Genetics: TIG* 22 (1): 56–63.

Oyaizu, H., Y. Adachi, T. Okumura, M. Okigaki, N. Oyaizu, S. Taketani, K. Ikebukuro, S. Fukuhara, and S. Ikehara. 2001. “Proteasome Inhibitor 1 Enhances Paclitaxel-Induced Apoptosis in Human Lung Adenocarcinoma Cell Line.” *Oncology Reports* 8 (4): 825–29.

Oyake, T., K. Itoh, H. Motohashi, N. Hayashi, H. Hoshino, M. Nishizawa, M. Yamamoto, and K. Igarashi. 1996. “Bach Proteins Belong to a Novel Family of BTB-Basic Leucine Zipper Transcription Factors That Interact with MafK and Regulate Transcription through the NF-E2 Site.” *Molecular and Cellular Biology* 16 (11): 6083–95.

Pan, Xuewen, Daniel S. Yuan, Dong Xiang, Xiaoling Wang, Sharon

Sookhai-Mahadeo, Joel S. Bader, Philip Hieter, Forrest Spencer, and Jef D. Boeke. 2004. “A Robust Toolkit for Functional Profiling of the Yeast Genome.” *Molecular Cell* 16 (3): 487–96.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. “Scikit-Learn: Machine Learning in Python.” *Journal of Machine Learning Research: JMLR* 12 (November): 2825–30.

Pérez, Fernando, and Brian E. Granger. 2007. “IPython: A System for Interactive Scientific Computing.” *Computing in Science & Engineering* 9 (3): 21–29.

Petrini, Iacopo, Paul S. Meltzer, In-Kyu Kim, Marco Lucchi, Kang-Seo Park, Gabriella Fontanini, James Gao, et al. 2014. “A Specific Missense Mutation in GTF2I Occurs at High Frequency in Thymic Epithelial Tumors.” *Nature Genetics* 46 (8): 844–49.

Pilleron, Sophie, Diana Sarfati, Maryska Janssen-Heijnen, Jérôme Vignat, Jacques Ferlay, Freddie Bray, and Isabelle Soerjomataram. 2019. “Global Cancer Incidence in Older Adults, 2012 and 2035: A Population-Based Study.” *International Journal of Cancer. Journal International Du Cancer* 144 (1): 49–58.

Pishvaian, M. J., R. Slack, A. Witkiewicz, A. R. He, J. J. Hwang, A. Hankin, K. Dorsch-Vogel, et al. 2011. “A Phase II Study of the PARP Inhibitor ABT-888 plus Temozolomide in Patients with Heavily Pretreated, Metastatic Colorectal Cancer.” *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 29 (15_suppl): 3502.

Polyak, Kornelia, and Judy Garber. 2011. “Targeting the Missing Links for Cancer

Therapy." *Nature Medicine* 17 (3): 283–84.

Poyner, R. R., and G. H. Reed. 1992. "Structure of the Bis Divalent Cation Complex with Phosphonoacetohydroxamate at the Active Site of Enolase." *Biochemistry* 31 (31): 7166–73.

Radovich, Milan, Curtis R. Pickering, Ina Felau, Gavin Ha, Hailei Zhang, Heejoon Jo, Katherine A. Hoadley, et al. 2018. "The Integrated Genomic Landscape of Thymic Epithelial Tumors." *Cancer Cell* 33 (2): 244–58.e10.

Radovich, Milan, Jeffrey P. Solzak, Bradley A. Hancock, Madison L. Conces, Rutuja Atale, Ryan F. Porter, Jin Zhu, et al. 2016. "A Large microRNA Cluster on Chromosome 19 Is a Transcriptional Hallmark of WHO Type A and AB Thymomas." *British Journal of Cancer* 114 (4): 477–84.

Rangarajan, L. N. 1992. "Kautilya: The Arthashastra." *Basu Mudran: Penguin Books India*.

Rees, Matthew G., Brinton Seashore-Ludlow, Jaime H. Cheah, Drew J. Adams, Edmund V. Price, Shubhroz Gill, Sarah Javaid, et al. 2016. "Correlating Chemical Sensitivity and Basal Gene Expression Reveals Mechanism of Action." *Nature Chemical Biology* 12 (2): 109–16.

Reva, Boris, Yevgeniy Antipin, and Chris Sander. 2011. "Predicting the Functional Impact of Protein Mutations: Application to Cancer Genomics." *Nucleic Acids Research* 39 (17): e118.

Robinson, D. R., Y. M. Wu, and S. F. Lin. 2000. "The Protein Tyrosine Kinase Family of the Human Genome." *Oncogene* 19 (49): 5548–57.

