THE UNIVERSITY OF CHICAGO

PREDICTING CLINICAL DRUG COMBINATION EFFICACY USING *IN VITRO* MONOTHERAPY DATA

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON CANCER BIOLOGY

BY

ALEXANDER LEE LING

CHICAGO, ILLINOIS

DECEMBER 2019

This work is dedicated to my family. To my wife, Jasmine, who has been a constant source of support and love throughout this research and who has graciously put up with my absence during many late nights of research. To my parents, Brian and Michelle, who have nurtured my curiosity from my infancy and who have been an unwavering source of love and support through every stage of my life. To my brother, Jeff, who filled my childhood with joy and adventure and whose patience and kindness have made me naïve of sibling rivalries. And to my heavenly Father, who sacrificed more than I know to save my soul and who is my eternal source of grace, hope, peace, and joy. I love you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Note: Supplementary Tables S1 – S5 are available online.

ACKNOWLEDGEMENTS

I must begin this section by expressing my sincerest gratitude to my mentor, Dr. R. Stephanie Huang. I am immensely grateful for the countless hours she spent providing her insights into this work and for her constant encouragement to generate and explore my own scientific ideas. It has been an honor to work in her lab. I would also like to thank Dr. Kenan Onel for allowing me to work in his lab while he was at the University of Chicago; I appreciate the chance he gave me to develop my technical and computational skills during the early years of my PhD training.

I must also thank the many past and present lab members from both Stephanie's and Ken's labs who have been sources of scientific discourse, advice, and friendship. In particular, I would like to thank Andrew Skol for helping me to learn how to use R and the bash command line, Paul Geeleher for providing advice and ideas for my thesis work, Robert Gruener for being an excellent partner in much of the work which comprises chapter 2 of this thesis, and Aritro Nath for his continuous willingness to share and debate scientific ideas. I owe much to all of you.

I also owe a great deal to Dr. Geoffrey Greene, Dr. Barbara Stranger, Dr. Lixing Yang, and Dr. Kay Macleod, who kindly agreed to serve as my thesis committee and who provided me with invaluable advice and direction during my time in the program. Their efforts have done much to change this work for the better.

Dr. Kay Macleod also served as the chair of the Committee on Cancer Biology during my time in the program, and I sincerely thank her and all of the other administrators and faculty of the committee for the opportunity and training they gave me to pursue this research. I also thank the many faculty who served as my teachers and who allowed me to do research rotations in their labs—I learned a great deal from each of them.

I am also grateful to the University of Minnesota for allowing me to complete the final years of my PhD on their campus following Stephanie's move to UMN. I appreciate the many efforts they made to welcome me and the opportunities they gave me to integrate into their programs.

I would also like to thank Emily Kurtz at the University of Minnesota Statistical Consulting Center and Dr. Hae Kyung Im at the University of Chicago for taking the time to provide feedback on my method of estimating uncertainties in the efficacy predictions produced by IDACombo. I also thank the Minnesota Supercomputing Institute (MSI, http://www.msi.umn.edu) at the University of Minnesota for providing resources that contributed to the research results reported in this thesis.

I am also grateful to the NIH for providing support for this work and for my training with NIH/NCI Grant 1R01CA204856-01A1 and with the University of Chicago Cancer Biology Training program grant [T32 CA009594].

I would also like to sincerely thank the many peers I had in the Biological Sciences Division who shared their friendship and scientific interests with me. I can't imagine having gone through the program without them.

Finally, I would like to thank the many cancer patients who bravely agreed to test experimental treatments by enrolling in the hundreds of clinical trials whose data I used for my research. The advances made in this work would have been impossible without them.

ABBREVIATIONS

CCL (Cancer Cell Line); CTRPv2 (Cancer Therapeutics Response Portal Version 2); GDSC (Genomics of Drug Sensitivity in Cancer); HR (Hazard Ratio); Highest Single Agent (HSA); IDA (Independent Drug Action); NCI (National Cancer Institute); OS (Overall Survival); PFS (Progression-Free Survival); TTP (Time to Progression); WHO (World Health Organization)

ABSTRACT

Drug combinations are a cornerstone of cancer therapy, but the vast number of possible drug combinations makes it infeasible to screen them all experimentally when identifying new therapies—for example, testing all possible 4-drug combinations for 200 compounds in 100 cell lines would require more than 6 billion experiments, each requiring multiple drug concentrations and replicate measurements. To overcome this problem, efforts have been made to develop computational models capable of predicting drug combination efficacy to select lead candidates prior to experimentally testing them. While these models have traditionally aimed to predict drug synergy, recent evidence has emerged suggesting that many cancer drug combinations may derive their efficacy from independent drug action (IDA), where patients only receive benefit from the single most effective drug in a drug combination.

In this thesis, I present my work to develop a method capable of using the IDA model to predict clinical drug combination efficacy using *in vitro* monotherapy data. This work resulted in the creation of IDACombo, an R package which enables IDA based predictions of drug combination efficacy using monotherapy data from high-throughput cancer cell line (CCL) screens. I show that IDACombo predictions closely agree with measured drug combination efficacies both *in vitro* (Pearson's correlation = 0.94 when comparing predicted efficacies to measured efficacies for >5000 combinations) and in a systematically selected set of clinical trials (accuracy > 88% for predicting PFS/TTP or OS benefit in 26 first line therapy trials). This work provides a framework for translating monotherapy cell line data into clinically meaningful efficacy predictions for hundreds of thousands of 2-drug combinations and millions of combinations of 3 or more drugs.

CHAPTER 1: BACKGROUND AND SIGNIFICANCE

CANCER INCIDENCE, MORTALITY, AND COSTS

Cancer is one of the most significant causes of loss of human life. According to the American Cancer Society, it is estimated that there will be more than 1.7 million new cancer cases and 600,000 cancer fatalities in the United States in 2019 (Siegel et al., 2019). This makes cancer the second leading cause of death in the US, surpassed only by heart disease, with approximately 39% of US men and 38% of US women expected to be diagnosed with invasive cancer during their lifetimes. Worldwide, the World Health Organization (WHO) estimates that there were 18.1 million new cancer cases with 9.6 million cancer deaths (approximately one death every 3.3 seconds) in 2018 (Bray et al., 2018). By WHO estimates, this made cancer the first or second leading cause of premature mortality (death before age 70) in 91 of 172 countries in 2015.

Beyond the loss of human lives caused by cancer, the economic costs of cancer are also large. These costs can be broadly divided into two categories—the lost economic productivity from people who have died from cancer before the end of their working years and the medical costs associated with treating cancer patients. There were an estimated $94.4 billion in lost earnings for persons aged 16 to 84 who died from cancer in the US in 2015 (Islami et al., 2019), and it is estimated that the US will spend roughly $185 billion on medical care costs for cancer patients in 2020 (Mariotto et al., 2011, adjusted for inflation). As such, it can be estimated that cancer costs approximately $280 billion a year in the US alone.

Fortunately, progress has been made in reducing cancer incidence and mortality over the past 45 years. Improved cancer therapies and measures to reduce cancer incidence have led to an estimated 2.6 million averted cancer deaths in the US since the annual number of cancer deaths

peaked in 1991 (Siegel et al., 2019). Much of this progress is due to a decrease in lung cancer incidence following efforts to improve awareness of the risks associated with tobacco and lung cancer development. This has had such a profound effect because lung cancer is the most common type of cancer in the US and is very deadly, with an estimated 5-year survival rate of only 25.1% between 2009 and 2015 (Howlader et al., 2019). Tobacco risk awareness efforts, along with improvements in lung cancer therapies, have led to a 27.4% reduction in lung cancer incidence rates and a 34.7% reduction in lung cancer death rates from 1991 to 2016.

For many other cancers, however, improvements in survival have been achieved based solely on improvements in cancer therapies. Leukemia, for example, has had a relatively stable US incidence rate since 1975, but 5-year survival rates for leukemia have improved from 34.6% in 1975 to 66.5% in 2011. Likewise, myeloma has seen a 42% increase in its US incidence rate since 1975—possibly due to an aging population (Turesson et al., 2018)—but 5-year survival rates for myeloma have improved from 24.6% from 1975-1977 to 53.7% from 2009-2015 (Howlader et al., 2019). Much of the credit for these improvements in cancer outcomes is due to the success of drug combinations for treating cancer.

THE USE OF DRUG COMBINATIONS FOR CANCER TREATMENT

The development of drug combinations for cancer treatment began more than 60 years ago. Drug combinations were tested in mouse models of leukemia in the early 1950's (Law, 1952; Skipper et al., 1954; Venditti et al., 1956), and, by the late 1950's, clinical trials of drug combinations had demonstrated that they could be safely administered in human leukemia patients (Frei et al., 1958). While an early drug combination trial in acute leukemia patients identified no significant benefit from the combination (6-Mercaptopurine + Azaserine, Heyn et al., 1960), later studies of different drug combinations (6-Mercaptopurine + Methotrexate and 6-

Mercaptopurine + Prednisone) demonstrated that drug combinations could improve outcomes for leukemia patients versus monotherapy (Frei et al., 1961, 1965). This laid the groundwork for the use of drug combinations in cancer, and, in the decades following this discovery, drug combinations transformed clinical outcomes for breast cancer, Hodgkin's disease, leukemia, ovarian cancer, testicular cancer, and many others (Devita et al., 1975; DeVita and Chu, 2008; Bukowska et al., 2015). Given their hugely successful clinical record, the development of new cancer drug combinations remains an important area of ongoing research, as do the continued efforts to understand the ways in which drugs interact when used in combination.

HISTORICAL MODELS OF DRUG COMBINATION EFFICACY

Drug Interaction Definitions

Over the past 150 years, numerous theories have been proposed for how drugs can interact when administered as a combination, each with different underlying assumptions and often leading to different conclusions when applied to experimental data (Vlot et al., 2019; Wooten et al., 2019). These theories have been principally concerned with creating ways to classify drug combinations into distinct classes of drug interactions, with different theories proposing different numbers of drug interaction classes. While some theories propose up to 11 interaction classes (Berenbaum, 1989), most suffice with just three, which can be broadly defined as follows:

1. Antagonism: where one drug opposes the effect of another drug.
2. Independence/Additivity: where the combined drugs do not interact in any way that affects their individual activities—the choice of "independence" or "additivity" for this class is determined by the theory which defines the term and its underlying assumptions.

3. Synergy: where interactions between the combined drugs results in an effect beyond that which would be expected from the combined independent activities of the drugs.

While an exhaustive summary of all of the proposed drug combination interaction models is beyond the scope of this thesis, many detailed reviews have been written on this topic over the past 30 years. In my opinion, some of the most helpful have been written by Berenbaum (1989), Greco et al. (1995), Foucquier and Guedj (2015), and Pemovska et al. (2018). Given the high level of interest in understanding and developing novel drug combinations, however, new models of defining drug antagonism, additivity, and synergy are constantly being proposed, including models that are in preprint as of the writing of this thesis (Wooten et al., 2019). That said, I will provide a brief introduction to the development of the Loewe Additivity and Bliss Independence, which are the most well established and widely used models, as well as independent drug action model (IDA), which is also known as Gaddum's independence or the highest single agent (HSA) model and which is the model with which my work is primarily concerned.

Loewe Additivity and Bliss Independence

The beginnings of Loewe additivity trace back to a 19th century lecture by Fraser (1872a, 1872b) at the Royal College of Physicians in Edinburgh, where he presented experiments into the use of drug antagonists as physiological antidotes to counteract poisons. One of the ways he presented his results was by plotting the pairs of concentrations of a combined poison and antidote which resulted in the death of rabbits treated with the combined poison and antidote (administered 5 minutes apart). 56 years later, Loewe (1928) named such plots isobolograms and he named the curves produced by such plots as isoboles. Using these plots and lines, he introduced formal definitions for drug antagonism, additivity, and synergy by hypothesizing the shape that a reference isobole should have when independently acting drugs are combined to

produce an additive effect and by then classifying deviations from this reference isobole (i.e. as observed in experimental results) as synergy or antagonism. In more recent years, mathematical definitions have been created for this reference additivity isobole (Chou and Talalay, 1984) which may be used to test whether screening results for a drug combination are classified as antagonistic, additive, or synergistic based on Loewe's model.

An alternative definition for the expected additive effect of a combination of independently acting drugs was proposed a decade later by Bliss (1939). This method, often termed "Bliss Independence", defines the expected independent (additive) effect of a drug combination by using probabilities calculated for each monotherapy in a combination which represent likelihoods that each monotherapy will cause a treated individual to reach a predefined efficacy endpoint. These monotherapy probabilities are then used to calculate the expected probability that at least one of the drugs in a combination will cause a treated individual to reach the efficacy endpoint when the drugs are administered together.

These two drug combination models, Loewe Additivity and Bliss Independence, have traditionally been the primary models used for defining drug antagonism, additivity, or synergy, and the suitability of each has been a source of much debate over the past century, with each gathering support from different groups of researchers. Given their overlapping but differently defined terminology, an agreement was reached in 1992 called the Saariselkä agreement, which defined a set of consensus terminology between the two models and stipulated that researchers should explicitly state whether their analyses utilize the Loewe or Bliss model (Greco et al., 1995; Tang et al., 2015).

Gaddum's Independence/Independent Drug Action

A third model of drug combination interactions was proposed, perhaps inadvertently, only a year after the Bliss model. In the first edition of his book, *Pharmacology*, Gaddum included a section titled "The combined effect of two drugs" which he prefaced by saying that the diagrams he uses to describe the effects of drug combinations are from the work of Loewe (Gaddum, 1940). He then proceeded to describe an isobologram similar to those presented in Loewe's work (Loewe, 1928), but which significantly differs in the regions of the isobologram which are classified as antagonistic or synergistic. Gaddum's isobologram indicates that a combination should be considered synergistic in any case where the effect of the combination is greater than the effect that can be achieved with either monotherapy, even if the combination effect is less than additive. While Loewe did make a clear distinction between combinations where effect of a drug combination is less than additive but more than either monotherapy alone vs combinations where the effect of the drug combination is less than can be achieved with either monotherapy (Loewe, 1928), he still considered both of these situations to be forms of antagonism. This makes Gaddum's interpretation a departure from Loewe's, and it effectively shifts Gaddum's non-interaction reference away from drug additivity (as was used by both Loewe and Bliss) to instead be the effect of the single most effective monotherapy.

It is not clear, however, if this change was intentional on Gaddum's part, as Gaddum stated in the text that, "The simplest way of determining how the interaction of the two drugs should be classified is to administer half the dose of A necessary for a given effect with half the corresponding dose of B. If this combination does not cause the effect the drugs are antagonists" (Gaddum, 1940, p. 354). This statement is consistent with Loewe's model, but it is inconsistent with the labels on Gaddum's isobologram, which is directly adjacent to the statement. This

description remained in *Pharmacology* through the third edition, which was published in 1948, but it was omitted in the fourth edition of the text, which was published in 1953. The section on drug combinations was omitted in its entirety starting with the sixth edition of the book, which was published in 1968 and renamed to *Gaddum's Pharmacology*, being now revised by Burgen and Mitchell given Gaddum's death a few years earlier (Burgen and Mitchell, 1968). The inconsistent and changing description of antagonism and synergy in *Pharmacology* raises the possibility that the original deviation from Loewe's theory was an oversight, possibly stemming from the process of translating Loewe's paper (which was published in German) into English. However, I have been unable to find any sources outside of Gaddum's book where he either expresses support for Loewe's theory or proposes the alterations to it that are inferred from the isobologram he included in *Pharmacology*, so I cannot conclusively determine whether or not Gaddum intended to introduce a new model of drug combination effect.

Regardless of his intentions, Gaddum's description of drug interactions in *Pharmacology* was later used by others as the basis for the model of drug combination interaction which has become known as Gaddum's Independence, HSA, or IDA (Berenbaum, 1989; Palmer and Sorger, 2017). The principle of this model is that the expected effect of a combination of one or more independently acting drugs (i.e. not antagonists or synergists) is simply the single effect of the most effective monotherapy in the combination. Despite the obvious consequence that this model predicts no benefit for individual patients when treated with a drug combination versus treating those same patients with the most effective monotherapies for each patient, IDA can confer large benefits from drug combinations at a population level by giving each patient multiple chances to receive at least one efficacious drug when the ideal drug for each patient is not known. This may often be the case, as patient responses to different drugs are heterogeneous,

so that it is unknown which drugs any individual patient will respond best to. The clinical relevance of this model was recently established by Palmer and Sorger (2017), who demonstrated that the clinical effects of many cancer drug combinations could be adequately modeled using IDA without any need for additivity or synergy. This report, along with the ease with which drug combination effects can be modeled using IDA, presented a clear opportunity for my research into developing a method to predict drug combination efficacy using monotherapy data from CCL screens. The results of this research are covered in detail in chapter 3.

Recent Availability of High Throughput Cancer Cell Line Screening Datasets

Before such a model could be created, however, I had to identify a source of monotherapy data from CCL screens, as screening the necessary number of cell lines and drugs myself was infeasible. Fortunately, numerous large CCL drug screening datasets have been publically released in the past decade, with several screens having tested hundreds of drugs in rough a thousand CCLs. Two other students in the lab and I wrote a thorough review on these publically available high-throughput CCL screens (Ling et al., 2018). Our review describes the current publically available CCL screening datasets and analyzes the properties of the drugs and cell lines available in each dataset and the overlap between datasets. These analyses, and their implications for the use of these datasets, are covered in detail in chapter 2.

Computational Approaches to Predict Drug Combination Efficacy

Given the established clinical success of using drug combinations in cancer, there is obvious utility in developing new drug combinations for cancer treatment. Unfortunately, this is a difficult task, because there are too many possible drug combinations (many orders of magnitude more than the number of possible monotherapies) to test them all experimentally. To

overcome this problem, efforts have been made to develop computational methods that can identify promising drug combinations before physically testing them. So far, these methods have mostly focused on estimating drug synergy, where the effect of a drug combination is greater than the additive effect of the drugs in the combination. Such models have been developed using a variety of modeling approaches, including those based on mechanistic understandings, drug similarity, known interaction frequencies, and machine learning (Bulusu et al., 2016; Weinstein et al., 2017).

A recently completed effort to improve these predictive models for drug synergy was initiated by DREAM Challenges (dreamchallenges.org) in partnership with Astrazeneca and the Sanger Institute (Menden et al., 2019). The challenge gave 160 research teams access to one of the largest available drug combination screens and tasked them with developing novel approaches for predicting drug synergy based on information such as gene expression, monotherapy response, drug structure, and drug mechanisms. While many of the developed methods performed near the limits of experimental reproducibility in the training dataset, applying the models to an independent screen by O'Neil et al. (2016) resulted in performance that was little better than random classification. These results suggest that significant challenges remain to be overcome before such methods become clinically useful on a large scale—though they certainly do nothing to limit the immense potential that predictive models of synergy have for transforming drug combination design in the future.

THESIS MOTIVATION AND HYPOTHESIS

Given the urgent need for computational models to predict drug combination efficacy, the challenges associated with creating such models based on drug synergy, and the large number of researchers already working to overcome the limitations of synergy based models, I chose to

explore non-synergy based approaches for predicting drug combination efficacy. Since the earliest drug combination trials in cancer, researchers have considered the possibility that drug combinations confer patient benefit via drug independence rather than drug synergy (Frei et al., 1961). The rationale for this idea is that patient populations are phenotypically heterogeneous such that different patients may respond best to different drugs. As a result, a patient population may benefit from a drug combination simply because the combination gives each patient multiple chances to receive the most effective drug for them and not because the drugs in the combination are interacting additively or synergistically.

As previously mentioned, this idea is consistent with the IDA model, which Palmer and Sorger (2017) recently showed to have high clinical relevance. In addition to its suggested clinical relevance, an algorithm based on IDA is attractive because of its ability to directly calculate drug combination efficacy using monotherapy drug screens without the need for large drug combination training datasets. Since numerous large monotherapy datasets already exist which have tested hundreds to thousands of compounds in dozens to hundreds of cell lines (Ling et al., 2018), this allows efficacy predictions to be made for hundreds of thousands of 2-drug combinations and hundreds of millions of 3- and 4-drug combinations using existing datasets.

Thus, I hypothesized that an algorithm could be constructed based on IDA which could use monotherapy CCL screening data to make clinically meaningful predictions of drug combination efficacy. In this thesis, I present the research resulting from that hypothesis in two primary parts. Part 1 is the characterization of the available CCL screening datasets necessary for the drug combination model. This work is presented in chapter 2. Part 2 of the research is the actual development, validation, and use of the drug combination model, which I named IDACombo. This work is presented in chapter 3.

CHAPTER 2: EVALUATING THE PROMISES AND PITFALLS OF HIGH THROUGHPUT CANCER CELL LINE SCREENS

INTRODUCTION

Purpose and Rationale of This Research

CCLs have been used in pre-clinical research for decades to evaluate drug efficacy prior to advancing to more costly and difficult in vivo studies. This is in part because CCLs represent an easy-to-manipulate system for high-throughput drug and genomic screens on a scale simply unattainable in animal and patient settings due to safety, ethical, and logistical concerns. Additionally, automated liquid handling systems have made it possible to quickly screen thousands of compounds against many hundreds of CCLs while technological advances in genome sequencing have allowed detailed genomic characterization of each cell line screened. Improvements in RNA interference and genome editing technologies have also enabled genome-wide shRNA and CRISPR-Cas9 CCL screens to interrogate the necessity of nearly every gene in the genome in hundreds of CCLs. This wealth of drug sensitivity and genomic data has led to the clinical approval of bortezomib for myeloma treatment (Shoemaker, 2006), the initiation of several ongoing clinical trials (Holbeck et al., 2017), and numerous attempts to discover biomarkers associated with cancer drug response.

Despite these accomplishments, there is still much work to be done in developing new drugs and drug biomarkers to treat what remains an expansive list of poor prognosis cancers. Given the critical role of CCL screens in this effort, Jessica Fessler, Robert Gruener, and I set out to create a comprehensive list of publically available CCL screening datasets and to examine how well these screens currently represent both the diversity of human cancers encountered in the clinic and the diversity of targetable pathways that are known to be deregulated in cancer.

Jessica Fessler began this project by identifying a number of available datasets and beginning a manuscript describing those datasets. Robert Gruener and I then identified additional screening datasets, harmonized the cell line and drug identifiers between each dataset, analyzed how well the CCLs and drugs used to create each dataset represented the spectrum of human cancers and the targetable pathways in cancer, and assessed the extent of overlap between each screen. These results, along with all of the figures and tables in this chapter, were published as a review article (Ling et al., 2018), and they ultimately provided us with a detailed knowledge of existing datasets and their strengths and limitations. Importantly, this project also provided us with a means to harmonize drug and CCL identifiers between screens so that I could use multiple screening datasets together in my eventual efforts to translate monotherapy screening data into predictions of drug combination efficacy.

A Brief History of Cancer Cell Line Screening

Numerous CCL screens have been performed over the last three decades, and the format and size of these screens has varied considerably. A very brief description to the history of these screens is provided in this section to demonstrate the progression of CCL screens. Given the large number of CCL screens available to date, we chose to focus most of our attention on large-scale screens with cell lines representing multiple cancer types. That said, several screens comprised of only a single cancer type are briefly mentioned by virtue of their being either relatively large or important in developing screening techniques. Furthermore, we only considered screens with publicly available screening data. A brief tabular summary of reviewed screens is included in Table 1. Individual summaries for each screen can be found in our published review on this subject (Ling et al., 2018).

Study Name	Type of Screen	Institution	# Cell Lines	# of Tested Reagents	Source
Pan-Cancer CCL Screens					
NCI60	compound	NCI	74	49,278 compounds	https://wiki.nci.nih.gov/display/NCIDTPdata/
GlaxoSmithKline	compound	GlaxoSmithKline	310	19 compounds	Greshock et al., 2010
CGP/GDSC	compound	Wellcome Trust Sanger Institute and Massachusetts General Hospital Cancer Center	1073	249 compounds	Garnett et al., 2012; Iorio et al., 2016
CCLE	compound	Broad Institute	503	24 compounds	Barretina et al., 2012
CTRP v1	compound	Broad Institute	242	354 compounds	Basu et al., 2013
CTRP v2	compound	Broad Institute	887	496 compounds	Seashore-Ludlow et al., 2015
gCSI	compound	Genentech	429	16 compounds	Haverty et al., 2016
FIMM	compound	Institute for Molecular Medicine Finland	50	52 compounds	Mpindi et al., 2016
NCI-ALMANAC	drug combination	NCI	60	104 compounds (5,334 combinations)	Holbeck et al., 2017
Single-Cancer CCL Screens					
Daemen et al. (Breast Cancer)	compound	Lawrence Berkeley National Laboratory	70	88 compounds	Daemen et al., 2013
Colorectal Cancer Organoid Screen	compound	Hubrecht Institute, Wellcome Trust Sanger Institute, Broad Institute	19 (organoids)	83 compounds	van de Wetering et al., 2015
NCI-Sarcoma Project	compound	NCI	64	440 compounds	Teicher et al., 2015
NCI-SCLC Project	compound	NCI	70	515 compounds	Polley et al., 2016
PRISM (NSCLC)	compound	Broad Institute	96	374 - 8,400 compounds	Yu et al., 2016
shRNA & CRISPR/Cas9 CCL Screens					
Achilles 2.0	shRNA	Broad Institute	102	54,020 shRNAs (targeting 11,194 genes)	Cheung et al., 2011
Achilles 2.4.3	shRNA	Broad Institute	216	54,020 shRNAs (targeting 11,194 genes)	Cowley et al., 2014
Achilles 2.20.2	shRNA	Broad Institute	501	107,523 shRNAs (targeting 25,579 genes)	Tsherniak et al., 2017
Achilles 3.3.8	CRISPR/Cas9	Broad Institute	33	123,411 sgRNAs (targeting 19,060 genes)	Aguirre et al., 2016
Munoz et al., 2016	CRISPR/Cas9 & shRNA	Novartis Institutes for Biomedical Research	5	51,413 sgRNAs (targeting 2707 genes)	Munoz et al., 2016
Tzelepis et al. (AML)	CRISPR/Cas9	Wellcome Trust Sanger Institute	7	90,709 sgRNAs (targeting 18,010 genes)	Tzelepis et al., 2016
Wang et al. (AML)	CRISRP/Cas9	Broad Institute	14	187,536 sgRNAs (targeting 18,543 genes)	Wang et al., 2017

Table 1. Available *in vitro* CCL screen datasets. This table provides summary information for the CCL screens we review in this article. Cell line and compound numbers reflect the latest releases of each dataset, with duplicated cell lines in each study being counted as a single cell line, and only cell lines with available screening data included.

CCL Drug Screens

To date, the majority of large-scale CCL drug screens have been performed either by the National Cancer Institute (NCI) or the Broad Institute, with other notable screens having also been performed by GlaxoSmithKline, the Sanger Institute, the MGH cancer center, Genentech, the Institute for Molecular Medicine Finland, Novartis, and Berkeley National Laboratory (see Table 1 for references). The NCI began one of the first major CCL drug screening efforts in the

1980's when it created the NCI60—an initiative to screen large numbers of known and novel compounds against a small group of CCLs. In its long history, the NCI60 has screened over 100,000 compounds and has been responsible for leading to a number of important cancer-related drug discoveries (Shoemaker, 2006). Among its breakthroughs, the NCI60 was used in the first study which integrated analysis of both molecular pharmacology and gene expression data for a large set of CCLs (Scherf et al., 2000). Since then, the genomes and transcriptomes of the CCLs used in drug screens have been extensively characterized, paving the way for studies which integrate genomic and drug sensitivity data.

Rather than focusing on screening huge compound libraries like the NCI-60, more recent large-scale CCL screens have focused on smaller compound sets screened in much larger numbers of CCLs. This approach has increased the genetic diversity of the screens, which has allowed researchers to draw links between genetic features and drug sensitivity. Tissue-specific cell line screens are also becoming more common. These screens usually contain more cell lines for a given cancer type compared to the pan-cancer screens and can be useful for interrogating diversity within a particular cancer type. Finally, recent advances in technology and screening techniques have allowed high-throughput screening of 2-drug combinations in Project NCI-ALMANAC (Holbeck et al., 2017) as well as the ability to pool CCLs together into single wells/tumors for both in vitro and in vivo drug screening using the PRISM method (Yu et al., 2016).

CCL shRNA/CRISPR-Cas9 Screens

The advent of high-throughput methods for introducing shRNAs and gene edits via CRISPR-Cas9 into CCLs has allowed scientists to assess individual gene function by knocking down/out genes in panels of CCLs. By far the largest project using these methods is Project

Achilles at the Broad Institute (see Table 1). Project Achilles has gradually increased the number of CCLs screened with shRNAs since first publishing in 2011 to now include 501 CCLs screened with shRNAs covering >25,000 gene products. Project Achilles also includes a CRISPR/Cas9 screen, which identified certain liabilities with CRISPR/Cas9 screening that could lead to a high rate of false-positive gene dependencies (Aguirre et al., 2016). This problem has been further researched by multiple groups, leading to the development of potential solutions and enabling the use of CRISPR/Cas9 screens to discover novel gene dependencies and drug sensitivities in CCLs (Munoz et al., 2016; Tzelepis et al., 2016; Wang et al., 2017).

MATERIALS AND METHODS

Identifying High-Throughput Cancer Cell Line Screens

Jessica Fessler began identifying CCL screens before I joined Stephanie's lab. After Jessica left and Robert Gruener and I took over this project, I used PubMed and online search engines to manually search the literature for high-throughput CCL screens that tested dozens to hundreds of compounds in dozens to hundreds of cell lines or which performed genome-wide shRNA or CRISPR/Cas9 screens in dozens to hundreds of cell lines. Priority was given to studies which screened multiple types of cancer, but some single cancer screens were included if they contained a relatively large number of cell lines for that cancer type or if that study provided technical innovations which impacted the field of CCL screening. As previously mentioned, the selected screens are summarized in Table 1. After identifying the studies, I wrote short text summaries for each study, a process which Jessica had begun for several studies before I joined the lab. These summaries are included in the review article which resulted from this project (Ling et al., 2018), but they are omitted in this thesis as they are not necessary for understanding either the analyses presented in this chapter or those presented in chapter 3.

Harmonizing Cell Line Identifiers between Cancer Cell Line Screens and Annotating Cell Lines

I began by downloading cell line information from Cellosaurus, which is an online repository of cell line information that includes detailed records for nearly 115,000 cell lines. I then downloaded the xml version of Cellosaurus (https://web.expasy.org/cellosaurus/) on 8/18/2017 and used R version 3.4.2 (R Core Team, 2017) with the XML R package (Lang and the CRAN Team, 2017) to parse the xml file to extract species, gender, disease type, ethnicity, commonly used names, Cellosaurus identifiers, and BioSample IDs for each cell line. Using R, I was then able to match the Cellosaurus identifiers to the cell line names used in each of the CCL screening datasets.

For further annotation, I downloaded the xml version of BioSample (Barrett et al., 2012) on 10/23/2017, which is an NCBI repository for descriptions of cell lines. I parsed the xml file to extract cell line BioSample identifiers, tissue source, disease type, sex, ethnicity, and age for each cell line using the same process as was used for Cellosaurus. This allowed to me to add the BioSample information for each cell line using the BioSample IDs which were present from both Cellosaurus and BioSample.

I then manually curated the cell line information, occasionally using COSMIC (Forbes et al., 2017) to fill in information that was missing from Cellosaurus or BioSample. This ultimately resulted in a dataset which included the original names used for each cell line in each dataset, a harmonized identifier from Cellosaurus, alternative cell line names, gender, ethnicity, age, and disease type information for each cell line used in any of the CCL screens selected for the study. This information is included in Table S1.

Harmonizing Compound Identifiers between Cancer Cell Line Screens and Annotating Compounds

The majority of the work to harmonize and annotate the compounds used in the CCL screens in our analysis was performed by Robert Gruener. He accomplished this by using PubChem's Identifier Exchange Service (https://pubchem.ncbi.nlm.nih.gov/idexchange/idexchange.cgi) to identify synonyms for all named compounds in the original datasets and to convert these synonyms to PubChem IDs. He then matched PubChem IDs back to their original name and used the PubChemID(s) identified to match the compounds among the datasets. This ensured the highest degree of overlap, and the results were manually checked and curated as needed to ensure correctness, an effort to which I helped to a very small degree. This method also identified overlap in highly related compounds, for example irinotecan and its active metabolite SN-38. In most circumstances, overlap based on very close relatedness such as this was considered appropriate and kept. Additionally, the Broad Drug Repurposing Hub (Corsello et al., 2017) was used to add information such as molecular targets and clinical phase to the information already provided for each compound by the screens in which they were used. These efforts resulted in a list of drugs with their original identifiers from each CCL screen, a harmonized identifier, their clinical phase, their general mechanism of action, and their specific molecular targets. This information is included in Table S2.

Efforts were also made to determine which biological pathways were affected by each compound. Robert and I both contributed to this effort, which was accomplished by utilizing the Broad Institute's MSigDB database (Liberzon et al., 2015; Subramanian et al., 2005). We used the database’s Canonical Pathways gene set (Fabregat et al., 2018; Kanehisa et al., 2017; Liberzon et al., 2015; Milacic et al., 2012) (http://www.biocarta.com/) to represent general cell

biology pathways as well as its Cancer Modules (Segal et al., 2004) and Oncogenic Pathways to represent cancer specific pathways. I then used R to match the molecular targets of each compound to the pathways which included those targets, allowing us to determine which pathways each compound targeted.

RESULTS

Screened Cancer Cell Lines Represent a Diverse Range of Cancers and Age Groups, but Gender and Ethnic Diversity Need to be Improved

Human cancers are a complex set of diseases which can arise from essentially every tissue in the body and vary widely in both incidence and mortality. As such, CCL screens are faced with the task of balancing a need to select diverse sets of CCLs which represent a wide range of human cancers with the need to collect a large enough set of CCLs for any individual cancer type to capture the genetic and phenotypic heterogeneity within that cancer. To assess how well current CCL screens have accomplished this task, we identified screened CCLs by name using Cellosaurus (https://web.expasy.org/cellosaurus/) and then matched age, gender, and ethnicity information from a combination of Cellosaurus, the BioSample database (Barrett et al., 2012), and COSMIC (Forbes et al., 2017). An examination of the cell lines used in the studies covered by this review (see Table S1) shows that there are over 1,600 unique cell lines screened among the 20 datasets, covering more than 30 tissues of origin and over 200 cancer types/subsets. These include both common cancers (e.g. breast cancers) and extremely rare cancers (e.g. leiomyosarcomas) as well as highly specific cancer subsets (e.g. B-cell prolymphocytic leukemia). Interestingly, the proportions of CCLs representing any given cancer type correlate well with the American Cancer Society's (Siegel et al., 2017) estimated number of deaths from those cancers (Figure 1A, R^2=0.71, data collected by Robert and plotted by myself).

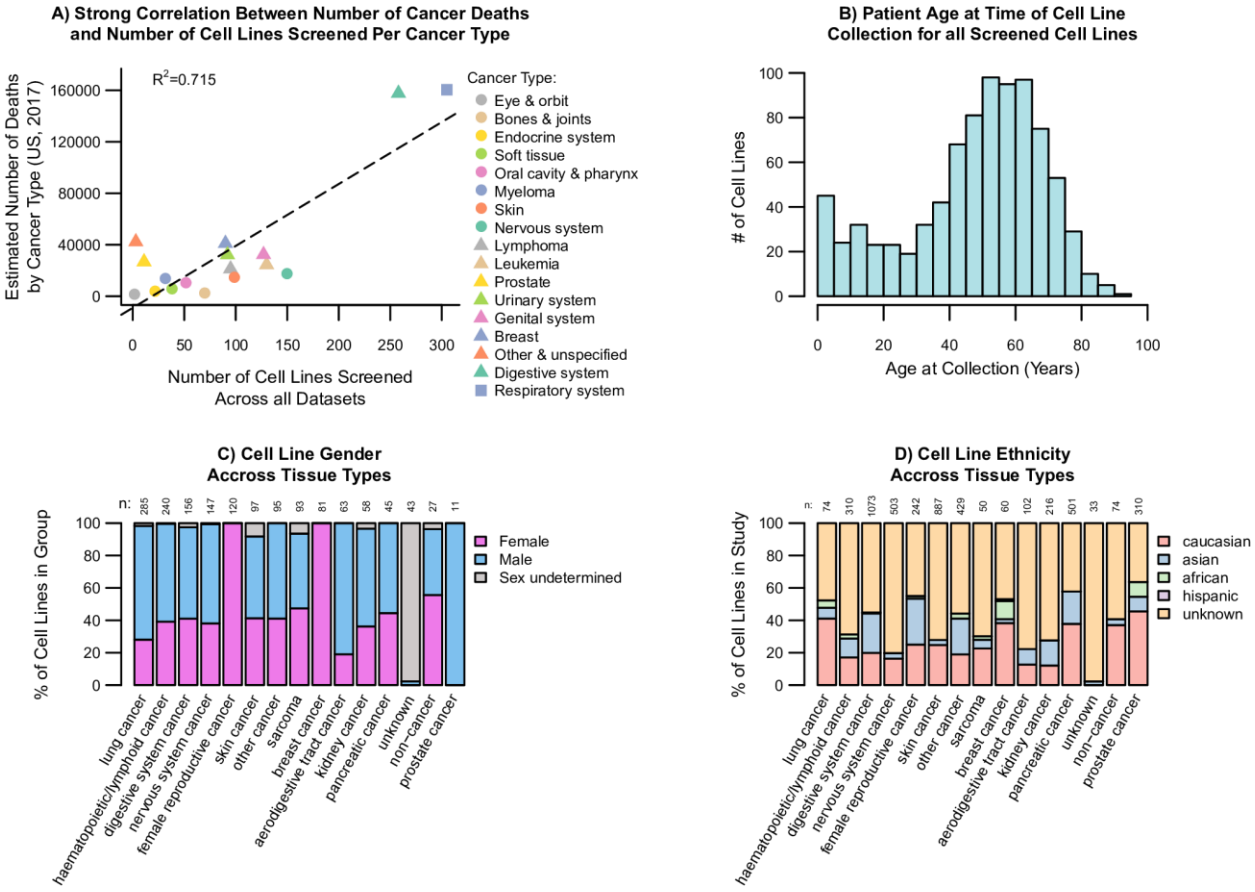

Figure 1. Screened CCLs correlate with cancer fatality while capturing age, gender, and ethnicity to varying extents. Data for this figure is included in Table S1. **A)** Correlation is shown between cancer mortality (obtained from Siegel et al., 2017) and the number of unique cell lines screened from each cancer type. Cancer type was determined by bioinformatic and manual curation using Cellosaurus, the BioSample database, COSMIC, or annotations provided by the datasets themselves. Only cell lines with available screening results are included. **B-D)** As with **part A**, age of collection, gender, and ethnicity for screened CCLs were determined by bioinformatic and manual curation using Cellosaurus, the BioSample database, and COSMIC. **Part B** shows the number of unique cell lines collected from patients at given ages, while **parts C and D** show the distribution of genders and ethnicities respectively for screened cell lines from each cancer type.

This suggests that current screens have been relatively successful in capturing the diversity of human cancers while prioritizing the cancers that cause the most deaths.

Beyond representing a diversity of human cancers, currently screened CCLs also represent a wide range of ages of onset in human cancer as well as relatively even proportions of male and female cancers (Figures 1B and 1C), enabling these datasets to be used to study age-

and gender-specific phenomenon in cancer drug sensitivity. Certain cancer types do, however, deviate significantly from clinical gender proportions. In lung cancer, for example, there are more than twice as many screened male CCLs as there are female CCLs (207 male, 84 female, Figure 1C). This disparity can be especially pronounced in cancer types with fewer screened CCLs such as liver cancer, for which there are 24 male CCLs and only 4 CCLs that we identified as female. If potential gender differences are to be studied in *in vitro* drug sensitivity for these cancers, care will need to be taken in designing future CCL screens to adequately represent both males and females in each cancer type.

Unlike age and gender, ethnicity is generally poorly annotated for CCLs. Based on the sources mentioned above, the ethnicity of many screened CCLs is unknown; however, most screened CCLs of known ethnicity are of either Caucasian or Asian descent (Figure 1D), with 87% of the Asian CCLs being Japanese. CCLs of African and Hispanic descent are particularly poorly represented. We were unable to identify any African CCLs in 7 of the 15 cancer categories, and Hispanic CCLs were absent in all but 1 of the categories.

Screened Drugs Target a Broad Array of Pathways Relevant to Cancer

Beyond including a diverse set of CCLs, it is also important for these screens to target a diverse set of molecular targets and to include compounds from all stages of the drug development pipeline. As for molecular targets, the shRNA and CRISPR screens easily cover the most molecular targets, with most screens targeting over 11,000 and even up to 25,000 gene products. Of the 14 compound-based screens, NCI-60 contained by far the most compounds (approximately 49,000) with the remaining screens collectively containing 1329 unique compounds. Given the large number of poorly characterized compounds screened in NCI-60, we were only able to identify molecular targets for 1.6% of NCI-60 compounds. For screens other

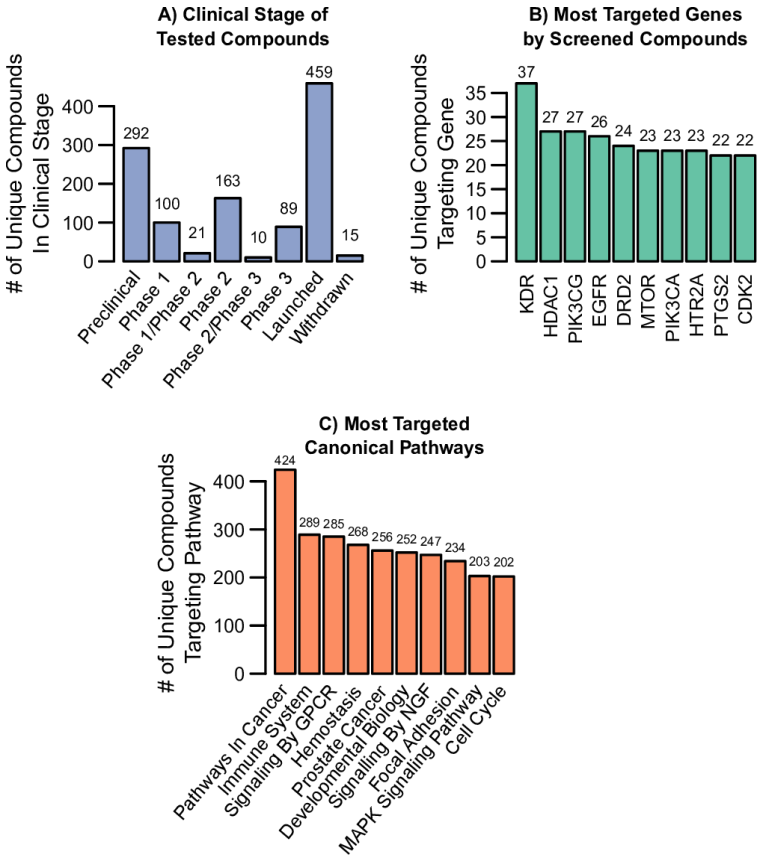

Figure 2. Targets and clinical stage of compounds in CCL screens. Compounds used in Figure 2 are the 1,207 unique compounds from the 14 CCL drug screens reviewed in this paper with targeted information from the CCL screens or the Broad DRH. **A)** The clinical stage distribution for the drugs with current clinical trial information from Broad DRH **B)** Shows the ten most commonly targeted genes and the number of unique compounds against them. **C)** Shows the 10 most commonly targeted pathways in MSigDB's Canonical Pathway Gene Set (C2:CP) based on the number of unique compounds which target at least one gene target in that pathway.

than NCI-60, we were able to identify molecular targets for 70% of tested compounds. Completion rates for finding clinical trial stage information were similar but slightly lower, with our results suggesting that 857 of the screened compounds are currently in or have completed testing in clinical trials (Figure 2A).

1,234 genes were impacted by these drugs, with many genes targeted by multiple compounds. Figure 2B shows the ten most frequently targeted genes, which were each targeted by at least 22 unique compounds in the CCL screens we reviewed**.** Encouragingly, many of these top gene targets are recognizable as important in cancer.

To further investigate the role of these targeted genes in general cell biology and cancer pathways, we investigated the pathways targeted by each compound. While the most commonly targeted canonical biology pathway was unsurprisingly "pathways in cancer," many other biologically significant pathways are also impacted (Figure 2C). Indeed, 1,234 of the 1,329

canonical biology pathways are impacted by at least one compound, with a median of 21 unique drugs impacting a given pathway. Regarding cancer specific pathways, 592 of the 620 pathways were impacted by at least one compound, with a median of 28 unique compounds per pathway. Overall, the coverage of the majority of the general biology pathways and cancer specific pathways along with the proportion of drugs approved or in clinical trials suggests that CCL screens have, in general, selected a relevant yet broad array of compounds for screening.

Extensive Cell Line Overlap Exists Between Different Cancer Cell Line Screens

To investigate how similar the CCL screens are in terms of the types of cancers they screened, we used the information in Table S1 to determine the relative abundance of a cancer type in the pan-cancer datasets. Figure 3 shows that the proportion of represented cancer types is largely similar across pan-cancer datasets, with the most prominent deviations from the average occurring in the datasets that screened the fewest cell lines. For instance, the NCI-60 and NCI-ALMANAC screening sets omit sarcomas and pancreatic cancer cell lines, the Achilles v3.3.8 dataset focuses more specifically on sarcomas and pancreatic cancers than other datasets, and the FIMM dataset has a larger proportion of haematopoietic/lymphoid, breast, and female reproductive cancers. However, on the whole, many of the large pan-cancer studies represent the various cancer types in similar proportions.

Some of this similarity can be explained by the fact that many of the pan-cancer datasets use highly overlapping sets of CCLs for their screens (Figure 4A). For example, CTRPv2 contains at least 79% of the cell lines used in any other Broad Institute screen, and NCI-

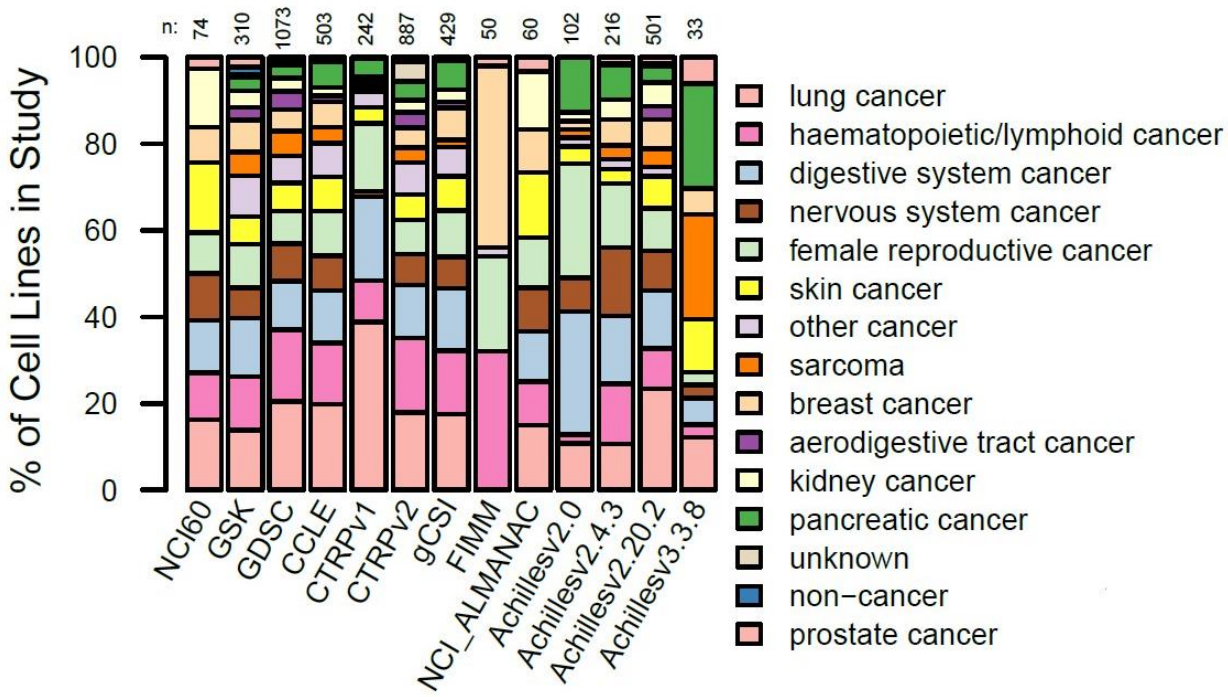

Figure 3. Many tissue types are represented in the assessed CCL screens. Cell line tissue type vs. dataset. Tissue type was determined by bioinformatic and manual curation using Cellosaurus, the BioSample database, COSMIC, or annotations provided by the datasets themselves, and then similar cancer types were grouped in the broad groups shown. The data for this figure is included in Table S1. Note that cancer types are colored at the right in the reverse order (from top to bottom) as they are plotted in the bargraph (from top to bottom), such that the top cancer type in the legend (lung cancer) is plotted as the lowest portion of each bargraph (0% to x).

ALMANAC is entirely composed of CCLs from NCI60. Of the 1,561 unique CCLs screened across all studies, 594 are unique to a single study, 229 are included in two studies, and 738 are included in three or more studies (Figure 4B), with some cell lines being included in most screens—such as the A549, a lung adenocarcinoma line which is included in 14 of the 20 datasets. The 594 CCLs that are unique to a single screen are mostly used in either GDSC or CTRPv2, the two datasets with the greatest number of cell lines, though several of the single-cancer screens also contribute a number of unique cell lines (Figure 4C).

Figure 4. Legend on page 25.

Figure 4, continued. Cell line overlap and frequency between CCL screens.
A) Heatmap of cell line overlap between reviewed studies. Overlap is based on Table S1, with each color scale being relevant to the amount of overlap each column study has with the study in that row. Note that only cell lines with available screening data (compound, shRNA, or CRISPR/Cas9) are counted for each dataset. Of particular importance, this means the overlap between CTRPv2 cell lines and CCLE cell lines with genomic data is much higher than that represented here. **B)** Distribution of the number of datasets in which each screened cell line is included. **C)** Number of cell lines that are unique to a single dataset. Note that this also only includes cell lines with available screening data, such that many of the unique CTRPv2 cell lines are unique in the sense of being used for a drug screen, but they are present without pharmacological profiling in CCLE.

Extensive Compound Overlap Also Exists Between Cancer Cell Line Screens

In total, there are over 50,000 unique screening agents with publicly available data in these datasets, most of which can be attributed to NCI-60. NCI-60 has data for close to 49,300 compounds with almost 49,000 of these agents being unique to the NCI-60 screen (Figure 5C). However, it should be noted that the majority of these compounds failed to meet NCI-60's screening standards by either missing the minimum range requirements, not passing a minimum consistency among replicates, or by having results for fewer than 35 cell lines. Taking this into consideration, only ~21,000 compounds are both publicly available from the NCI-60 and passed their standards. Comparatively, the other CCL screens we reviewed screened a combined total of approximately 2,800 agents, of which about 1,300 are unique (Table S2). Similar to the relatively high overlap among cell lines in these datasets, there was an appreciable amount of overlap among the drugs screened (Figure 5A). There are 766 compounds overlapped in at least 2 datasets, about 240 of which overlapped in 4 or more datasets (Figure 5B).

DISCUSSION

Over the past three decades, CCL screens have done much to aid in the development and understanding of cancer treatments, and they have also done much to improve themselves over time. That said, as with all cancer research, much remains to be done. While our results

Figure 5. Legend on page 27.

Figure 5, continued. Compound overlap and frequency between CCL screens.
A) Heatmap of compound overlap between reviewed studies. Overlap is based on data from Table S2, with each color scale being relevant to the amount of overlap each column study has with the study in that row. Note that only compounds with available screening data are counted for each dataset, and duplicate tests of the same compound were excluded. **B)** Distribution of the number of datasets in which each screened compound is included. Note that the 8,000 diversity-oriented synthesis molecules tested in PRISM are excluded in this plot. **C)** Number of compounds that are unique to each dataset. Note that the 8,000 diversity-oriented synthesis molecules tested in PRISM are also excluded in this plot.

demonstrate that numerous cancer types are represented in these screens in numbers proportionate to the mortality rate of each cancer, many cancer types are still represented by only a handful of CCLs. This seriously limits the power of any studies seeking to identify drug sensitivities or genomic associations that are unique to these poorly represented cancers, and it greatly increases the likelihood that the full range of disease specific genotypes and phenotypes are not captured by the few CCLs representing those diseases.

Likewise, while our analysis suggests that, while both male and female CCLs are well represented in most cancers, severe gender imbalances do exist in certain cancer types, such as liver cancer. This effectively prevents the use of these screens for efforts to identify gender-specific drug sensitivity associations in these cancer types. Perhaps even more concerning, ethnicity information for the CCLs used in these screens is largely missing, with nearly all recorded ethnicities being Caucasian or Asian across all cancer types. As it is becoming increasingly apparent that ethnicity affects cancer progression and treatment response (Sekine et al., 2008; Keenan et al., 2015; Costa and Gradishar, 2017), this raises two primary concerns about discoveries made using currently available CCL screens. First, the probable lack of ethnic diversity suggests that some of these discoveries may not translate to patients from underrepresented ethnic groups. Second, the high level of uncertainty in CCL ethnicity prevents researchers from properly controlling for ethnicity while searching for pharmacogenomic

associations—greatly reducing the ability of these datasets to detect ethnic-specific associations. Given its importance to the utility of these datasets, ethnicity will need to be carefully considered and recorded when designing the next generation of CCL screens, and efforts to improve ethnicity information for existing CCLs should be considered (i.e. by contacting labs/institutions who generated CCLs of unknown ethnicity or searching for literature describing the generation of these CCLs). Taken together, these results suggest that significant improvements are necessary in the number and diversity of CCLs used in these screens to adequately represent the diversity of human cancers. This will likely involve greatly expanding the number of CCLs screened in future studies, which may be especially important for identifying biomarkers relevant to targeted therapies, which are expected to only be effective in a subset of cell lines tested. Indeed, previous reports have suggested that up to 85% of the cell lines tested in some screens are insensitive to the majority of tested compounds (Bouhaddou et al., 2016), placing serious limits on the power to identify biomarkers associated with response to those treatments—limits which future screens will need to overcome.

Fortunately, others in the field are already aware of the need for additional *in vitro* cancer models to meet this challenge (Williams and McDermott, 2017). As large-scale sequencing efforts in patient tumors has revealed complex diversity and sub-grouping within cancer types, efforts have begun to generate *in vitro* cancer models which capture this diversity. Two large projects with this goal have emerged in recent years. One is the Cancer Cell Line Factory at the Broad Institute, which aims to generate more than 10,000 CCLs for use by the research facility (Boehm and Golub, 2015). The other is the Human Cancer Model Initiative collaboration between the NCI, Cancer Research UK, the Sanger Institute, and the foundation Hubrecht Organoid Technology, which aims to create as many as 1000 new *in vitro* cancer models with

detailed clinical information, carefully controlled culture condition, and modern culture techniques such as conditionally reprogrammed cells and organoids (https://ocg.cancer.gov/programs/HCMI). Patient-derived tumor xenografts have also been explored as a means of expanding the genetic diversity of pre-clinical drug screens (Gao et al., 2015). Hopefully, these efforts will greatly increase the diversity and clinical relevance of available pre-clinical cancer models for future screens.

Despite the limitations this analysis revealed in current CCL screens, however, our analysis also reveals a great many strengths to these datasets. These efforts have resulted in publically available datasets containing screening data for over 1,600 CCLs representing every major cancer type, >50,000 compounds targeting more than 93% of all known canonical and cancer related cellular pathways, and shRNAs and CRISPR/Cas9 screens which cover nearly every gene in the human genome. Furthermore, most of the CCLs used in these screens have undergone extensive molecular profiling—including whole-exome sequencing, RNA sequencing, mRNA array profiling, and proteomic profiling—which makes them ideal resources with which to identify and validate biomarkers of drug sensitivity. Furthermore, the high degree of overlap between these screens both in terms of screened compounds and CCLs provides a large number of technical replicates and the differences in screened compounds and CCLs between screens allows for independent datasets with which to validate findings from other screens.

Given these strengths, these screens remain an important tool in the effort to identify new treatments and biomarkers for cancer treatment, and the constantly ongoing work to improve the *in vitro* models and screening techniques use to generate these datasets ensures that this will continue to be the case. It is my hope that these results will help researchers in their efforts to

improve these screens, as well as in using already existing screens to their fullest potential. Whatever use other researchers find for these results, the efforts to harmonize cell line and drug identifiers between screens was essential for the success of my primary research project, which was to develop an algorithm capable of using the monotherapy data from these CCL screens to predict drug combination efficacy in the clinic. The results of that research are presented in the next chapter.

CHAPTER 3: CLINICAL TRIAL OUTCOMES FOR CANCER DRUG COMBINATIONS CAN BE PREDICTED BY MODELING INDEPENDENT DRUG ACTION USING CANCER CELL LINE SCREENS

INTRODUCTION

In the early 1960's, Frei et al. (1961, 1965) first demonstrated that drug combinations could be more effective than monotherapy in acute leukemia patients. In the decades following this discovery, drug combinations transformed the clinical outcomes for many other types of cancer beyond leukemia (Bukowska et al., 2015; DeVita and Chu, 2008; Devita et al., 1975), making drug combinations an essential part of modern cancer therapy. Unfortunately, the vast number of possible drug combinations (many orders of magnitude greater than the number of possible monotherapies) makes it infeasible to test them all experimentally when developing new therapies. To overcome this problem, efforts have been made to develop computational methods that can identify promising drug combinations before physically testing them.

As discussed in detail in chapter 1, these methods have mostly focused on estimating drug synergy using a variety of modeling approaches, including those based on mechanistic understandings, drug similarity, known interaction frequencies, and machine learning (Bulusu et al., 2016; Weinstein et al., 2017). Unfortunately, largescale efforts to use these models to predict drug synergy have resulted in performance that was little better than random classification (Menden et al., 2019). These results suggest that significant challenges remain to be overcome before such methods become clinically useful on a large scale—though they certainly do nothing to limit the immense potential that predictive models of synergy have for transforming drug combination design in the future. Given these challenges, and the large number of researchers

already working to overcome them, I chose to explore non-synergy based approaches for predicting drug combination efficacy.

Since the earliest drug combination trials in cancer, researchers have considered the possibility that drug combinations confer patient benefit via drug independence rather than drug synergy (Frei et al., 1961). The rationale for this idea is that patient populations are phenotypically heterogeneous such that different patients may respond best to different drugs. As a result, a patient population may benefit from a drug combination simply because the combination gives each patient multiple chances to receive the most effective drug for them and not because the drugs in the combination are interacting additively or synergistically.

While, as discussed in detail in chapter 1, there are multiple theories for how drug combination efficacies should be calculated when drugs act independently (Foucquier and Guedj, 2015), I chose to focus on IDA, which hypothesizes that the effect of a drug combination may be merely the effect of the single most effective drug in the combination. Evidence for the clinical relevance of this model was recently provided by Palmer and Sorger (2017), who examined 8 clinical trials that tested drug combinations in cancer and concluded that IDA could explain the results of 5 of the trials without the need for synergistic drug action. While identifying synergistic combinations likely remains an essential part of finding the most effective drug combinations, these results suggest that IDA is a clinically relevant mechanism by which drug combinations can improve patient outcomes in cancer.

In addition to its suggested clinical relevance, an algorithm based on IDA is attractive because of its ability to directly calculate drug combination efficacy using monotherapy drug screens without the need for large drug combination training datasets. Since numerous large monotherapy datasets already exist which have tested hundreds to thousands of compounds in

dozens to hundreds of cell lines (Ling et al., 2018), this allows efficacy predictions to be made for hundreds of thousands of 2-drug combinations and hundreds of millions of 3- and 4-drug combinations using existing datasets. In this paper, I present an IDA based algorithm for predicting drug combination efficacies using *in vitro* monotherapy screening data. I validate its predictions using both an *in vitro* dataset which measured monotherapy and drug combination efficacies and clinical results from a systematically selected set of phase III clinical trials. Furthermore, I use the algorithm to prospectively predict the efficacies of thousands of 2-drug combinations in 27 cancer types/subtypes and demonstrate how those predictions can be used to quickly identify candidate drug combinations for future clinical development.

MATERIALS AND METHODS

Data and Software Availability

The data and code necessary to reproduce the analyses in this paper, along with detailed readme files to aid investigators in navigating and understanding each analysis and script in the project, have been uploaded in their entirety to Open Science Framework (OSF, https://osf.io/). They are stored in the "IDACombo Paper" project, which can be accessed using this link. Figure 6 shows how the various analysis folders in this project relate to each other, and also gives a general overview of the experimental steps performed in this chapter. Since the readme files included in the OSF directories provide detailed information about the purpose of each script in the analysis, they serve as a detailed description of the experimental methods used to complete the analyses performed in this chapter. As such, I have included these readme files in Appendix I.

Most analyses were performed using R v3.4.2 (R Core Team, 2017) with Microsoft R Open v3.4.2 (Microsoft, R Core Team, 2017) and RStudio v1.1.463 (RStudio Team, 2015).

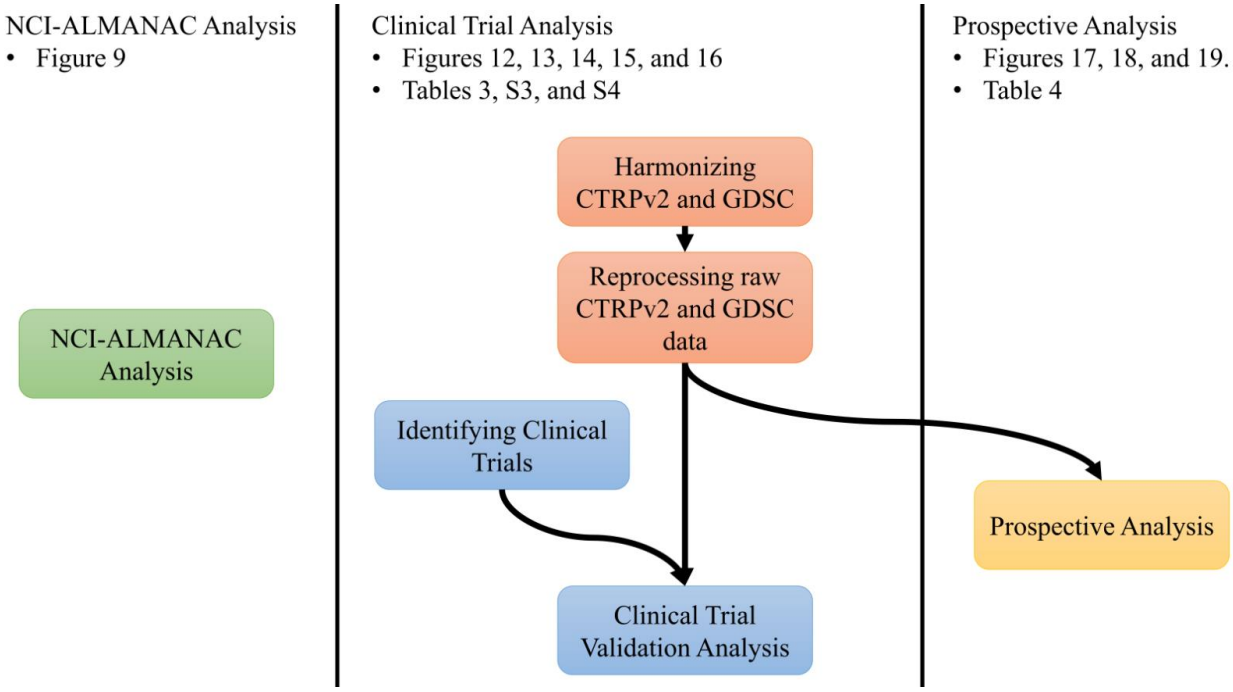

Figure 6: Analysis pipeline in "IDACombo Paper" OSF project. This figure outlines how the various analysis folders in the OSF project for the work complete in this chapter are related to each other and to the figures in this chapter. This is primarily included here to assist readers in interpreting Appendix I, but it also provides a general overview of the workflow for the analyses performed in chapter 3.

Processing of the raw dose-response data from CTRPv2 and GDSC was performed using the Mesabi compute cluster at the Minnesota Supercomputing Institute (MSI) at the University of Minnesota (http://www.msi.umn.edu) and R v3.4.4.

The IDACombo R package created for this analysis is available on GitHub. Additional R packages used in the analysis are listed in Table 1 along with their citations and web-links.

IDA and Bliss Independence drug combination efficacy predictions with IDACombo

As shown in Figure 7A, IDA predictions of drug combination efficacy are produced by predicting that the effect of a combination of 2 or more drugs on a cell line will be equal to the effect of the single most efficacious drug in the combination. The efficacy of the drug combination is then summarized by calculating the mean predicted efficacy across all cell lines

Package Name	Package Version	Package Citation	Package WebLink
car	2.1.5	Fox and Weisberg, 2011	https://CRAN.R-project.org/package=car
ComplexHeatmap	1.14.0	Gu et al., 2016	https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
drc	3.0.1	Ritz et al., 2015	https://CRAN.R-project.org/package=drc
IDACombo	1.0.0	This chapter.	
parallel	3.4.2	R Core Team, 2017	Created by the R Core team and included in R since R version 2.14.0.
pbapply	1.3.3	Solymos and Zawadzki, 2017	https://CRAN.R-project.org/package=pbapply
powerSurvEpi	0.0.9	Qiu et al., 2015	https://CRAN.R-project.org/package=powerSurvEpi
precrec	0.9.1	Saito and Rehmsmeier, 2017	https://CRAN.R-project.org/package=precrec
progress	1.1.2	Csárdi and FitzJohn, 2016	https://CRAN.R-project.org/package=progress
RColorBrewer	1.1.2	Neuwirth, 2014	https://CRAN.R-project.org/package=RColorBrewer
readr	1.1.1	Wickham et al., 2017	https://CRAN.R-project.org/package=readr
readxl	1.0.0	Wickham and Bryan, 2017	https://CRAN.R-project.org/package=readxl
rgl	0.98.1	Adler et al., 2017	https://CRAN.R-project.org/package=rgl
rvest	0.3.2	Wickham, 2016	https://CRAN.R-project.org/package=rvest
sandwich	2.4.0	Zeileis, 2004, 2006	https://CRAN.R-project.org/package=sandwich
xlsx	0.5.7	Dragulescu, 2014	https://CRAN.R-project.org/package=xlsx

Table 2: R packages used in chapter 3 analysis. This table lists all of the R packages used in the analyses in this chapter, along with their version numbers, citations, and weblinks.

being used in the analysis, and this average efficacy is used in downstream analyses. This equates to equation 1 below, where $\mu_{combo,IDA}$ is the mean IDA predicted efficacy of a combination of drugs A to Z in n cell lines and where EA_k and EZ_k are the respective efficacies of drugs A and Z in cell line k.

$$\mu_{combo,IDA} = \frac{\sum_{k=1}^{n} min(EA_k, \ldots, EZ_k)}{n} \quad (1)$$

Note that this is well defined for any efficacy metric where a decrease in the efficacy metric indicates a decrease in the ability of a drug to kill cells (i.e. for a metric such as viability relative to untreated cells). If a decrease in the efficacy metric indicates an increase in the ability

of a drug to kill cells (i.e. if the used metric is viability reduction, etc.), then the equation must be modified to equation 2.

$$\mu_{combo,IDA} = \frac{\sum_{k=1}^{n} max(EA_k, \ldots, EZ_k)}{n} \quad (2)$$

Bliss Independence predictions of drug combination efficacy are based on rearranged equations from Bliss et al. (Bliss, 1939) while assuming that the coefficient of association between drugs in a combination is equal to 0. This equates to equation 3 below, where $\mu_{combo,Bliss}$ is the mean Bliss Independence predicted efficacy of a combination of drugs A to Z in n cell lines and where PA_k and PZ_k are the respective probabilities of an individual cell surviving treatment with drugs A and Z in cell line k. These probabilities can be taken as the viabilities of cell line k when treated with drugs A or Z relative to an untreated control.

$$\mu_{combo,Bliss} = \frac{\sum_{k=1}^{n} PA_k \times \ldots \times PZ_k}{n} \quad (3)$$

Note that Bliss Independence is only defined for probabilities between 0 and 1, so any viabilities which fell below 0 or above 1 were rounded up to 0 or down to 1 respectively for Bliss Independence calculations in our analysis.

For both IDA and Bliss Independence based predictions, prediction uncertainties were estimated using Monte Carlo simulations with 10000 iterations each.

Efficacy Metrics: Percent Growth vs. Percent Viability

Percent relative growth (shortened to "percent growth" throughout this paper) is the primary efficacy metric available in NCI-ALMANAC. It can be interpreted such that -100% growth indicates complete cell death at the study's endpoint, 0% growth indicates no increase in viability at the study's endpoint relative to the viability measured at the start of the study (i.e.

time 0, the time a treatment was added to the cells), and 100% growth indicates identical viability to an untreated control at the study's endpoint. Notably, calculating this metric requires a cell viability measurement to be taken at time 0, when a treatment is first added to the cell line. A full description of how percent growth is calculated is available on the NCI-60 screening methodology webpage.

Percent relative viability (shortened to "percent viability" throughout this paper), on the other hand, is the primary metric used for CTRPv2 and GDSC. Unlike percent growth, percent viability is not calculated using a time zero (time of drug addition) timepoint. Instead, it is simply the ratio of the viability of a treated cell line at a study's endpoint divided by the viability of an untreated control at the study's endpoint. As such, it can be interpreted such that 0% viability indicates complete cell death at a study's endpoint and 100% viability indicates identical viability to an untreated control at the study's endpoint. Notably, this means that percent viability is not able to differentiate between treatments that are cytotoxic and treatments that are cytostatic.

Processing CTRPv2 and GDSC cell line drug screening data

CTRPv2 and GDSC often use slightly different names for the same drugs and cell lines, so these identifiers were matched between the two datasets using the harmonized identifiers provided by Ling et al. (2018). The code used to do this is included in the "Harmonizing GDSC and CTRPv2" folder of the "IDACombo Paper" project on OSF.

Following identifier harmonization, four-parameter log-logistic dose-response curves were fit to the raw drug response data using the drc R package v3.0.1 (Ritz et al., 2015) and the code included in the "Reprocessing raw CTRPv2 and GDSC data" folder of the "IDACombo Paper" OSF project. This was done because the available sources of processed dose-response

data for CTRPv2 and GDSC were generated using different algorithms between the two datasets. Recalculating the curves from the raw data allowed us to harmonize the analysis method for both datasets, and it allowed us to utilize information from all raw data points when estimating uncertainties in downstream analyses.

NCI-ALMANAC Analysis

The November 2017 release of NCI-ALMANAC was downloaded from this link on 5/17/2019. IDACombo was used to predict drug combination efficacies for the combinations included in NCI-ALMANAC using the monotherapy data in NCI-ALMANAC. To avoid evaluating the accuracy of IDACombo for the same drug combination more than once, predictions were only made for each drug combination using the maximum tested monotherapy concentrations for each drug in the combination. Since Holbeck et al. (2017) reported protocol differences between the different screening sites used to create NCI-ALMANAC, data for each monotherapy and drug combination was restricted to whichever site performed the most experiments for that monotherapy/combination. Furthermore, if multiple experiments were performed for the same treatment/cell line pair, the results of those experiments were averaged. The monotherapy based drug combination efficacy predictions for each cell line were then averaged across all cell lines to produce a mean predicted efficacy for each drug combination, and the measured efficacies in NCI-ALMANAC were also averaged to produce a mean measured efficacy for each drug combination. These predicted and measured mean efficacies were then compared. All data and code used for this analysis is included in the "NCI-ALMANAC Analysis" folder of the "IDACombo Paper" project on OSF.

Identifying Clinical Trials for IDACombo Clinical Validation

As outlined in Figure 9, the rvest R package v0.3.2 (Wickham, 2016) was used to search ClinicalTrials.gov with 9,165 search strings designed to identify trials that tested at least two of the drugs in CTRPv2 or GDSC. Search results were then compiled, resulting in the identification of 22290 clinical trial records. These records were filtered to identify only completed, phase III clinical trials, resulting in 1106 clinical trial records. Web scraping with rvest was then performed again on ClinicalTrials.gov to search the records of each trial for listed publications associated with the trial. This resulted in the identification of 1537 publications associated with 636 clinical trials. Web scraping with rvest was then performed on PubMed.gov to collect the abstracts for each of these publications, which were then manually inspected to determine if the trial met the following inclusion criteria: 1. Completed, phase III clinical trial; 2. ≥50 patients per trial arm; 3. All cytotoxic drugs in control and test therapies are available in at least one of either CTRPv2 or GDSC; 4. ≥50 cell lines available for predictions of tested control and test therapies; 5. Test therapy is control therapy plus one or more additional drugs; 6. Clinically relevant drug concentrations for each drug in a trial are not > 2x the tested drug concentrations in the dataset(s) necessary to predict that trial's efficacy (i.e. CTRPv2 and/or GDSC); and 7. Trial is not substantially the same as another selected trial (i.e. same treatment groups, doses, cancer type, patient population, and outcomes). After trials were selected based on publication abstracts, the full articles were downloaded and reviewed for final selection and collection of trial information. This resulted in the identification of 54 clinical trials for use in the validation analysis—48 of which reported PFS/TTP results and 50 of which reported OS results. These trials tested 62 unique drug treatments (46 unique control vs test treatment comparisons) involving 32 unique drugs. The selected trials and are listed in Table S3. The data and code used in this selection

process are included in the “Identifying Clinical Trials” folder of the “IDACombo Paper” project on OSF.

Identification of Clinically Relevant Drug Concentrations

In order to ensure that our drug combination efficacy predictions are clinically relevant, I surveyed the published literature to identify clinical plasma concentrations for all of the late-stage clinical drugs in CTRPv2 or GDSC. For drugs involved in the clinical trials identified for the clinical validation of IDACombo, I searched for plasma concentrations produced by the drug doses used in those trials. As such, multiple concentrations were identified for some drugs, each corresponding to a different dose of that drug. When multiple concentrations existed for a drug, the highest clinical concentration was used for the prospective analysis. All clinical concentrations as well as their corresponding citations are included in Table S4, with concentrations for the clinical trial analysis in the first sheet of the table and concentrations for the prospective analysis in the second sheet.

Clinical concentrations were defined using published clinical trials which measured patient plasma concentrations over time after drug administration. Since many drugs that are administered via bolus IV exhibit extremely high plasma concentrations at the time of administration with a very rapid decrease in concentration immediately after administration, I decided that Cmax values were not appropriate for use in our model. As such, I opted to define our clinical concentration as the maximum plasma concentration achieved at least 6 hours after drug administration, which I called Csustained. I chose 6 hours because I observed that the exponential decline in plasma concentration for bolus IV drugs is typically finished by 6 hours, though I also found that the idea of using 6 hour plasma concentrations to define clinical drug

activity is not unique to our study (Cantarovich et al., 1988). A graphical demonstration of how Csustained values were determined is included in Figure 11.

Estimating Clinical Trial Powers with IDACombo

Cell line viabilities were estimated at Csustained drug concentrations using the fitted four-parameter log-logistic dose-response curves calculated from the raw CTRPv2 and GDSC data. These monotherapy viabilities were used to estimate mean viabilities for the control and test treatments using IDACombo either using IDA based predictions or Bliss Independence based predictions. Since the efficacy predictions were made using viability, where 100% viability indicates a hazard of 1 (all cancer cells are alive relative to untreated) and 0% viability indicates a hazard of 0 (all cancer cells are dead), a hazard ratio could then be estimated for each control/test treatment comparison by dividing the mean test treatment viability by the mean control treatment viability. The estimated hazard ratios were then used to estimate PFS/TTP/OS power for each trial using the powerSurvEpi R package v0.0.9 (Qiu et al., 2015) and the number of PFS/TTP/OS events observed in each trial. Uncertainties for hazard ratios and trial powers were estimated using Monte Carlo simulations with 10000 iterations each. All clinical trial power predictions were performed using the data and scripts included in the "Clinical Trial Validation Analysis" folder of the "IDACombo Paper" project on OSF.

Prospective Analysis

The prospective analysis was performed using all drugs in CTRPv2 and GDSC that have reached phase III or IV clinical trials, with selected phase 2 drugs included based on our lab's interests. For each selected drug, cell line viabilities were estimated using drug concentrations from 0 to Csustained and the fitted four-parameter log-logistic dose-response curves calculated from the raw CTRPv2 and GDSC data. These monotherapy viabilities were then used to estimate

mean viabilities for the control and test treatments using IDACombo using IDA based predictions both with all available cell lines and with cancer-specific sets of cell lines. The

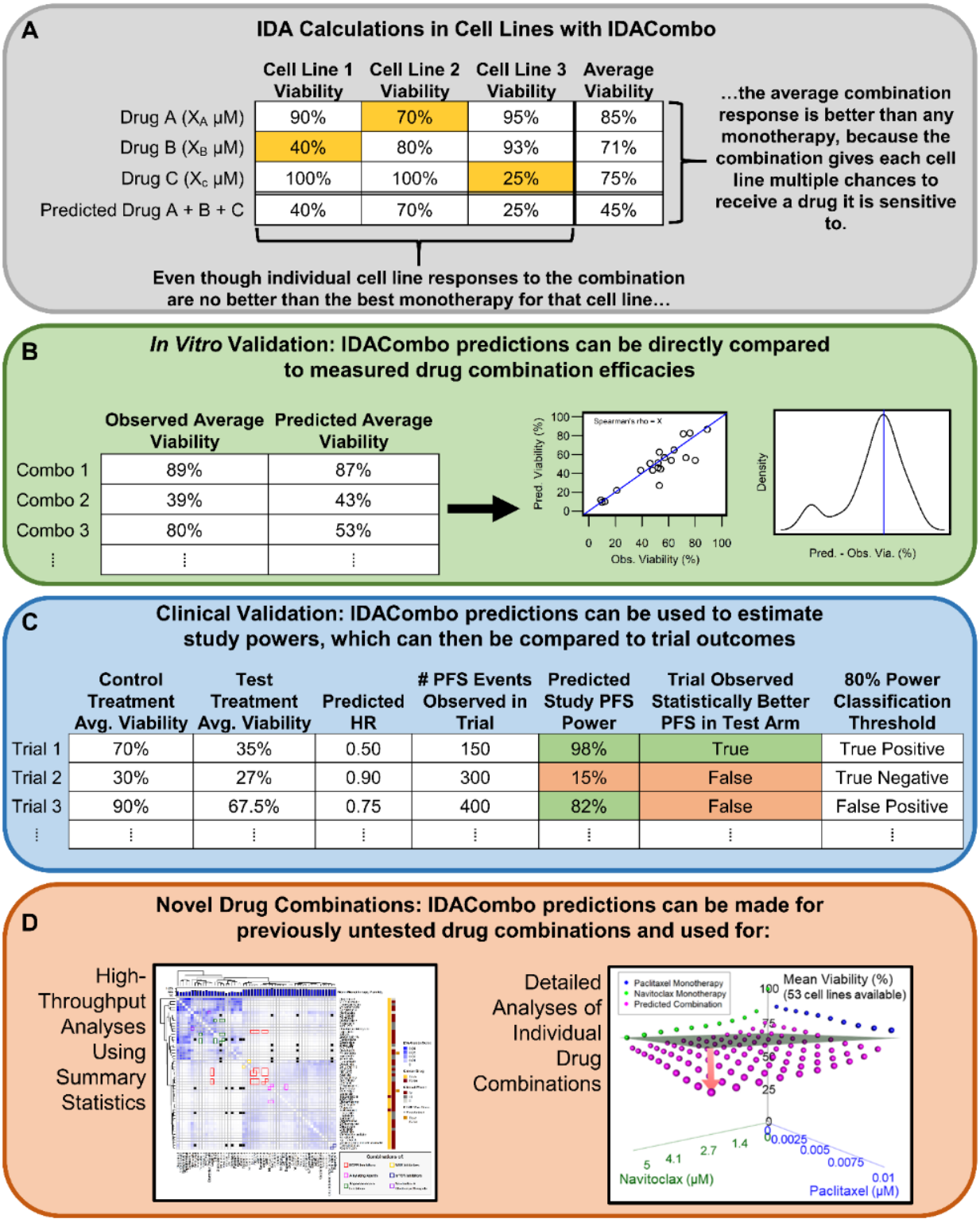

Figure 7. Legend on page 43.

Figure 7, continued. IDACombo allows drug combination efficacy predictions to be made using monotherapy cell line data, and these predictions can be validated against measured efficacies or used to identify novel efficacious drug combinations.
A) Example calculations demonstrating how IDACombo predicts drug combination efficacies based on IDA. In this example, three cell lines (cell lines 1 to 3) with measured efficacies for three monotherapies at their selected concentrations (drugs A to C) are used to predict the efficacy of the combination of drugs A + B + C. Highlighted cells indicate the best monotherapy for that cell line (i.e. provides greatest reduction in viability). **B)** Example describing strategy for validation of IDACombo efficacy predictions using *in vitro* measurements of efficacy. Measured and predicted average viabilities for each treatment can be directly compared by calculating their correlation and calculating prediction errors. **C)** Example describing strategy for validation of IDACombo efficacy predictions using published clinical trial results. Predicted combination efficacies can be used to calculate study powers, and a power threshold (80%) can be set to classify trials as likely or unlikely to detect a significant improvement in a trial outcome (i.e. PFS). These predictions can then be compared to observed trial outcomes. **D)** Intro to the analysis techniques available for using IDACombo predictions to identify novel efficacious drug combinations. High-throughput analyses using summary statistics can be used to compare efficacy predictions for many drug combinations at once, or detailed analyses can be used to explore the efficacy of a single drug combination at varying concentrations of each drug in the combination.

predicted mean drug combination viabilities were then used to calculate hazard ratios between the predicted drug combination efficacy and the best monotherapy efficacy ($HR_{C/Mbest}$) in the same way as was done for the clinical trial analysis. Since this hazard ratio would not allow for comparisons of drug combinations that did not share the same most effective monotherapy, I developed an IDAcomboscore metric which is calculated using equation 4, where Δ_{via} is equal to the mean viability when cell lines are treated with the best monotherapy (i.e. monotherapy resulting in the lowest mean viability) minus the mean viability when cell lines are treated with the drug combination.

$$\text{IDAcomboscore} = \Delta_{\text{via}} - \Delta_{\text{via}} \times \text{HR}_{\text{C/Mbest}} \quad (4)$$

The resulting metric is larger for drug combinations that are expected to be more efficacious, and it rewards drug combinations that maximally decrease the mean cell line

viability relative to monotherapy while also having a low HR relative to the most effective monotherapy in the combination.