Robson, Mark, Seock-Ah Im, Elżbieta Senkus, Binghe Xu, Susan M. Domchek,

Norikazu Masuda, Suzette Delaloge, et al. 2017. “Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation.” *The New England Journal of Medicine* 377 (6): 523–33.

Rognan, D. 2007. “Chemogenomic Approaches to Rational Drug Design.” *British Journal of Pharmacology* 152 (1): 38–52.

Rousseeuw, Peter J. 1987. “Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis.” *Journal of Computational and Applied Mathematics* 20 (November): 53–65.

Roychoudhuri, Rahul, Robert L. Eil, David Clever, Christopher A. Klebanoff, Madhusudhanan Sukumar, Francis M. Grant, Zhiya Yu, et al. 2016. “The Transcription Factor BACH2 Promotes Tumor Immunosuppression.” *The Journal of Clinical Investigation* 126 (2): 599–604.

Ruxton, Graeme D. 2006. “The Unequal Variance T-Test Is an Underused Alternative to Student’s T-Test and the Mann–Whitney U Test.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 17 (4): 688–90.

Rykowski, M. C., J. W. Wallis, J. Choe, and M. Grunstein. 1981. “Histone H2B Subtypes Are Dispensable during the Yeast Cell Cycle.” *Cell* 25 (2): 477–87.

Sasaki, Hidefumi, Nobuyuki Ide, Ichiro Fukai, Masanobu Kiriyama, Yosuke Yamakawa, and Yoshitaka Fujii. 2002. “Gene Expression Analysis of Human Thymoma Correlates with Tumor Stage.” *International Journal of Cancer. Journal International Du Cancer* 101 (4): 342–47.

Sasaki, S., E. Ito, T. Toki, T. Maekawa, R. Kanezaki, T. Umenai, A. Muto, et al. 2000. “Cloning and Expression of Human B Cell-Specific Transcription Factor BACH2

Mapped to Chromosome 6q15." *Oncogene* 19 (33): 3739–49.

Schaefer, Carl F., Kira Anthony, Shiva Krupa, Jeffrey Buchoff, Matthew Day, Timo Hannay, and Kenneth H. Buetow. 2009. "PID: The Pathway Interaction Database." *Nucleic Acids Research* 37 (Database issue): D674–79.

Schiff, P. B., J. Fant, and S. B. Horwitz. 1979. "Promotion of Microtubule Assembly in Vitro by Taxol." *Nature* 277 (5698): 665–67.

Schmidt, Esther E., Oliver Pelz, Svetlana Buhlmann, Grainne Kerr, Thomas Horn, and Michael Boutros. 2013. "GenomeRNAi: A Database for Cell-Based and in Vivo RNAi Phenotypes, 2013 Update." *Nucleic Acids Research* 41 (Database issue): D1021–26.

Schneeweiss, Andreas, Tjoung-Won Park-Simon, Joan Albanell, Ulrik Lassen, Javier Cortés, Veronique Dieras, Marcus May, et al. 2018. "Phase Ib Study Evaluating Safety and Clinical Activity of the Anti-HER3 Antibody Lumretuzumab Combined with the Anti-HER2 Antibody Pertuzumab and Paclitaxel in HER3-Positive, HER2-Low Metastatic Breast Cancer." *Investigational New Drugs*, 1–12.

Schultz, A., V. Hoffacker, A. Wilisch, W. Nix, R. Gold, B. Schalke, S. Tzartos, H. K. Müller-Hermelink, and A. Marx. 1999. "Neurofilament Is an Autoantigenic Determinant in Myasthenia Gravis." *Annals of Neurology* 46 (2): 167–75.

Scorsetti, Marta, Francesco Leo, Annalisa Trama, Rolando D'Angelillo, Danila Serpico, Marianna Macerelli, Paolo Zucali, Gemma Gatta, and Marina Chiara Garassino. 2016. "Thymoma and Thymic Carcinomas." *Critical Reviews in Oncology/hematology* 99 (March): 332–50.

Seashore-Ludlow, Brinton, Matthew G. Rees, Jaime H. Cheah, Murat Cokol, Edmund V. Price, Matthew E. Coletti, Victor Jones, et al. 2015. “Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset.” *Cancer Discovery* 5 (11): 1210–23.

Shah, Sohrab P., Andrew Roth, Rodrigo Goya, Arusha Oloumi, Gavin Ha, Yongjun Zhao, Gulisa Turashvili, et al. 2012. “The Clonal and Mutational Evolution Spectrum of Primary Triple-Negative Breast Cancers.” *Nature* 486 (7403): 395–99.