All data and code used to perform the prospective analysis is included in the "Prospective Analysis" folder of the "IDACombo Paper" project on OSF. Notably, this folder also includes a subfolder, "./Outputs/Cluster_Heatmaps/", with efficacy prediction plots and tables for predictions made with all cell lines and with 27 cancer type/subtype specific sets of cell lines in both CTRPv2 and GDSC. For plots and tables of all cell line predictions, combinations are only included if at least 50 cell lines were available for predicting the efficacy of that combination. For cancer type/subtype specific predictions, at least 3 cell lines were required for a combination to be plotted.

RESULTS

IDACombo R Package

To ascertain the utility of an IDA based algorithm for predicting drug combination efficacy, I developed the IDACombo R package, which uses experimentally measured *in vitro* monotherapy response data to estimate drug combination efficacies. As shown in Figure 7A, the algorithm relies on the principle of IDA, simply predicting that the efficacy of a drug combination in a given cell line or patient will be equal to the effect of the single best drug in that combination. Notably, while this approach cannot predict any benefit of a drug combination versus the most effective monotherapy in an individual cell line or patient, the approach can identify drug combinations that are predicted to be more effective than monotherapy when responses are averaged across a population of cell lines or patients (see "Average" column in Figure 7A). In this way, IDACombo is suitable for improving outcomes in situations where

precision medicine is not yet able to identify which of several available drugs will be most effective for an individual patient.

To ensure that the model is useful, I validate IDACombo predictions using both *in vitro* data and clinical trial data. *In vitro* validation is performed by directly comparing predicted combination efficacies to experimentally measured combination efficacies as shown in Figure 7B, and a detailed outline for our *in* vitro validation pipeline is included in Figure 8A. Clinical validation of IDACombo predictions is performed by using those predictions to estimate clinical trial powers and comparing those predicted powers to the published results of each trial as illustrated in Figure 7C. A detailed outline of our clinical validation pipeline is included in Figure 8B.

Beyond model validation, IDACombo predictions can also be used to identify novel efficacious drug combinations either by using summary statistics to compare many drug combinations at once or by performing detailed analyses of individual drug combinations where efficacies are predicted and analyzed for numerous concentrations of each drug in a combination (Figure 7D). A prospective analysis for EGFR-WT lung cancer is presented at the end of the results section.

In Vitro Validation of IDACombo

To validate the *in vitro* utility of the algorithm, I compared predictions made with IDACombo to measured combination efficacies for approximately 5000 drug combinations available in the NCI-ALMANAC dataset (Holbeck et al., 2017) using the analysis pipeline described in Figure 8A. Briefly, monotherapy data from NCI-ALMANAC was used to predict efficacies for the drug combinations in the dataset, and the predicted combination efficacies were compared to the measured combination efficacies (see materials and methods for more details).

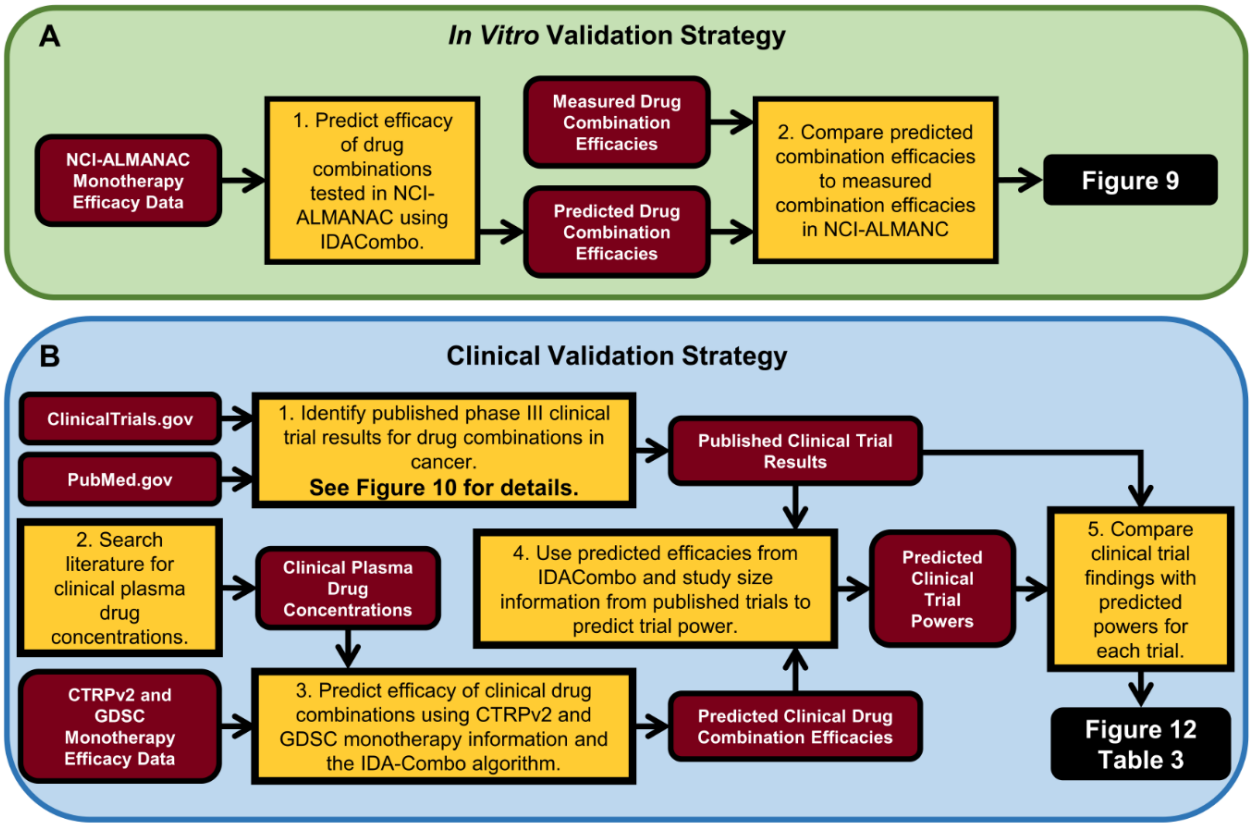

Figure 8. Pipelines to validate IDACombo predictions both in vitro and in clinical trial data. A) In vitro validation strategy. Monotherapy data from NCI-ALMANAC is used to predict drug combination efficacies, and these efficacies are compared to the measured combination efficacies that are also in NCI-ALMANAC. **B)** Clinical trials are systematically identified using ClinicalTrials.gov and PubMed.gov, and efficacy predictions are made for each treatment in each trial using clinical drug concentrations and monotherapy cell line data from CTRPv2 and/or GDSC. These predictions are used to estimate powers for each trial, and these powers are compared to clinical trial outcomes.

Notably, the efficacy metric used in NCI-ALMANAC is percent growth, which should not be confused with percent viability which is used for the clinical trial validation and prospective analyses later in this paper. These terms are defined and compared in the materials and methods.

As shown in Figure 9A, the predicted combination efficacies in NCI-ALMANAC strongly correlate with the measured efficacies (Pearson's r = 0.937, Spearman's rho = 0.929). Furthermore, the large majority of predicted efficacies were within 10% growth of the observed values with a median error of 4.99% growth (Figure 9B), and the predictions were slightly

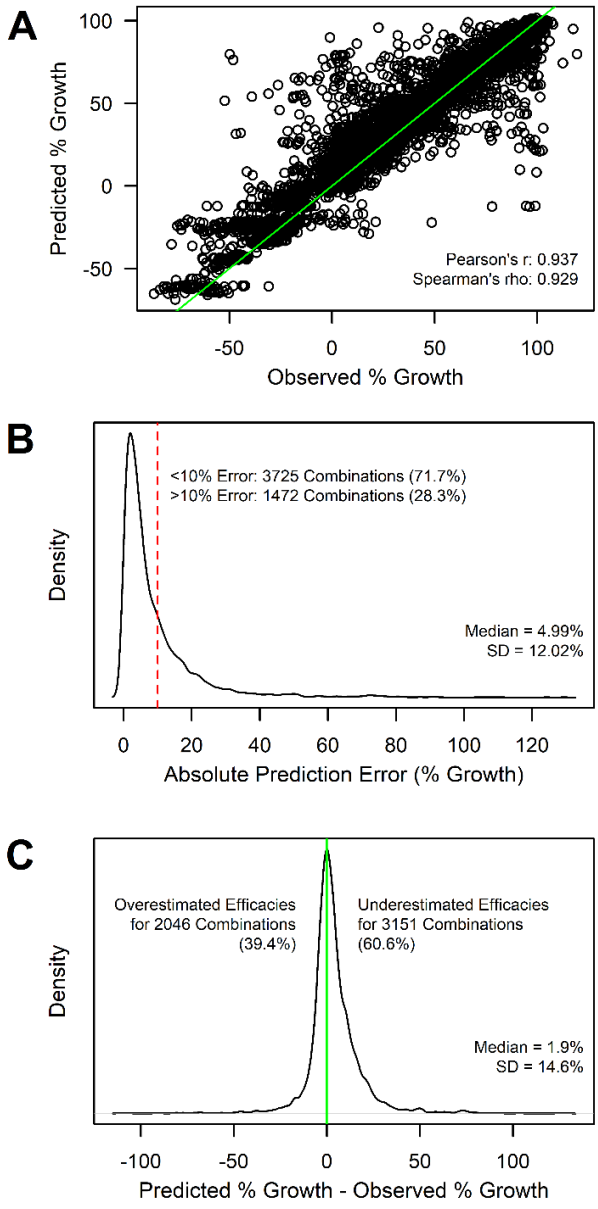

Figure 9. Agreement between predicted and observed combination efficacies in NCI-ALMANAC. A) Scatterplot showing the high correlation between the IDA-Combo predicted average percent growth in NCI-ALMANAC for each drug combination versus the experimentally observed average percent growth. The green line is a reference diagonal with slope = 1 and intercept = 0. Note that predictions were only made for the maximum concentration tested in NCI-ALMANAC for each drug. **B)** Density plot showing that the absolute values of the differences between the predicted percent growths and the observed percent growths for each drug combination are generally below 10%, with >50% of drug combinations having an absolute prediction error below 5%. The red line marks a difference of ±10% growth between predicted and observed values. **C)** Density plot showing that the differences between the predicted percent growths and the observed percent growths for each drug combination have a slight tendency towards being positive—indicating that IDA-Combo underestimates efficacy more often than it underestimates efficacies.

skewed towards being conservative estimates, with 60.6% of combination efficacies being predicted to be less effective than they actually are and 39.4% being predicted to be more effective than they actually are (Figure 9C). These results suggest that most drug combinations in NCI-ALMANAC can be accurately modeled via IDA, and they support the use of IDACombo to computationally predict drug combination efficacy using single agent screening data.

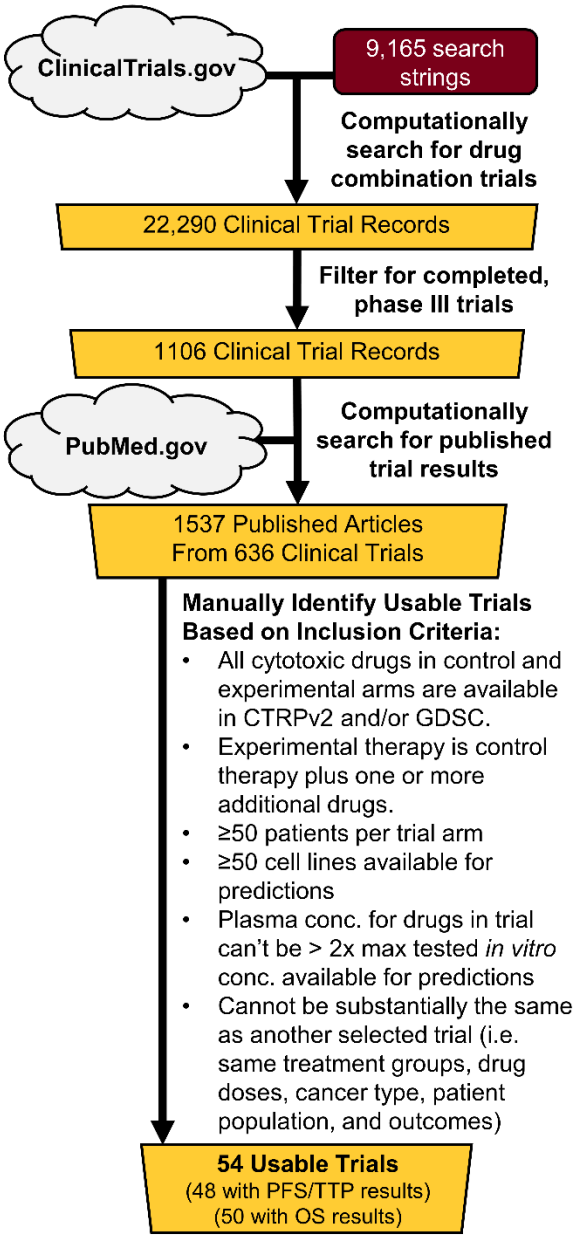

Figure 10. Selection pipeline for clinical trial validation. Flowchart detailing how completed, phase III clinical trials were selected for the clinical trial validation analysis. Searches of ClinicalTrials.gov and PubMed.gov were performed via web scraping (see materials and methods) to identify published results for trials that may meet our inclusion criteria, and the identified clinical trial publications were then manually inspected to identify trials that met the study's inclusion criteria.

Clinical Validation of IDACombo

While the *in vitro* validation results suggest that the IDACombo approach is promising, they do not establish the clinical utility of the model, as it is likely that most of the combinations tested in NCI-ALMANAC are not clinically useful. Thus, even though IDACombo accurately predicts efficacy for most of the NCI-ALMANAC combinations, it is possible that the clinically efficacious drug combinations tested in NCI-ALMANAC are all within the subset of combinations that are poorly predicted by IDACombo. To explore this possibility, I sought to validate IDACombo predictions of clinical trial efficacy against published clinical trial results. The analysis pipeline for this clinical validation is outlined in Figure 8B.

Identifying clinical trials and clinical drug concentrations for clinical validation of IDACombo

My first priority for evaluating the clinical utility of IDACombo was to identify a

diverse and unbiased set of clinical trials against which I could test IDACombo's predictions. To this end, I systematically searched clinicaltrials.gov for completed, phase III clinical trials that tested cancer drug combinations for which I could make IDACombo predictions using the Genomics of Drug Sensitivity in Cancer (GDSC) (Yang et al., 2013) and the Cancer Therapeutics Response Portal Version 2 (CTRPv2) (Basu et al., 2013) monotherapy CCL screening datasets. My pipeline for selecting clinical trials is illustrated in Figure 10 and described in detail in the materials and methods. Ultimately, this resulted in the identification of 54 clinical trials for our clinical validation of IDACombo which tested 62 unique treatments involving 32 drugs. A list of these trials is included in Table S3. To ensure the clinical relevance of IDACombo's predictions for these trials, I searched published phase I and II clinical trials to identify clinical plasma concentrations for each drug at the administered doses used in each trial. Since maximum plasma concentrations (Cmax) are extremely transient for some drugs, especially those administered via IV bolus, I decided to use the maximum plasma concentrations achieved at least 6 hours after drug administration (a metric I termed Csustained,6hr) as the concentrations for IDACombo predictions. Figure 11A illustrates how Csustained is calculated for drugs with constantly decreasing plasma concentrations over time, and Figure 11B illustrates how Csustained is calculated for drugs with increasing plasma concentrations beyond 6 hours. A more detailed description of this metric and why it was chosen is included in the materials and methods. Csustained values for each drug in the clinical trial analysis, along with the citations used to determine them, are included in Table S4.

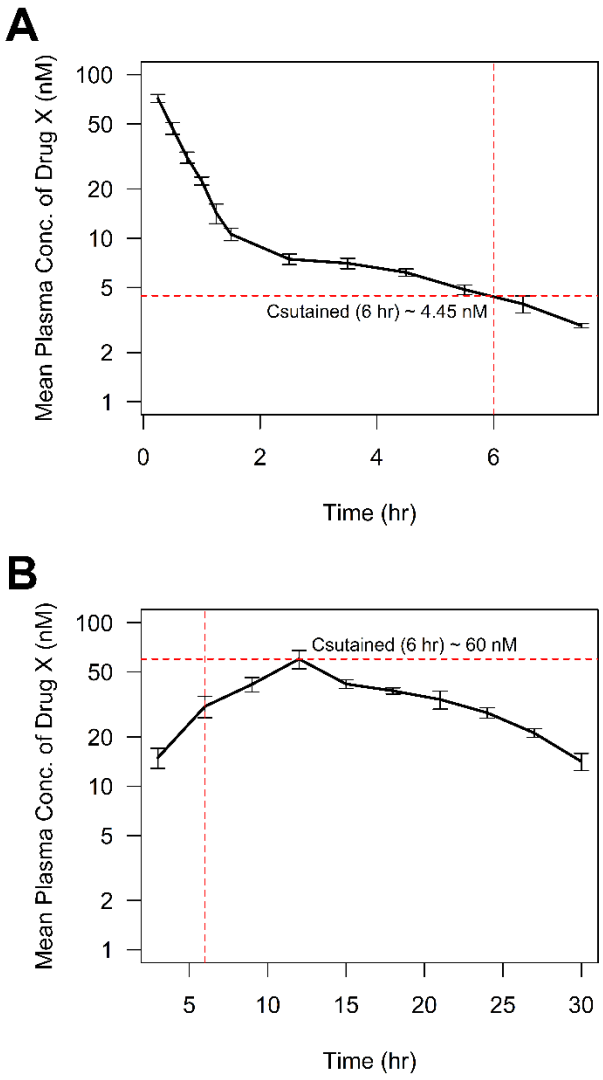

Figure 11. Calculating Csustained,6hr from clinical plasma concentration curves. This figure gives two examples to illustrate how Csustained is calculated from plasma concentration curves identified in published clinical trials. **A)** When mean plasma drug concentrations constantly decrease following administration of a drug, Csustained,6hr is simply the mean plasma concentration at 6 hours after drug administration. **B)** When mean plasma drug concentrations continue rising for more than 6 hours following administration of a drug, Csustained,6hr is the maximum plasma concentration achieved at least 6 hours after drug administration. Error bars represent mean ± standard error.

IDACombo predictions of clinical trial power closely agree with clinical trial results for trial in chemo-naïve patients, but not for trials in patients who had received previous chemotherapy

Following clinical trial and drug concentration selection, IDACombo was used to predict efficacies for the control and experimental treatments of each trial. The predicted efficacies were then used to calculate hazard ratios (HRs) between treatment groups in each trial, and the HRs

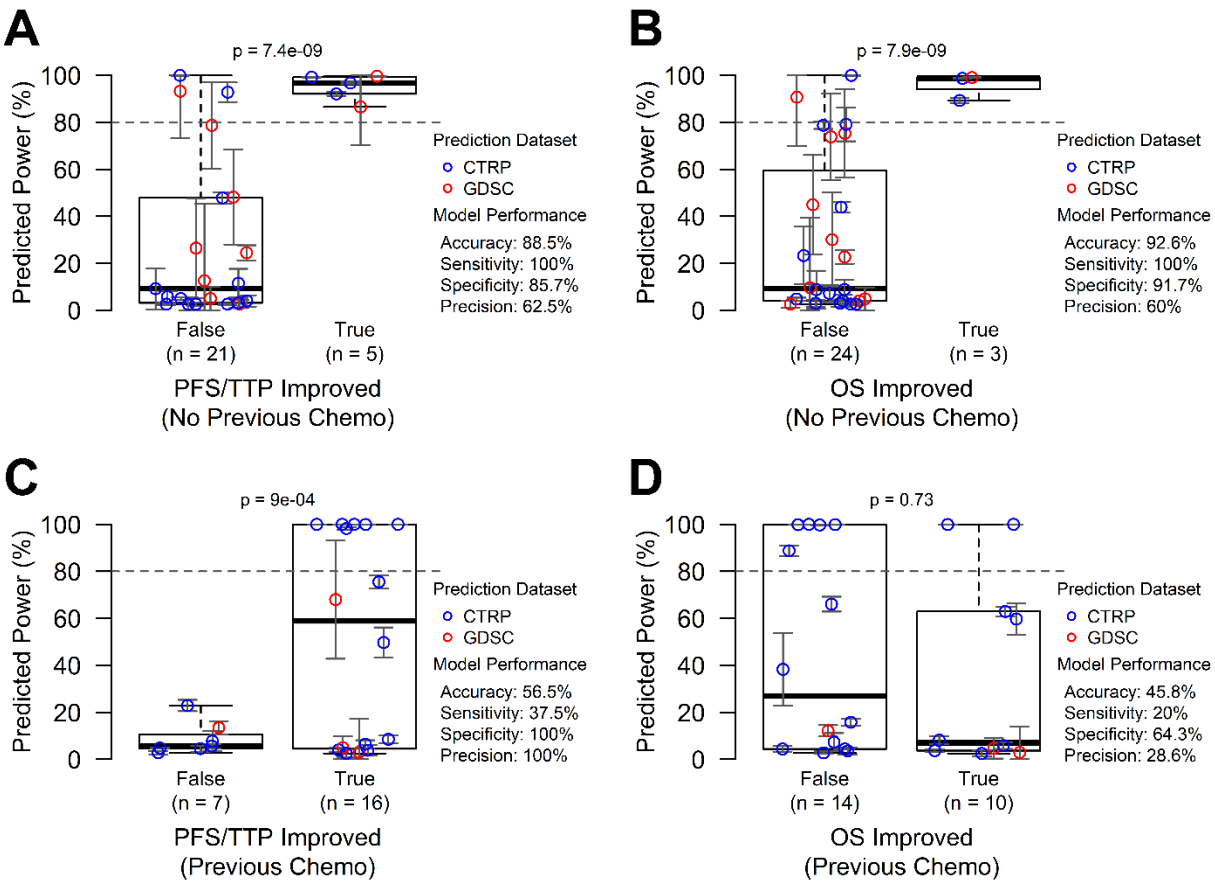

Figure 12. Clinical trial validation results show accurate efficacy predictions for trials in previously untreated patients but not for trials in previously treated patients. IDACombo was used to make efficacy predictions for the control and experimental arm treatments of the clinical trials selected using the pipeline in Figure 10. Hazard ratios were then calculated using these predictions, and study powers were calculated for each available comparison of a control therapy vs. an experimental therapy. These comparisons are separated based on whether or not the experimental arm statistically improved either PFS/TTP **(panels A & C)** or OS **(panels B & D)** in the published trial results. Predicted powers for each comparison on plotted on the y-axes, and an 80% power threshold is used to classify whether or not a comparison is expected to yield a statistically significant improvement. Comparisons are colored according to the dataset used to make predictions for the compared treatments. **Panels A & B** show results for trials in which patients had received no previous chemotherapy treatments. **Panels C & D** show results for trials in patients who had received previous chemotherapy. Error bars represent mean ± standard error.

were used to estimate the statistical power each trial had to detect significant improvements in progression-free survival (PFS), time to progression (TTP), or overall survival (OS). These powers are included in Table S3 and are plotted in Figure 12, with trials separated based on

whether or not a statistical improvement in PFS/TTP or OS was observed in the clinic. Encouragingly, using a standard 80% power cutoff to classify trials as likely or unlikely to detect a statistically significant improvement, our predicted powers for PFS/TTP correctly classified 88.5% of clinical trials in which patients had not received cancer drug treatment prior to trial entry (Figure 12A), with >85% sensitivity and specificity. For OS powers in treatment-naïve trials (Figure 12B), accuracy, sensitivity, and specificity were >90%, but it is difficult to confidently assess the suitability of IDACombo for predicting OS benefit, because I only identified 3 clinical trials in treatment-naïve patients which detected a statistically significant improvement in OS. Unfortunately, the model performed much more poorly for clinical trials in patients who had undergone cancer drug treatment prior to entering the trial (Figures 12C and 12D). While the reasons for this poor performance in previously treated patients are not immediately clear, several possible explanations are discussed in detail in the discussion section of this chapter.

Clinical IDACombo predictions for targeted therapies

It should be noted that the previous predictions were made using all of the available cell lines in CTRPv2 or GDSC, so the predictions were not based exclusively on cell lines of the same cancer type as was being tested in each trial. For targeted therapies, however, it is often the case that the therapy is only effective in a specific molecular subtype of cancer. Two of the clinical trials selected for our analysis tested targeted therapies and reported full study results for patients with and without the molecular features targeted by those therapies. To evaluate the suitability of IDACombo to predict the efficacy of targeted therapies, I made power predictions for these two trials using sets of cell lines with or without the relevant molecular features for each reported patient subgroup. The resulting predictions for these trials are shown in Table 3.

Di Leo et al., 2008: Adding Lapatinib to Paclitaxel for Breast Cancer

Cancer Subtype	Lapatinib Improved OS	Measured OS HR	Lapatinib Improved TTP	Measured TTP HR	Predicted HR	Predicted OS Power	Predicted TTP Power	# Available Cell Lines
HER2-	False	0.89	False	1.05	0.83	0.24	0.35	14
HER2+	False	0.74	True	0.35	0.61	0.33	0.55	6

Wu et al., 2013: Adding Erlotinib to Gemcitabine and Carboplatin for NSCLC

Cancer Subtype	Erlotinib Improved OS	Measured OS HR	Erlotinib Improved PFS	Measured PFS HR	Predicted HR	Predicted OS Power	Predicted PFS Power	# Available Cell Lines
EGFR-WT	False	0.77	False	0.97	0.93	0.06	0.07	124
EGFR-Mutant	True	0.48	True	0.25	0.64	0.35	0.53	7

Table 3. IDA-combo predictions can identify subtype-specific responses to combinations of targeted therapies. For two clinical trials that only identified test arm benefit for subsets of patients with specific molecular subtypes, clinical powers were calculated using IDACombo with groups of cell lines that either had or lacked those molecular subtypes. This table summarizes those predicted trial powers and the observed results of these trials.

Notably, IDACombo's predictions agreed with clinical findings that there is a higher expected benefit for patients with the molecular features targeted by the targeted therapies than for patients without those molecular features. However, the subtype-specific predictions did not reach the 80% power cutoff necessary to correctly classify the trials. This may be due to the fact that very few cell lines were available for these subtype-specific predictions, leading to relatively high prediction uncertainties and a relatively small population in which to detect phenotypic heterogeneity.

Clinical IDACombo predictions with cancer-specific sets of cell lines perform more poorly than predictions with all available cell lines

To further assess the utility of making predictions with sets of cell lines matched to patient phenotypes, I predicted clinical trial powers using cancer-specific sets of cell lines for each clinical trial (Figure 13). Note that clinical trials were excluded if fewer than 5 cancer-specific cell lines were available with which to make predictions. The cancer-specific predictions

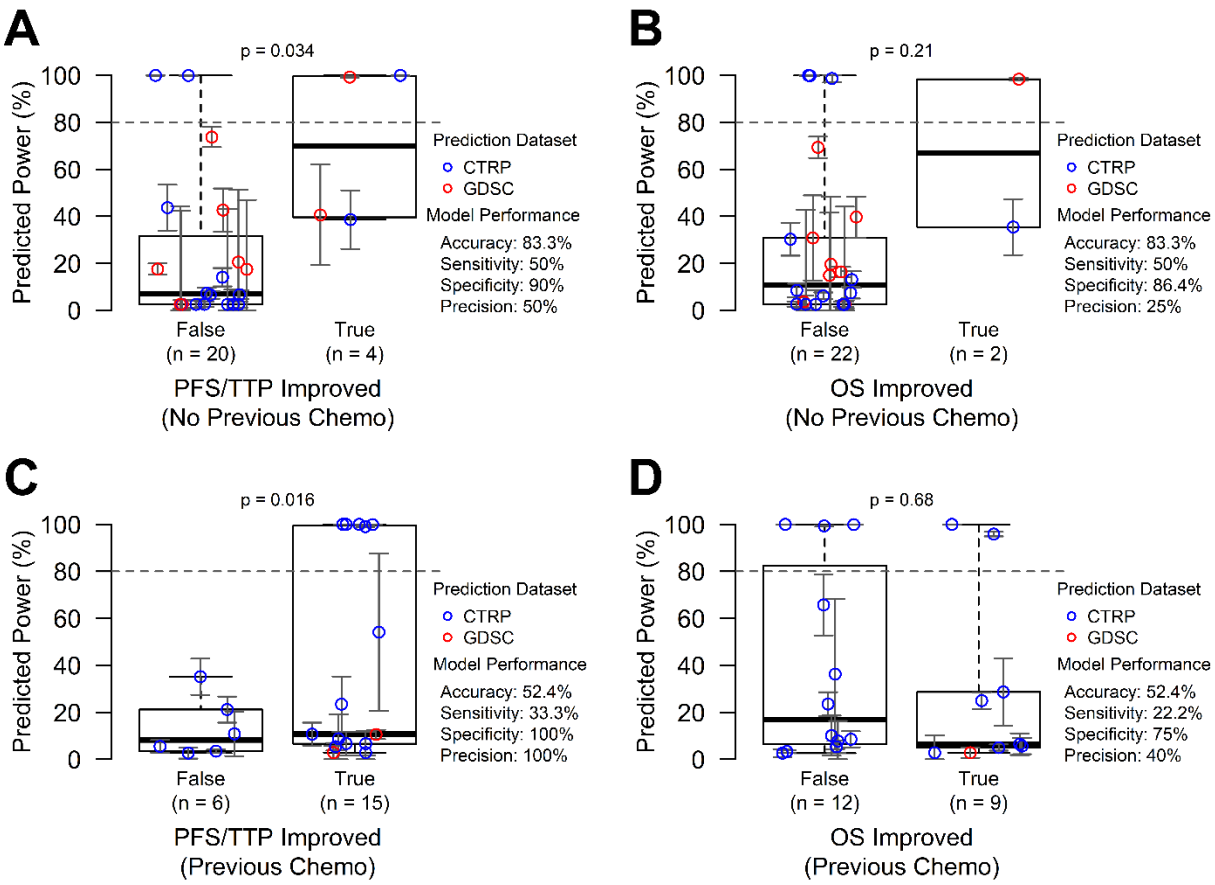

Figure 13. Using only cancer-specific cell lines does not improve model performance for clinical trial power predictions. Identical to Figure 12, except that predictions were made for each trial using sets of cell lines specific to the cancer type being studied in each trial. **A)** Predicted power of each trial in previously untreated patients to detect a significant improvement in PFS/TTP at an alpha of 0.05 versus whether or not the study actually detected a significant improvement in PFS/TTP. **B)** Predicted power of each study in previously untreated patients to detect a significant improvement in OS at an alpha of 0.05 versus whether or not the study actually detected a significant improvement in OS. **C)** Predicted power of each trial in previously treated patients to detect a significant improvement in PFS/TTP at an alpha of 0.05 versus whether or not the study actually detected a significant improvement in PFS/TTP. **D)** Predicted power of each study in previously treated patients to detect a significant improvement in OS at an alpha of 0.05 versus whether or not the study actually detected a significant improvement in OS. Error bars represent mean ± standard error.

resulted in accuracies > 80% for trials in chemo-naïve patients, but model performance was generally reduced and prediction uncertainties increased for cancer-specific predictions versus pan-cancer predictions. This result, along with the analysis of the two targeted therapy trials, suggests that predictions made using cancer and subtype-specific sets of cell lines could be

preferable to pan-cancer predictions if sufficient numbers of cell lines were available for each cancer type, but there are currently too few cell lines available for each cancer type in these datasets for this approach to be viable. In the meantime, pan-cancer predictions appear to be adequate for most of the drug combinations used in our clinical trial dataset.

Clinical IDACombo prediction accuracy drops when predicting efficacy for trials with drugs which have plasma concentrations beyond the tested in vitro concentrations

Beyond the selection of cell lines, I also wanted to investigate the importance of drug concentration selection for IDACombo predictions. Several of the trials identified from ClinicalTrials.gov tested drugs with Csustained concentrations above the tested concentrations for those drugs in CTRPv2 or GDSC, with several trials including drugs with Csustained concentrations > 2x the tested *in vitro* concentrations in GDSC (Figures 14A and 14C). To determine whether or not this would affect IDACombo based power predictions for these trials, I calculated model performance for both PFS/TTP and OS for these trials (specifically trials in chemo-naïve patients) and compared model performance to whether or not trials included drugs with Csustained concentrations higher than tested *in vitro* concentrations. Trials with at least one drug with a Csustained concentration > 2x the maximum tested *in vitro* concentration for that drug showed largely reduced accuracy, specificity, and precision in both PFS/TTP and OS predictions relative to trials with drugs that have Csustained concentrations ≤2x the maximum tested *in vitro* concentrations (Figures 14B and 14D). As a result of this finding, only trials with drugs that have Csustained concentrations ≤2x the maximum tested *in vitro* concentrations were included in the other clinical analyses in this paper.

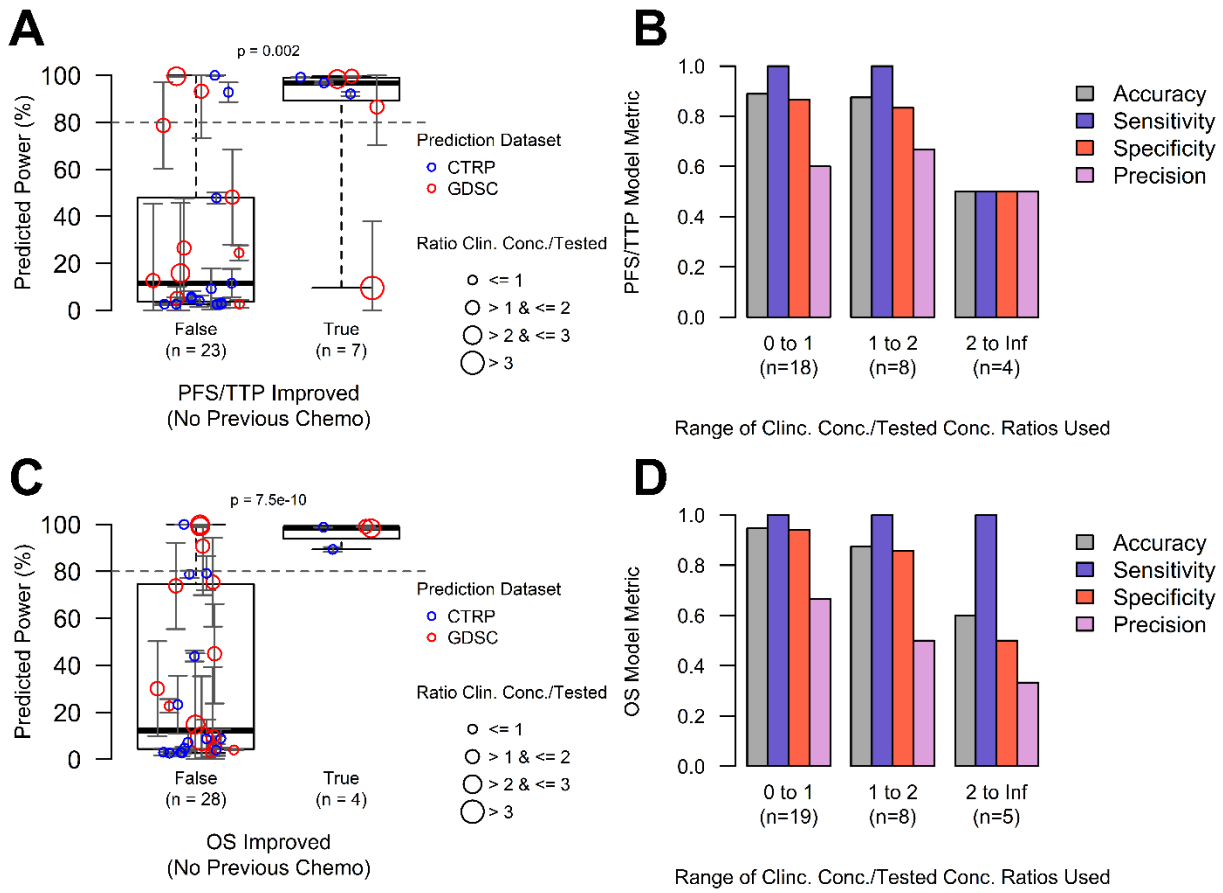

Figure 14. IDACombo predictions become less accurate when made using drug concentrations beyond the tested monotherapy concentration range. A) Similar to Figure 12A, this plot shows predicted clinical trial powers for PFS/TTP in trials with chemo-naïve patients, separated based on whether or not the trial actually observed a statistical improvement in PFS/TTP with the test treatment. Trial points are sized according to the maximum ratio of the Csustained concentrations used for the drugs in the trial to the maximum tested concentrations of those drugs in CTRPv2 or GDSC. Ratios above 1 indicate that the Csustained concentration is higher than the maximum available concentration in CTRPv2 or GDSC. Notably, most of the incorrectly classified trails have ratios > 1 and most of the correctly classified trials have ratios < 1. **B)** Barplot showing PFS/TTP model performance for trials with chemo-naïve patients that fall within three different ranges of ratios of drug Csustained concentration to tested concentration in CTRPv2 or GDSC. Notably, trials with a Csustained/tested concentration ratio > 2 are predicted much more poorly than trials with a ratio between 0 and 1 or with a ratio between 1 and 2. **C)** Same as **A**, except for OS in trials with chemo-naïve patients. **D)** Same as **B**, except for OS in trials with chemo-naïve patients. Error bars represent mean ± standard error.

Clinical IDACombo predictions are affected by selected drug concentrations, but remain accurate, sensitive, and specific across a range of concentrations

I further examined the importance of drug concentration selection by assessing whether or not prediction performance was harmed by using drug concentrations that deviated from clinical plasma concentrations. When predictions were made using the maximum concentrations tested for each drug in either CTRPv2 or GDSC rather than Csustained concentrations, prediction accuracies in treatment-naïve trials fell dramatically (69.2% accuracy for PFS/TTP and 70.4% accuracy for OS) (Figures 15A and 15B). Alternatively, when the Csustained concentrations for each drug in a trial were multiplied by factors between 0.1 and 10, I found that uniformly increasing drug concentrations kept the method's sensitivity high but decreased accuracy, specificity, and precision for both PFS/TTP and OS. Uniformly decreasing concentrations quickly reduced sensitivity and precision (Figures 15C and 15D). These results suggest that correctly identifying clinical drug concentrations is important for *in vitro* predictions using IDA, with underestimated concentrations decreasing model performance more than overestimated concentrations when clinical dose ratios between drugs are preserved.

Clinical predictions with Bliss Independence are less accurate than predictions with IDA

Another possibility I wished to explore was whether or not an alternative model of drug combination efficacy could perform better than IDA for predicting clinical trial outcomes in our dataset. One of the most well established models for calculating the expected efficacy of a combination of non-interacting drugs is Bliss Independence (Bliss, 1939), which is described in detail and compared to IDA in the materials and methods. To determine the utility of Bliss Independence for predicting clinical drug combination efficacy, I predicted clinical trial powers using Bliss Independence instead of IDA. As can be seen in Figure 16, Bliss Independence based predictions were generally less accurate, specific, and precise than the IDA based predictions in

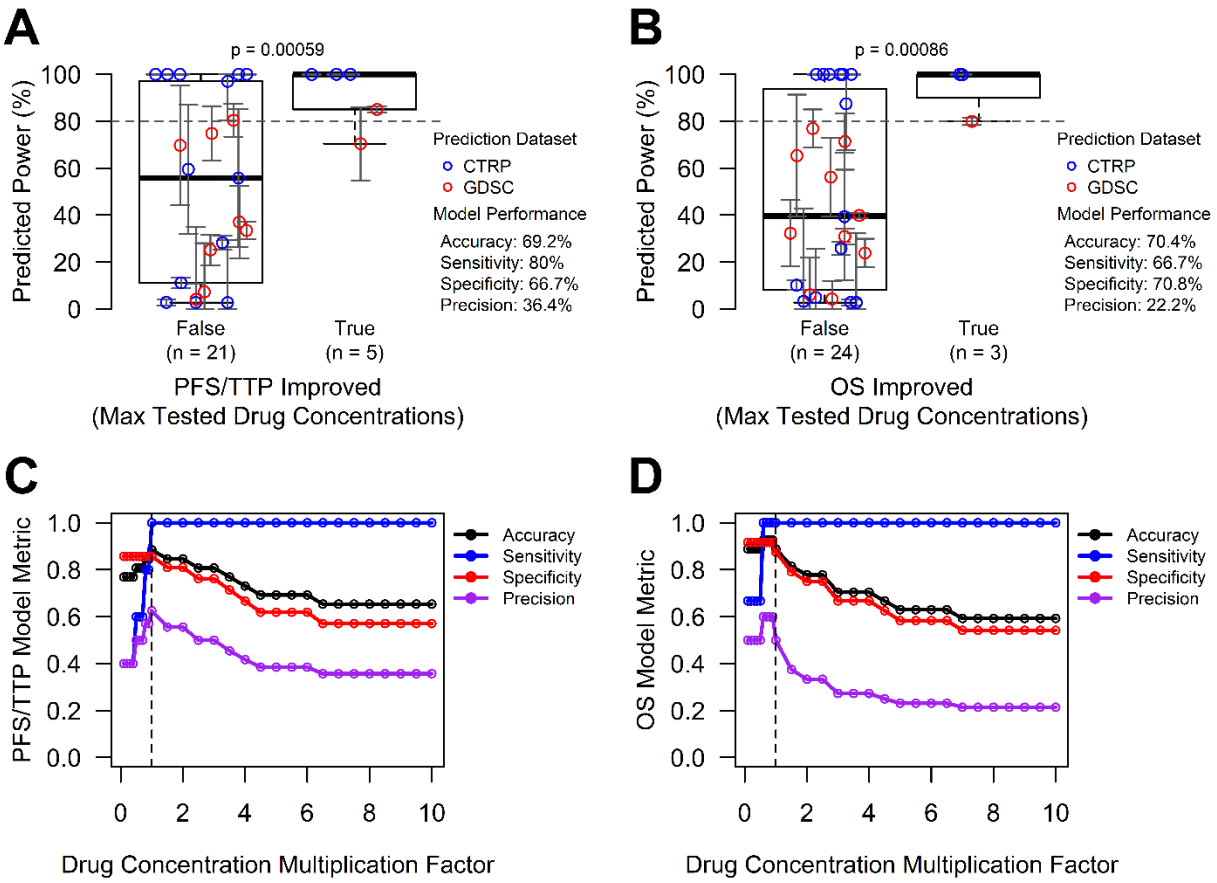

Figure 15. Clinical power predictions are dose-dependent. A&B) Similar to Figure 12A and 12B, all available cell lines were used to create predictions of study power for trials in chemo-naïve patients and compared to whether or not the trials saw a statistically significant improvement in PFS/TTP **(A)** or OS **(B)**. In this figure, however, maximum tested concentrations were used for each drug instead of Csustained concentrations. Notably, these predictions with the maximum tested concentration of each drug results in much poorer model performance than the Csustained predictions in Figure 12. **C&D)** In an effort to determine how sensitive our method is to dose perturbation, power predictions were made for each trial in chemo-naïve patients using Csustained drug concentrations which have been multiplied by a multiplication factor between 0.1 and 10. Model performance metrics for PFS/TTP **(C)** or OS **(D)** were then calculated using predictions from each dose multiplication factor, and those metrics are plotted here. Error bars represent mean ± standard error.

Figure 12, suggesting that IDA is a more useful model for clinical drug combinations, at least for the trials in our dataset.

Overall, the clinical trial validation results suggests that IDACombo is capable of making highly accurate clinical predictions for drug combinations in previously untreated patients using

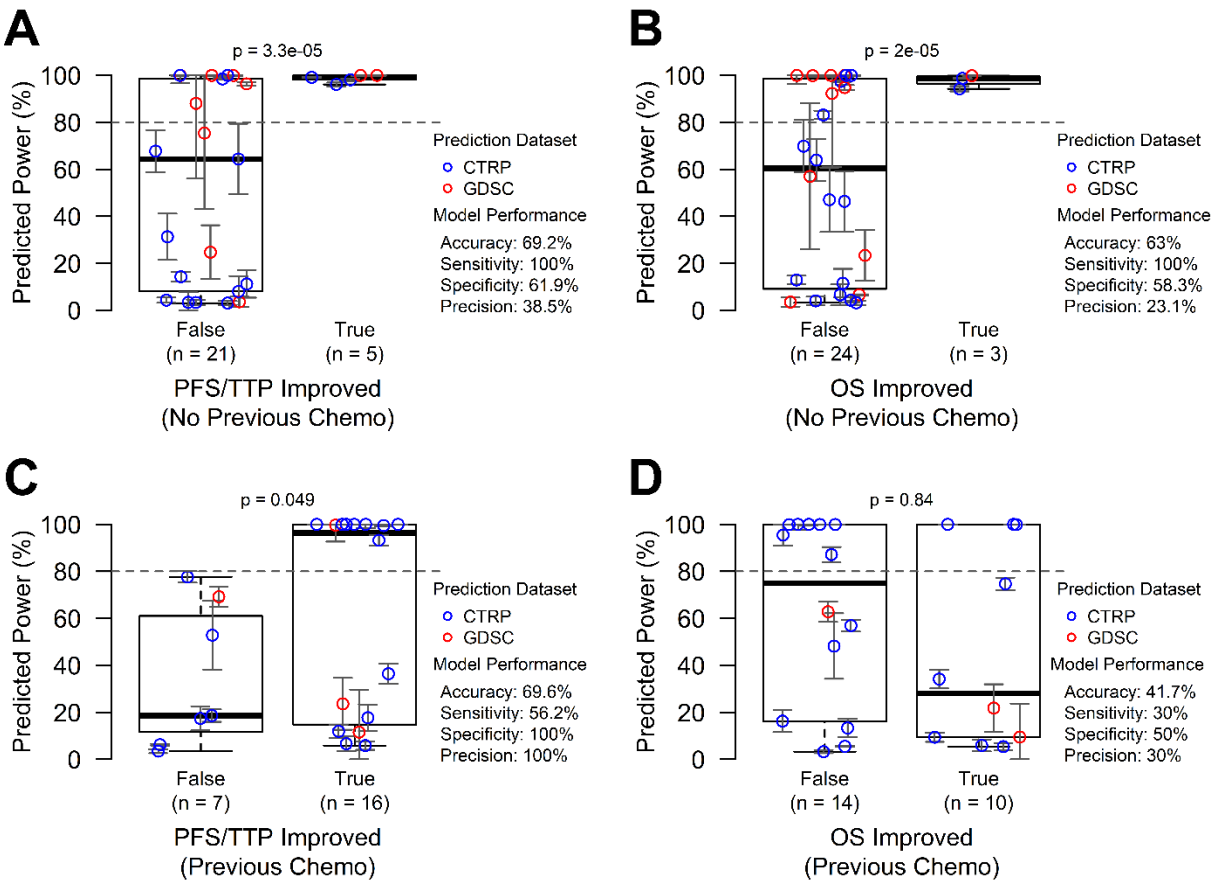

Figure 16. Predictions made using Bliss independence are less accurate than those made with independent drug action. Power predictions were made for the clinical trials shown in Figure 12, but using the Bliss independence model rather than the IDA model. In general, when compared to the IDA predictions in Figure 12, Bliss Independence inflates estimated powers, leading to decreased precision, specificity, and accuracy while providing marginal improvements in sensitivity. Error bars represent mean ± standard error.

only *in vitro* monotherapy information. The success of this tool bridges the gap not only between monotherapies and combinations, but also between cell lines and patients.

Prospective Efficacy Predictions

Pan-cancer IDACombo predictions reveal patterns based on drug mechanisms of action

Given the encouraging validation results both *in vitro* and in clinical trial data, I chose to create efficacy predictions for all possible 2-drug combinations of clinically advanced drugs available in CTRPv2 or GDSC. The analysis of these results necessitated a different analysis

strategy than was used for the clinical validation analysis, however, because power calculations were not convenient given the lack of knowledge about how many PFS/TTP or OS events would be observed in future trials of these combinations and the lack of knowledge about which control treatment each combination should be best compared to. As a result, I developed the IDAcomboscore, which is described in detail in the materials and methods. Generally, the IDAcomboscore can be interpreted such that a higher comboscore indicates a more effective drug combination relative to the most effective single drug in the combination.

IDAcomboscores were calculated for all possible 2-drug combinations of clinically advanced drugs in CTRPv2 and GDSC using either all available cell lines (pan-cancer) or cancer-specific sets of cell lines. The "Prospective Analysis" folder of the "IDACombo Paper" project on OSF (see Data and Software Availability in materials and methods) includes full plots of these scores, as well as plots comparing IDAcomboscores calculated using different sets of cell lines. Pan-cancer IDAcomboscores are plotted in Figure 17 for combinations of CTRPv2 drugs which had at least one IDAcomboscore >= 0.004 (this cutoff was determined based on heatmap cluster boundaries between drugs with higher and lower comboscores). Notably, the heatmap suggests that combinations of drugs which work via the same mechanisms of action are not expected to be more efficacious than the best monotherapy under IDA (see combinations of topoisomerase inhibitors, EGFR inhibitors, MEK inhibitors, mTOR inhibitors, or alkylating agents). This is not surprising, since IDA predicts that the best drug combinations will be comprised of drugs which target completely separate populations of cells/patients, and drugs that have the same mechanism of action are likely to target the same populations of cells/patients. An exception to this, however, can be found in the combination of navitoclax and obatoclax, which is predicted to be efficacious despite their both being classified as BCL inhibitors. The most

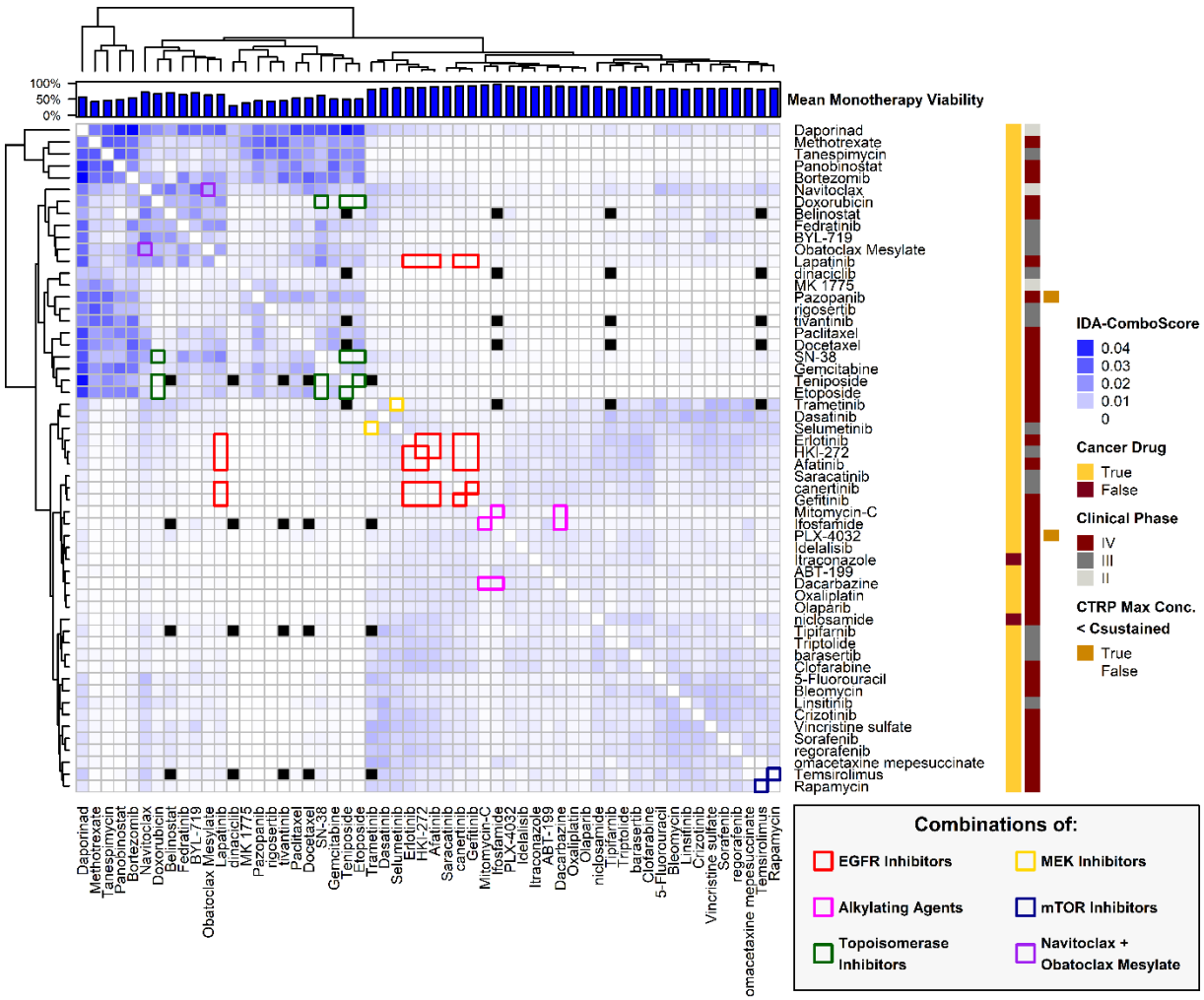

Figure 17. IDAcomboscore predictions for late-stage clinical drugs in CTRPv2. IDAcomboscores were calculated for all two-drug combinations of late-stage clinical drugs in CTRPv2 using all available cell lines for each drug combination. These comboscores are plotted here, with darker blue squares representing higher comboscores and, therefore, greater predicted drug combination efficacies relative to the constituent monotherapies. Black boxes represent missing values, where efficacies could not be predicted for a combination. The first bar, farthest left on the right side of the heatmap, indicates whether or not that drug is currently used for cancer treatment, the second bar indicates what stage of clinical trials that drug has reached, and the third bar indicates if the known Csustained concentration for the drug was higher than the maximum tested concentration in CTRPv2 such that predictions had to be made with a lower than clinical concentration. The barplot on top of the heatmap indicates the average viability achieved using that drug as a monotherapy (full bar indicates 100% viability). Late-stage clinical drugs that were not predicted to combine well with any other drugs (i.e. with a comboscore < 0.004) were excluded from this plot to save space, but a full heatmap with all late-stage clinical drugs can be found in the "IDACombo Paper" project on OSF (see materials and methods). Combinations of drugs with the same mechanisms of action are highlighted for several drug mechanisms.

likely explanation for this is that obatoclax has been found to have effects other than BCL inhibition. Indeed, it has been reported that obatoclax can be highly effective in cell lines that are relatively resistant to navitoclax (Stamelos et al., 2016).

An examination of combinations between drugs with different mechanisms of action (so called "class effect") is more difficult than assessing combinations of drugs with the same mechanisms of action, because most mechanisms of action are only represented by a single drug in Figure 17. This makes it impossible to be confident that the predicted combination efficacy of two drugs will be informative about the combination efficacy for two different drugs of the same mechanisms. One conclusion that can be made from the figure, however, is that there is clear structure in the data, with drugs clustering together based on similar combination partner preferences and efficacy intensities. Some of this structure can be explained by drug mechanisms of action, as drugs with the same mechanism of action often end up in the same hierarchical clusters (at least, this is the case for the few mechanisms of action for which I have more than one drug). This does not fully explain the clustering, however, as we can see with topoisomerase inhibitors and EGFR inhibitors, which are divided between several small clusters. A more detailed analysis of the drugs' mechanisms of action may partially explain this, as, for the topoisomerase inhibitors, drugs are separated by whether or not they inhibit topoisomerase I or II and whether or not they act by binding DNA or intercalating DNA. This is highly speculative, however, given the small number of drugs available for each mechanism of action. It is also notable that drugs which have similar average viabilities across all cell lines when used as a monotherapy tend to be more closely clustered. This suggests that the clustering observed in Figure 17 may be explained partially by similarity in drug mechanisms and partially by similarity in the average monotherapy efficacies of drugs at their clinical concentrations. Unfortunately, a

more detailed analysis of which mechanisms and monotherapy efficacies provide the most effective combinations is prevented by the limited number of drugs available for each drug mechanism.

The accuracy of cancer-specific IDACombo predictions is currently limited by the number of available cell lines for each cancer type

Another difficulty in this analysis is that clinical trials are typically designed to treat a single cancer type, and an analysis of pan-cancer predictions does not provide information about which cancer types each combination would be most effective in. The most direct solution to this problem is to analyze predictions made with cancer-specific sets of cell lines, but, as mentioned in the clinical trial validation analysis, many cancer types have very few available cell lines. To determine how much of a limitation this is, I sought to determine how many cell lines are necessary to create accurate predictions using IDACombo. Since the true efficacy of most drug combinations is not known, I decided to use agreement between predictions made using CTRPv2 and GDSC as a metric of prediction accuracy. Notably, I only compared CTRPv2 and GDSC predictions for combinations in which Csustained was available for both drugs in both datasets and which had at least 400 cell lines available to make predictions with—this resulted in comparisons for 351 drug combinations involving 27 compounds.

For the comparison, I calculated Spearman's ρ between CTRPv2 and GDSC predictions made with varying number of cell lines and plotted them in Figure 18A. This revealed that a ρ as high as 0.8 could be achieved using 250 or more cell lines, and that this correlation slowly decreased to ~0.7 as the number of cell lines was reduced to 50. With less than 50 cell lines, ρ decreased more rapidly, to ~0.6 with 25 cell lines and ~0.3 with 5 cell lines. This suggests that most cancer-specific predictions will be suboptimal, owing to their having less than 50 cell lines

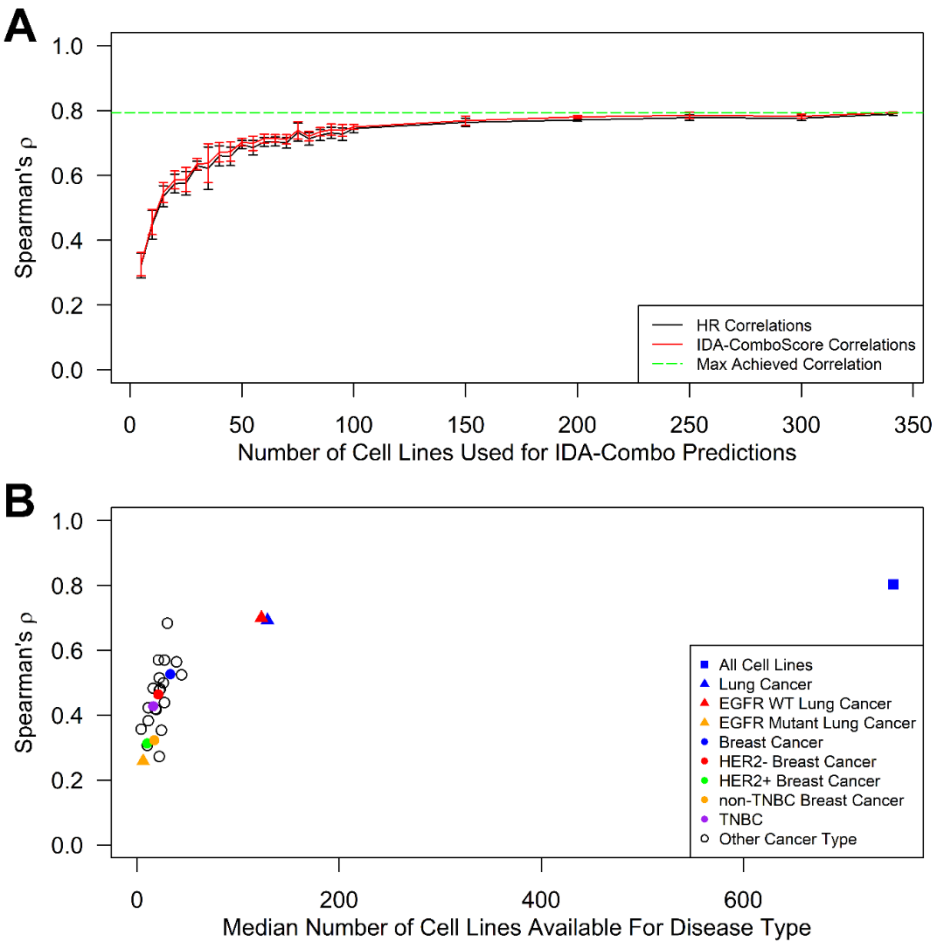

Figure 18. IDAcomboscore agreement between CTRPv2 and GDSC is affected by the number of cell lines available to make predictions with. In an effort to determine how many cell lines are required to estimate drug combination efficacy, IDAcomboscore correlations between CTRPv2 and GDSC are plotted versus the number of cell lines used to make those predictions. **A)** IDA-comboscore predictions were made using randomly sampled sets of cell lines of varying sizes. Sampling was performed three times for each number of cell lines, and error bars represent the standard deviation of the triplicate correlation measurements. Notably, agreement between CTRPv2 and GDSC decreases rapidly as the number of cell lines is reduced below 50. **B)** Correlations are plotted for predictions made using cancer-specific cell lines (see Figure 13, the "IDACombo Paper" project on OSF, and materials and methods). Note that the x-axis denotes the median number of cell line available for that cancer type for each drug combination, as the number of cell lines available for each cancer type varies from drug to drug. Subsets of breast cancer and lung cancer are highlighted in the plot. Note that for both panels **A** and **B**, correlations were only calculated for drug combinations that used drugs for which their clinical doses were available in both CTRP and GDSC so as to avoid calculating correlations between predictions made with different drug concentrations between the two datasets.

available to make predictions with, but it also suggests that there is some level of reproducibility using those numbers of cell lines. To quantify this reproducibility specifically for the cancer

types available in CTRPv2 and GDSC, I plotted Spearman ρ's between cancer-specific IDAcomboscores versus the median number of cell lines available for each of 25 cancer types/subtypes (Figure 18B). The results largely agreed with the downsampling approach in Figure 18A, showing that Spearman ρ's for cancer-specific predictions ranged from ~0.7 to ~0.3 depending roughly on how many cell lines were available for each cancer type. A full list of correlation coefficients for each cancer type can be found in Table S5. These findings suggest that highly reproducible cancer-specific predictions are currently possible for some cancer types, but IDACombo predictions for most cancer types would likely be significantly improved by increasing the number of cell lines available for those cancer types.

IDACombo predicts that navitoclax will efficaciously combine with taxanes in EGFR wild type lung cancer

The cancer subtype with the most available cell lines is currently *EGFR* wild type lung cancer. To demonstrate how cancer-specific predictions can be used to identify novel efficacious drug combinations, I performed an example analysis aimed at identifying efficacious 2-drug combinations with navitoclax, a BCL inhibitor currently in phase I/II clinical trials for lung cancer in various combinations, in *EGFR* wild type lung cancer. I did this by identifying the highest predicted IDAcomboscores for navitoclax combinations in *EGFR* wild type lung cancer (Figure 19A). The combination with the highest predicted efficacy was with daporinad (APO866), which is an NAMPT inhibitor that has yet to enter phase III trials after three phase II trials failed to show significant efficacy for daporinad monotherapy in refractory B-CLL, advanced melanoma, or cutaneous T-cell lymphoma (BioAlliance Pharma, 2013; Goldinger et al., 2015). Strikingly, the second and fourth best combinations were both with taxanes (paclitaxel and docetaxel). Given the stalled clinical development of daporinad and the shared mechanism of

action of paclitaxel and docetaxel, I decided to further investigate the combination of navitoclax with taxanes. To do this, I then calculated mean viabilities for the combinations of navitoclax + taxane using a range of concentrations from 0 μM up to the achievable clinically sustained plasma concentration for each drug. This analysis revealed that the navitoclax + taxane combination is predicted to be superior to the best achievable monotherapy efficacy across a wide range of drug concentrations for both combinations with docetaxel and paclitaxel (Figures 18B and 18C). In fact, the analysis predicts that the maximal monotherapy efficacy can be achieved using combinations of the drugs at much lower doses (approximately one third) than are required to achieve the same effect with monotherapy. This is important, because it suggests that combinations that are predicted to be efficacious by IDACombo may still be superior to monotherapy even if the clinical use of the combination requires lower doses to be used for each drug to limit toxicities.

Furthermore, a search of the published literature revealed that other groups have tested the combination of navitoclax with taxanes *in vitro, in vivo,* and in a phase I clinical trial. Tan et al. (2011) found that combining a taxane with navitoclax resulted in greater than additive benefit in a panel of non-small cell lung cancer (NSCLC) cell lines and that the combination of navitoclax + docetaxel showed statistically superior efficacy to monotherapy in three xenograft mouse models of NSCLC using two treatment schemas. Similarly, Chen et al. (2011) found that docetaxel showed greater than additive benefit when combined with navitoclax in a panel of human solid tumor cell lines, and experiments in a xenograft mouse model of ovarian cancer showed that the combination of navitoclax + docetaxel showed superior efficacy to monotherapy across a range of drug doses and treatment schedules. Furthermore, a phase I trial of navitoclax

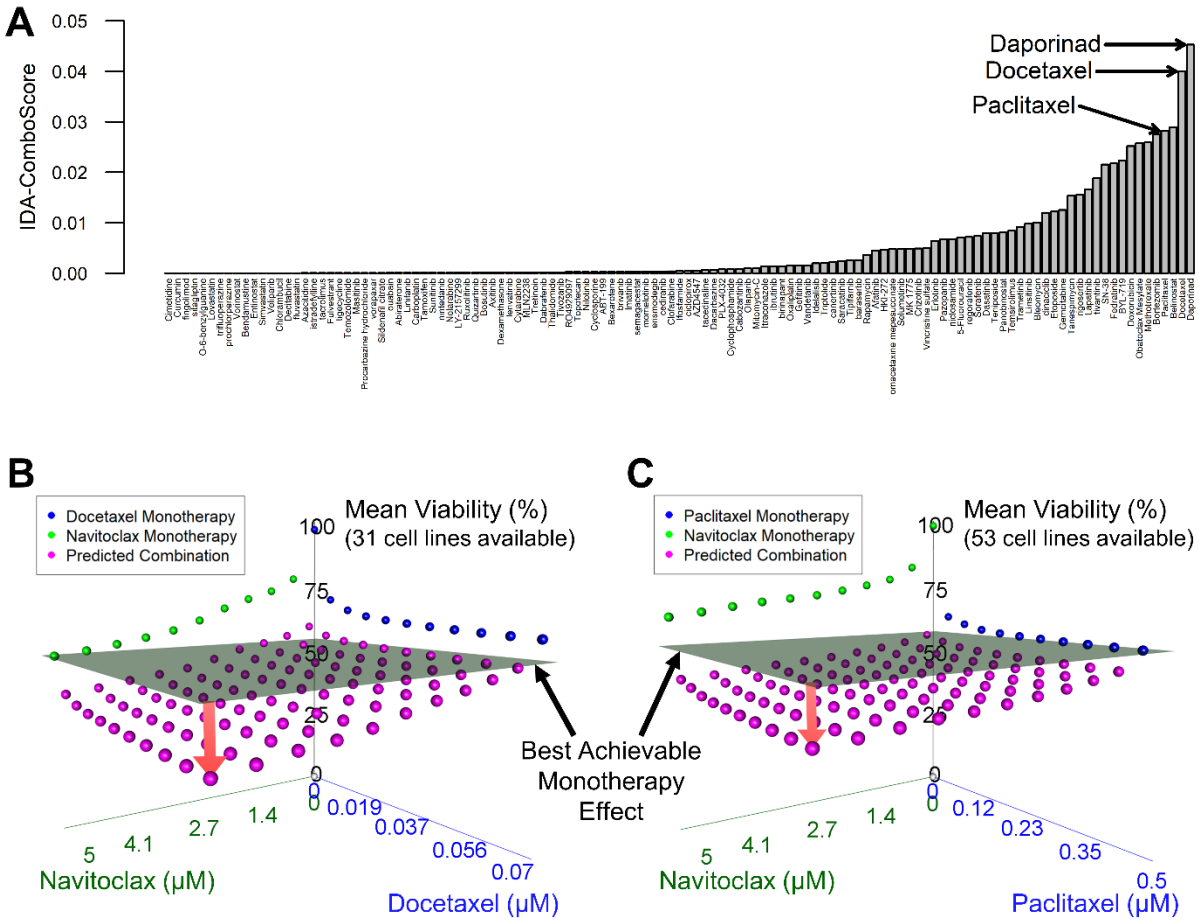

Figure 19. IDA-Combo predicts strong benefits for combinations of navitoclax and taxanes. A) An ordered bar plot of the IDA-comboscores predicted for combinations of navitoclax with other drugs that have reached late-stage clinical trials. Each bar represents a different combination of navitoclax with another drug. **B & C)** 3-D plots of measured and predicted average cell viabilities at different concentrations of navitoclax and docetaxel **(B)** or paclitaxel **(C)**. The transparent plane represents the lowest average viability achievable with monotherapy. The red arrow represents the difference between the best observed monotherapy effect and the best predicted combination effect, which suggests that the combination therapy will reduce tumor cell viability below what is achievable with monotherapy alone.

with docetaxel in patients with solid tumors showed that the combination could be safely administered (Puglisi et al., 2011). A separate phase I trial of the combination of nativoclax + paclitaxel reported modest activity of the combination in patients with solid tumors, though the study was discontinued due to toxicity issues partially resulting from the inclusion of carboplatin

in a second arm of the study (Vlahovic et al., 2014). To our knowledge, the combination of navitoclax with a taxane has not been clinically tested beyond a phase I trial.

Given these findings and our own predictions, I believe the combination of navitoclax with taxanes would be significantly more efficacious than either monotherapy alone in EGFR wild type lung cancer patients who have not received previous chemotherapy. More importantly, these findings demonstrate the feasibility of using IDACombo predictions to identify novel drug combinations for further development.

DISCUSSION

Our results demonstrate that IDACombo can be used with monotherapy cell line screening data to accurately predict drug combination efficacy both *in vitro* and in the clinic. While this does nothing to diminish the importance of continued efforts to understand and predict drug synergy and additivity, it does demonstrate that clinically meaningful predictions can be made using the simpler IDA hypothesis while methods which account for additivity and synergy continue to improve. One critical importance of our work lies in the fact that, while it is currently infeasible to experimentally test the vast number of possible cancer drug combinations, the algorithmic simplicity of IDACombo could allow researchers to computationally predict the efficacies of hundreds of millions of drug combinations in a matter of weeks to months. This enables data driven drug combination selection and could significantly speed up the rate of novel cancer drug combination discovery.

IDACombo works well despite the limited number of available cancer-specific cell lines and uncertainties in clinically relevant drug concentrations

Beyond the surprising utility of the relatively simple IDA model for predicting clinical trial outcomes, the close agreement of IDACombo's clinical predictions with published clinical trial results (at least for trials in chemo-naïve trials) is remarkable for two other reasons.

First, these results demonstrate that *in vitro* drug screening data can be used to generate clinically meaningful predictions for drug combination efficacies in patients, and, furthermore, they suggest that many of these predictions can be made using pan-cancer sets of cell lines. This is somewhat unexpected given the wide range of genetic and phenotypic diversities observed between different cancer types. On the other hand, our results suggest that it will be necessary to make predictions using cell lines of the appropriate cancer type/subtype for targeted therapies, and I believe it is likely that cancer-specific IDACombo predictions could be comparable to or better than pan-cancer predictions if not for the fact that many cancer types currently have relatively few available cell lines in CTRPv2 and GDSC. The solution to this problem, however, may be more complicated than simply increasing the number of cell lines for each cancer type. That is because it must also be noted that, beyond the limited numbers of cell lines available for many cancer types, the ethnic diversity of available CCLs is also very limited—particularly for ethnicities other than Caucasian or Asian (Ling et al., 2018). This means that caution will be necessary when applying the predictions made in this paper to ethnicities that are poorly represented in the cell lines currently available in CTRPv2 and GDSC. Fortunately, others in the field have already recognized the need to increase the number and genetic diversity of available CCLs (Boehm and Golub, 2015), and the Broad Institute has received an NCI contract to create new CCLs (https://portals.broadinstitute.org/cellfactory). This has already lead to the creation of

over 100 validated cancer models. The use of these models in future monotherapy drug screens may improve predictions made with IDACombo even further.

The second reason that the success of IDACombo is remarkable is that, despite our extensive efforts to identify clinical relevant drug concentrations for each drug in our analysis, these concentrations remain only rough estimates of true clinically relevant concentrations. Beyond the fact that measured plasma concentrations are simply unavailable for some drugs and doses for patients of each cancer type, this is largely because there is little available information about how plasma drug concentrations relate to intratumoral drug concentrations *in vivo.* Similarly, there is little available information about how media drug concentrations relate to intracellular drug concentrations *in vitro*. In the single study I was able to find that did examine these relationships, researchers found that the clinically relevant *in vitro* drug concentration for paclitaxel may be an order of magnitude below clinically measured plasma concentrations (Zasadil et al., 2014). Even with this information, the appropriate paclitaxel concentration to use for different cancer types is unclear, because the concentrations identified by Zasadil et al. were based on only two cell lines and six patients in a single cancer type. Given that our results suggest that varying drug concentrations can significantly affect prediction performance, it is possible that IDACombo predictions could be improved by future research aimed at identifying the *in vitro* drug concentrations that most closely mimic the drug exposure of tumor cells in the clinic. It is notable, however, that IDACombo works as well as it does with such high uncertainties in both the cell line populations and the drug concentrations I used to estimate clinical trial powers.

There are several limitations of IDACombo that may be addressed by future research

Despite the impressive accuracy of IDACombo for predicting clinical trial powers in this study, there are several limitations of the method that must be considered when using it in the future.

First, while IDACombo's predicted efficacies strongly correlate with measured efficacies in NCI-ALMANAC and deviations of predicted efficacies from measured efficacies are generally small, it is still obvious that examples can be found where the measured effect of a drug combination is significantly different from the predicted effect. These may represent true cases of drug synergy, additivity, or antagonism, and the drug interactions present in these combinations could have a significant impact on the clinical behavior of these treatments. Given this result and the fact that synergistic drug combinations are likely to outperform combinations that work via IDA (Palmer and Sorger, 2017), it is likely that predictions based on IDA will fail to identify a subset of highly effective drug combinations. Synergy and additivity based prediction methods will need to be developed to identify such combinations. Fortunately, however, the results of our clinical trial validation analysis suggest that this is not a problem for most clinical drug combinations, as the large majority of them were predicted well using IDACombo, at least for trials in previously untreated patients.

This brings us to a second, and perhaps more serious, limitation of the method, which is an apparent unsuitability of cell-line based IDA predictions for patients who have undergone previous cancer drug treatment. We do not have sufficient data from our analyses to definitively explain this finding, but I can propose several hypotheses for future testing. First, there is the possibility that the difference in model performance between previously treated and previously untreated patients is coincidental—merely due to the model working better for some drugs than

for others and to different drugs being tested in trials of previously treated or untreated patients. Upon a closer inspection of the drugs involved in misclassified trials, however, I believe this is unlikely to be the case. Of the 11 drugs involved in trials that were misclassified for PFS/TTP improvement, all except vandetanib (which was used in only one trial) were also used in trials that were correctly classified, and 8 of the 11 drugs were used in correctly classified trials at least as often as they were used in misclassified trials. A more likely explanation for this finding could be that the cell line models in CTRPv2 and GDSC may more accurately represent chemo-naïve tumors than previously treated tumors. It is well known that drug treatment can induce clonal selection in tumors in ways that alter the tumors' drug sensitivities (Ibragimova et al., 2017). While these altered sensitivities may be reflected in cell lines that were generated from the tumors of previously treated patients (Berendsen et al., 1988), it is likely that the cell lines in CTRPv2 and GDSC were derived under a diverse set of circumstances. As such, I would not expect our population of available cell lines to be a good representation of a population of tumors which had all recently received similar drug treatments. In the future, it may be possible to test this hypothesis by creating panels of cell lines that are derived from patients who had received the same prior therapies as the patients in the trials which were poorly predicted in this study and then test whether predictions made with these cell line panels agree with the clinical findings of those trials.

A third limitation of this study is that our method is currently unable to make predictions for combinations which include immunotherapies or drugs which function by acting systemically on non-tumor cells, such as drugs that act systemically to block hormone synthesis. This is because our predictions rely on *in vitro* drug screening data, and the *in vitro* systems that have been used for high-throughput CCL drug screens lack the ability to mimic immune responses or

non-tumor processes such as systemic hormone production. This does not mean, however, that IDA based predictions of drug combination efficacy are unsuitable for immunotherapies or drugs which act outside of the tumor. Efforts are underway to generate *in vitro* models which may be suitable for screening immunotherapies in the future (Dijkstra et al., 2018) and which could allow for IDA based predictions to be made for immunotherapy combinations. While those models mature, however, IDA based predictions of efficacy for combinations with immunotherapies/systemically acting therapies may be made using the results of monotherapy based clinical trials and the method developed by Palmer and Sorger (2017), providing that cross-resistance can be estimated between combined treatments.

IDACombo can quickly translate *in vitro* monotherapy drug screens into clinically relevant efficacy predictions for combinations of any number of drugs

Despite the limitations of this analysis, our results suggest that IDACombo predictions closely agree with the published results of clinical trials in previously untreated patients. This supports the notion that IDA is sufficient to explain the activity of many drug combinations used to treat cancer, and it provides a framework for translating monotherapy cell line data into clinically meaningful predictions of drug combination efficacy. As demonstrated in our prospective analysis, this can be done in a high-throughput fashion and recapitulates known mechanistic relationships between compounds. As available datasets grow in their number of compounds and cell lines, it is likely that the relationships revealed by IDACombo will become more complex. Furthermore, our example analysis to identify efficacious 2-drug combinations with navitoclax in EGFR wild type lung cancer demonstrates the ease with which IDACombo predictions can be assessed for expected efficacy to identify candidate drug combinations for further development. While this was only demonstrated for 2-drug combinations, this method

can also be used to predict the efficacy of combinations of more than two drugs, as was demonstrated in the clinical validation analysis. The only difference with such an analysis from that shown in the prospective analysis is that drug concentrations need to be varied for more than two drugs, which results in a higher dimensional dataset. While this makes visualization of the data more difficult, algorithmic processing of the data remains trivial.

As long as clinically relevant concentrations can be estimated for each drug, IDACombo can be used to generate predictions for hundreds of thousands of 2-drug combinations and millions of combinations of 3 or more drugs using existing *in vitro* datasets. As these datasets grow and novel *in vitro* models make *in vitro* testing feasible for more classes of therapy, the number of combination efficacies that can be predicted with IDACombo will also grow. It is our hope that these predictions will help researchers identify promising combinations for future clinical development and that they will ultimately lead to improved therapies for cancer patients.

CHAPTER 4: CONCLUSIONS: SIGNIFICANCE, INNOVATIONS, AND FUTURE DIRECTIONS

SIGNIFICANCE OF THIS WORK

The research I have presented in this thesis fills two important gaps in the field of cancer biology: 1. The ability to translate drug screening information from cell lines into clinically meaningful information about how patients will respond to those drugs, and 2. The ability to predict the efficacy of drug combinations using only monotherapy data.

The first development is critical for the field, because, as discussed in chapter 2, cell lines provide an extremely tractable model in which high-throughput screening methods can and have been used to screen tens of thousands of compounds relatively quickly and cheaply as compared to animal experiments and without the ethical concerns of testing new drugs in animals or patients. As such, the ability to translate cell line information into clinically relevant efficacy predictions has the potential to not only greatly accelerate the rate at which clinically useful compounds are identified, but it also has the potential to reduce the number of futile animal experiments and clinical trials. This is especially important given that the overall success rate for cancer drugs which enter clinical trials becoming FDA approved is only 11.4% (Wong et al., 2018).

The second development is equally critical for the field, because, as discussed in chapter 3, there are simply too many possible drug combinations to test them all experimentally. The ability to use monotherapy data rather than needing to generate drug combination data for the identification of new clinically useful drug combinations has the potential to save potentially millions or billions of dollars in screening costs. After all, research suggests that combinations including up to 4 drugs may be required to overcome some aggressive cancers (Horn et al.,

2016), and testing all possible 4-drug combinations for even 200 compounds in 1000 cell lines would require more than 64 billion experiments, each requiring multiple drug concentrations and replicate measurements. Furthermore, as discussed in chapter 2, the monotherapy data necessary for model I have developed is already largely mature and publically available, allowing us to make predictions for hundreds of millions of drug combinations using existing data.

INNOVATIONS THAT LEAD TO THE SUCCESS OF THIS RESEARCH

Accounting for Drug Concentrations in a Model of Drug Combination Efficacy

The success of this research depended on several innovations. The first innovation was to model drug responses in a dose dependent fashion. Many previous efforts to model drug combination efficacy have done so using generalized metrics such as IC50 or AUC which summarize efficacy into a single value for each drug/cell line pair (Bansal et al., 2014; Gilvary et al., 2019; He et al., 2018; Jaeger et al., 2017; Menden et al., 2019). Such metrics do not allow models to recognize when a drug combination provides benefit over monotherapy for only a specific range of drug concentrations. This is a critical limitation, because it is well established that the same drug combination may result in the drugs interacting antagonistically or synergistically depending on what concentrations are used for each drug (Meletiadis et al., 2007). Indeed, when our dose-aware model of drug combination efficacy is applied to different drugs, it is common to see different levels of benefit from the combination versus monotherapy depending on what concentrations are used for each drug.