Shankavaram, Uma T., William C. Reinhold, Satoshi Nishizuka, Sylvia Major, Daisaku Morita, Krishna K. Chary, Mark A. Reimers, et al. 2007. “Transcript and Protein Expression Profiles of the NCI-60 Cancer Cell Panel: An Integromic Microarray Study.” *Molecular Cancer Therapeutics* 6 (3): 820–32.

Shao, Diane D., Aviad Tsherniak, Shuba Gopal, Barbara A. Weir, Pablo Tamayo, Nicolas Stransky, Steven E. Schumacher, et al. 2013. “ATARiS: Computational Quantification of Gene Suppression Phenotypes from Multisample RNAi Screens.” *Genome Research* 23 (4): 665–78.

Sharma, Sreenath V., Daniel A. Haber, and Jeff Settleman. 2010. “Cell Line-Based Platforms to Evaluate the Therapeutic Efficacy of Candidate Anticancer Agents.” *Nature Reviews. Cancer* 10 (4): 241–53.

Shen, Wen Hong, Adayabalam S. Balajee, Jianli Wang, Hong Wu, Charis Eng, Pier Paolo Pandolfi, and Yuxin Yin. 2007. “Essential Role for Nuclear PTEN in Maintaining Chromosomal Integrity.” *Cell* 128 (1): 157–70.

Sjöblom, Tobias, Siân Jones, Laura D. Wood, D. Williams Parsons, Jimmy Lin,

Thomas D. Barber, Diana Mandelker, et al. 2006. “The Consensus Coding Sequences of Human Breast and Colorectal Cancers.” *Science* 314 (5797): 268–74.

Song, Min Sup, Leonardo Salmena, and Pier Paolo Pandolfi. 2012. “The Functions and Regulation of the PTEN Tumour Suppressor.” *Nature Reviews. Molecular Cell Biology* 13 (5): 283–96.

Souadjian, J. V., M. N. Silverstein, and J. L. Titus. 1968. “Thymoma and Cancer.” *Cancer* 22 (6): 1221–25.

Spring, Kevin, Farida Ahangari, Shaun P. Scott, Paul Waring, David M. Purdie, Philip C. Chen, Kevin Hourigan, et al. 2002. “Mice Heterozygous for Mutation in Atm, the Gene Involved in Ataxia-Telangiectasia, Have Heightened Susceptibility to Cancer.” *Nature Genetics* 32 (1): 185–90.

Stark, Chris, Bobby-Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and Mike Tyers. 2006. “BioGRID: A General Repository for Interaction Datasets.” *Nucleic Acids Research* 34 (Database issue): D535–39.

Stewart, Bwkp, Christopher P. Wild, and Others. 2014. “World Cancer Report 2014.” http://publichealthwell.ie/search-results/world-cancer-report-2014?source=relatedblock.

Stone, Eric A., and Arend Sidow. 2005. “Physicochemical Constraint Violation by Missense Substitutions Mediates Impairment of Protein Function and Disease Severity.” *Genome Research* 15 (7): 978–86.

Strollo, D. C., M. L. Rosado de Christenson, and J. R. Jett. 1997. “Primary

Mediastinal Tumors. Part 1: Tumors of the Anterior Mediastinum." *Chest* 112 (2): 511–22.

Sudhakar, Akulapalli. 2009. "History of Cancer, Ancient and Modern Treatment Methods." *Journal of Cancer Science & Therapy* 1 (2): 1–4.

Sudo, Hitomi, Atsushi B. Tsuji, Aya Sugyo, Masakazu Kohda, Chizuru Sogawa, Chisato Yoshida, Yoshi-Nobu Harada, Okio Hino, and Tsuneo Saga. 2010. "Knockdown of COPA, Identified by Loss-of-Function Screen, Induces Apoptosis and Suppresses Tumor Growth in Mesothelioma Mouse Model." *Genomics* 95 (4): 210–16.

Swisher, Elizabeth M., Kevin K. Lin, Amit M. Oza, Clare L. Scott, Heidi Giordano, James Sun, Gottfried E. Konecny, et al. 2017. "Rucaparib in Relapsed, Platinum-Sensitive High-Grade Ovarian Carcinoma (ARIEL2 Part 1): An International, Multicentre, Open-Label, Phase 2 Trial." *The Lancet Oncology* 18 (1): 75–87.