The challenge of making a clinically relevant dose-aware model of drug combination efficacy, however, was not trivial. While the algorithms and code for the model are relatively straightforward, identifying concentrations for each drug that are clinically relevant proved to be painstaking and time-intensive task which required me to read hundreds of phase I and II clinical

trials in which drug plasma concentrations had been measured in patients and, as was discussed in detail in chapter 3, to create what may be a new metric of drug plasma concentration (Csustained). Furthermore, even despite our extensive efforts, there is still a great deal of uncertainty about the appropriateness of the concentrations we identified. This is because: 1. Interactions between drugs can affect the pharmacokinetics of each drug in a combination, and it is impossible to measure plasma concentrations for each drug in every possible combination, 2. The relationships between plasma drug concentrations and tumor drug concentrations are largely unknown and are difficult to study, 3. The relationships between media drug concentrations and intracellular cell line concentrations are also largely unknown, 4. Drug concentration kinetics *in vitro* are often unknown and may not recapitulate clinical pharmacokinetic profiles, and 5. The most suitable time point at which to define drug concentrations is unknown and may vary from drug to drug. In the future, it may be possible to improve the performance of my model by further studying these factors to identify even more clinically relevant drug concentrations for each drug. Regardless of these difficulties, however, my research demonstrates that IDACombo produces highly accurate predictions of drug combination efficacy for trials in previously untreated patients regardless of uncertainties in the drug concentrations selected.

Metrics of Combination Efficacy

The second area of innovation that contributed to the success of my work was the techniques I developed to classify combination efficacy. By using average efficacies across populations of cell lines to calculate hazard ratios, I was able to directly use my efficacy predictions for clinical trial power calculations. This was essential given that the goal for this research was to aid in the development of new drug combination clinical trials, and because the success of clinical trials is determined by much more than the absolute efficacy of the

experimental treatment. The ratio of patients in the control and test arms, the number of observed endpoint events, and the efficacy of the control treatment relative to the experimental treatment are also critically important to whether or not a clinical trial will meet its endpoint criteria. Since, as with clinical trial results, efficacy predictions made by IDACombo are an average across a population, the results of my model can also be used to explore how the expected benefit from a drug combination differs in different populations. As discussed in chapters 2 and 3, however, this is currently difficult for many cancer types, subtypes, and patient ethnicities, because those populations are represented by only a handful of cell lines. As efforts continue to increase the number and diversity of CCL models, this will allow IDACombo to predict drug combination efficacy that will be relevant to an increasingly wide range of patient populations.

Constructing a Clinical Validation Dataset

A third innovation in my research was the construction of a clinical trial dataset to validate my algorithm's clinical utility. By systematically searching the published literature for clinical trials which tested cancer drug combinations, I was able to generate a validation dataset with which to directly benchmark the accuracy, sensitivity, specificity, and precision of IDACombo for predicting clinical trial outcomes. Furthermore, the unbiased nature of this dataset was ensured by establishing inclusion criteria for clinical trials before using IDACombo to predict clinical trial outcomes (with the exception of omitting trials that tested drugs that had clinical concentrations >2x their maximum tested concentrations *in vitro*, which was a problem that I was not initially aware we would encounter). Beyond validating IDACombo, this approach also provides a direct method for translating IDACombo predictions into established criteria (study power) used in designing new clinical trials, which may enable researchers to more accurately estimate study sizes in the future, potentially reducing the number of patients exposed

to unproven therapies or, alternatively, preventing the failure of clinical trials which are testing efficacious therapies with an inappropriately small number of patients.

As with my efforts to account for clinical drug concentrations, however, constructing a clinical trial dataset against which to validate IDACombo was difficult and time-consuming. Beyond the computational work to identifying candidate clinical trials, I had to screen more than a thousand clinical trial papers manually to determine whether or not each trial met the inclusion criteria and, if it did, to extract and summarize the study design elements and outcomes for each trial. As with my efforts to identify clinically relevant drug concentrations, my efforts to identify relevant clinical trials also have some significant limitations. The most notable limitation is that my search relied on ClinicalTrials.gov, which is a website which was not created until 1997 (it became publically available in 2000), and which not all clinical trials are required to register on. Given that much of the work which established the utility of drug combinations for cancer was performed before 1997 (see chapter 1), it is likely that this approach omits many relevant clinical trials. Despite this limitation, however, the approached identified dozens of relevant trials in an unbiased fashion, providing what I believe is a robust validation set. Future efforts to identify additional cancer drug combination trials may further improve the ability of the field to benchmark algorithms to predict clinical drug combination efficacy.

FUTURE DIRECTIONS

As I have discussed the future opportunities created by this research in detail at the ends of chapters 2 and 3, I will discuss previously mentioned opportunities in brief and general terms here so as to avoid unnecessary repetition.

As is often pointed out throughout this thesis, there are many research avenues which may yield further improvements on the presented results. Much may be done to improve our

confidence in the clinically relevant drug concentrations used with IDACombo, and doing so will likely require a rethinking of when and where we measure drug concentrations both in patients and in *in vitro* models. Given the importance of drug concentrations on the efficacy of drug combinations, I believe it is likely that useful insights may be obtained from detailed studies of how drug concentrations change over time in patient tumors and how *in vitro* drug concentration dynamics affect CCLs.

Further improvements to IDACombo's predictions may also be made by increasing the number and diversity of the *in vitro* cancer models used to generate the monotherapy data upon which IDACombo relies. Fortunately, this need is already understood (Williams and McDermott, 2017) and efforts to address it are well underway (Boehm and Golub, 2015; Gao et al., 2015). Beyond allowing IDACombo to make more accurate cancer type and subtype specific predictions, these efforts may also allow researchers to use IDACombo to identify gender and ethnic differences in drug combination sensitivities. Similarly, the development of new types of *in vitro* models and screening techniques may someday allow IDACombo to be useful for predicting the efficacy of combinations involving immunotherapies or therapies which act outside of the tumor as methods are developed which allow *in vitro* measurements of efficacy for those types of therapies.

While work in these areas may further improve the utility of IDACombo, the performance of the model is already remarkably good, and the most exciting future extensions of my research will likely involve using IDACombo to identify novel drug combinations for clinical development. As the scope of my thesis work was merely to create and validate this method, very little has been done so far to use IDACombo for the identification of novel drug combinations. The IDACombo package has been designed to make it easy for researchers to use

it to predict the efficacies of combinations of current standard of care treatments with one or more additional drugs, requiring only that monotherapy efficacy data be available in a sufficient number of CCLs (generally ≥50) for the necessary compounds at their desired concentrations. These predictions can then be used to calculate HRs for the novel combinations versus the standard of care treatments, and power calculations can be easily performed to estimate the likelihood of observing a statistically significant improvement in patient outcomes in a clinical trial. Beyond being used to select combinations for clinical development, these calculations can also be used to aid in the design of such trials by helping researchers determine the necessary sample sizes for each trial so as to avoid exposing an unnecessarily large number of patients to experimental treatments or, conversely, to prevent effective drug combinations from failing to gain FDA approval due to their being tested in trials with inappropriately small numbers of patients.

It is my hope that the ease with which IDACombo allows efficacy predictions to be made for large numbers of drug combinations will greatly increase the speed at which new efficacious drug combinations are identified for cancer treatment and ultimately, that these new treatments will improve outcomes for cancer patients.

REFERENCES

1. Abbas, R., Hug, B.A., Leister, C., Gaaloul, M.E., Chalon, S., and Sonnichsen, D. (2012). A phase I ascending single-dose study of the safety, tolerability, and pharmacokinetics of bosutinib (SKI-606) in healthy adult subjects. Cancer Chemother. Pharmacol. *69*, 221–227.

2. Adler, D., Murdoch, D., Nenadic, O., Urbanek, S., Chen, M., Gebhardt, A., Bolker, B., Csardi, G., Strzelecki, A., Senger, A., et al. (2017). rgl: 3D Visualization Using OpenGL. R package version 0.98.1. https://CRAN.R-project.org/package=rgl.

3. Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.-Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B., et al. (2016). Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. Cancer Discov. *6*, 914–929.

4. Albain, K.S., Barlow, W.E., Ravdin, P.M., Farrar, W.B., Burton, G.V., Ketchel, S.J., Cobau, C.D., Levine, E.G., Ingle, J.N., Pritchard, K.I., et al. (2009). A Randomized Trial of Adjuvant Chemotherapy and Tamoxifen Timing in Postmenopausal, Endocrine-Responsive, Node-Positive Breast Cancer. Lancet *374*, 2055–2063.

5. Amaravadi, R.K., Schilder, R.J., Martin, L.P., Levin, M., Graham, M.A., Weng, D.E., and Adjei, A.A. (2015). A Phase 1 Study of the SMAC-Mimetic Birinapant in Adults with Refractory Solid Tumors or Lymphoma. Mol. Cancer Ther. *14*, 2569–2575.

6. Aoki, D., Katsumata, N., Nakanishi, T., Kigawa, J., Fujiwara, K., Takehara, K., Kamiura, S., Hiura, M., Hatae, M., Sugiyama, T., et al. (2011). A phase II clinical trial of topotecan in Japanese patients with relapsed ovarian carcinoma. Jpn. J. Clin. Oncol. *41*, 320–327.

7. Aoki, T., Nishikawa, R., Mizutani, T., Nojima, K., Mishima, K., Adachi, J., and Matsutani, M. (2007). Pharmacokinetic study of temozolomide on a daily-for-5-days schedule in Japanese patients with relapsed malignant gliomas: first study in Asians. Int. J. Clin. Oncol. *12*, 341–349.

8. Araujo, J.C., Trudel, G.C., Saad, F., Armstrong, A.J., Yu, E.Y., Bellmunt, J., Wilding, G., McCaffrey, J., Serrano, S.V., Matveev, V.B., et al. (2013). Docetaxel and dasatinib or placebo in men with metastatic castration-resistant prostate cancer (READY): a randomised, double-blind phase 3 trial. Lancet Oncol. *14*, 1307–1316.

9. Attard, G., Reid, A.H.M., Yap, T.A., Raynaud, F., Dowsett, M., Settatree, S., Barrett, M., Parker, C., Martins, V., Folkerd, E., et al. (2008). Phase I Clinical Trial of a Selective Inhibitor of CYP17, Abiraterone Acetate, Confirms That Castration-Resistant Prostate Cancer Commonly Remains Hormone Driven. J. Clin. Oncol. *26*, 4563–4571.

10. Bajetta, E., Floriani, I., Di Bartolomeo, M., Labianca, R., Falcone, A., Di Costanzo, F., Comella, G., Amadori, D., Pinto, C., Carlomagno, C., et al. (2014). Randomized trial on adjuvant treatment with FOLFIRI followed by docetaxel and cisplatin versus 5-

fluorouracil and folinic acid for radically resected gastric cancer. Ann. Oncol. *25*, 1373–1378.

11. Baker, S.D., Zhao, M., Lee, C.K.K., Verweij, J., Zabelina, Y., Brahmer, J.R., Wolff, A.C., Sparreboom, A., and Carducci, M.A. (2004). Comparative Pharmacokinetics of Weekly and Every-Three-Weeks Docetaxel. Clin. Cancer Res. *10*, 1976–1983.

12. Balis, F.M., Thompson, P.A., Mosse, Y.P., Blaney, S.M., Minard, C.G., Weigel, B.J., and Fox, E. (2017). First-dose and steady-state pharmacokinetics of orally administered crizotinib in children with solid tumors: a report on ADVL0912 from the Children's Oncology Group Phase 1/Pilot Consortium. Cancer Chemother. Pharmacol. *79*, 181–187.

13. Bansal, M., Yang, J., Karan, C., Menden, M.P., Costello, J.C., Tang, H., Xiao, G., Li, Y., Allen, J., Zhong, R., et al. (2014). A community computational challenge to predict the activity of pairs of compounds. Nat. Biotechnol. *32*, 1213–1222.

14. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

15. Barrett, T., Clark, K., Gevorgyan, R., Gorelenkov, V., Gribov, E., Karsch-Mizrachi, I., Kimelman, M., Pruitt, K.D., Resenchuk, S., Tatusova, T., et al. (2012). BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. Nucleic Acids Res. *40*, D57–D63.

16. Basu, A., Bodycombe, N.E., Cheah, J.H., Price, E.V., Liu, K., Schaefer, G.I., Ebright, R.Y., Stewart, M.L., Ito, D., Wang, S., et al. (2013). An Interactive Resource to Identify Cancer Genetic and Lineage Dependencies Targeted by Small Molecules. Cell *154*, 1151–1161.

17. Bear, H.D., Anderson, S., Smith, R.E., Geyer, C.E., Mamounas, E.P., Fisher, B., Brown, A.M., Robidoux, A., Margolese, R., Kahlenberg, M.S., et al. (2006). Sequential Preoperative or Postoperative Docetaxel Added to Preoperative Doxorubicin Plus Cyclophosphamide for Operable Breast Cancer: National Surgical Adjuvant Breast and Bowel Project Protocol B-27. J. Clin. Oncol. *24*, 2019–2027.

18. Bellmunt, J., von der Maase, H., Mead, G.M., Skoneczna, I., De Santis, M., Daugaard, G., Boehle, A., Chevreau, C., Paz-Ares, L., Laufman, L.R., et al. (2012). Randomized Phase III Study Comparing Paclitaxel/Cisplatin/ Gemcitabine and Gemcitabine/Cisplatin in Patients With Locally Advanced or Metastatic Urothelial Cancer Without Prior Systemic Therapy: EORTC Intergroup Study 30987. J. Clin. Oncol. *30*, 1107–1113.

19. Berenbaum, M.C. (1989). What is synergy? Pharmacol. Rev. *41*, 93–141.

20. Berendsen, H.H., Leij, L. de, Vries, E.G.E. de, Mesander, G., Mulder, N.H., Jong, B. de, Buys, C.H.C.M., Postmus, P.E., Poppema, S., Sluiter, H.J., et al. (1988). Characterization

of Three Small Cell Lung Cancer Cell Lines Established from One Patient during Longitudinal Follow-up. Cancer Res. *48*, 6891–6899.

21. Bergman, A., Ebel, D., Liu, F., Stone, J., Wang, A., Zeng, W., Chen, L., Dilzer, S., Lasseter, K., Herman, G., et al. (2007). Absolute bioavailability of sitagliptin, an oral dipeptidyl peptidase-4 inhibitor, in healthy volunteers. Biopharm. Drug Dispos. *28*, 315–322.

22. Berkenblit, A., Eder, J.P., Ryan, D.P., Seiden, M.V., Tatsuta, N., Sherman, M.L., Dahl, T.A., Dezube, B.J., and Supko, J.G. (2007). Phase I Clinical Trial of STA-4783 in Combination with Paclitaxel in Patients with Refractory Solid Tumors. Clin. Cancer Res. *13*, 584–590.

23. BioAlliance Pharma (2013). English translation of French "Document E." http://www.onxeo.com/wp-content/uploads/2013/05/Document-E-VE.pdf. Accessed August 7, 2019.

24. Bliss, C.I. (1939). The toxicity of poisons applied jointly. Ann. Appl. Biol. *26*, 585–615.

25. Blumenschein, G.R., Reckamp, K., Stephenson, G.J., O'Rourke, T., Gladish, G., McGreivy, J., Sun, Y.-N., Ye, Y., Parson, M., and Sandler, A. (2010). Phase 1b Study of Motesanib, an Oral Angiogenesis Inhibitor, in Combination with Carboplatin/Paclitaxel and/or Panitumumab for the Treatment of Advanced Non–Small Cell Lung Cancer. Clin. Cancer Res. *16*, 279–290.

26. Bocci, G., Danesi, R., Di Paolo, A.D., Innocenti, F., Allegrini, G., Falcone, A., Melosi, A., Battistoni, M., Barsanti, G., Conte, P.F., et al. (2000). Comparative pharmacokinetic analysis of 5-fluorouracil and its major metabolite 5-fluoro-5,6-dihydrouracil after conventional and reduced test dose in cancer patients. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *6*, 3032–3037.

27. Boehm, J.S., and Golub, T.R. (2015). An ecosystem of cancer cell line factories to support a cancer dependency map. Nat. Rev. Genet. *16*, 373–374.

28. du Bois, A., Herrstedt, J., Hardy-Bessard, A.-C., Müller, H.-H., Harter, P., Kristensen, G., Joly, F., Huober, J., Avall-Lundqvist, E., Weber, B., et al. (2010). Phase III trial of carboplatin plus paclitaxel with or without gemcitabine in first-line treatment of epithelial ovarian cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *28*, 4162–4169.

29. du Bois, A., Kristensen, G., Ray-Coquard, I., Reuss, A., Pignata, S., Colombo, N., Denison, U., Vergote, I., del Campo, J.M., Ottevanger, P., et al. (2016). Standard first-line chemotherapy with or without nintedanib for advanced ovarian cancer (AGO-OVAR 12): a randomised, double-blind, placebo-controlled phase 3 trial. Lancet Oncol. *17*, 78–89.

30. Bonneterre, J., Chevalier, B., Focan, C., Mauriac, L., and Piccart, M. (2001). Phase I and pharmacokinetic study of weekly oral therapy with vinorelbine in patients with advanced breast cancer (ABC). Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. *12*, 1683–1691.

31. de Bono, J., Ramanathan, R.K., Mina, L., Chugh, R., Glaspy, J., Rafii, S., Kaye, S., Sachdev, J., Heymach, J., Smith, D.C., et al. (2017). Phase I, Dose-Escalation, Two-Part Trial of the PARP Inhibitor Talazoparib in Patients with Advanced Germline BRCA1/2 Mutations and Selected Sporadic Cancers. Cancer Discov. *7*, 620–629.

32. Bouhaddou, M., DiStefano, M.S., Riesel, E.A., Carrasco, E., Holzapfel, H.Y., Jones, D.C., Smith, G.R., Stern, A.D., Somani, S.S., Thompson, T.V., et al. (2016). Drug response consistency in CCLE and CGP. Nature *540*, E9–E10.

33. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. *68*, 394–424.

34. Breithaupt, H., Dammann, A., and Aigner, K. (1982). Pharmacokinetics of dacarbazine (DTIC) and its metabolite 5-aminoimidazole-4-carboxamide (AIC) following different dose schedules. Cancer Chemother. Pharmacol. *9*, 103–109.

35. Brooks, D.J., Srinivas, N.R., Alberts, D.S., Thomas, T., Igwemzie, L.M., McKinney, L.M., Randolph, J., Schacter, L., Kaul, S., and Barbhaiya, R.H. (1995). Phase I and pharmacokinetic study of etoposide phosphate. Anticancer. Drugs *6*, 637–644.

36. Bukowska, B., Gajek, A., and Marczak, A. (2015). Two drugs are better than one. A short history of combined therapy of ovarian cancer. Contemp. Oncol. *19*, 350–353.

37. Bulusu, K.C., Guha, R., Mason, D.J., Lewis, R.P.I., Muratov, E., Kalantar Motamedi, Y., Cokol, M., and Bender, A. (2016). Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. Drug Discov. Today *21*, 225–238.

38. Burgen, A.S.V., and Mitchell, J.F. (1968). Gaddum's Pharmacology (London: Oxford University Press).

39. Burz, C., Berindan-Neagoe, I.B.-N., Balacescu, O., Tanaselia, C., Ursu, M., Gog, A., Vlase, L., Chintoanu, M., Balacescu, L., Leucuta, S.E., et al. (2009). Clinical and pharmacokinetics study of oxaliplatin in colon cancer patients. J. Gastrointest. Liver Dis. JGLD *18*, 39–43.

40. Canal, P., Bugat, R., Michel, C., Roche, H., Soula, G., and Combes, P.F. (1985). Pharmacokinetics of teniposide (VM 26) after IV administration in serum and malignant ascites of patients with ovarian carcinoma. Cancer Chemother. Pharmacol. *15*, 149–152.

41. Cantarovich, F., Bizollon, C., Cantarovich, D., Lefrançois, N., Dubernard, J.M., and Traeger, J. (1988). Cyclosporine plasma levels six hours after oral administration. A useful tool for monitoring therapy. Transplantation *45*, 389–394.

42. Carrato, A., Swieboda-Sadlej, A., Staszewska-Skurczynska, M., Lim, R., Roman, L., Shparyk, Y., Bondarenko, I., Jonker, D.J., Sun, Y., De la Cruz, J.A., et al. (2013).

Fluorouracil, Leucovorin, and Irinotecan Plus Either Sunitinib or Placebo in Metastatic Colorectal Cancer: A Randomized, Phase III Trial. J. Clin. Oncol. *31*, 1341–1347.

43. Cavo, M., Tacchetti, P., Patriarca, F., Petrucci, M.T., Pantani, L., Galli, M., Di Raimondo, F., Crippa, C., Zamagni, E., Palumbo, A., et al. (2010). Bortezomib with thalidomide plus dexamethasone compared with thalidomide plus dexamethasone as induction therapy before, and consolidation therapy after, double autologous stem-cell transplantation in newly diagnosed multiple myeloma: a randomised phase 3 study. The Lancet *376*, 2075–2085.

44. Chabot, G.G., Abigerges, D., Catimel, G., Culine, S., de Forni, M., Extra, J.-M., Mahjoubi, M., Hérait, P., Armand, J.-P., Bugat, R., et al. (1995). Population pharmacokinetics and pharmacodynamics of irinotecan (CPT-11) and active metabolite SN-38 during phase I trials. Ann. Oncol. *6*, 141–151.

45. Chen, J., Jin, S., Abraham, V., Huang, X., Liu, B., Mitten, M.J., Nimmer, P., Lin, X., Smith, M., Shen, Y., et al. (2011). The Bcl-2/Bcl-XL/Bcl-w Inhibitor, Navitoclax, Enhances the Activity of Chemotherapeutic Agents In Vitro and In Vivo. Mol. Cancer Ther. *10*, 2340–2349.

46. Chen, T.-L., Passos-Coelho, J.L., Noe, D.A., Kennedy, M.J., Black, K.C., Colvin, O.M., and Grochow, L.B. (1995). Nonlinear Pharmacokinetics of Cyclophosphamide in Patients with Metastatic Breast Cancer Receiving High-Dose Chemotherapy followed by Autologous Bone Marrow Transplantation. Cancer Res. *55*, 810–816.

47. Cheng, A.L., Hsu, C.H., Lin, J.K., Hsu, M.M., Ho, Y.F., Shen, T.S., Ko, J.Y., Lin, J.T., Lin, B.R., Ming-Shiang, W., et al. (2001). Phase I clinical trial of curcumin, a chemopreventive agent, in patients with high-risk or pre-malignant lesions. Anticancer Res. *21*, 2895–2900.

48. Cheung, H.W., Cowley, G.S., Weir, B.A., Boehm, J.S., Rusin, S., Scott, J.A., East, A., Ali, L.D., Lizotte, P.H., Wong, T.C., et al. (2011). Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. Proc. Natl. Acad. Sci. *108*, 12372–12377.

49. Choo, S.P., Chowbay, B., Ng, Q.S., Thng, C.H., Lim, C., Hartono, S., Koh, T.S., Huynh, H., Poon, D., Ang, M.K., et al. (2013). A Phase 1 dose-finding and pharmacodynamic study of rapamycin in combination with bevacizumab in patients with unresectable hepatocellular carcinoma. Eur. J. Cancer *49*, 999–1008.

50. Chou, T.-C., and Talalay, P. (1984). Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. Adv. Enzyme Regul. *22*, 27–55.

51. Chu, N.-N., Chen, W.-L., Xu, H.-R., and Li, X.-N. (2012). Pharmacokinetics and Safety of Ezetimibe/Simvastatin Combination Tablet. Clin. Drug Investig. *32*, 791–798.

52. Colucci, G., Labianca, R., Di Costanzo, F., Gebbia, V., Cartenì, G., Massidda, B., Dapretto, E., Manzione, L., Piazza, E., Sannicolò, M., et al. (2010). Randomized Phase

III Trial of Gemcitabine Plus Cisplatin Compared With Single-Agent Gemcitabine As First-Line Treatment of Patients With Advanced Pancreatic Cancer: The GIP-1 Study. J. Clin. Oncol. *28*, 1645–1651.

53. Comandone, A., Passera, R., Boglione, A., Tagini, V., Ferrari, S., and Cattel, L. (2005). High dose methotrexate in adult patients with osteosarcoma: clinical and pharmacokinetic results. Acta Oncol. Stockh. Swed. *44*, 406–411.

54. Conley, B.A., Egorin, M.J., Sridhara, R., Finley, R., Hemady, R., Wu, S., Tait, N.S., Echo, D.A.V., and Echo, D.A.V. (1997). Phase I clinical trial of all-trans-retinoic acid with correlation of its pharmacokinetics and pharmacodynamics. Cancer Chemother. Pharmacol. *39*, 291–299.

55. Corsello, S.M., Bittker, J.A., Liu, Z., Gould, J., McCarren, P., Hirschman, J.E., Johnston, S.E., Vrcic, A., Wong, B., Khan, M., et al. (2017). The Drug Repurposing Hub: a next-generation drug library and information resource. Nat. Med. *23*, 405–408.

56. Costa, R.L.B., and Gradishar, W.J. (2017). Differences Are Important: Breast Cancer Therapy in Different Ethnic Groups. J. Glob. Oncol. *3*, 281–284.

57. Cowley, G.S., Weir, B.A., Vazquez, F., Tamayo, P., Scott, J.A., Rusin, S., East-Seletsky, A., Ali, L.D., Gerath, W.F., Pantel, S.E., et al. (2014). Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. Sci. Data *1*, 140035.

58. Csárdi, G., and FitzJohn, R. (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress.

59. Daemen, A., Griffith, O.L., Heiser, L.M., Wang, N.J., Enache, O.M., Sanborn, Z., Pepin, F., Durinck, S., Korkola, J.E., Griffith, M., et al. (2013). Modeling precision treatment of breast cancer. Genome Biol. *14*, R110.

60. Dahut, W.L., Aragon-Ching, J.B., Woo, S., Tohnya, T.M., Gulley, J.L., Arlen, P.M., Wright, J.J., and Figg, W.D. (2009). A Phase I Study of Oral Lenalidomide in Patients with Refractory Metastatic Cancer. J. Clin. Pharmacol. *49*, 650–660.

61. Danhauser, L.L., Freimann, J.H., Gilchrist, T.L., Gutterman, J.U., Hunter, C.Y., Yeomans, A.C., and Markowitz, A.B. (1993). Phase I and plasma pharmacokinetic study of infusional fluorouracil combined with recombinant interferon alfa-2b in patients with advanced cancer. J. Clin. Oncol. *11*, 751–761.

62. Davids, M.S., Roberts, A.W., Seymour, J.F., Pagel, J.M., Kahl, B.S., Wierda, W.G., Puvvada, S., Kipps, T.J., Anderson, M.A., Salem, A.H., et al. (2017). Phase I First-in-Human Study of Venetoclax in Patients With Relapsed or Refractory Non-Hodgkin Lymphoma. J. Clin. Oncol. *35*, 826–833.

63. DeVita, V.T., and Chu, E. (2008). A History of Cancer Chemotherapy. Cancer Res. *68*, 8643–8653.

64. Devita, V.T., Young, R.C., and Canellos, G.P. (1975). Combination versus single agent chemotherapy: A review of the basis for selection of drug treatment of cancer. Cancer *35*, 98–110.

65. Di Leo, A., Gomez, H.L., Aziz, Z., Zvirbule, Z., Bines, J., Arbushites, M.C., Guerrera, S.F., Koehler, M., Oliva, C., Stein, S.H., et al. (2008). Phase III, Double-Blind, Randomized Study Comparing Lapatinib Plus Paclitaxel With Placebo Plus Paclitaxel As First-Line Treatment for Metastatic Breast Cancer. J. Clin. Oncol. *26*, 5544–5552.

66. Diamant, Z., Samuelsson Palmgren, G., Westrin, B., and Bjermer, L. (2017). Phase I study evaluating the safety, tolerability and pharmacokinetics of a novel oral dissolvable film containing dexamethasone versus Fortecortin dexamethasone tablets. Eur. Clin. Respir. J. *4*.

67. Dijkstra, K.K., Cattaneo, C.M., Weeber, F., Chalabi, M., van de Haar, J., Fanchi, L.F., Slagter, M., van der Velden, D.L., Kaing, S., Kelderman, S., et al. (2018). Generation of Tumor-Reactive T Cells by Co-culture of Peripheral Blood Lymphocytes and Tumor Organoids. Cell *174*, 1586-1598.e12.

68. Do, K., Wilsker, D., Ji, J., Zlott, J., Freshwater, T., Kinders, R.J., Collins, J., Chen, A.P., Doroshow, J.H., and Kummar, S. (2015). Phase I Study of Single-Agent AZD1775 (MK-1775), a Wee1 Kinase Inhibitor, in Patients With Refractory Solid Tumors. J. Clin. Oncol. *33*, 3409–3415.

69. Doebele, R.C., Conkling, P., Traynor, A.M., Otterson, G.A., Zhao, Y., Wind, S., Stopfer, P., Kaiser, R., and Camidge, D.R. (2012). A phase I, open-label dose-escalation study of continuous treatment with BIBF 1120 in combination with paclitaxel and carboplatin as first-line treatment in patients with advanced non-small-cell lung cancer. Ann. Oncol. *23*, 2094–2102.

70. Doi, T., Tamura, K., Tanabe, Y., Yonemori, K., Yoshino, T., Fuse, N., Kodaira, M., Bando, H., Noguchi, K., Shimamoto, T., et al. (2015). Phase 1 pharmacokinetic study of the oral pan-AKT inhibitor MK-2206 in Japanese patients with advanced solid tumors. Cancer Chemother. Pharmacol. *76*, 409–416.

71. Dolan, M.E., Roy, S.K., Fasanmade, A.A., Paras, P.R., Schilsky, R.L., and Ratain, M.J. (1998). O6-benzylguanine in humans: metabolic, pharmacokinetic, and pharmacodynamic findings. J. Clin. Oncol. *16*, 1803–1810.

72. Douillard, J., Cunningham, D., Roth, A., Navarro, M., James, R., Karasek, P., Jandik, P., Iveson, T., Carmichael, J., Alakl, M., et al. (2000). Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. The Lancet *355*, 1041–1047.

73. Dragulescu, A.A. (2014). xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files. R package version 0.5.7. https://CRAN.R-project.org/package=xlsx.

74. Dueñas-González, A., Zarbá, J.J., Patel, F., Alcedo, J.C., Beslija, S., Casanova, L., Pattaranutaporn, P., Hameed, S., Blair, J.M., Barraclough, H., et al. (2011). Phase III, Open-Label, Randomized Study Comparing Concurrent Gemcitabine Plus Cisplatin and Radiation Followed by Adjuvant Gemcitabine and Cisplatin Versus Concurrent Cisplatin and Radiation in Patients With Stage IIB to IVA Carcinoma of the Cervix. J. Clin. Oncol. *29*, 1678–1685.

75. European Medicines Agency (2018). Assessment report: Alsitek. https://www.ema.europa.eu/en/documents/assessment-report/alsitek-epar-refusal-public-assessment-report_en.pdf. Accessed September 12, 2019.

76. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. Nucleic Acids Res. *46*, D649–D655.

77. Fakih, M.G., Fetterly, G., Egorin, M.J., Muindi, J.R., Espinoza-Delgado, I., Zwiebel, J.A., Litwin, A., Holleran, J.L., Wang, K., and Diasio, R.B. (2010). A phase I, pharmacokinetic, and pharmacodynamic study of two schedules of vorinostat in combination with 5-fluorouracil and leucovorin in patients with refractory solid tumors. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *16*, 3786–3794.

78. Falcone, A., Ricci, S., Brunetti, I., Pfanner, E., Allegrini, G., Barbara, C., Crinò, L., Benedetti, G., Evangelista, W., Fanchini, L., et al. (2007). Phase III Trial of Infusional Fluorouracil, Leucovorin, Oxaliplatin, and Irinotecan (FOLFOXIRI) Compared With Infusional Fluorouracil, Leucovorin, and Irinotecan (FOLFIRI) As First-Line Treatment for Metastatic Colorectal Cancer: The Gruppo Oncologico Nord Ovest. J. Clin. Oncol. *25*, 1670–1676.

79. Fishman, M.N., Srinivas, S., Hauke, R.J., Amato, R.J., Esteves, B., Cotreau, M.M., Strahs, A.L., Slichenmyer, W.J., Bhargava, P., and Kabbinavar, F.F. (2013). Phase Ib study of tivozanib (AV-951) in combination with temsirolimus in patients with renal cell carcinoma. Eur. J. Cancer Oxf. Engl. 1990 *49*, 2841–2850.

80. Flaherty, K.T., Lee, S.J., Zhao, F., Schuchter, L.M., Flaherty, L., Kefford, R., Atkins, M.B., Leming, P., and Kirkwood, J.M. (2013). Phase III Trial of Carboplatin and Paclitaxel With or Without Sorafenib in Metastatic Melanoma. J. Clin. Oncol. *31*, 373–379.

81. Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. *45*, D777–D783.

82. Foucquier, J., and Guedj, M. (2015). Analysis of drug combinations: current methodological landscape. Pharmacol. Res. Perspect. *3*, e00149.

83. Fox, J., and Weisberg, S. (2011). An R Companion to Applied Regression (Thousand Oaks, CA: SAGE Publications).

84. Fraile, R.J., Baker, L.H., Buroker, T.R., Horwitz, J., and Vaitkevicius, V.K. (1980). Pharmacokinetics of 5-Fluorouracil Administered Orally, by Rapid Intravenous and by Slow Infusion. Cancer Res. *40*, 2223–2228.

85. Fraser, T.R. (1872a). Lecture on the Antagonism between the Actions of Active Substances. Br. Med. J. *2*, 457–459.

86. Fraser, T.R. (1872b). Lecture on the Antagonism between the Actions of Active Substances. Br. Med. J. *2*, 485–487.

87. Frei, E., Holland, J.F., Schneiderman, M.A., Pinkel, D., Selkirk, G., Freireich, E.J., Silver, R.T., Gold, G.L., and Regelson, W. (1958). A Comparative Study of Two Regimens of Combination Chemotherapy in Acute Leukemia. Blood *13*, 1126–1148.

88. Frei, E., Freireich, E.J., Gehan, E., Pinkel, D., Holland, J.F., Selawry, O., Haurani, F., Spurr, C.L., Hayes, D.M., James, G.W., et al. (1961). Studies of Sequential and Combination Antimetabolite Therapy in Acute Leukemia: 6-Mercaptopurine and Methotrexate. Blood *18*, 431–454.

89. Frei, E., Karon, M., Levin, R.H., Freireich, E.J., Taylor, R.J., Hananian, J., Selawry, O., Holland, J.F., Hoogstraten, B., Wolman, I.J., et al. (1965). The Effectiveness of Combinations of Antileukemic Agents in Inducing and Maintaining Remission in Children with Acute Leukemia. Blood *26*, 642–656.

90. Fujisaka, Y., Horiike, A., Shimizu, T., Yamamoto, N., Yamada, Y., and Tamura, T. (2006). Phase 1 clinical study of pegylated liposomal doxorubicin (JNS002) in Japanese patients with solid tumors. Jpn. J. Clin. Oncol. *36*, 768–774.

91. Fujisaka, Y., Onozawa, Y., Kurata, T., Yasui, H., Goto, I., Yamazaki, K., Machida, N., Watanabe, J., Shimada, H., Shi, X., et al. (2013). First report of the safety, tolerability, and pharmacokinetics of the Src kinase inhibitor saracatinib (AZD0530) in Japanese patients with advanced solid tumours. Invest. New Drugs *31*, 108–114.

92. Fumoleau, P., Koch, K.M., Brain, E., Lokiec, F., Rezai, K., Awada, A., Hayward, L., Werutsky, G., Bogaerts, J., Marréaud, S., et al. (2014). A phase I pharmacokinetics study of lapatinib and tamoxifen in metastatic breast cancer (EORTC 10053 Lapatam study). The Breast *23*, 663–669.

93. Gaddum, J.H. (1940). Pharmacology (London: Oxford University Press).

94. Gadgeel, S.M., Gandhi, L., Riely, G.J., Chiappori, A.A., West, H.L., Azada, M.C., Morcos, P.N., Lee, R.-M., Garcia, L., Yu, L., et al. (2014). Safety and activity of alectinib against systemic disease and brain metastases in patients with crizotinib-resistant ALK-rearranged non-small-cell lung cancer (AF-002JG): results from the dose-finding portion of a phase 1/2 study. Lancet Oncol. *15*, 1119–1128.

95. Gandhi, V., Plunkett, W., Rodriguez, C.O., Nowak, B.J., Du, M., Ayres, M., Kisor, D.F., Mitchell, B.S., Kurtzberg, J., and Keating, M.J. (1998). Compound GW506U78 in

refractory hematologic malignancies: relationship between cellular pharmacokinetics and clinical response. J. Clin. Oncol. *16*, 3607–3615.

96. Gao, H., Korn, J.M., Ferretti, S., Monahan, J.E., Wang, Y., Singh, M., Zhang, C., Schnell, C., Yang, G., Zhang, Y., et al. (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat. Med. *21*, 1318–1325.

97. Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*, 570–575.

98. Gatzemeier, U., Pluzanska, A., Szczesna, A., Kaukel, E., Roubec, J., De Rosa, F., Milanowski, J., Karnicka-Mlodkowski, H., Pesek, M., Serwatowski, P., et al. (2007). Phase III Study of Erlotinib in Combination With Cisplatin and Gemcitabine in Advanced Non–Small-Cell Lung Cancer: The Tarceva Lung Cancer Investigation Trial. J. Clin. Oncol. *25*, 1545–1552.

99. Ghosal, K., Haag, M., Verghese, P.B., West, T., Veenstra, T., Braunstein, J.B., Bateman, R.J., Holtzman, D.M., and Landreth, G.E. (2016). A randomized controlled study to evaluate the effect of bexarotene on amyloid-β and apolipoprotein E metabolism in healthy subjects. Alzheimers Dement. Transl. Res. Clin. Interv. *2*, 110–120.

100. Giaccone, G., Herbst, R.S., Manegold, C., Scagliotti, G., Rosell, R., Miller, V., Natale, R.B., Schiller, J.H., von Pawel, J., Pluzanska, A., et al. (2004). Gefitinib in Combination With Gemcitabine and Cisplatin in Advanced Non–Small-Cell Lung Cancer: A Phase III Trial—INTACT 1. J. Clin. Oncol. *22*, 777–784.

101. Gianni, L., Baselga, J., Eiermann, W., Porta, V.G., Semiglazov, V., Lluch, A., Zambetti, M., Sabadell, D., Raab, G., Cussac, A.L., et al. (2009). Phase III Trial Evaluating the Addition of Paclitaxel to Doxorubicin Followed by Cyclophosphamide, Methotrexate, and Fluorouracil, As Adjuvant or Primary Systemic Therapy: European Cooperative Trial in Operable Breast Cancer. J. Clin. Oncol. *27*, 2474–2481.

102. Gill, S., Ko, Y.-J., Cripps, C., Beaudoin, A., Dhesy-Thind, S., Zulfiqar, M., Zalewski, P., Do, T., Cano, P., Lam, W.Y.H., et al. (2016). PANCREOX: A Randomized Phase III Study of Fluorouracil/Leucovorin With or Without Oxaliplatin for Second-Line Advanced Pancreatic Cancer in Patients Who Have Received Gemcitabine-Based Chemotherapy. J. Clin. Oncol. *34*, 3914–3920.

103. Gilvary, C., Dry, J.R., and Elemento, O. (2019). Multi-task learning predicts drug combination synergy in cells and in the clinic. BioRxiv 576017.

104. Gojo, I., Jiemjit, A., Trepel, J.B., Sparreboom, A., Figg, W.D., Rollins, S., Tidwell, M.L., Greer, J., Chung, E.J., Lee, M.-J., et al. (2007). Phase 1 and pharmacologic study of MS-275, a histone deacetylase inhibitor, in adults with refractory and relapsed acute leukemias. Blood *109*, 2781–2790.

105. Goldinger, S.M., Gobbi, S., Ding, M., Frauchiger, A.L., Fink-Puches, R., Klemke, C.-D., Dréno, B., Bagot, M., Assaf, C., and Dummer, R. (2015). A multicenter, open label, phase II study to assess the efficacy and safety of APO866 in the treatment of patients with refractory or relapsed cutaneous T-cell lymphoma. J. Clin. Oncol. *33*, e20044–e20044.

106. Graham, R.A., Lum, B.L., Cheeti, S., Jin, J.Y., Jorga, K., Von Hoff, D.D., Rudin, C.M., Reddy, J.C., Low, J.A., and LoRusso, P.M. (2011). Pharmacokinetics of Hedgehog Pathway Inhibitor Vismodegib (GDC-0449) in Patients with Locally Advanced or Metastatic Solid Tumors: the Role of Alpha-1-Acid Glycoprotein Binding. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *17*, 2512–2520.

107. Grahnén, A., von Bahr, C., Lindström, B., and Rosén, A. (1979). Bioavailability and pharmacokinetics of cimetidine. Eur. J. Clin. Pharmacol. *16*, 335–340.

108. Greco, W.R., Bravo, G., and Parsons, J.C. (1995). The search for synergy: a critical review from a response surface perspective. Pharmacol. Rev. *47*, 331–385.

109. Greene, R.F., Collins, J.M., Jenkins, J.F., Speyer, J.L., and Myers, C.E. (1983). Plasma Pharmacokinetics of Adriamycin and Adriamycinol: Implications for the Design of in Vitro Experiments and Treatment Protocols. Cancer Res. *43*, 3417–3421.

110. Greshock, J., Bachman, K.E., Degenhardt, Y.Y., Jing, J., Wen, Y.H., Eastman, S., McNeil, E., Moy, C., Wegrzyn, R., Auger, K., et al. (2010). Molecular Target Class Is Predictive of In vitro Response Profile. Cancer Res. *70*, 3677–3686.

111. Gridelli, C., Perrone, F., Gallo, C., Cigolari, S., Rossi, A., Piantedosi, F., Barbera, S., Ferraù, F., Piazza, E., Rosetti, F., et al. (2003). Chemotherapy for Elderly Patients With Advanced Non-Small-Cell Lung Cancer: The Multicenter Italian Lung Cancer in the Elderly Study (MILES) Phase III Randomized Trial. JNCI J. Natl. Cancer Inst. *95*, 362–372.

112. Grippo, J.F., Zhang, W., Heinzmann, D., Yang, K.H., Wong, J., Joe, A.K., Munster, P., Sarapa, N., and Daud, A. (2014). A phase I, randomized, open-label study of the multiple-dose pharmacokinetics of vemurafenib in patients with BRAFV600Emutation-positive metastatic melanoma. Cancer Chemother. Pharmacol. *73*, 103–111.

113. Grossman, S.A., Carson, K.A., Batchelor, T.T., Lesser, G., Mikkelsen, T., Alavi, J.B., Phuphanich, S., Hammour, T., Fisher, J.D., and Supko, J.G. (2006). The Effect of Enzyme-Inducing Antiseizure Drugs on the Pharmacokinetics and Tolerability of Procarbazine Hydrochloride. Clin. Cancer Res. *12*, 5174–5181.

114. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.

115. Gueorguieva, I., Cleverly, A., Desaiah, D., Azaro, A., Seoane, J., Braña, I., Sicart, E., Miles, C., Lahn, M.M., Mitchell, M.I., et al. (2016). Relative bioavailability of three

formulations of galunisertib administered as monotherapy in patients with advanced or metastatic cancer. Drugs Context *5*, 212303.

116. Gupta, N., Hanley, M.J., Venkatakrishnan, K., Perez, R., Norris, R.E., Nemunaitis, J., Yang, H., Qian, M.G., Falchook, G., Labotka, R., et al. (2016). Pharmacokinetics of ixazomib, an oral proteasome inhibitor, in solid tumour patients with moderate or severe hepatic impairment. Br. J. Clin. Pharmacol. *82*, 728–738.

117. Hageboutros, A., Rogatko, A., Newman, E.M., McAleer, C., Brennan, J., LaCreta, F.P., Hudes, G.R., Ozols, R.F., and O'Dwyer, P.J. (1995). Phase I study of phosphonacetyl-L-aspartate, 5-fluorouracil, and leucovorin in patients with advanced cancer. Cancer Chemother. Pharmacol. *35*, 205–212.

118. Haller, D.G., Rothenberg, M.L., Wong, A.O., Koralewski, P.M., Miller, W.H., Bodoky, G., Habboubi, N., Garay, C., and Olivatto, L.O. (2008). Oxaliplatin Plus Irinotecan Compared With Irinotecan Alone as Second-Line Treatment After Single-Agent Fluoropyrimidine Therapy for Metastatic Colorectal Carcinoma. J. Clin. Oncol. *26*, 4544–4550.

119. Hamberg, P., Boers-Sonderen, M.J., van der Graaf, W.T.A., de Bruijn, P., Suttle, A.B., Eskens, F.A.L.M., Verweij, J., van Herpen, C.M.L., and Sleijfer, S. (2014). Pazopanib exposure decreases as a result of an ifosfamide-dependent drug–drug interaction: results of a phase I study. Br. J. Cancer *110*, 888–893.

120. Hanna, N., Neubauer, M., Yiannoutsos, C., McGarry, R., Arseneau, J., Ansari, R., Reynolds, C., Govindan, R., Melnyk, A., Fisher, W., et al. (2008). Phase III Study of Cisplatin, Etoposide, and Concurrent Chest Radiation With or Without Consolidation Docetaxel in Patients With Inoperable Stage III Non–Small-Cell Lung Cancer: The Hoosier Oncology Group and U.S. Oncology. J. Clin. Oncol. *26*, 5755–5760.

121. Hartigh, J. den, McVie, J.G., Oort, W.J.V., and Pinedo, H.M. (1983). Pharmacokinetics of Mitomycin C in Humans. Cancer Res. *43*, 5017–5021.

122. Hauschild, A., Agarwala, S.S., Trefzer, U., Hogg, D., Robert, C., Hersey, P., Eggermont, A., Grabbe, S., Gonzalez, R., Gille, J., et al. (2009). Results of a Phase III, Randomized, Placebo-Controlled Study of Sorafenib in Combination With Carboplatin and Paclitaxel As Second-Line Treatment in Patients With Unresectable Stage III or Stage IV Melanoma. J. Clin. Oncol. *27*, 2823–2830.

123. Haverty, P.M., Lin, E., Tan, J., Yu, Y., Lam, B., Lianoglou, S., Neve, R.M., Martin, S., Settleman, J., Yauch, R.L., et al. (2016). Reproducible pharmacogenomic profiling of cancer cell line panels. Nature *533*, 333–337.

124. He, H., Tran, P., Gu, H., Tedesco, V., Zhang, J., Lin, W., Gatlik, E., Klein, K., and Heimbach, T. (2017). Midostaurin, a Novel Protein Kinase Inhibitor for the Treatment of Acute Myelogenous Leukemia: Insights from Human Absorption, Metabolism, and Excretion Studies of a BDDCS II Drug. Drug Metab. Dispos. *45*, 540–555.

125. He, X., Folkman, L., and Borgwardt, K. (2018). Kernelized rank learning for personalized drug recommendation. Bioinformatics *34*, 2808–2816.

126. Herbst, R.S., Prager, D., Hermann, R., Fehrenbacher, L., Johnson, B.E., Sandler, A., Kris, M.G., Tran, H.T., Klein, P., Li, X., et al. (2005). TRIBUTE: A Phase III Trial of Erlotinib Hydrochloride (OSI-774) Combined With Carboplatin and Paclitaxel Chemotherapy in Advanced Non–Small-Cell Lung Cancer. J. Clin. Oncol. *23*, 5892–5899.

127. Herbst, R.S., Sun, Y., Eberhardt, W.E.E., Germonpré, P., Saijo, N., Zhou, C., Wang, J., Li, L., Kabbinavar, F., Ichinose, Y., et al. (2010). Vandetanib plus docetaxel versus docetaxel as second-line treatment for patients with advanced non-small-cell lung cancer (ZODIAC): a double-blind, randomised, phase 3 trial. Lancet Oncol. *11*, 619–626.

128. Hexner, E., Roboz, G., Hoffman, R., Luger, S., Mascarenhas, J., Carroll, M., Clementi, R., Bensen-Kennedy, D., and Moliterno, A. (2014). Open-label study of oral CEP-701 (lestaurtinib) in patients with polycythaemia vera or essential thrombocythaemia with JAK2-V617F mutation. Br. J. Haematol. *164*, 83–93.

129. Heyn, R.M., Brubaker, C.A., Burchenal, J.H., Cramblett, H.G., and Wolff, J.A. (1960). The Comparison of 6-Mercaptopurine with the Combination of 6-Mercaptopurine and Azaserine in the Treatment of Acute Leukemia in Children: Results of a Cooperative Study. Blood *15*, 350–359.

130. Hidalgo, M., Siu, L.L., Nemunaitis, J., Rizzo, J., Hammond, L.A., Takimoto, C., Eckhardt, S.G., Tolcher, A., Britten, C.D., Denis, L., et al. (2001). Phase I and Pharmacologic Study of OSI-774, an Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor, in Patients With Advanced Solid Malignancies. J. Clin. Oncol. *19*, 3267–3279.

131. Holbeck, S.L., Camalier, R., Crowell, J.A., Govindharajulu, J.P., Hollingshead, M., Anderson, L.W., Polley, E., Rubinstein, L., Srivastava, A., Wilsker, D., et al. (2017). The National Cancer Institute ALMANAC: A Comprehensive Screening Resource for the Detection of Anticancer Drug Pairs with Enhanced Therapeutic Activity. Cancer Res. *77*, 3564–3576.

132. Holen, K., Saltz, L.B., Hollywood, E., Burk, K., and Hanauske, A.-R. (2008). The pharmacokinetics, toxicities, and biologic effects of FK866, a nicotinamide adenine dinucleotide biosynthesis inhibitor. Invest. New Drugs *26*, 45–51.

133. Homesley, H.D., Filiaci, V., Markman, M., Bitterman, P., Eaton, L., Kilgore, L.C., Monk, B.J., and Ueland, F.R. (2007). Phase III Trial of Ifosfamide With or Without Paclitaxel in Advanced Uterine Carcinosarcoma: A Gynecologic Oncology Group Study. J. Clin. Oncol. *25*, 526–531.

134. Horn, T., Ferretti, S., Ebel, N., Tam, A., Ho, S., Harbinski, F., Farsidjani, A., Zubrowski, M., Sellers, W.R., Schlegel, R., et al. (2016). High-Order Drug Combinations Are Required to Effectively Kill Colorectal Cancer Cells. Cancer Res. *76*, 6950–6963.

135. Howell, S.B., Schiefer, M., Andrews, P.A., Markman, M., and Abramson, I. (1987). The pharmacology of intraperitoneally administered bleomycin. J. Clin. Oncol. *5*, 2009–2016.

136. Howlader, N., Noone, A., Krapcho, M., Miller, D., Brest, A., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D., et al. (2019). SEER Cancer Statistics Review, 1975-2016 (Bethesda, MD: National Cancer Institute).

137. Huang, L., Lizak, P., Dvorak, C.C., Aweeka, F., and Long-Boyle, J. (2014). Simultaneous determination of fludarabine and clofarabine in human plasma by LC-MS/MS. J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci. *960*, 194–199.

138. Huang, S.-Y., Chang, C.-S., Liu, T.-C., Wang, P.-N., Yeh, S.-P., Ho, C.-L., Kuo, M.-C., Lin, H.-Y., Jong, J. de, Chen, J.-Y., et al. (2018). Pharmacokinetic study of bortezomib administered intravenously in Taiwanese patients with multiple myeloma. Hematol. Oncol. *36*, 238–244.

139. Hwang, J.J., Kuruvilla, J., Mendelson, D., Pishvaian, M.J., Deeken, J.F., Siu, L.L., Berger, M.S., Viallet, J., and Marshall, J.L. (2010). Phase I Dose Finding Studies of Obatoclax (GX15-070), a Small Molecule Pan-BCL-2 Family Antagonist, in Patients with Advanced Solid Tumors or Lymphoma. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *16*, 4038–4045.

140. Ibragimova, M.K., Tsyganov, M.M., and Litviakov, N.V. (2017). Natural and chemotherapy-induced clonal evolution of tumors. Biochem. Mosc. *82*, 413–425.

141. Ibrahim, N.K., Desai, N., Legha, S., Soon-Shiong, P., Theriault, R.L., Rivera, E., Esmaeli, B., Ring, S.E., Bedikian, A., Hortobagyi, G.N., et al. (2002). Phase I and Pharmacokinetic Study of ABI-007, a Cremophor-free, Protein-stabilized, Nanoparticle Formulation of Paclitaxel. Clin. Cancer Res. *8*, 1038–1044.

142. Ikeda, K., Terashima, M., Kawamura, H., Takiyama, I., Koeda, K., Takagane, A., Sato, N., Ishida, K., Iwaya, T., Maesawa, C., et al. (1998). Pharmacokinetics of Cisplatin in Combined Cisplatin and 5-Fluorouracil Therapy: A Comparative Study of Three Different Schedules of Cisplatin Administration. Jpn. J. Clin. Oncol. *28*, 168–175.

143. Ikeda, M., Okusaka, T., Mitsunaga, S., Ueno, H., Tamai, T., Suzuki, T., Hayato, S., Kadowaki, T., Okita, K., and Kumada, H. (2016). Safety and Pharmacokinetics of Lenvatinib in Patients with Advanced Hepatocellular Carcinoma. Clin. Cancer Res. *22*, 1385–1394.

144. Inoue, K., Kuroi, K., Shimizu, S., Rai, Y., Aogi, K., Masuda, N., Nakayama, T., Iwata, H., Nishimura, Y., Armour, A., et al. (2015). Safety, pharmacokinetics and efficacy findings in an open-label, single-arm study of weekly paclitaxel plus lapatinib as first-line therapy for Japanese women with HER2-positive metastatic breast cancer. Int. J. Clin. Oncol. *20*, 1102–1109.

145. Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. Cell *166*, 740–754.

146. Isah, A.O., Rawlins, M.D., and Bateman, D.N. (1991). Clinical pharmacology of prochlorperazine in healthy young males. Br. J. Clin. Pharmacol. *32*, 677–684.

147. Islami, F., Miller, K.D., Siegel, R.L., Zheng, Z., Zhao, J., Han, X., Ma, J., Jemal, A., and Yabroff, K.R. (2019). National and State Estimates of Lost Earnings From Cancer Deaths in the United States. JAMA Oncol. *5*, e191460–e191460.

148. Jaeger, S., Igea, A., Arroyo, R., Alcalde, V., Canovas, B., Orozco, M., Nebreda, A.R., and Aloy, P. (2017). Quantification of Pathway Cross-talk Reveals Novel Synergistic Drug Combinations for Breast Cancer. Cancer Res. *77*, 459–469.

149. Jain, L., Woo, S., Gardner, E.R., Dahut, W.L., Kohn, E.C., Kummar, S., Mould, D.R., Giaccone, G., Yarchoan, R., Venitz, J., et al. (2011). Population pharmacokinetic analysis of sorafenib in patients with solid tumours. Br. J. Clin. Pharmacol. *72*, 294–305.

150. Jin, F., Robeson, M., Zhou, H., Hisoire, G., and Ramanathan, S. (2015). The pharmacokinetics and safety of idelalisib in subjects with moderate or severe hepatic impairment. J. Clin. Pharmacol. *55*, 944–952.

151. de Jong, J., Sukbuntherng, J., Skee, D., Murphy, J., O'Brien, S., Byrd, J.C., James, D., Hellemans, P., Loury, D.J., Jiao, J., et al. (2015). The effect of food on the pharmacokinetics of oral ibrutinib in healthy participants and patients with chronic lymphocytic leukemia. Cancer Chemother. Pharmacol. *75*, 907–916.

152. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. *45*, D353–D361.

153. Kantarjian, H.M., Sekeres, M.A., Ribrag, V., Rousselot, P., Garcia-Manero, G., Jabbour, E.J., Owen, K., Stockman, P.K., and Oliver, S.D. (2013). Phase I study assessing the safety and tolerability of barasertib (AZD1152) with low-dose cytosine arabinoside in elderly patients with AML. Clin. Lymphoma Myeloma Leuk. *13*, 559–567.

154. Keenan, T., Moy, B., Mroz, E.A., Ross, K., Niemierko, A., Rocco, J.W., Isakoff, S., Ellisen, L.W., and Bardia, A. (2015). Comparison of the Genomic Landscape Between Primary Breast Cancer in African American Versus White Women and the Association of Racial Differences With Tumor Recurrence. J. Clin. Oncol. *33*, 3621–3627.

155. Kerbusch, T., Mathôt, R. a. A., Keizer, H.J., Kaijser, G.P., Schellens, J.H.M., and Beijnen, J.H. (2001). Influence of Dose and Infusion Duration on Pharmacokinetics of Ifosfamide and Metabolites. Drug Metab. Dispos. *29*, 967–975.

156. Keyvanjah, K., DiPrimeo, D., Li, A., Obaidi, M., Swearingen, D., and Wong, A. (2017). Pharmacokinetics of neratinib during coadministration with lansoprazole in healthy subjects. Br. J. Clin. Pharmacol. *83*, 554–561.

157. Kindler, H.L., Ioka, T., Richel, D.J., Bennouna, J., Létourneau, R., Okusaka, T., Funakoshi, A., Furuse, J., Park, Y.S., Ohkawa, S., et al. (2011). Axitinib plus gemcitabine versus placebo plus gemcitabine in patients with advanced pancreatic adenocarcinoma: a double-blind randomised phase 3 study. Lancet Oncol. *12*, 256–262.

158. Kisanga, E.R., Gjerde, J., Guerrieri-Gonzaga, A., Pigatto, F., Pesci-Feltri, A., Robertson, C., Serrano, D., Pelosi, G., Decensi, A., and Lien, E.A. (2004). Tamoxifen and Metabolite Concentrations in Serum and Breast Cancer Tissue during Three Dose Regimens in a Randomized Preoperative Trial. Clin. Cancer Res. *10*, 2336–2343.

159. Kitzen, J.J.E.M., de Jonge, M.J.A., Lamers, C.H.J., Eskens, F.A.L.M., van der Biessen, D., van Doorn, L., ter Steeg, J., Brandely, M., Puozzo, Ch., and Verweij, J. (2009). Phase I dose-escalation study of F60008, a novel apoptosis inducer, in patients with advanced solid tumours. Eur. J. Cancer *45*, 1764–1772.

160. Köhne, C.-H., van Cutsem, E., Wils, J., Bokemeyer, C., El-Serafi, M., Lutz, M. p., Lorenz, M., Reichardt, P., Rückle-Lanz, H., Frickhofen, N., et al. (2005). Phase III Study of Weekly High-Dose Infusional Fluorouracil Plus Folinic Acid With or Without Irinotecan in Patients With Metastatic Colorectal Cancer: European Organisation for Research and Treatment of Cancer Gastrointestinal Group Study 40986. J. Clin. Oncol. *23*, 4856–4865.

161. Kosoglou, T., Reyderman, L., Tiessen, R.G., van Vliet, A.A., Fales, R.R., Keller, R., Yang, B., and Cutler, D.L. (2012). Pharmacodynamics and pharmacokinetics of the novel PAR-1 antagonist vorapaxar (formerly SCH 530348) in healthy subjects. Eur. J. Clin. Pharmacol. *68*, 249–258.

162. Kovarik, J.M., Hartmann, S., Bartlett, M., Riviere, G.-J., Neddermann, D., Wang, Y., Port, A., and Schmouder, R.L. (2007). Oral-intravenous crossover study of fingolimod pharmacokinetics, lymphocyte responses and cardiac effects. Biopharm. Drug Dispos. *28*, 97–104.

163. Kubota, K., Hida, T., Ishikura, S., Mizusawa, J., Nishio, M., Kawahara, M., Yokoyama, A., Imamura, F., Takeda, K., Negoro, S., et al. (2014). Etoposide and cisplatin versus irinotecan and cisplatin in patients with limited-stage small-cell lung cancer treated with etoposide and cisplatin plus concurrent accelerated hyperfractionated thoracic radiotherapy (JCOG0202): a randomised phase 3 study. Lancet Oncol. *15*, 106–113.

164. Kudo, M., Ueshima, K., Yokosuka, O., Ogasawara, S., Obi, S., Izumi, N., Aikata, H., Nagano, H., Hatano, E., Sasaki, Y., et al. (2018). Sorafenib plus low-dose cisplatin and fluorouracil hepatic arterial infusion chemotherapy versus sorafenib alone in patients with advanced hepatocellular carcinoma (SILIUS): a randomised, open label, phase 3 trial. Lancet Gastroenterol. Hepatol. *3*, 424–432.

165. Kuhn, J.G., Chang, S.M., Wen, P.Y., Cloughesy, T.F., Greenberg, H., Schiff, D., Conrad, C., Fink, K.L., Robins, H.I., Mehta, M., et al. (2007). Pharmacokinetic and Tumor Distribution Characteristics of Temsirolimus in Patients with Recurrent Malignant Glioma. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *13*, 7401–7406.

166. Kurata, T., Tamura, T., Shinkai, T., Ohe, Y., Kunitoh, H., Kodama, T., Kakinuma, R., Matsumoto, T., Kubota, K., Omatsu, H., et al. (2001). Phase I and Pharmacological Study of Paclitaxel Given Over 3 h with Cisplatin for Advanced Non-small Cell Lung Cancer. Jpn. J. Clin. Oncol. *31*, 93–99.

167. Lacy, S.A., Miles, D.R., and Nguyen, L.T. (2017). Clinical Pharmacokinetics and Pharmacodynamics of Cabozantinib. Clin. Pharmacokinet. *56*, 477–491.

168. Lang, D.T., and the CRAN Team (2017). XML: Tools for Parsing and Generating XML Within R and S-Plus. R package version 3.98-1.9. https://CRAN.R-project.org/package=XML.

169. Lankheet, N.A.G., Kloth, J.S.L., Gadellaa-van Hooijdonk, C.G.M., Cirkel, G.A., Mathijssen, R.H.J., Lolkema, M.P.J.K., Schellens, J.H.M., Voest, E.E., Sleijfer, S., de Jonge, M.J.A., et al. (2014). Pharmacokinetically guided sunitinib dosing: a feasibility study in patients with advanced solid tumours. Br. J. Cancer *110*, 2441–2449.

170. Law, L.W. (1952). Effects of Combinations of Antileukemic Agents on an Acute Lymphocytic Leukemia of Mice. Cancer Res. *12*, 871–878.

171. Lee, J., Park, S.H., Chang, H.-M., Kim, J.S., Choi, H.J., Lee, M.A., Chang, J.S., Jeung, H.C., Kang, J.H., Lee, H.W., et al. (2012). Gemcitabine and oxaliplatin with or without erlotinib in advanced biliary-tract cancer: a multicentre, open-label, randomised, phase 3 study. Lancet Oncol. *13*, 181–188.

172. Lee, S., Yoon, S.H., Cho, J.-Y., Shin, S.-G., Jang, I.-J., and Yu, K.-S. (2010). Relative Bioavailability and Tolerability of Two Formulations of Bicalutamide 50-mg Tablets: A Randomized-Sequence, Open-Label, Two-Period Crossover Study in Healthy Korean Male Subjects. Clin. Ther. *32*, 2496–2501.

173. Lee, S.M., Woll, P.J., Rudd, R., Ferry, D., O'Brien, M., Middleton, G., Spiro, S., James, L., Ali, K., Jitlal, M., et al. (2009). Anti-angiogenic Therapy Using Thalidomide Combined With Chemotherapy in Small Cell Lung Cancer: A Randomized, Double-Blind, Placebo-Controlled Trial. JNCI J. Natl. Cancer Inst. *101*, 1049–1057.

174. Leijen, S., Soetekouw, P.M.M.B., Jeffry Evans, T.R., Nicolson, M., Schellens, J.H.M., Learoyd, M., Grinsted, L., Zazulina, V., Pwint, T., and Middleton, M. (2011). A phase I, open-label, randomized crossover study to assess the effect of dosing of the MEK 1/2 inhibitor Selumetinib (AZD6244; ARRY-142866) in the presence and absence of food in patients with advanced solid tumors. Cancer Chemother. Pharmacol. *68*, 1619–1628.

175. Levêque, D., Jehl, F., Quoix, E., and Breillout, F. (1992). Clinical Pharmacokinetics of Vinorelbine Alone and Combined with Cisplatin. J. Clin. Pharmacol. *32*, 1096–1098.

176. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. *1*, 417–425.

177. Lilenbaum, R.C., Herndon, J.E., List, M.A., Desch, C., Watson, D.M., Miller, A.A., Graziano, S.L., Perry, M.C., Saville, W., Chahinian, P., et al. (2005). Single-Agent Versus Combination Chemotherapy in Advanced Non–Small-Cell Lung Cancer: The Cancer and Leukemia Group B (study 9730). J. Clin. Oncol. *23*, 190–196.

178. Ling, A., Gruener, R.F., Fessler, J., and Huang, R.S. (2018). More than fishing for a cure: The promises and pitfalls of high throughput cancer cell line screens. Pharmacol. Ther. *191*, 178–189.

179. Liston, D.R., and Davis, M. (2017). Clinically Relevant Concentrations of Anticancer Drugs: A Guide for Nonclinical Studies. Clin. Cancer Res. *23*, 3489–3498.

180. Liu, Y.-M., Zhang, K.E., Liu, Y., Zhang, H.-C., Song, Y.-X., Pu, H.-H., Lu, C., Liu, G.-Y., Jia, J.-Y., Zheng, Q.-S., et al. (2012). Pharmacokinetic Properties and Bioequivalence of Two Sulfadoxine/Pyrimethamine Fixed-Dose Combination Tablets: A Parallel-Design Study in Healthy Chinese Male Volunteers. Clin. Ther. *34*, 2212–2220.

181. LoConte, N.K., Razak, A.R.A., Ivy, P., Tevaarwerk, A., Leverence, R., Kolesar, J., Siu, L., Lubner, S.J., Mulkerin, D.L., Schelman, W.R., et al. (2015). A Multicenter Phase 1 Study of γ -secretase inhibitor RO4929097 in Combination with Capecitabine in Refractory Solid Tumors. Invest. New Drugs *33*, 169–176.

182. Loewe, S. (1928). Die quantitativen Probleme der Pharmakologie. Ergeb. Physiol. *27*, 47–187.

183. Lu, K., Yap, H.-Y., and Loo, T.L. (1983). Clinical Pharmacokinetics of Vinblastine by Continuous Intravenous Infusion. Cancer Res. *43*, 1405–1408.

184. Ma, W.W., Messersmith, W.A., Dy, G.K., Weekes, C.D., Whitworth, A., Ren, C., Maniar, M., Wilhelm, F., Eckhardt, S.G., Adjei, A.A., et al. (2012). Phase I Study of Rigosertib, an Inhibitor of the Phosphatidylinositol 3-Kinase and Polo-like Kinase 1 Pathways, Combined with Gemcitabine in Patients with Solid Tumors and Pancreatic Cancer. Clin. Cancer Res. *18*, 2048–2055.

185. Macaulay, V.M., Middleton, M.R., Eckhardt, S.G., Rudin, C.M., Juergens, R.A., Gedrich, R., Gogov, S., McCarthy, S., Poondru, S., Stephens, A.W., et al. (2016). Phase I Dose-Escalation Study of Linsitinib (OSI-906) and Erlotinib in Patients with Advanced Solid Tumors. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *22*, 2897–2907.

186. Mariotto, A.B., Robin Yabroff, K., Shao, Y., Feuer, E.J., and Brown, M.L. (2011). Projections of the Cost of Cancer Care in the United States: 2010–2020. JNCI J. Natl. Cancer Inst. *103*, 117–128.

187. Martín, M., Ruiz, A., Muñoz, M., Balil, A., García-Mata, J., Calvo, L., Carrasco, E., Mahillo, E., Casado, A., García-Saenz, J.Á., et al. (2007). Gemcitabine plus vinorelbine versus vinorelbine monotherapy in patients with metastatic breast cancer previously treated with anthracyclines and taxanes: final results of the phase III Spanish Breast Cancer Research Group (GEICAM) trial. Lancet Oncol. *8*, 219–225.

188. Marumo, A., Miyawaki, S., Dan, N., and Ishiyama, K. (2017). Plasma Concentration of Itraconazole in Patients With Hematologic Malignancies Treated With Itraconazole Oral Solution. Ther. Drug Monit. *39*, 229–234.

189. Mavroudis, D., Pappas, P., Kouroussis, C., Kakolyris, S., Agelaki, S., Kalbakis, K., Androulakis, N., Souglakos, J., Vardakis, N., Nikolaidou, M., et al. (2003). A dose-escalation and pharmacokinetic study of gemcitabine and oxaliplatin in patients with advanced solid tumors. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. *14*, 304–312.

190. Mehrotra, S., Gopalakrishnan, M., Gobburu, J., Greer, J.M., Piekarz, R., Karp, J.E., Pratz, K., and Rudek, M.A. (2017). Population pharmacokinetics and site of action exposures of veliparib with topotecan plus carboplatin in patients with haematological malignancies. Br. J. Clin. Pharmacol. *83*, 1688–1700.

191. Mekhail, T., Masson, E., Fischer, B.S., Gong, J., Iyer, R., Gan, J., Pursley, J., Patricia, D., Williams, D., and Ganapathi, R. (2010). Metabolism, Excretion, and Pharmacokinetics of Oral Brivanib in Patients with Advanced or Metastatic Solid Tumors. Drug Metab. Dispos. *38*, 1962–1966.

192. Meletiadis, J., Stergiopoulou, T., O'Shaughnessy, E.M., Peter, J., and Walsh, T.J. (2007). Concentration-Dependent Synergy and Antagonism within a Triple Antifungal Drug Combination against Aspergillus Species: Analysis by a New Response Surface Model. Antimicrob. Agents Chemother. *51*, 2053–2064.

193. Melichar, B., Casado, E., Bridgewater, J., Bennouna, J., Campone, M., Vitek, P., Delord, J.-P., Cerman, J., Salazar, R., Dvorak, J., et al. (2011). Clinical activity of patupilone in patients with pretreated advanced/metastatic colon cancer: results of a phase I dose escalation trial. Br. J. Cancer *105*, 1646–1653.

194. Menden, M.P., Wang, D., Mason, M.J., Szalai, B., Bulusu, K.C., Guan, Y., Yu, T., Kang, J., Jeon, M., Wolfinger, R., et al. (2019). Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. Nat. Commun. *10*, 2674.

195. Microsoft, R Core Team (2017). Microsoft R Open. R package version 3.4.2. https://mran.microsoft.com/.

196. Midha, K.K., Korchinski, E.D., Verbeeck, R.K., Roscoe, R.M., Hawes, E.M., Cooper, J.K., and McKay, G. (1983). Kinetics of oral trifluoperazine disposition in man. Br. J. Clin. Pharmacol. *15*, 380–382.

197. Milacic, M., Haw, R., Rothfels, K., Wu, G., Croft, D., Hermjakob, H., D'Eustachio, P., and Stein, L. (2012). Annotating Cancer Variants and Anti-Cancer Therapeutics in Reactome. Cancers *4*, 1180–1211.

198. Minami, H., Ando, Y., Ma, B.B.Y., Hsiang Lee, J., Momota, H., Fujiwara, Y., Li, L., Fukino, K., Ito, K., Tajima, T., et al. (2016). Phase I, multicenter, open-label, dose-escalation study of sonidegib in Asian patients with advanced solid tumors. Cancer Sci. *107*, 1477–1483.

199. Minden, M.D., Hogge, D.E., Weir, S.J., Kasper, J., Webster, D.A., Patton, L., Jitkova, Y., Hurren, R., Gronda, M., Goard, C.A., et al. (2014). Oral ciclopirox olamine displays biological activity in a phase I study in patients with advanced hematologic malignancies. Am. J. Hematol. *89*, 363–368.

200. Moloney, M., Faulkner, D., Link, E., Rischin, D., Solomon, B., Lim, A.M., Zalcberg, J.R., Jefford, M., and Michael, M. (2018). Feasibility of 5-fluorouracil pharmacokinetic monitoring using the My-5FU PCM^{TM} system in a quaternary oncology centre. Cancer Chemother. Pharmacol. *82*, 865–876.

201. Moore, M.J., Goldstein, D., Hamm, J., Figer, A., Hecht, J.R., Gallinger, S., Au, H.J., Murawa, P., Walde, D., Wolff, R.A., et al. (2007). Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *25*, 1960–1966.

202. Morabito, A., Gebbia, V., Di Maio, M., Cinieri, S., Viganò, M.G., Bianco, R., Barbera, S., Cavanna, L., De Marinis, F., Montesarchio, V., et al. (2013). Randomized phase III trial of gemcitabine and cisplatin vs. gemcitabine alone in patients with advanced non-small cell lung cancer and a performance status of 2: the CAPPA-2 study. Lung Cancer Amst. Neth. *81*, 77–83.

203. Moreau, P., Avet-Loiseau, H., Facon, T., Attal, M., Tiab, M., Hulin, C., Doyen, C., Garderet, L., Randriamalala, E., Araujo, C., et al. (2011). Bortezomib plus dexamethasone versus reduced-dose bortezomib, thalidomide plus dexamethasone as induction treatment before autologous stem cell transplantation in newly diagnosed multiple myeloma. Blood *118*, 5752–5758.

204. Mpindi, J.P., Yadav, B., Östling, P., Gautam, P., Malani, D., Murumägi, A., Hirasawa, A., Kangaspeska, S., Wennerberg, K., Kallioniemi, O., et al. (2016). Consistency in drug response profiling. Nature *540*, E5–E6.

205. Mross, K., Richly, H., Schleucher, N., Korfee, S., Tewes, M., Scheulen, M.E., Seeber, S., Beinert, T., Schweigert, M., Sauer, U., et al. (2004). A phase I clinical and pharmacokinetic study of the camptothecin glycoconjugate, BAY 38-3441, as a daily infusion in patients with advanced solid tumors. Ann. Oncol. *15*, 1284–1294.

206. Mross, K., Frost, A., Steinbild, S., Hedbom, S., Büchert, M., Fasol, U., Unger, C., Krätzschmar, J., Heinig, R., Boix, O., et al. (2012). A Phase I Dose–Escalation Study of

Regorafenib (BAY 73–4506), an Inhibitor of Oncogenic, Angiogenic, and Stromal Kinases, in Patients with Advanced Solid Tumors. Clin. Cancer Res. *18*, 2658–2667.

207. Muggia, F.M., Braly, P.S., Brady, M.F., Sutton, G., Niemann, T.H., Lentz, S.L., Alvarez, R.D., Kucera, P.R., and Small, J.M. (2000). Phase III Randomized Study of Cisplatin Versus Paclitaxel Versus Cisplatin and Paclitaxel in Patients With Suboptimal Stage III or IV Ovarian Cancer: A Gynecologic Oncology Group Study. J. Clin. Oncol. *18*, 106–106.

208. Mukai, M., Uchimura, T., Zhang, X., Greene, D., Vergeire, M., and Cantillon, M. (2018). Effects of Rifampin on the Pharmacokinetics of a Single Dose of Istradefylline in Healthy Subjects. J. Clin. Pharmacol. *58*, 193–201.

209. Mukohara, T., Nagai, S., Koshiji, M., Yoshizawa, K., and Minami, H. (2010). Phase I dose escalation and pharmacokinetic study of oral enzastaurin (LY317615) in advanced solid tumors. Cancer Sci. *101*, 2193–2199.

210. Müller, M., Brunner, M., Schmid, R., Mader, R.M., Bockenheimer, J., Steger, G.G., Steiner, B., Eichler, H.G., and Blöchl-Daum, B. (1998). Interstitial methotrexate kinetics in primary breast cancer lesions. Cancer Res. *58*, 2982–2985.

211. Munoz, D.M., Cassiani, P.J., Li, L., Billy, E., Korn, J.M., Jones, M.D., Golji, J., Ruddy, D.A., Yu, K., McAllister, G., et al. (2016). CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. Cancer Discov. *6*, 900–913.

212. Muralidharan, G., Micalizzi, M., Speth, J., Raible, D., and Troy, S. (2005). Pharmacokinetics of Tigecycline after Single and Multiple Doses in Healthy Subjects. Antimicrob. Agents Chemother. *49*, 220–229.

213. Narasimhan, N.I., Dorer, D.J., Niland, K., Haluska, F., and Sonnichsen, D. (2013). Effects of Ketoconazole on the Pharmacokinetics of Ponatinib in Healthy Subjects. J. Clin. Pharmacol. *53*, 974–981.

214. Nemunaitis, J.J., Small, K.A., Kirschmeier, P., Zhang, D., Zhu, Y., Jou, Y.-M., Statkevich, P., Yao, S.-L., and Bannerji, R. (2013). A first-in-human, phase 1, dose-escalation study of dinaciclib, a novel cyclin-dependent kinase inhibitor, administered weekly in subjects with advanced malignancies. J. Transl. Med. *11*, 259.

215. Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1.2. https://CRAN.R-project.org/package=RColorBrewer.

216. Nichols, D.J., Muirhead, G.J., and Harness, J.A. (2002). Pharmacokinetics of sildenafil after single oral doses in healthy male subjects: absolute bioavailability, food effects and dose proportionality. Br. J. Clin. Pharmacol. *53*, 5S-12S.

217. Niesvizky, R., Flinn, I.W., Rifkin, R., Gabrail, N., Charu, V., Clowney, B., Essell, J., Gaffar, Y., Warr, T., Neuwirth, R., et al. (2015). Community-Based Phase IIIB Trial of

Three UPFRONT Bortezomib-Based Myeloma Regimens. J. Clin. Oncol. *33*, 3921–3929.

218. Nijenhuis, C.M., Hellriegel, E., Beijnen, J.H., Hershock, D., Huitema, A.D.R., Lucas, L., Mergui-Roelvink, M., Munteanu, M., Rabinovich-Guilatt, L., Robertson, P., et al. (2016). Pharmacokinetics and excretion of 14C-omacetaxine in patients with advanced solid tumors. Invest. New Drugs *34*, 565–574.

219. Nokihara, H., Yamamoto, N., Ohe, Y., Hiraoka, M., and Tamura, T. (2016). Pharmacokinetics of Weekly Paclitaxel and Feasibility of Dexamethasone Taper in Japanese Patients with Advanced Non–small Cell Lung Cancer. Clin. Ther. *38*, 338–347.

220. Ogura, M., Uchida, T., Taniwaki, M., Ando, K., Watanabe, T., Kasai, M., Matsumoto, Y., Shimizu, D., Ogawa, Y., Ohmachi, K., et al. (2010). Phase I and pharmacokinetic study of bendamustine hydrochloride in relapsed or refractory indolent B-cell non-Hodgkin lymphoma and mantle cell lymphoma. Cancer Sci. *101*, 2054–2058.

221. Oki, Y., Kondo, Y., Yamamoto, K., Ogura, M., Kasai, M., Kobayashi, Y., Watanabe, T., Uike, N., Ohyashiki, K., Okamoto, S., et al. (2012). Phase I/II study of decitabine in patients with myelodysplastic syndrome: A multi-center study in Japan. Cancer Sci. *103*, 1839–1847.

222. O'Neil, B.H., Scott, A.J., Ma, W.W., Cohen, S.J., Aisner, D.L., Menter, A.R., Tejani, M.A., Cho, J.K., Granfortuna, J., Coveler, L., et al. (2015). A phase II/III randomized study to compare the efficacy and safety of rigosertib plus gemcitabine versus gemcitabine alone in patients with previously untreated metastatic pancreatic cancer. Ann. Oncol. *26*, 1923–1929.

223. O'Neil, J., Benita, Y., Feldman, I., Chenard, M., Roberts, B., Liu, Y., Li, J., Kral, A., Lejnine, S., Loboda, A., et al. (2016). An Unbiased Oncology Compound Screen to Identify Novel Combination Strategies. Mol. Cancer Ther. *15*, 1155–1162.

224. Orlowski, R.Z., Nagler, A., Sonneveld, P., Bladé, J., Hajek, R., Spencer, A., San Miguel, J., Robak, T., Dmoszynska, A., Horvath, N., et al. (2007). Randomized Phase III Study of Pegylated Liposomal Doxorubicin Plus Bortezomib Compared With Bortezomib Alone in Relapsed or Refractory Multiple Myeloma: Combination Therapy Improves Time to Progression. J. Clin. Oncol. *25*, 3892–3901.

225. O'Shaughnessy, J., Miles, D., Vukelja, S., Moiseyenko, V., Ayoub, J.-P., Cervantes, G., Fumoleau, P., Jones, S., Lui, W.-Y., Mauriac, L., et al. (2002). Superior Survival With Capecitabine Plus Docetaxel Combination Therapy in Anthracycline-Pretreated Patients With Advanced Breast Cancer: Phase III Trial Results. J. Clin. Oncol. *20*, 2812–2823.

226. Palmer, A.C., and Sorger, P.K. (2017). Combination Cancer Therapy Can Confer Benefit via Patient-to-Patient Variability without Drug Additivity or Synergy. Cell *171*, 1678-1691.e13.

227. Park, Y.B., Kim, H.S., Oh, J.H., and Lee, S.H. (2001). The co-expression of p53 protein and P-glycoprotein is correlated to a poor prognosis in osteosarcoma. Int. Orthop. *24*, 307–310.

228. Patnaik, A., Eckhardt, S.G., Izbicka, E., Tolcher, A.A., Hammond, L.A., Takimoto, C.H., Schwartz, G., McCreery, H., Goetz, A., Mori, M., et al. (2003). A Phase I, Pharmacokinetic, and Biological Study of the Farnesyltransferase Inhibitor Tipifarnib in Combination with Gemcitabine in Patients with Advanced Malignancies. Clin. Cancer Res. *9*, 4761–4771.

229. Paz-Ares, L.G., Biesma, B., Heigener, D., von Pawel, J., Eisen, T., Bennouna, J., Zhang, L., Liao, M., Sun, Y., Gans, S., et al. (2012). Phase III, Randomized, Double-Blind, Placebo-Controlled Trial of Gemcitabine/Cisplatin Alone or With Sorafenib for the First-Line Treatment of Advanced, Nonsquamous Non–Small-Cell Lung Cancer. J. Clin. Oncol. *30*, 3084–3092.