Tarailo, Maja, Sanja Tarailo, and Ann M. Rose. 2007. "Synthetic Lethal Interactions Identify Phenotypic 'Interologs' of the Spindle Assembly Checkpoint Components." *Genetics* 177 (4): 2525–30.

Tatchell, K., D. T. Chaleff, D. DeFeo-Jones, and E. M. Scolnick. 1984. "Requirement of Either of a Pair of Ras-Related Genes of Saccharomyces Cerevisiae for Spore Viability." *Nature* 309 (5968): 523–27.

Thomas, Roman K., Alissa C. Baker, Ralph M. Debiasi, Wendy Winckler, Thomas Laframboise, William M. Lin, Meng Wang, et al. 2007. "High-Throughput Oncogene Mutation Profiling in Human Cancer." *Nature Genetics* 39 (3):

347–51.

Tong, A. H., M. Evangelista, A. B. Parsons, H. Xu, G. D. Bader, N. Pagé, M. Robinson, et al. 2001. “Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants.” *Science* 294 (5550): 2364–68.

Travis, William D., Elisabeth Brambilla, Allen P. Burke, Alexander Marx, and Andrew G. Nicholson. 2015. “Introduction to The 2015 World Health Organization Classification of Tumors of the Lung, Pleura, Thymus, and Heart.” *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 10 (9): 1240–42.

Tsherniak, Aviad, Francisca Vazquez, Phil G. Montgomery, Barbara A. Weir, Gregory Kryukov, Glenn S. Cowley, Stanley Gill, et al. 2017. “Defining a Cancer Dependency Map.” *Cell* 170 (3): 564–76.e16.

Turner, Nicholas C., Christopher J. Lord, Elizabeth Iorns, Rachel Brough, Sally Swift, Richard Elliott, Sydonia Rayter, Andrew N. Tutt, and Alan Ashworth. 2008. “A Synthetic Lethal siRNA Screen Identifying Genes Mediating Sensitivity to a PARP Inhibitor.” *The EMBO Journal* 27 (9): 1368–77.

Tutt, Andrew, Mark Robson, Judy E. Garber, Susan M. Domchek, M. William Audeh, Jeffrey N. Weitzel, Michael Friedlander, et al. 2010. “Oral poly(ADP-Ribose) Polymerase Inhibitor Olaparib in Patients with BRCA1 or BRCA2 Mutations and Advanced Breast Cancer: A Proof-of-Concept Trial.” *The Lancet* 376 (9737): 235–44.

Typas, Athanasios, Robert J. Nichols, Deborah A. Siegele, Michael Shales, Sean R. Collins, Bentley Lim, Hannes Braberg, et al. 2008. “High-Throughput,

Quantitative Analyses of Genetic Interactions in E. Coli." *Nature Methods* 5 (9): 781–87.

Van Der Walt, Stefan, S. Chris Colbert, and Gael Varoquaux. 2011. "The NumPy Array: A Structure for Efficient Numerical Computation." *Computing in Science & Engineering* 13 (2): 22.

Vaske, Charles J., Stephen C. Benz, J. Zachary Sanborn, Dent Earl, Christopher Szeto, Jingchun Zhu, David Haussler, and Joshua M. Stuart. 2010. "Inference of Patient-Specific Pathway Activities from Multi-Dimensional Cancer Genomics Data Using PARADIGM." *Bioinformatics* 26 (12): i237–45.

Vavouri, Tanya, Jennifer I. Semple, and Ben Lehner. 2008. "Widespread Conservation of Genetic Redundancy during a Billion Years of Eukaryotic Evolution." *Trends in Genetics: TIG* 24 (10): 485–88.

Vogel, Christine, and Edward M. Marcotte. 2012. "Insights into the Regulation of Protein Abundance from Proteomic and Transcriptomic Analyses." *Nature Reviews. Genetics* 13 (4): 227–32.

Wang, Hong, Haiyong Han, Spyro Mousses, and Daniel D. Von Hoff. 2006. "Targeting Loss-of-Function Mutations in Tumor-Suppressor Genes as a Strategy for Development of Cancer Therapeutic Agents." *Seminars in Oncology* 33 (4): 513–20.

Wang, Yisong, Anish Thomas, Christopher Lau, Arun Rajan, Yuelin Zhu, J. Keith Killian, Iacopo Petrini, et al. 2014. "Mutations of Epigenetic Regulatory Genes Are Common in Thymic Carcinomas." *Scientific Reports* 4 (December): 7336.

Weinstein, John N. 2012. "Drug Discovery: Cell Lines Battle Cancer." *Nature* 483

(7391): 544–45.

Welch, Bernard L. 1947. “The Generalization Ofstudent’s' Problem When Several Different Population Variances Are Involved.” *Biometrika* 34 (1/2): 28–35.

Whitehurst, Angelique W., Brian O. Bodemann, Jessica Cardenas, Deborah Ferguson, Luc Girard, Michael Peyton, John D. Minna, et al. 2007. “Synthetic Lethal Screen Identification of Chemosensitizer Loci in Cancer Cells.” *Nature* 446 (7137): 815–19.

Wiegand, Kimberly C., Sohrab P. Shah, Osama M. Al-Agha, Yongjun Zhao, Kane Tse, Thomas Zeng, Janine Senz, et al. 2010. “ARID1A Mutations in Endometriosis-Associated Ovarian Carcinomas.” *The New England Journal of Medicine* 363 (16): 1532–43.

Wilding, Jennifer L., and Walter F. Bodmer. 2014. “Cancer Cell Lines for Drug Discovery and Development.” *Cancer Research* 74 (9): 2377–84.

Wilkie, A. O. 1994. “The Molecular Basis of Genetic Dominance.” *Journal of Medical Genetics* 31 (2): 89–98.

Wilkins, E. W., Jr, and B. Castleman. 1979. “Thymoma: A Continuing Survey at the Massachusetts General Hospital.” *The Annals of Thoracic Surgery* 28 (3): 252–56.

Wishart, David S., Craig Knox, An Chi Guo, Dean Cheng, Savita Shrivastava, Dan Tzur, Bijaya Gautam, and Murtaza Hassanali. 2008. “DrugBank: A Knowledgebase for Drugs, Drug Actions and Drug Targets.” *Nucleic Acids Research* 36 (Database issue): D901–6.

Wold, Finn. 1971. “18 Enolase.” In *The Enzymes*, edited by Paul D. Boyer, Volume

5:499–538. Academic Press.

Wolfsberg, Tyra G. 2011. “Using the NCBI Map Viewer to Browse Genomic Sequence Data.” *Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.]* Chapter 18 (April): Unit 18.5.

Wong, Sharyl L., Lan V. Zhang, Amy H. Y. Tong, Zhijian Li, Debra S. Goldberg, Oliver D. King, Guillaume Lesage, et al. 2004. “Combining Biological Networks to Predict Genetic Interactions.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (44): 15682–87.

Wu, Jennifer N., and Charles W. M. Roberts. 2013. “ARID1A Mutations in Cancer: Another Epigenetic Tumor Suppressor?” *Cancer Discovery* 3 (1): 35–43.

Wychulis, A. R., W. S. Payne, O. T. Clagett, and L. B. Woolner. 1971. “Surgical Treatment of Mediastinal Tumors: A 40 Year Experience.” *The Journal of Thoracic and Cardiovascular Surgery* 62 (3): 379–92.

Ye, P., B. D. Peyser, X. Pan, and J. D. Boeke. 2005. “Gene Function Prediction from Congruent Synthetic Lethal Interactions in Yeast.” *Molecular Systems Biology*. http://msb.embopress.org.oca.ucsc.edu/content/1/1/2005.0026.abstract.

Young, L. C., B. G. Campling, S. P. Cole, R. G. Deeley, and J. H. Gerlach. 2001. “Multidrug Resistance Proteins MRP3, MRP1, and MRP2 in Lung Cancer: Correlation of Protein Levels with Drug Response and Messenger RNA Levels.” *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 7 (6): 1798–1804.

Yuen, Karen W. Y., Cheryl D. Warren, Ou Chen, Teresa Kwok, Phil Hieter, and Forrest A. Spencer. 2007. “Systematic Genome Instability Screens in Yeast and

Their Potential Relevance to Cancer." *Proceedings of the National Academy of Sciences of the United States of America* 104 (10): 3925–30.

Zalatan, Jesse G., Michael E. Lee, Ricardo Almeida, Luke A. Gilbert, Evan H. Whitehead, Marie La Russa, Jordan C. Tsai, et al. 2015. "Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds." *Cell* 160 (1-2): 339–50.

Zekeridou, Anastasia, Andrew McKeon, and Vanda A. Lennon. 2016. "Frequency of Synaptic Autoantibody Accompaniments and Neurological Manifestations of Thymoma." *JAMA Neurology* 73 (7): 853–59.

Zettl, A., P. Ströbel, K. Wagner, T. Katzenberger, G. Ott, A. Rosenwald, K. Peters, et al. 2000. "Recurrent Genetic Aberrations in Thymoma and Thymic Carcinoma." *The American Journal of Pathology* 157 (1): 257–66.