230. Pemovska, T., Bigenzahn, J.W., and Superti-Furga, G. (2018). Recent advances in combinatorial drug screening and synergy scoring. Curr. Opin. Pharmacol. *42*, 102–110.

231. Peng, B., Hayes, M., Resta, D., Racine-Poon, A., Druker, B.J., Talpaz, M., Sawyers, C.L., Rosamilia, M., Ford, J., Lloyd, P., et al. (2004). Pharmacokinetics and Pharmacodynamics of Imatinib in a Phase I Trial With Chronic Myeloid Leukemia Patients. J. Clin. Oncol. *22*, 935–942.

232. Perilongo, G., Maibach, R., Shafford, E., Brugieres, L., Brock, P., Morland, B., de Camargo, B., Zsiros, J., Roebuck, D., Zimmermann, A., et al. (2009). Cisplatin versus Cisplatin plus Doxorubicin for Standard-Risk Hepatoblastoma. N. Engl. J. Med. *361*, 1662–1670.

233. Pfisterer, J., Weber, B., Reuss, A., Kimmig, R., du Bois, A., Wagner, U., Bourgeois, H., Meier, W., Costa, S., Blohmer, J.-U., et al. (2006). Randomized Phase III Trial of Topotecan Following Carboplatin and Paclitaxel in First-line Treatment of Advanced Ovarian Cancer: A Gynecologic Cancer Intergroup Trial of the AGO-OVAR and GINECO. JNCI J. Natl. Cancer Inst. *98*, 1036–1045.

234. Polley, E., Kunkel, M., Evans, D., Silvers, T., Delosh, R., Laudeman, J., Ogle, C., Reinhart, R., Selby, M., Connelly, J., et al. (2016). Small Cell Lung Cancer Screen of Oncology Drugs, Investigational Agents, and Gene and microRNA Expression. JNCI J. Natl. Cancer Inst. *108*, djw122.

235. Prakash, S., Foster, B.J., Meyer, M., Wozniak, A., Heilbrun, L.K., Flaherty, L., Zalupski, M., Radulovic, L., Valdivieso, M., and LoRusso, P.M. (2001). Chronic Oral Administration of CI-994: A Phase I Study. Invest. New Drugs *19*, 1–11.

236. Puglisi, M., van Doorn, L., Blanco-Codesido, M., De Jonge, M.J., Moran, K., Yang, J., Busman, T., Franklin, C., Mabry, M., Krivoshik, A., et al. (2011). A phase I safety and pharmacokinetic (PK) study of navitoclax (N) in combination with docetaxel (D) in patients (pts) with solid tumors. J. Clin. Oncol. *29*, 2518–2518.

237. Qiu, W., Chavarro, J., Lazarus, R., Rosner, B., and Ma, J. (2015). powerSurvEpi: Power and Sample Size Calculation for Survival Analysis of Epidemiological Studies. R package version 0.0.9. https://CRAN.R-project.org/package=powerSurvEpi.

238. R Core Team (2017). R: A Language and Environment for Statistical Computing. R package version 3.4.2. https://www.R-project.org/.

239. Rajkumar, P., Mathew, B.S., Das, S., Isaiah, R., John, S., Prabha, R., and Fleming, D.H. (2016). Cisplatin Concentrations in Long and Short Duration Infusion: Implications for the Optimal Time of Radiation Delivery. J. Clin. Diagn. Res. *10*, XC01–XC04.

240. Ramalingam, S.S., Parise, R.A., Ramananthan, R.K., Lagattuta, T.F., Musguire, L.A., Stoller, R.G., Potter, D.M., Argiris, A.E., Zwiebel, J.A., Egorin, M.J., et al. (2007). Phase I and Pharmacokinetic Study of Vorinostat, A Histone Deacetylase Inhibitor, in Combination with Carboplatin and Paclitaxel for Advanced Solid Malignancies. Clin. Cancer Res. *13*, 3605–3610.

241. Reigner, B., Blesch, K., and Weidekamm, E. (2001). Clinical Pharmacokinetics of Capecitabine. Clin. Pharmacokinet. *40*, 85–104.

242. Richardson, P.G., Chanan-Khan, A.A., Alsina, M., Albitar, M., Berman, D., Messina, M., Mitsiades, C.S., and Anderson, K.C. (2010). Tanespimycin monotherapy in relapsed multiple myeloma: results of a phase 1 dose-escalation study. Br. J. Haematol. *150*, 438–445.

243. Ritz, C., Baty, F., Streibig, J.C., and Gerhard, D. (2015). Dose-Response Analysis Using R. PLoS ONE *10*, e0146021.

244. Robertson, J.F.R., Erikstein, B., Osborne, K.C., Pippen, J., Come, S.E., Parker, L.M., Gertler, S., Harrison, M.P., and Clarke, D.A. (2004). Pharmacokinetic Profile of Intramuscular Fulvestrant in Advanced Breast Cancer. Clin. Pharmacokinet. *43*, 529–538.

245. Rodon, J., Curigliano, G., Delord, J.-P., Harb, W., Azaro, A., Han, Y., Wilke, C., Donnet, V., Sellami, D., and Beck, T. (2018). A Phase Ib, open-label, dose-finding study of alpelisib in combination with paclitaxel in patients with advanced solid tumors. Oncotarget *9*, 31709–31718.

246. Rosiñol, L., Oriol, A., Teruel, A.I., Hernández, D., López-Jiménez, J., Rubia, J. de la, Granell, M., Besalduch, J., Palomera, L., González, Y., et al. (2012). Superiority of bortezomib, thalidomide, and dexamethasone (VTD) as induction pretransplantation therapy in multiple myeloma: a randomized phase 3 PETHEMA/GEM study. Blood *120*, 1589–1596.

247. RStudio Team (2015). RStudio: Integrated Development for R. R package version 1.1.463. http://www.rstudio.com/.

248. Rugo, H.S., Herbst, R.S., Liu, G., Park, J.W., Kies, M.S., Steinfeldt, H.M., Pithavala, Y.K., Reich, S.D., Freddo, J.L., and Wilding, G. (2005). Phase I Trial of the Oral

Antiangiogenesis Agent AG-013736 in Patients With Advanced Solid Tumors: Pharmacokinetic and Clinical Results. J. Clin. Oncol. *23*, 5474–5483.

249. Saito, T., and Rehmsmeier, M. (2017). Precrec: fast and accurate precision–recall and ROC curve calculations in R. Bioinformatics *33*, 145–147.

250. Saka, H., Kitagawa, C., Kogure, Y., Takahashi, Y., Fujikawa, K., Sagawa, T., Iwasa, S., Takahashi, N., Fukao, T., Tchinou, C., et al. (2017). Safety, tolerability and pharmacokinetics of the fibroblast growth factor receptor inhibitor AZD4547 in Japanese patients with advanced solid tumours: a Phase I study. Invest. New Drugs *35*, 451–462.

251. Salem, A.H., Koenig, D., and Carlson, D. (2014). Pooled Population Pharmacokinetic Analysis of Phase I, II and III Studies of Linifanib in Cancer Patients. Clin. Pharmacokinet. *53*, 347–359.

252. Sandmaier, B.M., Khaled, S., Oran, B., Gammon, G., Trone, D., and Frankfurt, O. (2018). Results of a phase 1 study of quizartinib as maintenance therapy in subjects with acute myeloid leukemia in remission following allogeneic hematopoietic stem cell transplant. Am. J. Hematol. *93*, 222–231.

253. San-Miguel, J.F., Hungria, V.T.M., Yoon, S.-S., Beksac, M., Dimopoulos, M.A., Elghandour, A., Jedrzejczak, W.W., Günther, A., Nakorn, T.N., Siritanaratkul, N., et al. (2014). Panobinostat plus bortezomib and dexamethasone versus placebo plus bortezomib and dexamethasone in patients with relapsed or relapsed and refractory multiple myeloma: a multicentre, randomised, double-blind phase 3 trial. Lancet Oncol. *15*, 1195–1206.

254. Savelieva, M., Woo, M.M., Schran, H., Mu, S., Nedelman, J., and Capdeville, R. (2015). Population pharmacokinetics of intravenous and oral panobinostat in patients with hematologic and solid tumors. Eur. J. Clin. Pharmacol. *71*, 663–672.

255. Scagliotti, G., Novello, S., von Pawel, J., Reck, M., Pereira, J.R., Thomas, M., Abrão Miziara, J.E., Balint, B., De Marinis, F., Keller, A., et al. (2010). Phase III Study of Carboplatin and Paclitaxel Alone or With Sorafenib in Advanced Non–Small-Cell Lung Cancer. J. Clin. Oncol. *28*, 1835–1842.

256. Scherf, U., Ross, D.T., Waltham, M., Smith, L.H., Lee, J.K., Tanabe, L., Kohn, K.W., Reinhold, W.C., Myers, T.G., Andrews, D.T., et al. (2000). A gene expression database for the molecular pharmacology of cancer. Nat. Genet. *24*, 236–244.

257. Schilcher, R.B., Young, J.D., Ratanatharathorn, V., Karanes, C., and Baker, L.H. (1984). Clinical pharmacokinetics of high-dose mitomycin C. Cancer Chemother. Pharmacol. *13*, 186–190.

258. Schweizer, M.T., Haugk, K., McKiernan, J.S., Gulati, R., Cheng, H.H., Maes, J.L., Dumpit, R.F., Nelson, P.S., Montgomery, B., McCune, J.S., et al. (2018). A phase I study of niclosamide in combination with enzalutamide in men with castration-resistant prostate cancer. PLoS ONE *13*, e0198389.

259. Seashore-Ludlow, B., Rees, M.G., Cheah, J.H., Cokol, M., Price, E.V., Coletti, M.E., Jones, V., Bodycombe, N.E., Soule, C.K., Gould, J., et al. (2015). Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. Cancer Discov. *5*, 1210–1223.

260. Segal, E., Friedman, N., Koller, D., and Regev, A. (2004). A module map showing conditional activity of expression modules in cancer. Nat. Genet. *36*, 1090.

261. Sekine, I., Yamamoto, N., Nishio, K., and Saijo, N. (2008). Emerging ethnic differences in lung cancer therapy. Br. J. Cancer *99*, 1757–1762.

262. Selden, R., Smith, T.W., and Findley, W. (1972). Ouabain Pharmacokinetics in Dog and Man. Circulation *45*, 1176–1182.

263. Sethi, V.S., Jackson, D.V., White, D.R., Richards, F., Stuart, J.J., Muss, H.B., Cooper, M.R., and Spurr, C.L. (1981). Pharmacokinetics of Vincristine Sulfate in Adult Cancer Patients. Cancer Res. *41*, 3551–3555.

264. Shapiro, G.I., Kristeleit, R.S., Burris, H.A., LoRusso, P., Patel, M.R., Drew, Y., Giordano, H., Maloney, L., Watkins, S., Goble, S., et al. (2019). Pharmacokinetic Study of Rucaparib in Patients With Advanced Solid Tumors. Clin. Pharmacol. Drug Dev. *8*, 107–118.

265. Shiah, H.-S., Chao, Y., Chen, L.-T., Yao, T.-J., Huang, J.-D., Chang, J.-Y., Chen, P.-J., Chuang, T.-R., Chin, Y.-H., Whang-Peng, J., et al. (2006). Phase I and pharmacokinetic study of oral thalidomide in patients with advanced hepatocellular carcinoma. Cancer Chemother. Pharmacol. *58*, 654–664.

266. Shirao, K., Matsumura, Y., Yamada, Y., Muro, K., Gotoh, M., Boku, N., Ohtsu, A., Nagashima, F., Sano, Y., Mutoh, M., et al. (2006). Phase I Study of Single-Dose Oxaliplatin in Japanese Patients with Malignant Tumors. Jpn. J. Clin. Oncol. *36*, 295–300.

267. Shoemaker, R.H. (2006). The NCI60 human tumour cell line anticancer drug screen. Nat. Rev. Cancer *6*, 813–823.

268. Siegel, R.L., Miller, K.D., and Jemal, A. (2017). Cancer statistics, 2017. CA. Cancer J. Clin. *67*, 7–30.

269. Siegel, R.L., Miller, K.D., and Jemal, A. (2019). Cancer statistics, 2019. CA. Cancer J. Clin. *69*, 7–34.

270. Sikma, M.A., van Maarseveen, E.M., van de Graaf, E.A., Kirkels, J.H., Verhaar, M.C., Donker, D.W., Kesecioglu, J., and Meulenbelt, J. (2015). Pharmacokinetics and Toxicity of Tacrolimus Early After Heart and Lung Transplantation: Tacrolimus Pharmacokinetics Posttransplant. Am. J. Transplant. *15*, 2301–2313.

271. Silvennoinen, R., Malminiemi, K., Malminiemi, O., Seppälä, E., and Vilpo, J. (2000). Pharmacokinetics of chlorambucil in patients with chronic lymphocytic leukaemia: comparison of different days, cycles and doses. Pharmacol. Toxicol. *87*, 223–228.

272. Simon, G.R., Garrett, C.R., Olson, S.C., Langevin, M., Eiseman, I.A., Mahany, J.J., Williams, C.C., Lush, R., Daud, A., Munster, P., et al. (2006). Increased Bioavailability of Intravenous Versus Oral CI-1033, a Pan erbB Tyrosine Kinase Inhibitor: Results of a Phase I Pharmacokinetic Study. Clin. Cancer Res. *12*, 4645–4651.

273. Skipper, H.E., Thomson, J.R., and Bell, M. (1954). Attempts at Dual Blocking of Biochemical Events in Cancer Chemotherapy. Cancer Res. *14*, 503–507.

274. Solymos, P., and Zawadzki, Z. (2017). pbapply: Adding Progress Bar to "*apply" Functions. R package version 1.3.3. https://CRAN.R-project.org/package=pbapply.

275. Speth, P.A.J., Linssen, P.C.M., Holdrinet, R.S.G., and Haanen, C. (1987). Plasma and cellular Adriamycin concentrations in patients with myeloma treated with ninety-six-hour continuous infusion. Clin. Pharmacol. Ther. *41*, 661–665.

276. Speyer, J.L., Green, M.D., Dubin, N., Blum, R.H., Wernz, J.C., Roses, D., Sanger, J., and Muggia, F.M. (1985). Prospective evaluation of cardiotoxicity during a six-hour doxorubicin infusion regimen in women with adenocarcinoma of the breast. Am. J. Med. *78*, 555–563.

277. Stamelos, V.A., Fisher, N., Bamrah, H., Voisey, C., Price, J.C., Farrell, W.E., Redman, C.W., and Richardson, A. (2016). The BH3 Mimetic Obatoclax Accumulates in Lysosomes and Causes Their Alkalinization. PLoS ONE *11*, e0150696.

278. Steele, N.L., Plumb, J.A., Vidal, L., Tjørnelund, J., Knoblauch, P., Buhl-Jensen, P., Molife, R., Brown, R., de Bono, J.S., and Evans, T.R.J. (2011). Pharmacokinetic and pharmacodynamic properties of an oral formulation of the histone deacetylase inhibitor Belinostat (PXD101). Cancer Chemother. Pharmacol. *67*, 1273–1279.

279. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550.

280. Sun, J.X., Niecestro, R., Phillips, G., Shen, J., Lukacsko, P., and Friedhoff, L. (2002). Comparative pharmacokinetics of lovastatin extended-release tablets and lovastatin immediate-release tablets in humans. J. Clin. Pharmacol. *42*, 198–204.

281. Swain, S.M., Tang, G., Geyer, C.E., Rastogi, P., Atkins, J.N., Donnellan, P.P., Fehrenbacher, L., Azar, C.A., Robidoux, A., Polikoff, J.A., et al. (2013). Definitive Results of a Phase III Adjuvant Trial Comparing Three Chemotherapy Regimens in Women With Operable, Node-Positive Breast Cancer: The NSABP B-38 Trial. J. Clin. Oncol. *31*, 3197–3204.

282. Tamura, K., Mukai, H., Naito, Y., Yonemori, K., Kodaira, M., Tanabe, Y., Yamamoto, N., Osera, S., Sasaki, M., Mori, Y., et al. (2016). Phase I study of palbociclib, a cyclin-dependent kinase 4/6 inhibitor, in Japanese patients. Cancer Sci. *107*, 755–763.

283. Tan, N., Malek, M., Zha, J., Yue, P., Kassees, R., Berry, L., Fairbrother, W.J., Sampath, D., and Belmont, L.D. (2011). Navitoclax Enhances the Efficacy of Taxanes in Non–Small Cell Lung Cancer Models. Clin. Cancer Res. *17*, 1394–1404.

284. Tanaka, C., Yin, O.Q.P., Sethuraman, V., Smith, T., Wang, X., Grouss, K., Kantarjian, H., Giles, F., Ottmann, O.G., Galitz, L., et al. (2010). Clinical Pharmacokinetics of the BCR–ABL Tyrosine Kinase Inhibitor Nilotinib. Clin. Pharmacol. Ther. *87*, 197–203.

285. Tang, J., Wennerberg, K., and Aittokallio, T. (2015). What is synergy? The Saariselkä agreement revisited. Front. Pharmacol. *6*.

286. Tchekmedyian, N.S., Egorin, M.J., Cohen, B.E., Kaplan, R.S., Poplin, E., and Aisner, J. (1986). Phase I clinical and pharmacokinetic study of cyclophosphamide administered by five-day continuous intravenous infusion. Cancer Chemother. Pharmacol. *18*, 33–38.

287. Teicher, B.A., Polley, E., Kunkel, M., Evans, D., Silvers, T., Delosh, R., Laudeman, J., Ogle, C., Reinhart, R., Selby, M., et al. (2015). Sarcoma Cell Line Screen of Oncology Drugs and Investigational Agents Identifies Patterns Associated with Gene and microRNA Expression. Mol. Cancer Ther. *14*, 2452–2462.

288. Terret, C., Erdociain, E., Guimbaud, R., Boisdron-Celle, M., McLeod, H.L., Féty-Deporte, R., Lafont, T., Gamelin, E., Bugat, R., Canal, P., et al. (2000). Dose and time dependencies of 5-fluorouracil pharmacokinetics. Clin. Pharmacol. Ther. *68*, 270–279.

289. The ICON and AGO Collaborators (2003). Paclitaxel plus platinum-based chemotherapy versus conventional platinum-based chemotherapy in women with relapsed ovarian cancer: the ICON4/AGO-OVAR-2.2 trial. The Lancet *361*, 2099–2106.

290. The ICON Group (2002). Paclitaxel plus carboplatin versus standard chemotherapy with either single-agent carboplatin or cyclophosphamide, doxorubicin, and cisplatin in women with ovarian cancer: the ICON3 randomised trial. Lancet Lond. Engl. *360*, 505–515.

291. Tomkinson, H., Kemp, J., Oliver, S., Swaisland, H., Taboada, M., and Morris, T. (2011). Pharmacokinetics and tolerability of zibotentan (ZD4054) in subjects with hepatic or renal impairment: two open-label comparative studies. BMC Clin. Pharmacol. *11*, 3.

292. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564-576.e16.

293. Tsimberidou, A.M., Said, R., Culotta, K., Wistuba, I., Jelinek, J., Fu, S., Falchook, G., Naing, A., Piha-Paul, S., Zinner, R., et al. (2015). Phase I study of azacitidine and

oxaliplatin in patients with advanced cancers that have relapsed or are refractory to any platinum therapy. Clin. Epigenetics *7*, 29.

294. Tsukada, H., Yokoyama, A., Goto, K., Shinkai, T., Harada, M., Ando, M., Shibata, T., Ohe, Y., Tamura, T., and Saijo, N. (2015). Randomized controlled trial comparing docetaxel–cisplatin combination with weekly docetaxel alone in elderly patients with advanced non-small-cell lung cancer: Japan Clinical Oncology Group (JCOG) 0207. Jpn. J. Clin. Oncol. *45*, 88–95.

295. Tsutsumi, H., Kumakawa, T., Hirai, M., Kikukawa, M., Arie, Y., Mori, M., Kodo, H., Nakamura, N., Murai, Y., and Mizutani, R. (1995). [Plasma concentration of cytosine arabinoside (Ara-C) in the elderly patients with hematological malignancy treated by Ara-C or cytarabine ocfosfate (SPAC)]. Nihon Ronen Igakkai Zasshi Jpn. J. Geriatr. *32*, 190–194.

296. Turesson, I., Bjorkholm, M., Blimark, C.H., Kristinsson, S., Velez, R., and Landgren, O. (2018). Rapidly changing myeloma epidemiology in the general population: Increased incidence, older patients, and longer survival. Eur. J. Haematol. *101*, 237–244.

297. Tzelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., Metzakopian, E., Dovey, O.M., Mupo, A., Grinkevich, V., Li, M., Mazan, M., et al. (2016). A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep. *17*, 1193–1205.

298. Valle, J., Wasan, H., Palmer, D.H., Cunningham, D., Anthoney, A., Maraveyas, A., Madhusudan, S., Iveson, T., Hughes, S., Pereira, S.P., et al. (2010). Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. N. Engl. J. Med. *362*, 1273–1281.

299. Van Cutsem, E., Moiseyenko, V.M., Tjulandin, S., Majlis, A., Constenla, M., Boni, C., Rodrigues, A., Fodor, M., Chao, Y., Voznyi, E., et al. (2006). Phase III Study of Docetaxel and Cisplatin Plus Fluorouracil Compared With Cisplatin and Fluorouracil As First-Line Therapy for Advanced Gastric Cancer: A Report of the V325 Study Group. J. Clin. Oncol. *24*, 4991–4997.

300. Van Cutsem, E., Labianca, R., Bodoky, G., Barone, C., Aranda, E., Nordlinger, B., Topham, C., Tabernero, J., André, T., Sobrero, A.F., et al. (2009). Randomized Phase III Trial Comparing Biweekly Infusional Fluorouracil/Leucovorin Alone or With Irinotecan in the Adjuvant Treatment of Stage III Colon Cancer: PETACC-3. J. Clin. Oncol. *27*, 3117–3125.

301. Van Veggel, M., Westerman, E., and Hamberg, P. (2018). Clinical Pharmacokinetics and Pharmacodynamics of Panobinostat. Clin. Pharmacokinet. *57*, 21–29.

302. van de Wetering, M., Francies, H.E., Francis, J.M., Bounova, G., Iorio, F., Pronk, A., van Houdt, W., van Gorp, J., Taylor-Weiner, A., Kester, L., et al. (2015). Prospective Derivation of a Living Organoid Biobank of Colorectal Cancer Patients. Cell *161*, 933–945.

303. Venditti, J.M., Humphreys, S.R., Mantel, N., and Goldin, A. (1956). Combined Treatment of Advanced Leukemia (L1210) in Mice with Amethopterin and 6-Mercaptopurine. JNCI J. Natl. Cancer Inst. *17*, 631–638.

304. Vermorken, J.B., Remenar, E., van Herpen, C., Gorlia, T., Mesia, R., Degardin, M., Stewart, J.S., Jelic, S., Betka, J., Preiss, J.H., et al. (2007). Cisplatin, Fluorouracil, and Docetaxel in Unresectable Head and Neck Cancer. N. Engl. J. Med. *357*, 1695–1704.

305. Verstovsek, S., Yeleswaram, S., Hou, K., Chen, X., and Erickson-Viitanen, S. (2018). Sustained-release ruxolitinib: Findings from a phase 1 study in healthy subjects and a phase 2 study in patients with myelofibrosis. Hematol. Oncol. *36*, 701–708.

306. Vlahovic, G., Karantza, V., Wang, D., Cosgrove, D., Rudersdorf, N., Yang, J., Xiong, H., Busman, T., and Mabry, M. (2014). A phase I safety and pharmacokinetic study of ABT-263 in combination with carboplatin/paclitaxel in the treatment of patients with solid tumors. Invest. New Drugs *32*, 976–984.

307. Vlot, A.H.C., Aniceto, N., Menden, M.P., Ulrich-Merzenich, G., and Bender, A. (2019). Applying drug synergy metrics to oncology combination screening data: agreements, disagreements and pitfalls. Drug Discov. Today.

308. Von Hoff, D.D., Ervin, T., Arena, F.P., Chiorean, E.G., Infante, J., Moore, M., Seay, T., Tjulandin, S.A., Ma, W.W., Saleh, M.N., et al. (2013). Increased Survival in Pancreatic Cancer with nab-Paclitaxel plus Gemcitabine. N. Engl. J. Med. *369*, 1691–1703.

309. Wada, T., Fukuda, T., Kawanishi, M., Tasaka, R., Imai, K., Yamauchi, M., Kasai, M., Hashiguchi, Y., Ichimura, T., Yasui, T., et al. (2016). Pharmacokinetic analyses of carboplatin in a patient with cancer of the fallopian tubes undergoing hemodialysis: A case report. Biomed. Rep. *5*, 199–202.

310. Wang, T., Yu, H., Hughes, N.W., Liu, B., Kendirli, A., Klein, K., Chen, W.W., Lander, E.S., and Sabatini, D.M. (2017). Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. Cell *168*, 890-903.e15.

311. Wang, X., Roy, A., Hochhaus, A., Kantarjian, H.M., Chen, T.-T., and Shah, N.P. (2013). Differential effects of dosing regimen on the safety and efficacy of dasatinib: retrospective exposure–response analysis of a Phase III study. Clin. Pharmacol. Adv. Appl. *5*, 85–97.

312. Weinstein, Z.B., Bender, A., and Cokol, M. (2017). Prediction of synergistic drug combinations. Curr. Opin. Syst. Biol. *4*, 24–28.

313. Wickham, H. (2016). rvest: Easily Harvest (Scrape) Web Pages. R package version 0.3.2. https://CRAN.R-project.org/package=rvest.

314. Wickham, H., and Bryan, J. (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl.

315. Wickham, H., Hester, J., and Francois, R. (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr.

316. Williams, S.P., and McDermott, U. (2017). The Pursuit of Therapeutic Biomarkers with High-Throughput Cancer Cell Drug Screens. Cell Chem. Biol. *24*, 1066–1074.

317. Willis, B.A., Zhang, W., Ayan-Oshodi, M., Lowe, S.L., Annes, W.F., Sirois, P.J., Friedrich, S., and Peña, A. de la (2012). Semagacestat Pharmacokinetics Are Not Significantly Affected by Formulation, Food, or Time of Dosing in Healthy Participants. J. Clin. Pharmacol. *52*, 904–913.

318. Wilson, W.H., O'Connor, O.A., Czuczman, M.S., LaCasce, A.S., Gerecitano, J.F., Leonard, J.P., Tulpule, A., Dunleavy, K., Xiong, H., Chiu, Y.-L., et al. (2010). Safety, Pharmacokinetics, Pharmacodynamics, and Activity of Navitoclax, a Targeted High Affinity Inhibitor of BCL-2, in Lymphoid Malignancies. Lancet Oncol. *11*, 1149–1159.

319. Wind, S., Schmid, M., Erhardt, J., Goeldner, R.-G., and Stopfer, P. (2013). Pharmacokinetics of Afatinib, a Selective Irreversible ErbB Family Blocker, in Patients with Advanced Solid Tumours. Clin. Pharmacokinet. *52*, 1101–1109.

320. Wong, C.H., Siah, K.W., and Lo, A.W. (2018). Estimation of clinical trial success rates and related parameters. Biostatistics *20*, 237–286.

321. Wooten, D.J., Meyer, C.T., Quaranta, V., and Lopez, C. (2019). A Consensus Framework Unifies Multi-Drug Synergy Metrics. BioRxiv 683433.

322. Wu, Y.-L., Lee, J.S., Thongprasert, S., Yu, C.-J., Zhang, L., Ladrera, G., Srimuninnimit, V., Sriuranpong, V., Sandoval-Tan, J., Zhu, Y., et al. (2013). Intercalated combination of chemotherapy and erlotinib for patients with advanced stage non-small-cell lung cancer (FASTACT-2): a randomised, double-blind trial. Lancet Oncol. *14*, 777–786.

323. Xin, Y., Shao, L., Maltzman, J., Stefanidis, D., Hemenway, J., Tarnowski, T., Deng, W., and Silverman, J.A. (2018). The Relative Bioavailability, Food Effect, and Drug Interaction With Omeprazole of Momelotinib Tablet Formulation in Healthy Subjects. Clin. Pharmacol. Drug Dev. *7*, 277–286.

324. Xue, C., Hong, S., Li, N., Feng, W., Jia, J., Peng, J., Lin, D., Cao, X., Wang, S., Zhang, W., et al. (2015). Randomized, Multicenter Study of Gefitinib Dose-escalation in Advanced Non-small-cell Lung Cancer Patients Achieved Stable Disease after One-month Gefitinib Treatment. Sci. Rep. *5*, 10648.

325. Yamamoto, N., Tamura, T., Yamamoto, N., Yamada, K., Yamada, Y., Nokihara, H., Fujiwara, Y., Takahashi, T., Murakami, H., Boku, N., et al. (2009). Phase I, dose escalation and pharmacokinetic study of cediranib (RECENTIN^{TM}), a highly potent and selective VEGFR signaling inhibitor, in Japanese patients with advanced solid tumors. Cancer Chemother. Pharmacol. *64*, 1165–1172.

326. Yamamoto, N., Nokihara, H., Yamada, Y., Goto, Y., Tanioka, M., Shibata, T., Yamada, K., Asahina, H., Kawata, T., Shi, X., et al. (2012). A Phase I, dose-finding and pharmacokinetic study of olaparib (AZD2281) in Japanese patients with advanced solid tumors. Cancer Sci. *103*, 504–509.

327. Yamamoto, N., Murakami, H., Hayashi, H., Fujisaka, Y., Hirashima, T., Takeda, K., Satouchi, M., Miyoshi, K., Akinaga, S., Takahashi, T., et al. (2013). CYP2C19 genotype-based phase I studies of a c-Met inhibitor tivantinib in combination with erlotinib, in advanced/metastatic non-small cell lung cancer. Br. J. Cancer *109*, 2803–2809.

328. Yamazaki, N., Tsutsumida, A., Takahashi, A., Namikawa, K., Yoshikawa, S., Fujiwara, Y., Kondo, S., Mukaiyama, A., Zhang, F., and Kiyohara, Y. (2018). Phase 1/2 study assessing the safety and efficacy of dabrafenib and trametinib combination therapy in Japanese patients with BRAF V600 mutation-positive advanced cutaneous melanoma. J. Dermatol. *45*, 397–407.

329. Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. *41*, D955–D961.

330. Yen, C.-J., Kim, T.-Y., Feng, Y.-H., Chao, Y., Lin, D.-Y., Ryoo, B.-Y., Huang, D.C.-L., Schnell, D., Hocke, J., Loembé, A.-B., et al. (2018). A Phase I/Randomized Phase II Study to Evaluate the Safety, Pharmacokinetics, and Efficacy of Nintedanib versus Sorafenib in Asian Patients with Advanced Hepatocellular Carcinoma. Liver Cancer *7*, 165–178.

331. Yoshioka, H., Azuma, K., Yamamoto, N., Takahashi, T., Nishio, M., Katakami, N., Ahn, M.J., Hirashima, T., Maemondo, M., Kim, S.W., et al. (2015). A randomized, double-blind, placebo-controlled, phase III trial of erlotinib with or without a c-Met inhibitor tivantinib (ARQ 197) in Asian patients with previously treated stage IIIB/IV nonsquamous nonsmall-cell lung cancer harboring wild-type epidermal growth factor receptor (ATTENTION study). Ann. Oncol. *26*, 2066–2072.

332. Yu, C., Mannan, A.M., Yvone, G.M., Ross, K.N., Zhang, Y.-L., Marton, M.A., Taylor, B.R., Crenshaw, A., Gould, J.Z., Tamayo, P., et al. (2016). High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. Nat. Biotechnol. *34*, 419–423.

333. Zasadil, L.M., Andersen, K.A., Yeum, D., Rocque, G.B., Wilke, L.G., Tevaarwerk, A.J., Raines, R.T., Burkard, M.E., and Weaver, B.A. (2014). Cytotoxicity of paclitaxel in breast cancer is due to chromosome missegregation on multipolar spindles. Sci. Transl. Med. *6*, 229ra43.

334. Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. J. Stat. Softw. *11*, 1–17.

335. Zeileis, A. (2006). Object-oriented Computation of Sandwich Estimators. J. Stat. Softw. *16*, 1–16.

336. Zhang, L., Li, S., Zhang, Y., Zhan, J., Zou, B.-Y., Smith, R., Martin, P.D., Jiang, Y., Liao, H., and Guan, Z. (2011). Pharmacokinetics and Tolerability of Vandetanib in Chinese Patients With Solid, Malignant Tumors: An Open-Label, Phase I, Rising Multiple-Dose Study. Clin. Ther. *33*, 315–327.

337. Zhang, M., Xu, C.R., Shamiyeh, E., Liu, F., Yin, J.Y., von Moltke, L.L., and Smith, W.B. (2014). A randomized, placebo-controlled study of the pharmacokinetics, pharmacodynamics, and tolerability of the oral JAK2 inhibitor fedratinib (SAR302503) in healthy volunteers. J. Clin. Pharmacol. *54*, 415–421.

APPENDIX I: README FILES FOR CHAPTER 3 ANALYSIS CODE

APPENDIX PREFACE

Please note that the code for the analyses performed in chapter 3 of this thesis and the readme files describing that code were written with the intention of the work being presented in a primary research article. As such, the figure and table numbers represented in the code on OSF and the readme files below are not consistent with the figure and table numbers used in this thesis. A mapping of the readme paper figure/table numbers to the same figures and tables in the thesis is provided in Table 4 below.

Identifier in Thesis	Identifier in OSF code and readme files
Table 2	Table S4
Figure 7	Figure 1
Figure 8	Figure 2
Figure 9	Figure 3
Figure 10	Figure 4
Figure 11	Figure S1
Table S3	Table S2
Table S4	Table S1
Figure 12	Figure 5
Table 3	Table 1
Figure 13	Figure S2
Figure 14	Figure S3
Figure 15	Figure S4
Figure 16	Figure S5
Figure 17	Figure 6
Figure 18	Figure S6
Table S5	Table S3
Figure 19	Figure 7

Table 4. Mapping of table and figure identifiers in this paper to their respective identifiers in the OSF repository.

NCI-ALMANAC ANALYSIS README

This folder contains the data and scripts used to perform IDACombo model validation against NCI-ALMANAC by predicting combination efficacy using the monotherapy data from NCI-ALMANAC and then comparing the predictions to the measured combination efficacies also present in NCI-ALMANAC.

The analysis was performed in two steps:

1. Combination efficacies were predicted using IDACombo and NCI-ALMANAC's monotherapy information, and the measured combination efficacies in NCI-ALMANAC were averaged across all cell lines for each combination.

2. The results of step 1 were plotted to create Figure 3: Agreement between predicted and observed combination efficacy in NCI-ALMANAC.

Step 1: IDACombo Predictions and Observed Efficacy Summary

Code:

1. "NCI-ALMANAC Analysis.R": Predict combination efficacy using NCI-ALMANAC monotherapy data and summarize measured combination efficacies.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. parallel v3.4.2: R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

2. pbapply v1.3.3: Peter Solymos and Zygmunt Zawadzki (2017). pbapply: Adding Progress Bar to '*apply' Functions. R package version 1.3-3. https://CRAN.R-project.org/package=pbapply

3. IDACombo: Chapter 3.

Input Files:

1. "./Inputs/ComboDrugGrowth_Nov2017.csv": NCI-ALMANAC data. Downloaded from "https://wiki.nci.nih.gov/download/attachments/338237347/ComboDrugGrowth_Nov2017.zip?version=1&modificationDate=1510057275000&api=v2" on 5/17/2019. This link was from "https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-ALMANAC".

2. "./Inputs/ComboCompoundNames_small.txt": List of drugs and drug identifiers used in NCI-ALMANAC.

Output Files:

1. "./Outputs/ALMANAC_Results.txt": Predicted and observed combination efficacies for NCI-ALMANAC combinations.

Step 2: Creating Figure 3: Agreement between predicted and observed combination efficacy in NCI-ALMANAC

Code:

1. "Making Figure 3_Agreement between predicted and observed combination efficacy in NCI-ALMANAC.R": Creates plots for Figure 3.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. None.

Input Files:

1. "./Outputs/ALMANAC_Results.txt": See step 1.

2. "./Inputs/ComboCompoundNames_small.txt": List of drugs and drug identifiers used in NCI-ALMANAC.

Output Files:

1. "./Outputs/Figure_3_Agreement between predicted and observed combination efficacy in NCI-ALMANAC.tiff": Figure 3.

HARMONIZING CTRPV2 AND GDSC README

README for "Harmonizing Datasets" code.

This code was written to harmonize the cell line and compound names between GDSC (2016) and CTRPv2.

Code:

1. "./Harmonizing_Datasets.R": adds harmonized drug and cell line identifiers to CTRPv2 and GDSC raw data.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr

2. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl

Input data:

1. "./Inputs/Table S2_Screened Cell Line Info_Ling et al_2018.xlsx" and "./Inputs/Table S3_Screened Drug Info_Ling et al_2018.xlsx" were obtained from Ling et al., 2018: Ling, A., Gruener, R.F., Fessler, J., and Huang, R.S. (2018). More than fishing for a cure: The promises and pitfalls of high throughput cancer cell line screens. Pharmacol. Ther.

2. "./Inputs/v17a_public_raw_data.xlsx" downloaded on 2/26/2018: ftp://ftp.sanger.ac.uk/pub4/cancerrxgene/releases/release-6.0/v17a_public_raw_data.xlsx

3. "./Inputs/GDSC_Screened_Compounds.xlsx" downloaded on 2/26/2018: ftp://ftp.sanger.ac.uk/pub4/cancerrxgene/releases/release-6.0/Screened_Compounds.xlsx

4. "./Inputs/GDSC_fitted_drug_data.xlsx" downloaded on 2/27/2018: ftp://ftp.sanger.ac.uk/pub4/cancerrxgene/releases/release-6.0/v17_fitted_dose_response.xlsx

5. "./Inputs/raw_GDSC.txt" is produced by the script "Harmonizing_Datasets.R" by removing some comment characters the first time the script is run (lines 22 and 23).

6. "v20.data.curves_post_qc.txt", "v20.data.per_cpd_well.txt", "v20.meta.per_cell_line.txt", "v20.meta.per_compound.txt", and "v20.meta.per_experiment.txt" were downloaded on 2/26/2018: ftp://anonymous:guest@caftpd.nci.nih.gov/pub/OCG-DCC/CTD2/Broad/CTRPv2.0_2015_ctd2_ExpandedDataset/CTRPv2.0_2015_ctd2_ExpandedDataset.zip

Output Data:

1. "./Outputs/Harmonized_raw_GDSC": raw GDSC drug response data with harmonized identifiers

2. "./Outputs/Harmonized_raw_CTRP": raw CTRPv2 drug response data with harmonized identifiers

3. "./Outputs/CTRP_Compound_Meta.txt": compound metadata for CTRPv2 with original and harmonized drug identifiers

4. "./Outputs/Harmonized_GDSC_Screened_Compounds.txt": list of GDSC compounds with original and harmonized drug identifiers

REPROCESSING RAW CTRPV2 AND GDSC DATA README

This folder contains the data and scripts used to reprocess the raw dose-response data from CTRPv2 and GDSC. The process was carried out in two steps. Note that the "Recalculating IC50s" file names are poorly chosen, as the scripts are actually fitting dose reponse curves and not simply calculating IC50s.

The first step was to fit dose-respones curves to the raw data using the drc package (Ritz et al., 2015). It was performed using the Mesabi compute cluster at the Minnesota Supercomputing Institute (https://www.msi.umn.edu). Files for this step are contained in the "Inputs", "MSI_Cluster_Script", and "MSI_Outputs" folders.

Code:

1. "./MSI_Outputs/Recalculating IC50s.R": Primary code for fitting dose-response curves to raw CTRPv2 and GDSC data

2. "./MSI_Outputs/job_Recalculate_GDSC_CTRP_Preds": job file to establish job parameters for scheduler

3. "./MSI_Outputs/Submit_Recalc_Jobs.sh": shell script used to initiate several jobs at once, allowing the process to be divided into multiple jobs that could run simultaneously.

Coding Environment:

R version 3.4.4

Mesabi compute cluster, Minnesota Supercomputing Institute (https://www.msi.umn.edu)

4 cores, single node

28 gb memory

Packages

1. drc v3.0.1: Ritz, C., Baty, F., Streibig, J. C., Gerhard, D. (2015) Dose-Response Analysis Using R PLOS ONE, 10(12), e0146021

2. progress v1.1.2: Gábor Csárdi and Rich FitzJohn (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress

3. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr

Input Files:

1. "./Inputs/Harmonized_raw_GDSC.txt": From "../Harmonizing CTRPv2 and GDSC/"

2. "./Inputs/Harmonized_raw_CTRP.txt": From "../Harmonizing CTRPv2 and GDSC/"

3. "./Inputs/v20.meta.per_assay_plate.txt" and "./Inputs/v20.meta.per_experiment.txt": downloaded on 2/26/2018 from: ftp://anonymous:guest@caftpd.nci.nih.gov/pub/OCG-DCC/CTD2/Broad/CTRPv2.0_2015_ctd2_ExpandedDataset/CTRPv2.0_2015_ctd2_ExpandedDataset.zip

Output Files:

1. pdf files of curve fits for each cell line and drug in CTRPv2 and GDSC. Organized by drug: ./MSI_Outputs/CurveFits/

2. rds files of drc fitted models: ./MSI_Outputs/FittedModels/

3. txt files of summary information for each fit, separated into clusters of information for ~100 drugs: ./MSI_Outputs/

The second step was to combine the txt data from the multiple Mesabi jobs into single output files as well as calculate AUC values based on the fitted model parameters.

Code:

1. "./Combining_MSI_Job_Outputs_and_Calculating_AUCs.R": Combines Mesabi txt outputs and calculates AUC values.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages:

1. progress v1.1.2: Gábor Csárdi and Rich FitzJohn (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress

2. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr

Input Files:

1. "./Inputs/Harmonized_raw_GDSC.txt": From "../Harmonizing CTRPv2 and GDSC/"

2. "./Inputs/Harmonized_raw_CTRP.txt": From "../Harmonizing CTRPv2 and GDSC/"

3. MSI_Output txt files: "./MSI_Outputs/*_Results_*_*.txt" created in first step.

Output Files:

1. "./Ouputs/Recalculated_CTRP_12_21_2018.txt"

2. "./Outputs/Recalculated_GDSC_12_21_2018.txt"

IDENTIFYING CLINICAL TRIALS README

README for "Identifying Clinical Trials" code.

This code was written to use webscraping to search clinicaltrials.gov and pubmed.gov for published phase 3 clinical trials that tested drug combinations for which we can make efficacy predictions using IDA-Combo and CTRPv2 or GDSC. The first step computationally identifies

papers from clinical trials that could be useful for our clinical trial validation analysis. In the second step, the identified papers are manually currated to identify trials that are useful for our IDA-Combo clinical validation analysis. Finally, a third step adds clinical plasma concentrations (Csustaioned 6 hours) to the trial information.

Step 1

Code:

1. "Gathering Clinical Trial Data.R": webscraping to search clinicaltrials.gov for phase 3 clinical trials that tested at least two drugs from CTRPv2 or GDSC for which we can make combination efficacy predictions, then identifies published papers from that trial and uses web scraping to search pubmed.gov for abstracts for those papers.

Notes:

The script was first run from line 1 to 31 (which downloads clinicaltrials.gov search records) and then lines 27-31 were commented out and a linux command line command (line 34) was used to concatenate the outputs of lines 27-31 (note that this command was run mannually outside of R). This concatenated output is loaded in line 37, and the script was subsequently completed.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages:

1. progress v1.1.2: Gábor Csárdi and Rich FitzJohn (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress
2. rvest v0.3.2: Hadley Wickham (2016). rvest: Easily Harvest (Scrape) Web Pages. R package version 0.3.2. https://CRAN.R-project.org/package=rvest

Input Files:

1. "./Inputs/CTRP_Results_Clin_Range_all_ccl_v2.tbd" and "./Inputs/GDSC_Results_Clin_Range_all_ccl_v2.tbd": These files are outputs from deprecated code which was originally used for the prospective analysis. They are only used to tell the script which drug combinations IDA-Combo can make predictions for using CTRPv2 and GDSC, so they were never updated--especially since we didn't want to go through an updated output from this script, which may add hundreds of papers to review and require re-running all downstream analyses. We have verified that these files contain all of the combinations included in the most recent version of the prospective analysis.

Output Files:

1. "./Outputs/Article.Info.P3&4_10_08_2018.txt": A list of all of the articles identified as potentially useful for our clinical trial IDA-Combo validation analysis. Includes abstracts for all papers.
2. "./Outputs/Completed.Info.P3&4_10_08_2018.txt: clinicaltrials.gov records for all identified completed phase 3 trials of potential utility for our analysis.

Step 2:

Manual curration of identified clinical trial publications. First, papers are judged for whether or not they meet the inclusion criteria based on abstracts (see "Article.Info.P3&4_10_08_2018_manual_updates.txt"). Next, papers that made it past the first step are downloaded in their entirety and a more thorough assesment is made for whether or not they meet the inclusion criteria (see "SelectedArticle.Info.P3&4_10_08_2018.xlsx"). Finally, detailed results from the selected papers are collated into a table for downstream analysis and a final determination of whether or not each trial meets the inclusion criteria is made (see "ClinicalTrialValidationInputs_10_08_2018_predose.xlsx").

Column Descriptors for "ClinicalTrialValidationInputs_10_08_2018_predose.xlsx"

A: NCT.Number: clinical trial identifier from clinicaltrials.gov for trial each paper is reporting results for

B: Citation: citation for papers

C: Authors: paper authors

D: Unique Study ID: our own assigned ID for each study

E:Title: paper title

F: Disease_Type: type of cancer studied in trial

G: Condensed_Cancer_Type: a more formatted version of Disease_Type which groups trials into broader cancer categories.

H:Skip_Study: a flag for whether or not the trial has met the final inclusion criteria. F for studies that meet the inclusion criteria, and T_* for trials that will be excluted from final analysis. Reason for exclusion is included.

I: Test in GDSC: Flag for whether or not predictions can be made for this trial using GDSC

J: Test in CTRP: Flag for whether or not predictions can be made for this trial using CTRPv2

K: Patients have undergone previous drug treatment: Information about whether or not patients in the trial had undergone drug treatment prior to entering the clinical trial.

L: Study_Type: Classification of what type of control group-test group comparison was being made in each trial.

M: Study Treatments: Information about treatments used in the trial

N: Control Drugs: List of drugs used in control group treatment for trial

O: Control Drug DoseKeys: Keys to allow "Add_Clinical_Concentrations_To_Clinical_Trials_Data.R" to automatically add control treatment drug plasma concentrations to the table in step 3.

P: Test Group Drugs: List of drugs used in the test group treatment for trial

Q: Test Group Drug DoseKeys: Keys to allow "Add_Clinical_Concentrations_To_Clinical_Trials_Data.R" to automatically add test treatment drug plasma concentrations to the table in step 3.

R: Specific_Outcomes: Early, crude summary of clinical trial outcomes

S/T: Control_OS (months)/Test_OS (months): Median Overall Survival (OS) times for control and test arms of trial in months.

U: OS_HR: OS hazard ratio (HR)

V: OS Significantly Improved in Test: Whether or not study reported statistically significant improvement in OS in test group vs control group ($p > 0.05$).

W-AT: Similar to OS columns (S-V) except for progression free survival (PFS), relapse free survival (RFS), time to progression (TTP), event free survival (EFS), disease free survival (DFS), or failure free survival (FFS).

AU/AV: Control_Objective_Response (%)/Test_Objective_Response (%): objective response rates in control arm and test arm respectively in percent.

AX: Objective Response Significantly Improved in Test: Whether or not objective response rate was reported to be statistically improved in test arm vs control arm ($p > 0.05$).

AY: Trial Considered a Success: Whether or not trial met primary criteria for trial success.

Az: General_Outcome: Summary of trial outcome

BA: Specific Population: Information about whether or not recorded results are for specific subgroup of cancer type

BB: n_Patients: number of patients in trial

BC: n_Test_Patients: number of patients in test arm

BD: n_Control_Patients: number of patients in control arm

BE-BF: n_Test_OS_Events/n_Control_OS_Events: Number of OS events obeserved in trial in test arm or control arm respectively

BG: Total_OS_Events: Number of total OS events observed in trial

BH-BY: Similar to BE-BG but for PFS, DFS, EFS, TTP, RFS, or FFS.

BZ-CC: Stage_*_Percent: Percentage of patients in study who had stage I, II, III, or IV cancer.

CD: Event Notes: Note for whether or not event numbers are exact or were estimated based on plotted Kaplan-Meier curves. Note, estimated numbers will tend to be overestimated as it is difficult to account for all censored individuals.

CE: Data Entry Complete: Check for whether data entry for that study is completed in this table. Note that this will be marked complete even with missing information if trial failed to meet inclusion criteria (i.e. see "Skip_Study" column H.

Step 3:

Code:

1. "Add_Clinical_Concentrations_To_Clinical_Trials_Data.R": A script that adds clinical trial plasma concentrations (Csustained 6 hours) to clinical trial data.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages:

1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
2. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl

Input Files:

1. "./Outputs/ClinicalTrialValidationInputs_10_08_2018_predose.xlsx": The clinical trial information currated in Step 2.
2. "./Inputs/Clinical Trial Drug Plasma Concentrations.xlsx": A list of clinical plasma concentrations for each drug at a given dose. Manually assembled. See "Source" column for paper from which each concentration is derived. Csustained is median maximum plasma drug concentration measured at least 6 hours after drug is administered to patients.

Output Files:

1. "ClinicalTrialValidationInputs_10_08_2018.RDS": R data file of clinical trial information with drug concentrations. Save as R data file instead of text file to avoid rounding errors for drug concentrations.

CLINICAL TRIAL VALIDATION ANALYSIS README

This folder contains the data and scripts used to perform the clinical trial analysis presented in the paper, where monotherapy data from CTRPv2 and GDSC was used to predict the efficacy of control and test treatments in the phase 3 clinical trials identified using the data and scripts in the "./Identifying Clinical Trials/" folder. This analysis pertains to Figures 5, S1, S2, S3, and S4 and Tables 1 and S2.

The analysis was performed in several steps:

1. Viability estimates were calculated at clinical concentrations for each drug to be used in the analysis
2. Drug combination efficacy predictions were made for each treatment in each trial based on Independent Drug Action using IDACombo.
3. Analysis summary figures and tables were created for the main clinical trial analysis. Figures 5, S2, and S3 and Tables 1 and S2.
4. Viability estimates were recalculated at more concentrations to allow for analysis of how perturbing drug concentrations would affect clinical trial validation results.

5. Drug combination efficacy predictions were remade for each treatment in each trial using perturbed drug concentrations. Predictions still based on Independent Drug Action using IDACombo.
6. Figure S4 created based on concentration perturbation analysis.
7. Drug combination efficacy predictions were remade for each treatment in each trial based on Bliss Independence using IDACombo.
8. Figure S5 was created based on Bliss Independence results.

Step 1: Viability calculations for clinical drug concentrations

Code:

1. "CalculateViabilityAndUncertaintyForTrialConcentrations.R": Calculate viability and viability uncertainty estimates from fitted curves for CTRPv2 and GDSC using clinical drug concentrations.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. drc v3.0.1: Ritz, C., Baty, F., Streibig, J. C., Gerhard, D. (2015) Dose-Response Analysis Using R PLOS ONE, 10(12), e0146021
2. progress v1.1.2: Gábor Csárdi and Rich FitzJohn (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress
3. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
4. sandwich v2.4.0: Achim Zeileis (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. Journal of Statistical Software 11(10), 1-17. URL http://www.jstatsoft.org/v11/i10/. AND Achim Zeileis (2006). Object-Oriented Computation of Sandwich Estimators. Journal of Statistical Software 16(9), 1-16. URL http://www.jstatsoft.org/v16/i09/.

Input Files:

1. "./Inputs/ClinicalTrialValidationInputs_10_08_2018.RDS": List of clinical trial results, treatments, and drug concentrations created using "../Identifying

Clinical Trials/Add_Clinical_Concentrations_To_Clinical_Trials_Data.R". (see "../Identifying Clinical Trials/README.txt")

2. "./Inputs/CTRP_GDSC_tested_compound_doses.txt": List of maximum and minimum drug concentrations tested for each drug in CTRPv2 and/or GDSC. Note, some drugs were tested at different dose ranges depending on the cell line. The minimums and maximums recorded in this file are median values based on the minimum and maximum concentrations used for each cell line. For some compounds, a handful of cell lines may have been tested with lower or higher concentrations than the majority of cell lines. These values are based on the concentration range used for the majority of cell lines.
3. "../Reprocessing raw CTRPv2 and GDSC data/Outputs/Recalculated_GDSC_12_21_2018.txt": Summary of reprocessed GDSC dose-response curves produced by the code in the "../Reprocessing raw CTRPv2 and GDSC data/" folder.
4. "C:/Users/Lingy/Desktop/Huang Lab/Drug Combo Project/Reprocessing raw CTRPv2 and GDSC data/Outputs/Recalculated_CTRP_12_21_2018.txt": : Summary of reprocessed CTRPv2 dose-response curves produced by the code in the "../Reprocessing raw CTRPv2 and GDSC data/" folder.
5. "../Reprocessing raw CTRPv2 and GDSC data/MSI_Outputs/FittedModels/*": Fitted dose-response models for each drug in CTRPv2 and GDSC.
6. "./Outputs/PFS_TTP_no_Chemo_Figure_5_Data_ratio_cutoff_2.rds": Information for clinical trials plotted in Figure 5A
7. "./Outputs/OS_no_Chemo_Figure_5_Data_ratio_cutoff_2.rds": Information for clinical trials plotted in Figure 5B

Output Files:

1. "./Outputs/Clinical_Trial_Info.rds": An updated version of "./Inputs/ClinicalTrialValidationInputs_10_08_2018.RDS" with additional information added. Saved as data frame R object in rds format.
2. "./Outputs/CTRP_Viabilities_and_Uncertainties.rds": Calculated CTRPv2 viabilities at defined concentrations for each drug/cell line pair. Saved as data frame R object in rds format.
3. "./Outputs/GDSC_Viabilities_and_Uncertainties.rds": Calculated GDSC viabilities at defined concentrations for each drug/cell line pair. Saved as data frame R object in rds format.

<u>Step 2: Drug combination efficacy predictions based on Independent Drug Action</u>

Code:

1. "Clinical_Trial_Validation.R": Uses viaibilites from Step 1 and IDACombo to predict efficacy of test and control treatments in the clinical trials. Then calculates hazard ratios and powers for each trial based on those predictions.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. IDACombo: Chapter 3.
2. progress v1.1.2: Gábor Csárdi and Rich FitzJohn (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress
    i. 3. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
3. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl
4. xlsx v0.5.7: Adrian A. Dragulescu (2014). xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files. R package version 0.5.7. https://CRAN.R-project.org/package=xlsx
5. powerSurvEpi v0.0.9: Weiliang Qiu, Jorge Chavarro, Ross Lazarus, Bernard Rosner and Jing Ma. (2015). powerSurvEpi: Power and Sample Size Calculation for Survival Analysis of Epidemiological Studies. R package version 0.0.9. https://CRAN.R-project.org/package=powerSurvEpi

Input Files:

1. "./Outputs/Clinical_Trial_Info.rds": See step 1 outputs.
2. "./Outputs/CTRP_Viabilities_and_Uncertainties.rds": See step 1 outputs.
3. "./Outputs/GDSC_Viabilities_and_Uncertainties.rds": See step 1 outputs.
4. "./Inputs/Table S2_Screened Cell Line Info_plus_BRCA_Info.xlsx": A manually modified version of Table S2 from Ling et al., 2018: Ling, A., Gruener, R.F., Fessler, J., and Huang, R.S. (2018). More than fishing for a cure: The promises and pitfalls of high throughput cancer cell line screens. Pharmacol. Ther. This version of the file is modified to identify cell lines that harbor EGFR, HER2, PR, and/or ER abberations.

Output Files:

1. "./Outputs/Clinical_Trial_Predictions.rds": Clinical trial information with added efficacy predictions. Saved as list R object in RDS format.
2. "./Outputs/Clinical_Trial_Predictions.xlsx": Clinical trial information with added efficacy predictions. Saved as xlsx format to be made into Table S2.

Step 3: Summary figures and tables for main clinical trial validation analysis

Step 3 part 1:

Code:

1. "Create Figure 5_Clinical Trial Validation Results.R": Creates Figure 5, which is a summary of clinical trial efficacy predictions made using Independent Drug Action and all available cell lines.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr

Input Files:

1. "./Outputs/Clinical_Trial_Predictions.rds": See step 2 outputs.

Output Files:

1. "./Outputs/Figures and Tables/Figure_5_Clinical trial validation results_ratio_cutoff_2.tiff": Figure 5.

Step 3 part 2:

Code:

1. "Create Figure S2_Cancer Specific Clinical Trial Validation.R": Creates Figure S2, which is a summary of clinical trial efficacy

predictions made using Independent Drug Action and cancer specific sets of cell lines.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr

Input Files:

1. "./Outputs/Clinical_Trial_Predictions.rds": See step 2 outputs.

Output Files:

1. "./Outputs/Figures and Tables/Figure_S2_Cancer Specific Clinical trial validation results_ratio_cutoff_ratio_cutoff_2.tiff": Figure S2.

Step 3 part 3:

Code:

1. "Create Figure S3_Clinical conentrations beyond tested range.R": Creates Figure S3, which is an analysis of how making predictions with concentrations beyond the tested concentrations ranges in CTRPv2 or GDSC affects model performance.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr

Input Files:

1. "./Outputs/Clinical_Trial_Predictions.rds": See step 2 outputs.

Output Files:

1. "./Outputs/Figures and Tables/Figure_S3_Clinical concentrations beyond tested range.tiff": Figure S3.

Step 3 part 4:

Code:

1. "Create Table 1_Subtype Trials.R": Gets data for Table 1, which is a summary of predictions made in trials that reported full results for patients with a given molecular subtype of cancer.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr

Input Files:

1. "./Outputs/Clinical_Trial_Predictions.rds": See step 2 outputs.

Output Files:

1. "./Outputs/Figures and Tables/Wu_2013_Table.txt": Data for Table 1 from Wu et al., 2013. (Lancet Oncol. 2013 Jul;14(8):777-86. doi: 10.1016/S1470-2045(13)70254-7. Epub 2013 Jun 17.).

2. "./Outputs/Figures and Tables/DiLeo_2008_Table.txt": Data for Table 1 from Di Leo et al., 2008 (J Clin Oncol. 2008 Dec 1;26(34):5544-52. doi: 10.1200/JCO.2008.16.2578. Epub 2008 Oct 27.).

Step 4: Viability calculations for perturbed clinical drug concentrations

Code:

1. "CalculateViabilityAndUncertaintyForDosePerturbationTrialConcentrations.R": Calculate viability and viability uncertainty estimates from fitted curves for CTRPv2 and GDSC using perturbed clinical drug concentrations.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. drc v3.0.1: Ritz, C., Baty, F., Streibig, J. C., Gerhard, D. (2015) Dose-Response Analysis Using R PLOS ONE, 10(12), e0146021
2. progress v1.1.2: Gábor Csárdi and Rich FitzJohn (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress
2. 3. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
3. 4. sandwich v2.4.0: Achim Zeileis (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. Journal of Statistical Software 11(10), 1-17. URL http://www.jstatsoft.org/v11/i10/. AND Achim Zeileis (2006). Object-Oriented Computation of Sandwich Estimators. Journal of Statistical Software 16(9), 1-16. URL http://www.jstatsoft.org/v16/i09/.

Input Files:

1. "./Inputs/ClinicalTrialValidationInputs_10_08_2018.RDS": List of clinical trial results, treatments, and drug concentrations created using "../Identifying Clinical Trials/Add_Clinical_Concentrations_To_Clinical_Trials_Data.R". (see "../Identifying Clinical Trials/README.txt")
2. "./Inputs/CTRP_GDSC_tested_compound_doses.txt": List of maximum and minimum drug concentrations tested for each drug in CTRPv2 and/or GDSC.

Note, some drugs were tested at different dose ranges depending on the cell line. The minimums and maximums recorded in this file are median values based on the minimum and maximum concentrations used for each cell line. For some compounds, a handful of cell lines may have been tested with lower or higher concentrations than the majority of cell lines. These values are based on the concentration range used for the majority of cell lines.

3. "../Reprocessing raw CTRPv2 and GDSC data/Outputs/Recalculated_GDSC_12_21_2018.txt": Summary of reprocessed GDSC dose-response curves produced by the code in the "../Reprocessing raw CTRPv2 and GDSC data/" folder.
4. "C:/Users/Lingy/Desktop/Huang Lab/Drug Combo Project/Reprocessing raw CTRPv2 and GDSC data/Outputs/Recalculated_CTRP_12_21_2018.txt": : Summary of reprocessed CTRPv2 dose-response curves produced by the code in the "../Reprocessing raw CTRPv2 and GDSC data/" folder.
5. "../Reprocessing raw CTRPv2 and GDSC data/MSI_Outputs/FittedModels/*": Fitted dose-response models for each drug in CTRPv2 and GDSC.

Output Files:

1. "./Outputs/Clinical_Trial_Info_Dose_Perturbation.rds": An updated version of "./Inputs/ClinicalTrialValidationInputs_10_08_2018.RDS" with additional information added. Saved as data frame R object in rds format.
2. "./Outputs/Dose_Perturbation_Dose_Multipliers.rds": A list of the dose multipliers used to perturb drug concentrations. Saved as data frame R object in rds format.
3. "./Outputs/CTRP_Viabilities_and_Uncertainties_Dose_Perturbation.rds": Calculated CTRPv2 viabilities at perturbed concentrations for each drug/cell line pair.
4. "./Outputs/GDSC_Viabilities_and_Uncertainties_Dose_Perturbation.rds": Calculated GDSC viabilities at perturbed concentrations for each drug/cell line pair.

<u>Step 5: Drug combination efficacy predictions based on Independent Drug Action with perturbed drug concentrations</u>

Code:

1. "Dose_Perturbation_Analysis.R": Uses viaibilites from Step 4 and IDACombo to predict efficacy of test and control treatments in the clinical trials using perturbed drug concentrations. Then calculates hazard ratios and powers for each trial based on those predictions.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. IDACombo: Chapter 3.
2. progress v1.1.2: Gábor Csárdi and Rich FitzJohn (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress
2. 3. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
3. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl
4. xlsx v0.5.7: Adrian A. Dragulescu (2014). xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files. R package version 0.5.7. https://CRAN.R-project.org/package=xlsx
5. powerSurvEpi v0.0.9: Weiliang Qiu, Jorge Chavarro, Ross Lazarus, Bernard Rosner and Jing Ma. (2015). powerSurvEpi: Power and Sample Size Calculation for Survival Analysis of Epidemiological Studies. R package version 0.0.9. https://CRAN.R-project.org/package=powerSurvEpi

Input Files:

1. "./Outputs/Clinical_Trial_Info.rds": See step 1 outputs.
2. "./Outputs/CTRP_Viabilities_and_Uncertainties.rds": See step 1 outputs.
3. "./Outputs/GDSC_Viabilities_and_Uncertainties.rds": See step 1 outputs.
4. "./Inputs/Table S2_Screened Cell Line Info_plus_BRCA_Info.xlsx": A manually modified version of Table S2 from Ling et al., 2018: Ling, A., Gruener, R.F., Fessler, J., and Huang, R.S. (2018). More than fishing for a cure: The promises and pitfalls of high throughput cancer cell line screens. Pharmacol. Ther. This version of the file is modified to identify cell lines that harbor EGFR, HER2, PR, and/or ER abberations.

Output Files:

1. "./Outputs/OS_Dose_Perturbation_Results.rds": Information about OS model performance when concentrations are multiplied by a range of multiplication factors. Saved in rds format.

2. "./Outputs/PFS_TTP_Dose_Perturbation_Results.rds": Information about PFS/TTP model performance when concentrations are multiplied by a range of multiplication factors. Saved in rds format.
3. "./Outputs/PFS_TTP_Max_Tested_Predictions.rds": Clinical trial information and efficacy predictions when efficacy is predicted using maximum tested concentrations for each drug. Only includes trials plotted in Figure 5A.
4. "./Outputs/OS_Max_Tested_Predictions.rds": Clinical trial information and efficacy predictions when efficacy is predicted using maximum tested concentrations for each drug. Only includes trials plotted in Figure 5B.

Step 6: Summary figure for dose perturbation analysis

Code:

1. "Create Figure S4_Dose perturbation results.R": Creates Figure S4, which is a summary of dose perturbation analysis.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. 1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr

Input Files:

1. "./Outputs/OS_Max_Tested_Predictions.rds": See step 5 outputs.
2. "./Outputs/PFS_TTP_Max_Tested_Predictions.rds": See step 5 outputs.
3. "./Outputs/PFS_TTP_Dose_Perturbation_Results.rds": See step 5 outputs.
4. "./Outputs/OS_Dose_Perturbation_Results.rds": See step 5 outputs.

Output Files:

1. "./Outputs/Figures and Tables/Figure_S4_Clinical power predictions are dose dependent.tiff": Figure S4.

Step 7: Drug combination efficacy predictions based on Bliss Independence

Code:

1. "Clinical_Trial_Validation_Bliss_Independence.R": Uses viaibilites from Step 1 and IDACombo to predict efficacy of test and control treatments in the clinical trials. Then calculates hazard ratios and powers for each trial based on those predictions.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. IDACombo: Chapter 3.
2. progress v1.1.2: Gábor Csárdi and Rich FitzJohn (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress
2. 3. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
3. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl
4. xlsx v0.5.7: Adrian A. Dragulescu (2014). xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files. R package version 0.5.7. https://CRAN.R-project.org/package=xlsx
5. powerSurvEpi v0.0.9: Weiliang Qiu, Jorge Chavarro, Ross Lazarus, Bernard Rosner and Jing Ma. (2015). powerSurvEpi: Power and Sample Size Calculation for Survival Analysis of Epidemiological Studies. R package version 0.0.9. https://CRAN.R-project.org/package=powerSurvEpi

Input Files:

1. "./Outputs/Clinical_Trial_Info.rds": See step 1 outputs.
2. "./Outputs/CTRP_Viabilities_and_Uncertainties.rds": See step 1 outputs.
3. "./Outputs/GDSC_Viabilities_and_Uncertainties.rds": See step 1 outputs.
4. "./Inputs/Table S2_Screened Cell Line Info_plus_BRCA_Info.xlsx": A manually modified version of Table S2 from Ling et al., 2018: Ling, A., Gruener, R.F., Fessler, J., and Huang, R.S. (2018). More than fishing for a

cure: The promises and pitfalls of high throughput cancer cell line screens. Pharmacol. Ther. This version of the file is modified to identify cell lines that harbor EGFR, HER2, PR, and/or ER abberations.

Output Files:

1. "./Outputs/Clinical_Trial_Predictions_Bliss_Independence.rds": Clinical trial information with added Bliss Independence efficacy predictions. Saved as list R object in RDS format.
2. "./Outputs/Clinical_Trial_Predictions_Bliss_Independence.xlsx": Clinical trial information with added Bliss Independence efficacy predictions. Saved as xlsx format for easy viewing.

Step 8: Summary figure for Bliss Independence predictions

Code:

1. "Create Figure S5_Bliss Independence Predictions.R": Creates Figure S5, which is a summary of efficacy predictions based on Bliss Independence.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr

Input Files:

1. "./Outputs/Clinical_Trial_Predictions_Bliss_Independence.rds": See step 7 outputs.

Output Files:

1. "./Outputs/Figures and Tables/Figure_S5_Bliss_Independence_ratio_cutoff_2.tiff": Figure S5.

PROSPECTIVE ANALYSIS README

This folder contains the data and scripts used to perform the prospective analysis presented in the paper, where monotherapy data from CTRPv2 and GDSC was used to predict the efficacy of 2-drug combinations of drugs that have reached late-stage clinical trials. This analysis pertains to figures 5, 6, and S6 and table S3.

The analysis was performed in several steps:

1. Viability estimates were calculated at specific concentrations for each drug to be used in the analysis
2. Drug combination efficacy predictions were made with IDACombo using all cell lines and using cell line sets specific to 27 cancer types/subtypes.
3. Analysis summary figures and tables were created for each cancer type (and the all cell line analysis). Note: these figures and tables are human readable and are stored in "./Outputs/Cluster_Heatmaps/".
4. Cell line downsampling analysis. Drug combination efficacy predictions were remade several times as in step 2, except using increasingly smaller numbers of cell lines in order to identify how agreement between GSDC and CTRPv2 predictions changed as fewer cell lines were available for predictions.
5. Figures and tables were created for use in the paper.

Step 1: Viability calculations

Code:

1. "CalculateViabilityAndUncertaintyFromClinicalConcentrations.R": Calculate viability and viability uncertainty estimates from fitted curves for CTRPv2 and GDSC.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. drc v3.0.1: Ritz, C., Baty, F., Streibig, J. C., Gerhard, D. (2015) Dose-Response Analysis Using R PLOS ONE, 10(12), e0146021
2. progress v1.1.2: Gábor Csárdi and Rich FitzJohn (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress
3. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
4. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl
5. sandwich v2.4.0: Achim Zeileis (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. Journal of Statistical Software 11(10), 1-17. URL http://www.jstatsoft.org/v11/i10/. AND Achim Zeileis (2006). Object-Oriented Computation of Sandwich Estimators. Journal of Statistical Software 16(9), 1-16. URL http://www.jstatsoft.org/v16/i09/.

Input Files:

1. "./Inputs/GDSC_CTRP_Clin_Drugs_cSustained.xlsx": Manually currated list of clinically relevant drug concentrations. Csustained is maximum plasma drug concentration achieved at least 6 hours after drug was administered to patient. See STAR Methods for details.
2. "./Inputs/CTRP_GDSC_tested_compound_doses.txt": List of maximum and minimum drug concentrations tested for each drug in CTRPv2 and/or GDSC. Note, some drugs were tested at different dose ranges depending on the cell line. The minimums and maximums recorded in this file are median values based on the minimum and maximum concentrations used for each cell line. For some compounds, a handful of cell lines may have been tested with lower or higher concentrations than the majority of cell lines. These values are based on the concentration range used for the majority of cell lines.
3. "../Reprocessing raw CTRPv2 and GDSC data/Outputs/Recalculated_GDSC_12_21_2018.txt": Summary of reprocessed GDSC dose-response curves produced by the code in the "../Reprocessing raw CTRPv2 and GDSC data/" folder.
4. "C:/Users/Lingy/Desktop/Huang Lab/Drug Combo Project/Reprocessing raw CTRPv2 and GDSC data/Outputs/Recalculated_CTRP_12_21_2018.txt": : Summary of reprocessed CTRPv2 dose-response curves produced by the code in the "../Reprocessing raw CTRPv2 and GDSC data/" folder.
5. "../Reprocessing raw CTRPv2 and GDSC data/MSI_Outputs/FittedModels/*": Fitted dose-response models for each drug in CTRPv2 and GDSC.

Output Files:

1. "./Inputs/Updated_ClinConc.rds": An updated version of "./Inputs/GDSC_CTRP_Clin_Drugs_cSustained.xlsx" with additional information added. Saved as data frame R object in RDS format.
2. "./Inputs/Recalculated_CTRP_Data_Viabilities.rds": Calculated CTRPv2 viabilities at defined concentrations for each drug/cell line pair in detailed format. Saved as data frame R object in RDS format.
3. "./Inputs/Recalculated_GDSC_Data_Viabilities.rds": Calculated GDSC viabilities at defined concentrations for each drug/cell line pair in detailed format. Saved as data frame R object in RDS format.
4. "./Inputs/Simple_Recalculated_CTRP_Data_Viabilities.rds": Calculated CTRPv2 viabilities at defined concentrations for each drug/cell line pair in simplified format for direct use with IDACombo. Saved as data frame R object in RDS format.
5. "./Inputs/Simple_Recalculated_GDSC_Data_Viabilities.rds": Calculated GDSC viabilities at defined concentrations for each drug/cell line pair in simplified format for direct use with IDACombo. Saved as data frame R object in RDS format.

Step 2: Drug combination efficacy predictions

Code:

1. "Predict_Drug_Combo_Efficacy_Tissue_Specific.R": Uses viaibilites from Step 1 and IDACombo to predict efficacy of 2-drug combinations of late-stage clinical drugs in CTRPv2 and GDSC. Makes cancer type/subtype specific predictions as well as predictions with all available cell lines.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. IDACombo: Chapter 3.
2. progress v1.1.2: Gábor Csárdi and Rich FitzJohn (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress

3. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
4. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl
5. pbapply v1.3.3: Peter Solymos and Zygmunt Zawadzki (2017). pbapply: Adding Progress Bar to '*apply' Functions. R package version 1.3-3. https://CRAN.R-project.org/package=pbapply
6. parallel v3.4.2: R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Input Files:

1. "./Inputs/Simple_Recalculated_CTRP_Data_Viabilities.rds": See step 1 outputs.
2. "./Inputs/Simple_Recalculated_GDSC_Data_Viabilities.rds": See step 1 outputs.
3. "./Inputs/Table S2_Screened Cell Line Info_plus_BRCA_Info.xlsx": A manually modified version of Table S2 from Ling et al., 2018: Ling, A., Gruener, R.F., Fessler, J., and Huang, R.S. (2018). More than fishing for a cure: The promises and pitfalls of high throughput cancer cell line screens. Pharmacol. Ther. This version of the file is modified to identify cell lines that harbor EGFR, HER2, PR, and/or ER abberations.

Output Files:

1. "./Outputs/CTRP_Combo_Prediction_Results_*.rds": CTRPv2 based drug combination efficacy predictions for 2-drug combinations of drugs that have reached late-stage clinical trials. Saved as list R object in RDS format.
2. "./Outputs/GDSC_Combo_Prediction_Results_*.rds": GDSC based drug combination efficacy predictions for 2-drug combinations of drugs that have reached late-stage clinical trials. Saved as list R object in RDS format.

Step 3: Summary figures and tables for combination efficacy predictions

Step 3 part 1:

Code:

1. "Making_Drug_Combo_Efficacy_Figures.R": Creates summary figures and tables for drug combination efficacy predictions created in step 2 using all available cell lines.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
2. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl
3. RColorBrewer v1.1.2: Erich Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer
4. ComplexHeatmap v1.14.0: Gu, Z. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics.

Input Files:

1. "./Inputs/Updated_ClinConc.rds": See step 1 outputs.
2. "./Outputs/CTRP_Combo_Prediction_Results_All_ccls.rds": See step 2 outputs.
3. "./Outputs/GDSC_Combo_Prediction_Results_All_ccls.rds": See step 2 outputs.
4. "./Inputs/Drug_Target_Info.xlsx": Manually currated list of drug targets for CTRPv2 and/or GDSC drugs that have reached late-stage clinical trials.

Output Files:

1. "./Outputs/Cluster_Heatmaps/": This script produces numerous summary figures and tables which are all contained within this folder. Specifically, this script populates the "./Outputs/Cluster_Heatmaps/All_ccls/" folder and adds two files to the "./Outputs/Cluster_Heatmaps/GDSC_CTRP_Correlation_Outputs/" folder. To shorten this README file, another README file has been created in the "./Outputs/Cluster_Heatmaps/" folder describing the files contained within that folder.

Step 3 part 2:

Code:

1. "Making_Drug_Combo_Efficacy_Figures_Tissue_Specific.R": Creates summary figures and tables for drug combination efficacy predictions created in step 2 using cancer type/subtype specific sets of cell lines. Also produces comparisons of cancer specific predictions to predictions made with all cell lines.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
2. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl
3. RColorBrewer v1.1.2: Erich Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer
4. ComplexHeatmap v1.14.0: Gu, Z. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics.

Input Files:

1. "./Inputs/Updated_ClinConc.rds": See step 1 outputs.
2. "./Outputs/Cluster_Heatmaps/All_ccls/CTRP_IDA_comboscore_matrix_All_ccls.rds": Step 3 part 1 output. See "./Outputs/Cluster_Heatmaps/README".
3. "./Outputs/Cluster_Heatmaps/All_ccls/GDSC_IDA_comboscore_matrix_All_ccls.rds": Step 3 part 1 output. See "./Outputs/Cluster_Heatmaps/README".

4. "./Outputs/Cluster_Heatmaps/All_ccls/CTRP_drug_targets_All_ccls.txt": Step 3 part 1 output. See "./Outputs/Cluster_Heatmaps/README".
5. "./Outputs/Cluster_Heatmaps/All_ccls/GDSC_drug_targets_All_ccls.txt": Step 3 part 1 output. See "./Outputs/Cluster_Heatmaps/README".
6. "./Inputs/Drug_Target_Info.xlsx": Manually currated list of drug targets for CTRPv2 and/or GDSC drugs that have reached late-stage clinical trials.
7. "./Outputs/CTRP_Combo_Prediction_Results_*.rds": See step 2 outputs.
8. "./Outputs/GDSC_Combo_Prediction_Results_*.rds": See step 2 outputs.

Output Files:

1. "./Outputs/Cluster_Heatmaps/": This script produces numerous summary figures and tables which are all contained within this folder. Specifically, this script populates the folders with cancer type/subtype specifc names and adds numerous files to the "./Outputs/Cluster_Heatmaps/GDSC_CTRP_Correlation_Outputs/" folder. To shorten this README file, another README file has been created in the "./Outputs/Cluster_Heatmaps/" folder describing the files contained within that folder.

Step 4: Cell line downsampling analysis

Code:

1. "Predict_Drug_Combo_Efficacy_Downsample_CCLs.R": Uses viabilities calculated in step 1 and IDACombo to predict drug combination efficacy for 2-drug combinations of drugs in CTRPv2 and GDSC that have reached late-stage clinical trials, but predictions are made using increasingly smaller numers of cell lines to determine how prediction agreement between CTRPv2 and GDSC is affected by the number of cell lines used to make the predictions.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. IDACombo: Chapter 3.
2. progress v1.1.2: Gábor Csárdi and Rich FitzJohn (2016). progress: Terminal Progress Bars. R package version 1.1.2. https://CRAN.R-project.org/package=progress
    i. 3. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
3. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl

Input Files:

1. "./Inputs/Simple_Recalculated_CTRP_Data_Viabilities.rds": See step 1 outputs.
2. "./Inputs/Simple_Recalculated_GDSC_Data_Viabilities.rds": See step 1 outputs.

Output Files:

1. "./Outputs/Downsample/CTRP_Combo_Prediction_Results*.rds": Prediction results based on different numbers of cell lines (note: predictions were made using each number of cell lines 5 times, randomly selecting the cell lines used each time). Saved as data frame R object in RDS format.

Step 5: Creating figures and tables for paper

Step 5 part 1:

Code:

1. "Making Figure 6_IDA_ComboScore predictions for late stage clinical drugs in CTRPv2.R": Creates figure 6.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
2. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl
3. RColorBrewer v1.1.2: Erich Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer
4. ComplexHeatmap v1.14.0: Gu, Z. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics.

Input Files:

1. "./Inputs/Updated_ClinConc.rds": See step 1 outputs.
2. "./Outputs/CTRP_Combo_Prediction_Results_All_ccls.rds": See step 2 outputs.
3. "./Outputs/GDSC_Combo_Prediction_Results_All_ccls.rds": See step 2 outputs.
4. "./Inputs/Drug_Target_Info.xlsx": Manually currated list of drug targets for CTRPv2 and/or GDSC drugs that have reached late-stage clinical trials.

Output Files:

1. "./Outputs/Figures_and_Tables/Figure_6_IDA_Comboscore_Predictions_for_late_stage_clinical_drugs_in_CTRPv2.tiff": Figure 6.

Step 5 part 2:

Code:

1. "Making Figure 7 Plots_IDA_Combo predicts strong benefits for combinations of navitoclax and taxanes.R": Creates plots for figure 7.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. readr v1.1.1: Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
2. readxl v1.0.0: Hadley Wickham and Jennifer Bryan (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl
3. rgl v0.98.1: Daniel Adler, Duncan Murdoch and others (2017). rgl: 3D Visualization Using OpenGL. R package version 0.98.1. https://CRAN.R-project.org/package=rgl
4. car v2.5.1: John Fox and Sanford Weisberg (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Input Files:

1. "./Outputs/CTRP_Combo_Prediction_Results_All_ccls.rds": See step 2 outputs.
2. "./Outputs/CTRP_Combo_Prediction_Results_EGFR_WT_lung_cancer.rds": See step 2 outputs.

Output Files:

1. "./Outputs/Figures_and_Tables/Figure_7A_Navitoclax_IDA_ComboScores_in_EGFR_WT_Lung_Cancer.tiff": Figure 7A.
2. "./Outputs/Figures_and_Tables/Navitoclax_Docetaxel_Viability.png": Figure 7B.
3. "./Outputs/Figures_and_Tables/Navitoclax_Paclitaxel_Viability.png": Figure 7C.

Step 5 part 3:

Code:

1. "Making Figure S6_GDSC_CTRP_Agreement_vs_n_Cell_Lines_and_Cancer_Types.R": Creates plots for figure S6.

Coding Environment:

R version 3.4.2 (2017-09-28) -- "Short Summer"

Platform: x86_64-w64-mingw32/x64 (64-bit)

Microsoft R Open 3.4.2

RStudio version 1.1.463

Packages

1. xlsx v0.5.7: Adrian A. Dragulescu (2014). xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files. R package version 0.5.7. https://CRAN.R-project.org/package=xlsx

Input Files:

1. "./Outputs/Downsample/*_Combo_Prediction_Results_*.rds": See step 4 outputs.
2. "./Outputs/Cluster_Heatmaps/GDSC_CTRP_Correlation_Outputs/*.txt": See step 3 outputs and "./Outputs/Cluster_Heatmaps/README".

Output Files:

1. "./Outputs/Figures_and_Tables/Figure_S6_GDSC_CTRP_Agreement_vs_n_Cell_Lines_and_Cancer_Type.tiff": Figure S6.
2. "./Outputs/Figures_and_Tables/Table_Sx_GDSC_CTRP_Agreement_vs_n_Cell_Lines_and_Cancer_Type.xlsx": Table S3.